

Programa de Doctorado en Biología Molecular y Celular Universidad Miguel Hernández

METAGENOMES AND GENOMES OF THE MEDITERRANEAN SEA PELAGIC MICROBIOTA

Tesis Doctoral presentada por el estudiante D. Jose Manuel Haro Moreno para optar al grado de Doctor.

Director: Francisco Eduardo Rodríguez Valera Codirector: Mario López Pérez

Instituto en Investigación, Desarrollo e Innovación en Biotecnología Sanitaria de Elche (iDiBE)

> San Juan de Alicante, España 2019



METAGENOMES AND GENOMES OF THE MEDITERRANEAN SEA PELAGIC MICROBIOTA

Trabajo realizado por el Licenciado Jose Manuel Haro Moreno, con DNI 74376824E, en el Grupo de Evolución Genómica (EGG) del Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández de Elche, para optar al grado de Doctor.

La siguiente Tesis se presenta como compendio de los siguientes trabajos previamente publicados:

- <u>Haro-Moreno J.M.</u>, López-Pérez M., de la Torre J.R., Picazo A., Camacho A. and Rodriguez-Valera F. (2018). Fine Metagenomic Profile of the Mediterranean Stratified and Mixed Water Columns Revealed by Assembly and Recruitment. Microbiome 6 (1): 128. doi:10.1186/s40168-018-0513-5.
- López-Pérez M., <u>Haro-Moreno J.M.</u>, de la Torre J.R. and Rodriguez-Valera F. (2018). Novel Caudovirales associated with Marine Group I Thaumarchaeota assembled from metagenomes. Environ Microbiol, 21:1980-1988. doi:10.1111/1462-2920.14462
- <u>Haro-Moreno J.M.</u>, Rodriguez-Valera F., López-García P., Moreira D. and Martin-Cuadrado A.B. (2017). New Insights into Marine Group III Euryarchaeota, from Dark to Light. The ISME Journal, 1–16. doi:10.1038/ismej.2016.188.

Adicionalmente, declaro que los artículos empleados como indicio de calidad para la presentación de la tesis doctoral en el marco del Programa de Doctorado en Biología Molecular y Celular del Instituto de Investigación, Desarrollo e Innovación en Biotecnología Sanitaria (IDiBE) de la Universidad Miguel Hernández de Elche, no han sido presentados ni serán presentados en ninguna otra tesis doctoral.

Durante la realización de esta Tesis, el doctorando fue financiado por las "Ayudas para contratos predoctorales para la formación de doctores 2014" (BES-2014-067828) dentro del Programa Estatal de Promoción del Talento y su empleabilidad en I+D+i, Ministerio de Economía y Competitividad. A su vez, las investigaciones recopiladas en este trabajo fueron financiadas por el Ministerio de Economía y Competitividad ("MEDIMAX" – BFPU2013-48007-P; "VIREVO" – CGL2016-76273-P; Acciones de dinamización "REDES DE EXCELENCIA CONSOLIDER" – CGL2015-71523-REDC), por la Generalitat Valenciana ("AQUAMET" – PROMETEO-II/2014/012) y por la Unión Europea ("MaCuMBA" Project 311975 of the European Commission FP7).





FRANCISCO EDUARDO RODRÍGUEZ VALERA, Catedrático de Microbiología de la Universidad Miguel Hernández de Elche

HAGO CONSTAR:

Que el presente trabajo titulado "METAGENOMES AND GENOMES OF THE MEDITERRANEAN SEA PELAGIC MICROBIOTA" ha sido realizado bajo mi dirección y recoge fielmente la labor desarrollada por el Licenciado Jose Manuel Haro Moreno para optar al grado de Doctor por la Universidad Miguel Hernández. Los trabajos reflejados en esta Tesis se han desarrollado íntegramente en el Departamento de Producción Vegetal y Microbiología de la Universidad Miguel Hernández de Elche.

Francisco Eduardo Rodríguez Valera

San Juan, 25 de julio de 2019.





MARIO LÓPEZ PÉREZ, Investigador Senior de la Universidad Miguel Hernández de Elche

HAGO CONSTAR:

Que el presente trabajo titulado "METAGENOMES AND GENOMES OF THE MEDITERRANEAN SEA PELAGIC MICROBIOTA" ha sido realizado bajo mi codirección y recoge fielmente la labor desarrollada por el Licenciado Jose Manuel Haro Moreno para optar al grado de Doctor por la Universidad Miguel Hernández. Los trabajos reflejados en esta Tesis se han desarrollado íntegramente en el Departamento de Producción Vegetal y Microbiología de la Universidad Miguel Hernández de Elche.

Mario López Pérez

San Juan, 25 de julio de 2019.



MH. Informe del Coordinador del programa de doctorado

UNIVERSITAS Miguel Hernández



Dr. Ricardo Mallavia Marin, Catedrático y Coordinador del Programa de Doctorado en Biología Molecular y Celular del Instituto de Investigación, Desarrollo e Innovación en Biotecnología Sanitaria (IDiBE) de la Universidad Miguel Hernández de Elche.

da su conformidad a la lectura de la Tesis Doctoral titulada "METAGENOMES AND GENOMES OF THE MEDITERRANEAN SEA PELAGIC MICROBIOTA", presentada por el Licenciado Jose Manuel Haro Moreno para optar al grado de Doctor.



Ricardo Mallavia Marín Coordinador del Programa de Doctorado en Biología Molecular y Celular.

Elche, 25 de julio de 2019







SUMMARY

The application of genomic tools, commonly known as metagenomics, has represented a revolution in the dawn of the XXI century Microbiology. It allows to directly study microbial assemblages in their habitat, without the isolation and culture of each one of the species present in the sample. Moreover, thanks to the widely used high-throughput sequencing, the assembly of multiple genomes obtained from metagenomes ("MAGs") have been achieved. The assembly of these MAGs allows a precise assignment of their phylogeny, distribution and, also, their ecology and metabolism. During the development of this Thesis, high-throughput sequencing has been used to study the prokaryotic community present in the off-shore water column of the Mediterranean Sea. The water column is thermally stratified during summer. However, during winter, the water column is mixed, triggering the upwelling of nutrients from mesopelagic waters. Six metagenomic samples were taken from the uppermost 100 m of the stratified (summer) water column at intervals of 15 m deep. Also, two more samples were taken during winter for comparison. By genome recruitment of the novel assembled MAGs and some reference microbes, results showed a marked distribution of microorganisms through the stratified water column. These microbes seemed to be found in layers of no more of 30-m. After comparing the two seasons, results showed a persistent prokaryotic community, although some microbes were to the season-associated, appearing only during summer or winter. Also, the analysis of rhodopsins indicated a sharp gradient of the copy number of this gene, being present in more than half of the prokaryotes in surface waters. From the pool of assembled contigs, a few fragments were detected with genes matching the genome of Ca. Nitrosopumilus, a marine archaeon belonging to the phylum Thaumarchaeota. Besides, in these contigs some common viral genes of the order Caudovirales were detected. Therefore, the second work included in this Thesis is the discovery of the first putative viruses capable of infecting this archaeal group. Lastly, in this Thesis is also included the discovery of eight novel MAGs of marine group III Euryarchaeota (MG-III), six of them collected from the photic zone, and two more from samples collected in the aphotic zone. Due to the fact that there are no cultured representatives of this group, the study of their genomes allowed to obtain information about their phylogeny and abundance in the ocean, indicating that they are ubiquitously distributed but in low numbers. Besides, from their genome, traits about their heterotrophic metabolism were attained.

RESUMEN

La aplicación de técnicas genómicas, comúnmente conocido como metagenómica, ha supuesto una revolución en la microbiología del siglo XXI debido a que permite el estudio directo de la comunidad microbiana en su entorno natural, sin la necesidad de aislar y cultivar cada una de las especies presentes en la muestra. Además, gracias a la secuenciación de alto rendimiento utilizada actualmente se ha conseguido con facilidad el ensamblaje de numerosos genomas obtenidos a partir de metagenomas ("MAG"), que permite asignar con una mayor precisión la filogenia del microorganismo, su distribución y su ecología a partir de su metabolismo. Para el desarrollo del primer trabajo de esta tesis se ha utilizado la secuenciación de alto rendimiento para el estudio de la microbiota presente en la columna de agua del mar Mediterráneo. La columna de agua de este hábitat se encuentra térmicamente estratificada durante los meses de verano. Sin embargo, en invierno la columna se homogeniza, con el consiguiente afloramiento de nutrientes provenientes de aguas mesopelágicas. Se tomaron seis muestras en los 100 primeros metros de la columna de agua estratificada (verano) y 2 muestras adicionales en invierno. Mediante el reclutamiento de los MAGs obtenidos y otros genomas de referencia, los resultados indicaron una marcada distribución de los microorganismos, encontrándose éstos delimitados en regiones dentro de la columna de agua de no más de 30 metros de anchura. Al comparar las dos estaciones, se observó que una gran parte de la población microbiana era persistente al cambio, mientras que la población restante era susceptible, apareciendo únicamente en una de las dos estaciones. Asimismo, el análisis de la rodopsina indicó un acusado gradiente del número de copias de este gen, resultando estar presente en más de la mitad de los procariotas en aguas superficiales. De la fracción de los cóntigos virales ensamblados de los metagenomas se detectó varios fragmentos con varias proteínas similares a aquellas codificadas en el genoma de Ca. Nitrosopumilus, una arquea marina perteneciente al filo Thaumarchaeota. Además, en estos cóntigos se detectaron genes típicos del orden viral Caudovirales. Por ello, en esta tesis se incluye como segundo trabajo el descubrimiento de los primeros virus detectados capaces de infectar a este grupo de arqueas. Por último, en esta tesis también incluye la obtención mediante ensamblaje de metagenomas de ocho nuevos representantes del grupo marino III Euryarchaeota, seis de ellos provenientes de muestras recogidas en la zona fótica y dos de la zona afótica. Debido a que en la actualidad no existe ningún representante cultivado, el estudio de sus genomas ha permitido obtener información sobre su filogenia y de su abundancia en los océanos y mares, indicando que se encuentran distribuidos ubicuamente pero que contribuyen, dentro de la comunidad microbiana, con una proporción relativamente pequeña. Además, el estudio de sus genes ha permitido obtener pistas sobre su metabolismo heterotrófico.

LIST OF ABREVIATIONS

		1	
AAI	Average Amino acid Identity	FISH	Fluorescence in situ hybridization
AMG	Auxiliary Metabolic Gene	Gb	Gigabases (10 ⁹)
amoC	Ammonia monooxygenase subunit C	GC (%)	Guanine-Citosine (%)
ANI	Average Nucleotide Identity	GH	Glycoside Hydrolases
AT (%)	Adenine-Thymine (%)	GOS	Global Ocean Sampling
ATP	Adenosine triphosphate	gyrA/B	DNA gyrase subunit a/b
Bathy	Bathypelagic	HDNA	High-content DNA
BLAST	Basic Local Alignment Search Tool	HGT	Horizontal Gene Transfer
bp	base-pair	HL	High Light
Ca.	Candidatus	HMM	Hidden Markov Models
CAZy	Carbohydrate-Active Enzymes	JGI	Joint Genome Institute
CCA	Canonical Correspondence Analysis	Kb	Kilobases (10 ³)
cDNA	complementary DNA	KEGG	Kyoto Encyclopedia of Genes and Genomes
CDS	Coding-DNA sequence	KM3	Metagenomic library 3000 m deep Mediterranean Sea
CG	Composite Genome	LDNA	Low-content DNA
1.0	adenosylcobinamide-GDP:alpha-		iecu -
cobS	ribazole ribazoletransferase	LL	Low Light
COG	Cluster of Orthologous Groups	LP	Lower Photic
	Conductivity, Temperature and		
CTD	Depth (pressure) measure instrument	LPS	Lipopolysaccharides
czcA	Cation efflux system	MAC	Marine actinobacterial clade
DCM	Deep Chlorophyll Maximum	MAG	Metagenome-assembled Genome
DMS	Dimethyl sulfide	Magrovirus	Marine Group II Euryarchaeota virus
DMSP	Dimethylsulfoniopropionate	Marthavirus	Marine Thaumarchaeota virus
DNA	Deoxyribonucleic acid	Mb	Megabases (10 ⁶)
dnaK	Chaperone Protein	MCP	Major Capsid Protein
DOM	Dissolved Organic Matter	MDA	Multiple-displacement Amplification
E-value	Expectation Value	MG	Marine Group
EDTA	Ethylenediaminetetraacetic acid	MicRhoDE	Microbial Rhodopsin Diversity and Evolution Database
EMP	Embden-Meyerhof-Parnas pathway	MIX	Mixed water column
Epi	Epipelagic	MUSCLE	Multiple Sequence Comparison by Log-Expectation
ETC	Electron Transport Chain	NCBI	NationalCenterforBiotechnologyInformation

NGS	Next-generation Sequencing	rRNA	Ribosomal RNA
NR	non-redundant database	SAG	Single-amplified Genome
PacBio	Pacific Biosciences	secY	Protein translocase
Pcb	Chlorophyll a/b-binding light- harvesting protein	SGC	Single-cell Genomics
PCR	Polymerase-chain reaction	SMRT	Single-Molecule Real-Time
PVC	Planctomycetes-Verrucomicrobia- Chlamydiae superphylum	SNP	Single nucleotide polymorphism
pVOG	Prokaryotic Virus Orthologous Groups	Tb	Terabases (10 ¹²)
radA	DNA repair protein	TCA	Tricarboxylic acid Cycle
RDP	Ribosomal Database Project	TOC	Total Organic Carbon
recA	DNA repair protein	Tris	tris(hydroxymethyl)aminomethane buffer
RNA	Ribonucleic acid	UP	Upper Photic
RNR	Ribonucleotide reductase	UV	Ultraviolet light
RPKG	Reads per Kilobase of Genome per Gigabase of Metagenome	w/v	weight/volume
rpoB	RNA polymerase subunit beta		









	Page
1 INTRODUCTION	number
1.1 Physical and chemical structure of the Ocean	1
1.1 1 Surface and deen water currents in the open occan	5 2
1.1.2 Ventical stratification of the water column	5
1.1.2 The Deep Chlorenhull Maximum (DCM)	5
1.1.5 The Deep Chiorophyli Maximum (DCM)	5
1.2 Marine meteoronomies	0
1.2 Marine metagenomics	8
1.2.1 First attempt: Metataxonomics	8
1.2.2 Second attempt: Shotgun metagenomics	10
1.2.3 Metagenome-assembled genomes (MAGs) and single- amplified genomes (SAGs)	11
1.2.4 Future perspectives: The Third-Generation Sequencing	14
1.3 Description of marine organisms	15
1.3.1 Common features of marine bacteria and archaea	15
1.3.2 Common marine prokaryotes in the water column	16
1.4 Marine Viruses	22
1.4.1 Isolation and metagenomic sequencing of marine viruses	22
1.4.2 Role of viruses in the environment and interaction with their	23
hosts	2.4
1.5 The Mediterranean Sea: a case study of the microbial diversity	24
1.5.1 Main features of the Mediterranean Sea: location, physicochemical properties and water currents	24
1.5.2 The water column in the Mediterranean Sea: water	26
stratification and seasonal dynamics	26
2. OBJECTIVES	29
3. MATERIAL AND METHODS	33
3.1 Sample collection, processing and sequencing	35
3.2 Metagenomic raw-read processing	36
3.3 Phylogenetic analysis of the 16S rRNA gene fragments derived from metagenomic reads	37
3.4 Cross-comparison of metagenomic samples	37
3.5 "De novo" assembly and gene annotation	37
3.6 "Binning" and genome reconstruction (MAGs) from	38
metagenomic datasets	
3.7 Phylogenomic classification of the reconstructed genomes	38
3.8 Metagenomic read recruitments	39
3.9 Functional assignment of assembled proteins	39
4. RESULTS	41
4.1 <u>Results derived from the work "Fine metagenomic profile of the</u> Mediterranean stratified and mixed water column"	43
4.1.1 The photic zone is delimited into three regions	43

4.1.2 Prokaryotic community structure of the stratified water		
column derived from 16S rRNA analysis and genome	45	
reconstruction and recruitment		
4.1.3 Seasonal variation in the prokarvotic community structure		
derived from 16S rRNA analysis and genome reconstruction and	46	
recruitment	-	
4.1.4 The microbial metabolism is modified by depth and season	46	
4.1.5 Rhodonsins showed a sharp gradient with depth	48	
4.2 Results derived from the work "Novel Caudovirales associated	10	
with Marine Group I Thaumarchaeota assembled from	49	
metagenomes"	12	
4.2.1 Novel Thaumarchaeota viruses were recovered from		
metagenomes and viromes	50	
122 Phylogenomic analyses revealed a senarate lineage from		
other archaeal Caudovirales	51	
A 2.3 Marthavirus as an code for the auxiliary metabolic gene cobs	51	
4.2.5 Matagenomic recruitments showed a patchy distribution	51	
with low intra population diversity	52	
A 2.5 Metatranscriptomics confirmed an active viral replication	52	
4.3 Results derived from the work "New insights into marine group	52	
III Furvarchaeota: from dark to light"	53	
4.3.1 Novel MG-III Furvarchaeota genomes were recovered from		
metagenomes	53	
A 3.2 Phylogenomic analysis divided MG-III genomes into small		
subgroups	54	
4.3.3 Genomic comparisons confirmed the division of MG-III into		
small groups	55	
A 3 A MG-III shares with MG-II a similar photoheterotrophic		
lifestyle	55	
4.3.5 MG-III represents a cosmopolitan and widespread group	57	
5 DISCUSSION	59	
5.1 Metagenomics to understand the taxonomical and functional	57	
prokaryotic composition of the water column	61	
5.2 The role of metagenomics to recover the active-replicating viral		
diversity	65	
5.3 Metagenomics recovery of novel genomes of the uncultured		
marine group III Eurvarchaeota	67	
6. CONCLUSIONS	71	
7 REFERENCES		
8 ANNEX 1		
9. ANNEX 2		
10. ANNEX 3	156	
	100	



1. INTRODUCTION



1.1 <u>Physical and chemical structure of the Ocean</u>

The ocean, with an average depth of 3.6 km, represents approximately 70% of the Earth surface and contains nearly 97% of the total water [1]. All kinds of organisms need water to live. Is, in this medium, where all the chemical reactions take place. For that reason, the ocean is a living ecosystem that comprises 90% of the biosphere, and thousands of species of macro-and micro-organisms live together. Due to their wide metabolism, they play an essential role in the biogeochemical cycles of carbon (i.e. carbon fixation, methane and CO₂ production), nitrogen (i.e. N₂ fixation, ammonia oxidation) and sulfur (i.e. sulfur oxidation, dimethyl sulfide production) [1]. However, the ocean is far from homogeneous, and differences on it are found both horizontally among water bodies, and vertically within the water column [1].

1.1.1 Surface and deep-water currents in the open ocean

Two types of flow contribute to the currents in the ocean. Surface currents are driven by the continuous friction that the wind generates over the surface, transmitting the momentum associated with air molecules to the water. However, these wind-driven surface currents affect only about 10% of the total ocean volume [1]. Winds are formed due to differences in the temperature between the equator and the poles. In warm latitudes, as in the equator and subtropics, the sun heats and expands the air, decreasing its density and originating a lowpressure zone. The wind generated from warm air rises and moves to the poles. Here, the air is cooled and increases its density (high-pressure zone), to the point that descends to the surface and generates friction over the ocean surface [1]. Nonetheless, water currents do not flow parallel to the wind, but instead, they bend to the right in the North Hemisphere and to the left in the South. This bending, called the Coriolis deflection, depends on the water speed and latitude, and have an impact on the different surface current pattern. The movement of the topmost water molecules in the surface sets in motion those that lie beneath them and the movement is transferred downwards the column [1]. Again, due to the Coriolis deflection, deeper layers move slower but to the right of the flow, showing a spiraling pattern, until no more energy is transferred from the upper layer to a lower layer. This spiral can extend to a depth of 100 to 200 meters, depending on the strength of the wind [1]. As a result, the surface current is displaced 90 degrees to the right of the wind in the North Hemisphere and the opposite direction in the South Hemisphere. This effect, named Ekman transport, plays a role in other types of surface water currents, such as the downwelling and upwelling. If the wind flows parallel to the edge of a shore in the proper direction, the resulting Ekman transport

moves surface water far from the shoreline, and deeper water rises to compensate the loss (upwelling) [1]. Conversely, if the wind flows in the opposite direction, surface water sinks (downwelling) [1]. These vertical currents can also occur in the open ocean. At midlatitudes, the surface currents of the circulation gyres converge and induce the downwelling of water. This phenomenon happens, as an example, in the very oligotrophic Sargasso Sea [1]. Conversely, the upwelling of water happens near the equatorial ring [1].

Subsurface currents, called the thermohaline circulation, arise from density differences between water masses generated by variation in the water temperature and salinity, rather than the action of the wind. Although these currents flow very slowly, they affect the 90% of the ocean's total volume. It is important to note that the processes modifying the temperature and salinity occur near the surface. Hence, the thermohaline circulation depends on climate and latitude. When dense, cold water coming from the poles travels to the equator, sinks below less dense, warmer surface waters from temperate zones. The cold water descends and flows horizontally, filling the deep ocean with water rich in nutrients and oxygen. At some point, the deep water goes upward and reaches the surface waters again (upwelling) and return nutrients to the surface. One proposed model tries to connect the deep and surface circulation, as the water flow follows an immense conveyor belt (Figure 1). This large-scale circulation has a tremendous effect on marine organisms.



Figure 1: An scheme of the ocean conveyor belt representing the thermohaline circulation [2].

1.1.2 Vertical stratification of the water column

Light is not only used to heat the surface layers of the oceans but also is to paramount importance to sustain life in such an oligotrophic environment. Certain microorganisms are capable of using the light to obtain energy in the form of ATP (rhodopsins) or to fix inorganic carbon (photosynthesis). Thus, in the open ocean, the water column is divided into two main layers. The photic zone extends from the water surface down to 100-150 m and corresponds with the epipelagic zone [1]. Below lies the aphotic zone, that comprises the largest part of the water in the ocean. This place is entirely dark, so the expression of light-related metabolic pathways is nearly absent. However, these waters are rich in nutrients and oxygen, and microorganisms can obtain energy from the oxidation of inorganic compounds, such as NH4⁺ or NO2⁻, and the oxidation of recalcitrant dissolved organic matter [1]. The aphotic zone comprises the mesopelagic (200-1000 m deep), bathypelagic (1000-2000 m deep), abyssalpelagic (2000-6000 m deep) and the hadalpelagic zones, the latter found only in deep-sea trenches [1].

Due to differences in temperature between the surface and deeper waters, the water column, mainly the photic zone, is strongly stratified and a gradient of physicochemical parameters, such as temperature, salinity and nutrients vary with depth. In the very top, surface waters are warmer, saltier and, typically less dense than waters below. These differences create the thermocline, halocline and pycnocline, respectively, and have profound implications for the microorganisms inhabiting in the water column [1]. They physically separate two water masses and impede the upwelling of nutrients from deeper waters to the surface. As a result, due to the continuous activity of phytoplankton and bacteria, surface waters become ultraoligotrophic, being depleted of nutrients, mostly nitrogen and phosphorus, limiting the growth and primary productivity of the microbial community [1].

1.1.3 The Deep Chlorophyll Maximum (DCM)

Within the photic zone, it is found the DCM (Figure 2), a layer where the highest concentrations of photosynthetic organisms are detected and, therefore, most of the primary productivity takes place [3]. The net primary productivity denoted as the amount of carbon fixed by photosynthesis that exceeds respiration occurs in the water column down to the compensation depth, below which there is no net productivity [3]. The compensation depth occurs where light levels are reduced to about 1 per cent of their surface value. In the open ocean, the compensation depth can reach down to approximately 100-150 m below the sea

Introduction

surface [1]. Furthermore, the DCM is also associated with the nutricline, a region in the water column where the greatest change in the nutrient concentration occurs, allowing to phytoplankton and heterotrophic bacteria to have access to nutrients coming up from mesopelagic waters [3]. Given the high number of microorganisms in this layer, the DCM plays a significant role in the cycling of nutrients, such as carbon, nitrogen and phosphorus [1]. The location and depth of the DCM depends mainly on two factors, light availability and temperature, being present below the thermocline. Thus, the DCM is a permanent feature in tropical waters, located down to 100 m deep [4], while in temperate waters, however, is a seasonal phenomenon present only from mid-April to late-October (see below), located down to 75 m deep [5].



Figure 2. Typical conditions in the subtropical ocean. The thermocline stratifies the upper water column. The deep chlorophyll maximum (DCM) appears where there is enough light for photosynthesis and nutrient supply from below [6].

1.1.4 Seasonal dynamics of the water column

Nevertheless, the stratification of the water column varies with latitude. In tropical and subtropical latitudes ($20^{\circ}N - 20^{\circ}S$) the stratification is a permanent feature, which is created by the persistent and robust heating of the water by the tropical sun throughout the year. There is no physical connection from the poor-nutrient surface and rich-nutrient deep waters, so the

surface layer remains ultraoligotrophic regardless of the season. In fact, tropical and subtropical waters harbor the lowest primary productivity in the ocean [1]. Only those waters close to i) the equatorial ring and ii) near the coast are productive due to the upwelling of nutrients from deep-rich water towards the ocean surface [1]. In the first case, Ekman transport causes water to move far from the equator due to the Coriolis deflection. Surface water is then replaced by nutrient-rich water coming from below. Secondly, winds blowing parallel to the edge of a land can generate Ekman transport that displaces surface water away from the coastline and induce the upwelling of nutrients [1].

In temperate waters (20-60 °N and 20-60 °S), the thermocline appears between 40 and 100 m, but it changes with the season, being present during warmer months (typically mid-April to late October [7]). However, the thermocline disappears during winter, which triggers the mixing of the water column and, therefore, the upwelling of nutrients from deeper layers [7]. Despite the availability of nutrients in winter, the weaker sunlight limits productivity. It is during spring that the combination of both light and the availability of nutrients induces phytoplankton blooms. Finally, during summer, phytoplankton decays because of the intense grazing pressure by zooplankton and the formation of the thermocline that limits the availability of nutrients in the surface [7].

What is clear is that the seasonal variation of the water column impacts the prokaryotic community composition. Several studies have illustrated the community change between seasons. For instance, the bacterial phyla Bacteroidetes and Verrucomicrobia are related to the blooming of phytoplankton during spring [7]. Members of these phyla are (photo)heterotrophic bacteria that excrete an unusually high number of extracellular enzymes [7, 8]. These enzymes catabolize the degradation of complex sugars (i.e. oligo- and polysaccharides) released to the media after phytoplankton death. Other fast-growing microorganisms are also benefited by the release and upwelling of nutrients [9]. On the other hand, other resilient microorganisms [10] seem to appear during summer, when conditions are extreme and disappear in winter [7, 11–14].

Lastly, in polar latitudes (60-90° N and S), the surface water is always too cold, so there is no thermocline, and the water column is thoroughly mixed and rich in nutrients [7]. However, the biological productivity in polar seas is limited not by nutrients but by solar energy. It is only in summer when there is a continuous flux of sunlight that the net primary productivity reaches its maxima [7].

1.2 <u>Marine metagenomics</u>

Initial marine microbiology studies used the isolation of culturable microorganisms as a method to understand their diversity and function in a given environment, but culture did not capture the enormous spectrum of the microbial diversity and, hence, those studies represented only a tiny part of the entire community. This fact was apparent when the numbers of plate counts and viable cells estimated by staining substantially differed by four to six orders of magnitude [15], indicating that only 0.1 to 1% of the microorganisms were culturable. Taking advantage of the development of the Sanger sequencing, it was in 1985 when Pace and colleagues first described the diversity of microorganisms in an environmental sample without culturing [16–18]. This new field comprised both the analysis of the community by using a gene marker, mainly the 16S ribosomal RNA (rRNA) sequence, and by randomly sequencing DNA fragments. However, some authors claim that only the latter approach can be called "metagenomics", first coined by Handelsman in 1998 [19], and we should refer the former as metataxonomics [20, 21]. Figure 3 shows a summary of the main milestones of these two disciplines in the ocean.



Figure 3. Timeline with the major milestones in metataxonomics and metagenomics.

1.2.1 First attempt: Metataxonomics

Metataxonomics has been widely used to sequence the 16S rRNA gene, due to its conserved regions that can be targeted by universal PCR primers, and its ubiquitous presence in the prokaryotic community. Furthermore, there are databases such as RDP [22] and SILVA

[23] that contain hundreds of thousands of 16S rRNA sequences and allow classifying taxonomically environmental sequences with simple bioinformatics tools. One caveat when dealing with metataxonomics is the low resolution to discriminate between species of some abundant taxa, such as Prochlorococcus and Ca. Pelagibacter. In Prochlorococcus, the comparison of three isolates [24, 25] at the level of 16S rRNA gene sequences shows that the three strains belong to the same species (>97% identity), although they have different genome sizes (1.7 to 2.4 Mb) and different ecological niches. Thanks to a robust whole-genome comparison among these three strains, which shared less than 95% of average nucleotide identity (ANI), it is known that they represent different species within the same genus. The same anomaly has been described within genomes of Ca. Pelagibacter, on which members of the same Ia subclade share 16S rRNA gene identity values higher than 98% and an average amino acid identity (AAI) lower than 80%, values that are representative at the genus level [26]. Moreover, other issues may appear when dealing with metataxonomics: i) the variation of the 16S rRNA gene copy number and their intragenomic (within the same microbe) heterogeneity in some taxa may overestimate the microbial diversity and relative abundance in a sample [27]; ii) depending on the primers used, results may be strongly biased, and some taxa cannot be detected [28, 29].

Nevertheless, we should not ignore the fact that most of the information regarding the hidden microbial community composition in different environments was started by several authors that applied metataxonomics. The first reports of mesophilic archaea inhabiting in the sea were published in the early 90s [30–32]. Before that, only some isolates from high temperature or high salinity environments were retrieved and was thought that these organisms were exclusive from such extreme habitats. From a natural sample taken in the Sargasso Sea [33], researchers were able to describe a high diversity, mostly dominated with the alphaproteobacterial clade SAR11, that numerically accounted for 15% of the prokaryotic rRNA fragments analysed. This group was followed by picocyanobacteria of the genera *Synechococcus* and *Prochlorococcus* and in a lesser extent with other previously unclassified bacteria such as SAR86, SAR116 or archaea. Other studies also detected the presence of ubiquitous organisms, such as members of the Bacteroidetes and Actinobacteria phyla [34–36]. These studies opened the door to understand that the microbial world is enormously rich, complex and sharply uneven, i.e. in which a few organisms are numerically dominant, whereas many others are represented at very low abundance and are known as the "rare biosphere" [37].

1.2.2 Second attempt: Shotgun metagenomics

However, only with the general adoption of shotgun metagenomics, first by cloning short environmental DNA fragments and later with the advent of the high-throughput sequencing, it has become possible to obtain, not only the community composition (who are there) but also their functional role in the environment (what are they doing). In the first approach, relatively large fragments (up to 100 Kb) of DNA from an environmental sample are cloned in fosmid or plasmid libraries. These DNA fragments contain several genes from the same genome they came on and, if a 16S rRNA gene is found in the fosmid clone, some metabolic traits can be associated with a specific species, genus or group.

The well-known study of this approach in the ocean and a milestone in marine ecology was carried out in 2004 in the Sargasso Sea [38]. This study, led by Craig Venter, marine samples were taken at three different stations off the coast of Bermuda, filtered and sequenced by automatic Sanger sequencers, generating up to 1.5 Gb of prokaryotic and viral DNA sequences, detecting roughly 70,000 novel genes, describing a novel ammonia oxidation pathway not photo-inhibited by UV light and characterizing an enrichment of genes for the uptake of organophosphates as a source of phosphorus (P) in a P-limited environment. In another illustrative example of shotgun metagenomics, Delong and colleagues cloned and sequenced an extensive collection of fosmids (~125,000 fosmids, 4.5 Gb of sequence) collected in the North Pacific Subtropical Gyre [4]. They took samples at seven depths, ranging from surface (10 m) to near seafloor (4000 m) and concluded the presence of a taxonomic and functional vertical distribution in the water column. Specifically, samples from the upper euphotic and DCM were dominated by members of SAR11 clade and Prochlorococcus, whereas in deep aphotic samples they were replaced for microorganisms belonging to Thaumarchaeota, Planctomycetes, Nitrospina and the SAR202 clade of Chloroflexi. Remarkably, they captured in the prokaryotic filter a large number of viral DNA, that was originated from the replication of phages infecting host cells. The highest recovery of viral DNA occurred in the photic zone, and decreased trough depth, due to almost 67% of them shared high similarities to cyanophages.

The linkage of the 16S rRNA marker gene to the metabolic potential of the fosmid clone resulted in the discovery of the proteorhodopsin gene in the yet uncultured cluster bacterium SAR86 [39], where the existence of this kind of proteins had only been detected so far in halophiles. When this protein is irradiated with blue/green light, it generates, in most of the cases, a proton gradient towards the outside of the cell. The coupling of this electrochemical

gradient to an ATP synthase serves to fill the cell with energy (ATP) [40]. Further metagenomic studies have increased our knowledge of the rhodopsins. For instance, in the Sargasso Sea metagenomic study, almost 800 new proteorhodopsin genes were retrieved, which gave insights into the widespread photoheterotrophic lifestyle of marine bacteria and archaea [38]. Indeed, diverse types of rhodopsins (H⁺, Na⁺ or Cl⁻ pumps) and families have been detected in metagenomic sequences opened the door to a novel branch, named "functional metagenomics", which mainly aimed to find novel genes that, after heterologous expression in a bacterial host, their product has an activity of interest, such as novel antibiotics [45], antibiotic-resistant genes [46] or enzymes involved in the degradation of complex compounds [47].

In the second approach, next-generation sequencing technologies (NGS), mostly, 454 pyrosequencing and Illumina, were used in metagenomic studies, where environmental DNA is randomly fragmented in pieces of 100 to 800 bp, depending on the technology, and massively sequenced. This approach has been widely accepted due to its low cost. Besides, the higher sequencing coverage achieved with NGS allowed researchers to i) discover a large number of low abundance microorganisms and ii) reconstruct their genomes to understand their metabolism and their interaction with nutrient cycles. Thus, this approach has been used to explore, in much finer detail than previously studied, the microbial diversity in several oceanic locations and depths [5, 37, 48–53]. Moreover, due to its advantages, as mentioned earlier, the application of NGS has been used in several sampling expeditions. The Global Ocean Sampling (GOS) was the first to collect surface samples from the Northwest Atlantic and Eastern tropical Pacific Oceans [54], producing more than 6 Gb of short-read sequences. Conversely, the Malaspina circumnavigation collected multiple prokaryotic and viral samples from the bathypelagic ocean [55]. Recently, the *Tara Oceans* consortium made a massive sampling effort [56], sequencing 7.2 Tb of DNA recovered from 243 samples from 68 locations at different depths comprising the surface, DCM and mesopelagic layers and from different filters (viral and prokaryotic free-living or particle attached).

1.2.3 *Metagenome-assembled genomes (MAGs) and single-amplified genomes (SAGs)*

After the metagenomic assembly of the sequenced short-reads (mostly Illumina), resulting contigs can be grouped according to their (i) genomic signatures (i.e. tetranucleotide frequencies and GC content), (ii) taxonomic affiliation, and (iii) their co-abundance values within metagenomic datasets. The resulting groups are called metagenome-assembled genomes

Introduction

(MAGs) and represent composite genomes of clonal lineages within species. Thanks to the public availability of several metagenomic datasets, it has been possible to reconstruct thousands of novel MAGs, which have served to significantly increase the number of novel species, genera and families [57–60]. In this way, a more reliable taxonomic classification based on the presence of several shared genes from genomes rather than the use of a single marker gene such as the 16S rRNA has been possible [61], which enabled an improved classification of uncultured bacteria. Through this genome-based classification, almost 60% of the genomes changed their existing taxonomy, and the prokaryotic realm was condensed to a total of 99 phyla [61].

The recovery of MAGs can be a powerful tool when no close relatives of a species are available, as sometimes happens with samples from unexplored ecological niches. Thus, MAGs can be used as references to recruit metagenomic reads and obtain a more reliable picture of the horizontal and vertical distribution of these microbes at the finer levels of diversity, such as species, ecotypes or even clonal lineages. Furthermore, it is possible to infer their biological role in a specific location or depth by reconstructing their core metabolic traits, that could explain their differential distribution among samples. For instance, Delmont and colleagues [59] studied the assemblages collected from the Tara Oceans sampling to characterise the MAGs involved in nitrogen fixation in the surface ocean, which corresponded to lineages within Proteobacteria and Planctomycetes, previously not identified as marine diazotrophs. Although widespread, they found that they were abundant in both the Pacific Ocean and the Atlantic Ocean northwest. Hawley and colleagues determined the global biogeographic distribution of the phylum Marinimicrobia along eco-dynamic gradients ranging from oxic to anoxic, together with sulfidic and methanogenic conditions [62]. There are many other examples of the metabolism characterisation, geographical distribution and ecological niche partition in different microbes such as Euryarchaeota [63, 64], SAR202 clade of Chloroflexi [65, 66], SAR11 [67–69] or the gammaproteobacterial clade SAR86 [70].

However, MAGs are not flawless. First, as mentioned above, MAGs are "composite" genomes. They do not represent a unique organism, but they are comprised from the consensus of the local assembly of several clones that differ in abundance, gene content and nucleotide composition (single nucleotide polymorphism – SNP). Second, the presence of multiple genomic repetitions, such as gene duplications, insertion sequences or multiple copies of the rRNA ribosomal operon, interrupt the assembly, and the genome is fragmented into hundreds of small contigs. A low sequencing coverage for the less abundant microbes also contributes to the fragmentation. Third, these contigs only represent the core genome, leaving the flexible

genome out of the bin. Therefore, joining the resulting small contigs of a given organism from a pool of thousands of microbes (metagenome) can become a challenging task. For that, binning algorithms have been applied that rely on co-abundance profiles among samples and in the shared genomic features in terms of GC content and k-mer frequencies [71]. However, very often when done automatically, MAGs can be contaminated with fragments of other organisms and, in the end, interfere in the correct phylogenetic placement of the bin and its metabolic reconstruction. Therefore, after this automatic binning it is recommended to carry out a manual curation of all the MAGs in order to avoid chimeras.

Single-cell genomics (SCGs) overcome some of the issues described before for MAGs. In this novel approach, cells collected from an environmental sample pass through a flow cytometer that selects, using different criteria (i.e. size or fluorescence), for a particular population (eukaryotic, prokaryotic or even viruses). Later, cells are distributed into a multiwell plate, and only one cell goes in one well, which are individually amplified, sequenced and assembled, generating single-amplified genomes (SAGs). Following this method, the sequenced DNA of a SAG truly represents one single clone collected from a drop of water. In addition, there is no need to bin contigs into genomes and, therefore, the risk to have significant levels of contamination is scarce. Because of that, the application of SCGs in microbiology has allowed to study the existing microdiversity between clones within the same species, as reflected in one study with Prochlorococcus in the North Pacific Ocean [72]. Authors concluded that the sample was composed of hundreds of divergent and stable subpopulations of Prochlorococcus. Additionally, they found that each subpopulation carried a small set of distinct genes grouped in cassettes within genomic islands involved in outer membrane modifications. This variation in the flexible gene content and its fine-scale resolution represented a new dimension of microdiversity within wild *Prochlorococcus* populations [72]. Another illustrative example resides in the recovery of SAGs belonging to SAR11 clade. Members of this group are oligotrophic, slow-growing bacteria, so the number of isolates is quite small (#18), considering that is the most abundant bacteria in the ocean (see below). Thus, SCG has provided with the first genomes of freshwater (clade IIIb – LD12) and marine aphotic (clade Ic) SAR11 bacteria, that enabled genomic comparisons between photic and aphotic, and between marine and freshwater sequences [73, 74].

It is important to note that, due to the small DNA concentration within one cell, an amplification step is needed in order to perform the sequencing of the SAG. Unfortunately, the multiple-displacement amplification (MDA) of the cell genome results in an uneven amplification, with parts with high coverage and others with low coverage or even unamplified.

Introduction

In the end, MDA generates very fragmented contigs and, in some cases, the percentage of the recovered genome is less than 10% [66]. The low genome completeness aggravates future genomic analysis, missing essential genes or even the flexible genome, generating an incomplete metabolic reconstruction.

1.2.4 Future perspectives: The Third-Generation Sequencing

In 2011, Pacific Biosciences (PacBio) commercialised the "Single Molecule Real Time (SMRT)" sequencing technology [75]. Three years later, in 2014, Oxford Nanopore released its first MinIon sequencer [76], a user-friendly device no bigger than a smartphone that permitted researchers to efficiently sequence samples anywhere, just with the help of a laptop. Although based on different approaches [77], both technologies promised very long reads, in the order from thousands to hundreds of thousands of bases [78, 79]. These long reads surpassed the limitations of the second-generation technologies, in which the retrieval and taxonomic assignment of the flexible genome and larger genomic fragments were hampered in complex populations, where multiple strains were abundant and highly diverse. The term flexible genome refers to the accessory genes that are not present in every one of the strains of a species [80]. These genes are diluted in the pool of multiple strains within the same sample, and usually contains repeat regions at the boundaries [81, 82], for this reason, the assembly based on Illumina short-reads is usually broken. These regions are often acquired through homologous gene transfer (HGT), which in most of the cases contains different GC-content and tetranucleotide frequencies, different genome coverage, and taxonomical assignment of genes; thus, they cannot be binned with the core genome into any MAG. The lack of the flexible genome in MAGs results in the loss of information of niche-defining genes that usually drive ecological speciation [83]. However, these problems can be overcome using long reads (PacBio and Nanopore), comprising both the core and the flexible genome (i.e. genomic island), which can ultimately be linked to one specific organism and, in the end, can be used to retrieve much more complete genomes from metagenomes.

However, these new technologies are more expensive and give much less throughput than the Illumina platform. Furthermore, the resulting long reads have high error rates (on the order of 5-15%), generally as insertions and deletions, which introduce frameshifts and premature stop codons [84]. This has resulted in the development of several methods to correct the error rate. One approach consists of a hybrid assembly, on which the almost error-free and short Illumina reads are used to correct the high error rate found in the long PacBio or Nanopore

reads. Other approaches consist of technical changes. For instance, PacBio improved the technique by repeatedly sequencing the same DNA fragment that has been circularised by adding adapters at the ends. In this way, the error rate can be reduced down to 0.01%, and the accuracy significantly increased, as announced on April 2019 for the new Sequel II system.

Eventually, the second sequencing generation will be replaced with the new, much better "Third-Generation Sequencing", and although nowadays is widely used for the sequence of individual genomes, in metagenomics is still lagging, and only a few studies have been published [85, 86].

1.3 Description of marine organisms

1.3.1 Common features of marine bacteria and archaea

Metataxonomics and metagenomics studies have pictured the ocean as a vastly rich and incredibly complex environment, with thousands of species living in a single drop of water. In fact, several studies indicated that in one millilitre live 10^5 to 10^7 prokaryotes. Thus, the oceans are populated by more than 10²⁹ microbial cells [87]. However, most of the abundant and widespread prokaryotes that inhabit the ocean have small cells and genome sizes, with low GCcontent, caused by an evolutionary streamlining of their genomes [88]. By these means, they are competitive in oligotrophic environments. Although these microorganisms seem to have less metabolic potential and a lower number of transporters for different substrates originated from the loss of genes during their evolutionary history, streamlining has provided them with some advantages. For instance, a smaller and AT-rich genome requires less nitrogen and phosphorus for cell replication; these essential elements are generally limited in the ocean [4, 89]. Besides, due to the small cell size, they benefit by a higher surface-to-volume ratio that bestows a better nutrient transport [90], which can be critical to success in nutrient-poor environments, like the oligotrophic open ocean [88]. Of course, organisms have to balance genome reduction and the need to keep enough genes to maintain a functional ability to succeed within a specific niche or even after responses to environmental changes [88]. In that way, some abundant organisms (i.e. Synechococcus or the Marine Group II Euryarchaeota) do not suffer a strong genome streamlining.

Very often, prokaryotes are divided based on their nutrient demands. The term oligotroph is used to those microbes that grow at very low-nutrient concentrations, such as those present in the open ocean. Thus, most of the marine microorganisms that are abundant and cosmopolitan are oligotrophs [91, 92]. Copiotrophic refers to bacteria that grow at high-nutrient concentrations, like in coastal waters or after the mixing of the water column during winter that upwells nutrients from bottom to surface layers. Examples of marine copiotrophic bacteria are *Alteromonas* or *Pseudomonas*, which appear forming blooms in rich-waters and are easily culturable. The terms *r*-strategist and *K*-strategist are often interchangeable with copiotroph and oligotroph, respectively, although sometimes their use can be misleading. *K*- and *r*-strategist terms refer to the speed of the replication when nutrients are abundant, differentiating between a low steady growth (*K*-) versus a fast bloomer (*r*-) [93].

1.3.2 *Common marine prokaryotes in the water column*

Independently of their genome or cell size, marine bacteria and archaea are considered the major contributors in several biogeochemical cycles. Some examples of cosmopolitan and abundant microorganisms that play a central role in these cycles are listed below. It has been calculated that approximately half of the oxygen on Earth is produced in the oceans by marine Cyanobacteria and eukaryotic phytoplankton, which also contribute to the carbon fixation [94]. Other chemolithoautotrophic microorganisms also contribute to the carbon fixation [95, 96]. However, almost all the oxygen and the newly synthesised organic matter is rapidly consumed in the photic zone by (photo)heterotrophic bacteria to CO₂ and recalcitrant dissolved organic matter, that continuously sinks to the bottom of the ocean, process known as the "biological pump" [97]. Some bacteria can degrade dimethylsulfoniopropionate (DMSP), an organosulfur compound that acts as an osmolyte in marine algae, to methanethiol and dimethyl sulfide (DMS) [98]. The former is a source of protein sulfur while the latter flows towards the atmosphere and have a significant impact on global warming [99]. Other bacteria and archaea contribute to the nitrogen cycling, by fixing atmospheric nitrogen [59], nitrifying ammonia to nitrate [95, 100] or denitrifying, in oxygen-limited zones, the dissolved nitrate to N₂O and N₂, well-known greenhouse gases [101].

a) Picocyanobacteria. Two genera, *Prochlorococcus* and *Synechococcus*, dominate marine picocyanobacteria. These bacteria represent the most abundant and widespread oxygenic primary producers in the ocean. However, due to the main oligotrophic condition of the oceans, *Prochlorococcus* are, by large, more abundant than *Synechococcus* [102]. Only in nutrient-rich waters, *Synechococcus* outnumbers *Prochlorococcus* [103]. Both genera had been successfully cultured, so their isolates had resulted in several studies on their physiological and genetic
diversity. They were first detected in 1979 by epifluorescence on natural seawater samples [104] and later classified based on their size and photosynthetic pigments. Prochlorococcus light-harvesting chromophore is a derivative of chlorophyll, divinyl chlorophyll-a and -b [91]. The ratio of these two pigments affects their vertical distribution across the stratified water column and separate the genus Prochlorococcus into two big groups or ecotypes, High-Light (HL) and Low-Light (LL). The term ecotype refers to organisms within the same species or genus that share the same ecological niche [91]. For instance, members of the LL clade have a higher ratio of divinyl chlorophyll b/a and a higher number of genes encoding the lightharvesting antenna protein Pcb than their HL counterparts [91]. This allows them to grow at very low light intensities [105]. Prochlorococcus possess a very streamlined genome, between 1.64 and 2.68 Mb and a GC-content between 31 and 51%. However, most of them are smaller than 2 Mb, and only the members of the LL-IV clade have high GC-content and large genome size [106]. Conversely, Synechococcus cells lack chlorophyll-a but rather possess the phycobilisome, a structure composed with the accessory pigments phycoerythrin, phycocyanin and allophycocyanin, that act as light-harvesting receptors [107]. Remarkably, there is considerable spectral diversity in the phycobilisome depending on the type and ratio of the chromophores present [107], that affects their global distribution [108]. Synechococcus genomes range in size between 2.2 and 2.86 Mb and a GC-content between 52 and 66 % [109].

These genera tend to co-occur in temperate and tropical areas. However, their pattern of abundance differs both spatially and seasonally [91, 103, 110]. Prochlorococcus is mainly found in temperate oligotrophic waters, with a maximal concentration in summer in deep waters of the photic zone, when the water column is stratified. On the other hand, Synechococcus prefers nutrient-rich waters, like those close to the coast, latitudes higher than 40° N or 40° S or after a mixing event in temperate waters (spring). Within *Prochlorococcus*, the aforementioned HL and LL groups are further subdivided into several ecotypes. Within the polyphyletic Synechococcus genus, the same subdivision occurs. These ecotypes are adapted to a different temperature, light intensity, depth and nutrient availability [110]. For instance, the HL-I P. marinus strain MED4 and the LL-I P. marinus strains NATL1A and NATL2A encode a set of genes to the high-affinity uptake of phosphorus and alkyl phosphonates in waters that are P-limited [111]. Nitrogen and ferrous limitation in some places also favoured the acquisition of specific genes in genomic islands to deal with [91, 112]. In that way, their global distribution is then delimited, and members of the HL-I Prochlorococcus and the Clade-III of Synechococcus dominate over the Mediterranean Sea, whereas members of the HL-II and Clade-II are the main picocyanobacterial contributors in the Red Sea or the Indian Ocean [110].

Introduction

b) SAR11. Like most of the marine microbes described here, the alphaproteobacterial clade SAR11 was firstly described after a 16S rRNA amplicon survey from a natural sample taken in the Sargasso Sea [33], and later confirmed by fluorescence in-situ hybridization (FISH) to be the most abundant microbe in the ocean's surface, accounting to 20-40% of all bacterioplankton in the photic zone and approximately 20% in the aphotic [113]. Indeed, it also contributes to the freshwater bacterioplankton composition. Some close members of the SAR11 clade, the LD12 subgroup IIIb, now in pure culture (*Fonsibacter*) [74] or a recently discovered MAG belonging to the Ia clade [114], were found in freshwater systems. Based on a 16S phylogenetic tree, several metagenomic surveys have divided this abundant and widespread group into nine subclades [113] which are discernible by season, depth and geographical location [115–117]. For instance, the subclade Ic has been only found in meso-and bathypelagic waters; the subclades IIa and IV mainly in the DCM; and the subclade Ia is further divided into Ia.1 and Ia.3 with differences in their abundance in cold and warm waters, respectively.

Since their discovery in 1991, eleven years had to pass so that they could be isolated in pure culture [118, 119], that allowed to recover information about their genome and ecophysiology. Due to their high abundance and metabolic activity [120, 121], members of the SAR11 clade play a central role in the cycling of dissolved organic matter (DOM) [113]. These microbes have a minimal genome and cell volume, with an average of 1.33 Mb and 0.015 μ m³, respectively [119]. They are photochemoorganotrophs, obtaining energy through light (proteorhodopsin) and the oxidation of reduced organic compounds [90, 122, 123]. Specifically, they can use a wide variety of labile low-molecular-weight dissolved organic matter, like pyruvate, C1 compounds (methanol, methylamine or formaldehyde, among others), amino acids, glycine betaine, taurine, polyamines (putrescine and spermidine). Furthermore, they can obtain reduced sulfur from the osmolyte DSMP and the dissolved gas methanethiol [98]; and some strains contain the operon for the assimilation of phosphonate in P-limited environments [90].

c) Thaumarchaeota. Based on rRNA studies, the domain Archaea was first discovered in the 70s [124] and subdivided into two groups, Euryarchaeota and Crenarchaeota. At that time, only isolates from extreme environments (thermophilic, acidophilic and hyperhalophilic) were known. It was not until the development of metagenomics when sequencing the 16S rRNA

amplicon from natural marine samples detected the presence of a new phylum, the Thaumarchaeota [30], which emerged as a sister group of the hyperthermophilic Crenarchaeota [32, 125].

The cultivation of numerous strains [95, 100, 126], together with the sequences from environmental metagenomes, MAGs, and SAGs [127, 128] have provided invaluable information on the ecology of this phylum. The two more abundant genera in the oceans, *Ca*. Nitrosopelagicus and *Ca*. Nitrosopumilus, have a small genome size (1.23-1.64 Mb) and low GC-content (33.2-34.2 %) and are examples of genome streamlining. In the open ocean, marine Thaumarchaeota are ubiquitously detected in deeper waters below the DCM [4] and are some of the most abundant microorganisms in the aphotic realm, accounting for up to 40% of the picoplankton [129, 130]. Besides, metatranscriptomic analysis demonstrated a high activity of this group in these waters [131, 132]. Members of this lineage are chemolithoautotrophs, which fix inorganic carbon [133] fueled by the oxidation of ammonia to nitrite [95], and are responsible for the majority of the aerobic nitrification measured in the ocean [134]. Furthermore, they globally contribute to cobalamin synthesis [135], an essential cofactor involved in amino acid and DNA synthesis. Their abundance and metabolism confirmed that the marine Thaumarchaeota are essential players in global carbon and nitrogen biogeochemical cycles.

d) Marine group II/III Euryarchaeota. In the same way as Thaumarchaeota, mesophilic Euryarchaeota evolved from thermophilic ancestors through adaptation to a mesophilic lifestyle [136]. Marine Euryarchaeota were discovered by 16S rRNA surveys in the early 90s [125]. They are represented by two deep and well-differentiated clades, named Marine Group II (MG-II) and Marine Group III (MG-III), distantly related with the order Termoplasmatales and the deep-sea hydrothermal vent Euryarchaeota group 2 (DHVE2) [137]. Further studies enhanced the phylogenetic resolution of this marker gene, and MG-II was divided into four groups, with two dominant clades (MG-IIA and MG-IIB) [138, 139] and two less-represented (MG-IIC and MG-IID) [140]. Given that it has not yet been possible to culture any representative of this group, a few studies with fosmid libraries tried to evaluate their metabolism, but only mere traits could be determined [139, 141–143]. However, with the development of high-throughput metagenomics and global expeditions [58–60, 144–146], several MAGs (>200 genomes) allowed a more reliable phylogenetic resolution of the MG-II Euryarchaeota clade. Thus, by using the concatenated of shared proteins among the several retrieved MAGs of MG-II, the phylogenetic tree resulted in up to 21 genera [63, 64], with

different biogeographic ranges and nutrient preferences. Besides, Rinke et al., [64] proposed that MG-II is an order level lineage named *Candidatus* Poseidoniales, comprising the families *Ca.* Poseidoniaceae (MG-IIA) and *Ca.* Thalassarchaeaceae (MG-IIB).

Furthermore, these recent analyses significantly improved the knowledge of their metabolism, ecology and biological function in the oceans. Ca. Poseidoniales genomes share a photoheterotrophic lifestyle. They contain one copy per genome of a proteorhodopsin, but two groups of rhodopsins with different absorbing spectra (blue or green) are found among the MG-II genomes [44]. It has been proposed that these differences in the spectra are due to their vertical distribution in the photic zone [44, 147]. They encode metabolic functions characteristic of heterotrophs, including glycolysis, a Tricarboxylic acid cycle (TCA), and electron transport chain [64]. To deal with the dissolved organic matter, Ca. Poseidoniales can degrade extracellular proteins and fatty acids, as they encode peptidases, acyl-CoA dehydrogenase and acetyl-CoA acetyltransferase [63, 64, 144, 145]. Their auxotrophy for specific amino acids explains their need to degrade dissolved proteins since they depend on external sources of amino acids [64]. They can also degrade complex sugars, but it appears that MG-IIA contains more and different families of glycoside hydrolases [63]. GH involved in the breakdown of algal oligosaccharides, including pectin, starch, and glycogen, are found exclusively amongst the MG-IIA [63]. Another differentiating factor is the presence of genes for the synthesis of the flagellum in all the MG-IIA genera, but not in MG-IIB. Some authors argue about the role of these proteins because genes involved in the chemotaxis were not found [64]. Therefore, whereas Tully et al. [63] proposed that MG-IIA is motile, Rinke et al. [64] proposed that the primary role of this cluster is the adhesion to particulate material.

Members of the order Ca. Poseidoniales are ubiquitously distributed throughout the major ocean basins, although there exist significant differences between Ca. Poseidoniaceae and Ca. Thalassarchaeaceae. The former family tend to dominate nutrient-rich surface coastal waters [144, 148, 149], while Ca. Thalassarchaeaceae is found at deeper waters, mostly at the DCM layer and below, in the open ocean [145, 148]. These results indicate that planktonic Euryarchaeota occupy diverse ecological sites.

Conversely, 16S rRNA studies suggested that MG-III is predominately found in the deep ocean at relatively low abundance [138, 150], although in rare occasions it has been detected in high numbers in the Arctic Ocean [151] or an aphotic sample in the Marmara Sea [152]. In surface waters, they can represent up to 10 % in the Mediterranean Sea [149]. Several archaeal fosmids from the deep Mediterranean [141], five MAGs from deep waters of the Guaymas basin and the Cayman Rise (located in the Gulf of California and the Caribbean Sea,

respectively) [146], and eight MAGs from the photic zone in the Mediterranean Sea [153] led to the conclusion that MG-III genomes share most of the (photo)heterotrophic metabolism with the sister clade MG-II. Since the MAGs of MG-III retrieved from the Mediterranean Sea is one of the topics in this Thesis, an extended analysis of the distribution and metabolism of this group can be found in the "Results" and "Discussion" sections.

e) Actinobacteria. With a lesser abundance, Actinobacteria is also a widespread phylum in the oceans. Depending on the season and depth, they can contribute approximately to 5 % of the total bacterial population in oligotrophic waters [11, 154, 155]. First detected in seawater by 16S rRNA assays [34], the phylum was later divided into two marine clades, the Acidimicrobiales (formerly OM1 clade) [154] and the new proposed *Ca*. Actinomarinales (previously known as the "marine actinobacterial clade" [MAC]) [156]. Members of the *Ca*. Actinomarinales are examples of streamlined bacteria, similar to the alphaproteobacterial SAR11 clade, with a low GC-content (33%), small genome size, estimated to be close to 1 Mb, and small intergenic spacers (3 bp). Remarkably, their volume size of 0.013 μ m³, the lowest described so far, makes the members of this group the smallest free-living marine microbes [156]. Conversely, members of the class Acidimicrobiales exhibit larger genomes (1.7-2.3 Mb) and have a moderate to high GC-content, between 40 to 50 %.

Despite the notable differences in genome size and GC-content, they share similar metabolic traits. Both are photochemoorganotrophs, containing transporters for sugars and amino acids; enzymes involved in the glycolysis, pentose phosphate and TCA pathways; and the light-harvesting rhodopsin [156, 157]. However, their rhodopsins are phylogenetically distant; while the rhodopsins encoded within the marine Acidimicrobiales cluster together with the freshwater counterparts [157], the rhodopsins within the *Ca*. Actinomarinales form a new group named MAC-rhodopsins [156]. Despite the similar metabolism, the larger genome size of the Acidimicrobiales allows them to use several other nutrients (DMSP, C2 compounds or CO, among others) as sources of carbon and energy [157].

These two groups co-occur in temperate and tropical waters [157], although only Acidimicrobiales have been detected from polar samples [158]. They can be found permanently through the photic zone in the water column [159] and, during summer stratification, they concentrate on the DCM and deeper layers [156, 157]. However, only members of Acidimicrobiales were found in meso- and bathypelagic samples [157].

f) Other less abundant marine bacteria. Indeed, the microbes described above, although they are abundant and ubiquitously distributed, there are many other bacterial groups in the ocean. For instance, two phyla, Bacteroidetes and Verrucomicrobia are always present in the water column. Members of these groups have been characterised as particle-attached microbes that degrade polymeric organic matter, mainly polysaccharides and proteins [8, 160]. Members of the recently designated SUP05 clade of marine gammaproteobacterial sulfur oxidisers are among the most abundant chemoautotrophs in the ocean and are mainly found in low-oxygen environments [161, 162], but also the aerobic water column [96]. The OM60/NOR5 clade contains aerobic anoxygenic photoheterotrophic bacteria commonly detected throughout the euphotic zone of marine environments. They are found very often in coastal waters [14], since the aerobic anoxygenic photosynthesis is not affected by photoinhibition and, therefore, plays an important role in the ocean's carbon cycle [163]. As the last example, SAR116 and SAR86 clades are photoheterotrophic bacteria (contain rhodopsin) widely distributed in the ocean. Both groups share similar metabolism with SAR11 in terms of nutrient uptake (C, P and N) [164, 165] as well as in the case of SAR116 with the degradation of DMSP [166].

1.4 Marine Viruses

If the abundance of bacterial cells in the ocean seem astonishing, their main predators, the viruses (also known as bacteriophages or phages), outnumber them by a factor of no less than 10 [167]. Indeed, the viral abundance often varies along with the prokaryotic abundance, decreasing in the open ocean and deep layers of the water column [168]. The first estimates of the viral abundance based on electron microscopy started to appear in the late 1980s [169]. Epifluorescence microscopy, which emerged as a more reliable technique to quantify the abundance of viruses, displaced electron microscopy in further studies [170, 171].

1.4.1 Isolation and metagenomic sequencing of marine viruses

Phage genome sequencing relies on host isolation and phage purification. However, considering that most of the prokaryotes in the ocean remains uncultured, viral diversity and abundance have been poorly characterised. Unfortunately, viruses lack universal marker genes, like the 16S rRNA gene in prokaryotes, that prevent a robust phylogenetic classification. However, metagenomics has provided a powerful tool to unveil the structure of viral communities, which appears to be more diverse than previously appreciated. By using fosmid

libraries, shotgun metagenomics and viromics, thousands of novel viral species have been recently published [172, 173], including the recently discovered putative phage genomes infecting the archaeal phyla Euryarchaeota (magrovirus [174]). Single-cell sequencing is also an useful approach when it is applied to separate and sequence viral genomes [175]. Nevertheless, although it has been exponentially increased the number of viral genomes, there is still a gap in determining the host. Certain protocols use the presence of tRNAs and specific host-related metabolic proteins encoded in the phage genome, the high-similarity match to a CRISPR spacer in a given bacterial host, or the co-occurrence of bacterial and viral sequences within samples [173].

1.4.2 Role of viruses in the environment and interaction with their hosts

The isolation and reconstruction of phage genomes have increased our knowledge about the function of these biological entities. It was already known that the viral lysis modifies the prokaryotic community both directly and indirectly. Directly because it kills the organism, so the phage regulates the cell concentration of its host and maintains the diversity of the ecosystem, avoiding overpopulation of a few microorganisms [176], and indirectly because the lysis of their hosts releases nutrients into the environment that become available to the microbial community, in a process called "viral shunt" [177] (Figure 4). This process restricts the uptake of nutrients to high trophic levels (predation) or a few microbial species. The viral killing of its host was estimated to release a large amount of carbon, in the order of 10⁹ tons per day [178]. Other studies highlighted the importance of viral shunt in nitrogen cycling [179]. Furthermore, genomic studies have determined the presence of auxiliary metabolic genes (AMGs) encoded in the phage genome. These AMGs are thought to enhance during the infection the metabolic potential of the host. In this sense, phages encode genes related to the DNA replication [180], photosynthesis in cyanobacteria [181], nitrogen [182] or sulfur [183] metabolisms, among others.

Consequently, marine viruses are critical players in the ecosystem, influencing over the ocean productivity and biogeochemical cycles. But it should not be forgotten that bacteriophages are also involved in bacterial evolution, either due to the selective pressure of the virus-host interaction in terms of viral infection and defence [184], and the outer layer of host cells (i.e. LPS in gram-negative bacteria) suffers changes in its chemical composition, that is encoded in the flexible bacterial genome. Alternatively, due to phages can act as vehicles to

horizontally transfer fragments of bacterial DNA from one cell to another (transduction) [185], providing the acceptor cell with novel genes that may modify its phenotype.



Figure 4. Schematics of the viral shunt [186].

1.5 The Mediterranean Sea: a case study of the microbial diversity

1.5.1 Main features of the Mediterranean Sea: location, physicochemical properties and water currents

The Mediterranean Sea ("the sea between the lands") is a semi-closed basin, with an estimated area of 2.5 million km². It is confined between the temperate latitudes 30° and 46° N and longitudes 5.50° and 36° E. The maximum depth of 5,267 m is found in the Calypso Deep, although the average depth is 1.5 km. A submarine ridge near Sicily and the African Coast divides the Mediterranean Sea into western and eastern parts. The western part is connected with the North Atlantic Ocean through the Strait of Gibraltar, while the Eastern Mediterranean connects in the north with the Sea of Marmara and in the south with the Red Sea through the Suez Canal. Moreover, the sea is subdivided in several basins (Figure 5), each one with different temperature, salinity and nutrient concentrations [187]. Due to the high evaporation in the surface and the insufficient input of freshwater coming from land (rivers) and precipitations, salinity is, on average, slightly higher in the Mediterranean (3.8% w/v) than in the rest of the oceans (3.5% w/v) (Figure 6). In summer, this value is exceptionally high, reaching 4% in the eastern Mediterranean [1]. Surface temperature significantly varies among

seasons. In the eastern Mediterranean, surface temperatures in summer can surpass the 30 °C. In winter, the average temperature is found between 14 and 15 °C [188]. Remarkably, the temperature in the Mediterranean water column does not decrease below 13 °C, in contrast to the average temperature of 4 °C found in the aphotic realm in the Ocean [189]. Overall, surface temperatures are higher in the east than in the west. Surface chlorophyll measures showed a decreasing gradient between Western and East, with values of 0.4 to 0.05 mg/m3, respectively [190]. Nutrient concentrations (carbon, nitrogen or phosphorus) are correlated with chlorophyll-a concentrations, as in the western part of the basin are approximately two times higher than in the eastern basin [187] (Figure 6). In the end, the Mediterranean Sea has overall low productivity, and the total phytoplankton biomass is dominated by small picocyanobacteria [191].



Figure 5. The Mediterranean Sea geography and nomenclature of the major sub-basins and straits [192].

Due to the Mediterranean characteristic hot and dry climate, surface and subsurface currents are controlled by the intense surface evaporation [1]. As a result, the relative cold, normal-salinity surface water from the North Atlantic Ocean flows through the Gibraltar sill into the Mediterranean Sea, while the water mass gets warmer and saltier as it moves eastward [1]. In the Eastern Mediterranean, the dense water sinks and fills the depths of the basin as it moves westward. As it happens in other temperate zones, the downwelling of water in the eastern Mediterranean by the effect on the land (Ekman transport) removes nutrients from the surface [1]. Hence, the low nutrient concentrations in the photic zone keep net productivity low [1]. The deep water continues until overtops the shallow sill at Gibraltar strait and then returns

Introduction

out into the Atlantic Ocean as subsurface outflow [1]. These subsurface currents make the deep Mediterranean unique, as there is no contact with the dense cold waters from the open Ocean, and the mean deep temperature does not decrease below 13 °C [1]. Calculations of residence time indicate that the water that flows into the Mediterranean Sea from the Atlantic Ocean remains in the basin for about 100 years before it returns to the Atlantic [1].



Figure 6. A longitudinal transect in the Mediterranean Sea showing the variations of a) salinity, b) temperature, c) nitrates and d) phosphates through depth. Figure adapted from [187].

1.5.2 *The water column in the Mediterranean Sea: water stratification and seasonal dynamics*

Due to its temperate location, the water column of the Mediterranean Sea is seasonally stratified [193]. Since April until November, differences in water temperature from surface and deep water causes the appearance in the photic zone of a thermocline layer around 40 m deep that separates the warm surface waters (20-25 °C) from the deep colder waters (14.5 °C). During these months, below the thermocline appears the DCM, a maximum in chlorophyll concentration associated with the increase in bioavailable pools of nitrogen and phosphorus [194]. Although light intensity decreases by depth, the availability of these nutrients boosts the growth of autotrophic and heterotrophic microorganisms within this layer. In the Mediterranean, the DCM often appears between 45 and 70 m deep, depending on the light intensity and season [5]. During the late autumn and winter, temperature decreases near the surface breaks the thermocline and leads to vertical mixing of the water column. This effect promotes the upwelling of nutrients from the mesopelagic zone [7, 195]. During this period,

the DCM disappears, and the temperature, nutrients and chlorophyll concentrations remain homogenous throughout the photic zone [196]. The bioavailability of these nutrients ultimately results, during spring, in phytoplankton and heterotrophic blooms [195] near the surface, as detected by chlorophyll-a measurements [197].

Metataxonomic and metagenomic studies have been carried out in the Mediterranean Sea to characterise the microbial community inhabiting at different depth and seasons. For instance, results published by Martin-Cuadrado and colleagues [189] showed that the Mediterranean bathypelagic community (3000 m deep, 13 °C) resembled more to the mesopelagic community in the Pacific Ocean (500 m deep, 7.2 °C) than samples gathered from the cold bathypelagic ocean (4000 m deep, 1.4 °C) [4]. These results indicated that, in the absence of light, the temperature might be considered as a major factor in the aphotic water column [189]. Besides, it showed the dominance of heterotrophic bacteria, such as Alteromonas sp. or the ammonia oxidising Thaumarchaeota. In another study, an unusual number of CO dehydrogenase genes was found encoded in fosmids collected at 3000 m, indicating that the oxidation of CO is a source of energy in bathypelagic habitats [198]. In the photic zone, due to the easiness to retrieve and process the sample, several studies defined the prokaryotic community in surface waters, dominated by heterotrophic bacteria (mainly SAR11) [35, 199-201]. These studies highlighted a prokaryotic community enriched in rhodopsin genes to cope with the low nutrient concentration in this layer [201]. Moreover, some studies indicated that the surface microbial community varies by season, with some phylotypes disappearing during winter [12, 158, 202]. The Mediterranean DCM has also been thoroughly studied. By using fosmid libraries and shotgun sequencing, Ghai and colleagues [5] revealed a remarkable number of similarities with other DCM samples from other oceans [4, 38], dominated with picocyanobacteria. However, they described a local dominance of the Prochlorococcus ecotype HL-I over others, resulting in a niche differentiation between the Mediterranean and the North Pacific Ocean [110]. Besides, other heterotrophic prokaryotes contributing to the DCM community, like the alphaproteobacterial clade SAR11 or planktonic marine Euryarchaeota were detected. Remarkably, from the Mediterranean DCM, novel planktonic microbes were first recovered, as the Ca. Actinomarina minuta [156], Acidimicrobiales [157] or the Ca. Thalassoarchaeota (MG-IIB) [145].





2. OBJECTIVES



There are many previous studies that have characterised the prokaryotic community in the Mediterranean Sea. However, most of them relied in low-resolution molecular biology tools (i.e. 16S rRNA sequencing, denaturing gel gradient electrophoresis or fluorescence in situ hybridization), or only sampled one depth, mainly surface water, ignoring the fact that different ecotypes are distributed through the water column, and their location in the column and abundance varied with the season. Besides, with the application of these techniques, only a small glimpse about the taxonomy of the microbiota was possible, missing information about their metabolism and, hence, their role in the ecosystem.

The aim of this Thesis was to analyse the fine-scale variations in the marine microbiome from the Mediterranean Sea using high-throughput metagenomic sequencing. To achieve this goal:

- We compare the samples collected during summer, when the water column is thermally stratified, against samples retrieved during winter, after the mixing of the water column. Besides, we analyse, in a finer detail, the variation of the microbiota in the uppermost 100 m of the water column.
- We analyse the genomic fragments generated after metagenomic assembly, to find novel microorganisms (prokaryotic and viral sequences) inhabiting in these samples.
- We use genome-resolving metagenomics of these fragments to focus on the assembly of novel metagenome-assembled genomes (MAGs) (i.e. members of the marine group III Euryarchaeota, among others), to analyse their distribution through the water column and in other oceanic regions, their phylogeny and their metabolism.
- We characterise the variation of some functional genes (i.e. rhodopsin) through the water column.





3. MATERIAL AND METHODS



3.1 Sample collection, processing and sequencing

Two sampling sites have been used to recover the microbial diversity off the coast of Alicante, Spain (Figure 7). These places were far enough to avoid the effect of the coast in order to characterise the real prokaryotic community inhabiting in the open sea. In the first sampling site, located at 20 nautical miles off the coast of Alicante (38.07° N, 0.23° W; bottom depth of 200 m), seven metagenomic samples were collected in summer, during a period of strong water stratification. From 2012 to 2015, four samples were retrieved from the DCM (between 55 to 75 m deep). Another sample from the same depth was collected in 2007. Two more samples, at 15 m and 30 m deep, were collected at the same time in 2014. Also, three more samples were collected at two consecutive winters, one sample collected at 20 m deep in December 2013 and two samples at 20 and 80 m deep in January 2015. Thanks to the help of the research vessel "García del Cid", eight samples from different depths were taken in October 2015 from the second sampling site, located at approximately 60 nautical miles off the coast of Alicante (37.35° N, 0.29 °W; bottom depth of 2600 m). Six of them were collected from the uppermost 100 m at 15 m intervals using a hose attached to a CTD (Seabird) connected to a water pump, to directly transfer seawater from the selected depth to the filtration system. The last two samples were collected at 1000 and 2000 m deep in two casts (100 L each) using the CTD rosette. Additionally, on the same cruise, three metatranscriptomes were made from samples collected at 15, 60 and 90 m deep.



Figure 7. Location and number of samples taken from the Mediterranean Sea.

Seawater samples were sequentially filtered on board through 20, 5, and 0.22 μ m pore size polycarbonate filters (Millipore). All filters were immediately frozen on dry ice and stored at -80 °C until processing. For the RNA sample, seawater was filtered onboard through a 0.22 μ m polyethersulfone filter that was suspended with RNAlater and kept on dry ice until storage at -80 °C.

To retrieve the free-living prokaryotic community, one-quarter of the 0.22 µm filter (containing approximately the biomass of 50 L of seawater) was thawed on ice, cut into small pieces with previously autoclaved material, and submerged in 5 mL of lysis buffer (40mM EDTA, 50mM Tris/HCl, 0.75M sucrose). Next, the solution was treated with final concentrations of 1 mg/mL lysozyme during 45 min at 37 °C, followed with 0.2 mg/mL proteinase-K during 60 min at 55 °C. During the process, autoclaved glass beads were added to the solution to mechanically increase, by vortexing, the lysis rate of the biomass. Nucleic acids were extracted from the aqueous phase using phenol-chloroform isoamyl alcohol centrifugation (twice the volume of the lysis solution). A second step of phenol-chloroform isoamyl alcohol followed with chloroform-isoamyl alcohol centrifugation was applied to increase the purity of the recovered DNA. DNA was precipitated overnight using absolute ethanol amended with 0.1 volumes of sodium acetate 3M, followed with a second precipitation of the DNA using a solution containing 70% ethanol. Nucleic acids were concentrated using a vacuum contentrator (miVac, Genevac, UK) and DNA integrity was checked by agarose gel electrophoresis. DNA was quantified with Qubit (ThermoFisher) and sent it to sequence. The sequencing company performed metagenomic library construction and sequencing. Metagenomes from the first and second sampling sites were sequenced using Illumina Hiseq 2000 (100 bp paired-end) (BGI - Hong Kong) and Illumina Hiseq-4000 150 bp paired-end (Macrogen – Republic of Korea), respectively.

3.2 <u>Metagenomic raw-read processing</u>

To remove those reads with errors generated during the sequencing from the bulk of metagenomic raw-reads, trimmomatic [203] was used with the following options (PE, - phred33, ILLUMINACLIP:adapters.fa:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, MINLEN:50) to remove these reads or to trim some regions within the read sequence (low-quality regions or adapters).

3.3 <u>Phylogenetic analysis of the 16S rRNA gene fragments derived from</u> <u>metagenomic reads</u>

To identify candidate 16S/18S rRNA gene fragments within the metagenomic reads, we prepared a non-redundant version of the RDP database [22] by clustering all available 16S/18S rRNA gene reads (*ca.* 2.3 million) into approximately 800,000 clusters at 90% identity level using UCLUST [204]. Using USEARCH [205], a subset of 10 millions of metagenomic reads were aligned to this database (*E*-value < 10^{-5}) and positive hits were considered potential 16S rRNA gene fragments. To further refine sequences from the bulk of the putative rRNA sequences, these candidates were then aligned to archaeal, bacterial, and eukaryal 16S/18S rRNA HMM models [206] using ssu-align [207]. Final 16S/18S rRNA sequences were compared to the entire RDP database and classified into a high-level taxon if the sequence identity was \geq 80% and the alignment length \geq 90 bp. Sequences failing these thresholds were discarded.

3.4 Cross-comparison of metagenomic samples

Two different approaches were used to compare similarities between metagenomic samples. First, a reciprocal global alignment of the short Illumina reads (in subsets of 2 million reads ≥ 50 bp) at $\geq 95\%$ identity was performed using USEARCH6 [205]. The results of the comparison were then clustered with the hclust package in R using a Euclidean distance matrix. In a second approach, subsets of 20 million reads ≥ 50 bp (where applicable) were taxonomically classified against the NR database using DIAMOND [208] with a minimum of 50% identity and 50% alignment. The resulting alignment was later analyzed with MEGAN6 Community Edition [209], and a canonical correspondence analysis (CCA) was inferred with the cluster analysis option and a Bray-Curtis ecological distance matrix.

3.5 "De novo" assembly and gene annotation

"De novo" assembly of the Illumina trimmed-reads into long contigs was performed individually for each metagenome using IDBA_UD [210]. Gene predictions on the assembled contigs were carried out using Prodigal [211]. tRNA and rRNA genes were predicted using tRNAscan-SE [212], ssu-align [207], and meta-RNA [213]. Taxonomic and functional assignment of the predicted protein sequences was performed by comparing against the NCBI-nr database using USEARCH6 [205], and against COG [214] and TIGRFAM [215] using

HMMscan [207]. GC content was calculated using the GeeCee program from the EMBOSS package [216].

3.6 "Binning" and genome reconstruction from metagenomic datasets

To bin the resulting assembled contigs into MAGs, only those longer than 10 Kb were used. Binning is the process of grouping contigs that share similar genomic traits and covariance values across samples into individual genomes. In complex metagenomes, these individual genomes are sometimes referred to as "composite" genomes, due to the contigs are assembled from a population of very closely related organisms. To do that, principal component analysis of tetranucleotide frequencies, GC-content and co-abundance values (retrieved by metagenomic read recruitment) within several samples were used. Moreover, we included in the analysis the taxonomic affiliation of each contig, obtained after the assignment of at least 50% of the genes that shared the same taxonomy at the phylum level (Proteobacteria was divided at the class level). Contigs were classified as unclassified if they did not meet the threshold. Tetranucleotide frequencies were computed using wordfreq program in the EMBOSS package [216] and the principal component analysis using the FactoMineR package [217] in R.

To guarantee the quality of the resulting MAGs, we analysed their completeness, which was estimated by comparison, using HMMscan [207], against three different universal gene sets, one with 35 genes [218], other with 111 genes [219] and another with 52 genes [52]. The degree of contamination was estimated by counting the number of marker genes that appeared more than one time over the total number of genes (in percentage). Also, the software CheckM [220] was also applied to estimate the completeness and degree of contamination. Only MAGs with >50% completeness and <5% contamination were kept for further analyses.

In order to improve the completeness and remove redundancy, a second assembly step was performed combining the genomic fragments with the short paired-end Illumina reads of the metagenomes from which they were assembled. For each genome, we used the BWA aligner [221] with default parameters to retrieve the short-paired reads that mapped onto the contigs. These reads were then pooled and assembled with the contigs using SPAdes [222].

3.7 <u>Phylogenomic classification of the reconstructed genomes</u>

Phylogenomic analysis was used to classify the reconstructed genomes. For each MAG, based on the gene annotation, "close" relatives were downloaded from the NCBI database and

pooled together with other known reference marine microbes. Phylogenomic trees were performed for each one of the class-level taxonomies. To retrieve the maximum number of shared proteins among genomes, we aligned the protein sequences against the COG database [214] using HMMscan. In the case of the composite genomes of marine group III Euryarchaeota, only the ribosomal proteins shared among them were used to classify the sequences phylogenomically. For each genome, shared proteins were concatenated and later aligned using Kalign [223]. A maximum-likelihood tree was then constructed using MEGA [224] with the following parameters: Jones-Taylor-Thornton model, gamma distribution with five discrete categories, and 100 bootstraps. Positions with less than 80% site coverage were eliminated.

3.8 Metagenomic read recruitments

To retrieve the relative coverage (abundance) of a given microbe in a sample, filtered metagenome reads are aligned (mapped) against a reference genome. This process, called metagenomic recruitment, allows to semi-quantify and normalise the abundance of genomes across several metagenomic datasets. In this way, genomes of known marine microbes in pure culture, MAGs or SAGs are used to recruit reads from the samples taken in our two sampling sites, as well as from samples collected elsewhere (i.e. *Tara Oceans* [56] or Malaspina expeditions [55]). To do that, reads are mapped using BLASTN [225], applying a cutoff of 99% nucleotide identity over a minimum alignment length of 50 nucleotides. The high identity (99%) applied here allowed us to recover only the closest bacterial genomes (highly similar clones) to a given reference and, therefore, we limited our analysis to a specific strain rather than the whole species (threshold of 95% nucleotide identity). The resulting numerical value is expressed as the number of reads per kilobase of genome per gigabase of metagenome (RPKG).

3.9 <u>Functional assignment of assembled proteins</u>

Since one part of the assembled contigs remained unbinned, we extracted all proteins encoded within the assembled contigs longer than 1 Kb and, for each metagenome, their putative functionality was inferred against the SEED subsystems [226] and KEGG [227] databases. These databases contain hundreds of thousands of sequences that have been manually curated and classified from general to concrete metabolic and ecologic pathways. Proteins were compared to the SEED database using DIAMOND [208] (blastp option, top hit,

 \geq 50% identity, \geq 50% alignment length, E value < 10⁻⁵). On the other hand, GhostKOALA [228] was used to classify the sequences against the KEGG database.

Also, from the bulk of the whole proteins, some single-gene families with biological interest were analyzed in more detail.

Rhodopsins. 168 rhodopsin sequences were extracted from all the metagenomes from the assembled contigs longer than 5 Kb. We imposed the minimum size of 5 Kb to have enough genes to apply the taxonomic assignment for the contig (see "Binning" and genome reconstruction (MAGs) from metagenomic datasets"). To infer their phylogeny, these sequences were pooled with 100 more rhodopsins of fungal, archaeal, viral, and bacterial origin obtained from databases. Sequences were aligned with MUSCLE [229], and a maximumlikelihood tree was constructed with MEGA (Jones-Taylor-Thornton model, gamma distribution with five discrete categories, and 100 bootstraps, positions with less than 80% site coverage were eliminated). Blue versus green light absorption was determined by looking at the L/Q point mutation in the third transmembrane helix, as described previously [147]. To compare the abundance of microbial rhodopsins with depth, we initially created a database containing our metagenomic rhodopsin sequences and approximately 7,900 rhodopsin genes obtained from the MicRhoDE database [230]. Rhodopsin sequences were recruited against metagenomic reads (in subsets of 20 million sequences) from the different metagenomes using BLASTN [225] (\geq 50 bp alignment, \geq 99% identity). Rhodopsin sequences that recruited \geq 1 RPKG were kept for further analyses.

In parallel, metagenomic reads were compared to the NR database using DIAMOND [208] (blastx option, top hit, \geq 50% identity, \geq 50% alignment length, *E*-value < 10⁻⁵). The abundance of rhodopsin genes in each metagenome was estimated from the number of reads matching rhodopsin sequences in NR, normalised by the number of reads matching the universal single-copy recA/radA sequences and by their respective gene length. Reads matching viral or eukaryotic proteins were not taken into account.

Glycoside Hydrolases. Predicted protein sequences from contigs longer than 5 Kb previously taxonomically classified were compared against the Carbohydrate-Active enZYmes (CAZy) database [231]. Using dbCAN [232], sequences that matched as glycoside hydrolases (GH) with an *E*-value $< 10^{-8}$ were kept for further analyses.



4. RESULTS



4.1 <u>Results derived from the work "Fine metagenomic profile of the</u> <u>Mediterranean stratified and mixed water column"</u>

Summary. In this study, eight samples were taken from a single off-shore location during the thermal stratification of the water column to analyse the fine-scale variations in the marine microbiome. Six were collected from the first 100 m at 15 m depth intervals, while the remaining samples were collected at 1000 and 2000 m deep. Furthermore, to compare the stratified and mixed water columns, two more samples were processed at 20 and 80 m deep during winter. To assess the variations in the community structure, we used genome-resolved metagenomics to measure the recruitment of 94 novel metagenome-assembled genomes (MAGs) and reference genomes at the different depths and conditions (stratified or mixed), at high similarity thresholds. We detected a marked stratification of ecotypes that reflects species adaptation to live at a defined depth range. The majority of microorganisms were confined to discreet horizontal layers of no more than 30 m (stenobathic). Only a few such as members of the SAR11 clade appeared at all depths (eurybathic). Furthermore, we detected a stable component of the photic zone microbial community, which was present regardless of the season or physicochemical parameters. Other microbes were more sensitive and appeared only in a specific season. For instance, during the winter mixing period, only some groups of bloomers such as Pseudomonas were favoured, while the SAR116 clade and some Bacteroidetes and Verrucomicrobia disappeared during the mixing period. Results also indicated a strong specialisation at the functional level for the manipulation and uptake of specific polysaccharides, or the transport of different substrates. Lastly, rhodopsin sequences (green or blue) also showed narrow depth distributions that correlated with the taxonomy of the microbe in which they were found but not with depth.

4.1.1 *The photic zone is delimited into three regions*

A vertical profile showed that, during the thermal stratification, several gradients for physical parameters and chemical compounds appeared within the first 100 meters of the water column (**Annex 1**:Table 1). For instance, the surface water temperature at 15 m was 22.9 °C, which decreased down to 14.5 °C and 13.8 °C at 60 and 90 m, respectively. At 1000m temperature was 13.1 °C. The seasonal thermocline was located between 30 and 45 m deep. The highest values of chlorophyll-a measured between 45 and 60 m deep indicated that the DCM occurred just below the seasonal thermocline. Chlorophyll-a reached 0.8 mg·m⁻³, almost one order of magnitude above those from surface waters and 100 times those from deep (1000

Results

m) waters. Small variations of dissolved O₂ concentration were detected, with a maximum at the DCM (9 mg/L) and a minimum at 2000 m deep (6.1 mg/L). NH₄⁺ concentration peaked at the DCM. Phosphates, nitrites and nitrates increased with depth, with a sharp variation found between 60 and 75 m. Conversely, a maximum of the concentration of total organic carbon (TOC) was found at 15 m (2.43 mgC/L), which decreased down to 0.84 mgC/L at 1000 m.

In the winter sampling, the vertical profile indicated that the seawater column was mixed, and no significant differences in temperature or any other physical or chemical features were observed. Moreover, values were more similar to those found between 45 and 60 m.

Using flow cytometry, we measured the absolute numbers of planktonic picoprokaryotes for the whole water column (**Annex 1**:Figure 1A). In the stratified period, the maximum of *Prochlorococcus* (nearly 3.2×10^4 cells mL⁻¹) and *Synechococcus* (1.35×10^4 cells mL⁻¹) were found in the DCM peak. These values were one to two orders of magnitude higher than those in surface waters. Besides, the distribution of *Prochlorococcus* cells was wider than *Synechococcus*. Heterotrophic bacteria were almost constant through the photic zone, with a slight increase in the DCM. In winter, values for both heterotrophic and autotrophic prokaryotes were highest in the shallowest sample (20 m), though the relative abundance of active heterotrophic bacterioplankton was higher in the deepest sample (80 m).

Using the number of similar reads (> 95% identity) among metagenomes, we examined the relationship between the nine sequenced samples (**Annex 1**:Figure 1B). The stratified samples were clustered by depth, with three main branches corresponding to (i) upper photic (UP, 15 and 30 m), (ii) DCM (45 and 60 m), and (iii) lower photic (LP, 75 and 90 m) layers. Despite the different depths at which the mixed samples (MIX) were obtained (20 and 80 m), both clustered together within the group of DCM samples. As an outgroup appeared, distant to all the photic zone samples, the bathypelagic 1000 m sample. Furthermore, the canonical correspondence analysis (CCA) of the read annotations and environmental parameters confirmed the clustering of samples according to the depth, and MIX samples with DCM group (**Annex 1**:Figure 1C). Inorganic nutrients (such as NOx and $PO4^{3-}$) increased with depth, while ammonia correlated closely with chlorophyll-a, and TOC increased at the surface together with water temperature.

Measurements of Simpson's Diversity Index at both genus and species levels (Annex 1:Figure S3) indicated that bacterial diversity increased continuously with depth only for the stratified season, being relatively constant during winter. GC content also varied with depth (Annex 1:Table 1), being lowest at 15 m (*ca.* 38.6%) and highest at 1000 m (45.9%).

4.1.2 Prokaryotic community structure of the stratified water column derived from 16S rRNA analysis and genome reconstruction and recruitment

Analysis of the metagenome-derived 16S rRNA fragments (Annex 1: Figure 1A and Table S1) revealed broad, depth-dependent variations in taxonomic ranges during stratification in the photic zone. The most striking difference appeared within Archaea, absent in the UP region but represented nearly 16% of the population at 90 m. In the DCM and LP samples, Euryarchaeota remained constant (ca. 5%), while Thaumarchaeota increased from 1% in the 45 m sample to 10% of all the rRNA reads in the 90 m sample. Some members of the bacterial community, such as Actinobacteria, Bacteroidetes, Cyanobacteria and Marinimicrobia, were present in the whole water column, while Deltaproteobacteria, Planctomycetes, Chloroflexi and Acidobacteria had a much more restricted range, appearing only in deeper layers of the photic zone. Verrucomicrobia were present at all depths except in the 45 m sample. Using a finer-scale taxonomic classification of the 16S rRNA sequences, we found that UP (15 and 30 m) Verrucomicrobia belonged to Puniceicoccaceae, whereas the members of Verrucomicrobiaceae were predominantly found below the DCM. The proportion of 16S rRNA gene reads assigned to unclassified bacteria also increased with depth, from 3% at 15 m to more than 10% at 90 m, indicating that a significant fraction of the microbes at the subsurface is still uncharacterized.

Metagenomic assembly and binning of contigs from samples collected in the Mediterranean Sea generated 94 MAGs (>50% completeness and <5% contamination) (Annex 1:Table S2), expanding most of the archaeal and bacterial phyla detected by 16S. These genomes, together with several selected genomes of well-known marine microbes, were used to infer changes in the prokaryotic community through the stratified photic zone. All the genomes recruited much more at one single specific depth and most (*ca.* 70%) recruited only from metagenomes sampled at either one or two consecutive depths (stenobathic) (Annex 1:Figure S13). This result indicates that the distribution of most of these microbes only extends over a 30-m-thick layer within the *ca.* 100 m deep photic zone. Most microbes were preferentially found at the UP or DCM depths except for some archaea. For example, members of the MGI Thaumarchaeota and some groups of Euryarchaea appeared to prefer the LP (Annex 1:Figure S13). For Cyanobacteria, the first 45 m were dominated by the HL-I clade (the pure culture Prochlorococcus MED4 and the MAG Prochlorococcus MED-G72) with a peak in abundance at approximately 30 m, which then decreased below this depth when clade

LL-I Prochlorococcus NATL1A and Prochlorococcus MED-G73) appeared (Annex 1:Figure S13). On the other hand, Synechococcus genomes were not detected deeper than 30 m.

4.1.3 Seasonal variation in the prokaryotic community structure derived from 16S rRNA analysis and genome reconstruction and recruitment

When comparing the late summer stratified and the winter mixed water columns as a whole, 16S rRNA-based analysis did not show any significant changes in the prokaryotic community at this high-level taxonomy (phylum and class levels), and a homogeneous distribution similar to the DCM and LP samples was observed in the MIX samples (**Annex** 1:Figure 1A).

However, an in-depth analysis using genome recruitment indicated that, despite the substantial variability in the physicochemical parameters (i.e. light, temperature and nutrients), only 47% of the recruited genomes were found in both the mixed and the stratified periods (**Annex 1**:Figure 2). Among the groups that were always present, 21 out of 49 (43%) were Alphaproteobacteria and Cyanobacteria (mainly the SAR11 clade and *Synechococcus*, respectively). Some less abundant, but resistant, taxa included members of the Actinobacteria families Acidimicrobiaceae (MedAcidi-G1, G2A, G2B, G3) and *Ca*. Actinomarinaceae, three SAR86 clade genomes within Gammaproteobacteria and the Bacteroidetes family Flavobacteriaceae (**Annex 1**:Figure 2).

Conversely, 53% of the recruited genomes were detected only in one season (Annex 1:Figure 2). Ten genomes, comprising the groups Actinobacteria, Gammaproteobacteria, Verrucomicrobia, Bacteroidetes, and Euryarchaeota were found exclusively during winter (Annex 1:Figure 2). These genomes were characterised by having a large estimated genome size (> 3.0 Mb) and a high GC content (> 50%) (Annex 1:Figure 3). 46 MAGs were only present during stratification, and most of them restricted to the UP region. Many of these MAGs were members of the phyla Bacteroidetes (12 genomes), Verrucomicrobia (4 genomes), members of the SAR116 clade of the Alphaproteobacteria and the OM60/NOR5 clade within the class Gammaproteobacteria. However, members of MG-II archaea, OM182, and SUP05 clades of Gammaproteobacteria, that also disappear in winter, came from deeper layers (DCM and LP) (Annex 1:Figure 2).

4.1.4 *The microbial metabolism is modified by depth and season*

In order to analyse at functional level the microbial community associated with the metagenomes, we used the assembled coding sequences collected from all the contigs > 1 Kb

obtained from the metagenomes against the SEED Subsystems database [226]. Clustering of level 1 subsystems (**Annex 1**:Figure 5) revealed a marked discontinuity between UP samples and the other samples, indicating unique characteristics of surface waters, while once again, the DCM and MIX samples clustered together, demonstrating that they are similar on a functional level. After comparing the number of proteins assigned to each subsystem, we found significant differences (based on the standard deviation among samples) mainly involving carbohydrates, membrane transport, and motility and chemotaxis.

a) Carbohydrates. To study the taxonomical and in-depth distribution of the genes encoding the glycoside hydrolases (GH) family of enzymes, which are involved in the breakdown of complex sugars, we compared all the proteins extracted from contigs larger than 5 Kb assigned to the phyla Actinobacteria, Bacteroidetes, Euryarchaeota, Thaumarchaeota, and Verrucomicrobia, as well as the classes Alphaproteobacteria and Gammaproteobacteria against the CAZy database [231]. The phylogenetic distribution of the CAZy genes was analysed, considering the number of GH per 1000 genes (EQ) and the abundance normalised by the percentage of 16S rRNA gene reads of each group (NORM) (Annex 1: Figure 5B). We could not detect any GH gene in Thaumarchaeota. Bacteroidetes was the group with more enzymes (74.3 GHs/1000 genes) and showed variations of the GH families among samples. For instance, we found that GH from Bacteroidetes from DCM and LP samples grouped and separated from UP, which in turn was close to the MIX samples. Verrucomicrobia represented the second group that included the most significant number of GH genes, with 54.3 GHs/1000 genes analysed. Results showed that the majority of the GHs present in Verrucomicrobia were different from Bacteroidetes, indicating that members of these phyla may be utilising different carbohydrate substrates (Annex 1:Figure 5C). Furthermore, as with Bacteroidetes, we found that the number of GH families was higher in Verrucomicrobia from UP than in DCM and LP. Cyanobacteria, Actinobacteria, Euryarchaeota and Alpha- and Gammaproteobacteria, followed the decreasing trend.

b) Membrane transport. PCA analysis showed that the mixed samples clustered together and separated from the stratified samples, which, in turn, were also clustered by depth for UP and LP samples, while the DCM samples showed a more dispersed distribution (**Annex** 1:Figure S15). In more detail, we found transport systems (ATP-binding cassettes and phosphotransferases) related to iron, phosphonate, polyamines (putrescine/spermidine), oligopeptides, and sugars, and several heavy metal resistances such as the cobalt-zinc-cadmium (CzcA) efflux system enriched in the winter-mixed samples. During the stratification, we found differences in the distribution of membrane transport systems. For instance, we detected in the

LP layer a higher proportion of generalistic ABC di/oligopeptide transporters together with some specific transporters for Archaea (A2 holin family). Conversely, TonB-dependent transporter proteins, as well as choline and betaine uptake proteins, were relatively abundant in UP.

c) Motility and chemotaxis. We detected a significant peak of proteins related to motility and chemotaxis in the UP. Manual inspection revealed enrichment in high GC-content microbes mainly from Alpha- (Sphingomonadadales and SAR116) and Gammaproteobacteria (Oceanospirillales) classes. Remarkably, within the group of MIX samples, bacteria from MedWinter-JAN2015-80m exhibited a significantly large number of genes involved in chemotaxis but not for the biosynthesis of the flagella in comparison with all the other samples.

4.1.5 *Rhodopsins showed a sharp gradient with depth*

Using metagenomic reads, we evaluated the number of reads classified as rhodopsins and calculated their frequency per genome, normalising them by the number of reads annotated as *recA* and *radA* genes (single-copy housekeeping genes) and by their gene length (**Annex** 1:Figure 4B). The total numbers of rhodopsin-assigned reads were correlated to light intensity, with a maximum at 15 m, where *ca*. 65% of the genomes contain rhodopsin, which then decreased with depth. Conversely, for winter samples, the number of rhodopsins was similar regardless of depth throughout the water column.

In another approach, we could assemble 168 rhodopsin genes throughout the stratified water column and 28 rhodopsin genes from the winter samples. Metagenomic recruitment of these genes showed that a total of 105 out of 196 rhodopsin genes (53%) recruited only during stratification, 46% in both seasons, and only one rhodopsin gene in winter. Besides, phylogenetic analysis revealed a broad diversity of this gene family, comprised of at least 11 major groups (**Annex 1**:Figure 4A). These rhodopsin cluster, a separate cluster including only sequences from Bacteroidetes was detected. Moreover, within the clusters, rhodopsin sequences were also grouped by depth, with many branches containing only upper or lower photic zone varieties. This result confirmed the stenobathic character of most groups at the finer level of diversity resolution.

Remarkably, when we recruited these genes together with those downloaded from the MicRhoDE database [230], we could not find any no correlation between the predicted absorption spectrum (blue versus green light) of the rhodopsins and of the depth from which

they recruited the most reads (**Annex 1**:Figure 4C). In contrast, we did see a consistent pattern of correlation between the absorption spectrum and the phylogenetic affiliation of the host genome; while genomes of Bacteroidetes and Actinobacteria all carried green rhodopsins, rhodopsins, the large phylum Proteobacteria mainly have the blue variety.

Lastly, within MAG Verrucomicrobia MED-G86 (3.19 Mb and 55% GC content), we found the unique rhodopsin that recruited only in the MIX samples but not in the stratified. This rhodopsin clustered together with a novel clade of freshwater rhodopsins [114] affiliated closely with the Exiguobacterium rhodopsins (**Annex 1**:Figure S14). Since this was the first marine representative, we searched in the *Tara Oceans* for similar members within this group and eight contigs were retrieved. Two of these contigs came from the Mediterranean Sea (stations 009 and 030), and the remaining six came from the North and South Pacific Oceans (stations 093, 094, 102, 109, 128, and 136). Detailed analysis indicated that within the novel clade, another rhodopsin subcluster was exclusively composed with sequences with low GC values (35 to 40%), evolutionary distant from the Verrucomicrobia MED-G86 and the other marine genomic fragments (45 to 60% GC). Taxonomical annotation of these contigs confirmed that these rhodopsins belonged to the Planctomycetes-Verrucomicrobia-Chlamydiae superphylum.

UNIVERSITAS Miguel Hermändez

4.2 <u>Results derived from the work "Novel Caudovirales associated with</u> Marine Group I Thaumarchaeota assembled from metagenomes"

Summary. In this study, we characterised several uncultivated viruses assembled from metagenomic samples that infect marine Thaumarchaeota, which we designated as "marthavirus". Thaumarchaeota are some of the most abundant autotrophic microorganisms in the deep ocean and responsible for much of the ammonia oxidation occurring in this environment. Therefore, their viruses should play a central role in shaping the biogeochemical cycles of nitrogen and carbon in the ocean. Overall, we could assemble 35 sequences from metagenomic (23 out of 35) and viromic (12 out of 35) datasets. Most of the sequences were obtained from cellular metagenomes confirming that they represent an essential tool to study environmental viral communities due to cells were retrieved while undergoing viral lysis. Phylogenetic analysis placed these novel viral genomes as a single group, distant to other archaeal viral sequences retrieved so far. On the other hand, metagenomic recruitment showed that this viral population is formed by very divergent entities with high intrapopulation homogeneity. Furthermore, metatranscriptomic analyses revealed a differential expression

profile with the capsid as a primary transcript, indicative of viruses during the lytic cycle. The cobalamin biosynthesis gene *cob*S, an auxiliary metabolic gene, was also highly expressed during the infection.

4.2.1 Novel Thaumarchaeota viruses were recovered from metagenomes and viromes

From the previous work [159, 233], more than 1,000 sequences could be classified as viruses. Among them, we detected a 69 Kb contig that had hits to genes encoded in Thaumarchaeota genomes (the majority of these hits were to the genus *Ca*. Nitrosopumilus). However, this contig also had hits to viral-related genes with a very low similarity (32-35%), like the predicted major capsid protein (MCP), portal protein, tail tape measure protein and the large subunit of the viral terminase. These viral-related genes gave hits to a complete unclassified archaeal virus (KY229235) recovered from a metagenomic assembly of a sample \sim 500 m below the seafloor [234].

To expand the repertoire of putative Thaumarchaeota viruses, the MCP, terminase and portal proteins of this new contig recovered from the metagenome Med-OCT2015-90m and KY229235 sequences were used as queries to search against several marine metagenomes and viromes, including the Mediterranean Sea dataset [153, 159], Tara Oceans [56] and Malaspina expeditions [55] and datasets publicly available at the Joint Genome Institute (JGI) database (https:// img.jgi.doe.gov/). In the end, 35 putative viral contigs (Annex 2:Table S1) could be classified and later manually curated to check for similarity to these reference proteins and thaumarchaeal genes. Seventeen sequences were found in the cellular fraction of our dataset (Med-OCT2015-75m and Med-OCT2015-90m), and one in the viral fraction (MedVir-OCT2015-60m). Furthermore, another large batch containing nine genomes was found in viromes from the Chesapeake Bay.

Genomic analyses showed that all of the recovered sequences had a low GC content, which varied from 30 to 37% (average: 33.6%). The genome size ranged from 69,519 to 8,022 bp (average: 21,988 bp), although only two of them (Marthavirus-1 and -2) had repeated sequences with more than 30 nucleotides at the ends of the 5' and 3' regions, indicating that they were complete. Annotation of the coding DNA sequences (CDS) indicated that only 14% of proteins showed significant homology to sequences present in the Prokaryotic Virus Orthologous Groups (pVOG) database. Sequence clustering of the total 1289 genes resulted in 684 protein clusters, where due to the incompleteness of most of the viral sequences, only eleven protein clusters were present in at least half of the sequences and formed the viral "soft"

core (**Annex 2**:Table S2). Five of the "soft" core protein clusters contained proteins involved in DNA metabolism (terminase, RadA, ATPase, PD-D/EXK nuclease and Ribonuclease H) and one sequence was an auxiliary metabolic gene (AMG), *cob*S, which catalyses the final step in cobalamin (vitamin B12) biosynthesis in prokaryotes. The remaining five 'soft' core clusters were hypothetical proteins, and no function could be inferred. Furthermore, no tRNA-encoding sequences or hallmarks of temperate phages, such as integrase or excisionase genes, were detected in any of the recovered genomes.

4.2.2 Phylogenomic analyses revealed a separate lineage from other archaeal Caudovirales

In order to establish the phylogenetic affiliation of these sequences and their relationship with other archaeal virus sequences, the terminase and the MCP proteins, characteristic of Caudovirales viruses, were used. Distant, but homologous genes were found in the reference genomes of the fosmid Oxic1 7 [235] and the provirus Nvie-Pro1 present in the genome of Nitrososphaera viennensis, a soil Thaumarchaeota [236], but could not be identified in the putative thaumarchaeal virus found in the single-cell genome AAA160-J20 [237] or in the putative provirus of Ca. Nitrosopumilus catalina SPOT01 [126]. Results showed a similar phylogenetic pattern for both terminase and MCPs (Annex 2:Figure 1), where the new sequences identified here formed a separate lineage from viruses infecting halophilic and marine mesophilic Euryarchaeota (named as haloviruses and magroviruses, respectively). Only the viral genome KY229235 was found close to the novel sequences, while Oxic1 7 and Nvie-Pro1, which clustered together, were found more closely related to haloviruses. Furthermore, phylogenetic analyses of the viral RadA and PD-D/EXK nuclease genes (both of which are also present in the genomes of Thaumarchaeota cells) (Annex 2: Figure 1) showed that those genes were strongly associated with Thaumarchaeota, and distinct from marine Euryarchaeota and their viruses. Remarkably, all the novel sequences retrieved in this work clustered as a single, monophyletic lineage in all four phylogenetic trees.

4.2.3 Marthaviruses encode for the auxiliary metabolic gene cobS

Pairwise comparison between the two complete genomes (Marthavirus-1 and -2) showed a high degree of divergence ([ANI 70.8%, coverage 6.16%]; [Average Amino Acid Identity (AAI) 53.5%; percentage of common proteins 40.43%]). Nevertheless, the alignment indicated that synteny among genomes was well-preserved, with two conserved regions, and structural region and DNA-related region (**Annex 2**:Figure 2A). These two regions were separated by a variable region, that in some viral sequences encoded the CobS protein. This gene was not detected within the other reference thaumarchaeal viral sequences nor in the magroviruses. Besides, a phylogenetic analysis containing the host and viral CobS protein for Thaumarchaeota and Cyanobacteria showed that Marthavirus-encoded *cob*S are not related to the archaeal *cob*S (**Annex 2**:Figure S4), and viral *cob*S sequences clustered together and separated from their hosts.

4.2.4 Metagenomic recruitments showed a patchy distribution with low intrapopulation diversity

To assess the abundance, distribution and genomic diversity of the novel group of viruses, we performed fragment recruitment analysis by comparing each sequence to 314 metagenomes from Mediterranean, Tara Oceans and Malaspina datasets (cellular and viral fraction) [55, 56, 153, 159, 233] with a sequence identity threshold of 70%. We considered only those samples where these viral genomes recruited more than 10 RPKG. As expected, the marthavirus genomes recruited from metagenomes containing Thaumarchaeal genomes, albeit at significantly lower levels and with more restricted distribution (**Annex 2**:Figure 2B). While reference genomes of Thaumarchaeota were detected in 65% of the metagenomic samples analysed, Marthaviruses were found only in 2% of the metagenomes and 12% of the viromes. Marthaviruses showed a patchy distribution. The majority of the samples where these viruses recruited came from the Mediterranean Sea and the South Atlantic Ocean (**Annex 2**:Figure 2). Interestingly, most of the viral genomes recruited reads at more than 99% nucleotide identity, with minimal coverage below 95% identity (**Annex 2**:Figure S5).

4.2.5 Metatranscriptomics confirmed an active viral replication

From the same seawater sample (Western Mediterranean Sea, 90 m) where we obtained 18 marthavirus genomes, we also performed a metatranscriptome sequencing. These data could provide clues about the prevailing activities during infection. cDNA reads were mapped onto the two complete genomes assembled from this sample (Marthavirus-1 and -2). Most abundant transcripts in both viruses corresponded to the MCPs, which is required for viral assembly (**Annex 2**:Figure 2C). Remarkably, we observed that the cobS, encoded within Marthavirus-1 genome, was also highly expressed in the metatranscriptome. Although no study of the structure and activity of the CobS-like viral proteins has been done, results of the mRNA transcripts indicate that the presence of this gene may have an essential role during the infection process.
Nine marthavirus genomes were recovered from viromic samples from the Chesapeake estuary. Consequently, we used the metagenomic, viromic and metatranscriptomic datasets collected there [238]. Similar results were obtained after analysing the transcripts for the two different genomes (Annex 2:Figure S6). Again, the MCP was the most expressed gene in Marthavirus-4. The *cob*S gene encoded within the Marthavirus-10 genome was expressed as well, although several genes, mostly hypothetical proteins but also an adhesin, which might mediate the virus-host adhesion, and a metallophosphatase were expressed.

4.3 <u>Results derived from the work "New insights into marine group III</u> <u>Euryarchaeota: from dark to light"</u>

Summary. While marine group II Euryarchaeota (MG-II) has been extensively studied, little is known about the ecophysiology and distribution of the less-abundant marine group III Euryarchaeota (MG-III). Recent work from deep hydrothermal vents was able to recover five genomes from this group partially. Using genome assembly from direct metagenome reads and metagenomic fosmid clones, we have identified six novels MG-III MAGs from the photic zone (Epi1-6) and two novel MAGs from deep-sea samples (Bathy1 and -2). Our photic-zone MG-III MAGs corresponded to novel groups with no similarity, and significantly lower GC content when compared with previously described deep-MG-III MAGs. Besides, they encoded for photolyases and rhodopsins genes, that were absent in the MAGs from deeper waters, and are an indicator of their epipelagic habitat. Moreover, they have genes for peptide and lipid uptake and degradation, suggesting a photoheterotrophic lifestyle. Phylogenetic analysis of these photolyases and rhodopsins, as well as their genomic context, suggests that these genes are of bacterial origin, supporting the hypothesis of an MG-III ancestor that lived in the dark ocean. Epipelagic MG-III occurs sporadically and in relatively small proportions in marine plankton, representing only up to 0.6% of the total microbial community reads in metagenomes. Based on differences in genome content and sequence identity, we propose the following nomenclature: Epipelagoarchaeales for the LowGC-MGIII and Bathypelagoarchaeales for the HighGC-MGIII.

4.3.1 Novel MG-III Euryarchaeota genomes were recovered from metagenomes

After an exhaustive search from *Tara* Oceans and our Mediterranean datasets, eight MAGs could be retrieved, accounting to a total of 10.5 Mb of sequence. Six of them, named CG-Epi1 to CG-Epi6, were isolated from metagenomes of the photic zone. The remaining two genomes,

Results

Bathy1 and Bathy2, were isolated from the aphotic zone (**Annex 3**:Table 1 and Table 2). Manual inspection of the differential coverage of the sequences in each bin identified three subgroups very similar to each other (93–96% ANI) within CG-Epi2, referred to as Epi2A, Epi2B and Epi2C (**Annex 3**:Figure S2).

Analysis of the degree of completeness indicated that CG-Epi1 was the most complete with 85% of the whole genome, followed by CG-Epi2 (75%) and the mesopelagic CG-Bathy1 (64%). Conversely, CG-Epi5 was the least complete of them (9.4%). Based on the number of different variants of single-copy genes, all our MAGs contained a single microbial species each. All MG-III bins had low GC content (36–36.8%) except Bathy2 (64.2%). Due to their differences in the GC content, MG-III MAGs were divided into two groups, lowGC and HighGC. Among the lowGC MAGs, ANI varied from 68% to 85.4%, whereas the highGC showed higher degrees of conservation, with ANIs ranging from 89.5% to 96.2% (Annex 3:Figure S4)

4.3.2 Phylogenomic analysis divided MG-III genomes into small subgroups

Phylogenetic trees for the 16S rRNA and 23S rRNA genes were performed (Annex **3**:Figure S5) to assess the phylogenetic relationships among the novel MAGs recovered in this study, together with those recently recovered from the hydrothermal vents of the Guaymas basin and Cayman rise [146]. However, due to the difficulty to assemble and bin these genes into MAGs, most of the assembled genomes did not have any of these gene markers. Therefore, we looked for other housekeeping genes that might be helpful to define the phylogenetic relationships of the novel MG-III with other archaea. We identified and constructed phylogenetic trees for RecA, RpoB, SecY, the geranylgeranylglyceryl phosphate synthase, DnaK and the two gyrase subunits, GyrA and GyrB (Annex 3:Figure S6 to S12). We also performed a phylogenomic analysis, including the concatenated ribosomal proteins shared among the genomes (Annex 3: Figure 1). The phylogenetic analysis of these genes revealed the same topology and clearly showed the split between MG-II and MG-III sequences. Besides, MAGs recovered from the epipelagic zone were divided into two well-defined groups. Accordingly, they were named LowGC1-MGIII, which contained the MAGs CG-Epi1, 3, 4 and 6, and LowGC2-MGIII formed with the MAG CG-Epi2 and the reference genome Guaymas32. Also, a separate clade, containing bins exclusively of bathypelagic origin (CG-Bathy2, Cayman92 and Guaymas31), corresponded with the HighGC-MGIII.

4.3.3 Genomic comparisons confirmed the division of MG-III into small groups

To examine the conservation of synteny across the different genome bins, we performed an all-versus-all genome comparison with the available sequences of MG-III (**Annex 3**:Figure 2A). Results showed the share of large fragments within MAGs of the two groups of LowGC. However, synteny blocks were not conserved between MAGs of LowGC-MGIII and HigGC-MGIII. In the case of LowGC-MGIII, the highest synteny was found between Epi1 and Epi4 (54 block alignments, 62% of Epi4 genome size). For LowGC-MGIII, only Epi2 and Guaymas32 showed a significant synteny (56 block alignments, 38% of CG-Epi2). The low level of synteny between Bathy1 and other bins confirmed that the microbes represented by this bin are very distant to the other LowGC-MGIII. Among the HighGC-MGIII bins, the highest synteny was found between Bathy2 and Guaymas31 (40 block alignments, 42% of CG-Bathy2) followed closely by Cayman92 and Guaymas31 (42 block alignments, 40.8% of the Cayman92 genome).

Moreover, for each MAG, we constructed a non-redundant set of proteins that were used to analyse the relationships among groups by the reciprocal best hit. Only the best hit for each protein (>80% similarity) was retained (**Annex 3**:Figure 2B). This protein content analysis supported the clustering observed in the phylogenomic tree. Bathy1 and SCGC-AAA-288-E19 appeared distantly associated with Guaymas32 and Guaymas31, respectively. MG-III Epi1 with Epi4 had the most significant percentage of shared proteins (34.8%), followed by Epi2B and Guaymas32 (24%) and then Bathy2 and Guaymas31 (25%). Only 8% of Epi1 proteins were conserved in Epi2 and 0.5% in Bathy2.

4.3.4 MG-III shares with MG-II a similar photoheterotrophic lifestyle

In order to infer the metabolic potential of the different MG-III MAGs, the predicted open reading frames were functionally classified according to the arCOG [239] (a specific database for Archaea) categories and their frequencies in the different genomes compared (**Annex 3**:Table S5 to S7).

a) Central carbon metabolism. MAGs harboured enzymes for the glycolysis, the tricarboxylic acid cycle and oxidative phosphorylation, indicating aerobic respiration. We found genes for the complete tricarboxylic acid cycle in LowGC-MGIII, but three genes were absent in Bathy1. Remarkably, only the aconitase and the fumarase were found in Bathy2. MG-III appears to possess most of the enzymes of the Embden-Meyerhof-Parnas (EMP) pathway, except the first and the last enzymes. We found typical gluconeogenesis enzymes such as

phosphoenolpyruvate carboxykinases in the LowGC-MGIII and Bathy1 MAGs, as well as subunits of the pyruvate/oxaloacetate carboxyltransferase in all the MAGs.

Only a small number of amino-acid synthases were found in MG-III: cysteine in Bathy1 and Bathy2, glutamine in LowGC-MGIII, and for glutamate in all MG-III bins. Remarkably, many enzymes for de novo biosynthesis were missing, including those for synthesising methionine, arginine, threonine, histidine, aromatic amino acids and branched amino acids. However, we observed multiple genes related with the uptake and transformation of peptides or amino acids in our MG-III bins, indicating that these organisms are capable of taking up amino acids from the environment and incorporating them into their proteins. In this way, genes for permeases for lysine/arginine (all MAGs), histidine (Bathy2), glutamine (LowGC-MGIII and Bathy1), proline (LowGC-MGIII and Bathy1) and polar amino acids (Bathy2) were detected into their genomes. Furthermore, several ABC-transporter-systems were found for peptides and oligopeptides; for example, Dpp-ABC type dipeptide/oligopeptide transporters (in all of them) and Liv-ABC-type branch amino-acid transporters (only in LowGC-MGIII and Bathy1). Several enzymes involved in the degradation of amino acids were also found, including dehydrogenases for alanine and glutamate (all of them), threonine (LowGC-MGIII and Bathy2) and proline (LowGC-MGIII), as well as several aminotransferases for branchedchain amino acids (LowGC-MGIII and Bathy1) and aspartate/tyrosine/aromatic aminotransferases (LowGC-MGIII and Bathy1). Peptidases also showed differences in the distribution among the MAGs; i.e. dipeptidyl-aminopeptidases and several AprE-like subtilisins (LowGC-MGIII and Bathy1), C1A-peptidases (LowGC-MGIII), C25-peptidases (Bathy1) and Xaa-Pro aminopeptidases (Bathy2).

b) Light-related genes. We detected the presence of photolyases and rhodopsin genes among LowGC-MGIII genomes but not within the deep HighGC-MGIII MAGs (Annex 3:Figure 3). Photolyases are proteins capable of photorepairing ultraviolet-induced pyrimidine dimers in the presence of light (Essen, 2006; Essen and Klar, 2006). Five related genes, a phytoene synthase, a phytoene-desaturase, a histidine kinase, a sugar-epimerase and one hypothetical protein, were found adjacent to the photolyase gene. At the equivalent genomic position, the aphotic Guaymas32 had neither the photolyase nor the associated genes mentioned above. Phylogenetically, MG-III rhodopsins clustered with bacterial proteorhodopsins rather than with the euryarchaeal rhodopsins previously described for MG-II [144, 145] (Annex 3:Figure S14). The analysis of key residues showed that all these MG-III rhodopsins are proton pumps [42] with glutamine (Q) in the characteristic spectral tuning residue site indicating their ability to absorb light from the blue range (Annex 3:Figure S16). c) Structural proteins. Given the incompleteness of the MG-III MAGs, we could only recover some genes involved in cell wall biosynthesis, including several glycosyltransferases (type I/IV), polysaccharide synthases and genes for carbohydrate modification (acyltransferases and aminotransferases). Furthermore, we found several sequences containing two concatenated flaJ genes (implicated in archaeal flagellum assembly) followed by a flaI gene (a transcriptional activator).

d) Other metabolic traits. Lipo-oligosaccharide transport systems (nodI/J-like genes) and phosphonate transporters were found exclusively in the LowGC-MGIII. Besides, multidrug and antimicrobial peptide transporters (ABC-type) together with several permeases for drug/metabolites (RhaT-like family) were abundant in all the genomes.

4.3.5 MG-III represents a cosmopolitan and widespread group

To evaluate the relative abundance of the novel MG-II, I MAGs we used genome recruitment from >200 metagenomic datasets recovered from the open ocean (**Annex 3**:Figure 4 and Table S1). The majority of these samples came from the *Tara Oceans* expedition [56], although it also was included samples from our Mediterranean sampling site [159, 233]. Recruitment values showed a clear correlation of the two MG-III groups with depth. Representatives of LowGC-MGIII were only present in epipelagic collections, while the HighGC-MGIII, Bathyl and Guaymas32 were meso- (>200 m deep) and bathypelagic (>1000 m deep). In the epipelagic zone, the highest abundance was found for CG-Epi1, which accounted for 0.5% of the reads in the sample collected from the Mediterranean station TARA_018. CG-Epi1 seemed to be evenly distributed throughout the photic zone, while the remaining LowGC-MGIII increased their recruitment values at deeper waters within the photic zone (25–155 m, including the DCM).





5. DISCUSSION



5.1 <u>Metagenomics to understand the taxonomical and functional prokaryotic</u> <u>composition of the water column</u>

It is well-known that in temperate waters, the water structure and, therefore, the microbiota change within seasons. In the majority of the cases, these studies have characterised the abundance and variation of the community by (i) using 16S rRNA amplicon sequencing and fluorescence in situ hybridisation (FISH) [12, 202], and by (ii) comparing weekly to monthly samples collected from the same depth [240]. These approaches, however, can bias the detection of certain groups. For instance, it is known that when using FISH, mismatches on the probes can underestimate the abundance of the different prokaryotic groups [12]. Furthermore, variations within the community were predicted at low taxonomical resolution (best case scenario: genus), ignoring the fact that within the same species, different depths, such as *Prochlorococcus* HL and LL ecotypes [241]. On the other hand, during stratification, due to differentially through the water column [4, 49, 242], and microorganisms present at 5 m deep in winter may colonise deeper niches during summer stratification.

In this sense, to assess the in-depth variations in the community structure, we have applied high-throughput metagenomic sequencing over a narrow depth profile in the stratified and mixed water columns. During summer, results showed that samples were clustered by depth, with three main branches corresponding to (i) upper photic (UP, 15 and 30 m), (ii) DCM (45 and 60 m), and (iii) lower photic (LP, 75 and 90 m) layers. Remarkably, the mixed samples (MIX) clustered together within the group of DCM. In our study, DCM and MIX waters have similar physicochemical properties, which can explain the similarity of these two regions.

The clustering of samples into different layers of the water column also has implications in the taxonomic structure, analysed by two approaches, 16S rRNA metagenomic fragments and genome recruitment. The former method has a shallow taxonomic resolution (as mentioned above). Conversely, genome recruitment relies on the availability of genomic sequences to measure their abundance. However, metagenomic assembly is sometimes hampered with the intra-species diversity of a given microbe [26], as it happened to genomes of Cyanobacteria, Thaumarchaeota and Pelagibacterales, which although being the most abundant microbes they assembled poorly in our samples. Nevertheless, independently of the method used, results indicated a depth distribution of the prokaryotic community during summer. Indeed, around 70% of the recruited genomes were only detected from metagenomes collected at either one or

Discussion

two consecutive depths, reflecting their stenobathic nature. This result indicates that the distribution of most of these microbes only extends over a 30-m-thick layer within the first 100 m deep of the photic zone. Besides, most microbes were preferentially found at the UP or DCM depths, like *Synechococcus* that was not detected below 30 m deep; and only a few genomes, including most of the archaeal genomes and the *Prochlorococcus* LL ecotype, recruited below the DCM. The lack of more representatives recruiting at the LP can be explained by the 3-fold increase of the proportion of 16S rRNA genes reads assigned to unclassified and the lower recruitment coverage of all the genomes at 95% nucleotide identity (only 10% of the total community recruited - data not shown), which indicate that a significant fraction of the microbes at lower depths (LP) is still uncharacterized.

Furthermore, as long as the entire water column is taken into account rather than only a single depth, our results indicate that there are no significant changes in the prokaryotic diversity during seasonal fluctuations, and a homogeneous community distribution similar to the DCM and LP samples is observed in the MIX samples. We measured a temporal persistence of some taxa previously considered sporadic or rare. For example, it has been suggested, using pyrosequencing 16S rRNA gene PCR amplicons in a surface sample (3 m depth) in the northwestern Mediterranean Sea, that Thaumarchaeota MG-I and Euryarchaeota MG-IIB populations were more abundant during winter [240] and absent in summer. Our results show that archaea (MG-I, -II and -III) were always present and abundant throughout the water column during the winter but were almost absent in the UP region during the stratification. A similar observation was made using metatranscriptomes from the stratified water column in the Gulf of Agaba/Eilat [243, 244]. As the Planctomycetes or Chloroflexi only appeared below the DCM. This finding highlights the importance of collecting samples at different depths in the water column when comparing seasonal variations and has significant ramifications for global marine studies that most often take samples only from the surface or, at most, from one single subsurface photic zone depth.

The majority of the persistent microbes (summer and winter) represent widespread and abundant (photo)heterotrophic microbes characterised by small genome size and low GC content (and likely a small cell size and more efficient absorption of nutrients), following the streamlining theory [88]. The lower GC content observed in the near-surface stratified water has been suggested to be a natural adaptation to reduce nitrogen demand in these environments with a severe depletion of bioavailable nitrogen [245]. It seems likely that these capabilities allow for better adaptation to overcome the environmental disturbances produced during winter mixing and subsequent phytoplankton blooms. Conversely, almost the other half of the

community was present only during summer, where most of them were found to be restricted to the ultraoligotrophic UP layer. Our results are in agreement with other studies that detected the absence of these resilient microbes during winter when the water column was mixed, and a maximum in mid-summer, mostly limited to surface waters [12–14]. Moreover, one small part of the population (<10% of the genomes) recruited only during winter. Taxonomic and functional classification of these microorganisms designated them as opportunistic prokaryotes (*r*-strategists) that grow rapidly, taking advantage of the mixing of the water column and subsequent upwelling of nutrients. These genomes were characterised by having a large genome size (> 3.0 Mb) and a high GC content (>50%), contrastingly to the persistent prokaryotic community. Additionally, these genomes also possess multiple clusters for degrading a wide range of substrates as well as genes responsible for flagellum biogenesis and motility, which are typical metabolic properties of heterotrophic bacterial communities associated with high nutrient inputs [246].

We also characterised the distribution of rhodopsins in our samples. Since their first discovery in marine bacteria [39], several studies have vertiginously increased the knowledge of their function and diversity [38, 42, 43, 147, 156, 157, 238, 247–249]. Rhodopsins are among the most widespread genes in the photic zone worldwide [40, 44, 250]. They are very diverse and are distributed throughout most taxa. From our data, we could extrapolate that total numbers of rhodopsin-assigned reads were correlated to light intensity, with a maximum at 15 m, where ca. 65% of the genomes contain rhodopsin, followed with a decrease through depth. Our results seem to be different from the situation in the permanently stratified central North Pacific, where the maximum was found at the DCM [250]. However, we believe that our results are in agreement with observations about the microbial community and the physicochemical composition in the UP. It is in this ultraoligotrophic region where bacteria need to cope with the low DOM concentration of these waters. To do that, they rely on solar energy to keep up the energetic balance of the cell. The produced ATP (or even the H⁺ gradient) can be used to fuel the active transport of nutrients, while they use the very low dissolved organic matter to fill the cellular NAD(P)H, which is needed in anabolism (bacterial growth and cell division). For instance, Kim and colleagues [251] constitutively expressed a cyanobacterial rhodopsin in E. coli cells, supplemented with minimal medium with limited glucose as the sole carbon source and operated under three different illumination conditions (always illuminated and switched on and off every 15 min and 12 h). Only in those experiments were illumination was always on, the chemostats were not washed out. Conversely, the DCM is in contact with higher

Discussion

pools of inorganic nutrients from deeper waters and is more appropriate for photosynthetic microbes.

When we phylogenetically classified the assembled rhodopsin sequences, we detected that these genes clustered primarily by phylum and later by depth, with many branches containing only upper or lower photic zone varieties. This result confirms the stenobathic character of most groups at the finer level of diversity resolution of one widespread gene. Moreover, we found a consistent pattern of correlation between the absorption spectrum and the phylogenetic affiliation of the host genome. For instance, while Bacteroidetes and Actinobacteria all carried green rhodopsins, Proteobacteria mostly had the blue variety. The findings suggest, unlike previously reported findings [44, 252] that the spectral tuning of rhodopsins may not be related to depth adaptation but tend to be associated with the classification of the microbe instead.

Lastly, following the same approach, we characterised the changes in the abundance and types of the assembled GH. We detected that, overall, the number of GH genes was similar at the different layers of the water column. However, the GH families varied with depth, suggesting specialisation in the degradation of different polysaccharides that is likely connected with specific groups of algae or particles. Furthermore, between the two most GHencoding groups, Bacteriodetes and Verrucomicrobia, there was a minimal overlap of GH families, indicating that although they share the same niche (most of them are found in the UP region), they might be using different substrates. Remarkably, we found a few GH families in the autotrophic Cyanobacteria, which suggests a new source to obtain organic carbon. Previous studies described that picocyanobacteria (Prochlorococcus and Synechococcus) encode for an ample amount of amino acid, peptide and sugar transporters [253–255], which allow them to uptake organic compounds. This "heterotrophy" can supply the primary autotrophic metabolism (photosynthesis). In the end, it seems that mixotrophy is present in all the marine Synechococcus and Prochlorococcus, and globally distributed in the photic zone of the oceans [255]. Recently, it has been shown that mixotrophy can increase the viability of Prochlorococcus marinus during extended periods of darkness. A coculture with marine copiotroph, Alteromonas macleodii, may be supplying organic compounds to Prochlorococcus [256, 257]. Our results, together with previous studies, highlight the mixotrophic nature of marine picocyanobacteria, as several glycoside hydrolases are encoded in their genomes.

5.2 The role of metagenomics to recover the active-replicating viral diversity

Metagenomics (and viromics) have also shed light into the vast uncultured diversity of phages [86, 89, 172, 173, 175, 258, 259]. While viromics can be helpful to understand the total viral community, metagenomics represents an important tool to study only those viral communities that are undergoing viral lysis (their genomes are within the infected cells). Despite the differences of the filter pore size applied for the collection of viral DNA in these two methods, both serve to retrieve novel phages through metagenomic assembly and, in the end, allow to understand the role of the viral population in the biosphere ("viral shunt") and their metabolic capacities (AMGs). The most notable example of how high-throughput sequencing can provide new information about unknown viruses was the discovery of phages infecting MG-II Euryarchaeota [174]. The lack of MG-II isolates, as well as for other well-known uncultured marine microbes (i.e. *Ca.* Actinomarina minuta or the SAR86 clade, among others) has hampered the retrieval of many phages infecting them. This is still true even for cultured microbes, such as SAR11 or SAR116 clades, due to the fact that they are hard to get in pure culture and only a small number of phages have been isolated (18 phages in SAR11 [260, 261] and only one phage for SAR116 [183]).

However, for Thaumarchaeota, one of the most abundant phyla in the deep ocean [129, 130] and with some representatives in pure culture [95, 100, 126], to date, no marine virus has yet been isolated. Only two putative viral sequences had previously been retrieved by singlecell genomics [237], fosmid libraries [235], together with one putative provirus within the genome of Ca. Nitrosomarinus catalina SPOT01 [126]. Very recently, the discovery of a set of metagenomic sequences [262] (we named them marthavirus (MARine THAumarchaeota VIRUS)) has largely improved our knowledge of these viruses. Within these novel viral genomes, the presence of terminase, major capsid, prohead and portal proteins seemed to affiliate them to the head-tailed Caudovirales group, representing the first member of this group described so far for the crenarchaeal superphylum [263, 264]. Besides, phylogenetic analyses of common proteins among our marthavirus sequences and some other archaeal viruses, including some representatives from the phylum Euryarchaeota, indicated that they formed a monophyletic clade representing a separate lineage from haloviruses and magroviruses and evolutionarily distinct from the previous putative Thaumarchaeota viruses. Furthermore, host and viral genome recruitments in 314 samples (at 70% nucleotide identity) encompassing cellular and viral fraction, several seas and oceans, and different depths (photic and aphotic), showed differences in the number of samples that were present host and viruses

Discussion

(64% and 14%, respectively). This striking difference in numbers suggests that our set of viruses does not extend to the whole marthavirus community, which is likely composed by a vast population with low intra-population diversity, given that most of the viral genomes recruited reads at more than 99% nucleotide identity with minimal coverage below 95% identity, but with significant divergence among groups.

Interestingly, some of the marthaviruses seemed to encode the AMG cobalamin biosynthesis protein CobS in the variable region of the genome, which is involved in the final step of the cobalamin biosynthesis. Cobalamin (vitamin B₁₂) plays a vital role as a cofactor in the synthesis of amino acids or DNA (i.e. cobalamin-dependent methionine synthase and ribonucleotide reductase – RNR, respectively), as well as in other metabolic pathways [135]. However, its synthesis seems to be restricted to only a few taxa, including autotrophic microorganisms such as the Thaumarchaeota or Cyanobacteria [135]. Given the role of this essential cofactor in the central metabolism, the non-synthesizer organisms depend exclusively on the availability of cobalamin (or its precursor) in seawater [135]. Some studies have reported a relationship between the availability of vitamin B₁₂ and the distribution and growth of phytoplankton and bacterioplankton blooms [265]. This gene, as it happened in some marthaviruses, has also been found in cyanophages, where it has been theorised that its role could be potentially associated with RNR during the infection process [266]. However, phylogenetic analysis showed that *cobS* encoded in cyanophages and marthaviruses were not related to any cyanobacterial and archaeal *cobS*. Besides, given the low similarity between the viral and prokaryotic CobS proteins, no data is supporting the idea that viral cobS genes share the same function. Although undetected in our genomes, Ahlgren et al. reported, at the same time than us, the existence of another distinct set of putative virus-infecting Thaumarchaeota encoding for the AMG ammonia monooxygenase C (amoC) [182]. This enzyme is implicated in the aerobic oxidation of ammonia to nitrite, which fuels the thaumarchaeal cells with energy and reducing power [95]. However, in this case, this protein shares a high amino acid sequence similarity (>90%) to the cellular Thaumarchaeota, suggesting that the viral protein has the same function [182]. Nevertheless, our metatranscriptome of the cellular fraction (0.22-5 μ m) indicated that the *cobS* gene is actively transcribing during infection, as its transcription was coupled with the MCP, which in cyanophages it transcribes during the last step of the infection [267]. Therefore, this gene may have an essential role during the infection process that is thus far unclear.

5.3 <u>Metagenomics recovery of novel genomes of the uncultured marine group</u> <u>III Euryarchaeota</u>

One of the main topics of this Thesis has been the recovery of novel MAGs belonging to marine group III Euryarchaeota. As it has been widely used by other authors [29, 156, 157, 219], metagenomic assembly and subsequent binning of genomic fragments is a successful strategy for the discovery of novel microbial lineages. Indeed, this bioinformatics, culture-free approach allowed to recover the first two genomes of their sister clade, the yet uncultured MG-II [144, 145], which very recently has been largely expanded with >200 new genomes [63, 64]. In this sense, from samples collected at different depths and in different places, we were able to assemble eight MAGs belonging to MG-III. Analysis of the phylogeny, conserved synteny and pairwise comparison of non-redundant sets of proteins indicated three divergent groups. Six of them belonged to a novel lineage assembled from epipelagic waters, while the remaining two came from deeper waters in the aphotic zone. One of them (CG-Bathy2), composed only by fosmids coming from the KM3 (Mediterranean Sea, 3000 m deep), was related with a set of MG-III genomes retrieved from the very cold and deep ocean (Guaymas basin and Cayman ridge, >2000 m deep) [146]. Conversely, given that most of the genomic fragments included in CG-Bathy1 came from a sample collected at 600 m deep in the Mediterranean (highly saline, relatively warm and extremely oligotrophic) might explain the appearance of CG-Bathy1 as a separate basal branch from the photic and aphotic MG-III genomes, the low level of synteny and gene content among this MAG and the other genomes.

It is noteworthy that, except CG-Epi1, which seems to be a cosmopolitan microbe, the remaining MAGs were assembled with contigs/fosmids mostly coming from one sample. This might be due to the low abundance of these groups in the oceans that can hamper their assembly. Thus, only when these microbes are only abundant enough at specific sites (endemism) or during a transient environmental condition causing a significant growth (i.e. bacterial or phytoplankton blooms, replenish of nutrients after water mixing or upwelling of nutrients, among others), they can be assembled. Previous studies based on 16S rRNA surveys detected MG-III at very low abundance in the ocean [138, 150, 151, 268], and only in a small number of studies, MG-II was present in an unusual much higher proportion [140, 151, 152]. Contrastingly, compared to MG-III, their sister clade MG-II is, very often, more abundant in surface and deep waters [138, 139, 145, 146, 148, 269] and seems to be easily assembled from metagenomes. Thus, it might explain the astounding difference in the number of genomes publicly available in the NCBI (296 genomes of MG-II vs 19 genomes of MG-III, June 2019).

Discussion

Along these lines, our recruitment analysis indicated that the MG-III MAGs analysed during this Thesis i) were not well-represented in metagenomes from cold water regions, such as polar regions, the Baltic Sea or the northeast subarctic Pacific [158, 270, 271], while previous studies showed the presence of MG-III in these habitats; ii) they were partitioned into two very well-defined niches, while MAGs (CG-Epi1 to -6) assembled from the epipelagic ocean were only present in the photic zone and CG-Bathy1 and CG-Bathy2 were clearly meso-or bathypelagic; and iii) they represented a very small fraction, accounting, in the best-case scenario, up to 0.5 (CG-Epi1) and 1% (CG-Bathy2) of the prokaryotic community in the photic and aphotic regions, respectively. Our results confirmed the previous idea that MG-III are relatively minor components of the archaeal communities in the photic zones.

Previous studies have correlated the abundance of MG-II with eukaryotic microorganisms. For instance, an experimental study demonstrated that eukaryotic phytoplankton additions could stimulate the growth of MG-II in bottle incubations [272]. Furthermore, a metagenomics study analysed the succession of the marine microbiota, detecting the growth of MG-II after a phytoplankton bloom where diatoms, small flagellates and picophytoplankton dominated consecutively [273]. Thus, to know whether MG-II and MG-III respond to similar blooming patterns, we measured the recruitment of MG-II in the same samples which MG-III were assembled. However, we could not find any relationship between MG-II and MG-III, indicating that, despite being closely related and using similar substrates (see below), MG-II and MG-III do not bloom concurrently.

Inference of the functional metabolism has been extensively studied for MG-II [63, 64, 144, 145]. They are (photo)heterotrophs, encoding for metabolic functions characteristic of heterotrophs, including glycolysis, a TCA cycle, and electron transport chain (ETC). They can deal with extracellular carbohydrates, proteins and fatty acids, as they encode for hydrolytic enzymes. By contrast, given that the aforementioned low abundance of MG-III, the metabolism of MG-III was previously understudied. For instance, the analysis of three fosmids from the deep Mediterranean showed the presence of some fermentation-related genes, which led to the hypothesis that they could be facultative anaerobes [139]. However, recent analyses, including this Thesis [146, 153], suggest that they are aerobic heterotrophs. Besides, they encode for the same main metabolic pathways (glycolysis, TCA cycle, ETC) and extracellular enzymes that their sister clade MG-II, indicating that both groups (MG-II and MG-III) are functionally equivalent in the marine biosphere, mainly modifying the carbon cycle. Despite their similarity, they encode for different families of extracellular peptidases and GHs, which suggest a niche differentiation among members of the same family [63, 64, 153] or between MG-II and MG-

III [146]. This differences in the nutrient uptake and degradation might explain their variability in their recruitment and may be the reason for MG-II and MG-III not blooming concurrently.

Remarkably, as it happened in MG-II [63, 64], epipelagic MG-III encoded for photolyase and rhodopsin, genes that appeared absent in the aphotic MAGs. The phylogenetic relationships of these genes, together with the multiple putative horizontal gene transfer (HGT) events observed in the nearby genes, led us to hypothesise an ancestral "dark nature" for MG-III. In this way, these genes would have been "recently" transferred from mesophilic epipelagic bacteria to MG-III, probably long after the massive HGT events that have been detected before the diversification of several mesophilic archaeal clades but with sharing some metabolic traits among MG-II and MG-III [136, 141]. Therefore, the acquisition of proteorhodopsins, together with ultraviolet-protection photolyases, would have promoted a better adaptation to the oligotrophic surface waters allowing MG-III clades to expand into new photic niches.







6. CONCLUSIONS



- Physicochemical and metagenomic analyses detected a strong stratification of the water column during summer. In this sense, the first 100 m were divided into three regions: UP, DCM and LP. Conversely, during winter, water column represented a single mixed region (MIX).
- 2. Comparison of the 16S rRNA gene fragments derived from metagenomic reads concluded that the prokaryotic community during the summer stratification differed, at the level of phylum, along the three regions. The main difference appeared within Archaea, absent in the UP but present at significant numbers in the DCM and LP.
- 3. High-stringent metagenomic recruitment of reference genomes and MAGs indicated a stenobathic distribution of most microbes, covering a 30-m-thick layer of seawater. Therefore, vertical distribution should be considered as the major element to study the prokaryotic community, rather than comparing single depth samples collected from different places.
- 4. The comparison of the 16S rRNA gene fragments between summer and winter did not indicate any marked difference in the prokaryotic diversity as long as the entire water column is taken into account, rather than a single depth. Besides, genome recruitment indicated that nearly half of the community was resistant to the seasonal variation, including members that were mistakenly believed to disappear during winter.
- 5. The number of rhodopsin genes was correlated to light intensity, where more than half of the community in the UP contained a rhodopsin. Furthermore, the spectral tuning of rhodopsins was not related to depth adaptation. Instead, it was associated with the taxonomy of the microorganism.
- 6. Marthavirus sequences were phylogenetically classified as *bona fide* Caudoviralesinfecting Thaumarchaeota. These genomes showed a sparse distribution in several metagenomic and viromic datasets. However, most of the metagenomic reads recruited at more than 99% nucleotide identity, suggesting that marthaviruses may form a population with low intra-population diversity, but with significant divergence among groups.

- 7. Marthaviruses encoded for a *cob*S, distantly related to the archaeal *cob*S. Although its function has not been characterised, metatranscriptomic analysis showed an active expression of this gene during infection.
- 8. Using genome-resolved metagenomics, eight new groups of planktonic marine group III Euryarchaeota spanning different genera and families could be retrieved.
- 9. Genome comparisons between MG-III members showed a marked differentiation at taxonomical and functional levels of MAGs assembled from photic or aphotic samples. Genomes were then condensed into two big groups: Epipelagoarchaeota and Bathyarchaeota.
- 10. Genome analyses indicated that MG-III had a heterotrophic lifestyle, with most of the metabolic pathways and protein families for nutrient uptake and degradation shared with the sister clade MG-II.
- 11. The presence of photolyases and rhodopsin genes encoded in Epipelagoarchaeota supported the hypothesis that they were *bona fide* epipelagic microbes. The acquisition of these genes stimulated a better adaptation to the oligotrophic surface waters.



7. REFERENCES



- 1. Pinet R. Invitation to Oceanography, Seventh Ed. 2016. Jones & Bartlett Learning.
- 2. Nellemann C, Hain S, Alder J, United Nations Environment Programme., GRID--Arendal. In dead water : merging of climate change with pollution, over-harvest, and infestations in the world's fishing grounds. 2008. UNEP.
- 3. Cullen JJ. Subsurface Chlorophyll Maximum Layers: Enduring Enigma or Mystery Solved? *Ann Rev Mar Sci* 2015; 7: 207–239.
- 4. Delong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N, et al. Community Genomics among microbial assemblages in the Ocean's Interior. *Science (80-)* 2006; **311**: 496–503.
- 5. Ghai R, Martin-Cuadrado A-BB, Molto AG, Heredia IG, Cabrera R, Martin J, et al. Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME J* 2010; 4: 1154–1166.
- 6. Sigman DM, Hain MP. The Biological Productivity of the Ocean | Learn Science at Scitable. *Nat Educ Knowl* 2012.
- Giovannoni SJ, Vergin KL. Seasonality in Ocean Microbial Communities. Science (80-) 2012; 335: 671–676.
- 8. Fernández-Gómez B, Richter M, Schüler M, Pinhassi J, Acinas SG, González JM, et al. Ecology of marine Bacteroidetes: a comparative genomics approach. *ISME J* 2013; 7: 1026–37.
- Hou S, López-Pérez M, Pfreundt U, Belkin N, Stüber K, Huettel B, et al. Benefit from decline: the primary transcriptome of Alteromonas macleodii str. Te101 during Trichodesmium demise. *ISME J* 2018.
- 10. Shade A, Peter H, Allison SD, Baho DL, Berga M, B??rgmann H, et al. Fundamentals of microbial community resistance and resilience. *Front Microbiol* 2012; **3**: 1–19.
- Morris RM, Vergin KL, Cho J-C, Rappé MS, Carlson C a., Giovannoni SJ. Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Timeseries Study site. *Limnol Oceanogr* 2005; 50: 1687–1696.
- 12. Alonso-Sáez L, Balagué V, Sà EL, Sánchez O, González JM, Pinhassi J, et al. Seasonality in bacterial diversity in north-west Mediterranean coastal waters: Assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiol Ecol* 2007; **60**: 98–112.
- 13. Eilers H, Pernthaler J, Peplies J, Glöckner FO, Gerdts G, Amann R. Isolation of Novel Pelagic Bacteria from the German Bight and Their Seasonal Contributions to Surface Picoplankton. *Appl Environ Microbiol* 2001; **67**: 5134–5142.
- 14. Yan S, Fuchs BM, Lenk S, Harder J, Wulf J, Jiao NZ, et al. Biogeography and phylogeny of the NOR5/OM60 clade of Gammaproteobacteria. *Syst Appl Microbiol* 2009; **32**: 124–139.
- 15. Grimes DJ, Atwell RW, Brayton PR, Palmer LM, Rollins DM, Roszak DB, et al. The fate of enteric pathogenic bacteria in estuarine and marine environments. *Microbiol Sci* 1986; **3**: 324–9.
- 16. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* 1985; **82**: 6955–9.
- 17. Pace NR, Stahl DA, Lane DJ, Olsen GJ. The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences. 1986. Springer, Boston, MA, pp 1–55.
- 18. Stahl DA, Lane DJ, Olsen GJ, Pace NR. Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Appl Environ Microbiol* 1985; **49**: 1379–84.
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem Biol* 1998; 5: R245– R249.
- 20. Breitwieser FP, Lu J, Salzberg SL. A review of methods and databases for metagenomic classification

and assembly. Brief Bioinform 2017.

- 21. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome* 2015; **3**: 31.
- 22. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2014; **42**: 633–642.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* 2013; 41: D590– D596.
- 24. Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, Barbe V, et al. Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci* 2003; **100**: 10020–10025.
- 25. Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, et al. Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature* 2003; **424**: 1042–1047.
- Grote J, Thrash JC, Huggett MJ, Cameron Thrash J, Huggett MJ, Landry ZC, et al. Streamlining and Core Genome Conservation among Highly Divergent Members of the SAR11 Clade. *MBio* 2012; 3: 1– 13.
- López-Pérez M, Gonzaga A, Martin-Cuadrado AB, López-García P, Rodriguez-Valera F, Kimes NE. Intra- and Intergenomic Variation of Ribosomal RNA Operons in Concurrent Alteromonas macleodii Strains. *Microb Ecol* 2013; 65: 720–730.
- 28. Eloe-Fadrosh EA, Ivanova NN, Woyke T, Kyrpides NC. Metagenomics uncovers gaps in ampliconbased detection of microbial diversity. *Nat Microbiol* 2016; **1**: 15032.
- 29. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 2015; **523**: 208–211.
- 30. Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. Mesophilic crenarchaeota: Proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* 2008; **6**: 245–252.
- 31. DeLong EF. Archaea in coastal marine environments. PNAS 2006; 89: 5685–5689.
- Fuhrman JA, McCallum K. Novel major archaebacterial group from marine plankton. *Nature* 1992;
 356: 148–149.
- Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 1990; 345: 60–63.
- 34. Fuhrman JA, McCallum K, Davis AA. Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific oceans. *Appl Environ Microbiol* 1995; **61**: 4517.
- 35. Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol S, Malits A, et al. Changes in Bacterioplankton Composition under Different Phytoplankton Regimens. *Appl Environ Microbiol* 2004; **70**: 6753–6766.
- 36. Schmidt TM, DeLong EF, Pace NR. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* 1991; **173**: 4371–4378.
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial Diversity in the Deep Sea and the Underexplored 'Rare Biosphere'. *Proceedings of the National Academy of Sciences*. 2006. National Academy of Sciences, pp 12115–12120.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen J a, et al. Environmental Genome Sshotgun Sequencing of the Sargasso Sea. *Science* 2004; **304**: 66–74.
- 39. Béjà O, Suzuki MT, Hadd A, Nguyen LP, Spudich JL, Spudich EN, et al. Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea. *Science (80-)* 2000; **1902**: 1902–1907.
- 40. Fuhrman J a, Schwalbach MS, Stingl U. Proteorhodopsins: an array of physiological roles? *Nat Rev Microbiol* 2008; **6**: 488–494.
- 41. Kandori H. Ion-pumping microbial rhodopsins. Front Mol Biosci 2015; 2.

- 42. Inoue K, Ono H, Abe-Yoshizumi R, Yoshizawa S, Ito H, Kogure K, et al. A light-driven sodium ion pump in marine bacteria. *Nat Commun* 2013; **4**: 1678.
- 43. Gushchin I, Chervakov P, Kuzmichev P, Popov AN, Round E, Borshchevskiy V, et al. Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. *Proc Natl Acad Sci* 2013; **110**: 12631–12636.
- Pinhassi J, DeLong EF, Béjà O, González JM, Pedrós-Alió C. Marine Bacterial and Archaeal Ion-Pumping Rhodopsins: Genetic Diversity, Physiology, and Ecology. *Microbiol Mol Biol Rev* 2016; 80: 929–54.
- 45. Wrighton KH. Antibacterial drugs: Discovering antibiotics through soil metagenomics. *Nat Rev Drug Discov*. 2018., **17**: 240
- 46. Mullany P. Functional metagenomics for the investigation of antibiotic resistance. *Virulence* . 2014. Taylor & Francis. , **5**: 443–447
- 47. Healy FG, Ray RM, Aldrich HC, Wilkie AC, Ingram LO, Shanmugam KT. Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl Microbiol Biotechnol* ; **43**: 667–74.
- 48. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 2013; **499**: 431–437.
- 49. Walsh EA, Kirkpatrick JB, Rutherford SD, Smith DC, Sogin M, D'Hondt S. Bacterial diversity and community composition from seasurface to subseafloor. *ISME J* 2015; **10**: 1–11.
- 50. Haroon MF, Thompson LR, Parks DH, Hugenholtz P, Stingl U. A catalogue of 136 microbial draft genomes from Red Sea metagenomes. *Sci Data* 2016; **3**: 160050.
- 51. Sangwan N, Xia F, Gilbert JA. Recovering complete and draft population genomes from metagenome datasets. *Microbiome* 2016; **4**: 1–11.
- 52. Narasingarao P, Podell S, Ugalde J a, Brochier-Armanet C, Emerson JB, Brocks JJ, et al. De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J* 2012; **6**: 81–93.
- 53. Mende DR, Bryant JA, Aylward FO, Eppley JM, Nielsen T, Karl DM, et al. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat Microbiol* 2017.
- Thorpe J, Stewart C, Venter JECEC, Smith H, Nealson K, Eisen JA, et al. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol* 2007; 5: 0398– 0431.
- 55. Duarte CM. Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition. *Limnol Oceanogr Bull*. 2015. Wiley-Blackwell., **24**: 11–14
- 56. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Ocean plankton. Structure and function of the global ocean microbiome. *Science* 2015; **348**: 1261359.
- 57. Tully BJ, Sachdeva R, Graham ED, Heidelberg JF. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ* 2017; **5**: e3558.
- 58. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci Data* 2018; **5**: 170203.
- 59. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes.
- 60. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2017; **903**: 1–10.
- 61. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018; **36**: 996.

- 62. Hawley AK, Nobu MK, Wright JJ, Durno WE, Morgan-Lang C, Sage B, et al. Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients. *Nat Commun* 2017; **8**: 1507.
- 63. Tully BJ. Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. *Nat Commun* 2019; **10**: 271.
- 64. Rinke C, Rubino F, Messer LF, Youssef N, Parks DH, Chuvochina M, et al. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. nov.). *ISME J* 2019; **13**: 663–675.
- 65. Mehrshad M, Rodriguez-Valera F, Amoozegar MA, López-García P, Ghai R. The enigmatic SAR202 cluster up close: Shedding light on a globally distributed dark ocean lineage involved in sulfur cycling. *ISME J* 2018; **12**: 655–668.
- 66. Landry Z, Swa BK, Herndl GJ, Stepanauskas R, Giovannoni SJ. SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. *MBio* 2017; **8**: e00413-17.
- 67. Becker JW, Hogle SL, Rosendo K, Chisholm SW, Becker JW, Chisholm SW. Co-culture and biogeography of Prochlorococcus and SAR11 2 3. 2018; 1–69.
- 68. Kraemer S, Ramachandran A, Colatriano D, Lovejoy C, Walsh DA. Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. *bioRxiv* 2019; 517433.
- 69. Delmont TO, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappe MS, et al. The global biogeography of amino acid variants within a single SAR11 population is governed by natural selection. *bioRxiv* 2017; 170639.
- 70. Hoarfrost A, Nayfach S, Ladau J, Yooseph S, Arnosti C, Dupont CL, et al. Global ecotypes in the ubiquitous marine clade SAR86. *bioRxiv* 2019; 635185.
- 71. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, et al. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* 2015; **15**: 141–61.
- 72. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. *Science* 2014; **344**: 416–20.
- 73. Cameron TJ, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF, et al. Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J* 2014; **8**: 1440–1451.
- Zaremba-Niedzwiedzka K, Viklund J, Zhao W, Ast J, Sczyrba A, Woyke T, et al. Single-cell genomics reveal low recombination frequencies in freshwater bacteria of the SAR11 clade. *Genome Biol* 2013; 14: 1–14.
- 75. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. *Science (80-)* 2009; **323**: 133–138.
- Loman NJ, Watson M. Successful test launch for nanopore sequencing. *Nat Methods* 2015; 12: 303–304.
- 77. Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang X-J, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research* 2017; **6**: 100.
- 78. Laver T, Harrison J, O'Neill P a., Moore K, Farbos a., Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif* 2015; **3**: 1–8.
- 79. Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics, Proteomics Bioinforma*. 2015. Elsevier., **13**: 278–289
- 80. López-Pérez M, Rodriguez-Valera F. Pangenome Evolution in the Marine Bacterium Alteromonas.
- 81. Rodriguez-Valera F, Martin-Cuadrado A-B, López-Pérez M. Flexible genomic islands as drivers of genome evolution. *Curr Opin Microbiol* 2016; **31**: 154–160.
- 82. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat Biotechnol*

. 2015. Nature Publishing Group. , **33**: 296–302

- 83. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF, et al. Genomic Islands and the Ecology and Evolution of Prochlorococcus. *Science (80-)* 2006; **311**: 1768–1770.
- 84. Watson M, Warr A. Errors in long-read assemblies can critically affect protein prediction. *Nat Biotechnol* 2019; **37**: 124–126.
- 85. Singer E, Bushnell B, Coleman-Derr D, Bowman B, Bowers RM, Levy A, et al. High-resolution phylogenetic microbial community profiling. *ISME J* 2016; **10**: 2020–2032.
- 86. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, et al. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* 2019; 7: e6800.
- 87. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* 1998; **95**: 6578–83.
- 88. Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for microbial ecology. *ISME J* . 2014. Nature Publishing Group. , **8**: 1553–1565
- 89. Mizuno CM, Ghai R, Saghaï A, López-García P, Rodriguez-Valeraa F. Genomes of abundant and widespread viruses from the deep ocean. *MBio* 2016; 7: 1–9.
- 90. Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF, et al. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J* 2009; **3**: 93–105.
- 91. Scanlan DJ. Marine picocyanobacteria. *Ecology of Cyanobacteria II: Their Diversity in Space and Time*. 2012. Springer Netherlands, Dordrecht, pp 503–533.
- 92. Eiler A, Mondav R, Sinclair L, Fernandez-Vidal L, Scofield DG, Schwientek P, et al. Tuning fresh: Radiation through rewiring of central metabolism in streamlined bacteria. *ISME J* 2016; **10**: 1902–1914.
- 93. Klappenbach JA, Dunbar JM, Schmidt TM. rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* 2000; **66**: 1328–1333.
- 94. Williams PJ leB., Thomas DN (David N, Reynolds CS. Phytoplankton productivity : carbon assimilation in marine and freshwater ecosystems. 2002. Blackwell Science.
- 95. Könneke M, Bernhard AE, De La Torre JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 2005; **437**: 543–546.
- Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D, et al. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 2011; 333: 1296– 300.
- 97. De La Rocha CL, Passow U. Factors influencing the sinking of POC and the efficiency of the biological carbon pump. *Deep Sea Res Part II Top Stud Oceanogr* 2007; **54**: 639–658.
- 98. Tripp HJ, Kitner JB, Schwalbach MS, Dacey JWH, Wilhelm LJ, Giovannoni SJ. SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* 2008; **452**: 741–744.
- 99. Gunson JR, Spall SA, Anderson TR, Jones A, Totterdell IJ, Woodage MJ. Climate sensitivity to ocean dimethylsulphide emissions. *Geophys Res Lett* 2006; **33**: L07701.
- Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, et al. Genomic and proteomic characterization of "*Candidatus* Nitrosopelagicus brevis": An ammonia-oxidizing archaeon from the open ocean. *Proc Natl Acad Sci* 2015; **112**: 1173–1178.
- Tsementzi D, Wu J, Deutsch S, Nath S, Rodriguez-R LM, Burns AS, et al. SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature* 2016; 536: 179–183.
- 102. Partensky F, Hess WR, Vaulot D. Prochlorococcus, a marine photosynthetic prokaryote of global significance. *MicrobiolMol BiolRev* 1999; **63**: 106–127.
- 103. Partensky F, Blanchot J, Vaulot D. Differential distribution and ecology of Prochlorococcus and

Synechococcus in oceanic waters : a review. Bull l'Institut océanographique 1999; 19: 457-475.

- 104. Waterbury JB, Watson SW, Guillard RRL, Brand LE. Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* . 1979. Nature Publishing Group. , **277**: 293–294
- 105. Moore LR, Chisholm SW. Photophysiology of the marine cyanobacterium Prochlorococcus: Ecotypic differences among cultured isolates. *Limnol Oceanogr* 1999; **44**: 628–638.
- 106. Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, et al. Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. *PLoS Genet* 2007; **3**: 2515–2528.
- Six C, Thomas JC, Garczarek L, Ostrowski M, Dufresne A, Blot N, et al. Diversity and evolution of phycobilisomes in marine Synechococcus spp.: A comparative genomics study. *Genome Biol* 2007; 8: R259.
- 108. Stomp M, Huisman J, De Jongh F, Veraart AJ, Gerla D, Rijkeboer M, et al. Adaptive divergence in pigment composition promotes phytoplankton biodiversity. *Nature* 2004; **432**: 104–107.
- 109. Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, et al. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* 2008; **9**: R90.
- 110. Farrant GK, Doré H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M, et al. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci* 2016; **113**: E3365–E3374.
- 111. Martiny AC, Huang Y, Li W. Occurrence of phosphate acquisition genes in Prochlorococcus cells from different ocean regions. *Environ Microbiol* 2009; **11**: 1340–1347.
- 112. Martiny AC, Kathuria S, Berube PM. Widespread metabolic potential for nitrite and nitrate assimilation among Prochlorococcus ecotypes. *Proc Natl Acad Sci* 2009; **106**: 10787–10792.
- 113. Giovannoni SJ. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Ann Rev Mar Sci* 2017; **9**: 231–255.
- Cabello-Yeves PJ, Zemskay TI, Rosselli R, Coutinho FH, Zakharenko AS, Blinov V V, et al. Genomes of novel microbial lineages assembled from the sub-ice waters of Lake Baikal. *Appl Environ Microbiol* 2018; 84.
- 115. García-Martínez J, Rodríguez-Valera F. Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine Archaea of Group I. *Mol Ecol* 2000; **9**: 935–48.
- Carlson C a, Morris R, Parsons R, Treusch AH, Giovannoni SJ, Vergin K. Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J* 2009; 3: 283–295.
- 117. Brown M V, Lauro FM, DeMaere MZ, Muir L, Wilkins D, Thomas T, et al. Global biogeography of SAR11 marine bacteria. *Mol Syst Biol* 2012; **8**: 595.
- 118. Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, et al. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 2002; **420**: 806–810.
- 119. Rappé MS, Connon SA, Vergin KL, Giovannoni SJ. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 2002; **418**: 630–633.
- 120. Malmstrom RR, Kiene RP, Cottrell MT, Kirchman DL. Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic ocean. *Appl Environ Microbiol* 2004; **70**: 4129–35.
- Malmstrom RR, Cottrell MT, Elifantz H, Kirchman DL. Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Appl Environ Microbiol* 2005; 71: 2979–2986.
- 122. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome streamlining in a cosmopolitan oceanic bacterium. *Science (80-)* 2005; **309**: 1242–5.
- 123. Tripp HJ. The unique metabolism of SAR11 aquatic bacteria. J Microbiol 2013; 51: 147–153.

- 124. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 1977; **74**: 5088–90.
- 125. Delong EF. Archaea in coastal marine environments. Proc Natl Acad Sci USA 1992; 89: 5685–5689.
- 126. Ahlgren NA, Chen Y, Needham DM, Parada AE, Sachdeva R, Trinh V, et al. Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ Microbiol* 2017; **19**: 2434–2452.
- 127. Mehrshad M, Amoozegar MA, Ghai R, Shahzadeh Fazeli SA, Rodriguez-Valera F. Genome reconstruction from metagenomic data sets reveals novel microbes in the brackish waters of the Caspian Sea. *Appl Environ Microbiol* 2016; 82: 1599–1612.
- 128. Luo H, Tolar BB, Swan BK, Zhang CL, Stepanauskas R, Ann Moran M, et al. Single-cell genomics shedding light on marine Thaumarchaeota diversification. *ISME J* 2014; **8**: 732–736.
- 129. Fuhrman JA, Ouverney CC. Marine microbial diversity studied via 16S rRNA sequences: Cloning results from coastal waters and counting of native archaea with fluorescent single cell probes. *Aquat Ecol* 1998; **32**: 3–15.
- 130. Karner MB, Delong EF, Karl DM. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 2001; **409**: 507–510.
- 131. Stewart FJ, Ulloa O, Delong EF. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ Microbiol* 2012; **14**: 23–40.
- 132. Baker BJ, Lesniewski RA, Dick GJ. Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J* 2012; **6**: 2269–2279.
- 133. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, et al. Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* . 2010. Nature Publishing Group. , **8**: 447–460
- 134. Santoro AE, Casciotti KL. Enrichment and characterization of ammonia-oxidizing archaea from the open ocean: phylogeny, physiology and stable isotope fractionation. *ISME J* 2011; **5**: 1796–1808.
- 135. Doxey AC, Kurtz DA, Lynch MDJ, Sauder LA, Neufeld JD. Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME J* 2015; **9**: 461–471.
- 136. López-García P, Zivanovic Y, Deschamps P, Moreira D. Bacterial gene import and mesophilic adaptation in archaea. *Nat Rev Microbiol* 2015; **13**: 447–56.
- 137. Reysenbach AL, Liu Y, Banta AB, Beveridge TJ, Kirshtein JD, Schouten S, et al. A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. *Nature* 2006; **442**: 444–447.
- Massana R, DeLong EF, Pedrós-Alió C. A Few Cosmopolitan Phylotypes Dominate Planktonic Archaeal Assemblages in Widely Different Oceanic Provinces. *Appl Environ Microbiol* 2000; 66: 1777– 1787.
- 139. Martin-Cuadrado A-B, Rodriguez-Valera F, Moreira D, Alba JC, Ivars-Martínez E, Henn MR, et al. Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *ISME J* 2008; 2: 865–886.
- Belmar L, Molina V, Ulloa O. Abundance and phylogenetic identity of archaeoplankton in the permanent oxygen minimum zone of the eastern tropical South Pacific. *FEMS Microbiol Ecol* 2011; 78: 314–326.
- 141. Deschamps P, Zivanovic Y, Moreira D, Rodriguez-Valera F, López-García P. Pangenome evidence for extensive inter-domain horizontal transfer affecting lineage-core and shell genes in uncultured planktonic Thaumarchaeota and Euryarchaeota. *Genome Biol Evol* 2014; **6**: 1549–1563.
- 142. Frigaard NU, Martinez A, Mincer TJ, DeLong EF. Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* 2006; **439**: 847–850.
- 143. Moreira D, Rodríguez-Valera F, López-García P. Analysis of a genome fragment of a deep-sea uncultivated Group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several

energy metabolism genes. Environ Microbiol 2004; 6: 959-969.

- 144. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. Untangling genomes from metagenomes: Revealing an uncultured class of marine euryarchaeota. *Science (80-)* 2012; **335**: 587–590.
- 145. Martin-Cuadrado A-B, Garcia-Heredia I, Moltó AG, López-Úbeda R, Kimes N, López-García P, et al. A new class of marine Euryarchaeota group II from the mediterranean deep chlorophyll maximum. ISME J 2014; 1–16.
- 146. Li M, Baker BJ, Anantharaman K, Jain S, Breier J a, Dick GJ. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat Commun* 2015; Accepted: 1–6.
- 147. Man D, Wang W, Sabehi G, Aravind L, Post AF, Massana R, et al. Diversification and spectral tuning in marine proteorhodopsins. *EMBO J* 2003; **22**: 1725–1731.
- 148. Zhang CL, Xie W, Martin-Cuadrado AB, Rodriguez-Valera F. Marine Group II Archaea, potentially important players in the global ocean carbon cycle. *Front Microbiol* . 2015. Frontiers. , **6**: 1108
- Galand PE, Gutiérrez-Provecho C, Massana R, Gasol JM, Casamayor EO. Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). *Limnol Oceanogr* 2010; 55: 2117–2125.
- 150. López-Garcia P, López-López A, Moreira D, Rodriguez-Valera F. Diversity of free-living prokaryotes from a deep-sea site at the Antarctic Polar Front. *FEMS Microbiol Ecol* 2001; **36**: 193–202.
- 151. Galand PE, Casamayor EO, Kirchman DL, Potvin M, Lovejoy C. Unique archaeal assemblages in the arctic ocean unveiled by massively parallel tag sequencing. *ISME J* 2009; **3**: 860–869.
- 152. Quaiser A, Zivanovic Y, Moreira D, López-García P. Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. *ISME J* 2011; **5**: 285–304.
- 153. Haro-Moreno JM, Rodriguez-Valera F, López-García P, Moreira D, Martin-Cuadrado A-B. New insights into marine group III Euryarchaeota, from dark to light. *ISME J* 2017; 1–16.
- 154. Morris RM, Frazar CD, Carlson CA. Basin-scale patterns in the abundance of SAR11 subclades, marine Actinobacteria (OM1), members of the Roseobacter clade and OCS116 in the South Atlantic. *Environ Microbiol* 2012; **14**: 1133–1144.
- 155. Treusch AH, Vergin KL, Finlay LA, Donatz MG, Burton RM, Carlson CA, et al. Seasonality and vertical structure of microbial communities in an ocean gyre. *ISME J* 2009; **3**: 1148–1163.
- 156. Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F. Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci Rep* 2013; **3**: 2471.
- 157. Mizuno CM, Rodriguez-Valera F, Ghai R. Genomes of planktonic acidimicrobiales: Widening horizons for marine actinobacteria by metagenomics. *MBio* 2015; **6**: e02083-14.
- 158. Alonso-Saez L, Waller AS, Mende DR, Bakker K, Farnelid H, Yager PL, et al. Role for urea in nitrification by polar marine Archaea. *Proc Natl Acad Sci* 2012; **109**: 17989–17994.
- 159. Haro-Moreno JM, López-Pérez M, de la Torre JR, Picazo A, Camacho A, Rodriguez-Valera F. Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* 2018; **6**: 128.
- Cardman Z, Arnosti C, Durbin A, Ziervogel K, Cox C, Steen AD, et al. Verrucomicrobia are candidates for polysaccharide-degrading bacterioplankton in an Arctic fjord of Svalbard. *Appl Environ Microbiol* 2014; 80: 3749–3756.
- Glaubitz S, Kießlich K, Meeske C, Labrenz M, Jürgens K. SUP05 Dominates the gammaproteobacterial sulfur oxidizer assemblages in pelagic redoxclines of the central baltic and black seas. *Appl Environ Microbiol* 2013; **79**: 2767–2776.
- 162. Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG, et al. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* 2009; **326**: 578–582.

- 163. Jiao N, Zhang F, Hong N. Significant roles of bacteriochlorophylla supplemental to chlorophylla in the ocean. *ISME J* 2010; **4**: 595–597.
- 164. Dupont CL, Rusch DB, Yooseph S, Lombardo MJ, Alexander Richter R, Valas R, et al. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* 2012; **6**: 1186–1199.
- 165. Oh HM, Kwon KK, Kang I, Kang SG, Lee JH, Kim SJ, et al. Complete genome sequence of 'Candidatus puniceispirillum marinum' IMCC1322, a representative of the SAR116 clade in the Alphaproteobacteria. *J Bacteriol* 2010; **192**: 3240–3241.
- 166. Choi DH, Park KT, An SM, Lee K, Cho JC, Lee JH, et al. Pyrosequencing revealed sar116 clade as dominant dddp-containing bacteria in oligotrophic nw pacific ocean. *PLoS One* 2015; **10**.
- 167. Suttle CA. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* 2007; **5**: 801–12.
- 168. Cochlan W, Wikner J, Steward G, Smith D, Azam F. Spatial distribution of viruses, bacteria and chlorophyll a in nentic, oceanic and estuarme environments. *Mar Ecol Prog Ser* 1993; **92**: 77–87.
- 169. Bergh Ø, BØrsheim KY, Bratbak G, Heldal M. High abundance of viruses found in aquatic environments. *Nature* 1989; **340**: 467–468.
- 170. Suttle CA, Chan AM, Cottrell MT. Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* 1990; **347**: 467–469.
- 171. Noble R, Fuhrman J. Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat Microb Ecol* 1998; **14**: 113–118.
- 172. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the Marine Virosphere Using Metagenomics. *PLoS Genet* 2013; **9**: e1003987.
- López-Pérez M, Haro-Moreno JM, Gonzalez-Serrano R, Parras-Moltó M, Rodriguez-Valera F. Genome diversity of marine phages recovered from Mediterranean metagenomes: Size matters. *PLoS Genet* 2017; 13: e1007018.
- 174. Philosof A, Yutin N, Flores-Uribe J, Sharon I, Koonin E V., Béjà O. Novel Abundant Oceanic Viruses of Uncultured Marine Group II Euryarchaeota. *Curr Biol* 2017; **27**: 1362–1368.
- 175. Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* 2017; 8: 15892.
- Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pasić L, Thingstad TF, Rohwer F, et al. Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* 2009; 7: 828– 36.
- 177. Wilhelm SW, Suttle CA. Viruses and Nutrient Cycles in the Sea. *Bioscience* 1999; 49: 781–788.
- 178. Breitbart M. Marine Viruses: Truth or Dare. Ann Rev Mar Sci 2011; 4: 425-448.
- 179. Tsai A-Y, Gong G-C, Huang Y-W. Importance of the Viral Shunt in Nitrogen Cycling in Synechococcus Spp. Growth in Subtropical Western Pacific Coastal Waters. *Terr Atmos Ocean Sci* 2014; 25: 839.
- Rohwer F, Segall A, Steward G, Seguritan V, Breitbart M, Wolven F, et al. The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. *Limnol Oceanogr* 2000; 45: 408–418.
- Mann NH, Cook A, Millard A, Bailey S, Clokie M. Bacterial photosynthesis genes in a virus. *Nature* 2003; 424: 741.
- Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA. Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode amoC nitrification genes. *ISME J* 2018; 1.
- 183. Kang I, Oh H-M, Kang D, Cho J-C. Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc Natl Acad Sci* 2013; **110**: 12343–12348.

- 184. Koskella B, Brockhurst MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev* 2014; **38**: 916–931.
- 185. Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brüssow H. Phage as agents of lateral gene transfer. *Curr Opin Microbiol* . 2003. Elsevier Current Trends. , **6**: 417–424
- 186. Weitz JS, Wilhelm SW. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol Rep* 2012; **4**: 17.
- 187. Tanhua T, Hainbucher D, Schroeder K, Cardin V, Álvarez M, Civitarese G. The Mediterranean Sea system: A review and an introduction to the special issue. *Ocean Sci*. 2013., **9**: 789–803
- 188. Emeis KC, Struck U, Schulz HM, Rosenberg R, Bernasconi S, Erlenkeuser H, et al. Temperature and salinity variations of Mediterranean Sea surface waters over the last 16,000 years from records of planktonic stable oxygen isotopes and alkenone unsaturation ratios. *Palaeogeogr Palaeoclimatol Palaeoecol* 2000; **158**: 259–280.
- 189. Martín-Cuadrado AB, López-García P, Alba JC, Moreira D, Monticelli L, Strittmatter A, et al. Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. *PLoS One* 2007; **2**.
- 190. Bosc E, Bricaud A, Antoine D. Seasonal and interannual variability in algal biomass and primary production in the Mediterranean Sea, as derived from 4 years of SeaWiFS observations. *Global Biogeochem Cycles* 2004; **18**: n/a-n/a.
- 191. Uitz J, Stramski D, Gentili B, D'Ortenzio F, Claustre H. Estimates of phytoplankton class-specific and total primary production in the Mediterranean Sea from satellite ocean color observations. *Global Biogeochem Cycles* 2012; **26**: 1–10.
- 192. Robinson AR, Leslie WG, Theocharis A, Lascaratos A. Mediterranean Sea Circulation. *Encycl Ocean Sci Second Ed* 2008; 710–725.
- 193. Lavigne H, D'Ortenzio F, Ribera D'Alcalà M, Claustre H, Sauzède R, Gacic M. On the vertical distribution of the chlorophyll a concentration in the Mediterranean Sea: A basin-scale and seasonal approach. *Biogeosciences* 2015; **12**: 5021–5039.
- 194. Denaro G, Valenti D, La Cognata A, Spagnolo B, Bonanno A, Basilone G, et al. Spatio-temporal behaviour of the deep chlorophyll maximum in Mediterranean Sea: Development of a stochastic model for picophytoplankton dynamics. *Ecol Complex* 2013; **13**: 21–34.
- 195. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science (80-)* 2012; **336**: 608–611.
- 196. Mignot A, Claustre H, Uitz J, Poteau A, D'Ortenzio F, Xing X. Understanding the seasonal dynamics of phytoplankton biomass and the deep chlorophyll maximum in oligotrophic environments: A Bio-Argo float investigation. *Global Biogeochem Cycles* 2014; 28: 856–876.
- 197. Marty J-C, Chiavérini J, Pizay M-D, Avril B. Seasonal and interannual dynamics of nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED time-series station (1991– 1999). Deep Sea Res Part II Top Stud Oceanogr 2002; 49: 1965–1985.
- Martin-Cuadrado AB, Ghai R, Gonzaga A, Rodriguez-Valera F. CO dehydrogenase genes found in metagenomic fosmid clones from the deep Mediterranean Sea. *Appl Environ Microbiol* 2009; 75: 7436– 7444.
- Schauer M, Massana R, Pedrós-Alió C. Spatial differences in bacterioplankton composition along the Catalan coast (NW Mediterranean) assessed by molecular fingerprinting. *FEMS Microbiol Ecol* 2000; 33: 51–59.
- 200. Feingersch R, Suzuki MT, Shmoish M, Sharon I, Sabehi G, Partensky F, et al. Microbial community genomics in eastern Mediterranean Sea surface waters. *ISME J* 2010; **4**: 78–87.
- 201. Dubinsky V, Haber M, Burgsdorf I, Saurav K, Lehahn Y, Malik A, et al. Metagenomic analysis reveals unusually high incidence of proteorhodopsin genes in the ultraoligotrophic Eastern Mediterranean Sea. *Environ Microbiol* 2017; **19**: 1077–1090.

- 202. Schauer M, Balagué V, Pedrós-Alió C, Massana R. Seasonal changes in the taxonomic composition of bacterioplankton in a oligotrophic coastal system. *Aquat Microb Ecol* 2003; **31**: 163–174.
- 203. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.
- 204. Robert C Edgar. UCLUST: Extreme high-speed sequence clustering , alignment and database search. 2010.
- 205. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010; **26**: 2460–2461.
- 206. Eddy SR. Multiple alignment using hidden Markov models. *Proc Int Conf Intell Syst Mol Biol* 1995; **3**: 114–120.
- 207. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011; 7: e1002195.
- 208. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015; **12**: 59–60.
- Huson DH, Beier S, Flade I, G??rska A, El-Hadidi M, Mitra S, et al. MEGAN Community Edition -Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput Biol* 2016; 12.
- 210. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012; **28**: 1420–1428.
- 211. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**: 119.
- 212. Lowe TM, Eddy SR. TRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1996; **25**: 955–964.
- Huang Y, Gilna P, Li W. Identification of ribosomal RNA genes in metagenomic fragments. Bioinformatics 2009; 25: 1338–1340.
- Tatusov RL, Natale DA, Garkavtsev I V, Tatusova TA, Shankavaram UT, Rao BS, et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 2001; 29: 22–8.
- 215. Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, et al. TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* 2001; **29**: 41–43.
- Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 2000; 16: 276–277.
- Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. J Stat Softw 2008; 25: 1–18.
- 218. Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P. Prediction of effective genome size in metagenomic samples. *Genome Biol* 2007; **8**: R10.
- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* 2013; **31**: 533–8.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; 25: 1043–55.
- 221. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.
- 222. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 2012; 19: 455–477.

- 223. Lassmann T, Sonnhammer ELL. Kalign--an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* 2005; **6**: 298.
- 224. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016; **33**: msw054.
- 225. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res*. 1997., 25: 3389– 3402
- 226. Overbeek R, Begley T, Butler RM, Choudhuri J V, Chuang HY, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 2005; **33**: 5691–5702.
- 227. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016; **44**: D457–D462.
- 228. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* . 2016., **428**: 726–731
- 229. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; **32**: 1792–1797.
- 230. Boeuf D, Audic S, Brillet-Guéguen L, Caron C, Jeanthon C. MicRhoDE: A curated database for the analysis of microbial rhodopsin diversity and evolution. *Database* 2015; **2015**.
- 231. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014; **42**.
- 232. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. DbCAN: A web resource for automated carbohydrateactive enzyme annotation. *Nucleic Acids Res* 2012; **40**.
- López-Pérez M, Haro-Moreno JM, Gonzalez-Serrano R, Parras-Moltó M, Rodriguez-Valera F. Genome diversity of marine phages recovered from Mediterranean metagenomes: Size matters. *PLOS Genet* 2017; 13: e1007018.
- 234. Nigro OD, Jungbluth SP, Lin HT, Hsieh CC, Miranda JA, Schvarcz CR, et al. Viruses in the oceanic basement. *MBio* 2017; **8**: e02129-16.
- 235. Chow CET, Winget DM, White RA, Hallam SJ, Suttle CA. Combining genomic sequencing methods to explore viral diversity and reveal potential virus-host interactions. *Front Microbiol* 2015; **6**: 1–15.
- 236. Krupovic M, Spang A, Gribaldo S, Forterre P, Schleper C. A thaumarchaeal provirus testifies for an ancient association of tailed viruses with archaea. *Biochem Soc Trans* 2011; **39**: 82–88.
- 237. Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, et al. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J* 2015; **9**: 2386–2399.
- Maresca JA, Miller KJ, Keffer JL, Sabanayagam CR, Campbell BJ. Distribution and diversity of rhodopsin-producing microbes in the Chesapeake Bay. *Appl Environ Microbiol* 2018; 84: AEM.00137-18.
- 239. Makarova KS, Galperin MY, Koonin E V. Comparative genomic analysis of evolutionarily conserved but functionally uncharacterized membrane proteins in archaea: Prediction of novel components of secretion, membrane remodeling and glycosylation systems. *Biochimie* 2015; **118**: 302–312.
- 240. Hugoni M, Taib N, Debroas D, Domaizon I, Jouan Dufournel I, Bronner G, et al. Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proc Natl Acad Sci US A* 2013; **110**: 6004–9.
- Garczarek L, Dufresne A, Rousvoal S, West NJ, Mazard S, Marie D, et al. High vertical and low horizontal diversity of Prochlorococcus ecotypes in the Mediterranean Sea in summer. *FEMS Microbiol Ecol* 2007; 60: 189–206.
- 242. Ferreira AJSS, Siam R, Setubal JC, Moustafa A, Sayed A, Chambergo FS, et al. Core microbial functional activities in ocean environments revealed by global metagenomic profiling analyses. *PLoS*
One 2014; 9.

- 243. Miller DR, Pfreundt U, Elifantz H, Hess WR, Berman-Frank I. Microbial metatranscriptomes from the thermally stratified Gulf of Aqaba/Eilat during summer. *Mar Genomics* 2017; **32**: 23–26.
- Miller D, Pfreundt U, Hou S, Lott SC, Hess WR, Berman-Frank I. Winter mixing impacts gene expression in marine microbial populations in the Gulf of Aqaba. *Aquat Microb Ecol* 2017; 80: 223– 242.
- 245. Luo H, Thompson LR, Stingl U, Hughes AL. Selection maintains low genomic GC content in marine SAR11 lineages. *Mol Biol Evol* 2015; **32**: 2738–2748.
- 246. Buchan A, LeCleir GR, Gulvik CA, González JM. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol*. 2014. Nature Publishing Group., **12**: 686– 698
- 247. Inoue K, Ito S, Kato Y, Nomura Y, Shibata M, Uchihashi T, et al. A natural light-driven inward proton pump. *Nat Commun* 2016; 7: 13415.
- 248. Ugalde JA, Podell S, Narasingarao P, Allen EE. Xenorhodopsins, an enigmatic new class of microbial rhodopsins horizontally transferred between archaea and bacteria. *Biol Direct* 2011; **6**: 52.
- 249. Pushkarev A, Béjà O. Functional metagenomic screen reveals new and diverse microbial rhodopsins. *ISME J* 2016; **10**: 2331–2335.
- 250. Olson DK, Yoshizawa S, Boeuf D, Iwasaki W, DeLong EF. Proteorhodopsin variability and distribution in the North Pacific Subtropical Gyre. *ISME J* 2018.
- 251. Kim HA, Kim HJ, Park J, Choi AR, Heo K, Jeong H, et al. An evolutionary optimization of a rhodopsin-based phototrophic metabolism in Escherichia coli. *Microb Cell Fact* 2017; **16**: 111.
- Sabehi G, Kirkup BC, Rozenberg M, Stambler N, Polz MF, Béjà O. Adaptation and spectral tuning in divergent marine proteorhodopsins from the eastern Mediterranean and the Sargasso Seas. *Isme J* 2007; 1: 48–55.
- 253. Munoz-Marin M d. C, Luque I, Zubkov M V, Hill PG, Diez J, Garcia-Fernandez JM. Prochlorococcus can use the Pro1404 transporter to take up glucose at nanomolar concentrations in the Atlantic Ocean. *Proc Natl Acad Sci* 2013; **110**: 8597–8602.
- 254. Zubkov M V., Tarran GA, Fuchs BM. Depth related amino acid uptake by Prochlorococcus cyanobacteria in the Southern Atlantic tropical gyre. *FEMS Microbiol Ecol* 2004; **50**: 153–161.
- 255. Yelton AP, Acinas SG, Sunagawa S, Bork P, Pedrós-Alió C, Chisholm SW. Global genetic capacity for mixotrophy in marine picocyanobacteria. *ISME J* 2016; 1–12.
- 256. Coe A, Ghizzoni J, LeGault K, Biller S, Roggensack SE, Chisholm SW. Survival of Prochlorococcus in extended darkness. *Limnol Oceanogr* 2016; **61**: 1375–1388.
- 257. Biller SJ, Coe A, Roggensack SE, Chisholm SW. Heterotroph Interactions Alter Prochlorococcus Transcriptome Dynamics during Extended Periods of Darkness. *mSystems* 2018; **3**: e00040-18.
- 258. Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan MB, et al. Numerous cultivated and uncultivated viruses encode ribosomal proteins. *Nat Commun* 2019; **10**: 752.
- 259. Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, et al. Environmental Viral Genomes Shed New Light on Virus-Host Interactions in the Ocean. *mSphere* 2017; **2**: e00359-16.
- 260. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11 viruses in the ocean. *Nature* 2013; **494**: 357–360.
- 261. Zhao Y, Qin F, Zhang R, Giovannoni SJ, Zhang Z, Sun J, et al. Pelagiphages in the Podoviridae family integrate into host genomes. *Environ Microbiol* 2018; **21**: 1989–2001.
- 262. López-Pérez M, Haro-Moreno JM, de la Torre JR, Rodriguez-Valera F. Novel Caudovirales associated with Marine Group I Thaumarchaeota assembled from metagenomes. *Environ Microbiol* 2018.
- 263. Pietilä MK, Demina TA, Atanasova NS, Oksanen HM, Bamford DH. Archaeal viruses and

bacteriophages: Comparisons and contrasts. Trends Microbiol 2014; 22: 334-344.

- 264. Rachel R, Bettstetter M, Hedlund BP, Häring M, Kessler A, Stetter KO, et al. Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Arch Virol* 2002; **147**: 2419–2429.
- 265. Sañudo-Wilhelmy SA, Gobler CJ, Okbamichael M, Taylor GT. Regulation of phytoplankton dynamics by vitamin B12. *Geophys Res Lett* 2006; **33**: L04604.
- 266. Helliwell KE, Lawrence AD, Holzer A, Kudahl UJ, Sasso S, Kräutler B, et al. Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin B12. *Curr Biol* 2016; **26**: 999–1008.
- 267. Doron S, Fedida A, Hernndez-Prieto MA, Sabehi G, Karunker I, Stazic D, et al. Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* 2016; **10**: 1437–1455.
- 268. López-García P, Moreira D, López-López A, Rodríguez-Valera F. A novel haloarchaeal-related lineage is widely distributed in deep oceanic regions. *Environ Microbiol* 2001; **3**: 72–78.
- 269. Herndl GJ, Reinthaler T, Teira E, van Aken H, Veth C, Pernthaler A, et al. Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. *Appl Environ Microbiol* 2005; **71**: 2303–9.
- Larsson J, Celepli N, Ininbergs K, Dupont CL, Yooseph S, Bergman B, et al. Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* 2014; 8: 1892–1903.
- Allers E, Wright JJ, Konwar KM, Howes CG, Beneze E, Hallam SJ, et al. Diversity and population structure of Marine Group A bacteria in the Northeast subarctic Pacific Ocean. *ISME J* 2013; 7: 256– 268.
- 272. Orsi WD, Smith JM, Wilcox HM, Swalwell JE, Carini P, Worden AZ, et al. Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *ISME J* 2015; 1–17.
- 273. Needham DM, Fuhrman JA. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat Microbiol* 2016; 1: 16005.



8. ANNEX 1



RESEARCH





Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment

Jose M. Haro-Moreno^{1†}, Mario López-Pérez^{1†}, José R. de la Torre², Antonio Picazo³, Antonio Camacho³ and Francisco Rodriguez-Valera^{1*}

Abstract

Background: The photic zone of aquatic habitats is subjected to strong physicochemical gradients. To analyze the fine-scale variations in the marine microbiome, we collected seven samples from a single offshore location in the Mediterranean at 15 m depth intervals during a period of strong stratification, as well as two more samples during the winter when the photic water column was mixed. We were able to recover 94 new metagenome-assembled genomes (MAGs) from these metagenomes and examine the distribution of key marine microbes within the photic zone using metagenomic recruitment.

Results: Our results showed significant differences in the microbial composition of different layers within the stratified photic water column. The majority of microorganisms were confined to discreet horizontal layers of no more than 30 m (stenobathic). Only a few such as members of the SAR11 clade appeared at all depths (eurybathic). During the winter mixing period, only some groups of bloomers such as *Pseudomonas* were favored. Although most microbes appeared in both seasons, some groups like the SAR116 clade and some Bacteroidetes and Verrucomicrobia seemed to disappear during the mixing period. Furthermore, we found that some microbes previously considered seasonal (e.g., Archaea or Actinobacteria) were living in deeper layers within the photic zone during the stratification period. A strong depth-related specialization was detected, not only at the taxonomic level but also at the functional level, even within the different clades, for the manipulation and uptake of specific polysaccharides. Rhodopsin sequences (green or blue) also showed narrow depth distributions that correlated with the taxonomy of the microbe in which they were found but not with depth.

Conclusions: Although limited to a single location in the Mediterranean, this study has profound implications for our understanding of how marine microbial communities vary with depth within the photic zone when stratified. Our results highlight the importance of collecting samples at different depths in the water column when comparing seasonal variations and have important ramifications for global marine studies that most often take samples from only one single depth. Furthermore, our perspective and approaches (metagenomic assembly and recruitment) are broadly applicable to other metagenomic studies.

Keywords: Photic zone, Deep chlorophyll maximum, Mediterranean, Stratification, Stenobathic

* Correspondence: frvalera@umh.es

[†]Jose M. Haro-Moreno and Mario López-Pérez contributed equally to this work. ¹Evolutionary Genomics Group, División de Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan de Alicante, 03550 Alicante, Spain Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Background

Stratified systems are widespread on Earth, from microbial mats to meromictic lakes and the temperate ocean. A common factor of all stratified systems is strong vertical physicochemical gradients [1, 2]. The large scale of the oceanic environment makes the upper 100 m seem relatively small. Nevertheless, this layer of the water column is one of the most biologically productive microbial habitats in the biosphere [3]. The open ocean is far from homogenous, and environmental conditions are strongly affected by the depth in the water column [4, 5]. As the depth increases, temperature declines, salinity increases, and the availability of nutrient dwindles. Among these factors, light attenuation is of paramount importance. The main divide in aquatic environments tends to be between the photic zone, where light allows for photosynthesis, and the aphotic zone, which is beyond the compensation depth and where the available light (if any) is insufficient to drive photosynthesis. The availability of light is critical for primary productivity and hence is the main limiting factor for organic matter production throughout the water column [6]. The differences in the microbiome between the photic and aphotic zones are well known using a variety of approaches [4, 7-9]. Studies at global ocean scales such as those derived from the Sorcerer II Global Ocean Sampling [10] or the more recent Tara Oceans expedition [11] have provided essential information on the composition, dynamics, and spatial distribution of surface ocean microbial communities. However, much less attention has been devoted to the differences in the vertical distribution of microbial communities. This lack of attention is particularly true of the microbial assemblages within the photic zone, where samples from a single depth are often considered representative of the complete photic water column. However, most offshore oceanic waters are permanently or seasonally stratified, sometimes as deep as hundreds of meters, which creates strong gradients of environmental parameters.

In the Mediterranean, the water column is seasonally stratified, typically from March to November. A characteristic and extensively studied phenomenon associated with this stratification is the formation of the deep chlorophyll maximum (DCM) [1], a maximum in chlorophyll concentration that is associated with an increase in bioavailable pools of nitrogen (N) and phosphorus (P) diffusing from the mixed layer below the seasonal thermocline [12]. In tropical waters, the DCM is a permanent feature, whereas in the Mediterranean and other temperate waters, the DCM is a seasonal phenomenon [13] that often appears between 45 and 70 m deep [14], depending on the degree of light penetration (dictated by the season of the year and biological productivity).

During late autumn and winter, the temperature decrease near the surface leads to vertical mixing of the water column and promotes the upwelling of nutrients (mainly dissolved organic carbon (DOC), P and N) from the mesopelagic to the euphotic zone [15]. The availability of these nutrients results in phytoplankton blooms during spring [15]. When these blooms decay, a large amount of nutrients is released, and this ecological disturbance reshapes the composition of the prokaryotic community [16–19]. The Mediterranean Sea is characterized by a relatively high temperature (>13 °C) throughout the entire water column. Although the mixing depth is variable depending on the year, it is normally located beyond 200 m [20].

Previous studies have used denaturing gradient gel electrophoresis (DGGE) [21], catalyzed reporter depositionfluorescence in situ hybridization procedure (CARD-FISH), and clone libraries [22] to demonstrate the seasonal variability of the prokaryotic community in the northwestern Mediterranean observatory located in Blanes Bay. However, most of these studies were based at one single depth (surface). Furthermore, variations within the community were predicted at the level of a class or at most at the level of genera, ignoring the fact that within the same species, different ecotypes have different niche specialization, and therefore, they are found at different depths, such as Prochlorococcus high-light and low-light ecotypes [23]. Besides, many used PCR of 16S rRNA genes introducing unknown biases, and many others relied on FISH where mismatches on the probes can underestimate the abundance of the different prokaryotic groups. Metagenome shotgun sequencing, genome reconstruction, and metagenomics recruitment can give us a glimpse of the uncultured community inhabiting in this region, and changes in their concentration among different samples can be followed at a much finer level.

Here, we have analyzed two temporal sampling efforts, one with samples collected during the stratified period every 15 m throughout the photic zone (down to 90 m) and the other with samples collected during the winter when the water column was mixed (at two depths, 20 and 80 m). To assess the variations in the community structure, we used genome-resolved metagenomics [24] to measure the recruitment of reconstructed and reference genomes at the different depths and conditions (stratified or mixed), at high similarity thresholds. This allowed the discrimination of different ecotypes within the same species. We detected marked stratification of ecotypes that reflects species adaptation to live at defined depth range. Furthermore, we detected a stable component of the photic zone microbial community, which was present regardless of the season or physicochemical parameters. Other microbes were more sensitive and appeared only in a specific season. Our results highlight the importance of collecting and comparing samples from multiple depths to understand the dynamics between mixed and stratified waters.

Results and discussion

Seawater samples from six depths in the photic zone were collected at 15 m intervals (15, 30, 45, 60, 75, and 90 m) on a single day (between 8 am and noon) during the stratification period (15 October 2015). For the comparison between the photic and aphotic regions, we also collected another sample from 1000 m (the next day starting at 8 am) from the same site. Another two samples were collected on 27 January 2015, during the winter mixing at 20 and 80 m. With the exception of the 1000 m sample, all the samples were collected using a hose directly connected to the filtration apparatus to minimize processing time and to avoid bottle effects (Additional file 1: Figure S1). The metadata and sequencing results are described in Table 1.

Variability of environmental parameters

Because of the prolonged physical isolation and accumulation of settled organic matter during the summer stratification period, bottom waters were richer in both dissolved and particulate inorganic nutrients (N and P) (Table 1 and Additional file 1: Figure S2). However, layers within the DCM were typically the richest of the photic zone, in terms of biomass accumulation. The surface water temperature for the samples taken during the stratification period was 22.9 °C, decreasing to 14.5 °C in deeper layers beyond the DCM. However, a similar temperature was found to be constant through the entire water column in winter (Table 1). Chlorophyll-a measurements indicated that the DCM occurred between depths of 40-60 m, just below the seasonal thermocline (Fig. 1a). Chlorophyll-a reached 0.8 mg m^{-3} , almost one order of magnitude above those from surface waters and 100 times those from deep (1000 m) waters. Using flow cytometry, we measured the absolute numbers of planktonic picoprokaryotes for the whole water column (Fig. 1a). In the stratified period, the maximum of *Prochlorococcus* (nearly 3.2×10^4 cells mL⁻¹) and Synechococcus $(1.35 \times 10^4 \text{ cells mL}^{-1})$ were found in the DCM peak. These values were one to two orders of magnitude higher than those in surface waters. The distribution of Prochlorococcus cells was wider than that of Synechococcus (Fig. 1a), as previously described [25]. In the winter sampling, the seawater column was mixed, and no significant differences in temperature or any other physical or chemical features were observed (Table 1). We also measured the abundances of both heterotrophic and autotrophic prokaryotes, which were highest in the shallowest sample (20 m), though the relative abundance of active heterotrophic bacterioplankton was higher in the deepest sample (80 m) (Fig. 1a).

Depth and seasonal variation of the prokaryotic community

Using the number of similar reads (> 95% identity) among metagenomes, we examined the relationship between the

nine sequenced samples (Fig. 1b). The stratified samples were clustered by depth, with three main branches corresponding to (i) upper photic (UP, 15 and 30 m), (ii) DCM (45 and 60 m), and (iii) lower photic (LP, 75 and 90 m) layers. As shown in Fig. 1b, despite the different depths at which the mixed samples (MIX) were obtained (20 and 80 m), both clustered together within the group of DCM samples. As expected, the bathypelagic 1000 m sample appeared as an outgroup compared to all the photic zone samples. Independently, the canonical correspondence analysis (CCA) of the read annotations and environmental parameters confirmed the clustering of samples according to the depth and MIX with DCM (Fig. 1c). Inorganic nutrients (such as NOx and PO43-) increased with depth, while ammonia correlated closely with chlorophyll-a, and total organic carbon (TOC) increased at the surface together with water temperature (Fig. 1c).

Measurements of Simpson's Diversity Index at both genus and species levels clearly indicated that bacterial diversity increased continuously with depth only for the stratified season (Additional file 1: Figure S3). The 15 m sample was the least diverse of those from the photic zone and was markedly predominated by Pelagibacterales. At this depth, high light intensity and nutrient depletion generate conditions that can be considered extreme and that may result in the survival of only a few microbial taxa. In deeper waters, diversity increased with depth, particularly at the species level, reaching a maximum in our deepest sample at 1000 m (Additional file 1: Figure S3). The larger diversity of microbes in bathypelagic waters might correlate with a capacity to degrade or use a larger number of different substrates [26, 27]. While diversity increased with depth during stratification, the constant value in the winter samples was similar to that of the photic region at both the genus and species levels suggesting that the disturbances in the environment produce variations in the bacterial community that diminish the diversity in favor of more adapted species.

Together with diversity, other genomic parameters also varied with depth, such as the GC content (Table 1). The GC content was lowest at 15 m (*ca.* 38.6%) and highest at 1000 m (*ca.* 45.9%) while remaining relatively constant throughout the photic zone deeper than 15 m and in the winter samples (*ca.* 41%). The lower GC content observed in the near-surface stratified waters has been suggested to be a natural adaptation to reduce N demand in these environments with a severe depletion of bioavailable N pools [28].

Metagenome-derived 16S rRNA profiles revealed broad, depth-dependent variations in taxonomic ranges during stratification (Fig. 1a). Archaea, absent in the UP region, represent nearly 16% of the population at 90 m. In the DCM and LP samples, Euryarchaeota remained

	MedWinter- JAN2015- 20m	MedWinter- JAN2015- 80m	Med- OCT2015- 15m	Med- OCT2015- 30m	Med- OCT2015- 45m	Med- OCT2015- 60m	Med- OCT2015- 75m	Med- OCT2015- 90m	Med- OCT2015- 1000m
Sampling parameters									
Sampling data	1/27/2015		10/15/2015	-)					10/16/2015
Collection depth (m)	20	80	15	30	45	60	75	90	1000
Sea bottom depth (m)	200		2647						
Size fraction (μm)	5–0.22		5–0.22						
Latitude (N)	38.06851		37.35361						
Longitude (W)	0.231994		0.286194						
Environmental parameters									
Temperature (°C)	14.50	14.40	22.90	18.40	15.80	14.50	14.00	13.80	13.10
Chlorophyll (mg/m ³)	0.46	0.21	0.10	0.24	0.78	0.36	0.26	0.08	0.01
Oxygen (mg/L)	9.59	9.42	7.09	8.77	9.00	7.66	7.16	6.88	6.10
TOC (mgC/L)	1.23	1.03	2.43	2.17	1.46	1.43	1.36	1.35	0.84
PO4 ³⁻ (µM)	0.12	0.08	0.06	0.07	0.10	0.08	0.12	0.22	0.39
Total P (μM)	0.15	0.12	0.10	0.12	0.14	0.12	0.16	0.25	0.45
NO_{x} (μ M)	3.26	3.89	0.20	0.21	0.25	0.23	5.79	6.23	8.24
NH4 ⁺ (μM)	0.13	0.11	0.13	0.12	0.14	0.15	0.09	0.08	0.03
Total N (µM)	3.68	4.36	0.40	0.41	0.48	0.46	6.33	6.90	8.89
N:P ratio	23.81	35.74	4.00	3.42	3.43	3.83	39.56	27.60	19.76
Sequencing statistics									
Total bp (Gb)	15.9	15.3	19.9	15.1	3.1	15.3	16.9	15.0	14.8
Mean read length (bp)	92	91.9	121.4	117.2	112.3	119.8	121.4	120.2	117.0
Mean GC (%)	40.7	41.1	38.6	41.4	40.4	40.6	41.3	41.1	45.9
Assembly statistics									
Total bp (Mb)	203	201	738.5	500.3	79.5	577.3	701.3	613.5	490.0
Mean GC (%)	38	39	38.5	35.7	35.5	38.6	38.9	40	42.2
Maximum contig length (Kb)	137	109	251	165	118	235	186	140	450
Contigs > 1 Kb	39,794	38,491	172,484	125,642	17,760	145,031	173,153	153,245	115,711
Contigs > 10 Kb	869	885	4648	2198	202	2360	2789	2556	1807

 Table 1 Summary statistics of the sampling, sequencing, and assembly parameters

constant (*ca.* 5%), while Thaumarchaeota increased from 1% in the 45 m sample to 10% of all the rRNA reads in the 90 m sample. The abundance of Thaumarchaeota correlated with a sharp decrease in ammonia concentration, although the main increase in Thaumarchaeota occurred at 60 m, while ammonia concentrations showed the lowest values at 75 m (Fig. 1a and Table 1).

Whereas Actinobacteria, Bacteroidetes, Cyanobacteria, and Marinimicrobia were present in the whole water column, Deltaproteobacteria, Planctomycetes, Chloroflexi, and Acidobacteria had a much more restricted range, appearing only in deeper layers of the photic zone (Fig. 1a). Interestingly, Verrucomicrobia were present at all depths except in the 45 m sample. Using finer-scale taxonomic classification of the 16S rRNA sequences, we found that UP (15 and 30 m) Verrucomicrobia belonged to *Puniceicoccaceae*, whereas the members of *Verrucomicrobiaceae* were predominantly found below the DCM (Additional file 2: Table S1). Although the results clearly reveal ecologically distinct lineages that occupy different niches, we still know very little about the ecophysiology of these Verrucomicrobia lineages in seawater. The proportion of 16S rRNA gene reads assigned to unclassified bacteria also increased with depth, from 3% at 15 m to more than 10% at 90 m, indicating that a significant fraction of the microbes at the subsurface is still uncharacterized.

Furthermore, our results indicate no significant changes in prokaryotic diversity during seasonal fluctuations as long as the entire water column is taken into account rather than only a single depth. A homogeneous community distribution similar to the DCM and LP samples was observed in the MIX samples. For example, it has been



suggested that through using pyrosequencing 16S rRNA gene PCR amplicons in a surface sample (3 m depth) in the northwestern Mediterranean Sea, Thaumarchaeota MGI and Euryarchaeota MGII-B populations were more abundant during winter [29]. However, our results show that archaea were always present and abundant throughout the water column during the winter but were almost absent in the UP region during the stratification. Similar observation was made using metatranscriptomes from the stratified water column in the Gulf of Aqaba/Eilat [30]. In the same way as the Planctomycetes or Chloroflexi that only appeared below the DCM (Fig. 1a). Even at lower taxonomic ranks, the distribution was similar, except for some specific families such as *Sphingomonadaceae*, *Alteromonadaceae*, and *Pseudomonadaceae*, that predominantly increased in the deeper layers during the winter (Additional file 2: Table S1). This finding highlights the importance of collecting samples at different depths in the water column when comparing seasonal variations and has important ramifications for global marine studies that most often take samples only from the surface or, at most, from one single subsurface photic zone depth.

Metagenome-assembled genomes

The broad organismal distributions detected by 16S rRNA genes or raw sequence annotation methods described

above, however, do not shed light on the more subtle but ecologically significant variations in community structure or metabolic function that likely occur at the finer levels of diversity, such as ecotypes, or even clonal lineages, within species [31–34]. To investigate the distribution of the major ecotypes present in the water column, we used stringent recruitment of metagenomic reads for assemblies of locally predominant metagenome-assembled genomes (MAGs). We have used the same approach with several metagenomes obtained closer to the sampling site [35–37].

Overall, using a combination of different parameters, such as GC content, metagenomics read coverage, and tetranucleotide frequencies, we have retrieved new MAGs belonging to phyla for which we obtained more than 5 Mb of assembled contigs (Additional file 3: Table S2 and Additional file 1: Figure S4). These genomes were classified phylogenomically using concatenated sequences of conserved proteins (Additional file 1: Figures S5–S12). In the end, we were able to obtain 94 novel MAGs.

In general, genome assembly improved proportionally with the abundance of the phylum. However, we found that genomes of representatives from Bacteroidetes, Actinobacteria, and Acidobacteria assembled better than expected based on their abundance by 16S rRNA gene fragments recovered (Additional file 1: Figure S4 and Additional file 2: Table S1). On the other hand, Cyanobacteria, Thaumarchaeota, and Pelagibacterales assembled much more poorly. Both picocyanobacteria and Pelagibacterales are known to possess enormous intra-species diversity [38], which might be the reason why the assembly for these two major components of the bacterioplankton was very poor.

Relative abundance of the prokaryotic community

To examine the patterns of relative abundance and diversity of the microbial communities among the metagenomes, we performed metagenomic recruitments of the reads over the MAGs as well as several reference genomes from public databases, taking into account only the reads that match the genomes with a similarity \geq 99% in our metagenomic samples, thus representing finer levels of diversity. To simplify, we set a threshold of three reads per kilobase of genome and gigabase of metagenome (RPKG) in at least one sample to establish the presence of these genomes.

Fine taxonomic profile (eurybathic and stenobathic)

Additional file 1: Figure S13 shows the recruitment of MAGs obtained from the stratified metagenomic samples of this study (from MED-G01 to MED-G44), MAGs from other metagenomic samples previously described from the same site (from MED-G45 to MED-G94), and several selected genomes of marine isolates sourced from public databases (21 recruited more than 3 RPKG

and are shown in the figure). It is remarkable how uneven the recruitment depth profile was for the vast majority of the microbial genomes, particularly considering that all the samples were collected on the same day (except the 1000 m). All the genomes recruited much more at one single specific depth and most (ca. 70%) recruited only from metagenomes sampled at either one or two consecutive depths (stenobathic). This result indicates that the distribution of most of these microbes only extends over a 30-m-thick layer within the ca. 100 m deep photic zone. Only one of the photic zone genomes, Sphingomonadaceae MED-G03, recruited in the 1000 m sample. This genome actually recruited more at this depth and it may be the only truly eurybathic microbe among the ones assembled here. The actinobacterial genomes seemed to be the next most eurybathic, and although they always appeared to be more prevalent at a single depth, they were detectable at four depths, with the lone exception of the single cell genome SCGC-AAA015 -M09 [39] (only found at 15 and 30 m). Alphaproteobacteria (with the exception of Pelagibacterales), such as most Bacteroidetes and Gammaproteobacteria, were only detected at one or two depths. Most microbes were preferentially found at the UP or DCM depths except for some archaea. For example, members of the MGI Thaumarchaeota and some groups of Euryarchaea appear to prefer the LP (Additional file 1: Figure S13). Ca. Nitrosopelagicus brevis [40] and Nitrosopumilus MED-G94 possess the complete cluster for ammonia oxidation and are expected to increase with depth due to the much higher availability of its major substrate (ammonia). Moreover, their abundance in this region is also correlated with the light intensity attenuation in deeper waters due to the ammonia oxidation photoinhibition [41]. We utilized the relatively large collections of available pure culture genomes of picocyanobacteria and used the ones that had contigs with high similarity (close to 100%) as proxies of local genomes. Synechococcus MAGs were practically identical (> 99.2% average nucleotide identity [ANI]) to the isolated genomes, whereas Prochlorococcus MAGs where closely related but not identical (97-98% ANI) (Additional file 1: Figure S8). Recruitments of cultured picocyanobacteria occurred over a range similar to the locally assembled genomes, and again, the clear depth preferences were apparent. In Cyanobacteria, there are low/high light-adapted ecotypes, as has been repeatedly described in several oceanic regions [23, 42]. The first 45 m were dominated by the HLI clade (the pure culture Prochlorococcus MED4 and the MAG Prochlorococcus MED-G72) with a peak in abundance at approximately 30 m, which then decreased below this depth when clade LLI (Prochlorococcus NATL1A and Prochlorococcus MED-G73) appeared. On the other hand, Synechococcus genomes were not detected deeper than 30 m (Additional file 1: Figure S13).

Seasonal dynamics of the community structure

To analyze the impact of the strong winter convection on the community, we included two more samples obtained in winter (January 2015) at 20 and 80 m depth, during the period when the water column is fully mixed. Using the previous criteria, we found that despite the strong variability in the physicochemical parameters (light, temperature, and nutrients), 47% of the genomes were found in both the mixed and the stratified periods. In fact, based on relative abundances, microbes that were found only in small ranges of 15 to 30 m deep during the stratified period were present at similar values at both depths (20 and 80 m) when the water column was mixed (Fig. 2a). Among the groups that were always present, 21 out of 49 (43%) were Alphaproteobacteria and Cyanobacteria (mainly the SAR11 clade and Synechococcus, respectively). Some less abundant, but nevertheless resistant, taxa (always present) included members of the Actinobacteria families Acidimicrobiaceae (MedAcidi-G1, G2A, G2B, G3) and Ca. Actinomarinaceae (Ca. Actinomarina minuta), three SAR86 clade genomes within Gammaproteobacteria and the Bacteroidetes family *Flavobacteriaceae* (Fig. 2a, b). Most of these groups represent ubiquitous and abundant heterotrophic microbes characterized by a small genome size and low GC content (and likely a small cell size and more efficient absorption of nutrients) (Fig. 3), highly adapted to oligotrophic environments by metabolic streamlining (K-strategists) [43–47]. This persistence could be due also to their ability to use organic matter along with light energy through light-dependent proton pumps (rhodopsins). It seems likely that these capabilities allow for better adaptation to overcome the environmental disturbances produced during winter mixing and subsequent phytoplankton blooms. Unlike what the literature has described so far, we measured a temporal persistence of some taxa previously considered sporadic or rare, such as archaea [22, 29, 48, 49]. Although these taxa only recruit below the photic zone during the stratified period, members of the ammonia-oxidizing group I Thaumarchaeota and groups II and III Euryarchaeota were always found to be present in the water column.

Conversely, most of the microbes that we found only during winter could be considered opportunistic (*r*-strategists or bloomers) and are microbes that grow rapidly, taking advantage of the sporadic inputs of organic matter that appear in the environment. However, although they can be easily retrieved in pure culture, they are usually rare in seawater. These microbes could be associated with the decay of the phytoplankton blooms and higher nutrient levels [50, 51]. We were able to assemble seven genomes only found in winter that were classified within the Actinobacteria, Gammaproteobacteria, Verrucomicrobia, Bacteroidetes, and Euryarchaeota phyla (Fig. 2). As was previously shown in Fig. 3, these genomes were characterized by having a large estimated genome size (> 3.0 Mb) and a high GC content (> 50%). Additionally, these genomes also possess multiple clusters for degrading a wide range of substrates as well as genes responsible for flagellum biogenesis and motility, which are typical metabolic properties of heterotrophic bacterial communities associated with these phytoplankton blooms [50].

Remarkably, 46 MAGs were only present during stratification, being totally absent in winter (Fig. 2). Many of these MAGs were members of the phyla Bacteroidetes (12 genomes), Verrucomicrobia (4 genomes), members of the SAR116 clade of the Alphaproteobacteria (Additional file 1: Figure S6) and the OM60/NOR5 clade within the class Gammaproteobacteria (Additional file 1: Figure S10). The vast majority of these genomes were found to be restricted to the UP layer. However, members of MGII archaea, OM182, and SUP05 clades of Gammaproteobacteria, that also disappear in winter, came from deeper layers (DCM and LP). A seasonal analysis carried out in surface waters of Blanes [22], Bermuda [18], and the North Sea [52] showed variations in the concentration of members of these clades throughout the year, with a maximum in mid-summer and a near absence in winter when the water column was mixed and which were mostly limited to surface waters [22, 52, 53] in agreement with our data.

Depth stratification of rhodopsins

Rhodopsins have been shown to be among the most widespread genes in the photic zone worldwide [54–56]. They are very diverse and are distributed throughout most taxa. We found 28 rhodopsin genes in both winter samples, but just one gene recruited only in winter and not during stratification. This rhodopsin (within the MAG Verrucomicrobia MED-G86) was analyzed in detail (see below) and belonged to the Planctomycetes-Verrucomicrobia-Chlamydiae (PVC) superphylum. In the end, a total of 105 out of 196 rhodopsin genes (53%) recruited only during stratification, 46% in both, and just 1 rhodopsin gene only in winter.

We evaluated the numbers of rhodopsins among the individual reads and calculated their frequency per genome, normalizing them by the number of single copy housekeeping genes (*recA* and *radA*) and by their gene length (Fig. 4b). The total numbers of rhodopsin-assigned reads were clearly correlated to light intensity, with a maximum at 15 m, where ca. 65% of the genomes contain a rhodopsin, which then decreased with depth. This result is different from the situation in the permanently stratified central North Pacific, where the maximum was found at the DCM [56]. Conversely, for winter samples, the number of rhodopsins was similar regardless of depth



throughout the water column, as was expected due to the mixing event.

We assembled 168 rhodopsin genes throughout the stratified water column. All of the genes were classified at

least at the phylum level based on the flanking genes (Fig. 4a). The phylogenetic analysis revealed a large diversity of this gene family, and at least 11 major evolutionary lineages were detected. All the assembled rhodopsin genes



clustered with previously described groups, indicating that surveys may have achieved saturation with the extant diversity of rhodopsins, at least in the oligotrophic ocean photic zone. Rhodopsin sequences clustered primarily by phylum, with the exception of euryarchaeal rhodopsins as previously reported [36, 57]. Within the proteorhodopsin cluster, we clearly differentiated a separate cluster including only Bacteroidetes sequences (Fig. 4a). Within the clusters, rhodopsin sequences were also grouped by depth, with many branches containing only upper or lower photic zone varieties. This result confirms the stenobathic character of most groups at the finer level of diversity resolution.

Rhodopsin genes from our metagenomic assemblies and from the MicRhoDE database [58] were used to recruit reads from the different depths (Fig. 4c). We observed no correlation between the predicted absorption spectrum (blue versus green light) of the rhodopsins and of the depth from which they recruited the most reads. In contrast, we did see a consistent pattern of correlation between the absorption spectrum and the phylogenetic affiliation of the host genome; Bacteroidetes and Actinobacteria all carry green rhodopsins, while Proteobacteria largely have the blue variety. The findings suggest, as previously reported [55, 59], that the spectral tuning of rhodopsins may not be related to depth adaptation but tend to be associated with the classification of the microbe instead.

Interestingly, within the MAG Verrucomicrobia MED-G86 (3.19 Mb and 55% GC content), we found the unique rhodopsin that recruited only in the MIX samples but not in the stratified. This is the first marine rhodopsin that clustered together with a novel clade of freshwater rhodopsins [60, 61] affiliated closely with the Exiguobacterium

rhodopsins [62], confirming that this group is a characteristic of the Planctomycetes-Verrucomicrobia PVC superphylum (Additional file 1: Figure S14). Since this is the first marine representative, we searched in the Tara Oceans assembled contigs > 5 Kb for similar members in this group. Eight genomic fragments containing rhodopsin that clustered with this novel branch were retrieved (Additional file 1: Figure S14). It is remarkable that although two sequences came from the Mediterranean Sea (stations 009 and 030), the remaining six came from the North and South Pacific Oceans (stations 093, 094, 102, 109, 128, and 136). Furthermore, within the novel clade, we found another rhodopsin subcluster formed only with Tara sequences. However, the contigs that contained these sequences differed from the others in GC content, with low values between 35 and 40% instead of the high GC values found in Verrucomicrobia MED-G86 and the freshwater MAGs (Additional file 1: Figure S14). Unfortunately, we failed to classify taxonomically these contigs due to the ambiguous annotation of their proteins (proteins were annotated either as Verrucomicrobia or Planctomycetes).

Functional analysis of the stratified and mixed water column To make a functional characterization of the microbial community associated with the metagenomes, we used the assembled coding sequences collected from contigs > 1 Kb against the SEED Subsystems database [63]. Clustering of level 1 subsystems (Fig. 5a) revealed a marked discontinuity between UP samples and the other samples, indicating unique characteristics of surface waters, while once again, the DCM and MIX samples clustered together, demonstrating that they are similar on a functional level, in concordance with the taxonomic clustering





rhodopsin normalized by the number and length of recA+radA genes (estimated rhodopsin genes per genome) are indicated by red squares. **c** Recruitment of all rhodopsins including those obtained in this work together with the publicly available in the MicRhoDE database. Left panel, rhodopsins classified according to their taxa of origin. Right panel, average number of blue/green rhodopsins for each depth. Numbers within the brackets indicate the number of rhodopsins based on PCA (Fig. 1a). After comparing the number of proteins assigned to each subsystem, we found significant differences (based on the standard deviation among samples) mainly involving carbohydrates, membrane transport, and motility and chemotaxis.

Carbohydrates

To study the taxonomical and in-depth distribution of the genes encoding the glycoside hydrolases (GH) family of enzymes, which are involved in the breakdown of complex sugars, we compared all the proteins extracted from contigs larger than 5 Kb assigned to the phyla Actinobacteria, Bacteroidetes, Euryarchaeota, Thaumarchaeota, and Verrucomicrobia, as well as the classes Alphaproteobacteria and Gammaproteobacteria (all of which comprised more than 85% of the metagenomic 16S rRNA gene reads for all the samples), against the CAZy database [64].

The phylogenetic distribution of the CAZy genes was analyzed, considering the number of GH per 1000 genes (EQ) and the abundance normalized by the percentage of 16S rRNA gene reads of each group (NORM). Figure 5b shows that the abundance varied across bacterial phyla, and most of the genes were mainly derived from Verrucomicrobia, Bacteroidetes, and Cyanobacteria. Thaumarchaeota showed no GHs within the contigs, demonstrating an inability to degrade complex polysaccharides, as was expected from chemolithoautotrophs [40, 65]. Notably, within each group, the number of GH genes was similar at the different layers of the water column, although the types of GHs were different, suggesting specialization in the degradation of different polysaccharides that is likely connected with specific groups of algae or particles.

As expected [16, 66], Bacteroidetes was the group with more enzymes (74.3 GHs/1000 genes) (Fig. 5b). Clustering based on the abundance of the samples showed that Bacteroidetes from DCM and LP samples grouped together and separated from UP, which in turn was close to the MIX samples. We found some predominant GH families in winter, the two most abundant were endo- β -1,3-glucanases of the families GH5 and GH17, and a GH30 exo- β -1,6-glucanase (Fig. 5c). These enzymes are involved in the cleavage of the main storage polysaccharide (β -glucan) present in brown algae (laminarin) and in diatoms (chrysolaminarin) [67].

Verrucomicrobia represented the second group that included the largest number of GH genes, with 54.3 GHs/ 1000 genes analyzed. The results showed that the majority of the GHs present in Verrucomicrobia were different from Bacteroidetes, indicating that members of these phyla may be utilizing different carbohydrate substrates (Fig. 5c). As with Bacteroidetes, we found that the number of GH families was higher in Verrucomicrobia from UP than in DCM and LP. This result suggests that deeper Verrucomicrobia shows less variability in degrading potential substrates. Specifically, we found an overrepresentation of alpha- and beta-galactosidases, xylanases, fucosidases, agarases, and endoglucanases in UP Verrucomicrobia. The most abundant family at all depths was GH109, with the only known activity being that of an α -*N*-acetylgalactosaminidase that might degrade the peptidoglycan of the cell walls [68].

Remarkably, Cyanobacteria was the third group with a higher number of GHs (Fig. 5b). Unlike the previous cases, the type of GH family was similar in all the samples and was associated with amylose degradation (GH13 and GH57—α-amylase; GH77—amylomaltase). These three GH families were also found in Actinobacteria (UP and MIX) and in Euryarchaeota (LP), which shared the same metabolic potential. Additionally, clustering showed that Cyanobacteria from DCM and MIX shared similar values for the families GH19 and GH24, both with chitinase/lysozyme activities. Thus, the degradation of complex sugars (i.e., amylose or chitin) increases their capability to obtain organic carbon. It has been described that both Cyanobacteria (Prochlorococcus and Synechococcus) also harbor genes that encode a wide number of amino acid, peptide, and sugar transporters [69–71], which allow them to uptake organic compounds, that together with the ability to obtain energy using the sunlight (mixotrophy) seems to be present in all the marine Synechococcus and Prochlorococcus, and globally distributed in the photic zone of the oceans [71]. Recently, it has been shown that mixotrophy can increase the viability of Prochlorococcus marinus during extended periods of darkness, due to the coculture with a marine copiotroph, Alteromonas macleodii, which may be supplying organic compounds to Prochlorococcus. Our results, together with previous studies, highlight the mixotrophic nature of marine picocyanobacteria, as several glycoside hydrolases are encoded in their genomes.

Although Alpha- and Gammaproteobacteria comprised > 50% of the prokaryotic community (based on the metagenomic 16S rRNA gene reads), they possessed very low numbers of GHs (14.1 and 16.6 GHs/1000 genes, respectively), indicating a different functional role in the marine ecosystem.

Membrane transport

We analyzed the abundance of genes affiliated with membrane transport using KEGG modules. PCA analysis was performed to determine the clustering of the samples (Additional file 1: Figure S15). The results showed that the mixed samples clustered together and separated from the stratified samples, which, in turn, were also clustered by depth for UP and LP samples, while the DCM samples showed a more dispersed distribution. In terms of nutrient acquisition, we found transport systems (ATP-binding cassettes and phosphotransferases) related to iron, phosphonate, polyamines (putrescine/spermidine),



oligopeptides, and sugars, and several heavy metal resistances such as the cobalt-zinc-cadmium (CzcA) efflux system, which are typically components of the flexible genome of some bloomers [72], were enriched in the winter-mixed samples. This wide variety of transporters might allow for uptake and use of a large quantity of phytoplankton-derived compounds. During the stratification, in the lower layers of the water column beyond the DCM, in addition to putative specific transporters for Archaea (A2 holin family), we found a higher proportion of ABC di/oligopeptide transporters. TonB-dependent transporter proteins are relatively abundant particularly in UP. These transporters allow the uptake of scarce resources (i.e., iron complexes and other nutrients [73]) from nutrient-limiting environments such as surface layers due to their high affinity. Choline and betaine uptake proteins that play an important role in bacterial osmoregulation and stress tolerance were also abundant in the UP [74]. For instance, the SAR11 clade, which based on 16S rRNA data is the most abundant here, was enriched in these transporters that are highly active based on transcriptome data [75].

Motility and chemotaxis

Motility is another adaptation that differentiates copiotrophs from oligotrophs [45]. Despite that UP presents the highest value in abundance of genes related to the SEED category "motility and chemotaxis," this region is dominated by members of the SAR11 clade, which have no genes encoding for flagellar synthesis or chemotaxis proteins. Manual inspection of the contigs including these proteins revealed an enrichment in high GC-content microbes mainly from Alpha- (Sphingomonadadales and SAR116) and Gammaproteobacteria (Oceanospirillales) classes. These genomes probably assembled better due to the lower intra-species diversity. Remarkably, within the group of MIX samples, bacteria from MedWinter-JAN2015-80m exhibited a significantly large number of genes involved in chemotaxis but not for biosynthesis of the flagella in comparison with all the other samples (Additional file 4: Table S3). These results suggest that a mechanism to sense and respond to the chemicals likely released by phytoplankton is an important competitive advantage for opportunistic bacteria during winter when the access to nutrients increases. Other functions reflected the interaction with phytoplankton blooms, for example, the inclusion of modules involved in the detoxification of reactive oxygen species (ROS) since phytoplankton are the most important source of ROS in the water column [76] or peptidases to process phytoplankton-derived organic matter [77]. Many studies have demonstrated that there is a mutualistic or parasitic interaction between bacteria and phytoplankton [50].

Conclusions

The photic zone of aquatic habitats is subjected to several strong gradients of environmental parameters. In the Mediterranean, similar to in most temperate seas, thermal stratification appears only during warmer periods, typically from May to November, while in winter, the water column is mixed and the gradient of nutrients disappears. Although there is abundant information about the prokaryotic community composition during the stratified and mixed periods, most previous works derive from 16S rRNA-related techniques (such as FISH or barcoding approaches). These approaches do not have enough resolution at the species or ecotype levels. Here, we have used metagenome recruitment as an alternative to detect specific MAGs and some previously described genomes to assess the distribution of specific microbial genomes in a fine depth profile (every 15 m) from an stratified and also in a mixed water column during winter. We found major depth-associated shifts in the community structure during the stratified period and that, particularly at the level of fine variation, most microbes had a distribution covering only a ca. 30-m-thick layer of seawater and were stenobathic. During the stratified period, it is necessary to consider the vertical distribution as the major element instead of comparing single depth samples. Thus, we found that some microbes previously considered rare or seasonal (such as archaea or Actinobacteria) are actually resistant to seasonal variations. These microbes generally live in deeper layers within the photic zone during the stratification period (Fig. 6). Our results also indicate a strong specialization not only at the taxonomic level but also at the functional level, even within the different clades, for the manipulation and uptake of specific polysaccharides and likely for the succession of different bloom events. This finding has important ramifications for global marine studies that most often take samples only at the surface or, at most, from one single subsurface photic zone. Moreover, the description of seasonal dynamics within the water column has important implications in the analysis and response to future alterations in the water conditions due to climate change. Mainly, an increase of the seawater temperature will produce a change of the physical mixing dynamics, where the upwelling of nutrients from the deeper layers to the surface will be prevented, reshaping the microbial community structure. As a result, this will have direct consequences on microbial metabolism, which will modify the marine global biogeochemical cycles (mostly carbon and nitrogen) [78].

Methods

Sampling, sequencing, assembly, and annotation

Six samples from different depths were taken for metagenomic analyses on 15 October 2015 at a single site from the western Mediterranean (37.35361° N, 0.286194°



W), at approximately 60 nautical miles off the coast of Alicante, Spain, from the research vessel "García del Cid." These seawater samples (200 L each) were collected from the uppermost 100 m at 15 m intervals using a hose attached to a CTD (Seabird) connected to a water pump, to directly transfer seawater from the selected depth to the filtration system, and thus minimize sample storage time and potential bottle effects (Additional file 1: Figure S1). Each sample was filtered in less than 30 min. Another sample from a depth of 1000 m was taken the next day (16 October) in two casts (100 L each) using the CTD rosette. Two more samples were collected on 27 January 2015, at 20 and 80 m depth, at 20 nautical miles off the coast of Alicante (38.068° N, 0.232° W).

All seawater samples were sequentially filtered on board through 20, 5, and 0.22 μ m pore size polycarbonate filters (Millipore). All filters were immediately frozen on dry ice and stored at – 80 °C until processing. DNA extraction was performed from the 0.22 μ m filter as previously described [79]. Metagenomes were sequenced using Illumina Hiseq-4000 (150 bp, paired-end read) (Macrogen, Republic of Korea). Individual metagenomes were assembled using IDBA-UD [80]. The resulting genes on the assembled contigs were predicted using Prodigal [81]. tRNA and

rRNA genes were predicted using tRNAscan-SE [82], ssu-align [83], and meta-RNA [84]. Predicted protein sequences were compared against NCBI NR databases using USEARCH6 [85] and against COG [86] and TIG-FRAM [87] using HMMscan [88] for taxonomic and functional annotation. GC content and richness in each sample were calculated using the gecee program from the EMBOSS package [89] and MEGAN6 Community Edition [90], respectively.

Vertical profiles and chemical features

Vertical profiles of several physical, chemical, and biological variables were determined in situ using a Seabird SBE 19 multiprobe profiler coupled to several fluorometric probes. Variables measured were temperature (SBE), dissolved oxygen (SBE43), pH (SBE27), chlorophyll-a concentration (WETStar), phycoerythrin (Seapoint) and phycocyanin (Turner) fluorescence, turbidity (Seapoint), and chromophoric dissolved organic matter (cDOM) concentration (Wetlabs). Other chemical variables, inorganic soluble forms of nitrogen (NOx and ammonium), and phosphorus (soluble reactive phosphorus), as well as total nitrogen (TN) and total phosphorus (TP), were performed following standard methods for water analyses [91]. Total organic

carbon (TOC) was determined on a Shimadzu TOC-VCSN Analyser. Quantitative determination of chlorophyll-a concentrations was determined by HPLC after extraction in acetone following [92].

Microbial counts

The abundance of heterotrophic and autotrophic picoplankton (Synechococcus and Prochlorococcus) were determined using a Coulter Cytomics FC500 flow cytometer (Brea, California, USA) equipped with two different lasers, an argon laser (488 nm excitation) and a red-emitting diode (635 nm excitation), and five detectors for fluorescent emission (FL1-FL5). Quantitative counts of heterotrophic bacterioplankton and its relative DNA content (HDNA versus LDNA cells, as a relative measure of activity) [93] were performed after cell DNA staining with Sybr Green I (Sigma-Aldrich, Missouri, USA) following [94]. Using the green fluorescence of Sybr Green I, the argon laser allowed detecting the cells with the FL1 detector (525 nm). The abundance of autotrophic picoplankton was determined by combining the argon laser and the red diode with the red fluorescence of chlorophyll-a and phycobiliproteins, using the FL4 detector for the identification of the populations of Synechococcus and Prochlorococcus. Their cells were differentiated by both their fluorometric signature and size features. Cytometric parameter settings were FSC (550), SSC (390), FL1 (600), FL2 (670), FL3 (670), FL4 (620), and FL5 (700). Analyses were run for 160 s at the highest possible single flow rate (128 μ L min⁻¹). Abundance of each population was calculated according to the formula: $N = (n \times 1000)/(q \times t)$, where q is the flow rate (microliter per minute), t is the duration (minutes) of the acquisition, *n* is the number of events counted by the flow cytometer, and N is the number of cells per milliliter. Data were collected using the Beckman Coulter software for acquisition "CXP Version 2.2 Acquisition," and the analysis of the data was performed using the Beckman Coulter software for analysis "CXP Version 2.2 Analysis."

Phylogenetic classification

A non-redundant version of the RDP database [95] was prepared by clustering all available 16S/18S rRNA gene reads (*ca.* 2.3 million) into approximately 800,000 clusters at 90% identity level using UCLUST [85]. This database was used to identify candidate 16S/18S rRNA gene sequences in the raw metagenomes (subsets of 10 million reads). Using USEARCH [85], sequences that matched this database (*E* value < 10^{-5}) were considered potential 16S rRNA gene fragments. These candidates were then aligned to archaeal, bacterial, and eukaryal 16S/18S rRNA HMM models [96] using ssu-align to identify true sequences [83]. Final 16S/18S rRNA sequences were compared to the entire RDP database and

classified into a high-level taxon if the sequence identity was $\ge 80\%$ and the alignment length ≥ 90 bp. Sequences failing these thresholds were discarded.

Binning and genome reconstruction

Assembled contigs longer than 10 Kb were assigned a high-level taxon classification if > 50% of the genes shared the same taxonomy. The rest of the contigs were grouped together as unclassified. To bin the contigs into MAGs, their taxonomic affiliation (including unclassified group) was used together with the principal component analysis of tetranucleotide frequencies, GC content, and coverage values within the metagenomes collected in this work, together with those described in [36, 37]. Tetranucleotide frequencies were computed using wordfreq program in the EMBOSS package [89]. The principal component analysis was performed using the FactoMineR package [97] in R. Completeness of the MAGs was estimated by comparison against two different universal gene sets, one with 35 genes [98] and another with 111 genes [99], and with CheckM, which also provides the degree of contamination [100]. In order to improve the completeness and remove the redundancy, a second assembly step was performed combining the genomic fragments with the short paired-end Illumina reads of the metagenomes from which they were assembled. For each genome, we used the BWA aligner [101] with default parameters to retrieve the short paired reads that mapped onto the contigs. These reads were then pooled and assembled together with the contigs using SPAdes [102].

Metagenomic read recruitments

The genomes of known marine microbes together with the genomes reconstructed in this study were used to recruit reads from our metagenomic datasets using BLASTN [103], using a cutoff of 99% nucleotide identity over a minimum alignment length of 50 nucleotides. Genomes that recruited less than three reads per kilobase of genome per gigabase of metagenome (RPKG) were discarded.

Phylogenomic trees of the reconstructed genomes

Phylogenomic analysis was used to classify and identify the closest relatives for all the reconstructed genomes. Using HMMscan, we aligned the sequences against the COG database. Shared proteins were concatenated and aligned using Kalign [104]. A maximum-likelihood tree was then constructed using MEGA 7.0 [105] with the following parameters: Jones-Taylor-Thornton model, gamma distribution with five discrete categories, and 100 bootstraps. Positions with less than 80% site coverage were eliminated.

Metagenomic cross-comparisons

Two different approaches were used to compare similarities between metagenomic samples. First, a reciprocal global alignment of the short Illumina reads (in subsets of 2 million reads \geq 50 bp) at \geq 95% identity was performed using USEARCH6. The results of the comparison were then clustered with the hclust package in R using a euclidean distance matrix. In a second approach, subsets of 20 million reads \geq 50 bp (where applicable) were taxonomically classified against the NR database using DIAMOND [106] with a minimum of 50% identity and 50% alignment. The resulting alignment was later analyzed with MEGAN6 Community Edition, and a canonical correspondence analysis (CCA) was inferred with the cluster analysis option and a Bray-Curtis ecological distance matrix.

Rhodopsins

One hundred sixty-eight rhodopsin sequences were extracted from all the metagenomes from assembled contigs longer than 5 Kb. These sequences were pooled with 100 more rhodopsins of fungal, archaeal, viral, and bacterial origin obtained from databases. Sequences were aligned with MUSCLE [107] and a maximum-likelihood tree was constructed with MEGA 7.0 (Jones-Taylor-Thornton model, gamma distribution with five discrete categories, and 100 bootstraps, positions with less than 80% site coverage were eliminated). Blue versus green light absorption was determined as described previously [108]. To compare the abundance of microbial rhodopsins with depth, we initially created a database containing our metagenomic rhodopsin sequences and approximately 7,900 rhodopsin genes obtained from the MicRhoDE database (http://micrhode.sb-roscoff.fr). Metagenomic reads (in subsets of 20 million sequences) were recruited to these rhodopsin sequences using BLASTN (\geq 50 bp alignment, \geq 99% identity). Rhodopsin sequences that recruited ≥ 1 RPKG were kept for further analyses. In parallel, metagenomic reads were compared to the NR database using DIAMOND (blastx option, top hit, \geq 50% identity, \geq 50% alignment length, *E* value < 10⁻⁵). The abundance of rhodopsin genes in each metagenome was estimated from the number of reads matching rhodopsin sequences in NR, normalized by the number of reads matching recA/radA sequences and by their respective gene length. Reads matching viral or eukaryotic proteins were not taken into account.

Analysis of glycoside hydrolases

Predicted protein sequences of contigs longer than 5 Kb previously taxonomically classified were compared against the Carbohydrate-Active enZYmes (CAZy) database [64]. Using dbCAN [109], sequences that matched as glycoside hydrolases (GH) with an *E* value $< 1e^{-8}$ were kept for further analyses.

Functional classification of the assembled proteins

All the proteins encoded within the assembled contigs > 1 Kb were selected, and their putative functionality was inferred against the SEED subsystems [63] and KEGG [110] databases for each metagenome analyzed. Proteins were compared to the SEED database using DIAMOND (blastp option, top hit, \geq 50% identity, \geq 50% alignment length, *E* value < 10⁻⁵). GhostKOALA [111] was used to classify the sequences against the KEGG database.

Additional files

Additional file 1: Figure S1. Method used for sampling. Water was pumped through a hose directly on to the filters instead of using the Niskin bottles rosette. Figure S2. Bar plot showing the concentration of inorganic nutrients in both stratified (blue) and mixed (red) samples. Figure S3. Simpson Diversity Index versus depth. Figure S4. Assembled contigs. A) Size of individual contigs to the left and total assembled size to the right for each phylum. Proteobacteria was divided into its class-level taxonomy. The number of contigs longer than 10 Kb that were taxonomically classified is indicated within brackets. B) Individual contribution of each metagenome to the total assembled size. Figure S5. Phylogenetic analysis of Actinobacteria metagenome-assembled genomes (MAGs). A maximum likelihood genome tree was constructed with 100 bootstraps using 31 conserved proteins among the 20 genomes compared. Black circles represent bootstrap values. Between brackets: ANI, average nucleotide identity; COV, percentage of genome sequence shared. In red, those MAGs retrieved in this work. Figure S6. Phylogenetic analysis of Alphaproteobacteria MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 46 conserved proteins among the 40 genomes compared. Black circles represent bootstrap values. Figure S7. Phylogenetic analysis of Bacteroidetes MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 21 conserved proteins among the 29 genomes compared. Black circles represent bootstrap values. Figure S8. Phylogenetic analysis of Cyanobacteria MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 286 conserved proteins among the 45 genomes compared. Black circles represent bootstrap values. Figure S9. Phylogenetic analysis of Euryarchaeota MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 26 conserved proteins among the 20 genomes compared. Black circles represent bootstrap values. Figure S10. Phylogenetic analysis of Gammaproteobacteria MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 32 conserved proteins among the 44 genomes compared. Black circles represent bootstrap values. Figure S11. Phylogenetic analysis of Verrucomicrobia MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 79 conserved proteins among the 25 genomes compared. Black circles represent bootstrap values. Figure S12. Phylogenetic analysis of Thaumarchaeota MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 129 conserved proteins among the 17 genomes compared. Black circles represent bootstrap values. Figure S13. Relative abundance of the reconstructed and reference genomes measured by recruitment (RPKG, reads per kilobase of genome and gigabase of metagenome) from the different depths of the stratified metagenomes. To show the relationships among genomes, a maximum likelihood genome tree was constructed using all the conserved proteins (number in colored square). Each MAG (in blue) has been assigned a name derived from their position in the phylogenomic tree built with the closest known relatives from databases and presented in Additional file 1: Figures S5–S12). Black genomes are from databases (cultures, single amplified genomes, SAGs, or MAGs). Figure S14. Phylogenetic analysis of a novel rhodonsin branch of Verrucomicrobia-Planctomycetes superphylum in marine waters. The evolutionary history was inferred by using the maximum likelihood method based on the JTT matrix-based model. A discrete gamma distribution was used to model evolutionary rate differences among sites (five categories). All positions with less than 80% site coverage were eliminated. Sequences in green were isolated from freshwater systems. Colored circles on the right side of sequences indicate the GC content (%) of the contig containing the rhodopsin. Protein sequences were downloaded from NCBI database

(www.ncbi.nlm.nih.gov). Tara contigs were downloaded from ENA database (www.ebi.ac.uk/ena). Accession numbers are within brackets. **Figure S15.** Abundance of genes affiliated with membrane transport function based on KEGG modules using principal component analysis (PCA) for each of the individual metagenomics samples. UP, upper photic; DCM, deep chlorophyll maximum; LP, lower photic; MIX, mixed water column (PDF 5881 kb).

Additional file 2: Table S1. Relative abundance of 16S rRNA reads. (XLSX 25 kb)

Additional file 3: Table S2. Summary statistics of the reconstructed genomes obtained from metagenomes. (XLSX 16 kb)

Additional file 4: Table S3. Relative abundance of functional gene categories related to motility and chemotaxis at subsystem level 3 (SEED database). The highest value for each one has been highlighted in red. (XLSX 13 kb)

Acknowledgements

Help from the crew and technicians of the CSIC R/V "Garcia del Cid" for the sampling is gratefully acknowledged. We thank Zachary Aanderud for providing helpful comments on the manuscript. This work was supported by grants "MEDIMAX" BFPU2013-48007-P, "VIREVO" CGL2016-76273-P [AEI/FEDER, EU], (co-founded with FEDER funds); Acciones de dinamización "REDES DE EXCELENCIA" CONSOLIDER CGL2015-71523-REDC from the Spanish Ministerio de Economía, Industria y Competitividad and PROMETEO II/2014/012 "AQUAMET" from Generalitat Valenciana. JHM was supported with a Ph.D. fellowship from the Spanish Ministerio de Economía y Competitividad (BES-2014-067828). MLP was supported with a postdoctoral fellowship from the Valencian Consellería de Educació, Investigació, Cultura i Esport (APOSTD/2016/051).

Availability of data and materials

Metagenomic datasets have been submitted to NCBI SRA and are available under BioProjects accession number PRJNA352798 (Med-OCT2015-15m [SRR5007106], Med-OCT2015-30m [SRR5007114], Med-OCT2015-45m [SRR5007115], Med-OCT2015-60m [SRR5007118], Med-OCT2015-75m [SRR5007138], Med-OCT2015-90 m [SRR5007139] and Med-OCT2015-1000m [SRR5007141]), and PRJNA257723 (MedWinter-JAN2015-20m [SRR3405540] and MedWinter-Jan2015-80m [SRR5877433]). The reconstructed genomes have been deposited as BioSample SAMN06890612 to SAMN06890655 and from SAMN08905455 to SAMN08905504 under BioProject PRJNA352798.

Authors' contributions

FRV conceived the study, helped with the analysis, and wrote the manuscript. JHM analyzed the data together with MLP and contributed to the writing of the manuscript. AC and AP helped in the sampling and analyzed all physicochemical and ecological parameters. JRT helped in analyzing the data and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Evolutionary Genomics Group, División de Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan de Alicante, 03550 Alicante, Spain.
²Department of Biology, San Francisco State University, San Francisco, CA 94132, USA. ³Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Burjassot, E-46100 Valencia, Spain.

Received: 29 March 2018 Accepted: 2 July 2018 Published online: 10 July 2018

References

 Christaki U, Van Wambeke F, Lefevre D, Lagaria A, Prieur L, Pujo-Pay M, et al. Microbial food webs and metabolic state across oligotrophic waters of the Mediterranean Sea during summer. Biogeosciences. 2011;8:1839–52.

- Bolhuis H, Cretoiu MS, Stal LJ. Molecular ecology of microbial mats. FEMS Microbiol Ecol. 2014;90:335–50.
- Scalan DJ, West NJ. Molecular ecology of the marine cyanobacteria genera Prochlorococcus and Synechococcus. FEMS Microbiol Ecol. 2002;40:1–12.
- Delong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N, et al. Community genomics among microbial assemblages in the ocean's interior. Science. 2006;311:496–503.
- Konstantinidis KT, Braff J, Karl DM, DeLong EF. Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific Subtropical Gyre. Appl Environ Microbiol. 2009;75:5345–55.
- Letelier RM, Karl DM, Abbott MR, Bidigare RR. Light driven seasonal patterns of chlorophyll and nitrate in the lower euphotic zone of the North Pacific Subtropical Gyre. Limnol Oceanogr. 2004;49:508–19.
- Ferreira AJSS, Siam R, Setubal JC, Moustafa A, Sayed A, Chambergo FS, et al. Core microbial functional activities in ocean environments revealed by global metagenomic profiling analyses. PLoS One. 2014;9:e97338.
- Walsh EA, Kirkpatrick JB, Rutherford SD, Smith DC, Sogin M, D'Hondt S. Bacterial diversity and community composition from seasurface to subseafloor. ISME J. 2015;10:1–11.
- Shi Y, Tyson GW, Eppley JM, DeLong EF. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. ISME J Nature Publishing Group;. 2011;5:999–1013.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, et al. The Sorcerer II Global Ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. PLoS Biol. 2007;5:0398–431.
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. Science. 2015; 1261359:348.
- Denaro G, Valenti D, La Cognata A, Spagnolo B, Bonanno A, Basilone G, et al. Spatio-temporal behaviour of the deep chlorophyll maximum in Mediterranean Sea: development of a stochastic model for picophytoplankton dynamics. Ecol Complex. 2013;13:21–34.
- Ghai R, Martin-Cuadrado A-B, Molto AG, Heredia IG, Cabrera R, Martin J, et al. Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and Fosmid library 454 pyrosequencing. ISME J. 2010;4:1154–66.
- Estrada M, Henriksen P, Gasol JM, Casamayor EO, Pedrós-Alió C. Diversity of planktonic photoautotrophic microorganisms along a salinity gradient as depicted by microscopy, flow cytometry, pigment analysis and DNA-based methods. FEMS Microbiol Ecol. 2004;49:281–93.
- Giovannoni SJ, Vergin KL. Seasonality in ocean microbial communities. Science. 2012;335:671–6.
- Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science (80-). American association for the Adv Sci. 2012;336:608–11.
- Hou S, López-Pérez M, Pfreundt U, Belkin N, Stüber K, Huettel B, et al. Benefit from decline: the primary transcriptome of Alteromonas macleodii str. Te101 during Trichodesmium demise. ISME J. 2018;12(4):981.
- Morris RM, Vergin KL, Cho J-C, Rappé MS, Carlson CA, Giovannoni SJ. Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. Limnol Oceanogr. 2005;50:1687–96.
- Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, et al. Fundamentals of microbial community resistance and resilience. Front Microbiol. 2012;3:1–19.
- Macias D, Garcia-Gorriz E, Stips A. Deep winter convection and phytoplankton dynamics in the NW Mediterranean Sea under present climate and future (horizon 2030) scenarios. Sci Rep Nat Publ Group. 2018;8:6626.
- Schauer M, Balagué V, Pedrós-Alió C, Massana R. Seasonal changes in the taxonomic composition of bacterioplankton in a oligotrophic coastal system. Aquat Microb Ecol. 2003;31:163–74.
- Alonso-Sáez L, Balagué V, Sà EL, Sánchez O, González JM, Pinhassi J, et al. Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. FEMS Microbiol Ecol. 2007;60:98–112.
- Garczarek L, Dufresne A, Rousvoal S, West NJ, Mazard S, Marie D, et al. High vertical and low horizontal diversity of Prochlorococcus ecotypes in the Mediterranean Sea in summer. FEMS Microbiol Ecol. 2007;60:189–206.
- 24. Barnum TP, Figueroa IA, Carlström CI, Lucas LN, Engelbrektson AL, Coates JD. Genome-resolved metagenomics identifies genetic mobility, metabolic

interactions, and unexpected diversity in perchlorate-reducing communities. ISME J. 2018;12(6):1568.

- Partensky F, Blanchot J, Vaulot D. Differential distribution and ecology of Prochlorococcus and Synechococcus in oceanic waters: a review. Bull l'Institut océanographique. 1999;19:457–75.
- Baltar F, Arístegui J, Sintes E, Van Aken HM, Gasol JM, Herndl GJ. Prokaryotic extracellular enzymatic activity in relation to biomass production and respiration in the meso- and bathypelagic waters of the (sub) tropical Atlantic. Environ Microbiol. 2009;11:1998–2014.
- Arnosti C. Patterns of microbially driven carbon cycling in the ocean: links between extracellular enzymes and microbial communities. Adv Oceanogr. 2014;2014:1–12.
- Luo H, Thompson LR, Stingl U, Hughes AL. Selection maintains low genomic GC content in marine SAR11 lineages. Mol Biol Evol. 2015;32:2738–48.
- Hugoni M, Taib N, Debroas D, Domaizon I, Jouan Dufournel I, Bronner G, et al. Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. Proc Natl Acad Sci U S A. 2013;110:6004–9.
- Miller DR, Pfreundt U, Elifantz H, Hess WR, Berman-Frank I. Microbial metatranscriptomes from the thermally stratified Gulf of Aqaba/Eilat during summer. Mar Genomics. 2017;32:23–6.
- Biller SJ, Berube PM, Lindell D, Chisholm SW. Prochlorococcus: the structure and function of collective diversity. Nat Rev Microbiol. 2014;13:13–27.
- Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. Science. 2014;344:416–20.
- Bendall ML, Stevens SL, Chan L-K, Malfatti S, Schwientek P, Tremblay J, et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J. 2016;10:1–13.
- Gonzaga A, Martin-Cuadrado AB, López-Pérez M, Mizuno CM, García-Heredia I, Kimes NE, et al. Polyclonality of concurrent natural populations of Alteromonas macleodii. Genome Biol Evol. 2012;4:1360–74.
- López-Pérez M, Kimes NE, Haro-Moreno JM, Rodriguez-Valera F. Not all particles are equal: the selective enrichment of particle-associated bacteria from the Mediterranean Sea. Front Microbiol. 2016;7:996.
- Haro-Moreno JM, Rodriguez-Valera F, López-García P, Moreira D, Martin-Cuadrado A-B. New insights into marine group III Euryarchaeota, from dark to light. ISME J. 2017;11(5):1102.
- López-Pérez M, Haro-Moreno JM, Gonzalez-Serrano R, Parras-Moltó M, Rodriguez-Valera F. Genome diversity of marine phages recovered from Mediterranean metagenomes: size matters. PLoS Genet. 2017;13:e1007018.
- Grote J, Cameron Thrash J, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, et al. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. MBio. 2012;3:1–13.
- Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, et al. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci U S A. 2013;110:11463–8.
- Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, et al. Genomic and proteomic characterization of "*Candidatus* Nitrosopelagicus brevis": an ammonia-oxidizing archaeon from the open ocean. Proc Natl Acad Sci. 2015;112:1173–8.
- Merbt SN, Stahl DA, Casamayor EO, Martí E, Nicol GW, Prosser JI. Differential photoinhibition of bacterial and archaeal ammonia oxidation. FEMS Microbiol Lett. 2012;327:41–6.
- Mella-Flores D, Mazard S, Humily F, Partensky F, Mahé F, Bariat L, et al. Is the distribution of Prochlorococcus and Synechococcus ecotypes in the Mediterranean Sea affected by global warming? Biogeosciences. 2011;8: 2785–804.
- Dupont CL, Rusch DB, Yooseph S, Lombardo MJ, Alexander Richter R, Valas R, et al. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. ISME J. 2012;6:1186–99.
- Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome streamlining in a cosmopolitan oceanic bacterium. Science. 2005; 309:1242–5.
- Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S, et al. The genomic basis of trophic strategy in marine bacteria. Proc Natl Acad Sci. 2009;106:15527–33.
- Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F. Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. Sci Rep. 2013;3:2471.

- Mizuno CM, Rodriguez-Valera F, Ghai R. Genomes of planktonic acidimicrobiales: widening horizons for marine actinobacteria by metagenomics. MBio. 2015;6:e02083–14.
- Galand PE, Gutiérrez-Provecho C, Massana R, Gasol JM, Casamayor EO. Interannual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). Limnol Oceanogr. 2010;55:2117–25.
- Miller D, Pfreundt U, Hou S, Lott SC, Hess WR, Berman-Frank I. Winter mixing impacts gene expression in marine microbial populations in the Gulf of Aqaba. Aquat Microb Ecol. 2017;80:223–42.
- Buchan A, LeCleir GR, Gulvik CA, González JM. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat Rev Microbiol. 2014;12:686–98.
- Fuhrman JA, Cram JA, Needham DM. Marine microbial community dynamics and their ecological interpretation. Nat Rev Microbiol. 2015;13:133–46.
- Eilers H, Pernthaler J, Peplies J, Glöckner FO, Gerdts G, Amann R. Isolation of novel pelagic bacteria from the German bight and their seasonal contributions to surface picoplankton. Appl Environ Microbiol. 2001;67:5134–42.
- Yan S, Fuchs BM, Lenk S, Harder J, Wulf J, Jiao NZ, et al. Biogeography and phylogeny of the NOR5/OM60 clade of Gammaproteobacteria. Syst Appl Microbiol. 2009;32:124–39.
- Fuhrman JA, Schwalbach MS, Stingl U. Proteorhodopsins: an array of physiological roles? Nat Rev Microbiol. 2008;6:488–94.
- Pinhassi J, DeLong EF, Béjà O, González JM, Pedrós-Alió C. Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. Microbiol Mol Biol Rev. 2016;80:929–54.
- Olson DK, Yoshizawa S, Boeuf D, Iwasaki W, DeLong EF. Proteorhodopsin variability and distribution in the North Pacific Subtropical Gyre. ISME J. 2018;12(4):1047.
- Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. Science. 2012;335:587–90.
- Boeuf D, Audic S, Brillet-Guéguen L, Caron C, Jeanthon C. MicRhoDE: a curated database for the analysis of microbial rhodopsin diversity and evolution. Database. 2015;2015:bav080. https://doi.org/10.1093/database/ bav080.
- Sabehi G, Kirkup BC, Rozenberg M, Stambler N, Polz MF, Béjà O. Adaptation and spectral tuning in divergent marine proteorhodopsins from the eastern Mediterranean and the Sargasso Seas. ISME J. 2007;1:48–55.
- Cabello-Yeves PJ, Ghai R, Mehrshad M, Picazo A, Camacho A, Rodriguez-Valera F. Reconstruction of diverse verrucomicrobial genomes from metagenome datasets of freshwater reservoirs. Front Microbiol. 2017;8:2131.
- Cabello-Yeves PJ, Zemskay TI, Rosselli R, Coutinho FH, Zakharenko AS, Blinov W, et al. Genomes of novel microbial lineages assembled from the sub-ice waters of Lake Baikal. Appl Environ Microbiol. 2018;84:e02132-17
- Gushchin I, Chervakov P, Kuzmichev P, Popov AN, Round E, Borshchevskiy V, et al. Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. Proc Natl Acad Sci. 2013;110: 12631–6.
- Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 2005;33:5691–702.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2013;42(D1):D490–5.
- Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ, et al. Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. Proc Natl Acad Sci U S A. 2010;107:8818–23.
- Fernández-Gómez B, Richter M, Schüler M, Pinhassi J, Acinas SG, González JM, et al. Ecology of marine Bacteroidetes: a comparative genomics approach. ISME J. 2013;7:1026–37.
- 67. Painter TJ. 4-Algal polysaccharides. In: Aspinall GO, editor. The polysaccharides. Academic press; 1983. p195–285.
- Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, et al. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. ISME J. 2013;7:2287–300.
- 69. del Carmen Munoz-Marin M, Luque I, Zubkov MV, Hill PG, Diez J, Garcia-Fernandez JM. Prochlorococcus can use the Pro1404 transporter to take up

glucose at nanomolar concentrations in the Atlantic Ocean. Proc Natl Acad Sci. 2013;110:8597–602.

- Zubkov MV, Tarran GA, Fuchs BM. Depth related amino acid uptake by Prochlorococcus cyanobacteria in the Southern Atlantic tropical gyre. FEMS Microbiol Ecol. 2004;50:153–61.
- Yelton AP, Acinas SG, Sunagawa S, Bork P, Pedrós-Alió C, Chisholm SW. Global genetic capacity for mixotrophy in marine picocyanobacteria. ISME J. 2016;10(12):2946.
- López-Pérez M, Gonzaga A, Martin-Cuadrado A-B, Onyshchenko O, Ghavidel A, Ghai R, et al. Genomes of surface isolates of Alteromonas macleodii: the life of a widespread marine opportunistic copiotroph. Sci Rep. 2012;2:1–11.
- Schauer K, Rodionov DA, de Reuse H. New substrates for TonBdependent transport: do we only see the 'tip of the iceberg'? Trends Biochem Sci. 2008;33(7):330–8.
- 74. Ziegler C, Bremer E, Krämer R. The BCCT family of carriers: from physiology to crystal structure. Mol Microbiol. 2010;78(1):13–34.
- Gifford SM, Sharma S, Booth M, Moran MA. Expression patterns reveal niche diversification in a marine microbial assemblage. ISME J. 2013;7:281–98.
- Wolfe-Simon F, Grzebyk D, Schofield O, Falkowski PG. The role and evolution of superoxide dismutases in algae. J Phycol. 2005;41(3):453–65.
- 77. Kaiser K, Benner R. Major bacterial contribution to the ocean reservoir of detrital organic carbon and nitrogen. Limnol Oceanogr. 2008;53:99–112.
- Hutchins DA, Fu F. Microorganisms and ocean global change. Nat Microbiol. 2017;2:1–11.
- Martin-Cuadrado A-B, Rodriguez-Valera F, Moreira D, Alba JC, Ivars-Martínez E, Henn MR, et al. Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. ISME J. 2008;2:865–86.
- Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28:1420–8.
- Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 2010;11:119.
- Lowe TM, Eddy SR. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1996;25:955–64.
- Nawrocki EP. Structural RNA homology search and alignment using covariance models. All Theses and Dissertations (ETDs). 2009;256. http://dx. doi.org/10.7936/K78050MP.
- Huang Y, Gilna P, Li W. Identification of ribosomal RNA genes in metagenomic fragments. Bioinformatics. 2009;25:1338–40.
- 85. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.
- Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 2001;29:22–8.
- Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, et al. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res. 2001;29:41–3.
- 88. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1002195.
- Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16:276–7.
- Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, et al. MEGAN Community Edition—interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput Biol. 2016;12(6):e1004957.
- American Public Health Association. Standard Methods for the Examination of Water and Wastewater. 21st ed. American Public Health Association (APHA), Washington DC. 2005; 1220p.
- Picazo A, Rochera C, Vicente E, Miracle MR, Camacho A. Spectrophotometric methods for the determination of photosynthetic pigments in stratified lakes: a critical analysis based on comparisons with HPLC determinations in a model lake. Limnetica. 2013;32:139–58.
- Gasol JM, Li Zweifel U, Peters F, Fuhrman JA, Hagström Å. Significance of size and nucleic acid content heterogeneity as measured by flow cytometry in natural planktonic bacteria. Appl Environ Microbiol. 1999; 65:4475–83.
- Marie D, Partensky F, Jacquet S, Vaulot D. Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. Appl Environ Microbiol. 1997;63:186–93.

- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal database project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42:633–42.
- Eddy SR. Multiple alignment using hidden Markov models. Proc Int Conf Intell Syst Mol Biol. 1995;3:114–20.
- Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Softw. 2008;25:1–18.
- 98. Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P. Prediction of effective genome size in metagenomic samples. Genome Biol. 2007;8:R10.
- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol. 2013;31:533–8.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to singlecell sequencing. J Comput Biol. 2012;19:455–77.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
- Lassmann T, Sonnhammer ELL. Kalign—an accurate and fast multiple sequence alignment algorithm. BMC Bioinf. 2005;6:298.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol Oxford University Press; 2016;33(7):1870–4.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60.
- 107. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
- Man D, Wang W, Sabehi G, Aravind L, Post AF, Massana R, et al. Diversification and spectral tuning in marine proteorhodopsins. EMBO J. 2003;22:1725–31.
- 109. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. DbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012; 40(W1):W445–51.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–62.
- Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428:726–31.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions





Fine metagenomic profile of the Mediterranean stratified and mixed water column revealed by assembly and recruitment

Jose M. Haro-Moreno^{1,4}, Mario López-Pérez^{1,4}, José R. de la Torre², Antonio Picazo³, Antonio Camacho³, Francisco Rodriguez-Valera^{1*}.

¹Evolutionary Genomics Group, División de Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan 03550, Alicante, Spain. ²Department of Biology, San Francisco State University, San Francisco, CA 94132, USA. ³Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Burjassot, E-46100 Valencia, Spain. ⁴These authors contributed equally to this work.

*Corresponding author: frvalera@umh.es

Evolutionary Genomics Group, División de Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan de Alicante, 03550 Alicante, Spain.

Phone +34-965919313, Fax +34-965 919457

Additional files



Figure S1. Method used for sampling. Water was pumped through a hose directly on to the filters instead of using the Niskin bottles rossette.



Figure S2. Bar plot showing the concentration of inorganic nutrients in both stratified (blue) and mixed samples (red).



Figure S3. Simpson diversity index versus depth.



Figure S4. Assembled contigs. A) Size of individual contigs to the left and total assembled size to the right for each phylum. Proteobacteria was divided into its class-level taxonomy. The number of contigs longer than 10 Kb that were taxonomically classified are indicated within brackets. B) Individual contribution of each metagenome to the total assembled size.

2.5 5.0 7.5

2.5

7.5

Mb assembled

0.25 0.50 0.75

A)



Figure S5. Phylogenetic analysis of Actinobacteria metagenome assembled genomes (MAGs). A maximum likelihood genome tree was constructed with 100 bootstraps using 31 conserved proteins among the 20 genomes compared. Black circles represent bootstrap values. Between brackets: ANI, average nucleotide identity; COV, percentage of genome sequence shared. In red, those MAGs retrieved in this work.



Figure S6. Phylogenetic analysis of Alphaproteobacteria MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 46 conserved proteins among the 40 genomes compared. Black circles represent bootstrap values. In red, those MAGs retrieved in this work.



Figure S7. Phylogenetic analysis of Bacteroidetes MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 21 conserved proteins among the 29 genomes compared. Black circles represent bootstrap values. In red, those MAGs retrieved in this work.



Figure S8. Phylogenetic analysis of Cyanobacteria MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 286 conserved proteins among the 45 genomes compared. Black circles represent bootstrap values. In red, those MAGs retrieved in this work.



Figure S9. Phylogenetic analysis of Euryarchaeota MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 26 conserved proteins among the 20 genomes compared. Black circles represent bootstrap values. In red, those MAGs retrieved in this work.



Figure S10. Phylogenetic analysis of Gammaproteobacteria MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 32 conserved proteins among the 44 genomes compared. Black circles represent bootstrap values. In red, those MAGs retrieved in this work.



Figure S11. Phylogenetic analysis of Verrucomicrobia MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 79 conserved proteins among the 25 genomes compared. Black circles represent bootstrap values. In red, those MAGs retrieved in this work.


Figure S12. Phylogenetic analysis of Thaumarchaeota MAGs. A maximum likelihood genome tree was constructed with 100 bootstraps using 129 conserved proteins among the 17 genomes compared. Black circles represent bootstrap values. In red, those MAGs retrieved in this work.



Figure S13. Relative abundance of the reconstructed and reference genomes measured by recruitment (RPKG, Reads per Kilobase of genome and Gigabase of metagenome) from the different depths of the stratified metagenomes. To show the relationships among genomes, a maximum likelihood genome tree was constructed using all the conserved proteins (number in colored square). Each MAG (in blue) has been assigned a name derived from their position in the phylogenomic tree built with the closest known relatives from databases and presented in Figures S5 to S11). Black genomes are from databases (cultures, single amplified genomes, SAGs or MAGs).



Figure S14. Phylogenetic analysis of a novel rhodopsin branch of Verrucomicrobia-

Planctomycetes superphylum in marine waters. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories). All positions with less than 80% site coverage were eliminated. Sequences in green were isolated from freshwater systems. Colored circles on the right side of sequences indicate the GC content (%) of the contig containing the rhodopsin. Protein sequences were downloaded from NCBI database (www.ncbi.nlm.nih.gov). Tara contigs were downloaded from ENA database (www.ebi.ac.uk/ena). Accession numbers are within brackets.



Figure S15. Abundance of genes affiliated to membrane transport function based on KEGG modules using Principal Component Analysis (PCA) for each of the individual metagenomic samples. UP: upper photic, DCM: deep chlorophyll maximum, LP: lower photic and MIX: mixed water column.

Demain	Dhuduum	Class	Family	MedWinter-	MedWinter-	Med-	Med-	Med-	Med-	Med-	Med-
Domain	Phylum	Class	Family	JAN2015-	JAN2015-	0C12015- 15m	OC12015-	0C12015-	OC12015-	OC12015-	OC12015-
Archaoa				20m	5.04	0.03	0 70	4011	11 12	13 50	15 78
Alciidea	Funvarchagota			1.50	2.72	0.03	0.79	5.04	5 50	10.00	5.81
	Euryarchaeota	Thormonlaemata		4.45	1.04	0.02	0.72	1.56	2.35	4.97	2.68
		mermoplasmata	Methanomassiliicoccaceae	0.36	0.12	0.00	0.23	0.02	0.16	0.16	0.20
			unc Thermonlasmata	1 49	0.92	0.00	0.18	1.54	2 19	1.83	2.46
		Halobacteria		0.03	0.00	0.00	0.00	0.00	0.02	0.08	0.02
		Indiosaotoria	Halobacteriaceae	0.03	0.00	0.00	0.00	0.00	0.02	0.08	0.02
		unc Eurvarchaeota		2.47	1.66	0.00	0.47	3.46	3.22	2.91	3.10
	Thaumarchaeota			2.83	2.32	0.00	0.05	1.34	5.51	8.58	9.95
			Nitrosopumilaceae	2.82	2.32	0.00	0.05	1.34	5.51	8.56	9.85
	unc_Archaea		·	0.00	0.00	0.02	0.02	0.29	0.00	0.00	0.02
Bacteria				92.70	94.96	99.97	99.21	93.33	88.88	86.41	84.22
	Acidobacteria			0.08	0.04	0.00	0.00	0.00	0.12	0.21	0.60
		Acidobacteria_Gp6		0.01	0.00	0.00	0.00	0.00	0.08	0.08	0.38
		Acidobacteria_Gp21		0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.10
		unc_Acidobacteria		0.00	0.00	0.00	0.00	0.00	0.04	0.06	0.10
	Actinobacteria			2.38	3.42	1.06	1.19	1.41	2.37	3.24	2.72
		Actinobacteria		2.38	3.42	1.06	1.19	1.41	2.37	3.24	2.72
			Acidimicrobineae_incertae_sedis	0.00	0.00	0.00	0.00	0.00	0.08	0.10	0.00
			lamiaceae	0.62	0.52	0.00	0.02	0.13	0.30	0.27	0.32
			Microbacteriaceae	0.07	0.87	0.07	0.02	0.02	0.02	0.00	0.00
			Mycobacteriaceae	0.13	0.13	0.09	0.11	0.18	0.24	0.39	0.30
			unclassified_"Acidimicrobineae"	0.12	0.18	0.02	0.00	0.59	1.17	1.83	1.78
			unclassified_Actinobacteria	1.11	1.35	0.72	0.86	0.33	0.28	0.31	0.08
	Bacteroidetes			8.33	7.03	6.95	5.90	8.90	5.02	3.55	2.64
		Bacteroidia		0.25	0.19	0.26	0.09	0.20	0.06	0.06	0.02
			Bacteroidaceae	0.12	0.05	0.14	0.05	0.18	0.04	0.02	0.00
		Cytophagia		0.49	0.40	0.03	0.45	0.99	0.73	0.41	0.34
			Flammeovirgaceae	0.03	0.06	0.02	0.38	0.66	0.40	0.19	0.18
			unclassified_Cytophagales	0.46	0.32	0.00	0.07	0.31	0.28	0.21	0.16
		Flavobacteriia	- ·	/.1/	5.90	5.68	4.34	7.25	3.79	2.96	2.06
			Cryomorphaceae	1.67	1.12	0.38	0.56	0.92	0.83	0.62	0.34
			Flavobacteriaceae	4.62	4.13	4.99	3.42	5.81	2.53	2.16	1.58
		Cabinachesterile	unc_Flavobacteriales	0.88	0.65	0.31	0.36	0.51	0.43	0.18	0.14
		Springobacterna	Chitingsharessa	0.09	0.07	0.40	0.45	0.13	0.14	0.02	0.00
			Chilinophagaceae	0.02	0.00	0.43	0.41	0.09	0.02	0.00	0.00
		una Pastaraidatas	Saprospiraceae	0.00	0.02	0.00	0.05	0.02	0.10	0.02	0.00
	condidate division 7P2	unc_bacteroidetes		0.04	0.45	0.07	0.54	0.33	0.30	0.10	0.22
	Chloroflevi			0.40	0.28	0.07	0.07	0.13	0.02	1 11	2 14
	onioronexi	unc Chloroflexi		0.38	0.27	0.02	0.00	0.10	0.65	1.03	1.96
	Cvanobacteria		And the state of the state of the state of the	6.54	5.08	5.83	9.79	4.89	5.35	7.92	3.24
	ojanobaotona	Cvanobacteria	the first the last the second	6.54	5.08	5.83	9.79	4.89	5.35	7.92	3.24
		-,	Family I	0.22	0.13	0.00	0.05	0.02	0.02	0.06	0.02
			Family II	6.18	4.90	5.75	9.61	4.67	5.26	7.80	3.10
			Family IX	0.14	0.03	0.03	0.05	0.11	0.02	0.00	0.04
	Deinococcus-Thermus			0.13	0.15	0.05	0.02	0.02	0.00	0.08	0.02
		Deinococci		0.13	0.15	0.05	0.02	0.02	0.00	0.08	0.02
			Deinococcaceae	0.12	0.15	0.05	0.02	0.02	0.00	0.04	0.02
	Firmicutes			0.27	0.53	0.12	0.25	0.64	0.57	0.49	0.88
		Bacilli		0.09	0.29	0.03	0.05	0.20	0.14	0.12	0.20
			Alicyclobacillaceae	0.00	0.13	0.02	0.00	0.02	0.00	0.02	0.08
		Clostridia		0.00	0.18	0.07	0.14	0.33	0.36	0.31	0.48
			Clostridiaceae 1	0.00	0.04	0.02	0.05	0.00	0.10	0.06	0.12
			Peptococcaceae 1	0.00	0.01	0.00	0.00	0.04	0.06	0.12	0.08
		Negativicutes		0.00	0.03	0.00	0.02	0.09	0.08	0.04	0.16
			Veillonellaceae	0.00	0.00	0.00	0.02	0.09	0.08	0.04	0.16
	Marinimicrobia			2.48	2.32	1.17	3.11	5.42	4.58	5.17	6.27
	Nitrospinae			0.16	0.14	0.00	0.02	0.04	0.38	1.09	0.92
		Nitrospinia		0.16	0.14	0.00	0.02	0.04	0.38	1.09	0.92
	Disastenus		Nitrospinaceae	0.00	0.14	0.00	0.02	0.04	0.38	1.09	0.92
	Planctomycetes	Dhuadaalaasaa		1.14	0.96	0.34	0.23	0.15	0.77	1.15	1.02
		Phycisphaerae	Discional	0.11	0.10	0.14	0.11	0.04	0.26	0.35	0.34
		Dispetantusetia	Phycisphaeraceae	0.11	0.10	0.14	0.11	0.04	0.26	0.35	0.54
		Planctomycetia	Disastemuseteese	0.00	0.64	0.17	0.11	0.11	0.51	0.60	0.04
			Planctomycetaceae	0.00	0.77	0.17	0.11	0.09	0.49	0.64	0.40
	Brotochastoria		unc_Plancionlycella	64.96	70.47	0.00	72.02	66.00	61.42	55.20	56.01
	FIOLEODACLEIIA	Alphaprotophastoria		41.60	10.47	67.02	52.05	45.50	42.26	24.92	24.06
		Alphaproteopacteria	Aastabastaraasa	41.62	44.29	0.02	52.05	45.54	42.20	34.62	34.06
			Bartopollacoao	0.00	0.14	0.02	0.00	0.02	0.00	0.00	0.04
			Bradyrhizobiacoao	0.08	0.14	0.03	0.00	0.02	0.00	0.02	0.02
			Caulobacteraceae	0.00	0.14	0.12	0.09	0.07	0.12	0.10	0.10
1			Frythrobacteraceae	0.06	0.23	0.03	0.05	0.02	0.00	0.12	0.04
			Hyphomonadaceae	0.12	0.57	0.00	0.00	0.02	0.00	0.00	0.00
			Hyphomicrobiaceae	0.06	0.05	0.03	0.00	0.02	0.06	0.06	0.04
			Magnetococcaceae	0.08	0.02	0.17	0.14	0.22	0.08	0.06	0.06
			Methylocystaceae	0.02	0.01	0.03	0.02	0.00	0.02	0.00	0.00
			Parvularculaceae	0.01	0.05	0.00	0.00	0.00	0.16	0.10	0.00
			Phyllobacteriaceae	0.09	0.13	0.03	0.05	0.02	0.00	0.02	0.00
			Rhodobacteraceae	2.77	2.06	2.92	2.59	4.29	1.60	0.70	0.42
			Rhodospirillaceae	2.10	1.50	2.83	2.21	1.23	1.84	1.07	1.02
			SAR11	24.80	27.62	39.30	37.87	32.68	30.01	26.18	26.17
			Sphingomonadaceae	0.38	1.81	0.74	0.16	0.13	0.18	0.55	0.26
-				-	3						

Table S1. Relative abundance of 16S rRNA reads.

		unc Alphaproteobacteria	9.88	8.39	10.72	8.10	5.11	6.86	5.09	5.27
		unc Rhizobiales	0.24	0.19	0.33	0.23	1.41	0.91	0.35	0.24
		unc Rhodospirillales	0.23	0.06	0.15	0.25	0.13	0.20	0.10	0.24
	Betaproteobacteria		0.66	0.88	0.39	0.29	0.44	0.40	0.35	0.16
	Dotaprotoobaotona	Burkholderiaceae	0.05	0.21	0.00	0.00	0.02	0.04	0.04	0.00
		Methylophilaceae	0.00	0.16	0.00	0.05	0.02	0.04	0.04	0.00
		Phodocyclacoao	0.13	0.10	0.00	0.00	0.24	0.10	0.19	0.10
			0.04	0.07	0.00	0.00	0.02	0.00	0.00	0.02
		unc_Betaproteobacteria	0.14	0.09	0.02	0.05	0.07	0.08	0.02	0.02
	Delterrete else sterie	unc_Burkholderlaies	0.03	0.02	0.36	0.10	0.02	0.10	0.02	0.00
	Deltaproteobacteria	Desteriousses	0.02	0.05	0.05	0.70	0.01	1.00	2.13	2.36
		Bacteriovoracaceae	0.03	0.05	0.05	0.05	0.00	0.22	0.12	0.16
		Desulturomonadaceae	0.06	0.02	0.02	0.05	0.00	0.00	0.08	0.04
		unc_Deitaproteobacteria	0.62	0.52	0.39	0.50	0.77	1.23	1.75	1.96
	Gammaproteobacteria	1	20.37	23.48	21.34	20.08	18.04	13.40	12.63	13.56
		Alcanivoracaceae	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.00
		Alteromonadaceae	0.04	0.52	0.10	0.18	0.18	0.34	0.21	0.16
		Alteromonadales_incertae_sedis	0.22	0.69	0.05	0.02	0.04	0.00	0.02	0.00
		Coxiellaceae	0.06	0.08	0.27	0.07	0.02	0.10	0.14	0.08
		Ectothiorhodospiraceae	0.14	0.09	0.02	0.00	0.02	0.14	0.14	0.22
		Enterobacteriaceae	0.23	0.18	0.21	0.11	0.13	0.06	0.08	0.16
		Halomonadaceae	0.09	0.17	0.07	0.00	0.04	0.12	0.04	0.04
		Legionellaceae	0.28	0.31	0.03	0.02	0.04	0.04	0.08	0.06
		Oceanospirillaceae	0.06	0.07	0.12	0.07	0.13	0.02	0.02	0.04
		Pseudoalteromonadaceae	0.21	0.09	0.07	0.07	0.07	0.06	0.06	0.04
		Pseudomonadaceae	0.69	6.49	0.57	0.27	0.33	0.40	0.37	0.30
		Shewanellaceae	0.15	0.14	0.03	0.07	0.00	0.00	0.02	0.00
		Thiotrichales_incertae_sedis	0.00	0.00	0.00	0.00	0.04	0.02	0.00	0.00
		Vibrionaceae	0.11	0.10	0.17	0.20	0.07	0.06	0.00	0.12
		Xanthomonadaceae	0.03	0.07	0.00	0.02	0.02	0.00	0.02	0.02
		unc_Chromatiales	0.41	0.24	0.02	0.14	0.02	0.02	0.06	0.08
		unc_Gammaproteobacteria	16.37	13.04	18.97	17.78	15.44	10.85	10.37	11.05
		unc_Gammaproteobacteria_incertae	0.16	0.17	0.03	0.02	0.20	0.43	0.49	0.74
		unc Oceanospirillales	0.03	0.12	0.00	0.02	0.02	0.08	0.00	0.04
	unc_Proteobacteria		1.04	0.80	0.63	0.77	2.03	3.75	5.42	5.63
Spirochaetes			0.18	0.16	0.09	0.23	0.13	0.04	0.02	0.00
	Spirochaetia		0.18	0.16	0.09	0.23	0.13	0.04	0.02	0.00
	·	Spirochaetaceae	0.17	0.16	0.05	0.18	0.11	0.04	0.02	0.00
errucomicrobia			2.10	1.58	1.85	2.18	0.62	2.04	1.35	1.56
	Opitutae		0.78	0.68	1.72	1.69	0.35	0.49	0.21	0.24
		Puniceicoccaceae	0.73	0.64	1.70	1.69	0.35	0.49	0.21	0.24
	Subdivision3		0.35	0.34	0.03	0.11	0.09	0.24	0.04	0.02
	Verrucomicrobiae		0.64	0.42	0.07	0.20	0.04	0.61	0.60	0.70
		Verrucomicrobiaceae	0.63	0.42	0.07	0.20	0.04	0.61	0.58	0.64
	unc Verrucomicrobia		0.31	0.14	0.03	0.18	0.13	0.65	0.47	0.58
			0.01	0	0.00	0.10	0.10	0.00	0.11	0.00
inc Bacteria			3.47	2 70	1.53	1 94	2.88	5 16	5 11	5 75
			0.47	2.10	1.00	1.04	2.00	0.10	0.11	0.70

	#contigs	Size (Mb)	%GC (±SD)	%Completeness Raes / Albertsen / CheckM	% Contamination CheckM
Bacteria					
Bacteria MED-G45	22	0.45	31.9 (1.1)	88.6 / 50.0 / 45.9	0.0
Bacteria MED-G46	24	1.23	35.0 (1.4)	94.3 / 65.7 / 68.1	0.0
Bacteria MED-G47	36	0.72	28.2 (1.3)	82.9 / 56.8 / 51.7	0.0
Actinobacteria					
Acidimicrobiales MED-G01	37	1.96	53.1 (0.5)	85.7 / 63.1 / 87.7	0.8
Ca. Actinomarinales MED-G02	22	0.49	34.1 (1.5)	88.6 / 56.8 / 45.0	1.3
Microbacterium MED-G48	199	1.66	66.6 (2.1)	60.0 / 44.1 / 55.1	0.0
Alphaproteobacteria					
Sphingomonadaceae MED-G03	69	3.15	66.3 (1.0)	100.0 / 83.8 / 95.3	0.2
SAR116 MED-G04	65	1.56	57.4 (1.4)	74.3/67.6/79.5	0.5
SAR116 MED-G05	58	1.66	57.1 (1.3)	80.0 / 61.3 / 67.3	0.0
SARTI6 MED-G06	85	1.67	61.1 (1.2)	42.9 / 52.25 / 80.5	0.5
Rhodobacteraceae MED-GU7	41	1.1	40.2 (0.9)	17.1/33.3/50.0	0.0
	39	2.21	42.3 (1.2) 22.0 (1.2)	77 1 / 64 0 / 56 6	0.3
	10	0.75	32.9 (1.2) 20.2 (1.6)	01 4 / 72 0 / 74 7	0.4
Hyphomonas MED-G10	42	0.99	58.8 (1.7)	91.4 / 73.5 / 74.7	0.8
PS1 MED-G50	50	1 32	45 5 (1.9)	97 1 / 80 2 / 75 3	0.0
Alphanroteobacteria MED-G51	38	1.52	29.5 (1.5)	88.6/63.1/68.8	0.0
Rhodobacteraceae MED-G52	48	1.10	39.9 (1.6)	74 3 / 57 7 / 62 7	0.0
SAR116 MED-G54	37	1.34	41 4 (1 2)	91 4 / 75 7 / 80 1	0.0
PS1 MED-G55	42	1.35	46.3 (1.3)	91 4 / 68 5 / 83 4	0.0
Alphaproteobacteria MED-G56	28	1.56	37 2 (0.8)	97 1 / 79 3 / 81 1	0.5
Bacteroidetes	20	1.00	01.2 (0.0)	0111770.070111	0.0
Cryomorphaceae MED-G11	30	0.85	30 1 (1 2)	91 4 / 65 8 / 73 9	0.0
Rhodothermaeota MED-G12	54	1.38	41 4 (1.9)	80 0 / 67 6 / 76 7	0.0
Bacteroidetes MED-G13	32	1.06	30.2 (1.8)	88.6 / 71.2 / 63.3	0.3
Cryomorphaceae MED-G14	43	1.06	28.6 (1.3)	85.1 / 61.3 / 66.5	0.0
Flavobacteriales MED-G15	55	0.93	37.4 (1.1)	17.1 / 31.5 / 55.2	1.7
Rhodothermaeota MED-G16	39	1.01	29.5 (2.2)	80.0 / 60.4 / 54.9	0.6
Bacteroidetes MED-G17	17	1.67	35.4 (0.8)	97.1 / 82.0 / 95.6	0.6
Rhodothermaeota MED-G18	38	1.24	31.4 (2.1)	88.6 / 66.7 / 63.8	0.4
Rhodothermaeota MED-G19	25	1.31	29.8 (1.8)	94.3 / 76.6 / 68.4	0.9
Bacteroidetes MED-G20	60	1.27	30.1 (1.2)	40.0 / 45.1 / 65.9	0.7
Bacteroidetes MED-G21	66	1.40	34.2 (1.9)	91.4 / 72.1 / 74.4	0.0
Flavobacteriales MED-G22	32	1.06	36.5 (1.2)	88.6 / 57.8 / 55.9	0.0
Flavobacteriales MED-G58	78	1.66	35.8 (1.1)	42.9 / 52.3 / 72.9	0.0
Flavobacteriales MED-G59	19	1.46	39.4 (1.5)	94.3 / 81.1 / 87.5	0.0
Cryomorphaceae MED-G60	27	1.08	30.3 (1.1)	91.4 / 72.1 / 67.7	0.1
Cryomorphaceae MED-G61	36	0.87	28.2 (1.4)	80.0 / 66.6 / 66.9	0.0
Bacteroidetes MED-G62	35	0.79	30.1 (1.7)	85.7 / 54.9 / 47.6	0.0
Cryomorphaceae MED-G63	38	0.83	29.1 (1.3)	45.7 / 45.1 / 61.9	0.0
Cryomorphaceae MED-G65	43	1.54	29.5 (1.4)	91.4 / 71.2 / 81.2	0.2
Cyanobacteria					
Synechococcus MED-G67	30	1.68	62.2 (1.5)	94.3 / 70.3 / 81.5	0.3
Synechococcus MED-G70	36	1.80	63.1 (2.5)	77.1/64.9/84.4	0
Prochlorococcus MED-G72	22	1.28	30.5 (1.4)	85.7 / 67.6 / 79.4	0
Prochlorococcus MED-G73	37	0.78	35.7 (1.6)	74.3 / 54.1 / 55.8	0
Gammaproteobacteria					
SUP05 MED-G23	36	0.85	55.6 (1.4)	(1.1/5/.7/62.6	0.0
OM182 MED-G24	86	2.09	55.2 (0.7)	40.0 / 51.4 / 75.2	1.1
SUP05 MED-G25	33	1.61	51.3 (1.7)	88.6 / 77.5 / 86.3	1.2
OM60/NOR5 MED-G26	62	1.28	53.7 (0.8)	17.1 / 39.6 / 55.2	0.0
OM60/NOR5 MED-G27	68	2.02	52.9 (1.4)	91.4 / 69.4 / 81.6	0.1
OM182 MED-G28	25	2.95	43.9 (0.9)	94.3 / 82.9 / 88.7	0.5
SAR92 MED-G29	53	1.29	42.5 (1.2)	100.0 / 76.6 / 71.1	0.0
Pseudomonas MED-G/4	304	4.06	63.9 (3.0)	40.0 / 51.4 / /3.4	1.1

 Table S2.- Summary statistics of the reconstructed genomes obtained from metagenomes.

	40	0.02	526(07)	77 1 / 52 2 / 53 2	0.0
	40 56	1 10	32.0(0.7)	01 4 / 76 6 / 60 2	0.0
SARSE MED C78	25	0.75	47.0 (1.0) 33.1 (1.3)	91.4 / 70.0 / 09.3	0.0
SARGO MED CZO	23	1.00	33.1(1.3)	04.2 / 70.2 / 74.2	0.0
SAROO MED-G79	20	1.00	35.1(0.0)	94.3 / 79.3 / 74.7	0.0
	29	1.31	35.5 (1.1)	94.3 / 73.9 / 63.0	1.4
SAR86 MED-G82	38	0.80	35.2 (1.3)	74.3 / 55.9 / 51.2	0.0
SAR86 MED-G83	11	0.94	36.6 (0.9)	94.3 / 80.2 / 75.4	0.0
SAR86 MED-G84	33	0.86	35.9 (1.2)	94.3 / 66.7 / 50.8	0.1
SAR86 MED-G85	15	1.06	31.3 (0.8)	94.3 / 76.6 / 63.3	0.0
Verrucomicrobia					
Puniceicoccaceae MED-G30	67	1.73	52.6 (2.2)	97.1 / 73.9 / 89.9	0.0
Puniceicoccaceae MED-G31	40	1.1	42.9 (2.1)	94.3 / 63.1 / 65.2	0.0
Puniceicoccaceae MED-G32	30	0.67	38.3 (1.5)	91.4 / 60.4 / 69.6	0.0
Verrucomicrobia MED-G86	101	3.19	55.0 (1.2)	77.1 / 73.9 / 95.0	0.7
Puniceicoccaceae MED-G87	52	1.68	49.8 (1.9)	85.7 / 73.9 / 92.1	0.0
Verrucomicrobia MED-G88	75	1.44	38.5 (2.2)	40.0 / 34.2 / 51.8	0.0
Euryarcheaota					
EUII MED-G33	66	1.27	48.3 (1.2)	60.0 / 19.8 / 57.3	0.0
EUII MED-G34	31	1.03	53.8 (1.6)	57.1 / 24.3 / 59.4	0.0
EUII MED-G35	29	1.20	52.7 (3.3)	68.6 / 26.1 / 65.4	0.0
EUII MED-G36	40	0.89	43.8 (1.4)	74.3 / 21.6 / 56.4	0.0
EUII MED-G37	25	1.28	45.4 (1.3)	85.7 / 29.7 / 71.7	0.0
EUII MED-G38	21	1.37	35.9 (0.9)	54.3 / 26.1 / 73.9	0.0
EUII MED-G90	23	1.46	49.5 (3.9)	91.4 / 31.5 / 85.9	0.8
EUII MED-G91	19	1.73	48.6 (1.4)	88.6 / 29.7 / 79.7	0
EUII MED-G92	11	1.50	51.9 (1.3)	91.4 / 30.6 / 81.6	0
EUII MED-G93	20	1.37	38.1 (0.9)	88.6 / 32.4 / 82.0	0
Thaumarchaeota					
Nitrosopumilus MED-G94	16	1.24	31.2 (0.8)	97.1 / 32.4 / 100.0	0
Pelagibacterales					
Pelagibacterales MED-G39	28	0.53	30.5 (1.3)	65.7 / 36.0 / 43.6	0.2
Pelagibacterales MED-G40	8	0.91	29.6 (1.4)	74 3 / 70 3 / 89 8	0.0
Pelagibacterales MED-G41	18	0.37	28.0 (1.8)	77.1 / 52.3 / 35.7	0.0
Pelagibacterales MED-G42	26	0.83	28 8 (1 7)	31 4 / 52 3 / 77 4	0.0
Pelagibacterales MED-G43	22	0.42	29 4 (1 7)	74.3 / 58.6 / 57.8	0.0
Pelagibacterales MED-G44	28	0.53	29 7 (2 7)	20.0 / 29.7 / 39.6	0.0
	20	0.00	23.1 (2.1)	20.0/20.1/00.0	0.0

Flagellar motor rotation protein MotB Flagellar regulatory protein FleQ Flagellin protein FlaA	15m	Med-OCT2015- 30m	Med-OCT2015- 45m	Med-OCT2015- 60m	Med-OCT2015- 75m	Med-OCT2015- 90m	MedWinter-JAN2015 20m	MedWinte JAN2015-8
Flagellar regulatory protein FleQ Flagellin protein FlaA	0.053	0.036	0.005	0.024	0.012	0.011	0.023	0.023
Flagellin protein FlaA	0.044	0.015	0.005	0.012	0.006	0.008	0.005	0.011
	0.040	0.026	0.022	0.020	0.019	0.014	0.003	0.002
Flagellar hook protein FlgE	0.033	0.020	0.000	0.009	0.004	0.003	0.000	0.007
Flagellar biosynthesis protein FlhA	0.029	0.026	0.011	0.015	0.008	0.006	0.005	0.009
Flagellar basal-body rod protein FlgG	0.028	0.015	0.005	0.011	0.005	0.004	0.002	0.011
Flagellar biosynthesis protein FlhB	0.027	0.028	0.005	0.013	0.007	0.006	0.005	0.011
Flagellar P-ring protein Flgl	0.027	0.011	0.005	0.010	0.007	0.002	0.000	0.005
Flagellum-specific ATP synthase Flil	0.025	0.017	0.011	0.010	0.007	0.006	0.002	0.007
Flagellar motor rotation protein MotA	0.025	0.021	0.022	0.013	0.011	0.011	0.000	0.014
Flagellar motor switch protein FliG	0.025	0.019	0.011	0.010	0.005	0.005	0.002	0.011
Flagellar hook-length control protein FliK	0.023	0.026	0.027	0.038	0.050	0.047	0.023	0.013
Flagellar M-ring protein FliF	0.023	0.014	0.005	0.007	0.004	0.001	0.000	0.005
RNA polymerase sigma factor for flagellar operon	0.023	0.014	0.005	0.012	0.009	0.006	0.003	0.004
Flagellin protein FlaB	0.022	0.013	0.005	0.008	0.005	0.005	0.000	0.002
Flagellar basal-body rod protein FlgC	0.022	0.014	0.005	0.010	0.006	0.004	0.000	0.004
Flagellar biosynthesis protein FliP	0.021	0.016	0.005	0.010	0.007	0.004	0.005	0.007
Flagellar L-ring protein FlgH	0.021	0.009	0.000	0.005	0.002	0.001	0.002	0.005
Flagellar basal-body rod protein FlgB	0.020	0.012	0.000	0.008	0.004	0.004	0.000	0.007
Flagellar biosynthesis protein FliQ	0.019	0.016	0.005	0.008	0.004	0.003	0.003	0.007
Flagellar biosynthesis protein FliR	0.019	0.016	0.005	0.007	0.003	0.004	0.002	0.007
Flagellar synthesis regulator FleN	0.019	0.014	0.005	0.007	0.005	0.003	0.000	0.005
Elagellar basal-body rod modification protein ElgD	0.019	0.014	0.000	0.005	0.002	0.002	0.002	0.005
Elagellar motor switch protein EliM	0.019	0.016	0.005	0.007	0.007	0.005	0.003	0.009
Elagellar motor switch protein EliN	0.018	0.019	0.022	0.011	0.008	0.008	0.002	0.009
Elagellar hook-basal body complex protein EliE	0.018	0.012	0.005	0.009	0.008	0.007	0.000	0.009
Elagellar protein Elg.l	0.017	0.006	0.005	0.009	0.005	0.002	0.000	0.005
MotA/ToIQ/ExbB proton channel family protein	0.014	0.009	0.005	0.006	0.004	0.003	0.005	0.005
Elagellar protein ElgP	0.014	0.006	0.000	0.004	0.000	0.000	0.002	0.000
Elagellar basal-body rod protein ElgE	0.013	0.009	0.000	0.004	0.002	0.002	0.002	0.013
Elagellar biosynthesis protein Flis	0.010	0.007	0.000	0.005	0.002	0.001	0.000	0.004
Elagellar biosynthesis protein Flil	0.008	0.007	0.000	0.003	0.000	0.002	0.002	0.007
	Mad Oct 2045	Mad OCT2045	Mad OCT2045	Med OCTODAS	Mad OCTOM	Mad OCTODAS	Ma diatan LANO045	Ma allafinati
Chemotaxis protein	15m	30m	45m	60m	75m	90m	20m	JAN2015-8
Chemotaxis protein methyltransferase CheR	0.008	0.006	0.000	0.005	0.004	0.003	0.005	0.032
hemotaxis response regulator methylesterase CheB	0.004	0.008	0.005	0.004	0.007	0.008	0.003	0.014
Methyl accepting chemotaxis protein I	0.004	0.003	0.000	0.005	0.007	0.009	0.002	0.088
weary-accepting chemotaxis protein i	0.004	0.004	0.005	0.007	0.004	0.008	0.003	0.022
Chemotaxis regulator CheY	0.004	0.005	0.005	0.005	0.008	0.007	0.002	0.020
Chemotaxis regulator CheY Signal transduction histidine kinase CheA	11.11.14			0.000	0.000	0.000	0.002	0.022
Signal transduction histidine kinase CheA Aerotaxis sesuor receptor protein	0.004	0.001	0.000	0.000	0.000	0.000	0.002	0.022
Chemotaxis regulator CheY Signal transduction histidine kinase CheA Aerotaxis sensor receptor protein Chemotaxis protein CheV	0.001	0.001	0.000	0.001	0.000	0.000	0.002	0.022









environmental microbiology

Environmental Microbiology (2019) 21(6), 1980-1988



Novel *Caudovirales* associated with Marine Group I Thaumarchaeota assembled from metagenomes

Mario López-Pérez ^(b),^{1†*} Jose M. Haro-Moreno,^{1†} José R. de la Torre² and Francisco Rodriguez-Valera¹ ¹Evolutionary Genomics Group,

División de Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan, Alicante, 03550, Spain.

²Department of Biology, San Francisco State University, San Francisco, CA, 94132, USA.

Summary

Marine Group I (MGI) Thaumarchaeota are some of the most abundant microorganisms in the deep ocean and responsible for much of the ammonia oxidation occurring in this environment. In this work, we present 35 sequences assembled from metagenomic samples of the first uncultivated Caudovirales viruses associated with Thaumarchaeota, which we designated marthavirus. Most of the sequences were obtained from cellular metagenomes confirming that they represent an important tool to study environmental viral communities due to cells retrieved while undergoing viral lysis. Metagenomic recruitment showed that this viral population is formed by very divergent entities with high intrapopulation homogeneity. However, metatranscriptomic analyses revealed the same differential expression profile with the capsid as major transcript, indicative of viruses during the lytic cycle. The cobalamine biosynthesis gene cobS, an auxiliary metabolic gene, was also highly expressed during the infection. These analyses expand our understanding of the global diversity of archaeal viruses.

Introduction

Marine Thaumarchaeota, initially discovered 26 years ago by 16S rRNA gene surveys (Fuhrman and McCallum, 1992; Delong, 1992), are some of the most abundant microorganisms in the deep ocean, accounting for up to 40% of the bacterioplankton below the euphotic zone (Karner et al., 2001; Fuhrman and Ouverney, 1998; Church et al., 2003). This abundance and the discovery that members of this lineage derive energy from the oxidation of ammonia (Könneke et al., 2005) and are able to fix inorganic forms of carbon (Berg et al., 2010) argue that the marine Thaumarchaeota are important players in global Carbon (C) and Nitrogen (N) biogeochemical cycles. Recent studies have shown that these marine archaea are responsible for the majority of the aerobic nitrification measured in marine environments and may be a significant source of the greenhouse gas nitrous oxide (Santoro et al., 2011). The cultivation of numerous strains, as well as sequences from environmental metagenomes and single-cell genomes have provided invaluable information on the ecology and evolution of this diverse lineage (Luo et al., 2014; Swan et al., 2014; Santoro et al., 2015). However, despite these efforts, remarkably little is known about their viruses. The vast majority of archaeal viruses that have been isolated so far came from either hyperthermophilic or hyperhalophilic environments, where Crenarchaeota or Euryarchaeota dominate (Snyder et al., 2015). Conversely, in the case of the mesophilic archaea, only the advent of highthroughput sequencing has provided novel information of the unknown viruses infecting archaea (Vik et al., 2017; Roux et al., 2016), expanding viral diversity far beyond that established by traditional methods for virus isolation. The recently discovered Marine Group II Euryarchaeota viruses (magrovirus) group from assembled sequences (Philosof et al., 2017), which infects the ubiquitous and abundant but yet uncultured Marine Group II Euryarchaeota, is an example of the benefits of metagenomics. However, to date no marine thaumarchaeal virus has yet been isolated probably because their host are also difficult to obtain in pure culture. Only two putative viral sequences have been retrieved by single-cell genomics (Labonté et al., 2015) and fosmid libraries (Chow et al., 2015). Additionally, a putative provirus has been found within the genome of Ca. Nitrosomarinus catalina SPOT01 (Ahlgren et al., 2017). Previous studies showed that viruses infecting Thaumarchaeota from the deep ocean were more active than bacterial viruses, contributing to their cell lysis and, hence, modifying the biogeochemical cycles of N and C (Danovaro et al., 2016).

Received 19 September, 2018; revised 23 October, 2018; accepted 24 October, 2018. *For correspondence. E-mail mario.lopezp@umh.es; Tel. +34-965919313; Fax +34-965 919457. [↑]These authors contributed equally to this work.

^{© 2018} Society for Applied Microbiology and John Wiley & Sons Ltd.

In this work, we present 35 sequences assembled from metagenomic samples of the first uncultivated viruses associated with marine Thaumarchaeota. Sequences of this newly identified viral population are locally distributed with low intra-population diversity. Metatranscriptomic analyses showed that they were retrieved while undergoing viral lysis and apart from the capsid, the essential structural component, expression of the cobalamine biosynthesis gene *cob*S, an auxiliary metabolic gene, was also high during the infection. Our analyses provide important insights into the genomic diversity of this new marine viral population, which remain uncultivated, expanding our understanding of the global diversity of archaeal viruses.

Results and discussion

Recently, we characterized variations in the marine microbiome at different depths within the photic zone during a period of strong thermal stratification of the water column (Haro-Moreno et al., 2018). Results showed that marine Thaumarchaeota were only found below the deep chlorophyll maximum (accounted for up to 10% of the community at 90 m), coinciding with the increase of available ammonia that is practically non-existent at shallower depths. Analyses of the assembled contigs from these samples showed the presence of a 69 kb contig that had hits to a few Thaumarchaeota genomes (the majority of these hits were to the genus Ca. Nitrosopumilus) but also to viral-related genes, including predicted major capsid proteins (MCP), portal proteins, tail tape measure proteins and the large subunit of viral terminases. Both the terminase and the MCP proteins gave hits with low identity (32-35%) to a complete, unclassified archaeal virus (KY229235) recovered from a metagenomic assembly of a sample ~550 m below the seafloor (Nigro et al., 2017). Like in the case of KY229235, our contig had identical repeated sequences (>30 nucleotides) at the 5' and 3' terminal regions suggesting a complete viral genome. It has been demonstrated that cellular metagenomes (>0.22 µm size fraction) are a source of bacterial and archaeal viruses that are undergoing the lytic cycle and actively replicating their DNA (López-Pérez et al., 2017).

New Thaumarchaeota viruses recovered from metagenomic samples

In order to expand the repertoire of putative Thaumarchaeota viruses we used the MCP, terminase and portal proteins of this new contig and KY229235 sequences as queries to search against several marine metagenomes and viromes, including the Mediterranean Sea dataset (Haro-Moreno et al., 2018; López-Pérez et al., 2017), *Tara* Oceans (Sunagawa et al., 2015) and Malaspina

Novel Thaumarchaeota viruses 1981

expeditions (Duarte, 2015) and datasets publicly available at the Joint Genome Institute (JGI) database (https:// img.jgi.doe.gov/). In the end, we identified 35 putative viral contigs (Supporting Information Table S1) manually curated to check for similarity to these reference proteins and thaumarchaeal genes (see Material and Methods). Most contigs with similarity to these proteins (#18) were found in our own dataset (Med-OCT2015-75m and Med-OCT2015-90m), 17 in the cellular and one in the viral fraction (MedVir-OCT2015-60m). Interestingly, another large batch (9 genomes) was found in viromes from the Chesapeake Bay, an estuary where strong ammonia gradients are also found (Maresca et al., 2018). All had a GC content varying from 30% to 37%, as expected from the low-GC content of marine Thaumarchaeota genomes (Ahlgren et al., 2017; Supporting Information Table S1). Based on the presence of terminal inverted repeats >30 nucleotides only two viral genomes were complete. A total of 1289 open reading frames were identified in all the sequences. However, only 14% showed significant homology to sequences present in the pVOGs (Prokaryotic Virus Orthologous Groups) database (Grazziotin et al., 2017), as typical for novel viruses. Clustering of the sequences based on similarity resulted in 684 protein clusters, 11 of which formed the viral 'soft' core (they were present in at least half of the sequences) (Supporting Information Table S2). Five of the 'soft' core protein clusters contained proteins involved in DNA metabolism (terminase, RadA, ATPase, PD-D/EXK nuclease and Ribonuclease H) and one sequence was an auxiliary metabolic gene (AMG), cobS, that encodes a protein that catalyses the final step in cobalamin (vitamin B₁₂) biosynthesis in prokaryotes, which has previously only been found in cyanophages (Sullivan et al., 2005). Unfortunately, the remaining five 'soft' core clusters were hypothetical proteins and no function could be inferred. Furthermore, no tRNA-encoding sequences or hallmarks of temperate phage, such as integrase or excisionase genes, were detected in any of the recovered genomes. In addition to the terminase, we found other clusters indicating the Caudovirales affiliation of these viruses such as prohead or portal proteins. This is to our knowledge the first group of head-tail viruses described for the Crenarchaeal superphylum, although they are relatively common in Euryarchaea (Rachel et al., 2002; Pietilä et al., 2014).

Phylogeny and host assignment

We next sought to establish the phylogenetic affiliation of these sequences and their relationship with other archaeal virus sequences. We used two characteristic *Caudovirales* marker genes, the terminase and the MCP. Homologous (although less than 30% nucleotide identity)

1982 M. López-Pérez, J.M. Haro-Moreno, J.R. de la Torre and F. Rodriguez-Valera

terminases and MCPs were found in the fosmid Oxic1 7 (Chow et al., 2015) and the provirus Nvie-Pro1 present in the genome of Nitrososphaera viennensis, a soil Thaumarchaeota (Krupovic et al., 2011), but could not be identified in the putative thaumarchaeal virus found in the single-cell genome AAA160-J20 (Labonté et al., 2015) or in the putative provirus of Ca. Nitrosopumilus catalina SPOT01 (Ahlgren et al., 2017). Both Nvie-Pro1 and Oxic1 7 encoded a multifunctional MCP fused with a protease sequence. This particular fusion between the two domains has not been seen in any of our recovered virus genomes. For the phylogenetic analyses, we selected only the MCP domain of the mentioned reference sequences. Results showed a similar phylogenetic pattern for both terminase and MCPs, where the new sequences identified here formed a separate lineage from haloviruses and magroviruses (Fig. 1A and B). Only the viral genome KY229235 was found close to our sequences, while Oxic1_7 and Nvie-Pro1, which clustered together, were found more closely related to haloviruses. Additionally, we carried out phylogenetic analyses of the viral RadA and PD-D/EXK nuclease genes (both of which are also present in the genomes of Thaumarchaeota cells) and appeared to be strongly associated with the Thaumarchaeota, and distinct from Euryarchaeota group II and their viruses (Fig. 1C and D). Our results confirm the association of these novel viruses to the marine archaeal phylum. Neither Oxic1_7 or Nvie-Pro1 encoded these genes in their sequence. Remarkably, all 35 putative viral sequences clustered as a single, monophyletic lineage in all four phylogenetic trees, indicating that they are a novel clade of marine archaeainfecting Caudovirales, evolutionarily distinct from previous putative Thaumarchaeota viruses. These findings have led us to name this group of new viruses marthavirus (MARine THAumarchaea viruses). The terminase, PD-D/EXK nuclease and the combination of RadA/ ATPase protein sequences of only the marthavirus were aligned, and a phylogenetic tree was constructed for each of them (Supporting Information Fig. S1). However, we could not identify any clustering of the sequences or a distinct pattern linking genomic phylogeny and place of isolation.

In order to gain more insights into the putative host of the new viruses, we first identified all the Thaumarchaea genomes including pure culture, Single-Cell Genomes (SAGs) and Metagenomic Assembled-Genomes (MAGs) available. Only those SAGs and MAGs with an estimated completion \geq 70% and \leq 5% contamination were considered. In total, we analysed 94 genomes that were classified into 12 clusters based on pairwise comparisons of average nucleotide identity (ANI; Supporting Information Fig. S2). Clusters A-G form an independent clade composed of strains from marine origin belonging to the order Nitrosopumilales and unclassified Thaumarchaeota. The other clade contained a mix between genomes recovered from soil metagenomes, mostly members of the genus Nitrososphaera (clusters H-J) and marine samples (K and L) (Supporting Information Fig. S2). Metagenomic recruitment of a representative of each cluster in the same metagenomic samples where marthavirus recruited showed that clusters F and G were the most prevalent (Supporting Information Fig. S3). Cluster G is represented by members of the genus *Ca.* Nitrosopelagicus. Only one representative of this group has been recovered by pure culture and showed a ubiquitous distribution in oligotrophic marine waters (Santoro et al., 2015).

Viral genomic features

Despite the high degree of sequence and gene-content divergence ([ANI 70.8%, coverage 6.16%]; [Average Amino Acid Identity (AAI) 53.5%; percentage of common proteins 40.43%]), the alignment of the two complete genomes showed that synteny was remarkably well preserved, also among the other sequences, with two clearly conserved genomic regions (structural and DNA related) (Fig. 2A), separated by a variable region that in Marthavirus-1 contains the auxiliary metabolic gene cobalamin biosynthesis protein (cobS). This gene catalyses the final step in the cobalamin biosynthesis in Thaumarchaeota, but not in marine Group II/III Euryarchaeota. Cobalamin (vitamin B₁₂) plays an important role in all three domains of life as a cofactor in the synthesis of amino acids (cobalamindependent methionine synthase) or DNA (ribonucleotide reductase-RNR), as well as in other metabolic pathways (Doxey et al., 2015), but only a few taxa are capable to synthetize it (Doxey et al., 2015). A recent study has implicated the Thaumarchaeota as important producers of cobalamin in aquatic environments (Doxey et al., 2015). In fact, some studies have reported a relationship between the availability of vitamin B₁₂ and the distribution and growth of phytoplankton and bacterioplankton blooms (Sañudo-Wilhelmy et al., 2006). Furthermore, this gene has been found in cyanophages, suggesting that it could be potentially associated with RNR during nucleotide metabolism (Helliwell et al., 2016) boosting the replication of viral DNA. Phylogenetic analysis showed that Marthavirus-encoded cobS is not related to archaeal cobS (Supporting Information Fig. S4). Remarkably, viral cobS sequences clustered together and separated from their hosts, suggesting a different evolutionary history.

Distribution and genomic diversity

To assess the abundance, distribution and genomic diversity of the novel group of viruses, we performed fragment recruitment analysis by comparing each sequence to



Fig. 1. Unrooted Maximum Likelihood phylogenetic trees of the (A) terminase large subunit, (B) major capsid, (C) DNA repair RadA and (D) PD-D/ EXK nuclease proteins. Marthavirus gene sequences were compared against the putative thaumarchaeal reference genomes KY229235, Oxic1_7 and Nvie-Pro1 (coloured in red). Additionally, viral sequences of magrovirus, halovirus and cyanophages and archaeal cellular sequences of MG-I Thaumarchaeota and MGII/III Euryarchaeota were also included in the analysis. [Color figure can be viewed at wileyonlinelibrary.com]

314 metagenomes from Mediterranean, *Tara* Oceans and Malaspina datasets (cellular and viral fraction) with a sequence identity threshold of 70%. We considered only those samples where these viral genomes recruited more than 10 Reads per Kilobase of genome and Gigabase of

metagenome (RPKG). As expected, the marthavirus genomes recruited from metagenomes containing Thaumarchaeal genomes, albeit at significantly lower levels and with a more restricted distribution (Fig. 2B). While reference genomes of Thaumarchaeota were detected in 65% of the

1984 M. López-Pérez, J.M. Haro-Moreno, J.R. de la Torre and F. Rodriguez-Valera



Fig. 2. A. Whole-genome translated nucleotide (tBlastX) comparison between the two complete Thaumarchaeota viruses, Marthavirus-1 and Marthavirus-2. Genome size and GC content are indicated between brackets. Hypothetical and annotated proteins are coloured in green and blue respectively.B. Recruitments of the novel Marthaviruses within the different metagenomic and metaviromic datasets of the Mediterranean Sea (MED), *Tara* Oceans (TARA) and Malaspina (MP) expeditions. On the left, pie charts indicating the percentage of datasets where MG-I Thaumarchaeota and Marthaviruses recruited >10RPKG with a threshold of 70% identity. On the right, a heatmap showing the abundance, measured in RPKG of the selected metagenomes and viromes where the genomes recruited >10RPKG with a threshold of 70% identity. C. Metagenome and Metatranscriptome analysis of Marthavirus-1 and -2. In the upper panel, a mapping of the metatranscriptomic raw reads from the sample Med-OCT2015-90m_MT (>99% identity, 100 bp window) is represented. In the lower panel, a recruitment plot using metagenomic raw reads from the sample Med-OCT2015-90m_MT (>99% identity). Genes are coloured according to (A). HP: hypothetical protein, MCP: major capsid protein. [Color figure can be viewed at wileyonlinelibrary.com]

metagenomic samples analysed, marthaviruses were found only in 2% of the metagenomes and 12% of the viromes contrasting with the global abundance of magroviruses in the *Tara* Oceans samples (Philosof et al., 2017), marthaviruses showed a patchy distribution. In fact, the majority of the samples where these viruses recruited came from the Mediterranean Sea and the South Atlantic Ocean (Fig. 2B).

Interestingly, most of the viral genomes recruited reads at more than 99% nucleotide identity, with minimal coverage below 95% identity (Supporting Information Fig. S5). These results suggest that marthaviruses may form a population with low intra-population diversity, but with significant divergence among groups.

Metatranscriptome analysis

From the same seawater sample (Western Mediterranean Sea, 90 m) where we obtained 18 marthavirus genomes (Supporting Information Table S1) we also performed a

metatranscriptome sequencing. These data could provide clues about the prevailing activities during infection. cDNA reads were mapped onto the two complete genomes assembled from this sample (Marthavirus-1 and -2). Most abundant transcripts in both viruses corresponded to the MCPs, which is required for viral assembly (Fig. 2C). In cyanophages, transcription of the structural genes, including MCP, tail and putative tail fiber proteins, is highest during the final phase of infection (Doron et al., 2016). These data confirm the active viral replication in our sample.

Remarkably, we observed that the *cobS*, encoded within Marthavirus-1 genome, was also highly expressed in the metatranscriptome. Although no study of the structure and activity of the CobS-like viral proteins has been done, results of the mRNA transcripts indicate that the presence of this gene may have an important role during the infection process. The acquisition of AMGs has been repeatedly seen in both bacterial and archaeal viral genomes (Rosenwasser et al., 2016), and their presence

modulate host metabolism to favour a more efficient viral replication.

Nine marthavirus genomes were recovered from viromic samples from the Chesapeake estuary. Consequently, we used the metagenomic, viromic and metatranscriptomic datasets collected there (Maresca et al., 2018). Similar results were obtained after analysing the transcripts for the two different genomes (Supporting Information Fig. S6) that recruited the most (Supporting Information Fig. S7). Again, the MCP was the most expressed gene in Marthavirus-4. The *cobS* gene encoded within the Marthavirus-10 genome was expressed as well, although several genes, mostly hypothetical proteins but also an adhesin, which might mediate the virus-host adhesion, and a metallophosphatase were expressed.

In summary, this study characterized several uncultivated viruses assembled from metagenomic samples that infect marine Thaumarchaeota, which we designated marthavirus. It is important to emphasize that several (23 out of 35) of the sequences were obtained from cellular (>0.2 µm) metagenomes reinforcing the idea that they are an important tool to study environmental viral communities containing complementary information which is sometimes missing in viromes. The cellular fraction obviously contains abundant viral material due to the cells retrieved while undergoing viral lysis (López-Pérez et al., 2017). Due to the ecological importance of marine Thaumarchaeota, which are important components in the global nitrogen and carbon nutrient cycling, the study of the thaumarchaea-infecting viruses comprises a key element to understand the dynamics of marine Thaumarchaeota in the ocean

While this article was in revision, another set of viruses linked to Thaumarchaeaota was reported. They were identified as contigs that encode the viral capsid and thaumarchaeal ammonia monooxygenase genes (*amoC*), highlighting the potential impact of these viruses on N cycling in the oceans (Ahlgren et al., 2018). However, those genomes are very different from the ones described here (only one sequence had 3.3 Kb 99% similar to the marthavirus-13). In addition, we have not found in our dataset any gene encoding AmoC that was the search criterium used by these authors (Ahlgren et al., 2018). Together with our results, the discovery of these viruses highlights the likely enormous diversity of Thaumarchaeota viruses present in the ocean.

Experimental procedures

Sample collection and processing

Six metagenomic samples from a depth profile in the Mediterranean Sea were taken on 15 October 2015. Information about the location and sampling procedure can be found in

Novel Thaumarchaeota viruses 1985

Haro-Moreno et al. (2018). Additionally, in the same cruise, a metatranscriptome was made from a sample collected at 90 m deep. For the RNA sample, 200 liter of seawater was collected and immediately filtered in a shaded area onboard through a 0.22 μ m polyethersulfone filter that was suspended with RNAlater and kept on dry ice until storage at -80 °C. RNA extraction was performed according to the phenolic PGTX (Miller et al., 2017). Metatranscriptome was sequenced using Illumina Hiseq-4000 (150 bp, paired-end read) (Macrogen, Republic of Korea).

Genome annotation

The resulting genes on the assembled contigs were predicted using Prodigal (Hyatt et al., 2010). tRNA and rRNA genes were predicted using tRNAscan-SE (Lowe and Eddy, 1996), ssu-align (Nawrocki, 2009) and meta-RNA (Huang et al., 2009). Predicted protein sequences were compared against NCBI NR databases using USEARCH6 (Edgar, 2010) and against COG (Tatusov et al., 2001) and TIGFRAM (Haft et al., 2001) using HMMscan (Eddy, 2011) for taxonomic and functional annotation.

Identification of novel archaeal viruses

MCP, terminase and portal proteins of marthavirus-1 and KY229235 sequences were used as queries to search against several marine metagenomes (Haro-Moreno et al., 2018; López-Pérez et al., 2017; Duarte, 2015; Sunagawa et al., 2015) using DIAMOND (blastp option, top hit, \geq 30% identity, \geq 50% alignment length, *E* value < 10⁻⁵; Buchfink et al., 2015). Only contigs larger than 8Kb were taken into account. These sequences were also filtered using VirFinder (Ren et al., 2017) to confirm the viral origin.

Metagenomic read recruitments

Genomes of known marine Thaumarchaeota (available up to May 2018 in the NCBI database) and the Marthavirus recovered in this work were used to recruit reads from our metagenomic and metaviromic datasets (Haro-Moreno et al., 2018; López-Pérez et al., 2017), together with those retrieved from the *Tara* Oceans (Sunagawa et al., 2015) and Malaspina expeditions (Duarte, 2015) and the Chesapeake estuary (Maresca et al., 2018), using BLASTN (Altschul et al., 1997), with a cut-off of 70% nucleotide identity over a minimum alignment length of 50 nucleotides. Metagenomic samples where archaeal and viral genomes recruited less than 10 reads per kilobase of genome per gigabase of metagenome (RPKG) were discarded.

1986 M. López-Pérez, J.M. Haro-Moreno, J.R. de la Torre and F. Rodriguez-Valera

Phylogenetic trees of hallmark proteins

Manual inspection of the viral genomes was used to retrieve the amino acid sequences of the Terminase, Major Capsid, RadA, PD-D/EXK nuclease and CobS proteins. To infer their taxonomic relationships, sequences coming from marine Thaumarchaeota and Euryarchaeota genomes, as well as from other archaeal viruses (magroviruses and halo-viruses) were used. For the CobS protein, we also included sequences coming from cyanobacterial genomes and phages. Sequences were aligned with MUSCLE (Edgar, 2004) and a Maximum-Likelihood tree was constructed with MEGA 7.0 (Kumar et al., 2016). Jones-Taylor-Thornton model, gamma distribution with five discrete categories, 100 bootstraps, positions with less than 80% site coverage were eliminated.

Thaumarchaeota diversity

Genome completeness and degree of contamination was estimated with CheckM (Parks et al., 2015). The ANI between strains was calculated using JSpecies software package v1.2.1 using default parameters (Richter and Rossello-Mora, 2009).

Acknowledgements

This work was supported by grants 'MEDIMAX' BFPU2013-48007-P, 'VIREVO' CGL2016-76273-P [AEI/FE-DER, EU], (co-founded with FEDER funds); Acciones de dinamización 'REDES DE EXCELENCIA' CONSOLIDER CGL2015-71523-REDC from the Spanish Ministerio de Economía, Industria y Competitividad and PROMETEO II/2014/012 'AQUAMET' from Generalitat Valenciana. JHM was supported with a Ph.D. fellowship from the Spanish Ministerio de Economía y Competitividad (Grant No. BES-2014-067828). MLP was supported with a Postdoctoral fellowship from the Valencian Consellería de Educació, Investigació, Cultura i Esport (Grant No. APOSTD/2016/051).

Author contributions

MLP conceived the project. MLP and JHM performed bioinformatic analyses. MLP, JHM, FRV and JRT wrote the manuscript with contributions from all authors to data analysis, figure generation and the final manuscript.

Data availability

Data (viral sequences and metatranscriptome) presented in this manuscript has been submitted to NCBI and are available under BioProject accession numbers PRJNA352798 and PRJNA484324. The metatranscriptomic sample has been deposited in the SRA database (Med-OCT2015-90m_MT – SRR7633016).

References

- Ahlgren, N. A., Chen, Y., Needham, D. M., Parada, A. E., Sachdeva, R., Trinh, V., et al. (2017) Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ Microbiol* **19**: 2434–2452.
- Ahlgren, N. A., Fuchsman, C. A., Rocap, G., and Fuhrman, J. A. (2018) Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode amoC nitrification genes. *ISME J* 1. https://doi.org/10.1038/s41396-018-0289-4.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
- Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., *et al.* (2010) Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* 8: 447–460.
- Buchfink, B., Xie, C., and Huson, D. H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12: 59–60.
- Chow, C. E. T., Winget, D. M., White, R. A., Hallam, S. J., and Suttle, C. A. (2015) Combining genomic sequencing methods to explore viral diversity and reveal potential virus-host interactions. *Front Microbiol* **6**: 1–15.
- Church, M. J., DeLong, E. F., Ducklow, H. W., Karner, M. B., Preston, C. M., and Karl, D. M. (2003) Abundance and distribution of planktonic Archaea and bacteria in the waters west of the Antarctic peninsula. *Limnol Oceanogr* **48**: 1893–1902.
- Danovaro, R., DellAnno, A., Corinaldesi, C., Rastelli, E., Cavicchioli, R., Krupovic, M., et al. (2016) Virus-mediated archaeal hecatomb in the deep seafloor. Sci Adv 2: 1–10.
- Delong, E. F. (1992) Archaea in coastal marine environments. Proc Natl Acad Sci USA 89: 5685–5689.
- Doron, S., Fedida, A., Hernández-Prieto, M. A., Sabehi, G., Karunker, I., Stazic, D., *et al.* (2016) Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* **10**: 1437–1455.
- Doxey, A. C., Kurtz, D. A., Lynch, M. D. J., Sauder, L. A., and Neufeld, J. D. (2015) Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME J* 9: 461–471.
- Duarte, C. M. (2015) Seafaring in the 21st century: the Malaspina 2010 circumnavigation expedition. *Limnol Oceanogr Bull* 24: 11–14.
- Eddy, S. R. (2011) Accelerated profile HMM searches. *PLoS Comput Biol* **7**: e1002195.
- Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Edgar, R. C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Fuhrman, J. A., and McCallum, K. (1992) Novel major archaebacterial group from marine plankton. *Nature* 356: 148–149.
- Fuhrman, J. A., and Ouverney, C. C. (1998) Marine microbial diversity studied via 16S rRNA sequences: cloning results from coastal waters and counting of native archaea with fluorescent single cell probes. Aquat Ecol 32: 3–15.
- Grazziotin, A. L., Koonin, E. V., and Kristensen, D. M. (2017) Prokaryotic virus orthologous groups (pVOGs): a resource

for comparative genomics and protein family annotation. *Nucleic Acids Res* **45**: D491–D498.

- Haft, D. H., Loftus, B. J., Richardson, D. L., Yang, F., Eisen, J. A., Paulsen, I. T., and White, O. (2001) TIGR-FAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* 29: 41–43.
- Haro-Moreno, J. M., López-Pérez, M., de la Torre, J. R., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2018) Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* **6**: 128.
- Helliwell, K. E., Lawrence, A. D., Holzer, A., Kudahl, U. J., Sasso, S., Kräutler, B., *et al.* (2016) Cyanobacteria and eukaryotic algae use different chemical variants of vitamin B12. *Curr Biol* **26**: 999–1008.
- Huang, Y., Gilna, P., and Li, W. (2009) Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* 25: 1338–1340.
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.
- Karner, M. B., Delong, E. F., and Karl, D. M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**: 507–510.
- Könneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., and Stahl, D. A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543–546.
- Krupovic, M., Spang, A., Gribaldo, S., Forterre, P., and Schleper, C. (2011) A thaumarchaeal provirus testifies for an ancient association of tailed viruses with archaea. *Biochem Soc Trans* **39**: 82–88.
- Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33: msw054.
- Labonté, J. M., Swan, B. K., Poulos, B., Luo, H., Koren, S., Hallam, S. J., *et al.* (2015) Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J* 9: 2386–2399.
- López-Pérez, M., Haro-Moreno, J. M., Gonzalez-Serrano, R., Parras-Moltó, M., and Rodriguez-Valera, F. (2017) Genome diversity of marine phages recovered from Mediterranean metagenomes: size matters. *PLoS Genet* **13**: e1007018.
- Lowe, T. M., and Eddy, S. R. (1996) TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25: 955–964.
- Luo, H., Tolar, B. B., Swan, B. K., Zhang, C. L., Stepanauskas, R., Ann Moran, M., and Hollibaugh, J. T. (2014) Single-cell genomics shedding light on marine Thaumarchaeota diversification. *ISME J* 8: 732–736.
- Maresca, J. A., Miller, K. J., Keffer, J. L., Sabanayagam, C. R., and Campbell, B. J. (2018) Distribution and diversity of rhodopsin-producing microbes in the Chesapeake Bay. *Appl Environ Microbiol* 84: 00137–00118.
- Miller, D. R., Pfreundt, U., Elifantz, H., Hess, W. R., and Berman-Frank, I. (2017) Microbial metatranscriptomes from the thermally stratified gulf of Aqaba/Eilat during summer. *Mar Genomics* **32**: 23–26.
- Nawrocki, E. P. (2009) Structural RNA Homology Search and Alignment Using Covariance Models. PhD Thesis,

Washington University, St. Louis. p. 256. Available at https://openscholarship.wustl.edu/etd/256. doi:https://doi.org/10.7936/K78050MP

- Nigro, O. D., Jungbluth, S. P., Lin, H. T., Hsieh, C. C., Miranda, J. A., Schvarcz, C. R., *et al.* (2017) Viruses in the oceanic basement. *MBio* **8**: e02129-16.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.
- Philosof, A., Yutin, N., Flores-Uribe, J., Sharon, I., Koonin, E. V., and Béjà, O. (2017) Novel abundant oceanic viruses of uncultured Marine Group II Euryarchaeota. *Curr Biol* 27: 1362–1368.
- Pietilä, M. K., Demina, T. A., Atanasova, N. S., Oksanen, H. M., and Bamford, D. H. (2014) Archaeal viruses and bacteriophages: comparisons and contrasts. *Trends Microbiol* **22**: 334–344.
- Rachel, R., Bettstetter, M., Hedlund, B. P., Häring, M., Kessler, A., Stetter, K. O., and Prangishvili, D. (2002) Remarkable morphological diversity of viruses and viruslike particles in hot terrestrial environments. *Arch Virol* 147: 2419–2429.
- Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., and Sun, F. (2017) VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 5: 69.
- Richter, M., and Rossello-Mora, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* **106**: 19126–19131.
- Rosenwasser, S., Ziv, C., van Creveld, S. G., and Vardi, A. (2016) Virocell metabolism: metabolic innovations during host–virus interactions in the ocean. *Trends Microbiol* 24: 821–832.
- Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., *et al.* (2016) Ecogenomics and potential biogeochemical impacts of globally abundant viruses. *Nature* 537: 689–693.
- Santoro, A. E., Buchwald, C., McIlvin, M. R., and Casciotti, K. L. (2011) Isotopic signature of N2O produced by marine ammonia-oxidizing Archaea. *Science* 333: 1282–1285.
- Santoro, A. E., Dupont, C. L., Richter, R. A., Craig, M. T., Carini, P., McIlvin, M. R., et al. (2015) Genomic and proteomic characterization of "*Candidatus* Nitrosopelagicus brevis": an ammonia-oxidizing archaeon from the open ocean. Proc Natl Acad Sci USA 112: 1173–1178.
- Sañudo-Wilhelmy, S. A., Gobler, C. J., Okbamichael, M., and Taylor, G. T. (2006) Regulation of phytoplankton dynamics by vitamin B12. *Geophys Res Lett* **33**: L04604.
- Snyder, J. C., Bolduc, B., and Young, M. J. (2015) 40 years of archaeal virology: expanding viral diversity. *Virology* 479–480: 369–378.
- Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F., and Chisholm, S. W. (2005) Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. *PLoS Biol* **3**: 0790–0806.
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., *et al.* (2015) Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**: 1261359.

^{© 2018} Society for Applied Microbiology and John Wiley & Sons Ltd., Environmental Microbiology, 21, 1980–1988

1988 M. López-Pérez, J.M. Haro-Moreno, J.R. de la Torre and F. Rodriguez-Valera

- Swan, B. K., Chaffin, M. D., Martinez-Garcia, M., Morrison, H. G., Field, E. K., Poulton, N. J., *et al.* (2014) Genomic and metabolic diversity of Marine Group I Thaumarchaeota in the mesopelagic of two subtropical gyres. *PLoS One* **9**: e95380.
- Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., *et al.* (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 29: 22–28.
- Vik, D. R., Roux, S., Brum, J. R., Bolduc, B., Emerson, J. B., Padilla, C. C., *et al.* (2017) Putative archaeal viruses from the mesopelagic ocean. *PeerJ* 5: e3428.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Genomic properties, isolation source and accession number for the 35 marthaviruses.

Fig. S1. Maximum likelihood phylogenetic trees for terminase, RadA+ATPase and PD-D/EXK nuclease protein. ColoredColoured circles indicate origin of the metagenome

Fig. S2. Pairwise comparison among all the available Thaumarchaeota genomes and the ones assembled here using average nucleotide identity (ANI) distances. Fig. S3. Metagenomic recruitment of a representative of each cluster in the same metagenomics samples where mathavirus recruited the most.

Fig. S4. Unrooted Maximum Likelihood phylogenetic tree of the CobS protein.

Fig. S5. Recruitments of selected marthaviruses among several metagenomic and metaviromic datasets. Red line represents the 95% species identity threshold. MED-MG: Mediterranean metagenomes; MED-MV: Mediterranean metavirome; T-MV: TARA metavirome; T-MG: TARA metagenome; MP-MG: Malaspina metagenome; MP-MV: Malaspina metavirome. M#: Marthavirus-#.

Fig. S6. Metagenome and Metatranscriptome analyses of Marthavirus-4 and -10. In the upper panel, a mapping of the metatranscriptomic raw reads from the sample SRR5830089 (>99% identity, 100 bp window) is represented. In the lower panel, a recruitment plot using metagenomic raw reads from the sample SRR5468101 (>70% identity, >50 bp long) is shown. Each dot represents a mapped raw read. Red line indicates the species threshold (95% identity). HP: hypothetical protein.

Fig. S7. A) Heatmap showing the abundance, measured in RPKG, of the novel marthaviruses within the different metagenomic and metaviromic samples collected in the Chesapeake Bay. Only those samples where at least one of the genomes recruited >10 RPKG are considered. Red circles indicate those viral genomes recovered from the Chesapeake Bay. B) Box plot of the selected samples in A).



Novel Caudovirales associated to Marine Group I Thaumarchaeota assembled from metagenomes

Mario López-Pérez^{1,3,*} Jose M. Haro-Moreno^{1,3}, José R. de la Torre², and Francisco Rodriguez-Valera¹.

¹Evolutionary Genomics Group, División de Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan 03550, Alicante, Spain. ²Department of Biology, San Francisco State University, San Francisco, CA 94132, USA. ³These authors contributed equally to this work.

*Corresponding author: <u>mario.lopezp@umh.es</u>

Evolutionary Genomics Group, División de Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan de Alicante, 03550 Alicante, Spain. Phone +34-965919313, Fax +34-965 919457

Running title: Novel Thaumarchaeota viruses

mentary Table S1. Genomic properties, isolation source and accession number for the 35 marthaviruses. Supp

Marine Thaumarchaeota	GC	Size	#ORFs	Isolation source	sampling type	Metagenome Biosample	Original contig name
Marthavirus-1	33.87	69519	94	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-2	34.32	59498	104	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-3	33.19	41184	68	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-4	33.71	38949	72	North Atlantic Ocean	Metavirome	SAMN06265875; SAMN06343780; SAMN06264725; SAMN06266357; SAMN06264725; SAMN06265875	(2) Ge0208161_1000420; Ge0209846_1001419; Ge0109605_1000506; Ge0208795_1000104 Ge0209851_1002456; Ge0208160_1002395; Ge0196901_1000213; Ge020816_1000184; Ge0209851_1000246; Ge020981_0101506; Ge0209848_1001103; Ge020816_11000564; Ge0209848_1002491; Ge020846_1001506; Ge0209848_1001103; Ge0209846_1002356; Ge0209848_1002491; Ge020821_1001396; Ge020985_1003354; Ge020986_1002356
Marthavirus-5	33.72	38734	77	North Pacific Ocean	Metagenome fosmid library	SAMN05422230	(2) JGI24914J38485-100191
Marthavirus-6	34.92	28827	42	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-7	32.07	28697	41	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-8	33.16	27843	63	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-9	37.04	27413	34	Mediterranean Sea	Metagenome	SAMN04325110	(1)
Marthavirus-10	33.54	27114	37	North Atlantic Ocean	Metavirome	SAMN06265905; SAMN06343780; (SAMN06265875; SAMN06264725; C SAMN06266357; SAMN06266062 (2) Gə0129342_1001450; Gə0196905_1000181; Gə0208160_1000372; Gə0208795_1000399; sə2028161_1000461; Gə0196901_1000383; Gə0208019_1000380; Gə0129342_1000513; sə0136656_1001370
Marthavirus-11	32.81	26201	46	Mediterranean Sea	Metavirome	SAMN09755890	(1)
Marthavirus-12	33.07	26021	38	North Atlantic Ocean	Metagenome	SAMEA2623756	(3) TARA_148b_MES_0.22-3_scaffold144681_1
Marthavirus-13	33.97	25293	44	North Atlantic Ocean	Metavirome	SAMN06343813; SAMN06343812; SAMN06264918; SAMN06343810	(2) Ga0208899_1003631; Ga0208899_1007809; Ga0070749_10002414; Ga0070751_100234 Ga0070751_1001009; Ga0208767_1001848
Marthavirus-14	33.55	24765	37	North Atlantic Ocean	Metavirome	SAMN06343780	(2) Ga0196905_1000181
Marthavirus-15	33.62	22109	37	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-16	34.99	21137	33	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-17	33.48	20709	46	Red Sea	Metagenome	SAMEA2619907	(3) TARA_034_DCM_0.22-1.6_scaffold241908_1
Marthavirus-18	33.92	16137	17	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-19	33.69	16134	25	Red Sea	Metagenome	SAMEA2619907	(3) TARA_034_DCM_0.22-1.6_scaffold68959_2
Marthavirus-20	33.88	15273	21	Mediterranean Sea	Metagenome	SAMN05992383	(1)
Marthavirus-21	34.88	15223	29	North Atlantic Ocean	Metavirome	SAMN06343810	(2) Ga0070746_10001264
Marthavirus-22	32.95	14724	38	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-23	30.59	13497	36	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-24	31.72	13132	23	South Pacific Ocean	Metagenome	SAMEA2622837	(3) TARA_125_SRF_0.22-0.45_scaffold436574_1; TARA_125_MIX_0.1-0.22_scaffold87032_
Marthavirus-25	36.55	12141	21	Red Sea	Metagenome	SAMEA2619907	(3) TARA_034_DCM_0.22-1.6_scaffold514352_1
Marthavirus-26	34.03	11785	15	Red Sea	Metavirome	SAMEA2619923	(3) TARA_034_DCM_<-0.22_scaffold10072_1
Marthavirus-27	33.18	11481	21	North Atlantic Ocean	Metavirome	SAMN06343912	(2) Ga0075460_10000942
Marthavirus-28	32.51	11331	23	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-29	32.17	10663	16	Mediterranean Sea	Metagenome	SAMN05992383	(1)
Marthavirus-30	37.61	9845	14	Mediterranean Sea	Metavirome	SAMN06266140	(4)
Marthavirus-31	31.66	9539	19	North Atlantic Ocean	Metavirome	SAMN06343812	(2) Ga0070749 10004237
Marthavirus-32	31.39	9408	17	Mediterranean Sea	Metagenome	SAMN05992384	(1)
Marthavirus-33	33.14	8743	13	Mediterranean Sea	Metagenome	SAMN05992383	(1)
Marthavirus-34	32.99	8517	14	North Atlantic Ocean	Metavirome	SAMN06343813	(2) Ga0070750_10003644
Marthavirus-35	34.6	8022	14	North Atlantic Ocean	Metavirome	SAMN06264587	(2) Ga0070747 1002923



Supplementary Table S2. Protein clusters present in at least half of the Mathavirus genomes

Cluster	Presence in Marthavirus Genomes	putative anotation	VOG cluster	Host Domain	Order
1	21/35	DNA repair and recombination protein RadA	VOG0025	Bacteria and Archaea	Caudovirales
2	19/35	hypothetical protein			
3	18/35	cobalamin biosynthesis protein CobS			
4	17/35	ATPase	VOG9957	Archaea	Caudovirales
5	17/35	terminase large subunit	VOG4544	Bacteria and Archaea	Caudovirales
6	17/35	PD-D/EXK nuclease	VOG3505	Bacteria and Archaea	Caudovirales
7	17/35	hypothetical protein			
8	17/35	ribonuclease H	VOG4714	Bacteria	Caudovirales
9	17/35	hypothetical protein			
10	17/35	hypothetical protein			
11	17/35	hypothetical protein			



Terminase



Supplementary Fig. S1. Maximum likelihood phylogenetic trees for terminase, RadA+ATPase and PD-D/EXK nuclease protein. Colored circles indicate origin of the metagenome



Supplementary Fig. S2. Pairwise comparison among all the available Thaumarchaeota genomes and the ones assembled here using average nucleotide identity (ANI) distances.



Supplementary Fig. S3. Metagenomic recruitment of a representative of each cluster in the same metagenomics samples where mathavirus recruited the most.

UNIVERSITAS Miguel Hermand



Supplementary Fig. S4. Unrooted Maximum Likelihood phylogenetic tree of the CobS protein.



Supplementary Fig. S5. Recruitments of selected marthaviruses among several metagenomic and metaviromic datasets. Red line represents the 95% species identity threshold. MED-MG: Mediterranean metagenomes; MED-MV: Mediterranean metavirome; T-MV: TARA metavirome; T-MG: TARA metagenome; MP-MG: Malaspina metagenome; MP-MV: Malaspina metavirome. M#: Marthavirus-#.



Supplementary Fig. S6. Metagenome and Metatranscriptome analyses of Marthavirus-4 and -10. In the upper panel, a mapping of the metatranscriptomic raw reads from the sample SRR5830089 (>99% identity, 100 bp window) is represented. In the lower panel, a recruitment plot using metagenomic raw reads from the sample SRR5468101 (>70% identity, >50bp long) is shown. Each dot represents a mapped raw read. Red line indicates the species threshold (95% identity). HP: hypothetical protein.





Supplementary Fig. S7. A) Heatmap showing the abundance, measured in RPKG, of the novel marthaviruses within the different metagenomic and metaviromic samples collected in the Chesapeake Bay. Only those samples where at least one of the genomes recruited >10 RPKG are considered. Red circles indicate those viral genomes recovered from the Chesapeake Bay. B) Box plot of the selected samples in A).





10. ANNEX 3



www.nature.com/ismej

ORIGINAL ARTICLE New insights into marine group III Euryarchaeota, from dark to light

Jose M Haro-Moreno^{1,3}, Francisco Rodriguez-Valera¹, Purificación López-García², David Moreira² and Ana-Belen Martin-Cuadrado^{1,3}

¹Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, Alicante, Spain and ²Unité d'Ecologie, Systématique et Evolution, UMR CNRS 8079, Université Paris-Sud, Orsay Cedex, France

Marine Euryarchaeota remain among the least understood major components of marine microbial communities. Marine group II Euryarchaeota (MG-II) are more abundant in surface waters (4-20% of the total prokaryotic community), whereas marine group III Euryarchaeota (MG-III) are generally considered low-abundance members of deep mesopelagic and bathypelagic communities. Using genome assembly from direct metagenome reads and metagenomic fosmid clones, we have identified six novel MG-III genome sequence bins from the photic zone (Epi1-6) and two novel bins from deep-sea samples (Bathy1-2). Genome completeness in those genome bins varies from 44% to 85%. Photic-zone MG-III bins corresponded to novel groups with no similarity, and significantly lower GC content, when compared with previously described deep-MG-III genome bins. As found in many other epipelagic microorganisms, photic-zone MG-III bins contained numerous photolyase and rhodopsin genes, as well as genes for peptide and lipid uptake and degradation, suggesting a photoheterotrophic lifestyle. Phylogenetic analysis of these photolyases and rhodopsins as well as their genomic context suggests that these genes are of bacterial origin, supporting the hypothesis of an MG-III ancestor that lived in the dark ocean. Epipelagic MG-III occur sporadically and in relatively small proportions in marine plankton, representing only up to 0.6% of the total microbial community reads in metagenomes. None of the reconstructed epipelagic MG-III genomes were present in metagenomes from aphotic zone depths or from high latitude regions. Most low-GC bins were highly enriched at the deep chlorophyll maximum zones, with the exception of Epi1, which appeared evenly distributed throughout the photic zone worldwide.

The ISME Journal (2017) 11, 1102–1117; doi:10.1038/ismej.2016.188; published online 13 January 2017

Introduction

Marine archaea are important marine microbes in terms of their metabolic activity and abundance (Karner *et al.*, 2001; Li *et al.*, 2015). Ammoniaoxidizing Thaumarchaeota (Brochier-Armanet *et al.*, 2008) are the most abundant archaeal phylum in the oceans and have a key role in the marine nitrogen cycle (Konneke *et al.*, 2005; Qin *et al.*, 2014). Studies have also identified three major groups of marine Euryarchaeota: (i) group II (MG-II) (DeLong, 1992; Fuhrman *et al.*, 2000), (ii) group III (MG-III) (Fuhrman and Davis, 1997; Lopez-Garcia *et al.*, 2001a), and (iii) group IV (MG-IV) (Lopez-Garcia *et al.*, 2001b). So far,

there are no cultured representatives of marine Euryarchaeota and little is known about their physiology and ecological role in the oceans. MG-II are widely distributed within the euphotic zone of temperate waters. MG-II are the dominant archaeal community not only in the surface and in the deep chlorophyll maximum (DCM) (Massana et al., 2000; Karner et al., 2001; Herndl et al., 2005; DeLong et al., 2006; Galand et al., 2010; Belmar et al., 2011; Martin-Cuadrado et al., 2015) but have also been found in deep-sea waters (Lopez-Garcia et al., 2001a, Martin-Cuadrado et al., 2008; Li et al., 2015). The other two marine Euryarchaeota groups, MG-III and MG-IV, are considered to be rare components of deep-sea communities (Lopez-Garcia et al., 2001a,b; Galand et al., 2009).

MG-III were first described by Fuhrman and Davis, 1997 from deep marine plankton samples and have subsequently been found in 16S-rRNA gene surveys from most deep oceanic regions, albeit at very low abundance (Massana *et al.*, 2000; Lopez-Garcia *et al.*, 2001a,b), and by metagenomics throughout the water column in the central Pacific gyre (DeLong *et al.*, 2006).

Correspondence: A-B Martin-Cuadrado, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, APARTADO 18, CAMPUS SAN JUAN, CP. 03550, San Juan de Alicante, Alicante 03550, Spain. E-mail: amartin@umh.es

³These authors contributed equally to this work.

Received 9 May 2016; revised 25 November 2016; accepted 5 December 2016; published online 13 January 2017

However, occasionally, they have been identified at much higher proportions. For instance, 16S-rRNA sequences from MG-III represented one of the largest archaeal groups in the deep Arctic Ocean (>40% of tag sequences) (Galand et al., 2009) and between 30% and 50% of the archaeal sequences from a deep (500-1250 m) Marmara Sea metagenome (Quaiser et al., 2011). They were also relatively abundant (ca.18% of the total archaeal population) in the oxygen minimum zone (50-400 m) in the Eastern tropical South Pacific (Belmar et al., 2011). Only a few studies report the presence of MG-III in the photic zone. They represented 0.4% of all the archaeal sequences obtained in surface Arctic waters (Galand et al., 2009) and up to 10% in samples recovered along 4.5 years in the Mediterranean DCM (Galand et al., 2010).

The initial analysis of three MG-III fosmids from deep-sea metagenomic libraries allowed a first glance at their metabolic potential (Martin-Cuadrado et al., 2008). The presence of some fermentation-related genes led to the hypothesis that they could be facultative anaerobes. In a recent study, up to 3% of the single amplified genomes of archaea recovered from mesopelagic waters from South-Atlantic and North-Pacific gyres belonged to MG-III (Swan et al., 2014). However, only two single amplified genomes classified as MG-III, SCGC-AAA-007-O11 (isolated at 800 m in the South-Atlantic sub-tropical gyre) and SCGC-AAA-288-E19 (from 770 m in the North-Pacific sub-tropical gyre), have been deposited in GenBank. Only the SCGC-AAA-288-E19 partial genome had ribosomal RNA genes that corresponded to MG-III, but contig annotation showed contamination with Chloroflexi (32 genome fragments out of the 102). Complete archaeal fosmids (452 adding up to 16 Mb of sequence) from deep Mediterranean samples belonging to MG-II/III have been published (Deschamps et al., 2014) and five MG-III partial genomes (31–65% completeness) were assembled from metagenomes from the Guaymas basin (1993 m, Gulf of California) and the Mid-Cayman Rise (2040-2238 m and 4869-4946 m, Caribbean Sea) (Li et al., 2015). Based on the genes present in these genomes, it was proposed that the microbes they represented are motile heterotrophs with different mechanisms for scavenging organic matter.

Binning the assembled fragments by oligonucleotide frequencies, GC content and differential recruitment in metagenomes is a successful strategy for the discovery of novel microbial lineages (Tyson *et al.*, 2004; Ghai *et al.*, 2012; Iverson *et al.*, 2012; Narasingarao *et al.*, 2012; Martin-Cuadrado *et al.*, 2015; Li *et al.*, 2015; Vavourakis *et al.*, 2016). We applied this approach to recover MG-III sequences using several metagenomic fosmid libraries from the Mediterranean Sea (collections KM3, AD1000 (Martin-Cuadrado *et al.*, 2008) and MedDCM-OCT2007 (Ghai *et al.*, 2010)) and from the assemblies of 16 metagenomes (four collections from the Mediterranean: MedDCM-JUL2012 (Martin-Cuadrado *et al.*, 2015), MedDCM-SEP2014 (this work),

Novel epipelagic marine Euryarchaeota group III JM Haro-Moreno et al

Med-Io7-77mDCM and Med-Ae2-600mDeep (Mizuno et al., 2016) and 12 from TARA microbiomes (Sunagawa et al., 2015)). We obtained a total of eight different MG-III genome bins. Six of them belong to novel surface MG-III lineages distantly related to the previously described deep MG-III sequence bins (Li et al., 2015). They are the first near-complete genomes of MG-III living in the photic zone. Some of them appear to be widespread in the ocean; their distribution in different water masses has been analyzed.

Materials and methods

Sampling and sequencing

A fosmid metagenomic library of ca. 13 000 clones was constructed with biomass recovered in October 2007 (50 m deep) at the Mediterranean DCM (38°4′6.64″N 0° 13'55.18"W). Partial results of almost 7000 fosmid sequences have been described previously in Ghai et al. (2010) and Martin-Cuadrado et al. (2015). Metagenomes were also sequenced from samples recovered at the same location and at a similar depth the following years (MedDCM-JUL2012 (Martin-Cuadrado et al., 2015) and MedDCM-SEP2014) from one sample recovered at the DCM from the Ionian Sea (Med-Io7-77mDCM) and from a sample collected from the deep Aegean Sea (Med-Ae2-600mDeep) (Mizuno et al., 2016). For these metagenomes, sea water was collected and sequentially filtered on-board using a positive pressure system through a 20 µm pore filter followed by a 5 µm pore size polycarbonate filter and, finally, 0.22 µm pore size Sterivex filters (Durapore; Millipore, Billerica, MA, USA). Filters were frozen on dry ice and kept at - 80 °C until processed in the laboratory. Filters were thawed on ice and then treated with 1 mg ml^{-1} lysozyme and 0.2 mg ml^{-1} proteinase K (final concentrations). Nucleic acids were extracted with phenolchloroform-isoamyl alcohol and chloroform-isoamyl alcohol. Sequencing was carried out using Illumina HiSeq2000 (PE, 100 bp) (Macrogen, Seoul, Korea and BGI, Hong Kong).

'De novo' assembly, gene annotation and binning of the MG-III sequences

A schematic of the assembly pipeline is shown in Supplementary Figure S1. The assembly of the fosmids from the MedDCM-OCT2007, KM3 and AD1000 metagenomic fosmid libraries has been previously described (Ghai et al., 2010; Deschamps et al., 2014; Martin-Cuadrado et al., 2015). Sequences from metagenomes MedDCM-JUL2012, MedDCM-SEP2014, Med-Io7–77mDCM and Med-Ae2–600mDeep were quality trimmed and assembled independently using IBDA-UD (Peng et al., 2012) with the following parameters: mink 70, -maxk 100, -step 10, -pre_correction. Gene predictions on the assembled sequences were carried out using Prodigal (Hyatt et al., 2010). Ribosomal genes were identified using ssu-align (Nawrocki, 2009) and meta_rna (Huang et al., 2009). Functional annotation was performed by comparing predicted protein
1104			
1104	CheckM ^c % contamination	2 22 2 2 5 2 2 5 2 2	
	Median gene size (bp)	752 792 765 765 767 767 767 767 704 704 704 705 705 705 705 705 701 702 701 701 701 701 701 701 701 701 701 701	
	Intergenic region (bp)	44444484888844444444444444444444444444	
	No. of CDS/ nr CDS ^b	$\begin{array}{c} 3058/1307\\ -//1106\\ 559/-\\ 555/-\\ 282/-\\ -//1182\\ 675/677\\ -/741\\ 675/677\\ -/741\\ 675/677\\ -/741\\ 675/677\\ -/741\\ 612/610\\ -/642\\ 1143/1007\\ -/542\\ 1143/1007\\ -/764\\ 1143/1007\\ -/704\\ 1023/654\\ -/704\\ \end{array}$	
mposite genomes CG-MGIII	Largest contig (Kb)	$\begin{array}{c} 120.8\\ 135.9\\ 797.4\\ 797.4\\ 701.9\\ 701.9\\ 71.7\\$	Bil
	No. of genomes	8	IVERS
	% Genome (1) (2) (3) ^a	$\begin{array}{c} 80.0/35.1/83.0\\ 85.7/34.2/84.9\\ 55.3/16.2/45.3\\ 57.1/18.0/54.7\\ 5.7/4.5/5.7\\ 8.0/39.6/75.5\\ 48.6/19.8/675.5\\ 48.6/19.8/475.3\\ 43.3/16.2/4.3/42.3\\ 34.3/16.2/4.3/42.3\\ 34.3/16.2/4.3/42.3\\ 52.9/3.6/7.6\\ 53.39.6\\ 57.1/21.6/60.4\\ 60.0/26.1/60.4\\ 68.6/18.9/58.5\\ 68.6/18.9/58.5\\ 68.6/18.9/58.5\\ 68.6/18.9/58.5\\ 68.6/18.9/58.5\\ 68.6/18.9/58.5\\ 68.6/18.9/58.5\\ 68.6/18.9/58.5\\ 68.6/21.6/58.5\\ 68.6/2$	redundant CDS. singarao <i>et al.</i> (2012). and 70% coverage.
bins and the c	No. of Mb	2.95 1.18 0.554 0.554 0.79 0.79 0.79 0.79 0.74 0.54 0.77 0.54 0.54 0.54 0.77 0.54 0.54 0.54 0.77 0.54 0.54 0.54 0.77 0.77 0.54 0.77 0.72 0.72 0.72 0.72 0.72 0.72 0.72	se; nr-CDS, non- (2013); (3) Naraa 80% similarity
s of the MG-III	$\% GC \pm s.d.$	$\begin{array}{c} 36.6\pm0.9\\ 36.6\pm0.8\\ 36.6\pm0.8\\ 36.5\pm0.9\\ 36.2\pm0.9\\ 36.1\pm0.9\\ 36.1\pm1.1\\ 35.9\pm1.2\\ 35.9\pm1.2\\ 36.5\pm0.8\\ 36.5\pm0.8\\ 36.5\pm0.8\\ 36.5\pm1.1\\ 36.5\pm1.2\\ 36.5\pm1.2\\$	ng DNA sequenc Albertsen <i>et al.</i> DS clustered at
neral feature	No. of contigs	1872222284 82222228233333328 187222228 187222228 187222228 187222 18722 19722	s: CDS, Codir <i>I.</i> (2007); (2) _ -redundant C 2014.
Table 1 Gen	MG-III bin	Epi1 CG-Epi1 Epi2A Epi2A Epi2B Epi2B Epi2B Epi2 Epi3 CG-Epi3 CG-Epi3 Epi6 Epi6 Epi6 CG-Epi5 Epi6 CG-Epi5 Epi6 CG-Epi5 Epi6 CG-Epi5 Bathy1 CG-Bathy2 CG-Bathy2	Abbreviation ^a (1) Raes <i>et a</i> ^b nr-CDS: non ^c Parks <i>et al.</i> ,

Novel epipelagic marine Euryarchaeota group III JM Haro-Moreno et al

> sequences against the NCBI-nr database, Pfam (Bateman *et al.*, 2004), arCOGS (Makarova *et al.*, 2015) and TIGRfams (Haft et al., 2001) (cutoff E-value 10⁻⁵). Based on sequence similarity against the nonredundant NCBI database, the best hit for each gene was determined and used to bin to top-level taxa. Bona fide Euryarchaeota genome fragments were defined as having >50% of the predicted open reading frames with best hits to other Euryarchaeota genes. The resulting sequences were used to screen for their presence in several metagenomes (in subsets of 20 million reads, where applicable): the TARA data sets (Sunagawa et al., 2015), the GOS collection (Rusch et al., 2007), the depth profiles collections from the subtropical gyres of North Atlantic (Bermuda Atlantic Time Series, BATS) and North Pacific (Hawaii Ocean Time-Series, HOT) (DeLong, 2006; Coleman and Chisholm, 2010), several Mediterranean Sea metagenomes at different depths (Ghai et al., 2010; Quaiser et al., 2011; Smedile et al., 2012; Martin-Cuadrado et al., 2015), and a number of deep ocean and cold waters metagenomes (Alonso-Saez et al., 2012; Larsson et al., 2014). The collections coming from the surroundings of hydrothermal vents published in Li et al. (2015) were also included. The screening was performed using Usearch6 (Edgar, 2010), with a cutoff of 95% identity over an alignment length of at least 50 bp (approximately species-level divergence, Konstantinidis and Tiedje, 2005). To compare the results among different data sets, the number of reads was normalized to the metagenome size and the sequence length. The final coverage results were expressed as the number of reads per kilobase of the fragment per gigabase of metagenome collection (rpkg). Only metagenomes in which any of the MG-III sequences recruited reads at over 3 rpkg, a total of 33 metagenomes, were used for genome assembly (Supplementary Table S1).

> All the sequences obtained from these assemblies were binned together in order to cluster them by their tetranucleotide frequencies, GC content and coverage values (Supplementary Figure S2 and Supplementary Table S1). Tetranucleotide frequencies were computed using the 'wordfreq' program from the EMBOSS package (Rice *et al.*, 2000) and the coverage values were calculated as rpkg as described before. Only those clusters with >10 sequences and containing at least one gene marker with a clear affiliation to MG-III were retained. The phylogenetic assignment to MG-III was determined by the presence of at least one housekeeping gene in the same bin (see below). Following this method, a total of 375 genomic fragments > 10 Kb could be classified into 10 different MG-III bins of sequences, Epi1, Epi2A, Epi2B, Epi2C, Epi3, Epi4, Epi5, Epi6, Bathy1 and Bathy2. We also considered 16 MG-III sequences that contained a ribosomal or a housekeeping gene but that could not be included in any of the bins by the criteria used (Supplementary Table S2).

> In order to improve the completeness and remove the redundancy present in the initial MG-III bins,

The ISME Journal

Table 2 Environmental collections from where MG-III sequences were assembled Bathy2 Depth (m) Fraction size (µm) Epi1 Epi5 Epi6 Epi2A Epi2B Epi2C Bathv1 Epi3 Epi4 Total, Kb 2950.4 707.0 631.3 259.7 848.7 542.7 564.7 305.0 1196.5 1061.4 ERR598993 (TARA_18)^a 0.22 - 1.6658.3 5 ERR599073 (TARA_18)^a 60 0.22 - 1.654.6ERR315859 (TARA 023) 55 0.22 - 01.611.7 ERR594297 (TARA 068)^a 5 0.45 - 0.825.3ERR594294 (TARA 068)* 50 0.22 - 0.45367.2 47.4 ERR594348 (TARA_068)* 0.45 - 0.850 159.3 0.45-0.8 ERR594335 (TARA_070)^a 5 41.9 ERR598942 (TARA_133)^a 707.0 45 0.22 - 360.9 ERR598983 (TARA_145)^a 0.22 - 3198.8 422.4 305.0 5 ERR598996 (TARA_150)^a 40 0.22 - 3128.0 ERR598976 (TARA_151)^a 5 0.22 - 3264.7ERR598986 (TARA 151)* 80 0.22 - 3216.5MedDCM-OCT2007 0.22 - 534.5 733.8 60 1034.7 MedDCM-JUL2012 0.22-5 75 542.7 142.3 MedDCM-SEP2014^d 60 0.22 - 5596.7 0.22-5 AD1000^e 1000 38.7 Med-Ae2-600mDeepf 600 0.22 - 51017.6 Med-Io7-77mDCMf 0.22 - 555.8 77 KM3^e 3000 0.22 - 5140.1 1059.6

"Sunaguawa et al. (2015). "Ghai et al. (2010). "Martin-Cuadrado et al. (2015). "This work. "Martin-Cuadrado et al. (2008). "Mizuno et al. (2016).

a second assembly was performed combining the sequences >10 Kb with the short paired-end Illumina reads of the metagenomes from where they were assembled (Tables 1 and 2 and Supplementary Figure S3). For each of the MG-III sequence bins, we used the BWA aligner (Li and Durbin, 2009; default parameters) to recover the short pair-reads that mapped onto the >10 Kb contigs. For each bin, these reads were then pooled and assembled together with the large DNA contigs previously assembled using SPAdes (Bankevich et al., 2012). The final assemblies were termed 'composite genomes' (CGs), as they belong to similar MG-III cellular lineages (defined by the MG-III bins) but from different samples (Supplementary Table S3). The completeness of the reconstructed archaeal genomes was estimated by three different criteria and based on the presence of essential/core genes using HMMER (35, 112 and 53 genes (Raes *et al.*, 2007; Narasingarao *et al.*, 2012; Albertsen *et al.*, 2013)). An *E*-value $<10^{-5}$ and an alignment coverage >65% were used as cutoffs to define homologs of the essential/core genes. Analysis of the contamination within the CGs was performed using CheckM (Parks et al., 2014) (Table 1). Average nucleotide identity (ANI) and conserved DNA fraction between reconstructed and/ or reference genomes were calculated based on the whole-genome sequence as in Goris et al. (2007) (Supplementary Figure S4). GC content was calculated using the 'geecee' tool from the emboss package (Rice et al., 2000).

Phylogenetic analysis

16S-rRNA and 23S-rRNA gene sequences detected in the MG-III genomic fragments were used to retrieve rRNA gene sequences from the most closely related euryarchaeal genomes and selected genome

fragments in GenBank using BLAST (Altschul et al., 1990). 16S-rRNA sequences from metagenome collections were screened and trimmed using ssualign (Nawrocki, 2009). Archaeal 16S-rRNA and 23SrRNA gene sequences were then aligned using MUSCLE (Edgar, 2004). Phylogenetic reconstructions were conducted by maximum likelihood using MEGA6-v.0.6 (Tamura-Nei model, 100 bootstraps, gamma distribution with (five discrete categories), all positions with <80% site coverage were eliminated) Tamura et al., 2013) (Supplementary Figure S5). For the protein trees of RecA, RpoB, SecY, geranylgeranylglyceryl phosphate synthase, DnaK, GyrA, GyrB, photolyase and rhodopsin (Supplementary Figures S6–S14), sequences were selected based on existing literature. Sequences were aligned using MUSCLE (Edgar, 2004) and a maximum likelihood tree was constructed using MEGA6-v.0.6 (Jones-Taylor-Thornton model, 100 bootstraps, gamma distribution with five discrete categories, positions with < 80%site coverage were eliminated). Taxonomic affiliation of the selected bins was also determined by a phylogenomic tree based on concatenates of several ribosomal proteins (L13, S9, L5, S8, L6, S5, S12, S7, L11, L3, L4, L2, L22, S3, L14, S17, L15 and L18). A balanced taxonomic representation of other archaeal genomes was included as reference. Shared proteins were concatenated and aligned using Kalign (Lassmann and Sonnhammer, 2005) and a maximum likelihood tree was made using MEGA6-v.0.6.

Genome comparisons

Synteny among the CG-MGIII was examined with CIRCOS (Krzywinski *et al.*, 2009) and defined as arrays of contiguous genes in tracts of DNA >5 Kb and having >70% of identity. For each of the MG-III bins, non-redundant protein databases were

Novel epipelagic marine Euryarchaeota group III

JM Haro-Moreno et al

Novel epipelagic marine Euryarchaeota group III JM Haro-Moreno et al

1106

constructed clustering the coding DNA sequences with UCLUST (Edgar, 2010) (cutoff: 80% similarity in 70% of their length). These subsets of proteins were compared among themselves using a reciprocal best-hit analysis of putative homologs by BLASTP. Reciprocal relations were plotted using CYTOSCAPE (Shannon *et al.*, 2003). In order to identify the unique proteins of each of the bins, UCLUST was used with a cutoff of 30% similarity along 70% of their length.

Accession numbers

Mediterranean metagenomes used for recruitment are available at NCBI-BioProjects: PRJNA257723 (MedDCM-SEP2014, MedDCM-JUL2012 and MedDCM-OCT 2007), PRJNA305355 (Med-Io7–77mDCM, Med-Io16– 70mDCM, Med-Io17–3500mDeep, Med-Ae1–75mDCM and Med-Ae2–600mDeep). Sequences >10 Kb and the reconstructed CGs genomes have been deposited in Bio-Project number: PRJNA335308. TARA metagenomes were downloaded from the European-Bioinformatics-Institute (http://www.ebi.ac.uk/services/tara-oceansdata).

Results and Discussion

General features of MG-III archaeal genomes

Following assembly and binning, we obtained 375 genomic fragments that clustered into 8 MG-III bins (Supplementary Figure S1). Six bins, Epi1–Epi6, were from epipelagic origin (photic zone) and contained a total of 386 genomic fragments with a total of 8.3 Mb. Two bins, Bathy1 and Bathy2, were from deep marine samples (aphotic zone) and contained 76 fragments for a total of 2.3 Mb. Manual inspection of the differential coverage of the sequences in each bin identified three subsets of Epi2, referred to as Epi2A, Epi2B and Epi2C. Further genomic comparisons indicated that these bins were very similar to each other (93–96% ANI, Supplementary Figure S4) and represent genomes from related species, likely within the same genus.

Remarkably, seven genome bins were formed by sequences primarily from a single sampling site (Table 2). The exception was Epi1, which includes sequences retrieved from nine different sites in the Mediterranean Sea, Atlantic and North-Pacific oceans. These findings suggest that the organisms represented by Epi1 are cosmopolitan in temperate epipelagic waters, whereas the other groups are only abundant enough to assemble from metagenomes at specific sites (endemic) or under transient environmental conditions causing significant growth (for example, blooms; see below).

To improve the analysis of each genome bin, a second assembly was performed and CGs were reconstructed using sequences from different samples and geographic origins (Supplementary Figure S1). These CGs are non-redundant and consist of genomic fragments from similar lineages of MG-III cells but not necessarily from the same sample. In this further assembly, subsets Epi2A, 2B and 2C were condensed into a single bin, CG-Epi2. Genomic features of the genome bins can be found in Tables 1 and 2 and the complete list of the MG-III contigs and the CGs are given in Supplementary Tables S2 and S3. Using the criteria of Narasingarao *et al.* (2012), the genome bins with highest degree of completeness were CG-Epi1 (85%), followed by CG-Epi2 (75%) and the mesopelagic CG-Bathy1 (64%). Based on the number of different variants of single copy genes in each bin, all our CGs contained a single microbial species each (Supplementary Table S4).

All MG-III bins had low GC content (36–36.8%) with the exception of Bathy2 (64.2%). Previously described MG-III sequences from different bathypelagic samples were all high GC (62.8%–65.4%) except for Guaymas32 (36.8%) (Li et al., 2015). It has been noted that GC content tends to increase with depth (Romero et al., 2009; Mizuno et al., 2016). Selection for less nitrogen demand has been proposed as the main drive toward low genomic GC content in free-living marine bacterioplankton. In epipelagic waters, nitrogen is more likely to be the limiting nutrient, in contrast to the dark, energy-limited but relatively nitrogen-rich, deep ocean (Dufresne et al., 2005; Swan et al., 2013; Batut et al., 2014; Giovannoni and Nemergut, 2014). Nevertheless, Bathy1 and Guaymas32 have similar low GC content to surface MG-III bins, suggesting that other factors might be also important.

In general, epipelagic MG-III bins were more genetically heterogeneous. Among the low GC-MGIII bins, the ANI varied from 68% to 85.4%, whereas the high GC-MGIII bins (Bathy2 is 90.8% similar to Cayman92) showed higher degrees of conservation, with ANIs ranging 89.5% to 96.2% (Supplementary Figure S4). This apparently higher diversity of the epipelagic groups may reflect the chemical and physical heterogeneity of surface water layers, which are submitted to stronger hydrodynamic, seasonal and geographical variations (Bryant *et al.*, 2015). In contrast, MG-III representatives from the deep ocean inhabit a more stable environment and might consequently be less diverse, with more homogenous genomes.

Phylogenetic affiliation of the genomic bins

Genes coding for rRNA are difficult to bin because (i) rRNA genes assemble poorly due to their conservation and duplication in genomes and (ii) they recruit metagenomic reads at much higher levels making coverage-based approaches impractical. Most of the rRNA sequences came from fosmid-libraries (Km3 and AD1000) and did not cluster within any of the bins described here. The only assigned 16S-rRNA sequence (372 bp) belonged to Bathy1 and it appears distantly related to the previously described OTU-D (Galand *et al.*, 2009) and DH148-W24 clusters (Lopez-Garcia *et al.*, 2001a,b) (Supplementary Figure S5a). A similar result was obtained with the 23S-rRNA gene identified in Bathy1 (Supplementary Figure S5b). Therefore, we

looked for other housekeeping genes that might be helpful to define the phylogenetic relationships of the novel MG-III with other archaea. We identified and constructed phylogenetic trees for RecA, RpoB, SecY, the geranylgeranylglyceryl phosphate synthase, DnaK and the two gyrase subunits, GyrA and GyrB (Supplementary Figures S6-S12). Although DnaK, GyrA and GyrB have a complex history of horizontal gene transfer (HGT) (Gribaldo et al., 1999; Petitjean et al., 2012; Raymann et al., 2014), their phylogenetic analysis clearly showed the split between MG-II and MG-III sequences. The MG-III housekeeping genes retrieved from epipelagic waters clustered into two groups, one represented only by Epi2 and the other including Epi1, 3, 4, 5 and 6. Bathy2 appeared as a separate cluster from the epipelagic MG-III, and Bathy1 sequences appeared as the most divergent and basal branch. The phylogenomic analysis of the concatenated ribosomal proteins revealed a similar topology (Figure 1). The two epipelagic clusters shared similar GC content. Accordingly, they were named LowGC-MGIII (comprising two subclades: LowGC1-MGIII

Novel epipelagic marine Euryarchaeota group III JM Haro-Moreno et al

(Epi1, 3, 4 and 6) and LowGC2-MGIII (Epi2 and Guaymas32)), and a separate clade, containing bins exclusively of bathypelagic origin (Bathy2, Cayman92 and Guaymas31), was named HighGC-MGIII. Bin Epi5 lacks the ribosomal operon, but it was included into the LowGC1-MGIII based on the phylogenetic analysis of the other housekeeping genes (Supplementary Figures S11 and S12). Bathy1 consistently appeared as a separate basal branch, which might reflect the intermediate depth (600 m), location (Aegean Sea) and physicochemical conditions (highly saline, relatively warm and extremely oligotrophic) of the samples contributing sequences to this genomic bin. The position of Guaymas32 (retrieved from 1993 m), which clusters with Epi2 (5–75 m), might be explained by the presence of two different microbial species in the Guaymas32 bin (Li et al., 2015). One appears to be most similar to the surface Epi2 sequences (80.8% ANI), while the other is closer to the deeper Bathv1 sequences (72.9% ANI) (also observed in the synteny plot of Figure 2a) (see below). Another plausible explanation is that Guaymas32 might be a surface



Figure 1 Maximum likelihood tree based on 18 ribosomal proteins concatenated present in draft MG-III archaeal genomes reconstructed from epipelagic and deep-sea metagenomes. Archaeal genomes from major orders of Euryarchaeota were included as references (accession number in brackets). Novel sequences from this work are shown in bold. Average GC content is shown on the right and colored depending on whether it is high or low GC. Only bootstrap values over >50% are shown.

The ISME Journal

Novel epipelagic marine Euryarchaeota group III



Figure 2 (a) Overview of genomic conserved synteny among the CG-MGIII genomes. Alignments > 5 Kb over 70% identity are shown. A color code is used for each MG-III bin. (b) Amino-acid comparison among the MG-III bins. Sets of non-redundant proteins (cutoff of 80% similarity over 70% of their length) were compared through reciprocal BLASTP and the average amino-acid similarity was plotted. Each circle represents a genome bin. Circles are interconnected as a function of the percentage of shared proteins and colored in accordance with their similarity. Size of the bins and width of the lines are explained in the legend. Proteins of the MG-II MG2-GG3 (Iverson *et al.*, 2012), Thalassoarchaea (Martin-Cuadrado *et al.*, 2015) and the deep-sea hydrothermal vent Euryarchaeota (DHVE2) Aciduliprofundum boneii T469 were included in the analysis.

organism dragged to the bottom by the continuous flux of surface microbes and particles into the deep. Indeed, Guaymas sediments are surprisingly enriched in surface planktonic microbes (Edgcomb *et al.*, 2002) when compared with other deep-sea sediments (Lopez-Garcia *et al.*, 2003). However, the lack of rhodopsins and photolyases (discussed below), together with higher recruitments from deep data sets, would suggest that Guaymas32 is a *bona fide* deep inhabitant.

Synteny and gene content

To examine the conservation of synteny across the different genome bins, we performed an all-versus-all genome comparisons with the available sequences of MG-III (Figure 2a). Within the two groups of LowGC-MGIII bins, large fragments have the same genomic context while synteny blocks are not conserved between LowGC-MGIII and HigGC-MGIII. In the case of LowGC-MGIII, the highest synteny was found between Epi1 and Epi4 (54 block alignments, 62% of Epi4 genome size). For LowGC-MGIII, only Epi2 and

Guaymas32 showed a significant synteny (56 block alignments, 38% of CG-Epi2). The low level of synteny between Bathy1 and other bins confirms that the microbes represented by this bin are very distant to the other LowGC-MGIII. Among the HighGC-MGIII bins, the highest synteny was found between Bathy2 and Guaymas31 (40 block alignments, 42% of CG-Bathy2) followed closely by Cayman92 and Guaymas31 (42 block alignments, 40.8% of the Cayman92 genome).

Non-redundant sets of proteins were obtained for each of the bins, including MG-II relatives, and compared between bins, retaining only the best hit for each protein and using a threshhold of 80% similarity. The relationships between bins were then plotted in the similarity network showed in Figure 2b. This protein content analysis supported the clustering observed in the phylogenomic tree (Figure 1). Bathy1 and SCGC-AAA-288-E19 appeared distantly associated with Guaymas32 and Guaymas31, respectively. MG-III bins Epi1 with Epi4 had the largest percentage of shared proteins (34.8%), followed by Epi2B and Guaymas32 (24%) and then Bathy2 and Guaymas31 (25%). Only 8% of Epi1 proteins were conserved in Epi2 and 0.5% in Bathy2. Although these numbers may be biased owing to the incomplete nature of the bins, they suggest that marine Euryarchaeota are very diverse and contain very different gene pools. Similar results were obtained by Deschamps *et al.* (2014) who found that the core genome of the MGII/III Euryarchaeota was only 15.6% of their pangenome, while their flexible genome was almost triple that of the Thaumarchaeota.

Metabolic functional inference

Several studies have suggested that marine Eurvarchaeota have a significant role in the degradation of dissolved organic matter in marine waters, for example, dissolved amino acids (Ouverney and Fuhrman, 2000) or carbohydrates (Boutrif et al., 2011). The presence of large peptidases related to protein degradation, together with enzymes for the use of fatty acids in the MG2-GG3 genome suggested that particles might be a habitat for MG-II Euryarchaeota (Iverson et al., 2012; Orsi et al., 2015). MG-II shared various features with the deep MG-III described by Li et al. (2015), suggesting that they might be aerobic heterotrophs that use proteins and polysaccharides as major energy source. In order to infer different lifestyles, the predicted open reading frames were functionally classified according the arCOG categories and their frequencies in the different genomes compared (Supplementary Tables S5 and S6 and Supplementary Figure S15).

Central carbon metabolism

MG-III genomes harbored enzymes for glycolysis, the tricarboxylic acid cycle and oxidative phosphorylation, indicating aerobic respiration (Supplementary Table S7). However, owing to the incomplete nature of these genomes, not all genes could be found, and some predictions need to be taken cautiously, especially for Bathy2. We found genes for the complete tricarboxylic acid cycle in LowGC-MGIII but three genes were absent in Bathy1. Remarkably, only the aconitase and the fumarase were found in Bathy2. As was observed in some MG-II (Martin-Cuadrado et al., 2015), MG-III appears to possess most of the enzymes of the Embden-Meyerhof-Parnas (EMP) pathway for metabolism of hexose sugars, with the exception of the first and the last enzymes of the pathway. We were unable to find any other enzyme that could serve as an alternative for the missing glucokinase. For the final step of the EMP, we propose that phosoenolpyruvate synthase, found in all of our MG-III bins, might be able to function bi-directionally and substitute for the missing pyruvate kinase, allowing the EMP to function in both directions, gluconeogenic and glycolytic. Likewise, we found typical gluconeogenesis enzymes such as phosphoenolpyruvate carboxykinases in the LowGC-MGIII and Bathy1 bins, as well as subunits of

Novel epipelagic marine Euryarchaeota group III JM Haro-Moreno et al

the pyruvate/oxaloacetate carboxyltransferase in all the MG-III bins. We were unable to find glucose 1-dehydrogenase, gluconolactonase and 2-keto-3deoxy gluconate aldolase homologs, suggesting that the Entner–Duodoroff hexose catabolic pathway is not present in the MG-III, unlike findings in other Euryarchaea (Makarova *et al.*, 1999; Makarova and Koonin, 2003; Hallam *et al.*, 2006).

Only a small number of amino-acid synthases were found in MG-III: cysteine in Bathy1 and Bathy2, glutamine in LowGC-MGIII, and for glutamate in all MG-III bins. Remarkably, many enzymes for *de novo* biosynthesis were missing, including those for synthesizing methionine, arginine, threonine, histidine, aromatic amino acids and branched amino acids (Supplementary Table S7). However, we observed multiple genes related with the uptake and transformation of peptides or amino acids in our MG-III bins, indicating that these organisms are capable of taking up amino acids from the environment and incorporating them into their proteins. For example, we found genes for permeases for lysine/ arginine (all bins), histidine (Bathy2), glutamine (LowGC-MGIII and Bathy1), proline (LowGC-MGIII and Bathy1) and polar amino acids (Bathy2). Also, several ABC-transporter-systems were found for peptides and oligopeptides; for example, Dpp-ABCtype dipeptide/oligopeptide transporters (in all) and Liv-ABC-type branch amino-acid transporters (In any (LowGC-MGIII and Bathy1). Several enzymes involved in the degradation of amino acids were also found, including dehydrogenases for alanine (all bins), glutamate (all bins), threonine (LowGC-MGIII and Bathy2) and proline (LowGC-MGIII), as well as several aminotransferases for branched-chain amino acids (LowGC-MGIII and Bathy1) and aspartate/ tyrosine/aromatic aminotransferases (LowGC-MGIII and Bathy1). These findings suggest that there may be differences in the substrates used by the different MG-III groups. Indeed, although several subtilasefamily proteases (arCOG00702 and arCOG02553) were present in all bins, some peptidases had limited distributions: dipeptidyl-aminopeptidases (LowGC-MGIII and Bathy1), C1A-peptidases (LowGC-MGIII), C25-peptidases (Bathy1), Xaa-Pro aminopeptidases (Bathy2), and several AprE-like subtilisins (arCO-G06823, present in LowGC-MGIII and arCOG03610 present in Bathy1) (Supplementary Table S6).

Carbohydrates can be important carbon sources and, with the exception of Bathy1, several proteins with sugar-binding domains were found in all the bins (lectin and laminin-like). In the Epi6 bin, a cutin-like hydrolase was found (37% similar to a hydrolase from the Bacteriodetes *Rufibacter* sp. DG15C). Cutin is a polyester composed of hydroxyl/hydroxyepoxy fatty acids present in plants, and cutinases are produced by pathogenic fungi as extracellular degradative enzymes (Chen *et al.*, 1997). Lipo-oligosaccharide transporters were found exclusively in the LowGC-MGIII. As observed in MG-II Thalassoarchaea (Martin-Cuadrado

et al., 2015), multidrug and antimicrobial peptide transporters (ABC-type) together with several permeases for drug/metabolites (RhaT-like family) were also abundant in all MG-III bins. Although the nature of the substrates is difficult to ascertain, these transporters may be involved in coping with high environmental concentrations of toxins such as those produced by cyanobacterial and algal blooms.

Oxygen. The presence of superoxide dismutase in all MG-III bins, together with several genes for alkyl-hydroperoxide reductases in LowGC1-MGIII and Bathy1, suggests that these microbes must cope with oxygen radicals. Complete cytochrome-C and B-B6 oxidase subunits operons were also found in LowGC1-MGIII and Bathy1 and Bathy2 bins. Copper-binding proteins and haloarchaeal-like halocyanins were found in proximity of these operons, an arrangement similar to that described for MG-II Thalassoarchaea (Martin-Cuadrado et al., 2015). It has been suggested that MG-II could be facultative anaerobes (Martin-Cuadrado et al., 2008; Belmar et al., 2011) and that sulfate could be used as terminal electron acceptor. Although no sulfate reductase-like proteins could be identified in our MG-III bins, several phosphate/sulfate permeases could be identified in Epi6 and Bathy2 and were also present in Guavmas31/32 and Cayman92. Pterin-based molybdenum enzymes (for example, sulfite oxidase, xanthine oxidase and dimethyl sulfoxide reductase) function under anaerobic conditions whereby their respective cofactors serve as terminal electron acceptors in respiratory metabolism (Schwarz et al., 2009). For Bathy2 (fosmid Km3-43-F08), a novel operon for the molybdopterin biosynthesis, was found (catalytic domains, MOCS1/S2/S3, have <55% similarities in the nr-database). However, we could not find any of the pterin-based enzymes.

Light-related genes. The presence of photolyases/ cryptochromes among the LowGC-MGIII bins supports our hypothesis that they are *bona fide* epipelagic microbes (Figure 3a). Photolyases are proteins capable of photorepairing ultraviolet-induced pyrimidine dimers in the presence of light (Essen, 2006; Essen and Klar, 2006). Cryptochromes are proteins structurally similar to photolyases that act as blue light photoreceptors or regulators of the circadian rhythm (Cashmore et al., 1999) but that have lost the enzymatic photolyase activity (Chaves et al., 2011). Up to now, seven major classes of photolyase/cryptochrome families have been found (Scheerer *et al.*, 2015). Interestingly, while the subunits found in Epi1 and Epi3 have similarity with eukaryotic cryptochromes (38–49%), the photolyases found in Epi2A and Epi2C bins have their highest similarities with Planctomycetales homologs (30-52%), suggesting potential interdomain HGT events. Five related genes, a phytoene synthase, a phytoene-desaturase, an histidine kinase, a sugar-epimerase and one hypothetical protein, were found adjacent to the photolyase gene. At the equivalent genomic position, the aphotic Guaymas32

had neither the photolyase nor the associated genes mentioned above (downstream from a 23S-rRNA gene) (Figure 3a). The phylogenetic origin of the genes flanking the photolyases was analyzed and, in several cases, were most closely related to homologs from Bacteriodetes/Planctomycetes, again suggesting instances of HGT. These included a chaperone involved in protein secretion that was 76% similar to a Rhodopirellula mairorica homolog, a nitroreductase that was 75% similar to a Gracilimonas tropica homolog and a sugar-epimerase next to the photolyase that was 58% similar to a Pirellula staleyi protein. Likewise, a hypothetical protein adjacent to the photolyase in Epi1 and Epi3 was most closely related to eukaryotic genes, suggesting that this pair of genes may have been transferred together.

Epipelagic bins Epi1-2-3 all contained rhodopsins (Figure 2b) indicative of a photoheterotrophic lifestyle (Beja et al., 2000; Fuhrman et al., 2008; Inoue et al., 2013). In contrast, and consistent with previous reports (Deschamps et al., 2014; Li et al., 2015), Bathy1 and Bathy2 did not have rhodopsins. Phylogenetically, MG-III rhodopsins cluster with bacterial proteorhodopsins rather than with the euryarchaeal rhodopsins previously described for MG-II (Iverson *et al.*, 2012; Martin-Cuadrado *et al.*, 2014), suggesting that they may have been acquired by HGT from bacteria (Supplementary Figure S14). The analysis of key residues showed that all of these MG-III rhodopsins are proton pumps (Inoue et al., 2013) with a glutamine (Q) in the characteristic spectral tuning residue site indicating their ability to absorb light from the blue range (Supplementary Figure S16). In deeper waters (down to 300 m), only blue light remains available and blue rhodopsins are more suitable for generating energy. Therefore, epipelagic MG-III archaea seem to prefer low-light environments rather than the highly irradiated uppermost surface. Indeed, epipelagic MG-III bins recruited better from DCM or subsurface pelagic metagenomes (\sim 50–70 m) than from surface (5 m) ones (see below). Genomic comparisons with MG-II rhodopsins (Martin-Cuadrado et al., 2014) revealed two new genomic contexts for this gene (Figure 3b). Interestingly, one of the clusters also contains one of the photolyase genes previously mentioned (Figure 3, contig Epi3-ERR598942-C530). Downstream from the rhodopsin genes, a gene for an unknown GYD domain protein was present. In cyanobacteria, proteins containing GYD and KaiC domains are involved in generating circadian rhythms (Chang et al., 2015). This raises the possibility that epipelagic MG-III Euryarchaeota may also have a circadian rhythm. A similar genome segment was found in two Guaymas32 sequences but, in these cases, the rhodopsin and the GYD domaincontaining protein were absent.

The phylogenetic relationships of photolyases and rhodopsins, their proximity in at least one of the MG-III bins, together with the multiple putative HGT events observed in the nearby genes, leads us to

Novel epipelagic marine Euryarchaeota group III



Figure 3 (a) Comparative genomic organization of MG-III sequences containing photolyases (in yellow). (b) Comparative genomic organization of MG-III sequences containing rhodopsins (in red) in context with other genomic fragments containing the MG-II Pop, Pop-1, Pop-2, Pop-3 and Pop-4 rhodopsins (bottom). Conserved genomic regions are indicated by gray shaded areas, gray intensity being a function of sequence similarity by TBLASTX. Particular open reading frames mentioned in the text are highlighted by a graphic code (see legend).

hypothesize an ancestral 'dark nature' for MG-III. These light-related genes would have been recently transferred from epipelagic bacteria to MG-III, probably long after the massive HGT events that have been detected prior to the diversification of several mesophilic archaeal clades, including MGII/ III (Deschamps *et al.*, 2014; Lopez-Garcia *et al.*, 2016). The acquisition of proteorhodopsins, together with ultraviolet-protection photolyases, would have promoted a better adaptation to the oligotrophic surface waters allowing MG-III clades to expand into new photic niches.

Structural components

Cell envelope. One of the advantages of generating environmental fosmid sequences is that they allow the unequivocal assembly and detection of the socalled 'metagenomic islands' (Coleman *et al.*, 2006; Cuadros-Orellana *et al.*, 2007; Rodriguez-Valera *et al.*, 2009). These are clone-specific genome areas that, owing to their low coverage, are rarely assembled from metagenomic data sets but can be easily identified in reference-genome recruitment plots in the form of empty (or little populated) areas

The ISME Journal

with virtually no environmental homologs. One example can be observed in CG-Epi1. The area of the genome shown in Figure 4b (labeled with an

asterisk) is enriched in genes needed for cell wall biosynthesis and contains several glycosyltransferases (type I/IV), together with polysaccharide



Figure 4 (a) Heat map of the number of rpkg of each CG-MGIII of this work together with the ones of Li *et al.* (2015), MG-II and other archaea genomes used as references in 106 different metagenomes from different geographical points and depths. Only those collections in which any of the MG-III sequences recruited rpkg > 1 were represented. (b) Recruitment plots of the CG-Epi1, CG-Epi2, CG-Bathy1 and CG-Bathy2 genomes in the metagenomes where they were better represented, from surface (< 200 m) and bathypelagic (> 500 m) (BLASTN-based, see Methods section). Rpkg and the percentages of the total of the reads with an identity bigger than >95% are indicated. (c) Worldwide distribution of the CG-MGIII determined by metagenomic fragment recruitment against public metagenomic databases. Only samples where the CGs recruited rpkg>5 are indicated in the map (cutoff: %identity >95% in >50 bp coverage). TARA spots are indicated by T#station.

The ISME Journal

synthases and genes for carbohydrate modification (acyltransferases and aminotransferases). The presence of several lipopolysaccharide biosynthesis proteins in all MG-III bins suggests a more complex cell envelope than a protein layer (S-layer). Adjacent to the CG-Epi1 island, we found a giant protein of 7258 amino acids with no similarity in sequence databases. These types of proteins have previously been observed in several bacterial and archaeal genomes (Reva and Tummler, 2008; Strom et al., 2011) and have been hypothesized to have a role in defense against predation or in cell adhesion. Although we could not predict any function for it, the presence of lectin/glucanase domains (laminin_G3), glycosyl-transferase domains (RfaB), several beta-helix repeats and copper-binding domains (NosD) suggest an extracellular function. Large proteins (>5000 amino acids) with similar domains were also detected in other bins (Epi2-3-5). The similarity found between the giant proteins present in Guaymas31 and Bathy2 (90%) was remarkable.

Flagellum/Pili. Many archaeal surface structures are assembled by mechanisms related to the assembly of bacterial type IV pili (Lassak et al., 2012). With the exception of Epi5, we found several sequences containing two concatenated *flaJ* genes (implicated in archaeal flagellum assembly) followed by a *flaI* gene (a transcriptional activator). Syntenic operons were also found among deep-MG-III in Li et al. (2015). However, these gene clusters are very different from the flagellar operon found in MG2-GG3 (Iverson et al., 2012) or in any other Eurvarchaeota described to date (Jarrell and McBride, 2008; Jarrell et al., 2010). Although it has been claimed that the genes found might be enough to build a functional flagellum (Li et al., 2015), the lack of a more complex gene cluster suggests that this operon might be involved in a secretion system translocating proteins rather than in cell motility.

Prevalence in the marine environment

To evaluate the relative abundance of the novel MG-III genomes, we used the non-redundant CGs to recruit reads from >200 metagenomic data sets that provide reasonably complete coverage of open-ocean waters from around the world. Among them, 106 gave values higher than one rpkg for any of the CGs tested (Figure 4 and Supplementary Table S1). Negative results are probably due to the small size of the data sets (for example, GOS) that may have poor representation of less abundant organisms. Although a considerable number of MG-III clones have been detected in cold waters such as the deep Atlantic layer of the central Arctic Ocean, (Galand et al., 2009), the MG-III bins described here were not well represented in metagenomes from cold water regions such as polar regions (Alonso-Saez et al., 2012), the Baltic (Larsson et al., 2014) or the northeast subarctic Pacific (Allers et al., 2013). This may suggest that there are other abundant MG-III

Novel epipelagic marine Euryarchaeota group III JM Haro-Moreno et al

groups present in high latitudes that have yet to be discovered. Even in warmer latitudes, our LowGC-MG-III bins only represent a small fraction of the total prokaryotic population of photic marine habitats. The highest abundance we found was for CG-Epi1 that accounted for 0.5% of the reads in the samples from the Mediterranean station TARA-018 (ERR599073 collection) (Figure 4b). The deep MG-III bins recruited slightly more. For instance, CG-Bathy2 recruited up to 1% of the reads in the deep sample Med-Io17 (3500 m).

Figure 4 shows a clear correlation of the two MG-III groups with depth (as already suggested by the origin of the assembled bins). Most LowGC-MGIII bins are only present in epipelagic collections, while the HighGC-MGIII plus the LowGC Bathy1 and Guaymas32 were clearly bathy or mesopelagic. CG-Epi1 seemed to be evenly distributed throughout the photic zone, but CG-Epi3, 5 and 6 increased at deeper waters (25–155 m, including the DCM) and the three CG-Epi2 showed an increase in even deeper photic zone waters. Bathy1 has its maximum at mesopelagic waters (Adriatic Sea 600 m), but it was also detected in colder bathypelagic waters (for example, the metagenomes from the Cayman-Rise and Guaymas Basin). CG-Bathy2 together with the Cayman and Guaymas bins revealed a strong correlation with deeper waters with much higher abundance in metagenomic collections <1000 m. These bins were more abundant in the warmer (13 °C) and saltier Mediterranean deep samples (KM3, 3000 m and Io17, 3500 m deep), although the temperature in most bathypelagic waters, where these microbes were detected (global ocean), typically decreases down to <5 °C. Overall, these numbers indicate that MG-III cells are relatively minor components of the archaeal communities in the photic and aphotic zones.

Using the Mediterranean DCM time series data sets, we found significant temporal variation in the abundance of the different GC bins despite a relatively constant abundance of reads attributable to euryarchaeal 16S rRNA genes (Supplementary Figure S17). For example, CG-Epi2A predominated in 2012, whereas CG-Epi6 was dominant in 2013 and CG-Epi4 in 2014. In the case of MG-II, it has been experimentally demonstrated that eukaryotic phytoplankton additions stimulate their growth in bottle incubations (Orsi et al., 2015). Also, MG-II became one of the most abundant organisms (up to 40% of prokaryotes) in a phytoplankton bloom where diatoms, small flagellates and picophytoplankton dominated consecutively (Needham and Fuhrman, 2016). In order to know whether MG-II and the genomes of MG-III described here respond to similar blooming patterns, we measured the recruitment of available MG-II genomes in the metagenomes from which MG-III were assembled. The results show very low numbers for MG-II genomes in these samples, close to 100 times less than for MG-III genomes (Supplementary Figure S18). These data indicate that, despite being closely related and using similar substrates, MG-II and MG-III do not bloom concurrently.

1114

Using published plankton-interactome data (Lima-Mendez *et al.*, 2015), we constructed an interaction network for MG-III archaea (Supplementary Figure S19). The results showed that MG-III coexists mainly with Metazoa and Dinophyta, which represented 50.6% and 23.5% of the total of interactions observed. These findings may indicate that MG-III cells could be attached to other organisms and only sporadically be released to the environment.

Conclusions

The photic zone of the oligotrophic ocean, one of the largest microbial habitats on Earth, has been extensively explored by molecular and genomic approaches (DeLong, 1992; DeLong et al., 1999; Venter et al., 2004; Rusch et al., 2007; Sunagawa et al., 2015). Nevertheless, many epipelagic microbes remain to be characterized. Using metagenomics, we have uncovered eight new groups of planktonic marine Euryarchaeota that likely represent novel taxonomic orders or at least families. Based on differences in genome content and sequence identity, we propose the following nomenclature: Epipelagoarchaeales for the LowGC-MGIII and Bathypelagoarchaeales for the HighGC-MGIII. A separate and basal clade with low GC content but apparently living in the dark ocean (Bathy1) has also been uncovered. Genome comparisons between these new groups together and previously described MG-III genomes (Li et al., 2015) showed a marked differentiation between MG-III from photic and aphotic layers. Genomic analysis indicates that at least some representatives Epipelagoarchaeales (Epi1–Epi6) are planktonic photoheterotrophs. Two other groups with the Epipelagoarchaeales, Bathy1 and Guaymas32, lack genes indicating photoheterotrophy and are likely mesopelagic microbes with diverse metabolic capabilities. We hypothesize that the low GC content characteristic of the Epipelagoarchaeales may be an adaptation to the nitrogen limitation of surface waters. It is remarkable that all marine Euryarchaeota appear to possess similar metabolic profiles based on heterotrophic degradation of polymers and proteins (Iverson et al., 2012; Martin-Cuadrado et al., 2014; Li et al., 2015; Orsi et al., 2015). The broad diversity of marine microbes exploiting this habitat is likely a reflection of the enormous diversity of metabolic substrates available. Our data suggest a possible interaction of MG-III with eukaryotic cells and, more specifically, with metazoa.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We are thankful to José de la Torre for editing the manuscript. We are thankful to L Gasperini and G Bortoluzzi of the Instituto di Geologia Marina (ISMAR), CNR, Bologna (Italy) for allowing PLG to participate in the Marmara2010 R/V Urania cruise during which part of the samples analyzed in this study were collected. This work was supported by projects MEDIMAX BFPU2013–48007-P from the Spanish Ministerio de Economía y Competitividad, MaCuMBA Project 311975 of the European Commission FP7, project AQUAMET II/2014/012 from the Generalitat Valenciana and by the French Agence Nationale de la Recherche (ANR-08-GENM-024–001,EVOL-DEEP). JHM was supported with a PhD fellowship from the Spanish Ministerio de Economía y Competitividad.

References

- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**: 533–538.
- Alonso-Saez L, Waller AS, Mende DR, Bakker K, Farnelid H, Yager PL et al. (2012). Role for urea in nitrification by polar marine Archaea. Proc Natl Acad Sci USA 109: 17989–17994.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. J Mol Biol 215: 403-410.
- Allers E, Wright JJ, Konwar KM, Howes CG, Beneze E, Hallam SJ *et al.* (2013). Diversity and population structure of Marine Group A bacteria in the Northeast subarctic Pacific Ocean. *Isme J* **7**: 256–268.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S *et al.* (2004). The Pfam protein families database. *Nucleic Acids Res* **32**(Dataabase issue): D138–D141.
- Batut B, Knibbe C, Marais G, Daubin V. (2014). Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat Rev Microbiol* **12**: 841–850.
- Beja O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP et al. (2000). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. Environ Microbiol 2: 516–529.
 Belmar L, Molina V, Ulloa O. (2011). Abundance and
- Belmar L, Molina V, Ulloa O. (2011). Abundance and phylogenetic identity of archaeoplankton in the permanent oxygen minimum zone of the eastern tropical South Pacific. FEMS Microbiol Ecol 78: 314–326.
- Boutrif M, Garel M, Cottrell MT, Tamburini C. (2011). Assimilation of marine extracellular polymeric substances by deep-sea prokaryotes in the NW Mediterranean Sea. *Environ Microbiol Rep* 3: 705–709.
- Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat Rev Microbiol 6: 245–252.
- Bryant JA, Aylward FO, Eppley JM, Karl DM, Church MJ, DeLong EF. (2015). Wind and sunlight shape microbial diversity in surface waters of the North Pacific Subtropical Gyre. *ISME J* **10**: 1308–1322.
- Cashmore AR, Jarillo JA, Wu YJ, Liu D. (1999). Cryptochromes: blue light receptors for plants and animals. *Science* **284**: 760–765.

- Coleman ML, Chisholm SW. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci USA* **107**: 18634–18639.
- Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF *et al.* (2006). Genomic islands and the ecology and evolution of Prochlorococcus. *Science* **311**: 1768–1770.
- Cuadros-Orellana S, Martin-Cuadrado AB, Legault B, D'Auria G, Zhaxybayeva O, Papke RT *et al.* (2007). Genomic plasticity in prokaryotes: the case of the square haloarchaeon. *ISME J* 1: 235–245.
- Chang YG, Cohen SE, Phong C, Myers WK, Kim YI, Tseng R et al. (2015). Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. Science 349: 324–328.
- Chaves I, Pokorny R, Byrdin M, Hoang N, Ritz T, Brettel K et al. (2011). The cryptochromes: blue light photoreceptors in plants and animals. Annu Rev Plant Biol 62: 335–364.
- Chen S, Su L, Chen J, Wu J. (1997). Cutinase: characteristics, preparation, and application. *Biotechnol Adv* **31**: 1754–1767.
- DeLong EF. (1992). Archaea in coastal marine environments. Proc Natl Acad Sci USA 89: 5685-5689.
- DeLong EF. (2006). Archaeal mysteries of the deep revealed. Proc Natl Acad Sci U S A 103: 6417–6418.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- DeLong EF, Taylor LT, Marsh TL, Preston CM. (1999). Visualization and enumeration of marine planktonic archaea and bacteria by using polyribonucleotide probes and fluorescent in situ hybridization. *Appl Environ Microbiol* **65**: 5554–5563.
- Deschamps P, Zivanovic Y, Moreira D, Rodriguez-Valera F, López-García P. (2014). Pangenome evidence for extensive inter-domain horizontal transfer affecting lineage-core and shell genes in uncultured planktonic Thaumarchaeota and Euryarchaeota. *Genome Biol Evol* 6: 1549–1563.
- Dufresne A, Garczarek L, Partensky F. (2005). Accelerated evolution associated with genome reduction in a freeliving prokaryote. *Genome Biol* 6: R14.
- Edgar RC. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
- Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Edgcomb VP, Kysela DT, Teske A, de Vera Gomez A, Sogin ML. (2002). Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc Natl Acad Sci USA* **99**: 7658–7662.
- Essen LO. (2006). Photolyases and cryptochromes: common mechanisms of DNA repair and light-driven signaling? *Curr Opin Struct Biol* **16**: 51–59.
- Essen LO, Klar T. (2006). Light-driven DNA repair by photolyases. *Cell Mol Life Sci* **63**: 1266–1277.
- Fuhrman JÅ, Davis AA. (1997). Widespread archaea and novel bacteria from the deep sea as shown by 16S rRNA gene sequences. *March Ecol Prog Series* **150**: 275–285.
- Fuhrman JA, McCallum K, Davis AA. (1992). Novel major archaebacterial group from marine plankton. *Nature* 356: 148–149.
- Fuhrman JA, Schwalbach MS, Stingl U. (2008). Proteorhodopsins: an array of physiological roles? Nat Rev Microbiol 6: 488–494.

- Galand PE, Casamayor EO, Kirchman DL, Potvin M, Lovejoy C. (2009). Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. *ISME J* **3**: 860–869.
- Galand PE, Gutiérrez-Provecho C, Massana R, Gasol JM, Casamayor EO. (2010). Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). *Limnol Oceanogr.* 55: 2117–2125.
- Ghai R, Martin-Cuadrado A, Gonzaga A, Garcia-Heredia I, Cabrera R, Martin J et al. (2010). Metagenome of the Mediterranenan deep Chlorophyl Maximum stuidied by direct and fomid library 454 pyrosequencing. *ISME* J 4: 1154–1166.
- Ghai R, McMahon KD, Rodriguez-Valera F. (2012). Breaking a paradigm: cosmopolitan and abundant freshwater actinobacteria are low GC. *Environ Microbiol Rep* 4: 29–35.
- Giovannoni S, Nemergut D. (2014). Ecology. Microbes ride the current. *Science* **345**: 1246–1247.
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.
- Gribaldo S, Lumia V, Creti R, Conway de Macario E, Sanangelantoni A, Cammarano P. (1999). Discontinuous occurrence of the hsp70 (dnaK) gene among Archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. J Bacteriol 181: 434–443.
- Haff DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT et al. (2001). TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res 29: 41–43.
- Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y, Sugahara J et al. (2006). Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. Proc Natl Acad Sci USA 103: 18296–18301.
- Herndl GJ, Reinthaler T, Teira E, van Aken H, Veth C, Pernthaler A *et al.* (2005). Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. *Appl Environ Microbiol* **71**: 2303–2309.
- Huang Y, Gilna P, Li W. (2009). Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **25**: 1338–1340.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.
- Inoue K, Ono H, Abe-Yoshizumi R, Yoshizawa S, Ito H, Kogure K *et al.* (2013). A light-driven sodium ion pump in marine bacteria. *Nat Commun* 4: 1678.
- Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* 335: 587–590.
- Jarrell KF, Jones GM, Kandiba L, Nair DB, Eichler J. (2010). S-layer glycoproteins and flagellins: reporters of archaeal posttranslational modifications. Archaea 2010.
- Jarrell KF, McBride MJ. (2008). The surprisingly diverse ways that prokaryotes move. *Nat Rev Microbiol* 6: 466–476.

The ISME Journal

- Karner MB, DeLong EF, Karl DM. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409: 507–510.
- Konneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437: 543–546.
- Konstantinidis KT, Tiedje JM. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102**: 2567–2572.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D et al. (2009). Circos: an information aesthetic for comparative genomics. Genome Res 19: 1639–1645.
- Larsson J, Celepli N, Ininbergs K, Dupont CL, Yooseph S, Bergman B *et al.* (2014). Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J.*
- Lassak K, Ghosh A, Albers SV. (2012). Diversity, assembly and regulation of archaeal type IV pili-like and nontype-IV pili-like surface structures. *Res Microbiol* 163: 630–644.
- Lassmann T, Sonnhammer EL. (2005). Kalign–an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics 6: 298.
- Li H, Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Li M, Baker BJ, Anantharaman K, Jain S, Breier JA, Dick GJ. (2015). Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat Commun* 6: 8933.
- Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F *et al.* (2015). Ocean plankton. Determinants of community structure in the global plankton interactome. *Science* **348**: 1262073.
- Lopez-Garcia P, Lopez-Lopez A, Moreira D, Rodriguez-Valera F. (2001a). Diversity of free-living prokaryotes from a deep-sea site at the Antarctic Polar Front. *FEMS Microbiol Ecol* 36: 193–202.
- Lopez-Garcia P, Moreira D, Lopez-Lopez A, Rodriguez-Valera F. (2001b). A novel haloarchaeal-related lineage is widely distributed in deep oceanic regions. *Environ Microbiol* **3**: 72–78.
- Lopez-Garcia P, Philippe H, Gail F, Moreira D. (2003). Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc Natl Acad Sci USA* **100**: 697–702.
- Lopez-Garcia P, Zivanovic Y, Deschamps P, Moreira D. (2016). Bacterial gene import and mesophilic adaptation in archaea. *Nat Rev Microbiol* **13**: 447–456.
- Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, Wolf YI *et al.* (1999). Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res* **9**: 608–628.
- Makarova KS, Koonin EV. (2003). Comparative genomics of Archaea: how much have we learned in six years, and what's next? *Genome Biol* **4**: 115.
- Makarova KS, Wolf YI, Koonin EV. (2015). archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between Thermococcales, Methanococcales, and MethanobacterialesLife (Basel) 5: 818–840.
- Martin-Cuadrado AB, Fontaine T, Esteban PF, del Dedo JE, de Medina-Redondo M, del Rey F *et al.* (2008).

Characterization of the endo-beta-1,3-glucanase activity of *S. cerevisiae* Eng2 and other members of the GH81 family. *Fungal Genet Biol* **45**: 542–553.

- Martin-Cuadrado AB, Garcia-Heredia I, Molto AG, Lopez-Ubeda R, Kimes N, Lopez-Garcia P et al. (2015). A new class of marine Euryarchaeota group II from the Mediterranean deep chlorophyll maximum. ISME J 9: 1619–1634.
- Martin-Cuadrado AB, Pasic L, Rodriguez-Valera F. (2014). Diversity of the cell-wall associated genomic island of the archaeon *Haloquadratum walsbyi*. BMC Genomics 16: 603.
- Martin-Cuadrado AB, Rodriguez-Valera F, Moreira D, Alba JC, Ivars-Martinez E, Henn MR et al. (2008). Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *ISME J* 2: 865–886.
- Massana R, DeLong EF, Pedros-Alio C. (2000). A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl Environ Microbiol* **66**: 1777–1787.
- Mizuno CM, Ghai R, Saghai A, Lopez-Garcia P, Rodriguez-Valera F. (2016). Genomes of abundant and widespread viruses from the deep ocean. *MBio* **7**: 4.
- Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ et al. (2012). De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. ISME J 6: 81–93.
- Nawrocki EP. (2009). Structural RNA homology search and alignment using covariance models, PhD thesis. Washington University in Saint Louis, School of Medicine, St. Louis, MO, USA.
- Needham DM, Fuhrman JA. (2016). Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat Microbiol*.
- Orsi WD, Smith JM, Wilcox HM, Swalwell JE, Carini P, Worden AZ *et al.* (2015). Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *ISME J* **9**: 1747–1763.
- Ouverney CC, Fuhrman JA. (2000). Marine planktonic archaea take up amino acids. *Appl Environ Microbiol* **66**: 4829–4833.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2014). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25: 1043–1055.
- Peng Y, Leung HC, Yiu SM, Chin FY. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28: 1420–1428.
- Petitjean C, Moreira D, Lopez-Garcia P, Brochier-Armanet C. (2012). Horizontal gene transfer of a chloroplast DnaJ-Fer protein to Thaumarchaeota and the evolutionary history of the DnaK chaperone system in Archaea. *BMC Evol Biol* **12**: 226.
- Qin W, Amin SA, Martens-Habbena W, Walker CB, Urakawa H, Devol AH et al. (2014). Marine ammoniaoxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. Proc Natl Acad Sci USA 111: 12504–12509.
- Quaiser A, Zivanovic Y, Moreira D, Lopez-Garcia P. (2011). Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. *ISME J* 5: 285–304.

1116

The ISME Journal

- Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P. (2007). Prediction of effective genome size in metagenomic samples. *Genome Biol* 8: R10.
- Raymann K, Forterre P, Brochier-Armanet C, Gribaldo S. (2014). Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biol Evol* 6: 192–212.
- Reva O, Tummler B. (2008). Think big-giant genes in bacteria. *Environ Microbiol* **10**: 768-777.
- Rice P, Longden I, Bleasby A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277.
- Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B et al. (2009). Explaining microbial population genomics through phage predation. Nat Rev Microbiol 7: 828–836.
- Romero H, Pereira E, Naya H, Musto H. (2009). Oxygen and guanine-cytosine profiles in marine environments. *I Mol Evol* 69: 203–206.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S et al. (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol 5: e77.
- Scheerer P, Zhang F, Kalms J, von Stetten D, Krauss N, Oberpichler I et al. (2015). The class III cyclobutane pyrimidine dimer photolyase structure reveals a new antenna chromophore binding site and alternative photoreduction pathways. J Biol Chem 290: 11504–11514.
- Schwarz G, Mendel RR, Ribbe MW. (2009). Molybdenum cofactors, enzymes and pathways. Nature 460: 839– 847.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504.
- Smedile F, Messina E, La Cono V, Tsoy O, Monticelli LS, Borghini M et al. (2012). Metagenomic analysis of hadopelagic microbial assemblages thriving at the

deepest part of Mediterranean Sea, Matapan-Vavilov Deep. *Environ Microbiol* **3**: 1462–2920.

- Strom SL, Brahamsha B, Fredrickson KA, Apple JK, Rodriguez AG. (2011). A giant cell surface protein in Synechococcus WH8102 inhibits feeding by a dinoflagellate predator. *Environ Microbiol.*
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G et al. (2015). Ocean plankton. Structure and function of the global ocean microbiome. Science 348: 1261359.
- Swan BK, Chaffin MD, Martinez-Garcia M, Morrison HG, Field EK, Poulton NJ *et al.* (2014). Genomic and metabolic diversity of Marine Group I Thaumarchaeota in the mesopelagic of two subtropical gyres. *PLoS One* 9: e95380.
- Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, Gonzalez JM et al. (2013). Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci USA 110: 11463–11468.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.
- Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM *et al.* (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.
- Vavourakis CD, Ghai R, Rodriguez-Valera F, Sorokin DY, Tringe SG, Hugenholtz P *et al.* (2016). Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline Soda Lake Brines. *Front Microbiol* 7: 211.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66–74.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)

SUPPLEMENTARY INFORMATION

Supplementary Figure S1. Schematic representation of the assembly and binning procedure.

Supplementary Figure S2. Binning of the MG-III genomic fragments. a. Principal component analysis of tetranucleotide frequencies of the MG-III DNA fragments. Reference sequences are shown as larger circles: *Nitrosopumilus maritimus SCM1*, *Aciduliprofundum boneii* T469, MGII-GG3, MGII-Thalassoarchaea, the single amplified genome (SAG) SCGC-AAA288-E19 and the DNA fragments from Cayman92, Cayman93, Guaymas31 and Guaymas32. **b.** Heat-map of the number of reads per kilobase per gigabase of metagenome collection (rpkg) of the MG-III sequences recruiting in 33 different metagenomes (only those in which any of the MG-III sequences recruited over rpkg>3 were considered for the binning analysis).

Supplementary Figure S3. Distribution of the assembled MG-III sequences among the metagenomes of the Mediterranean series and TARA collections ordered by depth. Size fraction is indicated above each column.

Supplementary Figure S4. Average nucleotide identity (ANI) among the MG-III genome bins (in bold the sequences published in this work). Dendrogram showing the similarity among the bins is shown in the *y* axis.

Supplementary Figure S5. 16S and 23S-rRNA phylogeny. a. Maximum-likelihood 16S-rRNA gene tree showing the relationship of the MG-III with other archaea. Circles at nodes in major branches indicate bootstrap support (see legend). Scale bar represents the estimated number of substitutions per site. The different MG-III clusters are indicated by different colours and named following Galand *et al.* 2008 and this work. Sampling location is indicated in each of the sequences. Those sequences from samples from more than 500m deep are underlined. **b.** Maximum likelihood 23S-rRNA

gene tree showing the relationship of the MG-III with other archaea. (Picture features as explained before).

Supplementary Figures S6-S12. Maximum likelihood phylogenetic trees for the housekeeping genes RecA, RpoB, SecY, geranylgeranylglyceryl phosphate synthase, DnaK, GyrA and GyrB. Protein sequences from this study are indicated in bold and coloured accordingly with the sequence bin (see legend in Supplementary Figure S4). Genomic bins of MG-II and MG-IIII groups are indicated. Bootstrap values over 50 are indicated.

Supplementary Figure S13. Maximum likelihood phylogenetic tree of the photolyases and cryptochromes found among the MG-III bins. Protein sequences from this study are indicated in bold and coloured accordingly with their kingdom affiliation (see legend). Bootstrap values over 50 are indicated.

Supplementary Figure S14. Maximum likelihood phylogenetic tree showing the relationship of the MG-III rhodopsins with other bacterial and archaeal rhodopsins. Protein sequences from this study are indicated in bold and coloured accordingly with the sequence bin (see legend in Supplementary Figure S4). Following the nomenclature of Iverson et al. (2013), Clade A and Clade B of rhodopsins is shown. In blue are marked Pop, Pop-1, Pop-2, Pop-3 and Pop-4 euryarchaeal rhodopsins previously described. Numbers at nodes in major branches indicate bootstrap support (shown as percentages and only those >50%). Scale bar represents the estimated number of substitutions per site.

Supplementary Figure S15. Distribution of arCOG functional classes. Percentage of arCOGs predicted in the MGIII bins described in this work and MG-II marine euryarchaeal genomes MG2-GG3 and Thalasoarchaea. All genes (**a**) and genes found only in one of the MGIII bins (**b**) are indicated. Asterisks indicate categories where a significant variation was found comparing the epipelagic and pelagic MGIII.

Supplementary Figure S16. Alignment of the MG-III rhodopsins with other cloned rhodopsins sequences. Identical residues are indicated in red. Residues in blue are conserved in more than 70% of the sequences. Key amino acids for rhodopsins functionality (listed herein with G. pallidula numbering) are marked by colours: Lys336 (K) binds retinal, and Asp164 (D) and Glu175 (E) function as Schiff base proton acceptor and donor, respectively. Glutamine (Q) in position 172 (*) in the MG-III rhodopsins sequences indicates an absorption maxima at the blue spectrum range. Letters (G) and (B) in the name of the sequences indicate the range of the spectrum. (The GenBank accession numbers of the sequences used for the alignment are as follows: Pop-2 HF10 3D09, 82548293; Pop-3 HF70 19B12, 82548286; Pop-4 HF70 59C08, 77024964; eBAC49C08, AAY82659; HF130 81H07, 119713419; HF10 49E08, 119713779; eBAC20E09, AAS73014; HOT75m4, AAK30179; eBAC31A08 (SAR86), AAG10475; SAR86E, WP_008490645; C. Pelagibacter ubique HTCC1062 (SAR11), YP 266049; Pelagibacter sp. HTCC7211, WP 008544914; gammaproteobacteria HTCC2207 (SAR92), EAS48197; G. pallidula, WP 006008821; Dokdonia donghaensis MED134, ZP 01049273; MedDCM-JUL2012-C3793, KP211832; KP211865; MedDCM-OCT2007-C1678, MedDCM-OCT-S08-C16, ADD93192; Exiguobacterium sibiricum 255-15, ACB60885; Exiguobacterium sp. AT1b, WP 012726785; Haloarcula marismortui ATCC 43049, YP 136594.

Supplementary Figure S17. Classification of the DCM-metagenomes reads using the RDP (16S-rRNA) database. Only those genera which represented more than 1% were represented.

Supplementary Figure S18. Number of reads per kilobase per gigabase of metagenome collection (rpkg) for the composite genome CG-Epi1 MG-III (ordered from minus to major) compared with those obtained for the MG-II (MG2-GG3 (Iverson *et al.* 2013) and the Thalassoarchaea (Martin-Cuadrado *et al.* 2014)). Other CG MG-III were also included in the graph.

Supplementary Figure S19. Interactions (showed as percentages) calculated by Mendez-Lima *et al.*, (2015) for the MG-III with other organisms. Data extracted from Supplementary table W7 in Mendez-Lima *et al.*, (2015).

SUPPLEMENTARY TABLES

Supplementary Table S1. List of metagenomes used for recruitments in Figure 4a and Supplementary Figure 2b, sorted by temperature and depth.

Supplementary Table S2. List of sequences of each of the MG-III sequence bins.

Supplementary Table S3. List of sequences of the MG-III composite genomes (CG-MGIII).

Supplementary Table S4. Housekeeping genes found in the MG-III bins and the CG-MG-III bins (as Narasingarao *et al.*, (2012).

Supplementary Table S5. CG-MGIII-protein categories based on the arCOG database.

Supplementary Table S6. Classification of the CG-MGIII unique CDSs based on the arCOG classification.

Supplementary Table S7. List of genes involved in MG-III metabolic pathways.



Supplementary Figure S2























Bacteria

Eukarya







а

| | 10 | 20

 | 30
 | 40
 | |
 | | 80 | 90
 | | 110 |
|---|---
--

--
--
--
--|---|--
---	--
--	
Pop-2 HF10 3D09 Eni-28 MedDCM-IIII 2012-C1059	

 | SKQRSLLLTA
 | LLTLLTFGTV
 | S A A N S | T L A
 | T N D M V G | | MMLASTVFFI
 | MERNNVADKW
VERSNVAPKW | |
| Epi-1 ERR598993-C1654 | | ML S

 | MRAMGITAFM
 | SLLLAGSASA
 | E T G | A I D
 | T S DN V A | ISFWIATA | MMLASTIFFL
 | VERNNVAPKW | RTSVTVAA-L |
| MedDCM-OCT2007-C4997 | | MLS

 | MRAMGITAFM
 | S L L L A G S A S A
 | E T G | A I D
 | T S DN V A | I S FWIATA | MMLASTIFFL
MMLASTVEEL
 | VERNNVAPKW | R T S V T V A A - L |
| Pop-4 HF70 59C08 | MR | FSESMDINPF

 | LKKRTAVVGG
 | ALALAAITIP
 | SAAAQTTETV | G L G
 | A T D Y V G | I S FWLATA | MMLATTVFFL
 | VERQNVAAKW | KTSMTVAA-L |
| Epi-2B ERR598983-C478 | | MTKIR

 | LAAFALCILS
 | IMILPNSAAT
 | KDSYSEDGFL | I L G
 | Q D D L V G | I T F W I A T A | AMLGSSIFFL
 | I E R S N V A P Q W | KTSMTIAT-L |
| MedDCM-OCT-S24-C2
Epi-1 MedDCM-OCT-S34-C62 | | MFSSLKT
MFFSLKN

 | TLFLIVCIFT
 | VELLPTSAAM
 | G D S Y S E D G F L
G D S Y S E D G F L | V L G
 | Q E D L V G | ITFWIATA | AMLASSLFFL
 | I ERANVAPOW | KTSMTIAT-L
KTSMTIAT-L |
| Epi-3 ERR598942-C530 | |

 | MCFFT
 | I L I L P S S V A T
 | GESYSEDGFL | V L G
 | Q D D L V G | I T F W I A T A | AMLASSLFFL
 | IERANVAPQW | KTSMTIAT-L |
| MedeBAC49C08
HF130 81H07 | |

 | MG E
 |
 | |
 | V D D Y V G | I S FWLAAA | IMLASTVFFF
 | V E R S D V P V K W | K T S L T V A G - L
K T S L T V A G - L |
| HF10 49E08 | |

 | MTFA
 | DDHADEETAA
 | T E P A E V V Y D G | S L H
 | P D D F V G | V S FWL A T A | MMLAATVFFF
 | IERDRVKGKW | K T S L T V A G - L |
| gamma EBAC20E09 (B)
gamma Hot 75m4 (B) | |

 | - MKLLLTMVG
 | ALALPVFAA-
 | 66 | DLD
 | T A DMT G | V S FWL V T A | AMLAATVEEL
 | V E R D R V H G K W | K T S L S V A A - L |
| gamma EBAC31A08 (G) | |

 | - MKLLLILGS
 | VIALPTFAAG
 | G G | DLD
 | A S D Y T G | V S FWL V T A | ALLASTVFFF
 | VERDRVSAKW | KTSLTVSG-L |
| SAR86E (G) | |

 | - MKLLMLLAG
 | VMAVPTFAAG
 | | D L D
 | P N D Y V G | V S FWL V T A | ALLAATVFFF
 | LERDNVSAKW | R T S L T V S G - L |
| Pelagibacter sp. HTCC7211(B) | | MKK

 | LKLFALTAVA
 | LLGVTGVANA
 | DA | MLA
 | Q D D F V G | I S FWV I SM | GMLAATAFFF
 | METGNVAAGW | R T S V I V A G - L |
| gamma HTCC2207 (G) | |

 |
 |
 | |
 | | | ALVAATAFFF
 | V E R D R V A G K W | K T S L T V S G - L |
| D.donghaensis MED134 (G) | |

 | M
 | NLLLLSTALD
 | | K A
 | S D D Y V A | F T F F V G A M | AMMAAAAFFF
 | LSMNSFDRKW | RTSILVSG-L |
| MedDCM-JUL2012-C3793 | | MN S K L S K F

 | SPRTNAIIAG
 | ATMLALMVTQ
 | NVSAATIDPE | ΤΥΥ
 | A G D A L G R A | TFFMFFIGYI | SMGAAFVFFM
 | S E R N N V A P E Y | R T T M T I S A - L |
| Pop MGII GG3 | |

 | SERKNVIVEI
 | TISLA-MAGT
 | VSAADVVDPS | V E Y
 | DGEPLOLV | TFFLFFVGYI | SMGAAFVFFM
 | SERSNVAPEF | RTTMTISA-L |
| MedDCM-OCT-S26-C80 | M S N M L N H | SDTKLRSLKL

 | NRRNSVIFAV
 | LMLSLTAIAM
 | PTAAEIVDPA | ККҮ
 | D G E P L Q L L | TFFLFFVGYI | SMGAAFVFFM
 | SERSNVAPEY | R T T M T I S A - L |
| E. sibiricum 255-15 | | MITTERF

 | MEEVNLLVLA
 | TQ
 | SAAGVGEMDI | EIGLPSFLSD
 | DLGNDSLELL | YMFWVGFV | GMAAGTLYFL
 | VERNSLAPEY | R STATVAA - L |
| Exiguobacterium sp. AT1b | |

 | MFVVG
 | Q F S L K GM GMM
 | N E |
 | E V N L L V L A | TQYMFWVGFV | GMAAATLYFL
 | VERSSLDPEY | R S V A T V A A - L |
| H. marismortui ATCC 43049 | |

 | M L
 | P L Q V S S L G V E
 | G E G |
 | | IWLALGTV | GML L GMV Y F M
 | A K GWD V Q D P E | QEEFYVITIL |
| | 120 | 130

 | 140
 | 150
 | 160 | D 170
 | * E/K 180 | 190 | 200
 | 210 | 220 |
| Pop-2 HF10 3D09 | VTGVAWYHYT | - YMR DHWANS

 | YAASGDAOV-
 | · · · · · · ·
 | D S P L V L - |
 | | AAIGVASAAL | FWRL-F
 | GASIVMLVAG | FLAEAN ED |
| Epi-2B MedDCM-JUL2012-C1059 | VTGVAWYHYT | - YMREHWVL-

 | · · · · · · · · · · · · · · · · · · ·
 | · · · · · N · · ·
 | E S P L V L - | RYVDWIITVP
 | LQVVEFYLIL | AAIGVASAML | FWRL-L
 | GASVVMLAFG | F L G E S G |
| Epi-1 ERR598993-C1654
MedDCM-OCT2007-C4997 | V T G V AWY H Y T
V T G V AWY H Y T | - YMRDHWVM-
- YMRDHWVM-

 |
 | G
 | E S P L V L -
E S P L V L - | K Y V DWL I T V P
R Y V DWL I T V P
 | | A A I G V A S AML
A A I G V A S AML | FWRL-L
 | GASIVMLAFG
GASIVMLAFG | F F G E S G
F F G E S G |
| Pop-3 HF70 19B12 | VTGVAWYHYT | - YMR E VWAQG

 | Y G V D D A -
 | · · · · · · v · · ·
 | G S P L V Y - | RYIDWLITVP
 | LQVVEFYLIM | AAIGVGTFAM | FRNL-M
 | GASIVMLVAG | F F G E S G A W E L |
| Pop-4 HF70 59C08
Epi-2B ERR598983-C478 | V T G I A WY H Y T
I T G I A F WH Y M | - YMRQHWVD -
YMREHWIL -

 |
 | G
 | S S P I V Y - | R Y I DWL L T V P
R Y V DWF I T V P
 | | A A I G V A S V V M
A A V T A V T A M I | FRNL-M
 | LASVVMLVAG
GASVLMLVFG | F F G E T G A W D L
F L G E T Q |
| MedDCM-OCT-S24-C2 | ITGIAFWHYM | - YMREHWII-

 |
 | G
 | E S P I V Y - | R Y V D W F L T V P
 | LQIVEFYFIL | AAVTTVTAML | FWR L - L
 | GSSVLMLVFG | FLGETQ |
| Epi-1 MedDCM-OCT-S34-C62
Epi-3 ERR598942-C530 | | - YMREHWII-

 |
 | G
 | ESPIVY- | R Y V DWFL T V P
 | | A A V T T V T A M L
A A V T T V T A M I | FWRL-L
 | G S S V L ML V F G | F L G E T Q F L G E T Q |
| MedeBAC49C08 | VTGVAFWHYL | YMR GVWIY

 | A .
 |
 | ETPTVF | RYIDWLITVP
 | LQIIEFYLII | AAVTAISSAV | FWKL-L
 | IASLVMLIGG | FIGEAG |
| HF130 81H07
HF10 49F08 | V T G I A F WH Y L | - YMR DVWVM -

 | · · · · · · · · · · · · · · · · · · ·
 | G
 | DTPTVY- |
 | | A A V T A I S A A I
A A V T K V S A N I | FWKL-L
 | TASIVMLVFG | YLGETG |
| gamma EBAC20E09 (B) | VTGIAFWHYL | YMRGVWVES

 | Y T G P G T -
 | G
 | | RYIDWLITVP
 | LQIIEFYLIL | KVCTNVGSGL | FWR L - L
 | GASLVMLVGG | FIGETG |
| gamma Hot 75m4 (B) | I T G I A F WHYL | - YMRGVWID-

 | · · · · · · · · · · · · · · · · · · ·
 | G
 | D T P T V F - | RYIDWLLTVP
 | | AACTSVAASL | FKKL-L
 | AGSLVMLGAG | |
| SAR86E (G) | VTGIALWHYL | - YMRGVWVD-

 | т.
 | G
 | T S P T V F - | RYIDWLLTVP
 | LLIVEFYLIL | RAVTDVAASL | FYKL-F
 | | YMGEAG |
| P.ubique HTCC1062 (G) | VTGIAFIHYM | - YMR DVWVM -

 | <u>T</u> -
 | · · · · · · · · · · · · · · · · · · ·
 | E S P T V Y - | RYIDWLITVP
 | LLMLEFYFVL | AAVNKANSGI | FWRL-M
 | IGTLVMLIGG | YLGEAG |
| gamma HTCC2207 (G) | VTLIAAVHYF | - YMR EVWVI-

 | т.
 | G
 | | RYIDWLITVP
 | | S A V G K A N S G M
S A V T K V P A G V | FWRL-L
 | IGTIAMLGFG | Y A G E A G |
| G.pallidula (B) | VTFIAAVHYF | - YMR D V WV A -

 | · · · · · · · · · · · · · · · · · · ·
 | · · · · · q · · ·
 | T T P T V Y - | R Y I <mark>D</mark> WL L T V P
 | LLMIEFYLIL | RAIGVASPGI | FWR L - L
 | LGTLVMLIPG | YMA E A G |
| MedDCM-JUL2012-C3793 | IVGIAAFHYY | - YMRDYWQA-

 |
 | G
 | AVSTDY- | RYMDWIITVP
 | LMALKFPSLV | GKDAITDNKV | LGLGFTGVCF
 | VGALIMIAFG | YLGEIG |
| MedDCM-OCT2007-C1678 | IVGIAAFLLL | YYMRGVYID-

 | · · · · · · · · · · · · · · · · · · ·
 | · · · · · · · · · · · · · · · · · · ·
 | A V S T D Y - | RYMDWIITVP
 | LMALKFPSLV | GKDAITDNKV | LGLGFTGVCF
 | VGALIMISFG | Y L G E T G |
| MedDCM-OCT-S26-C80 | IVGIAAFHYY | - YMR A V Y T D -

 |
 | G
 | | RYMDWIITVP
 | | GKDAITDAKF | AGLGFTGICF
 | TGALIMIGFG | Y A G E A G |
| Pop-1 MedDCM-OCT-S08-C16 | IVGIAAFHYY | - YMR G V Y A D -

 | F.
 | G
 | V V S I E Y - | RYMDWIITVP
 | LMALKFPSLV | GKEAITDKKV | AGLGFTGICF
 | TGALIMIGFG | Y L G E A G |
| E. sibiricum 255-15
Exiguobacterium sp. AT1b | V T F V A A I H Y Y
V T F V A A I H Y Y | - FMKDAVGT-

 |
 | GLLS
 | EIDGFPTEI-
EITGFPTEI- | RYIDWLVTTP
 | | G L K G R L G R P L
S L O G K M K T G L | LTKL-V
 | | Y I G E S S
Y L G E S S |
| H. marismortui ATCC 43049 | I A G I A A S S Y L | - SMFFGFGL-

 | · · · · · · · · T E
 | V E L V N G R
 | V I D V Y WA | RYADWLFTTP
 | LLLDIGLLA | GASNRDMASL |
 | TIDAFMIVTG | L A A T L M |
| | |

 |
 |
 | |
 | | |
 | | |
| | 220 | 240

 | 250
 | 260
 | 270 | 280
 | 200 | 200 | 210
 | 220 | 220 |
| | 230 | 240

 | 250
 | 260
 | 270 | 280
 | 290 | 300 | 310
 | 320 | 330 |
| Pop-2 HF10 3D09
Epi-28 MedDCM-JUL2012-C1059 | 230 | 240
D GMVWP L F A V
EMV AW A I

 | 250
 | 260
E I WA G
 | 270
- EAKESVS - G
- DAKAAAN - N | 280
A S E G T Q F A F K
A S E G V Q F A F T
 | 290 | 300 | 310

QGD
AGTLGDN
 | 320 | 330
DNL - NILYNI
DML - NIVYNI |
| Pop-2 HF10 3D09
Epi-2B MedDCM-JUL2012-C1059
Epi-1 ERR598993-C1654 | 230
I S A I S
AMN | 240
D GM V WP L F A V
EM V A W A I
A T I A W A I

 | 250
GML AWI Y I I Y
GMA AWI Y I I Y
GMA AWI Y I I Y
 | 260
E I WA G
E I WA G
E I WA G
 | 270
- E A K E S V S - G
- D A K A A A N - N
- D A K K A A S - E | 280
A S E G T Q F A F K
A S E G V Q F A F T
A S E G V Q F A F T
 | 290
AMALILTVGW
WMTYILTFGW
WMTYILTFGW | 300
A I Y P L G F I L G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G | 310
QGD
AGTLGDN
MDADA
 | 320
GGT
PDT
DGQ | 330
DNL-NILYNI
DML-NIVYNI
NML-NIVYNI |
| Pop-2 HF10 3D09
Epi-2B MedDCM-JUL2012-C1059
Epi-1 ERR598993-C1654
MedDCM-OCT2007-C4997
Pop-3 HF70 19812 | 230
I S A I S
- AMN
- AMD
F G D L S | 240
DGMVWPLFAV
EMVAWAI
ATIAWAI
ATIAWAI

 | 250
GMLAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
 | 260
E I WA G
E I WA G
E I WA G
E I WA G
E I WY G
F L WO G
 | 270
 | 280
A S E G T Q F A F K
A S E G V Q F A F T
A S E G V Q F A F T
A S E G V Q F A F T
G S D G V Q F A F K
 | 290
AMALILTVGW
WMTYILTFGW
WMTYILTFGW
WMTYILTFGW
AMRALVTFGW | 300
AIYPLGFILG
AIYPIGYLYG
AIYPIGYLYG
AIYPIGYLYG
AIYPLGYAMG | 310

QGD
AGTLGDN
MDADA
MDADA
NDML-SG
 | 320
GGT
PDT
DGQ
 | 330
DNL - NIL YNI
DML - NIV YNI
NML - NIV YNI
NML - NIV YNI
DDM - NIV YNI |
| Pop-2 HF10 3D09
Epi-2B MedDCM-JUL2012-C1059
Epi-1 ERR598993-C1654
MedDCM-0CT2007-C4997
Pop-3 HF70 19812
Pop-4 HF70 59C08 | 230
I S A I S | 240
D GM VWP L F A V
EM V AW A I
A T I AW A I
T E I WW A I
T E I WW A I

 | 250
GMLAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMAGWGYILY
GMAGWGYILY
GMAGWGYILY
 | 260
E I WA G E I WA G E
E I WA G E I WA G E
E I WY G E I FR G E I FR G
 | 270
- EAKESVS - G
- DAKAAAAN - N
- DAKKAAAS - E
- DAKKAAG - D
- DVKEASA - S
- SINDAAQ - K | 280
A S E G T Q F A F K
A S E G V Q F A F T
A S E G V Q F A F T
G S D G V Q F A F K
A S E S V Q F A F N
 | 290
AMALILTVGW
WMTYILTFGW
WMTYILTFGW
WMTYILTFGW
AMRAIVTFGW
SMRWIVLVGW | 300
AIYPLGFILG
AIYPIGYLYG
AIYPIGYLYG
AIYPLGYLYG
AIYPLGYLMG
AIYPLGYMMG | 310
QGD
AGTLGDN
MDADA
MDADA
NDMI-SG
NDML-PG
 | 320
 | 330
DNL - N I L YN I
DML - N I V YN I
NML - N I V YN I
NML - N I V YN I
DDM - N I I YN I
NDM - N I I YN L |
| Pop-2 HF10 3D09
Epi-28 MedDCM-JUL2012-C1059
Epi-1 ERRS9893-C1654
MedDCM-OCT2007-C4997
Pop-3 HF70 19812
Pop-4 HF70 55008
Epi-28 ERRS98983-C478
MedDCM-OCT-524-C2 | 230
I S A I S
- AMN
- AMD
F G D L S
F G D A S
- ML G G S
- ML G G S | 240
D GMVWP L F A V
EMV A W A I
A T I A W A I
A T I A W A I
T E I W A I
T E I W A I
S I I W TL

 | 250
GML AWI Y I I Y
GMAAWI Y I I Y
GMAAWI Y I I Y
GMAAWI Y I I Y
GMAGWI Y I I Y
GMAGWI Y I I Y
GMA CWF Y I I Y
GMY CWF Y I I Y
 | 260
E I WA G
 | 270
- E AK E S V S - G
- D AK A A A N - N
- D AK K A A S - E
- D AK K A A G - D
- D V K E A S A - S
- S I N D A A Q - K
- E A A I L N E N S
- E A A I L N E N S | AS EGTQFAFK
AS EGVQFAFT
AS EGVQFAFT
GSDGVQFAFK
GSDGVQFAFK
AS ESVQFAFN
GN EAGQFAFQ
GN EAGQFAFQ
 | 290
AMALILTVGW
WMTYILTFGW
WMTYILTFGW
MMTYILTFGW
SMRWIVLVGW
SMRWIVLVGW
ALKWIVTVGW
SLKWIVTVGW | 300
AIYPLGFLLG
AIYPIGYLYG
AIYPIGYLYG
AIYPIGYLYG
AIYPLGYAMG
AIYPLGYAMG
AIYPIGYLYG
AIYPIGYLYG | Q G D
 | 320
 | 330
DNL - N I L YN I
DML - N I V YN I
NML - N I V YN I
NML - N I V YN I
DDM - N I V YN I
DDM - N I I YN L
E LI - N I I YN L
E V I - N I YN L |
| Pop-2 HF10 3D09
Ep-128 MedDCM-JUL2012-C1059
Ep-1 ERN598993-C1654
MedDCM-0C72007-C4997
Pop-3 HF70 19812
Pop-4 HF70 59C08
Ep-128 ERN599983-C478
MedDCM-0CT-524-C2
Ep-1 MedDCM-0CT-534-C62 | 230
I S A I S
- AMD
- AMD
F G D L S
F G D A S
F G D A S
- ML G G S
- MI G G S | 240
D GMVWP L F A V
EMV A W A I
A T I A W A I
A T I A W A I
T E I WW A I
T E I WW A I
S I I WW T L
S I VWW T L

 | 250
GML AWI Y I I Y
GMAAWI Y I I Y
GMAAWI Y I I Y
GMAGWG Y I L Y
GMAGWG Y I L Y
GMA CWF Y I I Y
GMY CWF Y I I Y
 | 260
E I WA G E
E I WA G E
E I WA G E
E I WA G E
E I F R G E
 | 270
- E AK E S V S - G
- D AK A A A N - N
- D AK K A A S - E
- D AK K A A G - D
- D V K E A S A - S
- S I N D A A Q - K
- E A A I L N E N S
- E A A I I N K N S
- E A A I I N K N S | AS E GT Q F A F K
AS E GT Q F A F K
AS E GV Q F A F T
AS E GV Q F A F T
G S D GV Q F A F K
AS E SV Q F A F N
G N E A G Q F A F Q
G N E A G Q F A F Q
 | AMALILTYGW
WMTYILTFGW
WMTYILTFGW
MMTYILTFGW
SMRWIVLVGW
SMRWIVLVGW
SLKWIVTVGW
SLKWIVTYGW | 300
A I Y P L G F L I G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P L G Y MMG
A I Y P L G Y MMG
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G | QGD 310 QGD MDA MDADA MDA MDADA MDA MDADA MDA MDADA LOB NDMI-PG LDE LDETTIFG LEA
 | 320
G G T | 330
DNL - N I L YN I
DML - N I V YN I
NML - N I V YN I
NML - N I V YN I
DDM - N I V YN I
DDM - N I I YN L
E LI - N I I YN L
E V I - N I I YN L
E A I - N I I YN L |
| Pop-2 HF10 3D09
Epi-28 MedDCM-UL2012-CL059
Epi-2 RRS98993-CL654
MedDCM-OCT2007-C4997
Pop-3 HF70 15912
Pop-4 HF70 55038
Epi-28 ERRS9898-C478
MedDCM-OCT-524-C2
Epi-3 RRS98942-C530
MedBCA/GC08 | 2200
I S A I S | 240
D GMVWP L F A V
EMV AW A I
A T I AW A I
A T I AW A I
A T I AW A I
S I VWV A I
S I VWV T L
S I VWW T L
S I VWW T L

 | 250
CML AWI Y II Y
GMA AWI Y II Y
GMA AWI Y II Y
GMA GWG Y IL Y
GMA GWG Y IL Y
GMA CWF Y II Y
GMV CWF Y II Y
GMI CWF Y II Y
GMI AWI Y I Y
 | 260
E I WA G E
E I WA G E
I WA G E
I WY G E
E I F R G
 | 270
- E AK E S V S - G
- D AK A A A A N - N
- D AK K A A S - E
- D AK K A A G - D
- D V K E A S A - S
- S I N D A A Q - K
- E A A I L N E N S
- E A A I I N K N S
- E A A I I N K N S
- E A A I I N K N S
- E A A I I N K N S
- E A A I I N K N S
- E A A I N K N S
- E A | 280
AS E G TQ FAF K
AS E G VQ FAF T
AS E G VQ FAF T
G S D G VQ FAF T
G S D G VQ FAF K
G N E A G Q FAF Q
G N E A G Q FAF Q | 290
AMALILTVGW
WMTYILTFGW
WMTYILTFGW
MMTYILTFGW
AMRAIVTFGW
SLKWIVTVGW
SLKWIVTVGW
SLKWIVTIGW
SLKWIVTIGW
 | 300
A I Y P L G F L G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P L G Y A MG
A I Y P L G Y A MG
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y A I Y P I G | 310
G G D
A G T L G D N
MD A - D A
MD A - D A
NDM1 - S G
NDM1 - S G
L D E - T T I L G
L D E - T T I L G
L E A - T T I L G
L F G D G | 220

 | 30
DNL - NI LYNI
DML - NI VYNI
NML - NI VYNI
DML - NI YYNI
DM- NI YNL
ELI - NI IYNL
ELI - NI YNL
EAI - NI YNL
EAI - NI YNL |
| Pop-2 HF10 3D09
Epi-28 MedDCM-UL2012-C1059
Epi-1 ERK958939-C1654
MedDCM-OCT2007-C4997
Pop-3 HF70 35008
Epi-28 ERK958983-C478
MedDCM-OCT-524-C2
Epi-3 ERK958924-C300
MedBcM2-C30-
Hef130 S1H07 | 2230
I S A I S
- A M D
- F G D L S
F G D L S
F G D A S
M L G G S
- M I G G S
- I I G G S
- L G D | 240
D GMVWP L F A V
EMV AW A I
A T I AW A I
T E I WW A I
S I I WW ML
S I VWW ML
S I VWW T L
S V VWW I V
V V WW I V
V V L GW V I

 | 250
GMLAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMAGWGYIY
GMAGWGYIY
GMACWFYIIY
GMICWFYIIY
GMICWFYIIY
GMICWFYIIY
GMIAWLYIY
 | 260
E I WA G E
E I WA G
 | 270
- E AK E S V S - G
- D AK A A A A - N
- D AK K A A S - E
- D AK K A A S - E
- D X K B A S - S
- S I N D A A Q - K
- E A A I I N K N S
- E A A I I N K N S
- E A A I I N K N S
- E T A K A A A G S
- E T A K A A A G S | 280
A S E G T Q F A F K
A S E G V Q F A F T
A S E G V Q F A F T
A S E G V Q F A F T
A S E G V Q F A F K
A S E S V Q F A F K
G N E A G Q F A F Q
G N E A G Q F A F Q
G N E A G Q F A F Q
G N A A S Q Q A F T
 | 290
AMAL ILTYGW
WMTYILTFGW
WMTYILTFGW
AMRAIVTFGW
AMRAIVTFGW
ALKWIVTVGW
SLKWIVTVGW
SLKWIVTIGW
SLKWIVTIGW
TIKWIVTVGW | 300
A I Y P L G F L G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P L G Y AMG
A I Y P L G Y AMG
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y A WG | 310
GGD
AGTLGDN
MDADA
NDMI-SG
NDML-PG
LDETTILG
LDETTILG
LEATTILG
LEATTILG
YFGDG
YFGGG
 | 320
 | 30
DNL - NI L YN I
DML - NI L YN I
DML - NI V YN I
NML - NI V YN I
DDM - NI I YN I
E LI - NI I YN L
E AI - NI I YN L
E AI - NI I YN L
D A L - NI V YN L
D A L - NI V YN L |
| Pop-2 HED 3D09
Epi-28 MedDCA-UL3012-CL059
Epi-28 MedDCA-UL3012-CL059
MedDCA-OCT2007-C4997
Pop-3 HF0 19512
Pop-4 HF0 59508
Epi-28 ERR59898-C478
MedDCA-OCT-524-C2
Epi-3 ERR59898-C478
MedBCA-OCT-534-C5
Epi-3 MedDCA-OCT-534-C5
Epi-3 MedDCA-OCT-534-C5
Epi-3 MedDCA-0CT-534-C5
Epi-3 MedDCA-0CT-534-C5
Epi-3 MedDCA-0CT-534-C5
Epi-3 MedDCA-0CT-534-C5
Epi-3 MedDCA-0CT-534-C5
Epi-3 MedCA-0CT-534-C5
Epi-3 MedCA-0CT-544-C5
Epi-3 MedCA-0CT-544-C5 | 2230
I S A I S | 240
0 GMVWPLFAV
EMVAWAI
AT AWAI
AT AWAI
SI WWAI
SI WWTL
SI VWWTL
SI VWWTL
VVWWIV
VVWWVI
VVLGW-VI
VGPAFFV
 | 250
GMLAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMAGWGYILY
GMAGWGYILY
GMCWFYIIY
GMICWFYIIY
GMICWFYIIY
GMIAWLYIY
GMIAWLYIY
GMIAWLYIY
GATAWIYIY
 | 260
E I WA G E
E I WA G E
E I WA G E
E I WA G E
E I F R G

 | 270
- E AK E S V S - G
- D AK A A A N - N
- D AK K A A S - E
- D AK K A A S - E
- D AK K A A S - E
- D V K E A S - S
- S I N D A A Q - K
- E A A I L N K N S
- E A A I I N K N S
- E A A I I N K N S
- E T A K A N A G S
- E T A K A N A G S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I N E N S
- E A S Q I | 280
A S E G T Q F A F K
A S E G V Q F A F T
A S E G V Q F A F T
A S E G V Q F A F T
A S E G V Q F A F K
G N E A G Q F A F Q
G N E A G Q F A F Q
G N E A G Q F A F Q
G N E A G Q F A F Q
G N A A S Q Q A F T
G N A A S Q Q A F T | 290
AMAL ILTYGW
WMTYILTFGW
WMTYILTFGW
WMTYILTFGW
SMRWIVLYGW
SLKWIVTYGW
SLKWIVTYGW
SLKWIVTIGW
TIKWIVTYGW
TIKKIVTYGW
TIKLIYTYGW
TIRLIYTYGW
 | 300
A I Y P L G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P L G Y M MG
A I Y P L G Y M MG
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y P I G Y L Y G
A I Y | 310
QGD | 220
G G T

 | 30
DNL - NI L YNI
DML - NI VYNI
NML - NI VYNI
NML - NI VYNI
NMM - NI VYNI
DDM - NI I YNL
E LI - NI I YNL
E AI - NI YNL
D AL - NI VYNL
D AL - NI VYNL
SL - NI VYNL |
| Pop-2 HF10 3D09
Epi-28 MedDCM-UL2012-C1059
Epi-1 ERNS9893-C1654
MedDCM-OCT007-C4997
Pop-3 HF10 5020
Epi-28 ERNS9838-C478
MedDCM-OCT-524-C2
Epi-3 ERNS98342-C530
MedBCM-OCT-534-C5
Epi-1 MedDCM-OCT-534-C5
Epi-1 MedDCM-OCT-534-C5
Barbar EBAC2050
HF10 4500
gamma EBAC2050 (B)
gamma HcT2M-(B) | 230 I S A I S - AMD - FG D AS - MI G - K AD - L AP | 2440
D GMVWPLFAV
EMVAWAI
ATIAWAI
ATIAWAI
TEIWWAI
SIVWAI
SIVWAI
SIVWTL
SIVWIL
SIVWIL
SIVWIL
VVWIV
VTLGWVI
VTLGWFV
AMVCG-VI
 | 22500
GMLAWIYIIY
GMAAWIYIY
GMAAWIYIY
GMAGWGYILY
GMAGWGYILY
GMAGWFYIIY
GMCCFYIIY
GMCCFYIIY
GMICWFYIIY
GMIAWLYIY
GMIAWLYIY
GMIAWLYIY
GMAGWIYIY
ATLAWLYIY

 | 260
E I WA G
 | 2020
- DAKAAAN - N
- DAKKAAN - N
- DAKKAAN - N
- DAKKAAS - E
- DAKKAAS - E
- DAKKAAS - E
- AAI INKNS
- EAAI INKNS
- EAAI INKNS
- EAAI INKNS
- EAAI INKNS
- ETAKAAS
- ETAKAAS
- EASI INKNS
- EASI VANAS
- EASI VANAS | 280
A S E G T Q F A F K
A S E G V Q F A F T
A S E G V Q F A F T
A S E G V Q F A F T
G N E A G Q F A F Q
G N E A G Q F A F Q
G N E A G Q F A F Q
G N E A G Q F A F Q
G N E A G Q A F N
G N A A S Q Q A F N
G N A A S Q Q A F N
G N A A S Q Q A F N
G N A A S Q Q A F N
G N A A S Q Q A F N
G N A A S Q Q A F N
G N A A S S A F N
S S S A S N A S N A S N
S S S A S N S A S N S A S N
S S S A S N S A S N S A S N
S S S S S S S S S S S S S S S S S S | 230
MMAL I L T V G I
MMAT Y I L T F G W
WMT Y I L T F G W
WMT Y I L T F G W
WMT Y I L T F G W
MMT Y I L T F G W
MMT Y I L T F G W
MMT Y I L T F G W
S L K W I V T V G W
S L K W I V T V G W
S L K W I V T I G W
S L K W I V T I G W
T I K W I V T V G W
T I K W I V T V G W
T I K A I V T F G W
T I K A I V T F G W
T I K A I V T F G W
T I K A I V T F G W
T I K A I V T F G W
T I K A I V T F G W
T I K A I V T V G W
 | 300
A I YP LG F I LG
A I YP I G YL YG
A I YP LG YAMG
A I YP LG YAMG
A I YP I G YL YG
A I YP I G YL YG
A I YP I G YL YG
A I YP LG YAMG
A I YP LG YALA
A I YP I G YYLA
A I YP I G YYLA | 310 QGD AGTLGDN AGTLGDN MDA DA MDA MDA MDA DA MDA MO MG MG <t< th=""><th>320
G G T D G Q</th><th>330
DNL - N I V YN I
DML - N I V YN I
NML - N I V YN I
NML - N I V YN I
DM- N I YN I
DM- N I YN I
E I - N I YN L
E A I - N I YN L
S L - N I YN L
S L - N I YN L</th></t<> | 320
G G T D G Q
 | 330
DNL - N I V YN I
DML - N I V YN I
NML - N I V YN I
NML - N I V YN I
DM- N I YN I
DM- N I YN I
E I - N I YN L
E A I - N I YN L
S L - N I YN L
S L - N I YN L |
| Pop-2 HF10 3D09
Epi-28 MedDCM-UL2012-C1059
Epi-1 ERK958959-C1654
MedDCM-OC2007-C4997
Pop-3 HF70 15912
Epi-28 ERK958983-C478
MedDCM-OCT-534-C4
Epi-28 ERK958983-C478
MedDCM-OCT-534-C4
Epi-28 MedDCM-OCT-534-C4
Epi-28 MedDCM-OCT-544-C4
Epi-28 MedCM-O | 220
I S A I S | 260
D GM VWP LF AV
EMV AW
 | 250
GMAAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMACWIYIIY
GMACWIYIIY
GMACWIYIIY
GMACWIYIIY
GMICWFYIIY
GMICWFYIIY
GMIAWLYIY
GMIAWLYIY
GMIAWLYIY
ATLAWLYIY
GCAWYWIY
 | 260
E I WA G

 | 270
- EAK E SV S. G
DAKAAAA.
- DAKKAAAS.
- DAKKAAAS.
- DAKKAAAS.
- DAKKAAAS.
- DVKEASA.
- SINDAAQ.
- KAAAAAS.
- EAAI INKNS
- EAAI INKNS
- EAAI INKNS
- ETAKAANAGS.
- ETAKAANAGS.
- ETAKAANAGS.
- EGAGAKMAC | 280
A SE EGT Q F AFF
A SE EGT Q F AFF
A SE EGV Q F AFF
A SE EGV Q F AFF
A SE EGV Q F AFF
A SE SV Q F AFF
A SE SV Q F AFF
G N F A GQ F AFQ
G N F A GQ F AFQ
G N F A GQ F AFQ
G N A A SQ Q AFF
G N A A SQ A FT
G N A SQ A FT
G | 280
A MA L I L T V GW
WM T Y I L T F GW
S M R I V I V GW
S M R I V I V GW
S L KW I V T GW
S L KW I V T GW
S L KW I V T GW
T I KW I V T GW
T I KW I V T GW
T I K I V T GW
T I K I V T GW
T I R I V T GW
T M I V GW
T M I V I GW
T M I V I GW
T M I V GW
T M I V GW
T M I V I U V GW
T M I V I V GW
T M I V I U GW
T M I V GW
T M I V I U GW
T M I V GW
T M I V GW
T M I V I I V GW
T M I V G | 300
A 11YP L G 1 L G
A 11YP L G 1 L G
A 11YP I G Y L YG
A 11YP I G Y L YG
A 11YP L G Y L YG
A 11YP L G Y L YG
A 11YP I G Y L G
A 11YP I G Y L G
A 11YP I G Y L G
A 11YP A G Y L G Y L G
A 11YP A G Y L G Y L G
A 11YP A G Y L G | 310
QGD
 | 220
G G T | 330
 |
| Pop-2 HEID 3D09
Ep-28 MedDCA-HL2012-CL059
Ep-18 RM30993-CL564
MedDCA-OCT2007-C4997
Pop-3 HF70 19912
Pop-3 HF70 19912
Pop-3 HF70 19912
Ep-128 RH309398-C478
MedDCA-OCT-324-C2
Ep-13 ERH309382-C330
MedBA-OCT-324-C2
Ep-13 MedDCA-OCT-324-C2
Ep-13 MedDCA-OCT-324-C3
Ep-13 MedDCA-OCT-324-C3
Ep-13 MedDCA-OCT-324-C2
Ep-13 MedDCA-OCT-324-C3
Ep-13 MedDCA-0CT-324-C3
Ep-13 MedDCA-0CT-324-C3
Ep-13 MedDCA-0CT-324-C3
Ep-13 MedDCA-0CT-324-C3
Ep-13 MedDCA-0CT-324-C3
Ep-13 MedDCA-0CT-34-C3
Ep-13 MedDCA-0CT-34-C3 | 210
I S A I S
ANN
AND
F G D I S
F G D I S
M I G
G
I G G
I G G
G
I G G
I G G S
L I G
G
I M G
I M M
I M M M
I M M M M M M M M M M M M M M M M M M M | 242 D GM VWP LF AV EM VA W - F A V EM VA W - F A V AT I AW A I T I AW A I S I VWW A I S I VWW - T L S I VWW - T L V V UW - I A M V G - V V GP A F F V V L PA F I I AMP A F V | 2030 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAGWIYIIY GMACWIYIIY GMACWIYIIY GMCWFYIIY GMICWFYIIY GMIAWUYIY GMAWIYIY GMAWIY GMAWIY GMAWIY GMAWIY GMIY GMIY GMIY GMIY GMIY GMIY GMIY GMIY </th <th>200
E I WAG</th> <th>1270
1 = A KE SV S - G
D AK KAAA - N
D AK KAAA - D
D AK KAAA - D
D V K E ASA - S
S I ND AA C - K
E AA I L NENS
E AA I L NENS
E AA I I NENS
E AA I I NENS
E TAAANAGS
E AS G I NENS
E G AK LNASS
E G AG AKNAS
E AG GAKNAS</th> <th>LASE GU QA FAFT
ASE GU QAFT
ASE GU QAFT
ASE GU QAFT
GS DGU QAFT
GS DGU QAFT
GN EAG QAFT
GN FAG QAFT
GN FAG QAFT
GN FAG QAFT
GN FAG QAFT
GN</th> <th>280
AMA LILTY GW
WMTYILTF GW
WMTYILTF GW
WMTYILTF GW
AMR AI VT GW
AMR AI VT GW
SLKWIVTY GW
SLKWIVTY GW
SLKWIVT GW
TIKAIVTF GW
TIKAIVTF GW
AMRMIIVT GW
AMRMIIVT GW</th> <th>300
A 1 Y PL G Y L Y G
A 1 Y PL G Y M MG
A 1 Y PL G Y M MG
A 1 Y PL G Y L Y G
A 1</th> <th>310 QGD MDA -DA MDA -DA MDA -DA MDA -DA MDA -DA NDMI -SG LDE -TTILG LDE -TTILG YFG -GG YFG -GG YLM -GG YLM -GG YLM -GG </th> <th>320
6 GT
6 GT
7 GT
9 GT
9</th> <th>330
DNL N I LYNI
DML N I LYNI
DML N I YNNI
DML N I YNNI
DML N I YNNI
DMM N I YNNI
E LI N I YNL
E LI N I YNL
E LI N I YNL
E AI - N I YNL
B AL - N I YNL
B AL - N I YNL
S NL - N I YNL
S NL - N I YNL
G AL N I YNL
G AL N I YNL
G AL N I YNL</th> | 200
E I WAG | 1270
1 = A KE SV S - G
D AK KAAA - N
D AK KAAA - D
D AK KAAA - D
D V K E ASA - S
S I ND AA C - K
E AA I L NENS
E AA I L NENS
E AA I I NENS
E AA I I NENS
E TAAANAGS
E AS G I NENS
E G AK LNASS
E G AG AKNAS
E AG GAKNAS | LASE GU QA FAFT
ASE GU QAFT
ASE GU QAFT
ASE GU QAFT
GS DGU QAFT
GS DGU QAFT
GN EAG QAFT
GN FAG QAFT
GN FAG QAFT
GN FAG QAFT
GN FAG QAFT
GN | 280
AMA LILTY GW
WMTYILTF GW
WMTYILTF GW
WMTYILTF GW
AMR AI VT GW
AMR AI VT GW
SLKWIVTY GW
SLKWIVTY GW
SLKWIVT GW
TIKAIVTF GW
TIKAIVTF GW
AMRMIIVT GW
AMRMIIVT GW | 300
A 1 Y PL G Y L Y G
A 1 Y PL G Y M MG
A 1 Y PL G Y M MG
A 1 Y PL G Y L Y G
A 1 | 310 QGD MDA -DA MDA -DA MDA -DA MDA -DA MDA -DA NDMI -SG LDE -TTILG LDE -TTILG YFG -GG YFG -GG YLM -GG YLM -GG YLM -GG | 320
6 GT
6 GT
7 GT
9 | 330
DNL N I LYNI
DML N I LYNI
DML N I YNNI
DML N I YNNI
DML N I YNNI
DMM N I YNNI
E LI N I YNL
E LI N I YNL
E LI N I YNL
E AI - N I YNL
B AL - N I YNL
B AL - N I YNL
S NL - N I YNL
S NL - N I YNL
G AL N I YNL
G AL N I YNL
G AL N I YNL |
| Pop-2 HF10 3D09
Epi-28 MedDCM-JUL2012-C1059
Epi-1 RMS90939-C1624
Pop-4 HF70 59038
Epi-28 RMS90938-C478
Pop-4 HF70 59038
Epi-28 RMS90383-C478
MedDCM-OCT-524-C2
Epi-28 RMS90383-C478
MedBCM-OCT-524-C2
Epi-1 MedDCM-OCT-524-C2
Epi-1 MedDCM-0CT-524-C2
Epi-1 MedDCM-0CT-524-C2 | 220
I S A I S | 222
D GMVWP F A V
EMV XW - F A V
EMV XW - F A V
A T I AW A I
A T I AW A I
T I AW A I
T E I WW - A
S I VWW - T L
S I VWW - T L
A WY A A - I
A WP A F - I
A W A F - I
A T G - V I
A T L G - V I
A T L G - V I
A T L G - V I
 | 250
GMLAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMACWIYIIY
GMACWIYIIY
GMACWIYIIY
GMACWIYIIY
GMICWFYIIY
GMICWFYIIY
GMIAWIYIY
GMIAWIYIY
GAIAWIYIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GMACWYYIY
MACWYYIY
 | 200
E I WA G

 | 2020
1 = A K K A S V S - G
D A K A A A M - N
D A K K A A S - D
D A K K A A S - D
D X K A A S - D
D X K A A S - D
S - I M D A G K
S - I M D A G K
E A A I N N N
E A S I N K N
E A A I N N N
E A S I N K N
E A A I N N N
E A A I N N
E A I N N
E A I N N
E A I N N
E A A I N N
E A A I N N
E A I N N
E A A I N N
E A I N N
E A | ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GAFA
GN AASQ QAFK
GN AASQ AAFK
GN AAFK
GN AAFK
GN | 1990
AMALILTYGW
WMTYILTFGW
WMTYILTFGW
AMRAIVTGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTGW
SLKWIVTGW
TIKUVTGW
TIKUVTGW
TIKUVTGW
TIKUVTGW
AMMMILVTGW
AMMMILUGW | 340
A 1 Y PL G Y L Y G
A 1 Y PL G Y L Y G Y F M G
A 1 Y PL G Y L Y G Y F M G Y L Y G Y F M G Y L Y G Y F M G Y L Y G Y F M G Y L Y G Y F M G Y | 310 QGD AGT (GDA) AGT (GDA) MDA - DA LDE - TTILG LDE - TTILG YFG - GG
 | 220
G G T | 330
DNL - N I U YN I
DNL - N I U YN I
DML - N I U YN I
DML - N I U YN I
DDM - N I YN I
E U - N I I YN L
E A I - N I YN L
A NA - N I YN L
N SL - N I YN L
L N L YN L
E N L YN L
SN L - N L YN L
L N L YN L
E N L YN L
DML - N U YN L
SC L - N U YN L |
| Pop-2 HF10 3D09
Epi-28 MedDCM-UL2012-C1059
Epi-1 ERK930939-C1654
MedDCM-OCT2007-C4997
Pop-3 HF70 19812
Pop-3 HF70 19812
Pop-3 HF70 19812
Epi-3 HedDCM-OCT-324-C2
Epi-3 HeR050M-OCT-324-C2
Epi-3 HeR050M-OCT-334-C5
Epi-3 HeR050M-OCT-334-C5
Epi-3 HeR050M-OCT-334-C5
Epi-3 HeR050M-OCT-334-C5
Epi-3 HeR050M-OCT-334-C5
Epi-3 HeR050M-OCT-334-C5
Epi-3 HeR050M-OCT-334-C5
Epi-3 HeR050M-OCT-334-C5
Epi-3 HeR050M-OCT-34-C5
Epi-3 HeR050M-OCT-34-C5
Epi- | 210 I S A I S A NN A NN D A NN D A NN D F G D I S F G D I S F G D I S F G M I G G S I G K N D C S I G Y A D G N L A P I MA Y I N Y | 240
D GM VWP LF AV
EMV XAV A I
EMV XAV A I
AT I AV A I
T E I WW A I
S I VWW T L
S VWW T L
S VWW T L
V VW A I
S VWW T L
V VW A I
S VWW T L
S VWW T L
AWP A F I I
T T L G F V I
AKMA AF I I
AKMA AF I V
 | 2010
GMLAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMAGWIYIIY
GMAGWIYIIY
GMAGWIYIIY
GMAGWIYIIY
GMICWFYIIY
GMICWFYIIY
GMIAWIYIY
GMIAWIYIY
GMAGWIYIY
GCLAWYWIY
GCLAWYWIY
GCLAWYYIY
SMAGWFYILY
SMAGWFYILY
 | 200
E I WAG

 | 2020
 | ASE GTQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GS DGYQ FAFT
GS DGYQ FAFT
GN FAG GFAFQ
GN FAG GFAFG
GN FAG GGAFG
GN FAG GAFG
GN FAG GAG GAFG
GN FAG GAG GAFG
GN FAG GAG GAG GAG GAG GAG GAG GAG GAG
GN FAG GAG GAG GAG GAG GAG GAG GAG GAG GAG | 280
AMA LI LI V GW
WMT Y I LI F GW
WMT Y I LI F GW
WMT Y I LI F GW
AMR I V T GW
AMR I V T GW
S LKWI V T GW
S LKWI V T GW
T I KI V T GW
MMM I I V GW
AMMM I V T GW
AMR I V T GW
T MMY I I L GW
AMR I V T GW
T MMY I I L GW
AMR I V T GW
T MMY I V GW
T MY V T GW | 340
A I YP IG YL IG
A I YP IG YL YG
A I YP IG YL G
A I Y I Y IG YL G
A I Y I Y I Y I Y I Y I Y I Y I Y I Y I | 310 QGD MDA DA MDA DA
 MDA DA NDMI PG LDE -TTI LG LDE -TTI LG VFG OG YFG OG YMG GG YLM GG YLM GG YLM GG YLA GG YHT GG | 1200
0 G T | 330
DNL - N I L YN I
DNL - N I U YN I
DML - N I U YN I
DML - N I U YN I
DDM - N I YN I
DDM - N I YN I
DDM - N I YN I
E L I - N I YN L
E L I - N I YN L
DA L - N I YN L
DA L - N I YN L
N I - N L YN L
SS L - N U YN L
YML - N U YN L
SS L - N U YN L
YML - N U YN L
SS L - N U YN L
YML - N U YN L
SS L - N U YN L
YML - N U YN L
 |
| Pp.2 HEB 3D09
Ep.28 MedDCA-UL2012-CL059
Ep.28 MedDCA-UL2012-CL059
Pp.3 HF0 3959.CL654
MedDCA-OCT207-C4997
Pp.3 HF0 3950.2
Ep.28 ERR39898-C478
MedDCA-OCT-324-C2
Ep.13 ERR39898-C478
MedDCA-OCT-324-C2
Ep.13 MedDCA-OCT-324-C2
Ep.13 MedDCA-OCT-324-C3
Ep.13 MedDCA-UL2020
MedDCA-UL2020T/G (0)
D.donghaends MeD134 (6)
MedDCA-UL2021-C3793 | 220
1 S A S | 2000
D GM VWP LF A V
EMV X AV - A V
EMV X AV - A V
AT I AV A I
T E I WV A I
T E I WV A I
S I VWW - T L
S I VWW - T L
S I VWW - T L
V V WW - I
V V V - A I
A T A A V - A I
T C I G - V I
T L G - V I
A T G - V I
A T G - V I
A W
 | 253 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWIYIIY GMACWIYIIY GMACWIYIIY GMACWIYIIY GMICWFYIIY GMICWFYIIY GMACWIYIIY GMACWIYIIY GMACWIYIIY GMACWIYIIY GMACWIYIIY GMACWIYIIY GMACWIYIIY GMACWIYIIY GMACWYIYIY GMACWYIYIY GMACWYYIY SAAWFYIWY SGAAWFYIYIY SGAYFYIY SGAYFYIY
 | 200
E I WAG

 | 1270
1 = A KE SV S - G
D AK KAAA M - N
D AK KAAA M - N
D AK KAAA - D
D V KE ASA - S
S I ND AAA - K
E AA I L NE NS
E AA I L NE NS
E AA I L NE NS
E TAAANAG S
E AS Q I NE NS
E TAAANAG S
E AG K LANAS
E GAG KAASA
E AG KAAASA
E AG KAAASA
E AG C NAAKAS
E AG C NAASA
E AG C NAASA | LASE GTQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN EAG QFAFT
GN EAG QAFT
GN EAG QFAFT
GN EAG QAFT
GN FAG | 230
AMALILTYGW
WMTYILTFGW
WMTYILTFGW
MMTYILTFGW
AMRAIVTFGW
AMRAIVTGW
SLKWIVTYGW
SLKWIVTGW
SLKWIVTGW
TIKUVTYGW
TIKUVTYGW
TIKUTYGW
AMRMIVTGW
MTHLIFGW
TIKUTYGW
AMRWIVTGW
TIKUTYGW
AMRWIVTGW
TIKUTYGW
AMRWIVTGW
TIKUTYGW
AMRWIVTGW
TIKUTYGW
AMRWITGW
TIKUTYGW
AMRWITGW
TIKUTYGW
AMRWITGW
TIKUTYGW
AMRWITGW
TIKUTYGW
AMRWITGW
TIKUTYGW | 300
A 1 Y PL G F I LG
A 1 Y PL G Y L YG
A 1 Y PL G Y L G Y LG
A 1 Y PL G Y L G Y LG
A 1 Y PL G Y LG | 310 QGD - - - - - - - - - - - - - - - - - MDA - - - MDA - A - MDA - A - MDA - A - MDA - A MDM - A - MDA - A MDM - MDA - A MDM - MDM - A MDM - MDM - MDM - A - MDM - A - - - - T ME G -
 | 320
C G T | 330
D.N.L. N.I. Y.N.I
D.M.L. N.I Y.N.I
D.M.L. N.I Y.N.I
D.M.J. N.I Y.N.I
D.M.N.I Y.N.I
D.M.N.I Y.N.I
E.I. N.I Y.N.I
E.I. N.I Y.N.I
E.I. N.I Y.N.I
E.I. N.I Y.N.I
E.I. N.I Y.N.I
S.I. N.I Y.N.I Y.N.I
S.I. N.I Y.N.I Y.N.I
S.I. N.I Y.N.I Y.N.I
S.I. N.I Y.N.I Y.N.I Y.N.I
S.I. N.I Y.N.I Y.N.I Y.N.I Y.N.I Y.N |
| Pop-2 HED 3D09
Epi-28 MedDCM-UL2012-CL059
Epi-28 MedDCM-UL2012-CL059
Epi-28 MedDCM-UC207-C897
Pop-3 HF70 19702
Pop-3 HF70 19702
Pop-3 HF70 19702
Epi-28 ERK959898-C478
MedDCM-OCT-524-C2
Epi-2 HeRK95989-C753
MedDCM-OCT-524-C2
Epi-3 MedDCM-OCT-534-C2
Epi-3 MedDCM-OCT-534-C2
Epi-3 MedDCM-OCT-534-C2
Epi-3 MedDCM-OCT-534-C2
Barma EAC-276-06
(B)
Palagibacter sp. HFCC211(B)
gamma EAC-2767 (G)
G-palitiquia (B)
D-donghaensis MED134 (G)
MedDCM-OCI207-C1678 | 220
I S A I S
A AIND
A AIND
F G D A S
F G D A S
F G D A S
M I GG S

M I GG S

I I GG S

I A G
Y A D
Y A D
Y N D
Y I N
Y I N
Y I N
Y I N
Y N D
Y N D | 26 D GM VWP LF AV D GM VWP LF AV EM V AW A I AT I AW A I AT I AW A I T I SI I WW A I S I VWW T L S I VWW T L V VWW T L V VWW T L V V WW T L AMP A G F V AMP A F I I AMP A F I I ATL AG F V I ATL AG G V I
 | 253 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWFYIY GMACWFYIY GMICWFYIY GMICWFYIY GMIAWIYIY GMIAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GMAWIY GMAWIY GMAWIY GMAWIY GMAWIY GMAWIY GMAWIY GMAWIY GMAWIY GOLAWFYIIW GOLWAHIM GOLWAHIM GOLWAHIM GOLWAHIM GOLWAHIM
 | 200
E I WA G

 | 2020
1 = A K K S V S - G
D A K A AA A - N
D A K K A A S - D
D A K K A A S - D
S - A I I N K N S
E A A I I N K N S
E A A I I N K N S
E A A I I N K N S
E T A K A M A A S
E T A K A M A S
E T A K A M A S
E T A K A M A S
E A G K K A A K S
E A G K A A A K S
E A G K A A A K S
E A G K A A A K S
E A S C I N A S
E A G K A A A K S
E A S C I N A S
E A G K A A A K S
E A S C I N A S
E A S C A S C A S C A S
E A S C A S C A S C A S
E A S C A S C A S C A S C A S
E A S C | ASE GTQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GS DG YQ FAFT
GS DG YQ FAFT
GS DG YQ FAFT
GN FA GQ FAFQ
GN FA GQ FAFQ
GN FA GQ FAFQ
GN FA GQ FAFQ
GN AAS QQ FF
GN AAS QU FF
GN AAS AAS AAS AAS AAS A | 1233
AMALILTYGW
WMTYILTFGW
WMTYILTFGW
AMRAIVTFGW
AMRAIVTFGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKUIVTGW
TIKWIVTGW
TIKWIVTGW
MMMIIVTGW
AMMMIVTGV
AMMMIVTGU
AMRMIVTGW
AMRAIVTGU
AMRAIVTGU
AMRAIVTGU
AMRAIVTGU
AMRAIVTGU
AMRAIVTGU
AMRAIVTGU
AMRAIVTGU
AMRAIVTGU
AMRAIVTGU
AMRAIVTGU | 100
A I YP IG YL YG
A I YF IG YL YG
A | 310 GG D MDADA MDADA MDADA MDADA MDADA NDMIPG LDETTILG LEATTILG LEATTILG LPGTTILG VFGGG YFGGG YLMGG YLM YLM YLM
 | 220
 | 320 DN.L. N.I. VYN I DML - N. V. VYN I NML - N.I. VYN I DDM.N.I. VYN I DDM.N.I. YNL DAL - N.I. YNL E.I. N.I. YNL DAL - N.I. YNL DS.I. N.I. YNL SN.N.I. YNL SN.N.I. YNL SS.I. YNL YNL SS.I. N.V. YNL SS.I. N.V. YNL SS.I. N.V. YNL SS.I. N.V.YNL SS.I. N.I.YNL SS.I. N.I.YNL SS.I. N.I.YNL SS.I. N.I.YNL SS.I. N.I.YNL SS.I. N.V.YNL SS.I. N.Y.YNL |
| Po-2 HF10 3D09
Ep-12 MedDCM-UL2012-C1059
Ep-14 MedDCM-UL2012-C1059
Ep-14 MedDCM-OCT007-C4997
Po-p-14/F0 30008
Ep-128 Ent893983-C478
MedDCM-OCT-324-C2
Ep-13 MedDCM-OCT-334-C2
Ep-13 MedDCM-OCT-334-C2
Ep-13 MedDCM-OCT-334-C2
Ep-13 MedDCM-OCT-334-C2
Barma HeX20508
HF10 45008
garma BEAC20509 (B)
garma BEAC20509 (B)
garma BEAC20509 (B)
garma HCC0504 (G)
Publique HTCC1062 (G)
Publique HTCC1062 (G)
Publique HTCC1062 (G)
Publique HTCC1062 (G)
Publique HTCC1062 (G)
Publique HTCC1062 (G)
MedDCM-OCT-2526-C80 | 220
1 S A 1 S | 2020
D GM VWP F F A V
EM V AW A I
EM V AW A I
AT I AW A I
T I AW A I
T E I WW A I
T E I WW A I
S I VW A I
S I VW A I
S V V V A I
S V V V A I
S V V V A I
A T A A I
A I
 | 253 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWIYIYIY GMACWIYIYIY GMACWIYIYIY GMACWIYIYIY GMICWFYIYIY GMICWFYIYIY GMIAWIYIYIY GMAWIYIYIY GMAWIYIYIY GMAWIYIYIY GMACWIYIYIY GCLAWYMIY GCLAWYMIY GCLAWYMIY GGAY GGYGWAMIMI- GGYGWAMII-

 | 200
E I WAG
 | 1270
1 = A KE SVS - G
D AK KAAA M - N
D AK KAAA - D
D AK KAAA - D
D AK KAAA - D
D XK KAAA - D
D XK KAAA - D
S I NDAAQ - K
E AA I L NENS
S I NDAAQ - K
E AA I L NENS
E AA A A A A A
E AA I L NENS
E AA A A A A A A
E A A I NENS
E A A A A A A A S
E A A A A A A A A S
E A A A A A A A A S
E A A A A A A A A S
E A A A A A A A A A A A A A A A A A A A | ASE GTQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN FA GQ FAFQ
GN FA GQ FAFQ
GN FA GQ FAFQ
GN FA GQ FAFQ
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AAXQ TAYN
GN KAL YTAFG
GN YN | 200
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
MMTYILIFGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTVGW
SLKWIVTVGW
SLKWIVTVGW
TIKIVTYGW
TIKIVTYGW
TIKIVTYGW
AMMMILVYGW
AMMMILVYGW
AMMMIVTVGW
AMRWIVTVGW
AMRWIVTVGW
AMRWIVTVGW
AMRWIVTVGW
AMRWIVTVGW
ALRTFILVGW
ALRWFIVVGW | 200
A 1 YP LG YL LG
A 1 YP LG YL YG
A | 310 QGD
 | 1200
G G T | 330
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DML - N U YN
DMM - N U YN
E L N U YN
S L - N U YN |
| Pop-2 HEID 3D09
Ep.28 MedDCA-UL2012-CL059
Ep.28 MedDCA-UL2012-CL059
Pop-3 HF0 2095-CL654
MedDCA-OCT2007-C697
Pop-3 HF0 2098-C478
MedDCA-OCT-324-C2
Ep.12 ERR593983-C478
MedDCA-OCT-324-C2
Ep.13 MedDCA-OCT-324-C2
Ep.13 MedDCA-OCT-324-C3
Ep.13 MedDCA-OCT-324-C3
Ep.13 MedDCA-OCT-324-C3
Ep.13 MedDCA-OCT-324-C3
Ep.13 MedDCA-OCT-324-C3
Samma EBAC20509 (b)
gamma Hc1 75m4 (b)
gamma Hc1 75m4 (b)
gamma Hc1 75m4 (b)
gamma HC120207 (c)
Pelagibacter sp. HTCC7211(b)
gamma HC120207 (c)
D.donghaensis MED134 (c)
MedDCA-UC2007-C1678
MedDCA-OCT-0207-C1678
Pop MGII GG3 | 220
1 S A S
A MIO
F G D A S
F G D A S
F G D A S
MI GG S

MI GG S

MI GG S

I G
G A N

G A N

G A N

G N

G N

G N

G N

 | 240
D GM VWP F E A V
BM V AW - F A I
AT I AW A I
S I VWW A I
AW A A - A I
AW A A I
AM V G V I
AT I G I
AW A I
A I
AW A I
A I
A I
A
 | 253 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY
 GMAAWIYIIY GMAAWIYIIY GMAGWIYIIY GMAGWIYIIY GMACWIYIY GMACWIYIY GMACWIYIY GMICWFYIY GMICWFYIY GMAAWIYIY GMAAWIYIY GMAWIYIY GMAWIYIY GCLAWYMIY GCLAWYMIY SGAAWYY SGAWFY GC LAWYMIY SGAWFY GC CWAMIMI GC VGWAMIMI GC VGWAMII GC GWAMII GC GMWAMIY
 | 200
E I WAG
 | 2011
2 A KE SV S - G
D A KA AAA M - N
D A KK AAA S - D
D A KK AAA S - D
D V KE ASA - S
S I N D AA Q - K
E AA I L NE NS
E AA I NE NS
E AA G A AA A S - T
E GA Q A AN S - T
E GA Q A C S - T
E GA Q A S - T
S - A KK LA V - R
D G G C V DN S K
S G G C V DN S K | LASE GTQ FAFK
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFK
GN EAG QF AFG
GN EAG QA FT
GN | 230
AMA LI LI V GW
WMT Y I LI F GW
WMT Y I LI F GW
WMT Y I LI F GW
AMR I V T GW
AMR I V T GW
S LKW I Y T GW
S LKW I Y T GW
T I KW I Y T GW
ALK Y I L Y GW
AMR W Y T GW
T I KW I Y T GW
T I KW I Y T GW
T I KW I Y T GW
A K F I L Y GW
ALR Y F I L Y GW
ALR WF I Y GW | 300
A I YP I GY L YG
A I YP L GY L YG
A I YP L GY L YG
A I YP I GY L YG
A I YP I GY L YG
A I YP I GY L YG
A I YP L GY L GY L GY
A I YP L GY L GY L GY L GY
A I YP L GY L
 | 310 QGD - <th>320
G G T)
F D T)
F D T)
G G Q
F G G
F G G
F G G
V N I S N TA Y S V
V T I S N TA Y S V
V T I S N TA Y S V
V T I S N TA Y S V
C L N E
V D A
C G Q C
C Q P
A DM
C Q P
C Q D
C Q D</th> <th>330
DNL - N I U YN I
DML - N I U YN I
NML - N I U YN I
DML - N I U YN I
DMM - N I YN I
E U I - N I I YN L
E U I - N I I YN L
E AI - N I YN L
E AI - N I YN L
E AI - N I YN L
S I - N I YN L
S I - N I YN L
E I - N U YN L
S I - N U YN L YN L
S I - N U YN L YN L
S I - N U YN L YN L YN L YN L
S I - N U YN L YN L YN L YN L YN L YN L YN L</th> | 320
G G T)
F D T)
F D T)
G G Q
F G G
F G G
F G G
V N I S N TA Y S V
V T I S N TA Y S V
V T I S N TA Y S V
V T I S N TA Y S V
C L N E
V D A
C G Q C
C Q P
A DM
C Q P
C Q D
C Q D | 330
DNL - N I U YN I
DML - N I U YN I
NML - N I U YN I
DML - N I U YN I
DMM - N I YN I
E U I - N I I YN L
E U I - N I I YN L
E AI - N I YN L
E AI - N I YN L
E AI - N I YN L
S I - N I YN L
S I - N I YN L
E I - N U YN L
S I - N U YN L YN L
S I - N U YN L YN L
S I - N U YN L YN L YN L YN L
S I - N U YN L |
| Pop-2 HED 3D09
Epi-28 MedDCM-UL2012-CL059
Epi-28 MedDCM-UL2012-CL059
Epi-28 MedDCM-UC2007-C697
Pop-4 HFO 195028
Pop-4 HFO 195028
Epi-28 ERK959898-CA78
MedDCM-OCT-524-C2
Epi-28 ERK95998-CA78
MedBCM-OCT-524-C2
Epi-28 MERS9598-CA78
MedBCM-OCT-524-C2
Epi-28 MERS9598-CA78
MedBCM-OCT-524-C2
Epi-28 MERS9598-CA78
MedBCM-OCT-524-C2
Epi-28 MERS9598-CA78
MedBCM-OCT-524-C2
Gamma EAC2-526 (b)
Palagibacter sp. HFCC211(b)
gamma EAC207 (c)
G-palificial (b)
D-donghaenis MED124 (c)
D-donghaenis MED124 (c)
MedDCM-OCT-207 Clo57
Pop-1 MedDCM-OCT-594-C2
E, sibiricum 255-15
E-giaguebacterium sp. AT15 | 220
1 S A 1 S | 262 D GM VWP LF AV D GM VWP LF AV EM V AW A I AT I AW A I AT I AW A I AT I AW A I T I S I VWW I I VV WW T I VV WW I V VU WA I I MAM G V I AMP A I I AMP A I I TL G F V I AT L G F V I AT L G F V I AW L W L I AM C G V I AM C G V I AM A G I V L AM C G V L A G G W A L L L T G G L G W S L L L T G G G W G L L L T G G G W M C L L
 | 253 GMLAWIYIIY GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWFYIY GMACWFYIY
GMICWFYIY GMICWFYIY GMIAWIYIY GMIAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GMAWIY GMAWIY GMAWIY GMAWIY GWAWIY
 | 200
E I WAG
 | 2020
1 = A A K S X S - G
D A K A A A A A - D
D A K A A A A - D
D A K A A A A - D
D A K A A A - D
E A A I N K N S
E A A I I A A N A G
E A A A A A A G
E A G A A A A A C
E A G A A A A C
E A C A A A C A C
E A C A A A A C A A C
E A C A A A A C A A C
E A C A A A A A C A A A C
E A C A A A A A C A A A C A A A A A A A | ASE GTQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GS DG YQ FAFT
GS DG YQ FAFT
GS DG YQ FAFT
GN FA GQ FAFQ
GN AAS QQ FF
GN AAS QN F
GN AAS AN F
GN AAS AN F
GN AAS AN F
GN AAS AN F
G | 1233
AMALILTYGW
WMTYILTFGW
WMTYILTFGW
AMRAIVTEGW
AMRAIVTGW
ALKWITTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKWIVTGW
TIKWIVTGW
TIKWIVTGW
AMMMILVTGW
AMMMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRHIVTGW
ALRTFILGW | 100
A I YP IG YL YG
A I YF IG YL YG
A | 310 QG D
 | 220
240
240
240
240
240
240
240 | 330
DNL - N I U YN I
DNL - N I U YN I
DML - N I U YN I
DML - N I U YN I
DDM - N I YN I
DDM - N I YN I
E L - N I YNL
E L - N I YNL
E L - N I YNL
E AL - N I YNL H I YNL
E AL - N I YNL H I YNL H I YNL
E AL - N I YNL H I YNL |
| Po-2 HF10 3D09
Ep-12 MedDCM-UL2012-C1059
Ep-14 MedDCM-UL2012-C1059
Ep-14 MedDCM-OCT07-C4997
Po-94 HF70 35038
Ep-128 Ent893983-C478
MedDCM-OCT-324-C2
Ep-13 Ent895983-C478
MedDCM-OCT-324-C2
Ep-13 MedDCM-OCT-334-C2
Ep-13 MedDCM-OCT-334-C2
Ep-13 MedDCM-OCT-334-C2
By amma He12
MedeBCA-45C05
HF10 45E05
gamma EBAC20609 (8)
gamma EB | 220
1 S A 1 S | 226 D GM WVP F & A V EM VAW - F A T EM VAW - F A T AT I AW A T AT I AW A T T I AW - A T S I WW A T MW A A T S I WW T AMP A F T AWP A F T A T A G F - V T A W A G T A W A G T A T A G F - V T A T A G F - V T A T A G F - V T A T A G F - V T A G G W A G T A T A G W A G T A T A G W A G T A T A G W A G T A G W A G T A G W A G T A G W A G T A G G W A G T A G G W A G T A G G W A G T A G G W A G T A G G W A
 | 2030 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWIYIYIY GMACWIYIYIY GMACWIYIYIY GMACWIYIYIY GMACWIYIY GMICWFYIY GMICWFYIY GMICWFYIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GCLAWYMIY GCLAWYMIY GGUWWYNIY GGWWYNIY GGWWWIY GGWWWIY GGWWWMIY GGVGWAMIM GGVGWAMII- GGVGWAMIIY GCFAWIY GCFAWIY GY
 | 200
E I WAG

 | 1270
1 = A KE SVS - G
D AK KAAA - N
D AK KAAA - D
D AK KAAA - D
S I N DAAQ - K
E AA I L N ENS
E AA KA KAAA - D
E TAK AN AGS
E TAK AN AGS
E TAK AN AGS
E TAK AN AGS
E AA G L N ENS
E AGK AN AKS
E AGK NAAKS
E AGK NAAKS
E AGK NAAKS
E AGK NAAKS
E AGK OL NSK
S G G V DN SK
- S G G V DN SK
- V V K AA E SAAS
- V V V K E AAS D | ASE GTQ FAFT
ASE GTQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN FA GQ FAFQ
GN FA GQ FAFQ
GN FA GQ FAFQ
GN FA GQ FAFQ
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AAXQ TAYN
GN KAL VTAFG
GN KAL VTAFG
GN KAL VTAFG
GN KAL VTAFG
GN AAVQ FAYN
GN KAL VTAFG
GN FA GA GA GA GA
STA GA GA GA GA
STA GA GA GA GA
STA GA GA GA GA
STA GA GA GA GA GA GA GA GA GA
STA GA | 200
A MA & LI L U & GW
WMT Y I L T F GW
A MR AI V T G W
S L & W V T G W
S L & W V T G W
T I & W V T G W
A MMM I V V GW
A MMM I V V GW
A MM W V V GW
A MM W V V GW
A MM W V V GW
A M W V V GW
A L & W V T G W
A M W V V GW
A M W V V GW
A L & W V V V GW
A L & W V V V GW
A L & W V V V V V V V V V V V V V V V V V V | 100
A 1 YP LG YL LG
A 1 YP LG YL YG
A 1 YF LG YL YG
A | 310 QGD
 | 1200
G G T | 300
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DML - N U YN
DDM - N U YN
E L - N U YN
E M - N U |
| Pp-2 HEID 3D09
Ep-28 MedDCA-UL2012-CL059
Ep-28 MedDCA-UC207-C6997
Pp-3 HF0 19812
Pp-4 HF0 59088
Ep-28 ERR59898-C478
MedDCM-OCT-324-C2
Ep-13 ERR59898-C478
MedDCM-OCT-324-C2
Ep-13 MedDCA-OCT-324-C2
Ep-13 MedDCA-OCT-324-C2
Ep-13 MedDCA-OCT-324-C2
Ep-13 MedDCA-OCT-324-C2
Ep-13 MedDCA-OCT-324-C2
Ep-13 MedDCA-OCT-324-C2
Ep-13 MedDCA-OCT-324-C2
Ep-13 MedDCA-OCT-324-C2
Samma EBAC20609 (8)
gamma | 220
I S A IS | 240
D GM VWP LF AV
EMV XAW A Y
EMV XAW A Y
AT I AW A I
T I AW A I
T I AW A I
T I AW A I
S I VWW T L
S I VWW T L
S I VWW T L
S V VW T L
V V T L GW V I
AWP A F I I
T T L GF V I
AKM A F I P
T T V GF A F V I
AKMA F V I
AKMA F V I
AKMA F V I
AKMA G
 | 200
GMLAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMACWIYIY
GMACWIYIY
GMACWIYIY
GMACWIYIY
GMACWIYIY
GMACWIYIY
GMIUKUYIY
GMIUKUYIY
GMIUKUYIY
GMIUKUYIY
GMIUKUYIY
GMACWIYIY
GCLAWYYIYY
SMACWYYIY
SMACWYYIY
SMACWYYIY
SMACWYYIY
SMACWYYIY
SMACWYYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
GCLAWYYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
SGALAYFYIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIAMIY
STIA
 | 200
E I WAG

 | 11
2 A KE SVS - G
D A KAAAA - N
D A KAAAA - D
D A KAAAA - D
D A KAAAA - D
D X KAAA - D
D V KE ASA - S
S I N D AA Q - K
E AA I L N N S
E AA I L N N S
E AA I L N N S
E AA I N N S
E TAA I N N S
E AA I S
E AA I S
E A S
E AA I S
E A S | ASE GTQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN FAG GFAFQ
GN FAG GFAFG
GN FAG GN FAG
GN FAG GN FAG GN FAG
GN FAG GN FAG
GN FAG GN FAG GN FAG
GN FAG GN FAG
GN FAG GN FAG GN FAG
GN FAG GN FAG GN FAG GN FAG GN FAG
GN FAG GN | 289
AMALILITYGW
WMTYILTFGW
WMTYILTFGW
AMRAIVTGW
AMRAIVTGW
ALKWIVTYGW
SLKWIVTYGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKIVTYGW
ALKYIVTGW
TIKIVTGW
ALKYIVTGW
ALKYIVTGW
AMMMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
ALKTFILIGW
AMRHICGW
ALKFILIGW
MMRLFILIGW | 300
A I YP I GY LYG
A I YP I GY LGY FMG
A I YP LGY FMG
A I YP LGY FMG
A I YP LGY LGY FMG
A I YP LGY LGY FMG
A I YP LGY LGY LG
A I YP LGY LGY LGY
A I YP LGY LYG
A I Y LGY LYG
A I YP LGY LYG
A I Y LGY LYG
A I Y LGY LYG
A I YP LGY LYG
A I Y | 310 QGD - MDA DA MDA DA NDMI PG LDE - TITIFG LE LDE - VFG GG YHG GG YLM GG YLM GG YHT GG YH
 | 320
G G T | 330
DNL - N I L YN I
DML - N I VYN I
DML - N I VYN I
DML - N I VYN I
DDM - N I YN I
E L - N I YNL
E L - N I YNL
E L - N I YNL
E A - N I YNL
E S - N Y YNL
E S - N YNL
E S - |
| Pop-2 HED 3D09
Epi-28 MedDCM-UL2022-CL059
Epi-20 MedDCM-UL2022-CL059
Epi-28 Ent/S0207-C6937
Pop-4 HFO 195028
Epi-28 Ent/S03988-CA78
MedDCM-OCT-524-C2
Epi-28 Ent/S03988-CA78
MedBCM-OCT-524-C2
Epi-28 Ent/S03988-CA78
MedBCM-OCT-524-C2
Epi-28 Ent/S03988-CA78
MedBCM-OCT-534-C2
Epi-28 Ent/S0398-CA78
MedBCM-OCT-547-C2
Epi-28 Ent/S0398-CA78
MedBCM-OCT-547-C2
Failure (Control (C))
SAR86E (C)
Publique HTCC1052 (C)
Pelagibacter sp. HTCC211(8)
gamma EHC2070 (G)
G-pailidula (B)
MedDCM-OCT-257-C80
MedDCM-OCT-527-C80
MedDCM-OCT-527-C80
MedDCM-OCT-557-55
Esiguebacterium sp. AT15
H. marismortul ATCC 43049 | 220
I S A I S
A AIND
A AIND
F G D A S
F A D
F A D . | 240
D GM VWP LF AV
EMV XAV A I
AT 1 AV A I
AT 1 AV A I
T 1 AV A I
T 1 AV A I
T 2 I WV A I
S I VWV A I
S I VWV A I
S I VWV I
V V VWV I
V V VWV I
V V VWV I
T 2 GP A F F V
AMV G V I
AMP A F I
AMP A F I
T 1 G F V I
AT G F V I
AT G F V I
A C G V I
A C V C V V V V V V
 | 250
CML AWI YI I Y
CML CWF YI I Y
CML CWF YI I Y
CML AWI YI I Y
CML AWF Y I Y
CML AWF YI I Y
CML AWF Y
 | 200 E I WAG E I WG E I F G VA T G I F G VAT G I F M LL F T N LL F T N LL F T N VA
 | 1
 | ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GN FAGQ FAFQ
GN FAGQ FAFT
GN FAGQ FAGQ FAFT
GN FAGQ FAGQ FAGQ FAGQ FAGQ FAGQ FAGQ FAGQ | 1293
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKWIVTGW
TIKWIVTGW
TIKWIVTGW
AMMMILVTGW
AMMMIVTGW
AMMMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRHIVTGW
ALRTFILGW
MRTFILGW
VLRNILGW | 100
A 1 Y PL G Y L Y G
A 1 Y PL G Y L Y G Y L Y G
A 1 Y PL G Y L Y G Y L Y G
A 1 Y PL G Y L Y G Y L Y G Y L Y G
A 1 Y PL | 310 QGD
 | 120 G G T F D T F Q Q F Q Q F G E S F G E S F G E S F G E S F G E S F G E S V D T T N D S. T N D S. V D E S V D E S V D E S V D E S V D A S V D A S V D A S Q Q P S F G G S Q D G S Q D G S Q D G S Q D G S Q G G G G G G G G G G G G G G G G G G G | 300
DNL - N I U YN I
DNL - N I U YN I
DML - N I U YN I
DML - N I U YN I
DDM - N I YN I
E L - N I YN L
E L - N I YN L
E L - N I YN L
DAL - N I YN L
E L - N I YN L - N I YN L
E L - N I YN L - N I YN L
E L - N I YN |
| Pop-2 HF10 3D09
Epi-2 B MedDCM-JUL2012-C1059
Epi-1 RMS39039-C1654
MedDCM-JUL2012-C1059
Pop-4 HF70 59038
Epi-28 Ent893983-C478
MedDCM-OCT-524-C2
Epi-3 Ent895983-C478
MedDCM-OCT-524-C2
Epi-3 MedDCM-OCT-534-C5
Epi-3 MedDCM-OCT-534-C5
Byamma Hc7 Z040
Mede8AC45020
HF10 45008
gamma EBAC20508 (B)
gamma EBAC20508 (B)
gamma EBAC20508 (B)
gamma EBAC20508 (B)
gamma EBAC20508 (B)
gamma EBAC20508 (B)
gamma HC7C-1052 (C)
Pelaplikater sp. HTCC7111(B)
gamma HC7C0707 (C)
G.palildula (B)
D.donghaenis MED134 (G)
MedDCM-OCT-526-C80
Pop-3 MedDCM-OCT-535-L5
Edguabacter Jun 30-C1526
MedDCM-OCT-525-L5
Edguabacter Jun 30-C1526
MedDCM-OCT-526-C80
Pop-2 HF10 3D09
E-j-28 MedDCM-U12012-C1059 | 2200
I S A I S
A NIN
A NIN
F G D AS
F G D AS
F G D AS
M I G
G G S
M I G
G S
I G O
C S
M I G
G S
I G O
C S
M I G
G S
I G O
C S
I N
C S
I N
C S
I N
I N
C S
I N
I N
I N
C S
I N
I N
I N
C S
I N
I N
I N
I N
C S
I N
I N
I N
I N
I N
I N
I N
I N | 240
D GM VWP F E A V
EMV XW A V
EMV XW A V
AT I AW A I
T I AW A I
T I AW A I
T E I WW A I
S I VWW A I
A I A A I A I
A I A I A I
A I A I
A I A I A I
A I A I
A I A I A I
A I A I A I
A I A I
A I A I A I A I
A I A I A I A I
A I A I A I A I A I
A I A I A I A I A I
A I A I A I A I A I A I A I A I A I A I
 |
200
GMLAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMACWIYIIY
GMACWIYIIY
GMACWIYIIY
GMACWIYIIY
GMICWFYIIY
GMICWFYIIY
GMICWFYIYY
GMIAWIYIIY
GMIAWIYIIY
GMIAWIYIIY
GTLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCCAWYMI
GCCAWYMI
GCCAWYMI
GCCAWYMIY
GCCAWYMI
GCCAWYMIY
GCCAWYMI
GCCAWYMIY
GCCAWYMI
GCCAWYMIY
GCCAWYMIY
GCCAWYMIY
GCCAWMI
GCCAWMI
GCCAWMI
GCCAWMI
GCCAWMIY
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWAMI
GCCAWA
GCCAWAMI
GCCAWAMI
GCCAWA
GCCAWAMI
GCCAWAMI
G
 | 200
E I WAG
 | 12
2 A K K A S V S - G
D A K K A A S - D
D A K K A A S - D
D A K K A A S - D
D X K K A A S - D
D V K E A S A - S
S I N D A A C
E A A I L N N N
E A S I N N N S
E A S I I N N N
E A S I N N N S
E S S I N S I N S
E S S I N S I N S S
E S S I N S S I S S I S S S S S S S S S S | ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN EAG GA FAG
GN EAG GA FAG
GN EAG GA FAG
GN EAG GA FAG
GN AAS Q QA FK
GN AAA Q QA FK
GN AAA Q QA FK
GN AAA Q QA FK
GN AAA Q AT YN
GN KAL YT AFG
GN EAG GA FAG
GN EAG GA FAG
GN EAG GA FAG
GN AA Q GA FK
GN AAA Q FAFT
GN AAA Q FAFT
GN FAC GA FAG
GN | 200
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
AMRAIVTGW
ALKWIVTVGW
SLKWIVTVGW
SLKWIVTVGW
TIKLIVTVGW
TIKLIVTGW
TIKLIVTGW
ALKWIVTGW
AMMMILVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
ALKTFLUGW
ALRTFLUGW
ALRTFLUGW
ALRWFLVGW
ALRTFLUGW
 | 100
A 1 YP LG Y LG
A 1 YP LG | 310 QGD | 1200
C G T | 300
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DDM - N U YN
DDM - N U YN
E L - N U YN
S L - N U YN
S N - N U YN
S N - N U YN
S N - N U YN
E N U YN
E N U YN
E N U YN
E N U YN |
| Pop-2 HEID 3D09
Ep-28 MedDCA-UL2012-CL059
Ep-28 MedDCA-UC207-C697
Pop-3 HF70 19902
Pop-3 HF70 19912
Pop-3 HF70 19912
Pop-4 HF70 19908
Ep-128 ERR93983-C478
MedDCA-UC7-524-C2
Ep-13 ERR93983-C478
MedDA-UC7-524-C2
Ep-13 MedDCA-UC7-534-C5
Ep-13 MedDCA-UC7-534-C5
By MedDA-UC7-534-C5
By MedDA-UC7-534-C5
By MedDA-UC7-534-C5
By MedDA-UC7-534-C5
Pop-14 MedDCA-UC7-524-C5
D-donghaenis MED134 (G)
MedDCA-UC207-C1678
Pop MGII G5
E sibiricum 255-15
Esgluobacterium 3p. AT15
H. marismortul ATCC 43049
Pop-2 HEID 3D09
Ep-28 MedDCA-UC207-C1659
Ep-12 ERR93992-CL554 | 220
1 S A S
A MIO
A MIO
A MIO
FGD IS
FGD AS
MI G - GS
MI G - GS
MI G - GS
MI G - GS
I G - GS | 240 D GM WWP F E AV D GM WP F E AV F M V AW - A I AT I AW A I AT I AW A I AT I AW A I T I WW A I S I WW A I S I WW A I S I WW T L S I WW T L S I WW T L M V GG V I M Y GG V I AM GG I I AMP A F I I AMP A F I I AMY GG - V I A GG WAGL L GY LG L L A G GW AGL L GY LG L L L T Q L G LWSYL I T Q L G LWSYL I MM V W A A T I D MM WW A A T ID MM WW A A T ID MM WW A A T ID
 |
201
GMLAWIYIIY
GMAAWIYIIY
GMAAWIYIIY
GMACWFYIIY
GMACWFYIIY
GMACWFYIIY
GMCCWFYIIY
GMICWFYIIY
GMICWFYIIY
GMIAWIYIIY
GMIAWIYIIY
GMIAWIYIIY
GMIAWIYIIY
GCLAWYWIY
GCLAWYWIY
GCLAWYWIY
GCLAWFYILY
GMACWFYILY
GMACWFYILY
GMACWFYILY
GMACWFYILY
GMACWFYILY
GMACWFYILY
GMACWFYILY
GCLAWYWIY
GCLAWFYILY
GCLAWFYIIY
GCLAWFYIIY
GCLAWFYIIY
GCCAWIYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFYI
GCCAWFY
 | 200
E I WAG
 | 2020
1 = CAKESVS-G
DAKAAAN-N
DAKKAAS-C
DAKKAAS-C
DAKKAAS-C
DAKKAAS-C
DAKKAAS-C
SINDAAC.W
EAAIINKNS
EAAIINKNS
EAAIINKNS
EAAIINKNS
EAAIINKNS
EAAIANAGS
ETAKAANAGS
ETAKAANAGS
EGAGUANASS
EGAGUANASS
EAGKAASS
EAGKAASS
CASSACONSK
SGGCVDNSK
- VIKAAEN
- VIKAAEN
- VIKAAEN
- VIKAAEN | ASE GTQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN FA GQ FAFQ
GN FA GQ FAFT
GN FA GQ FAFT
GN FA GQ FAFQ
GN FA GQ FAFQ
GN FA GQ FAFQ
GN FA GQ FAFT
GN FA GQ FAFQ
GN FA GQ FA GQ FAFQ
GN FA GQ FA GQ FAFQ
GN FA GQ FA GQ FA GQ FA
GN FA GQ FA GQ FA GQ FA GQ FA
GN FA GQ |
289
AMALILITYGW
WMTYILTFGW
WMTYILTFGW
MMTYILTFGW
AMRAIVTFGW
SLKWIVTYGW
SLKWIVTYGW
SLKWIVTYGW
SLKWIVTTGW
TIKIVTYGW
ALKYIVTGW
TIKIVTYGW
ALKYIVTGW
ALKYFILVTGW
AMMMIVTYGW
AMRMIVTYGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVGW
AMRHIVTGW | 300
AIYPIGYIGYIG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG | 310 QGD - - - - - - - - - - MDA - A - A - A - A - A - A - A - A - A - A - A - A - A - A - A - A - A A - A A |
320
G G T | 330
DNL - N I L YN I
DML - N I VYN I
DML - N I VYN I
DML - N I VYN I
DDM - N I YN I
E L - N I YNL
E S L - N I YNL I YNL
E S L - N I YNL I YN |
| Pop-2 HED 3D09
Epi-28 MedDCM-UL2012-CL059
Epi-28 MedDCM-UL2012-CL059
Epi-28 MedDCM-UC207-C697
Pop-4 HFO 15007-C697
Pop-3 HFO 15007-C697
Pop-3 HFO 15007-C534-C2
Epi-28 ERK95988-C478
MedDCM-OCT-534-C2
Epi-3 MedDCM-OCT-534-C2
Epi-3 MedDCM-OCT-534-C2
Epi-3 MedDCM-OCT-534-C2
Barma EAC-20C609 (b)
garma EAC-20C609 (b)
garma EAC-20C609 (c)
Palagibacter sp. HTCC211(a)
garma EAC-20C609 (b)
Gapalitula (b)
D-donghaenis MED134 (c)
D-donghaenis MED134 (c)
D-donghaenis MED134 (c)
D-donghaenis MED134 (c)
D-donghaenis MED134 (c)
D-donghaenis MED134 (c)
D-donghaenis MED134 (c)
D-1 MedDCM-OCT-507-C6
E, sibiricum 255-15
E-glaubacter um, p. AT1b
H. marismortui ATCC 43049
Pop-2 HF10 3D09
Epi-28 MedDCM-UL2012-C1059
Epi-38 MedDCM-UL2012- | 220
1 S A 1 S
A AIN
A AIN
FGD AS
FGD AS
FGD AS
FGD AS
MI G
GG AS
MI G
GG AS
I G GS
I | 240
D GM VWP LF AV
EMV XM A I
AT 1 AW A I
AT 1 AW A I
T 1 AW A I
T 1 AW A I
T 1 AW A I
T 2 I WW A I
S I VWW A I
AMP A I
AMP A I
AMP A I
AMP A I
AMP A I
A I
A I
A I
A I
A I
A I
A I
 | 300 GMLAWIYIIY GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWFYIY GMACWFYIY GMY GMACWFYIY GMICWFYIY GMICWFYIY GMACWFYIY GMACWFYIY GMACWFYIY GMACWFYIY GMACWFYIY GMACWFYIY GMACWFYIY GMACWFYIY GMACWFYIY GGVGWAMIM GGVGWAMIMIY GCIAWFYIIY SGAAFVIY GMCWAMIY GCIAWFYIIY STIAMIYY SGMACWY SGMACWY GMCWAMIY GCIAWFYIIY SGMACWY SGMACWY SGMACWY GMCWAMIY GMCWAMIY SGMCWAMIY SGMCWAMIY SGMCWAMIY SGMCWAMIY SGASUY
 | 200 E I WAG E I WG E I F RG VA T G I F PG I VA T G I F PG I VA T G I F PG I VA T G I F WS V I V

 | 12 0 A KA AAA S - U
0 A KA AAAA S - U
0 A KA AAA S - U
0 A KA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA | ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GN EAS Q FAFQ
GN EAS Q FAFT
GN EAS Q FAFQ
GN EAS Q FAFT
GN | 1213
AMA LI L T Y GW
WMT Y I L T F GW
WMT Y I L T F GW
AMR AI V T G W
S L KWI Y I L Y GW
S L KWI Y T GW
S L KWI Y T GW
Y T I KWI Y T GW
T I KWI Y T GW
T I KWI Y T GW
T I KWI Y T GW
AMMM I L Y Y GW
AMMM I Y Y GW
AMMM I Y Y GW
AMMM I Y Y GW
AMMM I Y T GW
AMMM I Y T GW
AMMM I Y T GW
AMMM I Y T GW
AMMM I Y GW
AMM I Y T GW
AMM I Y GW
AMM I Y T GW
AMM I Y T GW
AMM I Y T GW
AMM I Y GW
AMM I Y T GW
AMM I Y T GW
AMM I Y GW
A | 100
A I YP I GY L YG
A I YP I GY L YG
I YF I YG Y L YG
I YF I YG Y L YG
I YF I YG Y YG Y YG
I YF I GY L YG
I YF I YG Y YG YG
I YF I GY L YG
I YF I YG Y YG YG
I YF I GY L YG
I YF I YG Y YG YG
I YF I YG Y YG
I YF I YG Y YG YG
I YF I YG Y YG YG
I YF I YG | 310 QGD |
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200 | 340
DNL - N I U YN I
DNL - N I U YN I
DML - N I U YN I
DML - N I U YN I
DDM - N I YN I
E L - N I YN L
E L - N I YN L
E A I - N I YN L
DAL - N I YN L
DAL - N I YN L
DAL - N I YN L
E T - N I YN L
D - N I YN L
D - N I YN L
D - N I YN L
E T - N I YN L
E - N I YN I
E - N I YN I
E - N I YN I
D - N I YN L
E - N I YN I
D - N I YN I
C - N |
| Pop-2 HF10 3D09
Epi-28 MedDCM-JUL2012-C1059
Epi-28 MedDCM-JUL2012-C1059
Epi-28 MedDCM-JUL2012-C1059
Pop-4 HF70 59008
Epi-28 FRR59898-C478
MedDCM-OCT-524-C2
Epi-28 FRR59898-C478
MedDCM-OCT-524-C2
Epi-28 MedDCM-OCT-524-C2
Epi-28 MedDCM-OCT-524-C2
Epi-28 MedDCM-OCT-524-C2
Epi-28 MedDCM-OCT-524-C2
Epi-28 MedDCM-OCT-524-C2
Barma HEAC20508 (b)
garma EBAC20508 (b)
garma EBAC20508 (b)
garma EBAC20508 (b)
garma EBAC20507 (c)
G-pailudua (c)
D-donghaenis MED124 (c)
MedDCM-OCT-526-C80
Pop-2 MedDCM-OCT-526-C80
Pop-2 MedDCM-OCT-526-C80
Pop-2 MedDCM-OCT-526-C80
Pop-2 HF10 3D09
Epi-28 MedDCM-UL2012-C1059
Epi-28 MedDCM-UL201 | 2200 1 S A 1 S A INN A IND A INN | 240
D GM WVP F & A Y
EM Y AW - A Y
EM Y AW - A Y
AT I AW A I
AT I AW A I
T = I WW - A
S I Y WW - T L
S I Y WY - T L
S I Y W - T L
S I Y WY - T L
S I Y W - T L Y W - T L
S I Y W - T L Y W -
 | 200
GMLAWIYIIY
GMLAWIYIIY
GMLAWIYIIY
GMLAWIYIIY
GMLAWIYIIY
GMLAWIYIIY
GMLAWIYIIY
GMLAWIYIYIY
GMLAWIYIY
GMLAWIYIY
GMLAWIYIY
GMLAWIYIY
GMLAWIYIY
GMLAWIYIY
GMLAWIYIY
GMLAWIYIY
GMLAWIYIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIYIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMIY
GCLAWYMY
GCLAWYMY
GCLAWYMIY
GCLAWYMIY
GCLAWYM
 | 200
E I WAG
E I WAG
E I WAG
E I WAG
E I WG
E I F R G
E I F R G
 | 12
2 A K K S V S - G
D A K K A A S - D
D A K K A A S - D
D A K K A A S - D
D X K K A A S - D
D V K E A S A - S
S I N D A A C
E A A I L N N N
E A S I N K N S
E A S I I N K N S
E A S I I N K N S
E A S I I N K N S
E A A I A M A G
E A A I A K N A S S
E A S I N K N S
E A S A N A S S
E A S C N A A K S
S A K K A A S Y
D G G V D N S K
S G G V D N S K
- V K K A A E T
V V G G A A S D
- V K K A A S D
 | ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN EAG GA FAFQ
GN EAG GA FAFQ
GN EAG GA FAFQ
GN EAG GA FAFQ
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AAXQ SAFQ
GN EAG GA FA
GN AAVQ TAYN
GN KAL YTAFG
GN AAVQ TAYN
GN KAL YTAFG
GN EAL YTAFG
GN EAL YTAFG
AS FAW SAYN
AS FAVQ SAYN
AS FAVQ SAYN
HEL LWSSAD
IAFELLWSSAD
IAFELWSSAD
IAFELWSSAD
IAFELMSAN
SAN FAFT
SAN | 200
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTEGW
AMRAIVTGW
SILWITYIGW
SILWITYGW
TILTFGW
AMRAIVTYGW
TIKIVTYGW
TIKIVTYGW
TIKIVTYGW
AMMMILVYGW
AMMMILVYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
ALRTFILVGW
ALRTFILVGW
ALRWFIVYGW
ALRTFILVGW
ALRWFIVYGW | 100
A I YP IG YL YG
A I YF IG YL YG
A | 310 QGD MDA - DA NDMI - PG LDE - TT I LG LE - TT I LG YFG - DG YFG - GG YLM - GG YLM - GG YLT - GG YLT - GG YFG - UL - GG YFG - DG YLT - GG YLT - GG YGE - LG - LG YFG - DG YFG - DG YLA - GG YLA - GG YLA - GG YLA - GG YCA - GG YCA - GG YCA - GG YCA - I CG - LG YCA - TG YCA - TG YCA - TG YCA - TG
 | 1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200 | 300
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DML - N U YN
E L - N U YN
S N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N - N U YN
E N U YN
E N U YN U YN
E N U YN
E N U YN U YN U YN
E N U YN U YN U YN U YN
E U U U U U U U U U |
| Pop-2 HEID 3D09
Ep-28 MedDCA-UL2012-CL059
Ep-28 MedDCA-UC207-C497
Pop-3 HF70 15903
Ep-28 RH59398-C454
MedDCA-UC7524-C2
Ep-32 RH59398-C478
MedDCA-UC7524-C2
Ep-32 RH59398-C478
MedDCA-UC7524-C2
Ep-32 RH59398-C478
MedDCA-UC7524-C2
Ep-32 RH59398-C478
MedDCA-UC7524-C2
Ep-32 RH59398-C478
MedDCA-UC7524-C2
Ep-32 RH59398-C478
MedDCA-UC207-C678
MedDCA-UC207-C678
MedDCA-UC207-C678
Pop-1 MedDCA-UC758-C60
MedDCA-UC207-C678
HeiD 400-UC758-C60
Pop-1 MedDCA-UC758-C60
Pop-1 MedDCA-UC758-C60
Pop-1 MedDCA-UC207-C678
Pop-2 HF10 3D09
Ep-28 MedDCA-UC207-C678
Ep-32 RH50259-C654
MedDCA-UC207-C678
Pop-2 HF10 3D09
Ep-32 RH50259-C654
MedDCA-UC207-C678
Pop-3 HF10 5D09
Ep-32 RH50250-C678
Pop-3 HF10 5D07
Ep-32 RH50250-C678
Pop-3 HF10 5D07
Ep-32 RH50250-C678
Pop-3 HF10 5D07
Ep-32 RH50250-C678
Pop-3 HF10 5D07
Ep-32 RH50250-C678
Pop-32 RH50250
Pop-32 RH50250-C678
Pop-32 RH50250
Pop-32 RH5025 | 220
1 S A S | 240 D GM WWP F E A V D GM WWP F E A V F M V AW - A I AT I AW A I AT I AW A I AT I AW A I T I WW A I S I WW T L M Y GE V I AM GE V I AT GE V I
 | 3050 GMLAWIYIIY GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWFYIIY GMACWFYIIY GMICWFYIIY GMIAWIYIIY GMIAWIYIY GMIAWIYIY GMIAWIYIY GMIAWIYIY GMIAWIYIY GMIAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GMAWIYIY GWAAMIY GOVGWAMIH GOVGWAMIY GCIAWIYIIY STIAMASE TAAASE TAAASE TAAASE CIAWIYIIY STIAASE TAAASE GMAKWIYIIY GMUAWIYIY GUNAWIY GOVGWAMIY GOVGWAMIY GOVGWAMIY
 | 200
E I WAG
E I WAG
E I WAG
E I WG
E I WG
E I WG
E I F G
E I F G
M F

 | 12
2 A K K A S Y S - G
D A K K A A S - D
D A K K A A S - D
S - N A A A S - D
S - S - N A A A A S - D
S - S - N A A A A A S - S
E A A I I N K N S
E A A C A A A A A S - S
E A A A A A A S - S
E A A A A A A S - S
E A A A A A A S - S
E A A C A A A A A S - S
E A A C A A A A A S - S
E A A C A A A A A S - S
E A A C A A A S - S
E A A C A A S - S
E A A C A A S - S
E A C A C A S - S - S - S - S - S - S - S - S - S | ASE GTQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GS DGYQ FAFT
GS DGYQ FAFT
GS DGYQ FAFT
GN FAG GF AFQ
GN FAG GF AFG
GN FAG GN FAG
GN FAG GN FAG GN FAG
GN FAG GN FAG
GN FAG GN FAG GN FAG GN FAG
GN FAG GN FAG GN FAG GN FAG GN FAG GN FAG
GN FAG GN FAG G | 280
AMALILITYGW
WMTYILTFGW
WMTYILTFGW
MMTYILTFGW
AMRAIVTGW
SLKWIVTYGW
SLKWIVTYGW
SLKWIVTGW
SLKWIVTGW
TIRIVTGW
ALKYIVTGW
ALKYIVTGW
ALKYIVTGW
ALKYIVTGW
ALKYIVTGW
AMMMIVTGW
AMMIVTUYGW
AMRIVTUYGW
AMRIVTUYGW
AMRIVTUYGW
ALRTFILIGW
ALRTFILIGW
NRLFILIGW
VLRNILUYW | 300
A I YP I GY L YG
A I YP I GY L GY I A
A I YP I GY L GY I FA
A I YP L GY L GY I FA
A I YP L GY L GY I FA
A I YP L GY L GY L GY
A I YP L GY L YG
A I YP L GY L YG A I YF L GY L YG
A I YF L GY L YG A I YF L GY L YG
A I YF L GY L YG A I YF L GY L YG A I YF L GY L YG A I | 310 QGD MDA - DA NDMI - FG LDE - TTI LG LDE - TTI LG YFG - OG YMG - GG YLM - GG YLM - GG YHT - GG YFT - GG YFT - GG YFM -
 | 320
G G T | 300
DNL - N I U YN I
DNL - N I U YN I
DML - N I U YN I
DML - N I U YN I
DDM - N I YN I
DDM - N I YN I
E LI - N I YN L
E LI - N I YN L
E AI - N I YN L
E AI - N I YN L
DA L - N I YN L
E N - N I YN L
S L - N I YN L - N I YN L
S L - N I YN L - N I YN L
S L - N I YN L - |
| Pop 2 HED 3D09
Epi-28 MedDOM-UL3D12-CL059
Epi-28 MedDOM-UL3D12-CL059
Epi-38 MedDOM-OCT307-C697
Pop 4 HF70 159C08
Epi-28 ERK95988-C478
MedDOM-OCT-524-C2
Epi-28 ERK95988-C478
MedDOM-OCT-524-C2
Epi-3 MedDOM-OCT-534-C2
Epi-3 MedDOM-OCT-534-C2
mma EAACJAD8 (6)
SAR86E (6)
Publique HTCC1062 (6)
Pelagibacter sp. HTCC211(8)
gamma HOT 25nd (8)
SAR86E (6)
Publique HTCC1062 (6)
Pelagibacter sp. HTCC211(8)
gamma HOT 25nd (8)
SAR86E (6)
Publique HTCC1062 (7)
Pelagibacter sp. HTCC211(8)
gamma HOT 25nd (8)
MedDOM-OCT 2507 (6)
E splitticum 255-15
E-giupobacterium sp. AT1b
H. marismortui ATCC 43049
Pop 2 HEID 3D09
Epi-28 MedDOM-OCT 2602
Epi-28 MedDOM-OCT 2602
Fop 31 HF0 15912
Pop 31 HF70 S9C08
Epi-28 HeID 39838-C478
MedDOM-OCT 524-C2
Fop 31 HF70 S9C08
Epi-28 HeID 59838-C478
MedDOM-OCT 524-C2
Epi-31 MedDOM-OCT 524-C2
Epi-31 MedD | 2200 1 S A 1 S A AIN | 240 D GM WWP LF AV D GM WWP LF AV AT I AW A I T I AW A I S I WW - T L S I WW T L Y WW - T L Y WW - T L MW CG F V AMU GG V I AMP A I P T L G F - V I AKM A F I P T L G F - V I AKM A F - I P T L G F - V I A G GWA GL I L L G G L W A T MD MW W A A T M
 | 253 GMLAWIYILY GMLAWIYILY GMAAWIYILY GMAAWIYILY GMAAWIYILY GMAAWIYILY GMAAWIYILY GMAAWIYILY GMAAWIYILY GMACWFYILY GMICWFYILY GMICWFYILY GMACWFYILY GMACWFYILY GMACWFYILY GMACWFYILY GMACWFYILY GMACWFYILY GGVGWAMIM GGVGWAMIM GGVGWAMIM GGVGWAMIMY GGCLAWFYILY SGAAFVILY GGVGWAMIMIY GGVGWAMIMY GGVGWAMIMY GGCLAWFYILY STIAMIYINY
 | 2000 E I WAG E I WGG E I F RG E V F SG U A T GT F WS V A T GT F WT L L F T N L L F T N V A T GT F WT L L F T N C C C C C C C C C C C C C C C C C C C

 | 12
2 A K K S V S - G
D A K K A A S - D
D A K K A A S - D
S - D A K A A S - D
S - D A K A A S - D
S - A I I N K N S
E A A K A A S - T
E G K A A K A S
E A G K A A A K S
E A G K A A A K S
E A S C V N S K
S - S G K O V N S K
- S G K G V D N S K
- S G K G V D N S K
- S G K G V D N S K
- S G K G V D N S K
- V T K A A E N
V V V G E A A S D | ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GN EASQ FAFQ
GN EASQ FAFT
GN EASQ FAFT | 1212
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKUIVTGW
TIKWIVTGW
TIKUIVTGW
AMMMILVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMRMIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
ALRTFILIGW
ALRTFILIGW
MRRFILIGW | 100
A I YP I GY L YG
A I YP I GY L GY
A I YP I GY | 310
QGD
MDA - DA
MDA - DA
NDMI - PG
LDE - TT ILG
LDE - TT ILG
LEA - TT ILG
LEA - TT ILG
CHA
YFG - DG
YFG
YFG
YFG
GG
YHM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
CLGLVG
 | 120
120
120
120
120
120
120
120 | 340
DNL - N I U YN I
DNL - N I U YN I
DML - N I U YN I
DML - N I U YN I
DDM - N I YN I
E L - N I YN L
E L - N I YN L
E L - N I YN L
E A L - N I YN L
E A L - N I YN L
E T - N I YN L
E - N I YN L - N I YN L
E - N I YN L - N I YN L
E - N I YN L - N I YN L
E - N I YN L - N I YN L
E - N I YN L - N I YN L
E - N I YN L - N I Y |
| Po-2 HF10 3D09
Epi-28 MedDCM-JUL2012-C1059
Epi-28 MedDCM-JUL2012-C1059
Epi-28 MedDCM-JUL2012-C1059
Po-4 HF70 59038
Epi-28 RenS93983-C478
MedDCM-OCT-324-C2
Epi-28 RenS93983-C478
MedDCM-OCT-324-C2
Epi-28 RenS93983-C478
MedBCM-OCT-324-C2
Epi-28 RenS93983-C478
MedBCM-OCT-324-C2
Epi-28 RenS93983-C478
MedBCM-OCT-324-C2
Epi-28 MedBCM-JUL2012-C1059
Po-28 HF10 3D09
Epi-28 MedDCM-JUL2012-C1059
Po-28 HF10 3D09
Epi-28 MedDCM-JUL2012-C1059
Epi-28 MedBCM-JUL2012-C1059
Epi-28 MedBCM-JUL20 | 2200
1 S A 1 S
A AIND
A AIND
F GD AS
F G GD AS | 240
D GM WVP F & A V
EM V AW - F A T
AT I AW A I
AT I AW A I
T A W - A A T
F I WW - A A T
F I WW - A A T
S I VWW - T L
S I VW - T L
S I V - T L V - T L
S I V - T L V - T L
S I V - T L V - T L V - T L V - T L V - T L V - T L V - T L V - T L V - T L V - T L V - T L V - T L V - T L V - T L V - T L V - T L V - T L
 | 2030 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMICWFYIY GMICWFYIY GMAWIYIY GMAWIYIY GMAWIYIY GMACWUYYIY GMACWUYYIY GMACWUYYIY GMACWUYYIY GMACWUYYIY GMACWUYYIY GMACWUYYIY GMACWUYYIY GOLWWYMIY GOLWWYMIY GOLWAMIMI- GOVGWAMIMI- GOVGWAMII- GOVGWAMIY CFAWIYIY
 | 200
E I WA G
E I WG
E I F R G
E I F R G
 | 12
2 A K K S V S - G
D A K K A A S - D
D A K K A A S - D
D A K K A A S - D
D X K K A A S - D
D V K E A S A - S
S I N D A A C
E A A I A N N S
E A S I I N N S
E A S I N S S
E A S I N S S
E A S S N S S
E A S S S S S S
E S S S S S S S S S S S S S S S S S S S
 | ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GAFA
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AASQ QAFK
GN AAXQ TAYN
GN KAL YTAFG
GN AAVQ TAYN
GN KAL YTAFG
TAFG
GN AAVQ TAYN
GN KAL YTAFG
GN AAVQ TAYN
GN KAL YTAFG
GN AAVQ TAYN
GN KAL YTAFG
GN AAVQ TAYN
GN KAL YTAFG
TAFG
GN AAVQ TAYN
GN KAL YTAFG
GN AAVQ TAYN
GN KAL YTAFG
GN AAVQ TAYN
GN KAL YTAFG
TAFG
TAFG
TAFG
TAFG
TAFG
TAFG
TAFG | 2020
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTVGW
SLKWIVTVGW
TIKLIVTGW
TIKLIVTGW
TIKLIVTGW
TIKLIVTGW
AMMMILVTGW
AMMMILVTGW
AMMMILVTGW
AMMMIVTVGW
AMMMIVTVGW
AMMMIVTVGW
AMMMIVTVGW
AMMMIVTVGW
AMMMIVTVGW
AMMMIVTVGW
AMMMIVTVGW
AMMMIVTVGW
AMMMIVTVGW
AMMMIVTVGW
AMMMIVTVGW
ALRTFILVGW
ALRTFILVGW
ALRTFILVGW
ALRTFILVGW | 100
A I YP I GY L YG
A I YP I GY L GY L GY
A I YP I GY L GY
A I YP I GY L GY
A I YP I GY L GY
A I YP L GY L GY
A | 310 QGD MDA - DA NDMI - PG LDE - TT I LG LDE - TT I LG YFG - DG TT I LG YFG - GG
 | 1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200 | 300
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DDM - N YN
E L - N YN
S L - N YN
S N - N YN
E L - N YN
S N YN
S N - N YN
S N |
| Pop-2 HEID 3D09
Ep-28 MedDCA-UL2012-CL059
Ep-28 MedDCA-UC207-C4997
Pop-3 HF70 15908
Ep-28 RH593983-C456
MedDCA-UC7524-C2
Ep-32 RH5939838-C478
MedDCA-UC7534-C2
Ep-32 RH593983-C478
MedDCA-UC7534-C2
Ep-32 RH593983-C478
MedDCA-UC7534-C2
Ep-32 RH593983-C478
MedDCA-UC7534-C2
Ep-32 RH593983-C478
MedDCA-UC7534-C2
Ep-32 RH593983-C478
MedDCA-UC7534-C2
Ep-32 RH59398-C478
MedDCA-UC207-C1578
Pop-1 MedDCA-UC7534-C2
Ep-31 RH502892-C530
MedDCA-UC207-C1578
Pop-31 HF10 3D09
Ep-28 MedDCA-UC7207-C1578
Pop-31 HF10 3D09
Ep-128 RH60CA-UC207-C478
Ep-12 RH593993-CL534
MedDCA-UC207-C478
Ep-32 RH50 3D09
Ep-32 RH50 3D09
Ep-3 | 220
1 S A S | 240 D GM WVP F E A V D GM WVP F A V AT I AV A I T I AV A I S I VWV - T L S I VWV - T L V V WW - T L MV G VI MV G VI AMP A I I AMP A I I AMP A I AMP A I AMP A I AMP A I AT L G VI A GU L L A GU L A GW A A TMD MMV WA A T MA A MW VA A TMD MMV WA A T MA A MW VA A TMD MMW WA A T MA A MW A TM MU VA A TMA MW VA A TMA MW VA A TMA MW VA A TMA MW VA A TMA M
 | 3050 GMLAWIYIIY GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWFYIIY GMCVFYIY GMCVFYIY GMICWFYIY GMAWIYIY GWAMMIY GUWAMIN GUWAMIY GUWAMIY GUWAMIY GUWAMIY GUWAMIY GUWAMIY GUWAMIY TAAAESKE TAAAESKE TAAAESKE TAAAESKE TAAAESKE TAAAESKE TKAAASCESERT
 | 2000 E I WAG E I F RG VA T G T PWS VAT G T PWT L I F TN L I F TN VAT G T PWT L I F TN VAT G T PWT W W C C C C C C M

 | 12
2 A K E S V S - G
D A K A AA N - N
D A K K AA S - D
D A K K AA S - D
D A K K AA S - D
D A K K AA S - D
S - N C A A C - V
S - S - N C A C - V
E A A I N K NS
E A A I N K NS
E A A I N K NS
E T A K A M A G S
E T A K A M A G S
E T A K A M A G S
E T A K A M A G
E A G A C A C - T
E G A G A C A C - T
E G A G A C A C - T
E G A G A C A C - T
E G A G A C A C - T
E G A G A C - T
E | ASE GUY CAPAFT
ASE GUY CAPAFT
ASE GUY CAPAFT
ASE GUY CAPAFT
GS DC UY CAPAFT
GS DC UY CAPAFT
GS DC UY CAPAFT
GN EAG CAPAFQ
GN EAG CAPAFQ
GN EAG CAPAFQ
GN EAG CAPAFQ
GN AAS QAFAF
GN AAS QAFAF
GN AAS QAFAF
GN AAS QAFA
GN AAS AAS AAS AAS AAS AAS AAS AAS AAS AA | 280
AMALILITYGW
WMTYILTFGW
WMTYILTFGW
MMTYILTFGW
AMRAIVTGW
SLKWIVTYGW
SLKWIVTYGW
SLKWIVTGW
SLKWIVTGW
TIKLIVTGW
TIKLIVTGW
ALKWIVTGW
ALKWIVTGW
ALKWIVTGW
ALKUVTGW
AMMMILVYGW
AMMIVTUVGW
AMMMIVTVGW
AMRWIVTGW
AMRWIVTGW
AMRWIVTGW
AMRWIVTGW
AMRWIVTGW
AMRWIVTGW
AMRWIVTGW
AMRWIVTGW
ALKTFILIGW
ALRWFIVJGW
ALRWFIVJGW
ALRWFIVJGW | 200
A 1 Y PL G FL I G
A 1 Y PL G Y L YG
A 1 Y PL | 310 QGD MDA -DA MDA -DA MDA -DA MDA -DA MDA -DA MDA -DA NDMI -FG LDE -TTILG VEG -TTILG YFG -GG YLG -GG YLM -GG YLM -GG YLM -GG YLM -GG YLM -GG YLM -GG YMT -GG YE -GG <
 | 1200
0 G T | 300
DNL - N I U YN I
DNL - N I U YN I
DML - N I U YN I
DML - N I U YN I
DDM - N I YN I
DDM - N I YN I
E LI - N I YN L
E LI - N I YN L
E AI - N I YN L
E AI - N I YN L
DA L - N I YN L
E AI - N I YN L
DA L - N I YN L
E AI - N I YN L
DA L - N I YN L
S L - N U YN L - N U YN L
S L - N U YN L - N U YN L
S L - N U YN L - |
| Pop 2 HED 3D09
Epi-28 MedDOM-UL3D12-CL059
Epi-28 MedDOM-UL3D12-CL059
Epi-38 MedDOM-OCT307-C697
Pop 3 HF70 19508
Epi-28 ERK95988-C478
MedDOM-OCT-524-C2
Epi-3 ERK959894-C530
MedDOM-OCT-524-C2
Epi-3 MedDOM-OCT-534-C2
mma EBAC3C02609 (b)
pamma EBAC3C02609 (c)
pamma EBAC3C0270 (c)
GapaIIIdula (d)
Dadonghaenis MED134 (c)
Dadonghaenis MED134 (c)
Dadonghaenis MED134 (c)
Pop J. MedDOM-OCT-526-C80
MedDOM-OCT-526-C80
Pop J. MedDOM-OCT-536-C80
Pop J. MedDOM-OCT-536-C80
Pop J. HeRDSM951-CL54
MedDOM-OCT-509-C154
E- sibiricum 255-15
E-sibiricum 255-15 | 2200
1 S A 1 S | 240 D GM WVP LF AV D GM WVP LF AV AT I AV A I T E I WW T L S I WW T L V VWW T V V WW T L V VWW T L MP A F I I AM A F I P T L G F - V I AKM A F I P T L G F - V I A G U A - T I MV WA A T MD MW WA A T MD MM WW A A T MD <t< th=""><th>333 GMLAWIYIIY GMLCWFYIIY GMLCWFYIIY GMLAWIYIYIY GGUCWAMIMIY GGUCWAMIMIY GCLAWFYNILY SGALYEN GGUCWAMIMIY GCLAWIYIYIY STIAAKSE TAAAKSE TAAAKSE TAAAKSE TAAAKSE TAAAKSE TAAAKSE TAAAKSE TAAKSE TAAKSE TAAKSE TAAKSE TAAKSE</th><th>2000 E I WAG E I WGG E I F RG E V F SG U V A T G L P F GL V A T G L P F GL V A T G T P WS V A T G T P WS L L F T N L L F T N V L V S MF - S S S S S S S S S S S S S S S<</th><th>12
2 A K K A S V S - G
D A K K A A S - D
D A K K A A S - D
S - D A K A A S - D
S - A I I N K N S
E A A I I N K N S
E A A I I N K N S
E A A I I N K N S
E T A K A N A G S
E A G K L N A S
E A G K L N A S
E A G K L N A S
E A G K A A K S
S - S G K G V D N S K
- S G K G V D N S K
- S G K G V D N S K
- S G K G V D N S K
- S G K G V D N S K
- S G K G V D N S K
- V T K A A E N
- V T K A A E N
- V T K A A E N
- V V V G E A A S D</th><th>ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GN EASQ FAFQ
GN EASQ FAFT
GN EASQ FAFT</th><th>2020
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKUIVTGW
TIKWIVTGW
TIKWIVTGW
AMMMILVTGW
AMMMILVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMRHIVTGW
AMRHIVTGW
ALRTFILIGW
ALRTFILIGW
ALRTFILIGW</th><th>100
A I YP I GY L YG
A I YF I GY L YG
A I YP I GY L YG
A I YF I GY L YG
A I YF</th><th>310
QGD
MDA - DA
MDA - DA
MDA - DA
NDMI - PG
LDE - TT I LG
LDE - TT I LG
LEA - TT I LG
LEA - TT I LG
GG
YFG - DG
YFG - GG
YFG - GG
YHM - GG
YLM - GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG</th><th>1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200</th><th>300
DNL - N I U YN I
DNL - N I U YN I
DNL - N I U YN I
DNL - N I U YN I
DDM - N I YN I
E L - N I YN L
E L - N I YN L
E A I - N I YN L
D A L - N I YN L
D A L - N I YN L
E A I - N I YN L
E I - N I YN I YN I YN I YN I YN I YN I Y</th></t<> | 333 GMLAWIYIIY GMLCWFYIIY GMLCWFYIIY GMLAWIYIYIY GGUCWAMIMIY GGUCWAMIMIY GCLAWFYNILY SGALYEN GGUCWAMIMIY GCLAWIYIYIY STIAAKSE TAAAKSE TAAAKSE TAAAKSE TAAAKSE TAAAKSE TAAAKSE TAAAKSE TAAKSE TAAKSE TAAKSE TAAKSE TAAKSE | 2000 E I WAG E I WGG E I F RG E V F SG U V A T G L P F GL V A T G L P F GL V A T G T P WS V A T G T P WS L L F T N L L F T N V L V S MF - S S S S S S S S S S S S S S S< | 12
2 A K K A S V S - G
D A K K A A S - D
D A K K A A S - D
S - D A K A A S - D
S - A I I N K N S
E A A I I N K N S
E A A I I N K N S
E A A I I N K N S
E T A K A N A G S
E A G K L N A S
E A G K L N A S
E A G K L N A S
E A G K A A K S
S - S G K G V D N S K
- S G K G V D N S K
- S G K G V D N S K
- S G K G V D N S K
- S G K G V D N S K
- S G K G V D N S K
- V T K A A E N
- V T K A A E N
- V T K A A E N
- V V V G E A A S D | ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GN EASQ FAFQ
GN EASQ FAFT
GN EASQ FAFT | 2020
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKUIVTGW
TIKWIVTGW
TIKWIVTGW
AMMMILVTGW
AMMMILVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMRHIVTGW
AMRHIVTGW
ALRTFILIGW
ALRTFILIGW
ALRTFILIGW | 100
A I YP I GY L YG
A I YF I GY L YG
A I YP I GY L YG
A I YF | 310
QGD
MDA - DA
MDA - DA
MDA - DA
NDMI - PG
LDE - TT I LG
LDE - TT I LG
LEA - TT I LG
LEA - TT I LG
GG
YFG - DG
YFG - GG
YFG - GG
YHM - GG
YLM - GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
YLM
GG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG | 1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200 | 300
DNL - N I U YN I
DNL - N I U YN I
DNL - N I U YN I
DNL - N I U YN I
DDM - N I YN I
E L - N I YN L
E L - N I YN L
E A I - N I YN L
D A L - N I YN L
D A L - N I YN L
E A I - N I YN L
E I - N I YN I YN I YN I YN I YN I YN I Y |
| Po-2 HEID 3D09
Epi-28 MedDCM-JUL2D12-CL059
Epi-28 MedDCM-JUL2D12-CL059
Epi-28 MedDCM-JUL2D12-CL059
Po-4 HF70 59C08
Epi-28 RenS9398-C478
MedDCM-OCT-324-C2
Epi-38 RenS9398-C478
MedDCM-OCT-324-C2
Epi-38 RenS9398-C478
MedBCM-OCT-334-C2
Epi-38 RenS9398-C478
MedBCM-OCT-334-C2
Epi-38 RenS9398-C478
MedBCM-OCT-324-C2
Epi-38 MedBCM-OCT-334-C2
By amma HeiDCADCE99 (8)
gamma EBAC20E09 (8)
gamma EBAC20E09 (8)
Gamma HeiDCADCE92 (6)
Capallique HCC052 (6)
Pelagibacter sp. HTCC7211(8)
gamma HCC207 (6)
G-pallique HCC052 (6)
Pelagibacter sp. HTCC7211(8)
gamma HCC207 (6)
G-pallique HCC052 (6)
Po-2 HFID 3D07 LE5
Po-2 HFID 3D07 LE5
Evijoum 25-27
Evijoum 25-27
Evijoum 25-27
Po-2 HFID 3D09
Epi-18 MedBCM-UCT-304-C3
Po-3 HFID 309
Epi-18 MedBCM-UCT-304
Evijoum 25-27
Epi-38 MedBCM-UCT-305-C1
Epi-38 MedBCM-UCT-304-C1
Epi-48 HS93993-C1654
MedDCM-UCT-304-C3
Epi-48 HS9393-C1654
MedDCM-UCT-304-C3
Epi-48 HS9393-C1654
MedDCM-UCT-304-C3
Epi-48 HS9393-C1654
MedDCM-UCT-304-C3
Epi-48 HS9393-C1654
MedDCM-UCT-304-C3
Epi-48 HS9393-C1654
MedDCM-UCT-304-C3
Epi-48 HS9393-C1654
MedDCM-UCT-304-C3
Epi-48 HS9394-C330
MedDCM-UCT-304-C3
Epi-48 HS9394-C330
MedDCM-UCT-304-C3
Epi-48 HS9394-C330
MedDCM-UCT-304-C3
Epi-48 HS9394-C330
MedDCM-UCT-304-C3
Epi-48 HS9394-C330
MedDCM-UCT-304-C3
Epi-48 HS9394-C | 2200
1 S A 1 S | 282 D GM WWP LF AV D GM WWP LF AV AT I AW A I T I AW A I S I VWW T L V WW T I V WW T I AMY A F I I AWP A F I I AWA A F I I AWA A F I I AWW A F I I AWW W T I AKMA F I I T T L G I V I AKMA F I I T L A L AW W W T I A KM W W T I A W W W T I A W W W T I A MW W T I A MW W T I A MW W T I MMW W T I MMW W T I MMW W T I
 | 2050 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWIYIIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMICWFYIY GMICWFYIY GMAWIYIY GMAWIYIY GMAWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYY GWAGWIYY GWAGWIYY GWAGWIYY GWAGWIY GGUCWAMIMI GGVCWAMIMI GGVCWAMI GGUCWAMIY GUC

 | 2000 E I WAG E I F RG VAT GI F WS L I F TN L I F TN L I F TN E I I F I I I I I I I I I I I I I I I I
 | 12
2 A K K S V S - G
D A K K A A S - D
D A K K A A S - D
D A K K A A S - D
D V K E A S A - S
S I N D A A G
E A A I A N N S
E A S I N N N S
E A S I I N N S
E A A I A N A G
E T A K A M A G
E T A K A M A G
E T A K A M A G
E A A G I N N S
E A G K A M S - T
E G A G A M A K S
E A G K S A C N - T
E G A G N M A K S
E A S G V D S K
S G G V D N S K
- S G G C V D S K
- V K K A A E T
- V V G E A A S D | ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN AASQ QAFK
GN AAAQ GAFA
GN AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA | 2020
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
ALKWIVTVGW
SLKWIVTVGW
SLKWIVTVGW
TIKLIVTGW
TIKUVTYGW
TIKUVTYGW
TIKUVTYGW
ALKWIVTGW
AMMMILVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
AMMMIVVTGW
ALKTFILVGW
ALRTFILVGW
ALRTFILVGW
ALRTFILVGW
ALRTFILVGW | 10
A I YP I GY L YG
A I YP I GY L GY L GY
A I YP I GY L GY
A I YP I GY L GY
A I YP I GY L GY
A I YP L GY L GY
A | 310 QGD MDA - DA NDMI - FG LDE - TT I LG LLE - TT I LG YFG - DG YFG - GG YLM - GG YLM - GG YLA - GG YLT - GG YFG - UL - GG YLA - GG YCA I LD PEV - AL LFA - YG LFA - YG LFA - YG TEG - LIGLVG
 | 1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200 | 300
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DDM - N YN
E L - N YN
S L - N YN
S N - N YN
E L - N YN
S N - N YN
E L - N YN
E L - N YN
S N - N YN
S N - N YN
D - N YN
S N - N YN
S N - N YN
S N - N YN
E L - N YN
D - N YN YN YN
D - N YN YN YN YN YN YN YN |
| Pop-2 HEID 3D09
Ep-28 MedDCA-UL2012-CL059
Ep-28 MedDCA-UC207-C4997
Pop-3 HF0 19812
Pop-4 HF0 19812
Pop-4 HF0 19812
Pop-4 HF0 19812
Pop-4 HF0 19812
Ep-12 8 (RH39898-C478
MedDCA-UC7-324-C2
Ep-13 ERK39898-C478
MedDCA-UC7-324-C2
Ep-13 MedDCA-UC7-334-C2
Ep-13 MedDCA-UC7-334-C3
Ep-13 MedDCA-UC7-334-C3
Barma EBAC2050 (8)
garma HeI 75m4 (8)
garma EBAC2050 (8)
garma HeI 75m4 (8)
MedDCA-UC207 (6)
D-donghaenis MED134 (6)
MedDCA-UC207 (6)
D-donghaenis MED134 (6)
MedDCA-UC207 (6)
D-donghaenis MED134 (6)
MedDCA-UC207 (7)
D-donghaenis MED134 (6)
MedDCA-UC207-C1578
Pop-1 MedDCA-UC7-39-C58
Ep-18 MedDCA-UC7-392-C58
Pop-1 HeID 3099
Ep-12 BIN359895-C1534
MedDCA-UC207-C4997
Pop-3 HF0 1912
Pop-3 HeID 2009-C159
Dop-3 HeIDCA-UC7-32-C38
Pop-3 HeIDSCA-C378
MedDCA-UC207-C4997
Pop-3 HeIDSCA-C378
MedDCA-UC207-C4997
Pop-3 HF0 1912
Pop-3 HF0 1912
Pop | 220
1 S A S | 240 D GM WVP F E AV D GM WVP F E AV AT I AV A I T I AV A I S I VWV - T L S I VWV - T L S I VWV - T L MV G VI MV G VI AMP A F I I AM G G I I AM G G I I AG GW AG I L GY A G G I I AG GW AG I L A GW AG A TMD MMV WY A TM A K ND
 | 3050 GMLAWIYIIY GMLAWIYIIY GMLAWIYIIY GMLAWIYIIY GMLAWIYIIY GMLAWIYIIY GMLAWIYIIY GMLAWIYIIY GMLAWIYIIY GMLCWFYIIY GMLCWFYIIY GMLAWIYIIY GMLAWIYIY GWLAWIYIY GWLAWIYIY GMLAWIYIY GWLAWIYIY GULAWIYIY SULAWIYIY SULAWIYIY SULAWIYIY SULAWIYIY SULAWIYIY SULAKIY
 | 1000 E I WAG E I F RG VA T G T PWS VAT G T PWS VAT G T PWT L I F TN VAT G T PWT VAT G T PWT W I V I F R I F R I F R I F R I F R I F R I F R </th <th>12
2 A K E S V S - G
D A K K A A S - D
D A K K A A S - D
S - N C A A C - V
S - S - N C A C - V
E A A I N K N S
E T A K A M A G S
E T A K A M A G S
E T A K A M A G S
E T A K A M A G
E T A K A M A G
E T A K A M A G
E A G A C A C - T
E G A G A C A S
E A G A C A C - T
E G A G A C A S
E A G C A C A S
E A G C A C A S
C - Y C A C A S
C - Y K A A E A S
C - Y K A A E A
V V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y
V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V C - Y - Y</th> <th>ASE GUY CAPAFT
ASE GUY CAPAFT
ASE GUY CAPAFT
ASE GUY CAPAFT
GS DC UY CAPAFT
GS DC UY CAPAFT
GS DC UY CAPAFT
GN EAG CAPAFQ
GN EAG</th> <th>20
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
MMTYILIFGW
AMRAIVTGW
SLKWIVTYGW
SLKWIVTGW
SLKWIVTGW
TIKNIVTGW
TIKNIVTGW
TIKNIVTGW
TIKNIVTGW
TIKNIVTGW
AMMMILVTGW
AMMMILVTGW
AMMMILVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNITTGW
AMMNIVTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW</th> <th>200
AIYPIGYIGYIG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYUGYIG
AIYPIGYUGYIG
AIYPIGYUGYIG
AIYPIGYUGYIG
AIYPIGYUGYIG
AIYPIGYUGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG</th> <th>310 QGD MDA - DA NDMI - FG LDE - TT I LG YFG - GG YHG - GG YLG - GG YLM - GG YLM - GG YHT - GG YHT - GG YFG - GG YFG - GG YLA - GG YFT - GG PGAELGL PGAELGL PGAELGL PEV - AL LFA - TG TEG - LGLVG</th> <th>1200
0 G G T</th> <th>300
DNL - N U YN
DNL - N U YN
NML - N U YN
DJM - N U YN
DJM - N U YN
DJM - N U YN
E L - N U YN
E L - N U YN
DAL - N U YN
CAL - N U YN
SSL - N U YN
SSL - N U YN
SSL - N U YN
SSL - N U YN
SSL - N U YN
SSL - N U YN
CH - A CAL SAN
SSL - N U YN
CH - A CAL SAN
CAL SAN U YN
DAL - N YN
DAL - N YN
DAL - N YN
DAL - N YN
SSL - N U YN
CAL SAN YN YN
CAL SAN YN </th> | 12
2 A K E S V S - G
D A K K A A S - D
D A K K A A S - D
S - N C A A C - V
S - S - N C A C - V
E A A I N K N S
E T A K A M A G S
E T A K A M A G S
E T A K A M A G S
E T A K A M A G
E T A K A M A G
E T A K A M A G
E A G A C A C - T
E G A G A C A S
E A G A C A C - T
E G A G A C A S
E A G C A C A S
E A G C A C A S
C - Y C A C A S
C - Y K A A E A S
C - Y K A A E A
V V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V G E A A S
C - Y V V C - Y - Y | ASE GUY CAPAFT
ASE GUY CAPAFT
ASE GUY CAPAFT
ASE GUY CAPAFT
GS DC UY CAPAFT
GS DC UY CAPAFT
GS DC UY CAPAFT
GN EAG CAPAFQ
GN EAG
 | 20
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
MMTYILIFGW
AMRAIVTGW
SLKWIVTYGW
SLKWIVTGW
SLKWIVTGW
TIKNIVTGW
TIKNIVTGW
TIKNIVTGW
TIKNIVTGW
TIKNIVTGW
AMMMILVTGW
AMMMILVTGW
AMMMILVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNIVTGW
AMMNITTGW
AMMNIVTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW
AMMNITTGW | 200
AIYPIGYIGYIG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYIYG
AIYPIGYUGYIG
AIYPIGYUGYIG
AIYPIGYUGYIG
AIYPIGYUGYIG
AIYPIGYUGYIG
AIYPIGYUGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG
AIYPIGYIGYIG | 310 QGD MDA - DA NDMI - FG LDE - TT I LG YFG - GG YHG - GG YLG - GG YLM - GG YLM - GG YHT - GG YHT - GG YFG - GG YFG - GG YLA - GG YFT - GG PGAELGL PGAELGL PGAELGL PEV - AL LFA - TG TEG - LGLVG
 | 1200
0 G G T | 300
DNL - N U YN
DNL - N U YN
NML - N U YN
DJM - N U YN
DJM - N U YN
DJM - N U YN
E L - N U YN
E L - N U YN
DAL - N U YN
CAL - N U YN
SSL - N U YN
SSL - N U YN
SSL - N U YN
SSL - N U YN
SSL - N U YN
SSL - N U YN
CH - A CAL SAN
SSL - N U YN
CH - A CAL SAN
CAL SAN U YN
DAL - N YN
DAL - N YN
DAL - N YN
DAL - N YN
SSL - N U YN
CAL SAN YN |
| Pop 2 HED 3D09
Epi-28 MedDOM-UL3D12-CL059
Epi-28 MedDOM-OCT307-C697
Pop 3 HF70 19508
Epi-28 ENR59898-C478
MedDOM-OCT524-C2
Epi-28 ERR59898-C478
MedDOM-OCT524-C2
Epi-28 ERR59898-C478
MedBOA-OCT524-C2
Epi-28 ERR59898-C478
MedBOA-OCT524-C2
Epi-28 ERR59898-C478
MedBOA-OCT524-C2
Epi-28 ERR59898-C478
MedBOA-OCT524-C2
Epi-28 ERR59898-C478
MedBOA-OCT524-C2
Epi-28 ERR59898-C478
MedBOA-OCT524-C2
Gapailloula (B)
Daonghaenis MED134 (G)
Daonghaenis MedDA-OCT-324-C2
E-13 ERR59893-C478
MedDA-OCT-324-C2
Epi-3 ERR59893-C478
MedDA-OCT-324-C2
Epi-3 ERR59893-C478
MedBACAC005
HT40 4650
Earmam EEAC2005 (H)
SAR865 (G)
Publique HCTC1625 (G) | 2200 220 220 220 220 220 220 220 220 22 | 240 D GM WVP LF AV D GM WVP LF AV AT I AV A I T E I WW T L S I WW T L VV WW T L VV WW T L MV W A F I I AM A F I P T L G F V I AKM A F I P T L G F V I AV WW A A T - D MV WA A T T D MM VW A A T MD MM WW A A T MD <t< th=""><th>253 GML AWIYIIY GMA CWFYIYIY GMA CWFYIYIY GMY GMA CWFYIY GMA CWFYIYIY GG CWMAHIM- GG CWMAHIM- GG CWMAHIM- GG CWMAHIM- GG CWMAHIY C C LAWFYHIYY ST LAMAS E T KAAAS E T AAAES C T AAAES C T AAAES C T AAAES C T KAAAS E T AAAES C T KAAES C T KAAES C T AAES C T KAS C T KAS C T KAS C T KAAES C <td< th=""><th>Bit Mage E I WAG E I WGG E I F G E I F RG E V F SG V A T GT F WS L L F T N L L F T N C C C C C C C C C C C C C C C C C C C</th><th>12
2 A K K A S V S - G
D A K K A A S - D
D A K K A A S - D
S - D A K A A S - D
S - A I I N K N S
E A A I I N K N S
E A A I I N K N S
E A A I I N K N S
E T A K A M A A S
E T A K A M A G
E A G A K A A S - T
E G K A A K A
E A G C V N S K
E A G K A A K A
S - S G K Q V M S K
- S G K G V M S K
- S G K K M A K S V
- S G K K M A K S V
- S K K K A K A K S V
- S K K K A K A K S V
- S K K K A K A K S V
- S K K K A K A K S V
- S K K K A K S V K S K K A K A K S V
- S K K K A K S V K S K K S K S K S K K S K S K K S K S</th><th>ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GN EAS GQ FAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GAFA
GN AAS QAFK
GN AAS QAFK
GN AAS QAFA
GN AS CAN GAFA
GN AS CAN GAF</th><th>2010
AMALILITYGW
WMTYILIFGW
AMRAIVIFGW
AMRAIVIFGW
AMRAIVIFGW
SLKWIVTIGW
SLKWIVTIGW
SLKWIVTIGW
SLKWIVTIGW
TIKUIVTGW
TIKUIVTGW
TIKUIVTGW
AMMMIIVTGW
AMMMIVTVGW
AMRMIVTGW
AMMMIVTVGW
AMRMIVTGW
AMRMIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
ALRTFILIGW
ALRTFILIGW
AMRHIVTGW</th><th>100
A I YP I GY L YG
A I YP I GY L YG
I YP I GY L YG
A I YF I GY L Y</th><th>310
QGD
MDA - DA
MDA - DA
MDA - DA
NDMI - PG
LDE - TTILG
LDE - TTILG
LEA - TTILG
LEA - TTILG
CHART
YFG - DG
YFG
YFG
YFG
CA
YFG
CA
YFG
CA
TTILG
LA
TTIG
LA
TTIG
LA
TTIG
LA
TTIG

YFG
PG
TTIG

YFG
CA
TTIG

YFG
CA
TTIG

YFG
CA
TTIG

YFG
CA
TTIG

YFG
CA
TG

YFG
CA
TG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

</th><th>1200
</th><th>300
DNL - N U YN
DNL - N U YN
NML - N U YN
DDL - N U YN
DDL - N U YN
DDL - N U YN
E L - N U YN
E L - N U YN
DDL - N U YN
E A A N YN
DDL - N U YN
E A A N YN
S S D YN
S S D YN
S S D YN
S S D YN
S A N YN
E A A U YN
C U A U YN U YN U YN
C U A U YN U YN U U YN
C U A U U YN U U U U U U U U U </th></td<></th></t<> | 253 GML AWIYIIY GMA CWFYIYIY GMA CWFYIYIY GMY GMA CWFYIY GMA CWFYIYIY GG CWMAHIM- GG CWMAHIM- GG CWMAHIM- GG CWMAHIM- GG CWMAHIY C C LAWFYHIYY ST LAMAS E T KAAAS E T AAAES C T AAAES C T AAAES C T AAAES C T KAAAS E T AAAES C T KAAES C T KAAES C T AAES C T KAS C T KAS C T KAS C T KAAES C <td< th=""><th>Bit Mage E I WAG E I WGG E I F G E I F RG E V F SG V A T GT F WS L L F T N L L F T N C C C C C C C C C C C C C C C C C C C</th><th>12
2 A K K A S V S - G
D A K K A A S - D
D A K K A A S - D
S - D A K A A S - D
S - A I I N K N S
E A A I I N K N S
E A A I I N K N S
E A A I I N K N S
E T A K A M A A S
E T A K A M A G
E A G A K A A S - T
E G K A A K A
E A G C V N S K
E A G K A A K A
S - S G K Q V M S K
- S G K G V M S K
- S G K K M A K S V
- S G K K M A K S V
- S K K K A K A K S V
- S K K K A K A K S V
- S K K K A K A K S V
- S K K K A K A K S V
- S K K K A K S V K S K K A K A K S V
- S K K K A K S V K S K K S K S K S K K S K S K K S K S</th><th>ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GN EAS GQ FAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GAFA
GN AAS QAFK
GN AAS QAFK
GN AAS QAFA
GN AS CAN GAFA
GN AS CAN GAF</th><th>2010
AMALILITYGW
WMTYILIFGW
AMRAIVIFGW
AMRAIVIFGW
AMRAIVIFGW
SLKWIVTIGW
SLKWIVTIGW
SLKWIVTIGW
SLKWIVTIGW
TIKUIVTGW
TIKUIVTGW
TIKUIVTGW
AMMMIIVTGW
AMMMIVTVGW
AMRMIVTGW
AMMMIVTVGW
AMRMIVTGW
AMRMIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
ALRTFILIGW
ALRTFILIGW
AMRHIVTGW</th><th>100
A I YP I GY L YG
A I YP I GY L YG
I YP I GY L YG
A I YF I GY L Y</th><th>310
QGD
MDA - DA
MDA - DA
MDA - DA
NDMI - PG
LDE - TTILG
LDE - TTILG
LEA - TTILG
LEA - TTILG
CHART
YFG - DG
YFG
YFG
YFG
CA
YFG
CA
YFG
CA
TTILG
LA
TTIG
LA
TTIG
LA
TTIG
LA
TTIG

YFG
PG
TTIG

YFG
CA
TTIG

YFG
CA
TTIG

YFG
CA
TTIG

YFG
CA
TTIG

YFG
CA
TG

YFG
CA
TG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

</th><th>1200
</th><th>300
DNL - N U YN
DNL - N U YN
NML - N U YN
DDL - N U YN
DDL - N U YN
DDL - N U YN
E L - N U YN
E L - N U YN
DDL - N U YN
E A A N YN
DDL - N U YN
E A A N YN
S S D YN
S S D YN
S S D YN
S S D YN
S A N YN
E A A U YN
C U A U YN U YN U YN
C U A U YN U YN U U YN
C U A U U YN U U U U U U U U U </th></td<> | Bit Mage E I WAG E I WGG E I F G E I F RG E V F SG V A T GT F WS L L F T N L L F T N C C C C C C C C C C C C C C C C C C C | 12
2 A K K A S V S - G
D A K K A A S - D
D A K K A A S - D
S - D A K A A S - D
S - A I I N K N S
E A A I I N K N S
E A A I I N K N S
E A A I I N K N S
E T A K A M A A S
E T A K A M A G
E A G A K A A S - T
E G K A A K A
E A G C V N S K
E A G K A A K A
S - S G K Q V M S K
- S G K G V M S K
- S G K K M A K S V
- S G K K M A K S V
- S K K K A K A K S V
- S K K K A K A K S V
- S K K K A K A K S V
- S K K K A K A K S V
- S K K K A K S V K S K K A K A K S V
- S K K K A K S V K S K K S K S K S K K S K S K K S K S | ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GN EAS GQ FAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GAFA
GN AAS QAFK
GN AAS QAFK
GN AAS QAFA
GN AS CAN GAFA
GN AS CAN GAF | 2010
AMALILITYGW
WMTYILIFGW
AMRAIVIFGW
AMRAIVIFGW
AMRAIVIFGW
SLKWIVTIGW
SLKWIVTIGW
SLKWIVTIGW
SLKWIVTIGW
TIKUIVTGW
TIKUIVTGW
TIKUIVTGW
AMMMIIVTGW
AMMMIVTVGW
AMRMIVTGW
AMMMIVTVGW
AMRMIVTGW
AMRMIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
ALRTFILIGW
ALRTFILIGW
AMRHIVTGW | 100
A I YP I GY L YG
A I YP I GY L YG
I YP I GY L YG
A I YF I GY L Y | 310
QGD
MDA - DA
MDA - DA
MDA - DA
NDMI - PG
LDE - TTILG
LDE - TTILG
LEA - TTILG
LEA - TTILG
CHART
YFG - DG
YFG
YFG
YFG
CA
YFG
CA
YFG
CA
TTILG
LA
TTIG
LA
TTIG
LA
TTIG
LA
TTIG

YFG
PG
TTIG

YFG
CA
TTIG

YFG
CA
TTIG

YFG
CA
TTIG

YFG
CA
TTIG

YFG
CA
TG

YFG
CA
TG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

YG

 | 1200
 | 300
DNL - N U YN
DNL - N U YN
NML - N U YN
DDL - N U YN
DDL - N U YN
DDL - N U YN
E L - N U YN
E L - N U YN
DDL - N U YN
E A A N YN
DDL - N U YN
E A A N YN
S S D YN
S S D YN
S S D YN
S S D YN
S A N YN
E A A U YN
C U A U YN U YN U YN
C U A U YN U YN U U YN
C U A U U YN U U U U U U U U U |
| Po-2 HEID 3D09
Epi-28 MedDCM-JUL2D12-CL059
Epi-28 MedDCM-JUL2D12-CL059
Epi-28 MedDCM-JUL2D12-CL059
Po-9 HF7D 19502
Po-9 HF7D 19502
Epi-28 ERK93983-C478
MedDCM-OCT-324-C2
Epi-38 ERK93983-C478
MedDCM-OCT-324-C2
Epi-38 MERS93983-C478
MedEAL-04503
Bamma HeID CAT-334-C2
Epi-38 MERS93983-C478
MedEAL-04503
Bamma HeID CAT-334-C2
Epi-38 MERS9398-C478
MedEAL-04503
Bamma HeID CAT-346-C2
Pi-38 MedDCM-OCT-34-C2
MedDCM-OCT-324-C2
Bamma HeID CAT-526-C3
MedDCM-OCT-526-C3
MedDCM-OCT-526-C3
MedDCM-OCT-526-C3
Po-2 MED 3D09
Epi-28 MedDCM-JUL2D12-C1059
Epi-28 MedDCM-JUL2D12-C1059
Epi-18 MedDCM-JUL2D12-C1 | 2200 220 220 220 220 220 220 220 220 22 | 2262 D GM WVP LF AV T I AW A I AT I AW A I AT I AW A I T I WW T L S I WW T L V WW T L V WW T L V WW T L AM A F I I AWP A F I I AWP A F I I AWP A F I I AW W W - A T I D AWW W A A T M D MW W A A T M D MM W W A A T M D
 | 3030 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWIYIIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMICWFYIY GMAWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GCLAWFWIY GCLAWFWIY GGLAWFWIY GGLAWFWIY GGLAWFWIY GGCAWWY GGCAWAMIM GCVGWAMIM GCGWWAMIY GCFAWIYIY GCGWAMIMI GGVGWAMI GGVGWAMI GGVGWAMI GCFAWIYIY GC
 | 2000 E I WAG E I F RG VAT G I F WS L I F TN L I F TN L I F TN E I F RG E I

 | 12
2 A K K S V S - G
D A K K A A S - D
D A K K A A S - D
D A K K A A S - D
D V K E A S A - S
S I M D A A G
E A A I A N N S
E A S I I N N S
E T A K A M A G S
E T A K A M A G S
E T A K A M A G
E T A K A M A G
E T A K A M A G
E A S G N M A K
S G K S A C N - T
E G A G A M A K
S G K C V D S K
- S G K C V D S K
- V K K A A E T
- V V G E A A S D | ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN AASQ QAFK
GN AAAQ GAFA
GN AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA | 2010
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTYGW
SLKWIVTYGW
TIKIVTYGW
TIKIVTYGW
TIKIVTYGW
TIKIVTYGW
AMMMILVTGW
AMMMILVTGW
AMMMILVTGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
AMMMIVTYGW
ALRTFILVGW
ALRTFILVGW
ALRTFILVGW
ALRTFILVGW | 10
A I YP LG Y LG
A I YP LG | 310 QGD MDA - DA NDMI - FG LDE - TT I LG LLE - TT I GG YFG - OGG YLM - GG YLM - GG YLM - GG YLA - GG YLT - GG YEG - LGLVG
 | 1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200 | 300
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DDM - N YN
E L - N YN
S L - N YN
S N - N YN
E L - N YN
S N YN
S N - N YN
S N |
| Pop-2 HEID 3D09
Ep-13 MedIOCA-UL2012-CL059
Ep-13 MedIOCA-UL2012-CL059
Pop-3 HF70 15912
Pop-3 HF70 15908
Ep-12 RENS9939-CL56
MedIOCA-OCT-207-C497
Pop-3 HF70 15908
Ep-13 RENS9939-CL53
MedIOCA-OCT-324-C2
Ep-13 RENS9939-CL53
MedIOCA-OCT-324-C2
Ep-13 MedIOCA-OCT-334-C5
Ep-13 MedIOCA-OCT-334-C5
Barma ERAC2050 (8)
garma HeIT-57n4 (8)
garma HIT-57n4 (8)
garma HIT-5704 (8)
garma HIT-5704 (8) | 2200 220 220 220 220 220 220 220 220 22 | 240 D GM WVP LF AV D GM WVP LF AV AT I AV A I T I AW A I T I AW A I S I WW A I S I WW I VWW I V WW I V WW I WW A F I AMP A F I AT L G F VI AT L G GW AG L L GY L G L L L T Q L G L WS VL I AG WA A T MD MM VW A A T MD A WW A A T MD A W A A T MA A WKD A W A A M K A M K D A W A A T MD<
 | 3050 GMLAWIYIIY GMLCWFYIIY GMLCWFYIIY GMLAWIYIY GWLAWIYIY GUVAMINIY GUVAMIY GUVAMIY GUVAMIY GUVAMIY GUVAMIY TAAAESKE TAAAESKE TAAAESKE TAAAESKE TAAAESKE TKAAASER SEASYA SEASYA
 | 1000 E I WAG E I F RG VA T G T P WS VA T G T P WS VA T G T P WS VAT G T P WS VAT G T P WS VAT G T P WS WF G WF G WF G G G G G G

 | 12
2 A K K A S Y S - G
D A K K A A S - D
D A K K A A S - D
S - N A A A A S - D
S - N A A A A S - D
S - S - N A A A A A S - D
E A A I I N K N S
E A A I N K N S - T
E G A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A S - T
E G A A A A A A A A S - T
E G A A A A A A A A S - T
E G A A A A A A A A S - T
E G A A A A A A A A S - T
E G A A A A A A A A A A A A A A A A A A | ASE GUYQ FAFT
ASE GUYQ FAFT
ASE GUYQ FAFT
GS DC UYQ FAFT
GS DC UYQ FAFT
GS DC UYQ FAFT
GS DC UYQ FAFT
GN FAG QF FAFQ
GN AAS QQ FF
GN AAS QD F
GN AAS AA GN AAS AAS AAS AAS AAS AAS AAS AAS AAS AA | 2010
AMALILITYGW
WMTYILIFGW
AMTYILIFGW
AMTYILIFGW
AMTYILIFGW
AMTYILIFGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKUIVTGW
TIKUIVTGW
ALKVIVTGW
AMMYILIGW
AMMYILIGW
AMTYICFGU
TIRLIVTYGW
AMTYICFGU
TIRLIVTYGW
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTYICFGU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUSTU
AMTUST | 200
AIYPIGYLGGILG
AIYPIGYLYG
AIYPIGYLYG
AIYPIGYLYG
AIYPIGYLYG
AIYPIGYLYG
AIYPIGYLYG
AIYPIGYLYG
AIYPIGYLYG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG
AIYPIGYLG | 310 QGD
 | 1200
 | 300
DNL - N U YN
DNL - N U YN
NML - N U YN
DJM - N U YN
DJM - N U YN
DJM - N U YN
E L - N U YN
E L - N U YN
DJL - N U YN
E L - N U YN
DJL - N U YN
E L - N U YN
DJL - N U YN
E L - N U YN
DJL - N U YN U YN
DJL - N U YN U YN
DJL - N U YN U YN U YN
DJL - N U YN |
| Pop 2 HED 3D09
Epi-28 MedDOM-UL3D12-CL059
Epi-28 MedDOM-UC3D7-C697
Pop 3 HF70 19508
Epi-28 ERK95988-C478
MedDOM-OCT-324-C2
Epi-28 ERK959894-C530
MedDOM-OCT-524-C2
Epi-3 MedDOM-OCT-534-C2
Epi-3 MedDOM-OCT-534-C2
Epi-3 MedDOM-OCT-534-C2
Epi-3 MedDOM-OCT-534-C2
Epi-3 MedDOM-OCT-534-C2
Epi-3 MedDOM-OCT-534-C2
Epi-3 MedDOM-OCT-534-C2
Epi-3 MedDOM-OCT-534-C2
MedDOM-OCT-524-C2
Epi-3 MedDOM-OCT-525-C80
MedDOM-OCT-525-C80
MedDOM-OCT-525-C80
MedDOM-OCT-525-C80
MedDOM-OCT-525-C80
MedDOM-OCT-525-C80
Pop-3 He7D S005
E- JaBintom gan.21
H-marismortui ATCC 43049
Pop-3 He7D S005
Epi-18 RR95983-C1544
MedDOM-OCT-524-C2
Epi-18 RR95983-C1544
MedDOM-OCT-524-C2
Epi-3 HeR05983-C1544
MedDOM-OCT-524-C2
Epi-3 He7D 5908
Epi-38 MedDOS
Epi-38 MedDOS
DF1-28 ERK95983-C354
MedDOM-OCT-524-C2
Epi-3 He7D 5908
Epi-38 MedDOS
Epi-38 MedDOS
E | 2200 1 S A 1 S | 240 D GM WVP LF AV D GM WVP LF AV AT I AW A I T E I WW T L S I WW T L V WW T L V WW T L M Y WW A I P M P A F I I AM A F I P T L G F V I AKM A F I P T W W A A I P T L G G L - V L A G L W A T I D M W W A A T M D MM WW A A T M A W K D A I W A A W A M A W K D A I W A A W A M A W A A W A D A I W A A W A A W A A W A A M A M M D
 | 253 GML AWIYIIY GMA CWFYIY GG CWAMIM GG CWAMIM GG CWAMIMIY GMA AS E T AAA & SC E T KAA & SC E
 | 1000 E I WAG E I WGG E I WGG E I F G E I F RG E V F SG V A T G T F WS V A T G T F WT L L F T N L L F T N V A T G T F WT L L F T N C C C C C C C C C C C C C C C C C C C

 | 12
E AKE SVS - G
D AK KAAS - E
D AK KAAS - E
E AA I IN KNS
E AA I IN KNS
E AA I IN KNS
E TAK AN AGS
E AG KANAKS
E AG KAAKS
E AG KAAKS
E AG KAAKS
E AG KAAKS
E AG KAAKS
- S G K G VONSK
- | ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GN FAG GAFA
GN FAG | 2010
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKWIVTGW
TIKWIVTGW
TIKWIVTGW
AMMMILVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMFIVGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
AMMRITUTGW | 100
A 1 Y PL G F L L G
A 1 Y PL G Y L Y G Y L Y G
A 1 Y PL G Y L Y G Y L Y G Y L Y G Y L Y G
A 1 Y PL G Y L Y G Y L Y G Y L Y G Y L Y G Y L Y | 310
QGD
MDA - DA
MDA - DA
MDA - DA
NDMI - PG
LDE - TTILG
LDE - TTILG
LEA - TTILG
LEA - TTILG
CHART
YFG - DG
YFG - GG
YFG - GG
YFG - GG
YIM - GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
YIM
GG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
CG
 | 120
120
120
120
120
120
120
120 | 300
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DDM - N U YN
DDM - N U YN
E L - N U YN
E A - N U YN
E A - N U YN
DA - N U YN
E A - N U YN
DA - N U YN
E - N U YN
E - N U YN
DA - N U YN
E - N U YN
C |
| Po-2 HF10 3D09
Ep-2 B MedDCM-JUL2012-CL059
Ep-1 BMR30293-CL654
Ep-1 BMR30293-CL654
Po-9 HF70 59C08
Ep-1 BRR302938-C478
MedDCM-OCT-324-C2
Ep-1 S RR503983-C478
MedDCM-OCT-324-C2
Ep-1 ABR02DCM-OCT-334-C2
Ep-1 ABR02DCM-OCT-334-C2
Ep-1 ABR02DCM-OCT-334-C2
Ep-1 ABR02DCM-OCT-334-C2
Ep-1 ABR02DCM-OCT-334-C2
Ep-1 ABR02DCM-OCT-334-C2
Ep-1 ABR02DCM-OCT-334-C2
Ep-1 ABR02DCM-OCT-334-C2
Ep-1 ABR02DCM-OCT-334-C2
Palagibacter sp. HTCC7211(8)
gamma HC2C0E20 (6)
Palagibacter sp. HTCC7211(8)
gamma HC2C0C2 (c)
Palagibacter sp. HTCC7211(8)
gamma HC2C0C7 (c)
G-palifuldu (6)
D-donghaensis MED134 (6)
MedDCM-OCT-324-C2
Egipobacterium sp. AT1b
1. marismortui ATCC 43049
Po-2 HF10 3D09
Ep-12 BKR59893-CL654
MedDCM-OCT-324-C2
Ep-13 MedDCM-OCT-344-C2
Ep-13 MedDCM-OCT-34-C2
Ep-13 MedDCM-OCT-34-C2
Ep-14 | 2200 1 S A 1 S | 2262 D GM WWP LF AV T I AW A I AT I AW A I AT I AW A I T I AW A I S I WW T L VWW T L VVWW T L VVWW T L VVWW T L AM V G F V AM G VI AW A F I I AW A F I I AW G VI AW M G VI AW M G VI AW W W - A T I AW W W - A T I MM WW A A T M
 | 2050 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWIYIIY GMACWIYIIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMICWFYIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GCLAWFYIW GCLAWFYIW GCLAWFYIW GCLAWFWIY GGUCWAMIM GCVGWAMIM GCVGWAMIM GCVGWAMIMI GCFAWIYIY GCFAWIYIY GCYGWAMIMI GCGVGWAMI GCFAWIYIY GCFAWIYIY GCFAWIYIY GCFAWIYIY GCFAWIY GCFAWIYIY GCFAWIYIY GCFAWIYIY GCFAWIYIY GCFAWIYIY GCFAWIYIY GCFAWIYIY GCFAWIYIY
 | 1000 E I WAG E I F RG VA T G T F WS LL F TN LL T TN E I F RG I I I T N E I I

 | 12
E AKE SVS - G
D AK KAAA - D
D AK KAAA - D
S I M D AAG K
E AA I N N NS
E TAK AN AG S
E AA Q A A A S T
E G AQ AAK NS - T
E G AQ AAK NS - T
E G AQ AAK AS - T
Y DG K G V DN SK
- S G K G V DN SK
- S G K G V DN SK
- V VK FAA E T
V V G E AAS D | ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN AASQ QAFK
GN AAAQ QAFK
GN AAAQ GAFN
GN AA | 2020
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
ALKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKIVTGW
TIKIVTGW
TIKUVTGW
TIKIVTGW
ALKWITGW
AMMMILVTGW
AMMMILVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
ALKTFILIGW
ALKTFILIGW
ALKTFILIGW
ALKTFILIGW
ALKTFILIGW | 10
A I YP LG Y LG
A I YP LG | 310
QGD
 | 1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200 | 300
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DDM - N YN
E L - N YN
S L - N YN
S N - N YN
E L - N YN
S N |
| Pop-2 HEID 3D09
Epi-28 MedioCA-UL2012-CL059
Epi-28 MedioCA-UL2012-CL059
Pop-3 HF70 15912
Pop-3 HF70 15912
Pop-3 HF70 15912
Pop-3 HF70 15912
Pop-3 HF70 15912
Pop-3 HF70 15928
Epi-28 ERR593983-C478
MedioCA-OCT-33-C62
Epi-3 ERR593983-C478
MedioCA-OCT-33-C62
Epi-3 MedioCA-OCT-33-C62
Epi-3 MedioCA-OCT-33-C62
Epi-3 MedioCA-OCT-33-C62
Epi-3 MedioCA-OCT-34-C62
Epi-3 MedioCA-OCT-34-C62
Epi-3 MedioCA-OCT-34-C62
Epi-3 MedioCA-OCT-34-C62
Epi-3 MedioCA-OCT-34-C62
Epi-3 MedioCA-OCT-34-C62
Epi-3 MedioCA-OCT-36-C63
MedioCA-OCT-36-C63
MedioCA-OCT-36-C63
Pop-1 MedioCA-OCT-36-C63
Epi-28 MedioCA-UC1-207-C1678
Pop-3 HF70 15028
Fop-3 HF70 15028
Fop- | 2200 220 220 220 220 220 220 220 220 22 | 242 D GM WVP LF AV D GM WVP LF AV AT I AV A I T I AW A I S I WW A I S I WW II S I WW II MV A F II AWP A F II AWP A F II AWP A F II AMP A F II AMP A F II AMP A F II AMP A GU ANG ALL A GU AV AG ALL A GU AV AG ALL A GU AV ALL
 | 3050 GMLAWIYIIY GMLCWFYIIY GMLGWFYIIY GGVGWAMII- GGVGWAMIY GCIAWFYIIY SGALAFE GUVGWAMIY GUVGWAMIY GUVGWAMIY GUVGWAMIY SCIAWIY SCIAWIY GUVGWAMIY GUVGWAMIY GUNAKASE TAAAESKE TAAAESKE TKAAASEKE TKAAASEKE SSNA SSNA SSNA SSN
 | 1000 E I WAG E I F RG E V F SG E I F RG VA T G T P VS V V VA T G T P VS V V VA T G T VS V V V V

 | 12
2 A K K A S Y S - G
D A K K A A S - D
D A K K A A S - D
S - N K A A S - D
S - N K A A S - D
S - S - N S - S - S - S - S - S - S - S | ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GS DG VQ FAFT
GN FAG QF FAFQ
GN FAG QF FAFQ
GN FAG QF FAFQ
GN FAG QAFK
GN AAS QAFK
GN FAC US SAFN
GN FAC US SAFN
G | 2010
AMALILITGW
WMTYILIFGW
WMTYILIFGW
AMRAIVFGW
AMRAIVFGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKWIVTGW
TIKWIVTGW
ALKWIVTGW
MMMYILIGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRMIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
AMRHIVTGW
ALKFIVGW
ALRFIVGW
ALRFIVGW
MMRLFILIGW
VLNNILLGW | 100
A 1 Y PL G F L L G
A 1 Y PL G Y L Y G Y L Y G
A 1 Y PL G Y L Y G Y L Y G Y L Y G Y L Y G Y L Y G Y L Y G Y L Y G Y L Y | 310
GGD
MDA - DA
MDA - DA

YE
YE
YF
YF
YF
YH
YMT
YMT
TGA
PQA E I LD
PQA E I LD
FA
TE
TE
 | 1200
 | 300
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DML - N U YN
DDM - N U YN
E L - N U YN
DAL - N U YN
E N - N U YN
E M - A U YN |
| Pop 2 HED 3D09
Epi-28 MedDO- UL302-CL059
Epi-28 MedDOM-OCT307-C697
Pop 3 HF70 19508
Epi-28 ERK95983-C478
MedDOM-OCT34-C2
Epi-28 ERK95983-C478
MedDOM-OCT-34-C2
Epi-28 ERK95983-C478
MedDOM-OCT-34-C2
Epi-28 ERK95983-C478
MedBOA-OCT-34-C2
Epi-28 ERK95983-C478
MedBOA-OCT-34-C2
Epi-28 ERK95983-C478
MedBOA-OCT-34-C2
Epi-28 ERK95983-C478
MedBOA-OCT-34-C2
Epi-28 ERK9598-C478
MedBOA-OCT-34-C2
Gamma ERAC-2007 (6)
G-pailloula (8)
D-AngMaents MED134 (6)
D-AngMaents MED134 (6)
D-Pop - MedDOM-OCT-32-C20
Fi-12 ERK9598-C1544
MedDOM-OCT-324-C2
E-315-12 ERK95983-C1544
MedDOM-OCT-324-C2
Epi-12 ERK95983-C458
MedDOM-OCT-324-C2
Epi-12 ERK95983-C478
MedDOM-OCT-324-C2
Epi-13 MedDOM-OCT-324-C2
Epi-13 M | 2200 1 S A 1 S | 240 D GM WYP LF AV D GM WYP LF AV AT I AW A I T E I WW T I S I WW T I S I WW T I Y WW A I P MY UW A I P T L G F V I AM A F I P T L G F V I AW W G V I AW W W A T - D MU W A A T MD MW W A A T MD | 2050 GML AWIYIIY GMA CWFYIY GG CWAMIM GG CWAMIM GG CWAMIM GG CWAMIMIY GMA CWFYIY SG AAFFVIY GMA CWAMIY GMA CWAMIY GMA CWAMIY GMA CWAMIY GMA CWAMIY GMA AKAS E TAAAK SC TAAAK SC < | 1000 E I WAG E I WGG E I WGG E I F G E I F RG E V F SG E V F SG <th>12
E AKE SVS - G
D AK KAAS - E
D AK KAAS - E
E AA I IN KNS
E AA I IN KNS
E AA I IN KNS
E TAK AN AGS
E AG KANAKS
E AG KAAKS
E AG KAAKS
E AG KAAKS
E AG KAAKS
- S G K G V DN SK
- V TK AA EN
- V TK AA EN
- V TK G E AAS D</th> <th>ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GSD GAVG FAFQ
GN FAGQ FAFT
GN FAGQ FAGQ FAFT
GN FAGQ FAGQ FAFT
GN FAGQ FAGQ FAGT
GN FAGQ FA</th> <th>2010
AMALILITGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKUVTGW
TIKUTTGW
AMMMILVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMTULTGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
AMMRIVTGW
ALRTFILGW
AMMRIVTGW
AMMTULTGW
ALRTFILGW
AMMRIVTGW
AMMRIVTGW
AMMRIVTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMM</th> <th>100
A 1 Y PL G F L L G
A 1 Y PL G Y L Y G Y L Y G
A 1 Y PL G Y L Y G Y L Y G Y L Y G Y L Y G
A 1 Y PL G Y L Y G Y L Y G Y L Y G Y L Y G Y L Y</th> <th>310
QGD
MDA - DA
MDA - DA
MDA - DA
NDMI - PG
LDE - TTILG
LDE - TTILG
LEA - TTILG
LEA - TTILG
CHART
YFG - DG
YFG
YFG
GG
YIM
GG
YIM
GG
TTILG
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL</th> <th>1200
</th> <th>300
DNL - N U YN
NML - N U YN
NML - N U YN
DJM N U YN
DJM N U YN
DJM N U YN
E L - N U YN
E A A N YN
E A A N YN
S S D A U YN
S D A U YN U YN
S D A U YN
S D A U YN U YN U YN
S D A U YN U YN </th> | 12
E AKE SVS - G
D AK KAAS - E
D AK KAAS - E
E AA I IN KNS
E AA I IN KNS
E AA I IN KNS
E TAK AN AGS
E AG KANAKS
E AG KAAKS
E AG KAAKS
E AG KAAKS
E AG KAAKS
- S G K G V DN SK
- V TK AA EN
- V TK AA EN
- V TK G E AAS D | ASE GAVQ FAFT
ASE GAVQ FAFT
ASE GAVQ FAFT
GSD GAVQ FAFT
GSD GAVQ FAFT
GSD GAVG FAFQ
GN FAGQ FAFT
GN FAGQ FAGQ FAFT
GN FAGQ FAGQ FAFT
GN FAGQ FAGQ FAGT
GN FAGQ FA | 2010
AMALILITGW
WMTYILIFGW
WMTYILIFGW
AMRAIVTGW
AMRAIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
SLKWIVTGW
TIKUVTGW
TIKUTTGW
AMMMILVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMMIVTGW
AMMTULTGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
ALRTFILGW
AMMRIVTGW
ALRTFILGW
AMMRIVTGW
AMMTULTGW
ALRTFILGW
AMMRIVTGW
AMMRIVTGW
AMMRIVTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMMTULTGW
AMM | 100
A 1 Y PL G F L L G
A 1 Y PL G Y L Y G Y L Y G
A 1 Y PL G Y L Y G Y L Y G Y L Y G Y L Y G
A 1 Y PL G Y L Y G Y L Y G Y L Y G Y L Y G Y L Y | 310
QGD
MDA - DA
MDA - DA
MDA - DA
NDMI - PG
LDE - TTILG
LDE - TTILG
LEA - TTILG
LEA - TTILG
CHART
YFG - DG
YFG
YFG
GG
YIM
GG
YIM
GG
TTILG
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL
TGL | 1200
 | 300
DNL - N U YN
NML - N U YN
NML - N U YN
DJM N U YN
DJM N U YN
DJM N U YN
E L - N U YN
E A A N YN
E A A N YN
S S D A U YN
S D A U YN U YN
S D A U YN
S D A U YN U YN U YN
S D A U YN |
| Pop-2 HF10 3D09
Epi-28 MedDCM-JUL2012-CL059
Epi-28 MedDCM-JUL2012-CL059
Epi-28 EMISSION
Pop-3 HF70 19502
Pop-3 HF70 19502
Pop-3 HF70 19502
Epi-28 EMISSIONSSIC-478
MedDCM-OCT-324-C2
Epi-3 EMISSIONSSIC-478
MedBCM-OCT-334-C2
Epi-3 EMISSIONSSIC-478
MedBCM-OCT-334-C2
Epi-3 EMISSION
Pop-2 HF10 3D09
Epi-28 MedDCM-UCT-34-C2
Pollogue HTCC02 (6)
Pelagibacter sp. HTCC211(8)
gamma EBAC20E09 (8)
gamma EBAC20E07 (6)
G-pallidula (8)
D-donghaenis MED134 (6)
D-donghaenis MED134 (6)
MedDCM-OCT-324-C2
E-30MiG G3
MedDCM-OCT-336-C3
E-30MiG G3
MedDCM-UL2012-C1598
Epi-28 MedDCM-UL2012-C1059
Epi-28 MedDCM-UL2012-C1059
Epi-28 MedDCM-UL2012-C1059
Epi-28 MedDCM-UL2012-C1059
Epi-28 MedDCM-OCT-34-C2
E-30MiG G3
MedDCM-OCT-324-C2
Epi-38 MedDCM-OCT-34-C2
Epi-38 MedDCM-OCT | 2200 1 S A 1 S | 282 D GM WVP LF AV T I AW A I AT I AW A I AT I AW A I AT I AW A I T I AW A I T I AW A I T I AW A I S I WW T L S I WW T L S I WW T L Y WW - T L S I WW T L Y WW - T L AMY AF I I AW AF I I AW AF I I AW AF I I AW MG I I AW W GF V I AT L GF V I AKM AF I P Y W W AF T AMW WY AF AT D MM WW AF AT D
 | 3050 GMLAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMAAWIYIIY GMACWIYIIY GMACWIYIIY GMACWIYIIY GMACWIYIIY GMICWFYIIY GMICWFYIIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GMACWIYIY GCLAWYMIY GCLAWYMIY GCLAWYMIY GGCWWAMING GCVGWAMIM GCYCWAMIM GCYCWAMIMI GCYCWAMIMI GCYCWAMIMI GCYCWAMINI
 | 1000 E I WAG E I F RG E V M GO E V M GO E V M GO E V M GO E V F RG V A T G I P KG V A T G T F WS V A T G T F WS V A T G T F WS L L T N V M G T M S M F C

 | 12
2 A K K S V S - G
D A K K A A S - D
D V K E A S A - S
S - I M D A A G
E A A I I N K N S
E A S I I N K N S
E A A I N K S A K A I S
E A A I N K S A K A I S
E A S O I N A G
E A S O I N A G
E A S O I N S C
S A K K I A V - R
Y D G K C V D N S K
- V T K A A E T
V V Y G E A S D N S
S A K K I A S T
V V Y G E A S D
S A K K I A S T
V V Y G E A S D
S A K K I A S T
V V Y G E A S D
S A K I A S T
S A K I | ASE GYQ FAFT
ASE GYQ FAFT
ASE GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GSD GYQ FAFT
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN EAG GFAFQ
GN AASQ QAFK
GN AAAQ QAFK
GN AAAQ GAFN
GN AAAQ GAFN
GN AAAQ GAFN
GN AAAQ GAFN
GN AAAQ SAFA
GN AAAQ SAFA
SEAAX SAFA
SEAAX SAFA
SAFA
SAFA SAFA
SAFA
SAFA SAFA
SAFA SAFA | 2020
AMALILITYGW
WMTYILIFGW
WMTYILIFGW
AMRAIVIFGW
AMRAIVIFGW
SLKWIYTGW
SLKWIYTGW
TIKIVITGW
SLKWIYTGW
TIKIVITGW
TIKUVITGW
AMMMILVITGW
AMMMILVITGW
AMMMILVITGW
AMMMILVITGW
AMMMILVIGW
AMMMILVIGW
AMMMILVIGW
AMMMILVIGW
AMMMILVIGW
AMRAFILUGW
ALRTFILUGW
ALRTFILUGW
ALRTFILUGW
ALRTFILUGW | 10
A 1 YP LG Y LG
A 1 YP LG | 310
QGD
 | 1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200
1200 | 300
DNL - N U YN
DNL - N U YN
DML - N U YN
DML - N U YN
DDM - N YN
E L - N YN
S L - N YN
S L - N YN
S S - N YN
S - N YN |




Supplementary Figure S19



		8	37.38 37.42	34.8 35.97	37.15	34.28	36.93	35.34	35.91	36.05	36.98	36.24	37.25	37.25	37.37	36.86	36.8	36.56	35.46	35.01	30.04																																		
5015	CL 1.0	Contig	ERR315 859-C350 ERR 599094-C1432	ERR 599094-C1899 ERR 599094-C1956	ERR599094-C1982 Made CMA.OCT7007-C68	MedDCM-OCT2007-C633	MedDCM-OCT2007-C188	MedDCM-OCT2007-C22970	MedD CM-OCT2007-C3050	MedDCM-OCT2007-C824	MedDCM-OCT2007-C352	MedD CM-OCT2007-C1482	MedDCM-OCT2007-C359	MedDCM-OCT2007-C359	MedDCM-OCT2007-C368	MedDCM-OCT2007-C360	MedDCM-OCT2007-C174	MedDCM-OCT2007-C2146 Med-lo7-77mDCM-C1113	Med-lo7-77mDCM-C1303	Med-lo7-77mDCM-C2049	MIEG-10/-//ШЛ/ГМІ-Г7773																																		
	11000	(bp)	16544 11461	11409	10464	21372	17947	16520	15490	14690	13848	13425	10823	10793	C7007																																								
		90	35.6 35.9	35.24 35.88	38.59	36.85	36.27	35.76	36.21	36.43	36.4	35.91	36.19	35.37	0.00																																								
	213	Contig	ERR598942-C1146 ERR598942-C2164	ERR598942-C2190 ERR598942-C2311	ERR598942-C2561	ERR 598983-C7 18	ERR 598983-C975	ERR598983-C1136 FBP508083-C1174	ERR598983-C1305	ERR598983-C1450	ERR598983-C1599	ERR598983-C1680	ERR598983-C2504	ERR598983-C2521	17077-FOE05FW17																																								
	1	(bp)	47235 40668	39153 31052	27517	24850	24120	20187	18345	18315	18089	17073	16955	16353	16025	15611	14538	14335	12756	12697	11717	10580	10193	10057	*******																														
		5	17.04	37	16.44	14.72	15.59	15.47	17.41	16.21	17.48	36.1	16.77	16.55	16.76	17.03	36.5	16.87	15.57	17.61	85.0	15.48	10.01	14.68	0/10/																														
501.4	*13	Contig	MedDCM-SEP2014-C87 3 Med DCM-SEP2014-C128 3	Med DCM-SEP2014-C148 Med DCM-SEP2014-C247	Med DCM-SEP2014-C304	Med DCM-SEP2014-C383	MedDCM-SEP2014-C411	MedDCM-SEP2014-C552	MedDCM-SEP2014-C667	MedDCM-SEP2014-C672	MedDCM-SEP2014-C690	Med DCM-SEP2014-C793	MedD CM-SEP2014-C803	MedDCM-SEP2014-C851	Med DCM-SEP2014-C888	MedDCM-SEP2014-C940	MedDCM-SEP 2014-C1047	MedDCM-SEP 2014-C1069	MedDCM-SEP2014-C1286	MedDCM-SEP 2014-C1299	MedDUM-SEP 2014-C1327	MedDCM-SEP2014-C1762	MedDCM-SEP2014-C1880	MedDCM-SEP2014-C1925																															
	And a local division of the local division o	(bp)	55782 43429	35446 33474	28224	27084	26753	25582	24908	24366	22564	19671	19191	18534	16478	15132	14753	14733	14168	13546	13491	12920	12827	11945	11189	10696	10233																												
		S	35.54 34.84	35.53 35.82	34.65	36.93	35.35	35.95	34,34	36.71	36.9	33.95	36.04	36.41	36.43	35.61	35.42	34.98	37.49	36.29	30.0L	38.68	35.55	35.3 35.47	38.12	35.03	36.25																												
010	2	Contig	ERR598942-C113 ERR598942-C202	ERR598942-C302 ERR598942-C339	ERR598942-C445	ERR598942-C481	ERR598942-C494	ERR598942-C530 FBB508042-C531	ERR598942-C563	ERR598942-C588	ERR598942-C669	ERR598942-C852	ERR 598942-C 899	ERR598942-C963	ERR598942-C1159	ERR598942-C1332	ERR598942-C1391	ERR598942-C1396 FRR598942-C1476	ERR598942-C1503	ERR598942-C1625	EKK598942-C1542	ERR598942-C1776	ERR598942-C1797	ERR598942-C2010 EBB508043-C2030	ERR598942-C2268	ERR598942-C2464	ERR 598942-C2658																												
-		th (bp)	1364 1365	1638 1616	030	053	185	543	749	605	51.75 50.35	875	686	645	116	1179	1037	838	277	843	2/02	279	707	1398	1393	196	030	th (bp)	058	1682	843	\$762	5223	045	1382	390	502.4	1057	954	704	1874	926	974	194	824	769	328		th (bp)	175	1440	5285	887	075 5740	953
		c leng	98 2. 94 2.	2 IS	14 2	1.2	26 20	8. 5	18 1	43 10	46	75	44 15	31 19	62	95 1-	1 01	E. 18	81	е: П	1 1	85	42 10	14 10	26 10	66	91 10	c leng	19 79	20 55	37 33	2	1. 5	56 20	80	2 22	52 10	57	9.84 1.11	1	9.9	87	28	85 11	38	47 1:	51 e 1 t		c leng	52 SS	88	5 5 98 98	4	- 5 - 5	
		ō	7 35. 8 34.	9 36. 9 36.	36.35	28 36	9 36.	96 ¥	。 36.	7 35,	2 30	6 35.	45 37,	28 36.	1 10	3 34.	0 37.	38 38	34.5	34	202 24	33.	37 36.	36.36. 35.35	35.	11 34.	34.	ō	36.	36	36.	36.	36	36.	36	37.	36.	¥. 5	9F 22	34.	36 1	35. 35.	35.	35. 35.	1 65 7 75	57 35,	26 36.		9	36.35	37.	36.33	36	8 8	
EDI 34	CFF 2A	Contig	MedDCM-JUL2012-C37 MedDCM-JUL2012-C36	MedDCM-JUL2012-C40 MedDCM-JUL2012-C51	MedDCM-JUL2012-C80	MedDCM-JUL2012-C12	MedDCM-JUL2012-C61	MedDCM-JUL2012-C64	MedDCM-JUL2012-C59	MedDCM-JUL2012-C74	MedDCM-JUL2012-C67	MedDCM-JUL2012-C69	MedDCM-JUL2012-C14-	MedD CM-J UL2012-C143	MedDCM-JUL2012-C/9	MedDCM-JUL2012-C83	MedDCM-JUL2012-C84	MedDCM-JUL2012-C85 MedDCM-IIII 2012-C86	MedDCM-JUL 2012-C89	MedD CM-J UL2012-C148	MedDUM-JULZUIZ-CI34	MedDCM-JUL2012-C148	MedD CM-J UL2012-C151	MedDCM-JUL2012-C129	MedDCM-JUL2012-C169	MedD CM-J UL2012-C174	MedDCM-JUL2012-C13	Contig	ERR598983-C49	ERR598983-C129	ERR 598983-C317	ERR 598983-C403	ERR 598983-C471 EDD 508083-C478	ERR 598983-C805	ERR 598983-C942	ERR598983-C1021 ERR598983-C1043	ERR598983-C1212	ER R59898 3-C 1565	ER R59898 3-C 1805 ER R59898 3-C 1824	ERR598983-C1894	MedDCM-JUL2012-C113	MedDCM-JUL2012-C105 MedDCM-JUL2012-C115	MedD CM-J UL2012-C110	MedD CM-J UL2012-C113 MedD CM-J UL2012-C1 0	MedDCM-JUL2012-C10	MedD CM-J UL2012-C10	MedD CM-J UL2012-C112 MedD CM-III 2012-C112	EPI-2C	Contig	ERR598983-C264 ERR598983-C264	ERR598983-C307	ERR 598983-C415 ERR 598983-C455	ERR 598983-C752	ERR598983-C870 FRR598983-C1176	ERR598983-C1232 ERR598983-C1232
		ngth (bp)	10318 10158	54729 46411	38768	32391	31966	29994	25250	24705	23521	20075	19611	12057	17255	16497	16346	13613	13048	12706	11142	11076	10870	10196	10145	10045	23347	17111	16664	16045	12697	10508	15307	11510	10667	120849	98497	88977	52149 44426	41300	38571	37963	37849	36482	33568	33097	32394	28213	27625	22218	16170	15511			
100		GC le	6.81 7.82	6.64 36.8	6.68	5.57	6.78	6.62	37.8	6.12	6.33	6.15	6.95	5.74	36	6.43	5.45	5.65	7.25	5.41	2017	6.42	7.21	5.57 5.06	117	6.96	5.91	1072	37.4	5.25	5.93	7.06	5.62	3.59	7.35	5.58	5.67	6.91	5.94	6.93	7.64	7.34	6.42	6.79	35.2	6.91	7.48	8.45	6.59	7.41	8.84	7.17			
olianhas			~ ~	m			m	m ··	.,,	m	m ··	' m	m		2	m	m		1 (7)	m			m	m m				n m				1 m		n m		13 2		62	14 E	148	566 26	3 2	198	00	1	967 3	114 22 2	12	530	50 90	818	576 3			
ו המכוו מו חוה ואומ-ווו		Contig	ERR 598986-C1752 ERR 598986-C1805	ERR598993-C136 ERR598993-C185	ERR598993-C260 EPP508003-C240	ERR598993-C343	ERR598993-C353	ERR598993-C391 FBP508003-C404	ERR598993-C549	ERR598993-C566	ERR598993-C612	ERR598993-C814	ERR598993-C863	ERR5 98993-C1010 EBBE 00003-C1033	ERR5 98993-C1094	ERR5 98993-C1183	ERR5 98993-C1197	ERR5 98993-C1499 FRP5 08003-C1654	ERR5 98993-C1784	ERR5 98993-C1876	289LJ-59892273555	ERR5 98993-C2381	ERR5 98993-C2473	ERR5 98993-C2653 ERP5 98993-C7553	ERR5 98993-C2785	ERR5 98993-C2847	ERR598996-C297	ERR598996-C539	ERR598996-C570	ERR598996-C611	ERR598996-C966	ERR5 98996-C1413	ERR5 99073-C63 EPP5 00073-C70	ERR599073-C169	ERR599073-C205	MedDCM-OCT2007-C	MedDCM-OCT2007-C	MedDCM-OCT2007-C	Med DCM- OCT 2007- C8 Med DCM- OCT 2007- C1	Med DCM- OCT 2007- C4	MedD CM-OCT2007-C6	MedD CM-OCT 2007-C3	MedDCM-OCT2007-C8	MedDCM-OCT2007-C1	MedDCM-OCT2007-C5	Med DCM-OCT2007-C9	MedDCM-OCT2007-C3	Med DCM- OCT 2007- C3	MedDCM-OCT2007-C1	MedDCM-OCT2007-C7 MedDCM-OCT2007-C3	MedDCM-OCT2007-C8	MedD CM-0 CT 2007-C6			
conugs (-	(bp)	37101 33760	24149 22103	21634	17143	15965	15847	15019	14354	13728	12240	12116	11778	10907	10770	10336	14137	11194	28852	16051	18426	15553	14481	13773	132.63	12441	12166	11346	24549	21149	18125	17084	17065	12960	12463	12288	12210	10626	10530	10510	10168	24737	23350	16494	14518	13981	12565	12163	12066	10564	10503			
	ľ	gc	37.16 35.99	36.13	36.61	37.01	36.67	35.19	37.48	36.35	36.5	36.45	36.28	37.13	35.58	36.31	37.1	36.01	37.13	36.69	10.15	36.93	37.34	35.29	34.51	35.26	35.29	37.53	38.2	36.29	35.63	35.61	36.28	37.48	38.58	37.32	35.26	37.01	36.74	37.65	36.22	36.04	37.44	35.74	35.9	35.12	34.91	37.1	37.34	35.29 37.48	36.33	37.79			
oupplementary table .	TLLT	Contig	ERR594294-C258 ERR594294-C291	ERR594294-C 495 ERR594294-C 597	ERR594294-C618	ERR594294-C928	ERR594294-C1053	ERR594294-C1074 EBB5-04704-C1114	ERR5 94294-C1164	ERR5 94294-C1249	ERR5 94294-C1343	ERR5 94294-C1639	ERR5 94294-C1668	ERR594294-C1732 EDDE04204-C1044	ERR594 294-C1971	ERR594294-C2015	ERR594294-C2167	ERR594 294-C2278 FRR504 297-C1107	ERR594 297-C1723	ERR594335-C415	0C010-555 #YCNN3 EDD C0A0A0-000	ERR594348-C443	ERR594348-C623	ERR594348-C739 FPP 504348-C758	ERR594348-C837	ERR594348-C893	ERR594348-C1015	ERR594 348-C1055	ERR594 348-C1205	ERR 598976-C3 79	ERR598976-C538	ERR 598976-C 738	ERR 598976-C803	ERR598976-C834	ERR598976-C1434	ERR598976-C1543 ERR598976-C1550	ERR598976-C1600	ERR598976-C1618	ERR598976-C1798 ERR598976-C2115	ERR598976-C2157	ERR598976-C2167	ERR598976-C2324	ERR 598986-C 308	ERR 598986-C357 FPP 508086-C457	ERR 598986-C7 03	ERR598986-C912	ERR 598986-C980 FRR598 986-C1168	ERR598986-C1198	ERR598986-C1274	ERR598986-C1293 ERR598986-C1344	ERR598986-C1671	ERR598986-C1685			

EPH6	EPI-5 EPI-6 length	EPI-5 EPI-5 EPI-6 EPI-6	4 EPI-5 EPI-5 EPI-6	EPI-4 EPI-5 EPI-6 EPI-6 EPI-6 EPI-6 EPI-6	EPI-4 EPI-6
GC (bp)	bp) Contig GC engu	c tengun Contig GC tengun (bp)	Contig GC ^{tengun} Contig GC ^{tengun} (bp)	(bp) Contig GC ^{tengtri} Contig GC ^{tengtri} (bp)	GC tengui Contig GC tengui Contig GC tengui (bp)
35.6 16544	7235 ERR598942-C1146 35.6 16544	24 47235 ERR598942-C1146 35.6 16544	WedDCM-SEP2014-C87 37.04 47235 ERR598942-C1146 35.6 16544	55782 MedDCM-SEP2014-C87 37.04 47235 ERR598942-C1146 35.6 16544	35.54 55782 MedDCM-SEP2014-C87 37.04 47235 ERR598942-C1146 35.6 16544
35.9 11461 E	3668 ERR598942-C2164 35.9 11461 E	26 40668 ERR598942-C2164 35.9 11461 E	wedDCM-SEP2014-C128 36.26 40668 ERR598942-C2164 35.9 11461 E	43429 MedDCM-SEP2014-C128 36.26 40668 ERR598942-C2164 35.9 11461 E	34.84 43429 MedDCM-SEP2014-C128 36.26 40668 ERR598942-C2164 35.9 11461 E
35.24 11409 E	9153 ERR598942-C2190 35.24 11409 E	7 39153 ERR598942-C2190 35.24 11409 E	vedDCM-SEP2014-C148 37 39153 ERR598942-C2190 35.24 11409 E	35446 MedDCM-SEP2014-C148 37 39153 ERR598942-C2190 35.24 11409 E	35.53 35446 MedDCM-SEP2014-C148 37 39153 ERR598942-C2190 35.24 11409 E
35.88 11050 E	1052 ERR598942-CZ311 35.88 11050 E	/1 31052 EKK598942-C2311 35.88 11050 E	VedDCM-SEP2014-C.24/ 36./1 31052 ERRS98942-C.2311 35.88 11050 E	334/4 MedUCM-5EPZ014-C24/ 36./1 31052 EKK598342-C2311 35.88 11050 E	35.82 334/4 MedDCM-5FP2014-C24/ 36//1 31052 EKK598942-C2311 35.88 11050 E
35.52 23368 Mad	1010 ERREGRANT 2010 25,52 23268 Med	13 77494 FRR598983-C544 35.52 23368 Med	WEUDUM-SEF2014-C304 30.44 2/31/ ENN333342-C201 30.33 20404 EN	20224 INEUDUNTSEFZUTATION 30A4 2/31/ ENN3034242-C201 36.33 20494 E1	34.03 26224 INRUDUNTSEFZUTATION 36.43 77494 EBP508083-CF04 35.57 23368 Mind
36.85 21372 Med	4850 ERR598983-C718 36.85 21372 Med	72 24850 ERR598983-C718 36.85 21372 Med	4edDCM-SEP2014-C383 34.72 24850 ERR598983-C718 36.85 21372 Med	27084 MedDCM-SEP2014-C383 34.72 24850 ERR598983-C718 36.85 21372 Med	36.93 27084 MedDCM-SEP2014-C383 34.72 24850 ERR598983-C718 36.85 21372 Med
36.27 17947 Med	(120 ERR598983-C975 36.27 17947 Med	39 24120 ERR598983-C975 36.27 17947 Med	0edDCM-SEP2014-C411 35.59 24120 ERR598983-C975 36.27 17947 Med	26753 MedDCM-5EP2014-C411 35.59 24120 ERR598983-C975 36.27 17947 Med	35.35 26753 MedDCM-5FP2014-C411 35.59 24120 FRR588983-C975 36.27 17947 Med
35.76 16520 MedDC	1187 ERR598983-C1136 35.76 16520 MedDC	47 20187 ERR598983-C1136 35.76 16520 MedDC	4ed DCM-SEP2014-C552 35.47 20187 ERR598983-C1136 35.76 16520 MedDC	25582 MedDCM-SEP2014-C552 35.47 20187 ERR598983-C1136 35.76 16520 MedDC	35.95 25582 MedDCM-SEP2014-C552 35.47 20187 ERR589893-C1136 35.76 16520 MedDC
35.36 16242 MedDi	1122 ERR598983-C1174 35.36 16242 MedD	18 20122 ERR598983-C1174 35.36 16242 MedD	/ied DCM-SEP2014-C554 37.18 20122 ERR598983-C1174 35.36 16242 MedDi	25568 MedDCM-SEP2014-C554 37.18 20122 ERR598983-C1174 35.36 16242 MedD	35.88 25568 MedDCM-SEP2014-C554 37.18 20122 ERR598983-C1174 35.36 16242 MedD
36.21 15490 MedDCI	1345 ERR598983-C1305 36.21 15490 MedDCI	11 18345 ERR598983-C1305 36.21 15490 MedDCI	Aed DCM-SEP2014-C667 37.41 18345 ERR598983-C1305 36.21 15490 MedDCI	24908 MedDCM-SEP2014-C667 37.41 18345 ERR598983-C1305 36.21 15490 MedDCI	34.34 24908 MedDCM-5EP2014-C667 37.41 18345 ERR598983-C1305 36.21 15490 MedDCI
36.43 14690 MedDCI	3315 ERR598983-C1450 36.43 14690 MedDCI	21 18315 ERR598983-C1450 36.43 14690 MedDCI	Aed DCM-SEP2014-C672 36.21 18315 ERR598983-C1450 36.43 14690 MedDCI	24366 MedDCM-5EP2014-C672 36.21 18315 ERR598983-C1450 36.43 14690 MedDCI	36.71 24366 MedDCM-SEP2014-C672 36.21 18315 ERR598983-C1450 36.43 14690 MedDCI
36.4 13848 MedDCI	3089 ERR598983-C1599 36.4 13848 MedDCI	48 18089 ERR598983-C1599 36.4 13848 MedDCI	Aed DCM-SEP2014-C690 37.48 18089 ERR598983-C1599 36.4 13848 MedDCI	22564 MedDCM-SEP2014-C690 37.48 18089 ERR598983-C1599 36.4 13848 MedDCI	36.9 22564 MedDCM-SEP2014-C690 37.48 18089 ERR598983-C1599 36.4 13848 MedDCI
37.02 13764 MedDCh	¹ 936 ERR598983-C1619 37.02 13764 MedDC ¹	58 17936 ERR598983-C1619 37.02 13764 MedDCP	Aed DCM-SEP2014-C703 37.68 17936 ERR598983-C1619 37.02 13764 Med DCP	19885 MedDCM-SEP2014-C703 37.68 17936 ERR598983-C1619 37.02 13764 MedDC	35.43 19885 MedDCM-5EP2014-C703 37.68 17936 ERR598983-C1619 37.02 13764 MedDC
35.91 13425 MedDCM	7073 ERR598983-C1680 35.91 13425 MedDCM	1 17073 ERR598983-C1680 35.91 13425 MedDCM	Aed DCM-SEP2014-C793 36.1 17073 ERR598983-C1680 35.91 13425 MedDCM	19671 MedDCM-SEP2014-C793 36.1 17073 ERR598983-C1680 35.91 13425 MedDCM	33.95 19671 MedDCM-SEP2014-C793 36.1 17073 ERR598983-C1680 35.91 13425 MedDCM
36.19 10823 MedDCM	5955 ERR598983-C2504 36.19 10823 MedDCN	77 16955 ERR598983-C2504 36.19 10823 MedDCM	AedDCM-SEP2014-C803 36.77 16955 ERR598983-C2504 36.19 10823 MedDCM	19191 MedDCM-SEP2014-C803 36.77 16955 ERR598983-C2504 36.19 10823 MedDCM	36.04 19191 MedDCM-SEP2014-C803 36.77 16955 ERR598983-C2504 36.19 10823 MedDCM
35.37 10793 MedDCM	3353 ERR598983-C2521 35.37 10793 MedDCM	55 16353 ERR598983-C2521 35.37 10793 MedDCM	Aed DCM-SEP2014-C851 36.55 16353 ERR598983-C2521 35.37 10793 MedDCM	18534 MedDCM-SEP2014-C851 36.55 16353 ERR598983-C2521 35.37 10793 MedDCM	36.41 18534 MedDCM-SEP2014-C851 36.55 16353 ERR598983-C2521 35.37 10793 MedDCM
35.3 10525 MedDCM-4	5283 ERR598983-C2624 35.3 10525 MedDCM-4	32 16283 ERR598983-C2624 35.3 10525 MedDCM-4	AedDCM-SEP2014-C857 34.82 16283 ERR598983-C2624 35.3 10525 MedDCM-4	17164 MedDCM-SEP2014-C857 34.82 16283 ERR598983-C2624 35.3 10525 MedDCM-4	36.96 17164 MedDCM-SEP2014-C857 34.82 16283 ERR598983-C2624 35.3 10525 MedDCM-
MedDCM	5025 MedDCM	76 16025 MedDCM	Aed DCM-SEP2014-C888 36.76 16025 Med DCM	16478 MedDCM-SEP2014-C888 36.76 16025 MedDCM	36.43 16478 MedDCM-SEP2014-C888 36.76 16025 MedDCM
MedDCM-	5611 MedDCM-	33 15611 MedDCM	Aed DCM-SEP2014-C940 37.03 15611 Med DCM-	15132 MedDCM-SEP2014-C940 37.03 15611 MedDCM-	35.61 15132 MedDCM-SEP2014-C940 37.03 15611 MedDCM-
MedDCM-6	1538 MedDCM-4	5 14538 MedDCM-4	tedDCM-SEP 2014-C1047 36.5 14538 MedDCM-4	14753 MedDCM-SEP 2014-C1047 36.5 14538 MedDCM-4	35.42 14753 MedDCM·SEP2014-C1047 36.5 14538 MedDCM·4
MedDCM-6	1335 MedDCM-C	37 14335 MedDCM-0	ledDCM-SEP 2014-C1069 36.87 14335 MedDCM-C	14733 MedDCM-SEP2014-C1069 36.87 14335 MedDCM-C	34.98 14733 MedDCM-5EP 2014-C1069 36.87 14335 MedDCM-C
Med-lo7-7	3922 Med-lo7-7	95 13922 Med-lo7-7	tedDCM-SEP2014-C1118 35.95 13922 Med-lo7-7	14312 MedDCM-SEP2014-C1118 35.95 13922 Med-lo7-7	36.49 14312 MedDCM-SEP2014-C1118 35.95 13922 MedDCM-SEP2014-C1118 35.95
Med-lo7-7	1756 Med-lo7-7	57 12756 Med-lo7-7	TedDCM-SEP2014-C1286 35.57 12756 Med-Io7-7	14168 MedDCM:SEP2014-C1286 35.57 12756 Med-k07-7	37.49 14168 MedDCM-SEP2014-C1286 35.57 12756 Med-lo7-7
Med-lo7-7	0.697 Med-lo 7-77	51 12697 Med-lo7-77	1edDCM-SEP 2014-C1299 37.61 12697 Med-b7-77	13546 MedDCM-SEP 2014-C1299 37.61 12697 Med-b7-77	36.29 13546 MedDCM-SEP 2014-C1299 37.61 12697 Med-b7-77
Med-Io7-77	3543 Med-lo7-77	98 12543 Med-lo7-77	ledDCM-SEP 2014-C1327 35.98 12543 Med-Io7-77	13491 MedDCM-SEP2014-C1327 35.98 12543 Med-lo7-77	36.61 13491 MedDCM-SEP2014-C1327 35.98 12543 Med-Io7-77
	1717	54 11717	ledDCM-SEP2014-C1485 37.54 11717	13056 MedDCM-SEP2014-C1485 37.54 11717	37.19 13056 MedDCM-SEP2014-C1485 37.54 11717
	1580	48 10580	ledDCM-SEP2014-C1762 35.48 10580	12920 MedDCM-SEP2014-C1762 35.48 10580	38.68 12920 MedDCM-SEP2014-C1762 35.48 10580
	1193	01 10193	ledDCM-SEP2014-C1880 36.01 10193	12827 MedDCM-SEP2014-C1880 36.01 10193	35.55 12827 MedDCM-SEP2014-C1880 36.01 10193
	0057	68 10057	redDCM-SEP2014-C1925 34.68 10057	11945 MedDCM-SEP2014-C1925 34.68 10057	35.3 11945 MedDCM-SEP2014-C1925 34.68 10057
	1544	78 34344	educm-Uci 2007-c1035 35.78 34544	11916 MedDCM-UCI200/-C1035 36./8 34544	25.4/ 1191b MedUXM-UCI200/-C1035 35./8 34344
				11189	38.12 11189
				10696	35.03 10696
				10/233	30,25 10,233

Supplement	tary Table 2. Li	st of meta	genomes used fo	r recruitments in Figure 4a and Supplementary Figure 2B, s	sorted by te	mpera	ture an	d dept	h.															
Metagenome	Size fraction	Depth	Temperature (°C)	Location	Work	Epi-1	Epi-3	Epi-4	Epi-5 E	Epi-6 Es	oi-2A	pi-2B E	pi-2C	Bathy-1	Bathy-2	Cavman9	Cavman9	Guavmas32	Guavmas31	SCGCAAA-228E19	Thalassoarchaea	MG2-GG3	A. boneii	N. maritimus
ERR599119	0.22 μm 0.22 μm	Sm Sm	26.8	South Pacific Subtropical Gyre Province, North and South	1	3.02	0.67	1.24	0.68	0.75	0.14	0.16	0.09	0.08	0.06	0.05	0.01	0.11	0.04	0.12	0.08	0.12	0.01	0.01
ERR599030	0.22 µm	Sm	26.6	North Pacific Equatorial Countercurrent Province	1	1.69	0.23	0.80	0.27	0.28	0.05	0.07	0.06	0.05	0.04	0.04	0.00	0.07	0.04	0.10	0.10	0.12	0.01	0.01
ERR594307	0.22 μm 0.22 μm	Sm Sm	26.5	South Pacific Subtropical Gyre Province, North and South South Pacific Subtropical Gyre Province, North and South	1	16.71	1.28	5.30	1.49	1.72	0.31	0.31	0.06	0.16	0.04	0.10	0.00	0.08	0.03	0.31	0.26	0.41	0.03	0.01
ERR599069	0.22 µm	Sm	26.5	South Pacific Subtropical Gyre Province, North and South	1	4.31	0.36	1.28	0.43	0.48	0.09	0.09	0.07	0.06	0.05	0.05	0.00	0.08	0.05	0.16	0.17	0.17	0.02	0.02
ERR598989 ERR599038	0.22 μm 0.22 μm	Sm Sm	26.1	Pacific Equatorial Divergence Province	1	3.19	0.12	1.18	0.15	0.46	0.02	0.03	0.03	0.08	0.02	0.02	0.00	0.03	0.03	0.12	0.21	0.32	0.02	0.01
ERR599142	0.22 µm	Sm	25.2	North Pacific Subtropical and Polar Front Provinces	1	11.42	0.53	2.61	0.72	0.87	0.12	0.13	0.09	0.07	0.05	0.05	0.01	0.08	0.04	0.10	0.10	0.10	0.01	0.00
ERR598984	0.22 μm 0.22 μm	Sm Sm	25.1	South Pacific Subtropical Gyre Province, North and South South Atlantic Gyral Province	1	4.59	0.22	1.10	0.27	0.44	0.06	0.07	0.05	0.03	0.02	0.02	0.00	0.05	0.02	0.06	0.07	0.05	0.01	0.01
ERR599136	0.22 µm	Sm	25	Caribbean Province	1	2.05	0.62	0.87	0.72	0.65	0.31	0.39	0.16	0.22	0.15	0.15	0.02	0.30	0.14	0.46	0.41	0.21	0.03	0.06
ERR598954 ERR594288	0.22 µm 0.22 µm	Sm Sm	24.2	South Pacific Subtropical Gyre Province, North and South Mediterranean Sea, Black Sea Province	1	2.15	0.32	0.48	0.16	0.24	0.15	0.03	0.02	0.09	0.07	0.08	0.01	0.04	0.03	0.02	0.05	0.08	0.01	0.00
ERR599024	0.22 µm	Sm	23.8	South Pacific Subtropical Gyre Province, North and South	1	2.08	0.20	0.75	0.22	0.32	0.05	0.06	0.03	0.03	0.02	0.02	0.00	0.05	0.03	0.06	0.06	0.07	0.01	0.01
ERR594286 ERR599077	0.22 μm 0.22 μm	Sm Sm	23.3 22.8	South Atlantic Gyral Province South Pacific Subtropical Gyre Province, North and South	1	4.74	0.74	3.32	0.34	1.17	0.07	0.09	0.04	0.04	0.03	0.03	0.00	0.06	0.03	0.16	0.08	0.08	0.01	0.01
ERR598970	0.22 µm	Sm	22.2	Eastern Africa Coastal Province	1	5.77	0.42	1.35	0.50	0.57	0.10	0.12	0.06	0.06	0.05	0.04	0.01	0.08	0.05	0.15	0.60	0.14	0.01	0.03
ERR598979 ERR598993	0.22 µm	Sm Sm	21.8	Eastern Africa Coastal Province Mediterranean Sea, Black Sea Province	1	3.83	0.35	1.08	0.50	6.87	0.09	0.10	0.07	0.07	0.06	0.06	0.01	0.11	0.06	0.23	0.97	0.15	0.02	0.07
ERR598955	0.22 µm	Sm	20.5	North Atlantic Subtropical Gyral Province	1	1.42	0.11	0.49	0.18	0.31	0.01	0.02	0.03	0.02	0.01	0.01	0.00	0.04	0.01	0.05	0.49	0.13	0.01	0.01
ERR599123 ERR594332	0.22 µm	Sm Sm	20.4	North Atlantic Subtropical Gyral Province South Atlantic Gyral Province	1	6.62 5.08	0.51	1.92	0.69	0.72	0.25	0.27	0.12	0.13	0.09	0.07	0.01	0.20	0.08	0.26	0.41	0.23	0.03	0.34
ERR594335	0.22 µm	Sm	19.8	South Atlantic Gyral Province	1	9.71	0.51	2.70	0.76	0.86	0.14	0.17	0.08	0.11	0.07	0.06	0.01	0.13	0.07	0.35	1.61	0.63	0.04	0.02
ERR598968	0.22 µm	Sm	19.1	North Atlantic Subtropical Gyral Province	1	2.44	0.34	0.81	0.44	0.44	0.32	0.33	0.14	0.16	0.08	0.07	0.01	0.23	0.07	0.24	0.37	0.17	0.03	0.79
ERR315858	0.22 µm	Sm	17.6	Mediterranean Sea, Black Sea Province	1	1.21	1.06	2.40	4.25	7.94	0.43	0.36	0.25	0.05	0.04	0.02	0.00	0.13	0.04	0.05	0.46	0.05	0.01	0.06
ERR599170	0.22 µm	Sm	17.6	North Atlantic Subtropical Gyral Province	1	5.25	0.37	1.40	0.86	1.31	0.20	0.18	0.11	0.05	0.03	0.03	0.00	0.09	0.03	0.16	2.79	0.44	0.02	0.10
ERR594297	0.22 µm	Sm	16.8	South Atlantic Gyral Province	1	7.87	0.76	2.37	1.76	2.54	1.30	1.03	0.50	0.21	0.11	0.10	0.02	0.31	0.10	0.30	3.73	0.20	0.02	0.19
ERR598973	0.22 µm	Sm	15	Benguela Current Coastal Province	1	3.05	1.01	1.21	3.41	3.91	1.01	0.81	0.67	0.07	0.03	0.03	0.00	0.11	0.03	0.15	15.02	0.35	0.01	0.51
ERR598983	0.22 μm 0.22 μm	Sm Sm	14.3	Gulf Stream Province	1	2.72	3.32	2.18	32.10	8.74	5.25	9.11	9.81	0.04	0.03	0.03	0.01	0.09	0.03	0.18	43.40	0.17	0.01	2.39
ERR599032	0.22 µm	40m	26.1	Pacific Equatorial Divergence Province	1	3.18	0.42	1.15	0.57	0.50	0.13	0.14	80.0	0.08	0.07	0.05	0.01	0.12	0.07	0.25	0.23	0.26	0.03	0.06
HOTS BATS	0.22 µm 0.22 µm	25m 50m	25	North Pacific Subtropical Gyre North Atlantic	5	8.08	0.17	1.70	0.29	0.45	0.09	0.17	0.12	0.04	0.00	0.00	0.00	0.09	0.03	0.03	0.06	0.12	0.01	0.00
ERR598990	0.22 µm	30m	21.8	Eastern Africa Coastal Province	1	3.88	0.38	1.03	0.52	0.45	0.11	0.12	0.08	0.11	0.08	0.10	0.04	0.20	0.10	0.36	0.91	0.20	0.02	0.13
ERR599014 ERR599081	0.22 µm	50m 50m	21.8	South Pacific Subtropical Gyre Province, North and South South Pacific Subtropical Gyre Province, North and South	1	2.98	0.76	2.00	0.72	0.64	0.12	0.16	0.10	0.08	0.07	0.07	0.01	0.11	0.07	0.29	1.44	0.46	0.03	0.02
ERR599007	0.22 µm	40m	19.6	Pacific Equatorial Divergence Province	1	1.22	1.71	3.49	0.52	0.73	0.08	0.11	0.08	0.04	0.03	0.02	0.01	0.07	0.03	0.11	1.34	0.14	0.01	0.01
ERR598987	0.22 µm	40m	18.6	North Atlantic Subtropical Gural Province	1	0.11	0.13	0.07	0.13	0.07	0.15	0.15	0.09	0.67	1.36	3.09	2.26	0.84	3.47	5.05	0.07	0.04	0.01	0.20
ERR594294	0.22 µm	50m	16.8	South Atlantic Gyral Province	1	14.74	1.31	4.54	2.64	3.84	1.39	1.11	0.59	0.26	0.14	0.13	0.02	0.45	0.12	0.45	2.91	0.17	0.03	0.93
ERR599094	0.22 µm	50m	15.2	Mediterranean Sea, Black Sea Province	1	2.19	0.78	2.66	2.86	4.85	0.96	0.67	0.39	0.04	0.03	0.02	0.00	0.11	0.03	0.06	0.43	0.06	0.01	0.82
ERR599001	0.22 µm	25m	13	North Atlantic Subtropical Gyral Province	1	1.04	0.40	0.60	1.18	1.79	1.37	0.88	0.59	0.06	0.04	0.04	0.00	0.11	0.04	0.20	22.98	0.19	0.02	0.48
ERR598942	0.22 µm	45m	13.2	North Pacific Subtropical and Polar Front Provinces	1	1.21	26.25	1.42	10.12	4.73	0.43	0.49	0.35	0.10	0.09	0.07	0.00	0.18	0.07	0.21	4.25	0.28	0.01	0.01
ERR599100	0.22 µm	120m 125m	23.2	Caribbean Province	1	2.00	0.70	1.03	0.86	0.81	0.32	0.45	0.15	0.25	0.19	0.18	0.01	0.34	0.16	0.50	0.44	0.26	0.02	0.09
ERR594342	0.22 µm	140m	23.7	South Pacific Subtropical Gyre Province, North and South	1	3.64	0.84	1.51	0.96	0.87	0.20	0.21	0.12	0.12	0.11	0.09	0.02	0.16	0.09	0.34	0.29	0.28	0.03	0.01
ERR598957 ERR598972	0.22 μm 0.22 μm	155m 65m	22.3	Eastern Africa Coastal Province, North and South	1	1.24	0.76	1.21	0.80	0.59	0.30	0.40	0.19	0.14	0.10	0.08	0.01	0.20	0.07	0.18	0.11	0.08	0.01	0.00
HOTS	0.22 µm	75m	22	North Pacific Subtropical Gyre	5	1.77	0.12	0.46	0.21	0.24	0.01	0.04	0.07	0.02	0.00	0.00	0.00	0.09	0.01	0.13	0.07	0.05	0.01	0.01
ERR594298 ERR598961	0.22 µm 0.22 µm	150m 90m	21.6	South Atlantic Gyral Province South Pacific Subtropical Gyre Province. North and South	1	8.28	1.37	2.12	1.40	1.19	0.46	0.91	0.34	0.39	0.17	0.14	0.03	0.60	0.13	0.37	0.23	0.11	0.03	0.15
ERR594336	0.22 µm	120m	19.3	South Atlantic Gyral Province	1	10.36	0.74	2.51	0.99	1.24	0.64	0.59	0.28	0.19	0.10	0.09	0.01	0.31	0.08	0.29	0.48	0.12	0.02	0.84
ERR599087 ERR318618	0.22 µm 0.22 µm	60m 70m	19	North Pacific Equatorial Countercurrent Province Mediterranean Sea, Black Sea Province	1	1.03	0.12	1.27	0.17	0.95	0.03	0.31	0.24	0.12	0.08	0.06	0.02	0.21	0.06	0.27	0.90	0.27	0.02	0.13
ERR599073	0.22 µm	60m	18.4	Mediterranean Sea, Black Sea Province	1	17.10	1.85	8.27	5.17	9.29	0.94	0.78	0.40	0.12	0.08	0.06	0.01	0.20	0.07	0.21	0.62	0.15	0.01	0.02
ERR598986 ERR594315	0.22 µm	80m 55m	16.8	North Atlantic Subtropical Gyral Province Mediterranean Sea, Black Sea Province	1	8.30 1.73	0.78	2.26	1.90	2.82	1.75	1.24	0.69	0.57	0.09	0.07	0.01	0.37	0.07	0.29	7.31	0.51	0.03	0.87
ERR315859	0.22 µm	55m	15.7	Mediterranean Sea, Black Sea Province	1	1.50	1.37	4.28	5.20	9.33	2.52	1.62	0.95	0.08	0.05	0.05	0.01	0.21	0.06	0.13	1.23	0.22	0.02	0.23
MedDCM-SEP2013 Med-Io16-70mDCM	0.22 µm	55m 70m	15.5	Mediterranean Sea	2	1.02	1.31	1.93	4.69	8.73	0.52	0.42	0.29	0.06	0.04	0.04	0.00	0.11	0.04	0.13	7.05	0.63	0.02	0.09
ERR598995	0.22 µm	115m	15.3	North Pacific Subtropical and Polar Front Provinces	1	1.39	2.15	1.35	2.49	3.07	4.31	2.91	1.72	0.32	0.15	0.13	0.02	0.57	0.12	0.36	1.78	0.21	0.02	0.11
MedDCM-SEP2014 Med-Ae1-75mDCM	0.22 µm	60m	15	Mediterranean Sea	3	3.32	1.60	16.47	5.31	9.61	0.70	0.62	0.36	0.09	0.07	0.05	0.00	0.17	0.06	0.11	0.86	0.06	0.01	0.03
Med-Io7-77mDCM	0.22 µm	77m	14.3	Mediterranean Sea	4	1.03	1.02	2.96	3.18	5.51	5.20	3.42	1.96	3.93	0.06	0.05	0.01	0.28	0.05	0.12	0.96	0.11	0.02	0.11
MedDCM-JUL2012	0.22 µm	75m	13.6	Mediterranean Sea North Atlantic Subtropical Coral Province	2	0.71	0.82	0.88	2.85	3.95	21.30	13.27	7.87	0.88	0.08	0.06	0.01	0.64	0.06	0.22	20.14	0.18	0.02	4.33
ERR599172	0.22 µm	270m	15.6	Indian Monsoon Gyres Province	1	0.04	0.05	0.03	0.07	0.02	0.05	0.05	0.02	0.45	1.83	3.18	2.32	0.22	2.90	2.31	0.01	0.01	0.02	0.05
ERR599109	0.22 µm	340m	15	Indian Monsoon Gyres Province	1	0.04	0.04	0.04	0.06	0.03	0.06	0.06	0.02	0.48	1.05	1.92	1.33	0.34	1.72	1.94	0.02	0.01	0.02	0.14
ERR598953	0.22 µm	177m	14	South Pacific Subtropical Gyre Province, North and South	1	0.24	0.33	0.21	0.39	0.24	0.49	0.55	0.29	1.55	0.52	1.10	0.76	5.07	1.27	2.08	0.14	0.05	0.02	1.23
ERR594290	0.22 µm	600m	12.1	Northwest Arabian Sea Upwelling Province	1	0.02	0.02	0.02	0.06	0.01	0.03	0.06	0.01	0.24	3.19	5.99	4.19	0.12	5.22	1.16	0.01	0.01	0.01	0.82
ERR599067 ERR599047	0.22 μm 0.22 μm	580m 640m	11.3	North Atlantic Subtropical Gyral Province	1	0.13	0.12	0.07	0.15	0.08	0.23	0.31	0.16	1.79	0.36	0.50	0.37	0.70	0.14	1.88	0.19	0.06	0.02	1.44
ERR599086	0.22 µm	350m	10.9	South Pacific Subtropical Gyre Province, North and South	1	0.21	0.16	0.12	0.20	0.11	0.23	0.28	0.13	1.36	0.39	0.30	0.28	2.65	0.43	4.51	0.25	0.08	0.03	1.12
ERR598985	0.22 µm	640m	9.8	Caribbean Province	1	0.40	0.08	0.06	0.13	0.07	0.13	0.23	0.07	1.10	0.25	0.51	0.50	0.57	0.35	1.84	0.19	0.06	0.03	1.55
ERR599055	0.22 µm	480m	9.2	Pacific Equatorial Divergence Province	1	0.12	0.14	0.07	0.15	80.0	0.13	0.27	0.08	0.72	0.11	0.09	0.09	2.90	0.14	1.16	0.16	0.05	0.02	1.98
ERR599152 ERR599004	0.22 μm 0.22 μm	375m 450m	8.9 8.2	North Pacific Equatorial Countercurrent Province North Pacific Equatorial Countercurrent Province	1	0.08	0.08	0.05	0.10	0.05	0.09	0.12	0.04	0.71	2.26	5.31 4.74	3.95	0.34	5.36	6.24	0.04	0.01	0.02	0.53
ERR594305	0.22 µm	600m	7.2	South Pacific Subtropical Gyre Province, North and South	1	0.33	0.07	0.10	0.10	0.07	80.0	0.25	0.04	0.57	0.33	0.34	0.29	0.29	0.44	4.12	0.21	0.07	0.02	3.45
ERR599166 ERR599115	0.22 µm 0.22 µm	590m 650m	5.1	Gult Stream Province North Pacific Subtropical and Polar Front Provinces	1	0.18	0.18	0.12	0.18	0.12	0.42	0.61	0.32	1.76	0.72	3.03	1.00	4.85	1.53	4.12	0.20	0.08	0.03	2.36
ERR598964	0.22 µm	740m	10.6	North Atlantic Subtropical Gyral Province	1	0.14	0.11	0.07	0.13	0.07	0.17	0.27	0.11	1.61	0.33	0.46	0.34	0.74	0.53	1.43	0.17	0.06	0.02	1.59
ERR598944 ERR598960	0.22 µm 0.22 µm	800m 850m	10.2 8.4	North Atlantic Subtropical Gyral Province Eastern Africa Coardal Province	1	0.12	0.11	0.06	0.14	0.08	0.15	0.25	0.07	1.52	0.54	1.01	0.71	0.60	0.74	1.51	0.19	0.07	0.03	1.31
ERR599021	0.22 µm	1000m	7.7	Eastern Africa Coastal Province	1	0.14	0.09	0.06	0.13	0.06	0.14	0.25	0.06	1.29	0.77	1.48	1.04	0.63	1.63	2.41	0.14	0.05	0.03	1.41
ERR598947 ERR599134	0.22 µm	700m	7	South Atlantic Gyral Province	1	0.18	0.16	0.10	0.20	0.10	0.32	0.50	0.23	1.95	0.46	0.88	0.61	3.75	1.00	1.06	0.20	0.06	0.03	2.13
ERR599000	0.22 µm	800m	4.7	South Atlantic Gyral Province	1	0.07	0.06	0.03	0.09	0.04	0.06	0.17	0.03	0.67	1.32	2.90	2.07	0.52	3.22	2.23	0.09	0.03	0.03	1.30
ERR599048	0.22 µm	800m	4.7	South Atlantic Gyral Province	1	2.29	0.19	0.54	0.25	0.24	0.07	0.10	0.04	0.32	0.30	0.52	0.47	0.20	0.61	1.40	0.11	0.05	0.02	0.50
ERR599125	0.22 μm 0.22 μm	790m	*.2 0.5	Antarctic Province	1	0.14	0.14	0.10	0.19	0.08	0.31	0.45	0.18	1.16	2.98	6.96	5.14	3.50	5.76	1.09	0.15	0.06	0.02	1.18
KM3 Medulo17, Monumer	0.22 µm	3000m	14	Mediterranean Sea	6	0.00	0.00	0.00	0.00	0.28	0.00	0.00	0.00	0.07	6.51	1.37	1.63	0.13	1.23	0.29	0.28	0.00	0.00	0.19
SRR488330	0.22 μm 0.22 μm	2000m	14 3	Guaymas Basin	4	0.03	0.04	0.03	0.06	0.02	0.11	0.15	0.00	0.14	0.14	29.25 0.55	0.38	2.72	25.92	5.35 0.17	0.10	0.03	0.03	2.24
SRR488331	0.22 µm	2000m	3	Guaymas Basin	7	0.05	0.07	0.07	0.21	0.01	0.19	0.16	0.17	0.17	0.17	0.72	0.27	3.06	1.18	0.28	0.15	0.05	0.04	2.44
SRR2046235 SRR2046236	0.22 µm 0.22 µm	2040m 2238m	3	Cayman Cayman	7	0.06	0.07	0.04	0.09 0.06	0.04 0.03	0.09 0.06	0.08	0.02 0.01	3.89 5.18	2.44 3.04	6.38 8.04	5.01 6.66	0.20	4.82 5.95	1.40	0.03	0.03	0.02	0.22
SRR2046222	0.22 µm	2327m	3	Cayman	7	0.07	0.07	0.04	0.09	0.04	0.10	0.11	0.02	0.74	4.41	11.36	9.04	0.24	8.53	1.69	0.05	0.03	0.02	0.23
SRR2046238 SRR2046237	0.22 μm 0.22 μm	4869m 4946m	1	Cayman Cayman	7	0.05	0.04	0.03	0.07 0.08	0.02	0.06 0.08	0.09 0.07	0.02 0.03	0.41	8.00 8.62	22.60 24.90	19.88 21.64	0.18	14.85 16.47	1.85	0.05	0.04	0.03	0.74
SRR2046221	0.22 µm	4950m	1	Cayman	7	0.10	0.07	0.05	0.11	0.04	80.0	0.13	0.04	4,77	12.39	36.11	31.20	0.27	23.89	2.91	0.08	0.05	0.03	0.23
(+, sunagawa et al. (2015); Highlighted in grey, list of t	(+) marun-Luadrac he 33 metagenomi	et al. (201 is that were	assembled and used f	or the differential coverage in the binning procedure.	o et al. (2007);	(7) Li et :	n. (2015)																	

G-FDI-1			ED1-2	P		CG-EDL-2			G-FDLA			2-ED1-5		-90	EDL6		5	BATHV-1			G_RATHV_2		
1-1-1-0	2	4	CL 1.5			0-EL-1-0				-	4		2	4	2	ē	4	T- 1111 VO		, 4 ^{to} uo	7-111100-00		d+not
Contig	ec le	bp)	Contig	g	length (bp)	Contig	gc	ength (bp)	Contig	gc	(bp)	Contig	GC let	igun pp)	Contig	ر ا ر 60	bp)	Contig	GC	engun (bp)	Contig	gc	(bp)
CG-Epi1-C1	34.8 13.	5963 C	CG-Epi2-C1	35.8	101994	CG-Epi3-C1	34.8	71763	CG-Epi4-C1	36.7 8	80064	CG-Epi5-C1	35.8 47	627 C	G-Epi6-C1	36.8 50	0314 0	CG-Bathy1-C1	36.6 2	11731	CG-Bathy2-C1	63.5	130682
CG-Epi1-C2	36.8 95	3555 C	3G-Epi2-C2	35.2	92741	CG-Epi3-C2	35.6	63159	CG-Epi4-C2	36.1 5	59597	CG-Epi5-C2	37.1 26	037 C	G-Epi6-C2	35.8 41	1924 0	CG-Bathy1-C2	36.7 1	.33274	CG-Bathy2-C2	63.5	85590
CG-Epi1-C3	36.6 91	1163 C	CG-Epi2-C3	36.1	78522	CG-Epi3-C3	35.8	49058	CG-Epi4-C3	36.8 5	59301	CG-Epi5-C3	36.7 25	001 C	G-Epi6-C3	37.0 41	1453 0	CG-Bathy1-C3	36.5	91842	CG-Bathy2-C3	67.2	63709
CG-Epi1-C4	36.4 75	3844 C	CG-Epi2-C4	37.4	65383	CG-Epi3-C4	35.5	47213	CG-Epi4-C4	37.0 4	47475	CG-Epi5-C4	36.5 22	352 C	G-Epi6-C4	36.2 35	0 6966	CG-Bathy1-C4	37.0	77586	CG-Bathy2-C4	64.2	63432
CG-Epi1-C5	36.8 73	3433 C	3G-Epi2-C5	35.5	63969	CG-Epi3-C5	35.8	41384	CG-Epi4-C5	36.3 4	40977	CG-Epi5-C5	36.8 21	574 C	G-Epi6-C5	36.0 37	7884 0	CG-Bathy1-C5	38.7	68229	CG-Bathy2-C5	62.6	62484
CG-Epi1-C6	37.9 65	3690 C	3G-Epi2-C6	36.1	51772	CG-Epi3-C6	35.6	34927	CG-Epi4-C6	34.8 3	37978	CG-Epi5-C6	36.3 18	:166 C	G-Epi6-C6	33.7 25	5846 0	CG-Bathy1-C6	38.1	65587	CG-Bathy2-C6	62.3	62026
CG-Epi1-C7	37.1 58	3515 C	CG-Epi2-C7	36.3	48636	CG-Epi3-C7	35.9	33678	CG-Epi4-C7	36.0 3	37801	CG-Epi5-C7	36.4 18	034 C	G-Epi6-C7	36.8 25	5634 0	CG-Bathy1-C7	43.6	43732	CG-Bathy2-C7	66.1	47436
CG-Epi1-C8	36.8 57	7181 C	CG-Epi2-C8	36.3	41423	CG-Epi3-C8	36.2	32361	CG-Epi4-C8	36.6 3	34409	CG-Epi5-C8	35.7 16	719 C	G-Epi6-C8	36.2 22	2443 0	CG-Bathy1-C8	38.3	41508	CG-Bathy2-C8	64.7	37420
CG-Epi1-C9	36.7 53	3175 C	CG-Epi2-C9	36.6	39937	CG-Epi3-C9	35.2	28243	CG-Epi4-C9	36.7 3	31390	CG-Epi5-C9	35.9 16	:426 C	G-Epi6-C9	36.6 21	1880	CG-Bathy1-C9	37.9	41315	CG-Bathy2-C9	65.0	33792
CG-Epi1-C10	37.2 50	7091 Ct	G-Epi2-C10	35.7	38179	CG-Epi3-C10	36.6	28004	CG-Epi4-C10	36.7 3	31342	CG-Epi5-C10	35.4 16	368 CC	3-Epi6-C10	36.2 21	1094 C	G-Bathy1-C10	36.0	37030	CG-Bathy2-C10	62.7	27467
CG-Epi1-C11	36.3 40	7290 CI	G-Epi2-C11	36.5	37752	CG-Epi3-C11	36.9	27182	CG-Epi4-C11	36.8 2	27609	CG-Epi5-C11	36.2 15	933 CC	3-Epi6-C11	35.1 20	3996 C	G-Bathy1-C11	43.0	35415	CG-Bathy2-C11	64.4	26425
CG-Epi1-C12	36.8 37	7293 CI	G-Epi2-C12	37.1	31968	CG-Epi3-C12	36.5	26391	CG-Epi4-C12	35.4 2	20476	CG-Epi5-C12	34.9 15	869 CC	5-Epi6-C12	36.9 20	0201 C	G-Bathy1-C12	37.4	27476	CG-Bathy2-C12	64.8	25944
CG-Epi1-C13	37.2 36	5578 CI	G-Epi2-C13	36.2	31324	CG-Epi3-C13	35.9	25790	CG-Epi4-C13	37.2 2	20393	CG-Epi5-C13	36.2 15	690 CC	5-Epi6-C13	36.4 19	9776 C	G-Bathy1-C13	36.5	23386	CG-Bathy2-C13	65.8	25194
CG-Epi1-C14	37.2 34	1349 C(G-Epi2-C14	36.2	30879	CG-Epi3-C14	34.3	25074	CG-Epi4-C14	36.6 1	19527	CG-Epi5-C14	36.6 14	852 CC	5-Epi6-C14	36.1 18	3182 C	G-Bathy1-C14	36.2	22942	CG-Bathy2-C14	66.2	24873
CG-Epi1-C15	36.4 33	3957 CI	G-Epi2-C15	37.4	27764	CG-Epi3-C15	36.7	24609	CG-Epi4-C15	36.1 1	18526	CG-Epi5-C15	38.1 13	655 CC	5-Epi6-C15	36.4 14	1526 C	G-Bathy1-C15	38.0	22193	CG-Bathy2-C15	66.1	21939
CG-Epi1-C16	36.1 25	3648 Ct	G-Epi2-C16	33.8	26756	CG-Epi3-C16	35.4	20644	CG-Epi4-C16	36.6 1	17553	CG-Epi5-C16	35.8 13	629 CC	3-Epi6-C16	37.3 13	3654 C	G-Bathy1-C16	36.7	21128	CG-Bathy2-C16	64.7	12380
CG-Epi1-C17	35.8 27	7053 Ct	G-Epi2-C17	36.3	23686	CG-Epi3-C17	37.3	20166	CG-Epi4-C17	36.8 1	17247	CG-Epi5-C17	37.0 13	:467 CC	3-Epi6-C17	37.1 12	2311 C	G-Bathy1-C17	35.3	18158	CG-Bathy2-C17	65.9	12098
CG-Epi1-C18	37.3 26	5563 Ct	G-Epi2-C18	37.1	22912	CG-Epi3-C18	36.1	20159	CG-Epi4-C18	37.0 1	16984	CG-Epi5-C18	35.6 12	483 CC	3-Epi6-C18	34.7 11	L537 C	G-Bathy1-C18	37.1	17274	CG-Bathy2-C18	67.5	10866
CG-Epi1-C19	37.4 26	5142 Ct	G-Epi2-C19	34.9	19544	CG-Epi3-C19	34.0	19853	CG-Epi4-C19	34.8 1	16641	CG-Epi5-C19	36.8 12	:156 CC	5-Epi6-C19	36.0 11	L219 C	G-Bathy1-C19	36.7	11671			
CG-Epi1-C20	36.8 23	3896 CI	G-Epi2-C20	36.8	19543	CG-Epi3-C20	36.4	18732	CG-Epi4-C20	36.5 1	16164	CG-Epi5-C20	34.7 11	809 CC	5-Epi6-C20	36.5 11	1161 C	G-Bathy1-C20	38.8	10952			
CG-Epi1-C21	37.5 23	3263 CI	G-Epi2-C21	37.6	19118	CG-Epi3-C21	37.0	17297	CG-Epi4-C21	37.0 1	15964	CG-Epi5-C21	35.7 11	221 CC	5-Epi6-C21	39.7 10	0474 C	G-Bathy1-C21	36.0	10381			
CG-Epi1-C22	37.6 21	1975 Ct	G-Epi2-C22	36.3	18555	CG-Epi3-C22	39.9	15409	CG-Epi4-C22	36.7 1	14428	CG-Epi5-C22	37.5 1C	057 CC	3-Epi6-C22	35.9 10	0049 C	G-Bathy1-C22	37.6	10120			
CG-Epi1-C23	35.8 15	3014 Ct	G-Epi2-C23	36.4	17861	CG-Epi3-C23	37.5	14390	CG-Epi4-C23	37.6 1	12998			U U	3-Epi6-C23	35.5 9	955						
CG-Epi1-C24	35.6 17	7494 Ct	G-Epi2-C24	36.7	17829	CG-Epi3-C24	36.6	13412	CG-Epi4-C24	36.1 1	12867			ö	3-Epi6-C24	38.0 9	081						
CG-Epi1-C25	35.2 14	1922 Ci	G-Epi2-C25	35.3	17776	CG-Epi3-C25	37.1	13266						ö	5-Epi6-C25	34.1 8	023						
		Ŭ	G-Epi2-C26	35.5	17562	CG-Epi3-C26	38.7	13151						ö	3-Epi6-C26	36.9 5	465						
		Ũ	G-Epi2-C27	36.2	16729	CG-Epi3-C27	35.5	12130															
		Ũ	G-Epi2-C28	37.2	16047	CG-Epi3-C28	35.3	12129															
		Ũ	G-Epi2-C29	34.8	15419	CG-Epi3-C29	35.0	10830															
		Ũ	G-Epi2-C30	36.8	14321	CG-Epi3-C30	36.2	10429															
		Ũ	G-Epi2-C31	36.5	14306																		
		Ũ	G-Epi2-C32	36.9	13761																		
		Ũ	G-Epi2-C33	34.3	12919																		
		Ű	G-Epi2-C34	35.0	12691																		
		Ú	G-Epi2-C35	35.9	12415																		
		Ú	G-Epi2-C36	34.4	12304																		
		σì	G-Epi2-C37	35.9	12225																		
			G-Epi2-C38	36.5	10771																		
		υ č	G-Epi2-C39 G-Epi2-C40	35.9	11337																		
			G-Eni2-C41	33.9	10850																		
			G-Eni2-C42	26.2	10816																		
		υ ŭ 	G-Epi2-C43	38.4	10309																		
]				_																	

				Numbe	r of hous	ekeeping	I genes fo	und in the	e MG-III bin	s		Number	of housekee	sping gene	s found in	the CG-MG	HII bins	Number	of housekeepin ger	g genes found in omes	reference
			Epi1 Ep	i2A Epi	2B Epi2(C Epi3	Epi4 E	Epi5 E	pi6 Bath	ny1 Bath	y2 Epi	1 Epi2	Epi3 E	pi4 Epi	5 Epi(Bathy1	Bathy2	Cayman 92	Cayman 93	Guaymas 31	Guaymas 32
		% Completness estimation (Rees et al. 2007) % Completness estimation (Nuberteen et al. 2013)	80 5	4.3 57	1 5.7	48.6	34.3	2.9 5	1.4 54.	3 68.0	6 85.	7 80	62.9 3	4.3 2.5	57.	60 75.4	68.6 24.5	71.4	34.3 4E 2	91.4	94.3 27 o
		% Completness estimation (Autorised et al. 2013) % Completness estimation (Narasincarao et al. 2013)	83.0 4	5.3 54	7 57	45.3	39.6	7.6 54	1.7 60.4	4 58.5	t a	20.0	623 4	3.4 9.4	80 A	64.2	2 1.0 7 8 5	67 G	35.0	00.5	9.10 B.0.6
		# of genome-species inside the bin	3.07 1.	1.1	13 1.00	1.00	1.00	1.00	48 1.0	9 2.32	2 1.1	8 1.43		1	1.035	1.14	-	-	- F	1	1.86
Ribosomal		16S RNA		H	•						'		•		'	-	•				с (
genes		23S RNA		-	•	•	•	,	-	1	'	'		1	'	-	•				e
	COG classification	Function																			
	COG0008	GInS COG0008 472 Glutamyl- and glutaminyl-tRNA synthetases	3	-	•	•	•		-	•	2	2	•	1	1	1	•	1	-	1	1
	COG0012	Predicted GTPase, probable translation factor			'		•		- 0	' *		· c	•	•	'	2					
	C0G0015	Alas COGUUT3 8/9 Alanyi-tikiya synthetase		_	•	-	-		-			N 0		-	'	-					- c
	CUG0016	PheS COG0016 335 Phenylalanyl-tRNA synthetase alpha subunit	n c		-							7		· ~	• •		-				7 0
	COG0018	Map COG0010 377 Alginyi-trivia syrititetase Map COG0024 355 Mathionine aminopantidase	4 12						- 0		4					- e					4 0
	COG0048	RpsL COG0048 129 Ribosomal protein S12	0 64	- ,		•			2 1			4 -			-						7 2
	COG0049	RpsG COG0049 148 Ribosomal protein S7	2		•	•	-		2 1	•	-	-	-	-	-	+	•	-		1	2
	COG0051	RpsJ COG0051 104 Ribosomal protein S10	- 0	-	'	•		-	2	•	-		•	•	- '	-	•	-		÷- ,	7
	000000	IIIes CUGUDDI 933 ISOIeucyi-tikiNA synthetase Ilinia COC0071 146 Miclarular chanacha (amall haat chack mictain)	7 6		•				- +		· ~	-		· · ·		• -		÷		- +	7 0
	COGOORD	Rolk COG000 1.140 Princecular staperorie (Striati ricat Strok protein)	4 C							4		~	-					•			1 0
	COG0081	RplA COG0081 228 Ribosomal protein L1	1 -	- - -						-		- 1			•		-		-		4 +
	COG0085	RpoB COG0085 1060 DNA-directed RNA polymerase, beta subunit/140 kD subunit		-	•	1	1		3 1	3	•	-	1	-	+	-	-	٢		1	2
	COG0086	RpoC COG0086 808 DNA-directed RNA polymerase, beta subunit/160 kD subunit			•	-	+		3	e	•	-	-	-	-	-	-	-	-	1	-
	COG0087	RpIC COG0087 218 Ribosomal protein L3	2	•	•		•		1	e	-	-	+	•	'	•	-	-	-	-	2
	COG0088	RpID COG0088 214 Ribosomal protein L4	с u		•				'	en e				•	'						~ ~
	000000	Rpid COG0090 2/3 NU0S011(al protein LZ RpiV COG0091 120 Rihosomal protein L2	о <i>и</i> .							n e.							-				- ~
	COG0092	RpsC COG0092 233 Ribosomal protein S3	2	-	•	-				° °	-				'		-				101
	COG0033	RpIN COG0093 122 Ribosomal protein L14	9	-	•	-			1	e	-	2	-	•	1	•	-	-	-	-	5
	COG0094	RpIE COG0094 180 Ribosomal protein L5	9	-	•	1			1 1	3	-	2	1		1	1	1	1		1	2
	COG0096	RpsH COG0096 132 Ribosomal protein S8	9	-	•	-	•		1	3	-	2	1		-	-	-	-		1	2
500	COG0097	RplF COG0097 178 Ribosomal protein L6P/L9E	9		•	- ,	•			en o		2	- ,	•	- ,	- ,		- ,		,	5
	COG0088	RpsE COG0098 181 Ribosomal protein S5	9 0		•	-	•		-		- 0	2	-	•	-	-		-			74 0
	0000000	Rpsin COGUUSS 121 Riposomal protein 513	7 C		•	•					v c	•		- +	•	•					ч с
	0000102	RpiM COG0100-128 Ribosomal protein 311	4		•					4 65	4	-				-	-				4 0
	COG0103	RpsI COG0103 130 Ribosomal protein S9	-	-	•	•	-		-	ŝ	-	-		·	-	-	-			-	0
	COG0124	HisS COG0124 429 Histidyl-tRNA synthetase		. .	•	•		1	2 1	•	-				'	-	•	-		+	2
	COG0130	TruB COG0130 271 Pseudouridine synthase			•	•	•		-		1		•		•	-	•		1	1	2
	COG0143	MetG COG0143 558 Methionyl-tRNA synthetase	5		-		•		-	2				•	-	• •	-	-		+ •	0
	COG0164	KinhB COG0164 199 Ribonuclease Hil	.7 +				• •		- 2	• *				' .	•	N *	• •				N 0
	COG0180	Trib COG0180 314 Trubtonham/-tRNA svnthetase			•						- ო					-					7 -
	COG0186	RpsQ COG0186 87 Ribosomal protein S17	9	-	•	-			-	S	-	2	-		-		-	-	-	-	0
	COG0197	RpIP COG0197 146 Ribosomal protein L16/L10E	5	2	1	-	+		-	•	-	-	÷-	- -	'	-	•			1	2
	COG0200	RpIO COG0200 152 Ribosomal protein L15	9 u		•	-	•			en e		c7 c	-	•	'			. .			20
	00000201	Dect COGOZUT 430 Freproterin italisiocase suburiit dect	n -		•		•			0		v c		· · ·	•	-	-	-		- +	4 0
	COG0256	Rulk COG0256 125 Ribosomal modelin 1 18	- 9			-				· ~	-	2			•		-	-			4 4-
	COG0441	ThrS COG0441 589 Threonyl-tRNA synthetase		- - -		• •			3	· ·		- 1			-		• •		-		- 0
	COG0459	GroL COG0459 524 Chaperonin GroEL (HSP60 family)	-		•	-	+		1 1	-	2	-	+	-	-	2	-	-	Ļ	2	4
	COG0468	RecA COG0468 279 RecA/RadA recombinase	4			-	1	+	4 1	-	-	2	1	1	2	-	-	+	1	1	4
	COG0480	FusA COG0480 697 Translation elongation factors (GTPases)	2		•	•	-		2 1	'	-	-	-	- -	-	-	•	-		1	1
	COG0495	LeuS COG0495 814 Leucyl-tRNA synthetase	-	-	•	•			•		-	-		•	'	'		-	-	÷	2
	COG0522	RpsD COG0522 205 Ribosomal protein S4 and related proteins	2		•		-			- c	2 +		• •	-	•	•		-			6 6
	0000000	Valo COGO22 of 7 Vary-Informe synthesizes		* 			•			، د		- c			•		- •				v +
	000032	ITTE COGUSSZ SUST FRANSIALION INITIALION LACTOR 2 (IF-2; GTP 358)	-	-	'	'	'		-	-	-	7	-	-		-	-		-		-

Supplementary Table 4. Housekeeping genes found in the MG-III bins and the CG-MG-III bins (as Naransigarao et al. 2012).

Supplementary Table 5. Protein categories based on arCOG database.

Ĺ							Epi	pelagic CG-	MGIII bins			Pelagic C	G-MGIII bins	Other Euryar	chaea	
	arCOG clasification	gene	product	טט. Lassification	pFAM domain	cdc classifica	ation E	oi1 Eoi	2 Epi3	Epi4	Epi5 Epi6	6 Bathv1	Bathv-2	A.boneii T46	-2010	i nalassoarch aea
INFORMATION STC	IRAGE AND PROCESSING															
arCOG00415	L Replication, recombination and repair	RecA	RecA/RadA recombinase	OG00468	pfam14520, pfam0842	cd01123 TIGR022	236	1 1		1	1		t1	1	1	1
arCOG04143	L Replication, recombination and repair		DNA topoisomerase VI, subunit A	0G01697	pfam04406	cd00223		1 1	1	1	1 1	1		1	1	5
arCOG04455	L Replication, recombination and repair	HYS2	Archaeal DNA polymerase II, small subunit/DNA polymerase delta, subunit B C	OG01311	pfa m01336, pfam0404	cd04490,cd07386		1	-1	1	1	1	1	1	1	1
arCOG00459	L Replication, recombination and repair	Nth	Endolli-related endonuclease	0G00177	pfam00730, pfam1057	cd00056 TIGR010	083	·		1	•	- ,		5	1	1
arcocou41/	L Replication, recombination and repair	Gurd	Tune IIA tonoicomersee (DNA rursee/tono II tonoicomersee IV). A cubinite - C	0600188	nfam0631 nfam0300	CGU1334 11GRU22	23/			-	7 F				-	
arCOG00872	L Replication, recombination and repair	MPH1	Type IIA topoisonner ase torwaigyrase/topoint, topoisonner ase tv), A subunit C	00000100	pfam00370. nfam0037	CUUDIE/ TIGRONE	543 TIGR	1 1			+ +					n c
arc0G04447	L Replication, recombination and repair		Archaeal DNA polymerase II. large subunit	0G01933	pfa m03833. pfam09845	TIGR003	354	1	•		-		4	. +		1
arCOG00551	L Replication, recombination and repair		DNA replication initiation complex subunit, GINS15 family	0G01711		cd11714		1				-		1		
arCOG04990	L Replication, recombination and repair		Predicted endonuclease, contains HTH and PD-DExK nuclease domains		pfam14811			1		1		1	-1		1	2
arCOG00467	L Replication, recombination and repair	CDC6	Cdc6-related protein, AAA superfamily ATPase	OG01474	pfam13401, pfam0907	cd00009,cd087 TIGR029	928	2 1		1	1	1	1	1	2	9
arCOG02610	L Replication, recombination and repair	ScpA	Chromatin segregation and condensation protein Rec8/ScpA/Scc1, kleisin famC	OG01354	pfam02616			1			1	1	1	1	1	2
arCOG00439	L Replication, recombination and repair	MCM2	Predicted ATPase involved in replication control, Cdc46/Mcm family	OG01241	pfam14551, pfam0049	cd00009			1	1	1 1	1	1	1	1	2
arCOG00427	L Replication, recombination and repair	RecJ/Cdc45	5 Single-stranded DNA-specific exonuclease RecJ	OG00608								1	-1	2		
arCOG02258	L Replication, recombination and repair		RPA family protein, a subunit of RPA complex in P.furiosus	0603390				1 1		1	1 1			1	1	9
arCOG04582	L Replication, recombination and repair	Dpo4/DinP	> Family Y DNA polymerase	OG00389	pfam00817, pfam1179	cd03586		1	1	1	1	1			1	1
arCOG01166	L Replication. recombination and repair	MutL	DNA mismatch repair enzyme (predicted ATPase)	0G00323	pfam13589.pfam0111	cd00075.cd007 TIG R005	585	1	-	1	-	-		1	-	2
arCOG01165	L Replication: recombination and repair		DNA topoisomerase VI. subunit B	0601389	pfam02518. pfam0348	cd00075.cd008 TIGR010	52		-		1			-		4
arCOG01486	I Renlication recombination and renair	RnmV	55 rRNA maturation endonuclease (Ribonuclease M5) contains TOPRIM dom 10	0601658	nfam01751	rd01027			-		-	-				· -
arCOG04121	I Realization recombination and repair	RohR		0600164	nfam01351	cd07180 TIGR007	002	+ -	-		,	• -				- -
arCOG01170	1 Realization recombination and renair	Hold H	ATDase involved in DMA replication Hold Targe subunit	1020000	Dfam0000	COULDO TIGRO3	202	1 0	-	+	+					- 1
ar COG04371	Replication, recombination and repair	Gurb	Tune IIA tonoicomerase (DNA surase/t-ono II tonoicomerase IV) B subunit	0,600187	nfam07518 nfam0020		100	4 F	+ -	4 -	4	+ +				4 0
arcoom371 arcn604110	L Replication recombination and repair	DRI1	Fight in toposoniciase (Drive Sylase) topologine as a ryly bautaine of Fight volume as a ryly bautaine of C	0601467	pfam01896	cd04860 TIGR003	35	1 -	- ~	•	-				- 6	n u
ar COGONA88	Replication, recombination and repair	Neol	DNA nolymerses cliding rismo subunit (DCNA homolog)	0000000	nfam00705 nfam0274	COUNTRY TIGBURS		4 -	4 -	+	- +				4 c	
ar COG 01 070	L Replication, recomplication and repair	DINN		2600000	4/ 20111b1d/c0/00111b1d		050	-		-	-			-		7
arcococaca	L Replication, recomplication and repair	-		060067	pramUU293		1000012-010		-	,				7	-	T
arc0600787	L Replication, recombination and repair	-	UvrD/Reptamily helicase tused to exonuclease family domain	0602887	ptam12/05	cd0963/ 11GK012	249,11GR003	1 1	+	1					,	e
arc060152/	L Replication, recombination and repair	TopA	Iopoisomerase IA	0600550	ptam01/51, ptam0113	cd03362,cd001 11GK010	/5/	+								7
arCOG04050	L Replication, recombination and repair	FEN1	5-3' exonuclease	0600258	ptam00752, ptam0086	cd09867 TIGR036	574					-		1	-	2
arCOG00558	L Replication, recombination and repair	SrmB	Superfamily II DNA and RNA helicase	OG00513	pfam00270,pfam0027	cd00268, cd000 TIGR013	389					1		1	e	e,
arCOG00129	L Replication, recombination and repair		DNA modification methylase	0G00863	pfam01555, pfam01555	TIGR011	177					1		1		
arCOG00328	L Replication, recombination and repair	PolB3	DNA polymerase PolB3 C	OG00417	pfam03104, pfam0013	cd05781,cd055 TIGR005	592	1 1				2	1	2	1	4
arCOG00905	L Replication, recombination and repair	,	Uracil-DNA glycosylase	0601573	pfam03167	cd10031 TIGR007	758	1				2	T.	2		
arCOG01510	L Replication, recombination and repair	RPA1	Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subuni	OG01599		cd04491,cd04491		2 3	2	2	1 2	2	2	'n	2	ę
arCOG01072	L Replication, recombination and repair		NUDIX family hydrolase	0600494	pfam00293	cd03426		1	1	1	1	2				
arCOG03013	L Replication, recombination and repair	PRI2	Eukaryotic-type DNA primase, large subunit	OG02219	pfam04104	cd06560		1 1	1			2		1	1	1
arCOG00469	L Replication, recombination and repair	HolB	ATPase involved in DNA replication HolB, small subunit	0G00470	pfam13177,pfam0854	cd00009 TIGR023	397	1 1	1			2		1	2	2
arCOG00553	L Replication, recombination and repair	BRR2	Replicative superfamily II helicase	OG01204	pfam00270, pfam0027	cd00046,cd000 TIGR041	121	2 3	1	2	1	m		1	1	4
arCOG01894	L Replication, recombination and repair	Nfo	Endonuclease IV C	OG00648	pfam01261	cd00019 TIGR005	587	1 1			1		1	1	1	4
arCOG01526	L Replication, recombination and repair	TopG2	Reverse gyrase	OG01110	pfa m00270, pfam0175	cd00046,cd033 TIGR010	054	1 2		1			1	1		2
arCOG00464	L Replication, recombination and repair	AIKA	3-methyladenine DNA glycosylase/8-oxoguanine DNA glycosylase	0G00122	pfam07934, pfam0073	cd00056 TIGR005	588	1	1		1		-1	1	1	
arCOG03142	L Replication, recombination and repair		Nuclease of RNase H fold, RuvC/YqgF family					1	1				-1	1	1	1
arCOG00368	L Replication, recombination and repair	SbcC	ATPase involved in DNA repair, SbcC	OG00419	pfam13476, pfam1351	cd03240,cd001 TIGR006	511,TIGR021	58 1			1		+	m		
arCOG02897	L Replication. recombination and repair	MutS	Mismatch repair ATPase (MutS family)	0600249	pfam01624, pfam0518	cd03284 TIGR010	070	1			2			1	1	1
arCOG02724	L Replication. recombination and repair	Ada	Methylated DNA-protein cysteine methyltransferase	0600350	pfam01035	cd06445 TIGR005	683	-						Ļ	-	1
arCOG00397	L Replication: recombination and repair	ShcD	DNA repair exonuclease. ShcD	0600420	nfam00149	cd00840							-	2		
arCOG02257	L Replication: recombination and repair		RPA family protein: a subunit of RPA complex in P.furiosus	0603390									-			
arCOG08649	I Renlication recombination and renair		Tonoisomerase IR	0603569	nfam0.201.0 nfam0.1.02	rd00660 rd00659	Ī	-		-		T	,		-	
ar COG0045	L Replication, recombination and real	5000	DNA mimoro (hortorial tumo)	0000000	pfam13663	cu0000/cu00033	101	- r		-				-		ç
ar COG 04 071	I Replication, recomplication and repair	DING		0000000	ZOCOUTINE		121	7							-	7
arcoco1247	 Replication, recompination and repair Positionization seconditionities and secondi- 	,		0001202	pidiliou295	CUU3424 119 KUUU	700			-						
ar COCOD 40	 Inspiration, recontinuation and repair Dautization consultination and consider 	2002	Air-ucpenueir. Urwiigase Deamuibaiimimidine abatalinee	1100000	00TOINBId/C/04000000		14		-			T		-	ſ	r
ar cou 2640	 Replication, recomplitation and repair 	LI D		10000012	Pidinuou 2, pidinu		000	v •							7,	`
60T CD 0000 JP	 Replication, recomplitation and repair 	DIIId		0000047	pidilious29		c/0		+						-	
ar cocos 600	 Replication, recomplitation and repair 			0000000	CCCTOUIPId	2000 O I A	00.		+							
ar COG01115	I Replication, recompination and repair	AND		00004294	pidin0201		670	-						Ŧ	•	Ŧ
CTTODOOJa	L Replication, recomplication and repair	-		00001112	- auc Printing - 2000 Printing	1000	240	+								
ar COGOD 12	 Inspiration, recontinuation and repair Daulisation second institution and second 			101000	one tillbid ope tillbid		0/0	ł	$\left \right $			T				
arc000000	 Replication, recomplitation and repair 	-		TECTODO	0/0TOUIPId	-100010 210001			+					-		7 0
ar c0000240	 Replication, recomplitation and repair 	VIIIA		00000110	5 - 00000 - 5	UUUUU 100000000000000000000000000000000	200		+					-	-	7
ar c000046	I Replication, recomplication and repair	RITIL	2 Z2S FRIMA G2442 IN2-IIIEUIYIASE KIIIILI	91100000	hidrinu2920, pidrinu1/			+								
ar COGOUAE2	L Replication, recombination and repair	Nern	Urit superioriiity priospriorigur orose/ exorraciese	0000000	plain0100c	ed10.4.4.1	I	$\left \right $								
arcocono30	L Replication, recombination and real			0001000	plann01040	T+++OT DO	I	$\left \right $								
ar COC01074	 Dedication recombination and repair 			02010000	CTCTOILIBID	7000717 70000	205									
arcOG01206	Replication, recombination and repair	MIIC81		0601048	pfam03733 nfam13836	TIGROOF	20K									
ar COG04200	Replication, recombination and repair	Mng	a-methyladanina DNA alveoselase	0000000	pfam02745	COUNCIL TIGROUS	292									
df.cog04257	L Replication, recombination and repair	ENDO36	ס-דוופנוויזין ומענינוויוב בעיצא פוירטאיומאב דאהייייהיליאוא 2 העמניניניומים DMA מעניביומנים	001050	Plainvz4+0	CUUD40 1101000	/00	+								Ī
arc0001050	L Replication, recombination and repair	Dold Dold		2100000		ravousu -	I	$\left \right $								
ar COG05336	L Replication, recombination and repair I Benlication recombination and repair	HimA	Bacterial purcharid DNA-binding protein C	1140000	nfa m00316	rd13831 TIGROOD	187	+						+ +		
ar COG0533	Replication, recombination and repair	5	Torrim domain and Zn-finger domain	0.0000	pfam01751 nfam0113	CULTURE TIG FULL	57									
arCOG06670	Replication recombination and repair		Nuclease-related protein NERD family		pfam08378		10	╞								
arCOG 14866	L Replication, recombination and repair		Hhh domain containing DNA-binding domain		pfam12836			$\left \right $. +		
arCOG00280	L Replication, recombination and repair	ŀ	Hera helicase	OG00433	pfam01935		T	╞	-					2		
arCOG01290	L Replication, recombination and repair	SpIB	DNA repair photolyase	OG01533	pfam04055	cd01335		╞						2		
arCOG00329	L Replication, recombination and repair	PolB2	DNA polymerase PolB2, inactivated	OG00417	pfam00136	cd05531 TIGR005	592								1	1
arCOG00802	L Replication, recombination and repair	RecB	ATP-dependent exoDNAse (exonuclease V) beta subunit (contains helicase an C	OG01074	pfam00580, pfam13361	L TIGR027	785	╞							1	1

arCOG02895	-	Replication, recombination and repair	MutS2	DNA structure-specific ATPase involved in suppression of recombination, Mu	t COG01193	pfam00488 cd03243 TIGR01069					1	1	Г
arCOG04748	-	Replication, recombination and repair	UvrB	Excinuclease ABC subunit B, helicase	COG00556	pfam04851, pfam1481 cd00046, cd000 TIG R00631					1	1	Π
arCOG00305 arCOG04694		Replication, recombination and repair Replication recombination and repair	POL4	DNA polymerase IV (family X) Evcinuclasca ARC subunit A ATDace	COG01796	rd03271 rd0321TIG800630						m r	T
arCOG00462		Replication, recombination and repair	MutY	A/G-specific DNA glycosylase	COG01194	pfam00730 cd00056 TIGR01084							Т
arCOG00873	-	Replication, recombination and repair		Excinuclease-associated helix-hairpin-helix domain	COG00322	pfam01541,pfam0215 cd10434,cd065 TIGR00194					-		П
arCOG03646 arCOG04513		Replication, recombination and repair Replication recombination and repair	XseB XseA	Exonuclease VII large cubinit Evonuclease VII large cubinit	COG01722	pfam02609 TIGR01280 TIGR01280 TIGR00237							Т
arCOG04754	-	Replication, recombination and repair	Lig	NAD-dependent DNA ligase	COG00272	pfam01653, pfam0312 cd00114, cd000 TIG R00575							Γ
arCOG07300		Replication, recombination and repair		Uncharacterized protein associated with inactivated PolB-like polymerase							1		
arCOG01082 arCOG01305		Replication, recombination and repair Replication recombination and repair	+	Tonoisomerase DNA hinding C4 zinc finger fised to uncharacterized N-termin	COG00494	pram00293 cd02885					T		Τ
arCOG02896	-	Replication, recombination and repair	MutS	Mismatch repair ATPase (MutS family)	COG00249	pfam01624, pfam0518 cd03284 TIGR 01070							Γ
arCOG02942	-	Replication, recombination and repair	RnhA	Ribonuclease HI	COG00328	pfam13456 cd09279						1	Π
arCOG04753		Replication, recombination and repair	UvrC	Excinuclease ABC subunit C	COG00322	pfam01541, pfam0215 cd10434, cd098 TIG R00194	,		,	,		4	T
arCOG04244 arCOG04280	××	Transcription	RPB10 NagC	DNA-directed RNA polymerase, subunit N (RpoN/RPB1U) Transcriptional regulator/sugar kinase	C0G01644 C0G01940	pfam01194 cd00012 TIGR00744 TIGR00744	1 1		1	न न			T
arCOG01863	×	Transcription	-	Predicted transcription factor, homolog of eukaryotic MBF1	COG01813	pfam01381 cd00093 TIGR00270			4 			2	Γ
arCOG04061	¥ :	Transcription	EGD2	Transcription factor homologous to NACalpha-BTF3	COG01308	cd14359 TIGR00264	1 2 1	1 1	1	с -	т.		П
arCOG01753	× 1	Transcription	Ssh10b	Archaeal DNA-binding protein	COG01581	pfam01918 TIGR00285				. е	•	2	Ţ
arcog04257	××	I ranscription Transcription	Rpoc/Rpo	 DNA-directed KNA polymerase subunit b DNA-directed RNA polymerase subunit A¹ 	COG00086	pram04565, pram0456 cd00653 116 K036 /0 pfam04997, pfam0062 cd02582 TIG R02390			п п			m	Т
arCOG02613	×	Transcription	-	Chromosome segregation and condensation protein B	COG01386	pfam04079 TIGR00281						,	Т
arCOG03112	¥ :	Transcription		Fic family protein	COG03177	pfam02661 TIGR02613		-	1, 1	н,		e	Т
arCOG01128	× ×	I ranscription Transcription	TenA	otti UNA-directed KNA polymerase subunit A Transcriptional activator TenA	COG00819	pramu4998 cdUb5.28 [1]GR04389 pram03070 TIGR04306	7	-			-	7	T
arCOG00579	×	Transcription	RPB9	DNA-directed RNA polymerase, subunit M/Transcription elongation factor TF	COG01594	TIGR01384	1 1 1	1 1	1 1		1 1	2	
arCOG04366	×	Transcription	-	ArsR family transcriptional regulator	COG04860	pfam09824	1 1 1	1	1 1		1		Π
arCOG01684	× >	Transcription	-	Transcriptional regulator, ArsR family	COG01777	pfam01022 cd00090,cd00090			2		1	4	Τ
arCOG01016	2 ×	Transcription	RD04	DNA-directed RNA polymerase. subunit Rpo4/RpoF	COG01460	premotes 1104003	1 1				1		Τ
arCOG01920	×	Transcription	NusG	Transcription antiterminator NusG	COG00250	pfam03439 cd09887,cd060 TIGR00405	1 2				1	2	Γ
arCOG00675	¥	Transcription	RPB7	DNA-directed RNA polymerase, subunit E//Rpb7	COG01095	pfam03876,pfam0057 cd04331,cd044 TIGR00448	1		-1		1 1		Π
arCOG04077	¥	Transcription	Spt4	Transcription elongation factor Spt4/RpoE2, zinc finger protein	COG02093	pfam06093	, ,				, ,		Τ
arCOG04258	× >	Transcription	RPB5 Becc/Bec2	DNA-directed RNA polymerase, subunit H, RpoH/RPB5	COG02012	pfam01191							Τ
arCOG02611	~ ~	Transcription	кроб/кро. -	 DNA-directed RNA polymerase suburit cyomega Predicted transcriptional regulator: containd two HTH domains 	CDG03398	pramu 1192			 -	6			Т
arCOG07561	×	Transcription		Predicted membrane-associated trancriptional regulator	COG02512		2 - 1		2 1	- ~	2 7	2	Τ
arCOG01580	×	Transcription		DNA-binding transcriptional regulator, Lrp family	COG01522	pfam13412, pfam0103 cd00090	1 1 1	1 1	2		5 2	9	Π
arCOG05161	¥	Transcription	WecD	Acetyltransferase (GNAT) family	COG00454		1	1	1 2				Π
arCOG04111	× >	Transcription	RPB11 CDT1E	DNA-directed RNA polymerase, subunit L	COG01761	pfam13656 cd06927	1 2 7		 	~		Ŧ	Τ
ar COG01981	~ ~	Transcription	SF1A7	Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor	COG01405	prerito0332, prerito033 C004316			7 t		- T		Т
arCOG02099	- -	Transcription	TroR	Mn-dependent transcriptional regulator (DtxR family)	COG01321	pfam01325, pfam02742, pfam04023	2 4		n m	,	- 2	m	Т
arCOG04060	×	Transcription	-	Predicted transcriptional regulator	COG01709	pfam01381 cd00093	1 1 1	L 1 1		1	1		
arCOG00770	×	Transcription	DinG	Rad3-related DNA helicase	COG01199	pfam06733, pfam13307 TIGR00604	1			1	1 1	2	Ι
arCOG04241	¥ :	Transcription	RpoA/Rpo	11 DNA-directed RNA polymerase subunit D	COG00202	pfam01193 cd07030	,	1		е	- -	1	T
arcocotean	∠ >	Transcription	Arsk	Transcriptional regulator containing FLFL domain, ArsK tamily Transcriptional regulator containing HTH domain, ArcD family	COG00640	Marrit 2840 caudoo							Т
arCOG02038	<u> </u>	Transcription	-	Sugar-specific transcriptional regulator TrmB	COG01378	0000000 7701181d	4			н е			Т
arCOG02037	¥	Transcription		Sugar-specific transcriptional regulator TrmB	COG01378	pfam01978	1 1 1	1			1	2	Π
arCOG04686	×	Transcription	VacB	Exoribonuclease R	COG00557	pfam00773 TIGR02063	1 1	1 1			1		Π
arCOG04341	¥ :	Transcription	RPC12	DNA-directed RNA polymerase, subunit RPC12/RpoP (contains C4-type Zn-fit	gC0G01996		,				_	ć	T
arCOG03182	× >	Transcription	Mark	Transcriptional regulator, Mark tamily	COG01846	ptam01047 cd00090	-1 •					77 0	Т
arcog05671	~ ~	Transcription	Cod	Transcriptional regulator luxe family	COG02771	OCONODA OCOTALIBIA					7	n	Т
arCOG00839	×	Transcription	WecD	Acetyltransferase (GNAT) family	COG00454	pfam13508 cd04301 TIGR01575						۲,	Γ
arCOG00608	×	Transcription	-	Predicted transcriptional regulator with C-terminal CBS domains	COG03620	pfam01381, pfam0057 cd00093, cd046 TIGR 03070, TIGI	R01137				1		Π
arCOG04248	×	Transcription	SIR2	NAD-dependent protein deacetylase, SIR2 family Transcriptional regulator Mare family, contains HTH domain	COG00846	pfam02146 cd01413							
arCOG00002	- -	Transcription		Transcriptional regulator, PadR family	COG01695	pfam03551 TIGR03433							Τ
arCOG00492	×	Transcription	ARO8	Transcriptional regulators containing a DNA-binding HTH domain and an ami	COG01167	pfam00155 cd00609 TIGR03947					1		Π
arCOG00610	¥ :	Transcription		Predicted transcriptional regulator, contains C-terminal CBS domains	COG02524	pfam03444, pfam0047 cd04588 TIGR00331, TIGF	R01302						T
arc0600742	× >	Transcription	Mark	I ranscriptional regulator, Mark tamily Dradicted transcriptional regulator	CUGU1846	pram13601 cd00090							Τ
arCOG00845	< ×	Transcription	WecD	Acetyltransferase (GNAT) family	COG00454	pfam00583 cd04301 TIGR01575							Τ
arCOG01055	×	Transcription		Transcriptional regulator, contains HTH domain	COG03432	pfam14947					1		Π
arCOG01586	¥	Transcription		DNA-binding transcriptional regulator, Lrp family	COG01522	pfam13404, pfam1340 cd00090, cd00090					1		Π
arCOG01683 arCOG01685	××	Transcription Transcription	ArsR	Transcriptional regulator containing HTH domain, ArsR family Transcriptional regulator containing HTH domain. AreR family	COG00640	pfam01978 cd00090 hfam12840 cd00090					-1 -		
arc0601760	~ ~	Transcription	NisA	Transcription elongation factor	COG0040	premitizeto cd02134.cd021TIGR01952							Т
arCOG01804	×	Transcription		Transcriptional regulator, contains HTH domain	COG03888	pfam13412 cd00090,cd13553						_	П
arCOG02808	¥	Transcription		Transcriptional regulator, contains HTH domain	COG04742						1		I
arCOG03698	×	Transcription	- TIDAO	Transciptional regulator, contains HTH domain	COG04344	pfam10007 cd00000 TIGB01241 TIG	207307						
arcoG04152	: ¥	Transcription	2	Predicted transcriptional regulator	COG01395	premotive curves memory premo	1002001					-	Т
arCOG04399	×	Transcription	ŀ	Predicted transcriptional regulator	COG01497								П
arCOG04479	~	Transcription	-+	Transcriptional regulator containing an HTH domain fused to a Zn-ribbon	COG03357								Т
arCOG04939 arCOG07764	××	Transcription Transcription	+	Transcriptional regulator, contains wHTH domain Dradicted transcriptional regulator	SC.00517 COG01318							+	Т
arcog13514	<u>د</u> ×	Transcription	+	Transcriptional regulator, contains N-terminal RHH domain	000000000000000000000000000000000000000								Т
			1										ļ

arCOG01117	×	Transcription -	Lrp/AsnC family C-terminal domain	0G01522	fam01037			2 3	9
arCOG00826	×	Transcription	Acetyltransferase (GNAT) family	0G00454	iam00583 cd04301 TIGR01575 cd04301			2	1
arCOG00001	¥	Transcription -	Transcriptional regulator, PadR family	OG01695	fam03551			2	
arCOG00998	¥ :	Transcription LSM1	Small nuclear ribonucleoprotein (snRNP) homolog	0G01958	fam01423 cd01726			2	
arcoc01345	¥ 1	Transcription	Predicted transcriptional regulator	06013388	5m012.25 mfsm0.27/1.2			7 (
	2 3	Transcription Chai	Commonant of chamotavic svetam accordated with archaellum contains Chaello	1751000				4 6	
arc0602394	∠ >	Transcription Cher	Component of chemotaxis system associated with archaeilum, contains cher-jo	0602409				7 (
arCOG03067	<u>د</u> ۲	Transcription ArsR	Transcriptional regulator containing HTH domain. Area family	0600640	fam1 2840			4 67	
arCOG02983	<u> </u>	Transcription	Cold shock protein. CspA family	0601278	am00313 cd04458 TIGR02381			, -	2
arCOG01875	<u> </u>	Transcription -	ParB-like nuclease domain	0G01475	am02195 cccccc fielder				4 m
arCOG02644	×	Transcription	Transcriptional regulator, TetR/AcrR family	OG01309	iam00440 TIGR03613 TIGR03613			1	
arCOG05152	×	Transcription -	Transcriptional regulator, contains HTH domain					1	
arCOG03748	¥	Transcription	Transcriptional regulator, MarR family	OG01846				2	1
arCOG02271	¥	Transcription -	Transcriptional regulator, contains HTH domain	OG03413	fam04967			2	4
arCOG02274	× :	Transcription	Transcriptional regulator, contains HTH domain	0603413	tam04967			m	,
arc0604280	¥ 1	Transcription	Iranscriptional regulator, contains HLH domain Superfamily II DNA/DNA helicase SNE2 family	0603413	2001 bfsm0765 cd00178 cd000 TIC 0015.87				
	2 -	Translation ribocomal structure and biogenesis	OUPERTAINING INVEXTIGATIONS SINCE TAINING ADMAIN ADMAIN	0600411				•	-
arc0600423	-	Translation, ribosonial structure and blogenesis - Translation, ribosomal structure and blogenesis -	ASCH UDITATIII, PLEUTCHEU KNA-DITUTING UDITATI Disortibonuclease NrnB or cAMP/cGMP phosphodiesterase. DHH superfamily C	0602404	am04200 cu00332				
arCOG00910	-	Translation, ribosomal structure and biogenesis	Predicted RNA met hvlase	0602263	am13659	1 1	-	1	1
arCOG00953	-	Translation, ribosomal structure and biogenesis	Homolog of Wybutosine (vW) biosynthesis enzyme. Fe-S oxidoreductase	OG00731	fam04055			1	
arCOG00990	-	Translation, ribosomal structure and biogenesis -	Queuine tRNA-ribosyltransferase, contain PUA domain	0G01549	am14810,pfam01472 TIGR00432 1 1	1 1 1		1 1	2
arCOG00991	-	Translation, ribosomal structure and biogenesis	tRNA modification protein, contains pre-PUA and PUA domains	OG01370	fam14810,pfam01472 TIGR00432 1 1 1		-1	1 1	2
arCOG01042	-	Translation, ribosomal structure and biogenesis	Exosome subunit, RNA binding protein with dsRBD fold	0G01325	am01877			1	1
arcOc01695	-	Translation, ribosomal structure and piogenesis - Translation ribosomal structure and biomenesis -	Prodicted comments of the ribecome cuality control (POC) complex VIoA /Talo	06012534	am012832 nfam016270 1		-	 	7 0
ar COG01861	-	Translation, ribosonial structure and blogenesis -	13-51 exortibonicies e YbaM. can participate in 235 rBNA maturation HD super	0603481	am01966 TIGR00277 L		-	7 6	7
arCOG04130	-	Translation, ribosomal structure and biogenesis	Predicted RNA-binding protein	OG01491	am04919 1 2	1 1 1		1	1
arCOG04318	_	Translation, ribosomal structure and biogenesis -	Predicted RNA-binding protein of the translin family	OG02178	fam01997 cd14820			1	
arCOG01255	-	Translation, ribosomal structure and biogenesis AlaS	Alanyl-tRNA synthetase	OG00013	fam01411, pfam0797 cd00673, cd077 TIGR03683 1 1 2	1 1 1	1	2 1	1
arCOG01254	-	Translation, ribosomal structure and biogenesis AlaX	Ser-tRNA(Ala) deacylase AlaX (editing enzyme)	0G02872	fam01411,pfam07973 TIGR03683	1 1		1 1	2
arCOG01256	-	Translation, ribosomal structure and biogenesis AlaX	Ser-tRNA(Ala) deacylase AlaX (editing enzyme)	0G02872	TIGR03683			1	
arCOG00487	_	Translation, ribosomal structure and biogenesis ArgS	Arginyl-tRNA synthetase	0600018	ta m00750, ptam0574 cd00671, cd079 TIGR00456 1 1	1 1	ŀ	, 1 1	1,
arc0600405	-	Translation, ribosomal structure and piogenesis Asno	Aspartyl/asparaginyl-tKNA synthetase	0600017	ramu1336, pramuu15 (d0431b, cd00 / 116K00458	, , , ,	-		
arC0G00407		Translation, ribosomal structure and blogenesis AsnS	Aspartyl/asparaginyi-tRNA synthetase	0000017	am01336, pram0015 (cd04319, cd00/7 TIG R00457 1 2 2				2 4
	-	Translation, ribosomal structure and biogenesis bud32 Translation siboromal structure and biogenesis CCA1	ItkiNA A-57 threeoffyticar barnoyt transferase component budst C *DNLA muchaetidi.item.neferance (CCA addime common) C	0601746	(affi00009 Cd03131 11GK03724 2 2 2			7	4
arCOG07197		Translation, ribosonial structure and blogenesis CCA1 Translation ribosomal structure and blogenesis [Gai121	knive itucieotiuyitiatistelase (Confautrig etizyitie)	0601617	am08617 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		-		
ar COGODE757	-	Translation ribosonial structure and biogenesis Celd	Everyone complex (Verizz BOU2ZNAEL) Everyone complex PNA-binding motelin Cold contains S1 and 7 miltion domain	1101000	1110001/ T T T T T	- -	-	+ +	÷
arCOG00486	-	Translation, ribosomal structure and biogenesis CvsS	Cysteinvi-tRNA synthetase	0600215	am01406.pfam0919_cd00672.cd079_TIGR00435111		7 7	1 1	2
arCOG04112	-	Translation, ribosomal structure and biogenesis DPH2	Diphthamide synthase subunit DPH2	OG01736	am01866 TIGR00322 1 1	1 1 1 2		1	2
arCOG04161	_	Translation, ribosomal structure and biogenesis DPH5	Diphthamide biosynthesis methyltransferase	0G01798	am00590 cd11647 TIGR00522 1	1		1	2
arCOG00035	-	Translation, ribosomal structure and biogenesis Dph6	Diphthamide synthase (EF-2-diphthineammonia ligase)	0G02102	am01902 cd01994 TIGR03679 1 1	1		1	1
arCOG00358	-	Translation, ribosomal structure and biogenesis DRG	Ribosome-interacting GTPase 1	OG01163	fam02824 cd01896,cd01666 1 2	1 1 1 2		1 1	1
arCOG00604	-	Translation, ribosomal structure and biogenesis DusA	tRNA-dihydrouridine synthase	OG00042	fam01207 cd02801 TIGR00737			1	9
arCOG04332	-	Translation, ribosomal structure and biogenesis EbsC	Cys-tR NA(Pro)/Cys-tRNA(Cys) deacylase, ybaK family C	OG02606	fam04073 cd04333 TIGR00011 cd04333				2
arCOG01988	-	Translation, ribosomal structure and biogenesis EFB1	Translation elongation factor EF-1beta	0G02092	fam00736 cd00292 TIGR00489 1 1	1 1 1		1 1	
arCOG04277	-	Translation, ribosomal structure and biogenesis Efp	Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A) C	0G00231	fam08207,pfam0128 cd04467 TIGR00037 1	1 1		1	
arCOG04122	-	Translation, ribosomal structure and biogenesis Emg1	rkNA pseudouridine-1189 N-methylase Emg1, Nep1/Mra1 family	0601/56		,	•	,	,
arCOG01742	-	Translation, ribosomal structure and biogenesis eKF1	Peptide chain release factor eKF1	0601503	ram03463, ptam03464, ptam03465 TIGR03676 4	, , ,			
arCOG04312	-	Translation, ribosomal structure and biogenesis Fcf1	rkNA-processing protein FCF1, contains PIN domain	0601412	Cd09879 ICCC011 1				2 *
arcocol 550	-	Translation, ribosomal structure and plogenesis FtsJ Translation ribosomal structure and biomenesis FtsJ	Z35 TRIVA UZ35Z (TID058-2 -U)-TITEUTIYIGSE KIITIE/ F1SJ Transflation alonnation factor G EE-G (GTDace)	0600480	amur/28 cau2440 Higk00438 L L L				- r
arCOG02466	-	Translation, ribosonial structure and blogenesis - Lush	RNA-binding protein involved in rRNA processing	0603277					7
arCOG01719	-	Translation. ribosomal structure and biogenesis GatE	Archaeal Glu-tRNAGIn a midotransferase subunit E (contains GAD domain)	0G02511	am02934.pfam02637 TIGR00134 1 1	1		1	m
arCOG01563	_	Translation, ribosomal structure and biogenesis GCD11	Translation initiation factor 2, gamma subunit eIF-2 gamma, GTPase C	0G05257	tam00009, pfam0917 cd01888, cd036 TIGR03680 1	1 1		1	4
arCOG00978	_	Translation, ribosomal structure and biogenesis GCD14	tRNA(1-methyladenosine) methyltransferase	0G02519	am08704 cd02440 TIGR03534 1 1 1	1 1 2 1	1	1 1	1
arCOG01124	-	Translation, ribosomal structure and biogenesis GCD2	Translation initiation factor 2B subunit, eIF-2B alpha/beta/delta family C	OG01184	am01008 TIGR00511			1	
arCOG01640	-	Translation, ribosomal structure and biogenesis GCD7	Translation initiation factor 2, beta subunit (eIF-2beta)/eIF-5 N-terminal domdC	OG01601	am01873 TIGR00311				
arCOG01616	-	Translation, ribosomal structure and biogenesis GEK1	D-aminoacyi-tRNA deacylase, involved in ethanol tolerance	0601650		, ,			ć
arc0604302	-	Translation, ribosomal structure and biogenesis 19105 Translation siboromal structure and biogenesis 1913	Ghadi ABNA sustistant (class II)	0600433	iamuU/45,piamu235 (ฉนช26/ 116KUU405 2 2 2				7
ar COGO0267	-	Translation ribosonial structure and biogenesis GAD1	Diposome-hindian ATBase VrhE GTD1/OBC family	000012	aiii00367, piaii10314 tu00774, tu009 iig N00367 1 1	-	-	+ +	
arC0600109	-	Translation, ribosonial structure and blogenesis OTEL	Methylase of nolynenticle chain release factors	0602890	am13659 cd02440 TIGR00537 1 2		-	- -	7
arCOG00404	-	Translation. ribosomal structure and biogenesis HisS	Histidul-tRNA svnthetase	0G00124	[am13393.pfam0312[cd00773.cd008]TIGR00442 1 1 1			1	2
arCOG00807	_	Translation, ribosomal structure and biogenesis IIeS	Isoleucyl-tRNA synthetase	0G00060	fam00133, pfam0826 cd00818, cd008 TIGR00392 1	1		1	4
arCOG01179	_	Translation, ribosomal structure and biogenesis InfA	Translation initiation factor 1 (IF-1)	OG00361	fam01176 cd05793 TIGR00523 1 1 1	1		1 1	2
arCOG01560	-	Translation, ribosomal structure and biogenesis InfB	Translation initiation factor 2 (IF-2; GTPase)	OG00532	fam00009,pfam0314 cd01887,cd037 TIGR00491 1 2	1 1 1	-1	1 1	
arCOG01183	-	Translation, ribosomal structure and biogenesis Kae1p/Tsa	2 Subunit of KEOPS complex, tRNA A37 threonylcarbamoyltransferase, contains C	OG00533	fam00814 TIGR03722	1 1 1		_	2
arCOG04063	-	Translation, ribosomal structure and biogenesis KptA	RNA:NAD 2'-phosphotransferase	0G01859	am01885 1 2	1 1		1, 1	2
arcoconeon	-	Translation, ribosomal structure and blogenesis Krr1	I KINA processing protein Krr1/Pho1, contains KH domain	0G01094	5				7
arC0601736	-	Translation, ribosomal structure and plogenesis Leus Translation ribosomal structure and biogenesis ligT	2-54, RNA ligase	0601514	amuutas,piamuezo cauveiz,cauva nekuuaso I. I. I. I.		-		
arCOG00485	-	Translation, ribosomal structure and biogenesis LvsS	Lusvi-tRNA svnthetase. class C	0G01384	am01921 cd00674 TIGR00467 1		11	1	1
arCOG00408	-	Translation, ribosomal structure and biogenesis LysU	Lvsvi-tRNA svnthetase (class II)	0G01190	am01336.pfam0015 cd04322,cd007 TIG R00499			1	
arCOG01001	-	Translation, ribosomal structure and biogenesis Map	Methionine aminopeptidase	0G00024	am00557 cd01088 TIGR00501 1 2	1 1 1 1 2		1 1	2
arCOG00810	-	Translation, ribosomal structure and biogenesis MetG	Methionyl-tRNA synthetase C	OG00143	fam09334, pfam0826 cd00814, cd079 TIGR00398 1	1 1	1	1 1	2
arCOG01358	Ţ	Translation, ribosomal structure and biogenesis MiaB	2-methylthioadenine synthetase	0G00621	am00919, pfam0405 cd01335 TIGR00089 1 1				
arCOG02286	-ſ-	Translation, ribosomal structure and biogenesis INCLE	Ribosome biogenesis protein, NOL1/NUP Z/ITMU TAMIIY	0603270	am13636 1 1 1 1		,	, 1	,
arC0G04149	-	Translation, ribosomal structure and biogenesis INIVIUS	NMD protein affecting ribosome stability and mRNA decay	0G01499	iam04981. 1 1 1	-	2	1	

arCOG00721	J Tran	anslation, ribosomal structure and biogenesis Nob1 Endonuclease Nob1, consists of a PIN domain and a	Zn-ribbon module COG01439		cd09876		1				-1		1	1	
arCOG00078	J Tran	anslation, ribosomal structure and biogenesis NOP1 Fibrillarin-like rRNA methylase	C0G01889	pfam01269			1	1 1					1	1	-
arCOG00906	J Tran	anslation, ribosomal structure and biogenesis Nop10 IrRNA maturation protein Nop10, contains Zn-ribbor	1 domain COG02260	pfam04135			1	1	1	1 1	1		1 1	2	2
arCOG04414	J Trar	anslation, ribosomal structure and biogenesis Pcc1 Subunit of KEOPS complex (Cgi121BUD32KAE1)	COG02892				£1 -		-						
arcogon /41	L Tar	ansiation, ribosomai structure and biogenesis PelA Release factor eRF1 and atom ribosomal structure and biogenesis DbAs DbAsv/Jabavia BNA swithbatses alpha subunit	COG016	pramu3463, pramu346	4, pramu3465			7 I	T	1				7 -	~ -
arcogona12	Tran	ansiauoni, ribosonial su ucure and biogenesis Prices Pricenylalany-tuwa synthetase apria subunit anciation ribosomal structure and biogenesis PheT Phenvlalamul-tRNA supthetase heta subunit	COG0013	pidiil0.1403 nfam03483 nfam0348	rd00769			7 T		+		Ŧ			
arCOG00784	Tran	anaouon, nuosonnai su uutare and biogenesis Princi Principalany-tutata synthetase beta suounit and ation ribosomal structure and biogenesis POP4 RNase P/RNase MRP subunit n/9	COG01588	pfarm01868		T/1000101		- T		4	-	-	-	4 C	-
arCOG01365	J Tran	anslation, ribosomal structure and biogenesis POP5 RNase P/RNase MRP subunit POP5	C0G01369	pfam01900			-								
arCOG00402	J Tran	anslation, ribosomal structure and biogenesis ProS ProlyI-tRNA synthetase	C0G00442	pfa m00587, pfam0312	cd00778,cd008 1	TIGR 00408	1	1 1	1	1 1	1		1 1	3	~
arCOG04228	J Trar	anslation, ribosomal structure and biogenesis Pth2 Peptidyl-tRNA hydrolase	COG01990	pfam01981	cd02430 1	TIGR00283	1	1 1	1	1	н,		1 2	2	2
arCOG01015	J Trar	anslation, ribosomal structure and biogenesis Pus10 ItRNA U54 and U55 pseudouridine synthase Pus10	COG01258			TIGR 01 2 13	-1	3 1	1	1 1	-		1	-1	1
arCOG00039 arCOG00039	J Tran	anslation, ribosomal structure and biogenesis QueA S-adenosylmethionine:tRNA ribosyltransferase-ison anslation ribosomal structure and biogenesis QueC 7-cvano-7-deazaguanine svuthase (queuosine biocv	nthesis) COG00603	ptam02547 nfam06508	rd01995	TIGR00113		-				-	1	-	
arCOG04210	J Tran	anslation, ribosomal structure and biogenesis QueFC NADPH-dependent 7-cyano-7-deazaguanine reduct	ase QueF, C-terminal dom COG00780	pfam14489		TIGR03139						1			
arCOG04125	J Tran	anslation, ribosomal structure and biogenesis RCL1 RNA 3 ⁻¹ terminal phosphate cyclase	COG00430	pfam01137	cd00874 1	TIGR03399							1		
arCOG00842	J Trar	anslation, ribosomal structure and biogenesis RimL Acetyltransferase, RimL family and the inhermal structure and historical Bits Translation initiation feature BUI contains Fo. 5 and	AAA ATDace domains COG01570	pfam13302	- CCOba - CCCOba	100000			,	-	-	н ,	,	ų	
arcOG05111	Tran	ansiation, ribosofiiai structure and piogenesis - KII. Iransiation imitation factor KLI, contains Fe-5 and a anciation-ribosomal structure and biogenesis - RIMH	AAA+ ATPase domains COGUI245	pramouuus, pramouuu nfam03590	1 causzs /, causz 1	TIG R00246	-	т 7	-	-	-	4	 -	Þ	
arCOG06660	J Tran	anslation, ribosomial structure and biogenesis Riuk Pseudouridylate synthase, 235 RNA-specific	COG00564	pfam00849	cd02869 1	TIG R00005							4	2	2
arCOG00547	J Tran	anslation, ribosomal structure and biogenesis RnjA mRNA degradation ribonuclease 11/12	COG00595	pfam12706		TIGR03675, TIGR0.	12651						1		
arCOG00501	J Trar	anslation, ribosomal structure and biogenesis Rnz Ribonuclease Z, beta-lactamase superfamily hydrol	ase COG01234	pfam12706		TIGR02651	ен ,	,	1		,		1	2 .	2
arcOG04209	L Tran	ansiation, ribosonial structure and plogenesis – Kipi – Kibonuciease PT anclation ribosomal structure and biogenesis – RPI 15A – Ribosomal protein 115F	COG01632	plam00827 nfam00827	0051100	C0U2UNU	-		-	-			1	4 (1	+ ~
arCOG00780	J Tran	anslation, ribosomal structure and biogenesis RPL18A Ribosomal protein L18E	COG01727	pfam00828		TIGR01071	1	1		1			. "	'n	
arCOG04089	J Tran	anslation, ribosomal structure and biogenesis RPL19A Ribosomal protein L19E	COG02147	pfam01280	cd01418		1	2 1		1	1	1	1 1	1	_
arCOG04175	J Tran	anslation, ribosomal structure and biogenesis RPL20A Ribosomal protein L20A (L18A)	C0G02157	pfam01775									1 1		
arCOG04129	J Trar	anslation, ribosomal structure and biogenesis RPL21A Ribosomal protein L21E	COG02139	pfam01157	0000	I		2 1	1	1		,	1	-1	1
arcOG01752	L Tran	ansiation, ribosonial structure and plogenesis Krt.24A Kibosonial protein 124E anclation ribosomal structure and biogenesis RPI 30 Ribosomal protein 130F	COG02075	plam01246 pfam01248	cau0472		-		-				 -	6	
arCOG04473	J Tran	anslation, ribosomal structure and biogenesis RPL31A Ribosomal protein L31E	C0G02097	pfam01198	cd00463		1	2	. 1	1		. 4			
arCOG00781	J Tran	anslation, ribosomal structure and biogenesis RPL32 Ribosomal protein L32E	COG01717	pfam01655	cd00513		1	2 1		1	1	1	1 1	1	_
arCOG04177	J Tran	anslation, ribosomal structure and biogenesis RPL39 Ribosomal protein L39E	COG02167	pfam00832									1	1	1
arCOG04049	J Trar	anslation, ribosomal structure and biogenesis RPL40A Ribosomal protein L40E	COG01552			I	•	,	1,	,			-		-
arCOG04109	J Trar	anslation, ribosomal structure and biogenesis RPL42A Ribosomal protein L44E	COG01631	pfam00935				,	1	1			.,	2 7	~
arCOG01751	I Tran	ansiation, ribosofiiai structure and piogenesis krk43A kibosofiiai pioteini L37AE/L43A anciation rihocomal structure and hineenesis Ru 7Ae Rihocomal nontain 17AE	COG01358	pidm01248			-1	+ +		-		÷	1 F	7 -	7 -
arCOG04289	I Tran	ansiation, ribosotilat structure and biogenesis – hprizee – hibosotilat protein LZAE and afton ribosomal structure and biogenesis – RolA – – Ribosomal nrotein 11	COG0081	pfam00687	rd00403	TIGR01169		+ +		-				- ~	
arCOG04067	J Tran	anslation, ribosomal structure and biogenesis RplB Ribosomal protein L2	COG00090	pfam00181, pfam0394	1	TIGR01171	4	1		1	•	н e		1 10	
arCOG04070	J Tran	anslation, ribosomal structure and biogenesis RpIC Ribosomal protein L3	COG0087	pfam00297		TIG R03626		1				. 4	1	2	2
arCOG04071	J Tran	anslation, ribosomal structure and biogenesis RpID Ribosomal protein L4	COG00088	pfam00573	_	TIGR03672	1	1 1				1	1 1	2	~
arCOG04092	J Trar	anslation, ribosomal structure and biogenesis RplE Ribosomal protein L5	COG00094	pfam00281, pfam0067	m		1	2 1		1		1	1	-1	-
arCOG04090	J Trar	anslation, ribosomal structure and biogenesis RpIF Ribosomal protein L6P	COG00097	pfam00347, pfam0034	7	TIGR03653		2		1			1,		-
arc0604288	1 Iran	ansiation, ribosomai structure and biogenesis kpiu kibosomai protein L10 and ation, ribosomai atructure and bionanaeis bulk bibosomai protain 111	COG00244	pram00466	CdU5/95	TIG D01637			-	-		-1		7 6	~
ar COG 043 / 2	Tran	ansiation, ribosoriiai su uuture ariu urugeriesis ji hyn. ji hubosoriiai proteini Li. antiation rihosomal structure and hiozenesis [RolM] Rihosomal rroteini 113	0000000	pfam00572	rd00392	TIGR01077		T						7	,
arCOG04095	J Tran	anslation, ribosomal structure and biogenesis RpIN Ribosomal protein L14	COG0003	pfam00238	-	TIGR03673		2 1			•	н с і	. "	2	2
arCOG00779	J Tran	anslation, ribosomal structure and biogenesis RpIO Ribosomal protein L15	COG00200	pfam00828			1	2 1		1	1	1	1 1	1	_
arCOG04113	J Trar	anslation, ribosomal structure and biogenesis RpIP Ribosomal protein L10AE/L16	COG00197	pfam00252	cd01433 1	TIGR00279	1	1 1	1		н,		1	1	1
arcOG04088	J Trar	anslation, ribosomal structure and biogenesis RplR Ribosomal protein L18 andation ribosomal structure and biogenesis DolV Dibosomal protein L23	C0G00256	ptam00861, ptam1420	cd00432	16001030		2 1			-1	ei e			_
ar COG 04072	J Tran	ansiation, ribosomial structure and biogenesis how Ribosomal protein 123 anslation, ribosomal structure and biogenesis RolW Ribosomal protein 123	COG0089	pfam00276	L	TIGR03636								2	
arCOG04094	J Tran	anslation, ribosomal structure and biogenesis RpIX Ribosomal protein L24	COG00198		cd06089 1	TIGR01080		2 1						2	
arCOG00785	J Tran	anslation, ribosomal structure and biogenesis RpmC Ribosomal protein L29	COG00255	pfam00831	cd00427 1	TIG R00012	1	2 1		1		1	1		
arCOG04086	J Trar	anslation, ribosomal structure and biogenesis RpmD Ribosomal protein L30	COG01841		cd01657 1	TIG R01309	гı .	2 1		1		Ţ,	1		_
arcOG04287	J Trar	anslation, ribosomal structure and biogenesis RPP1A Ribosomal protein L12E/L44/L45/RPP1/RPP2 andation: ribosomal structure and biogenesis DD2 DN2ee D culuinit DDD2	C0G02058	ptam00428	cd05832	TIGR03685			-			ei e		7	~ -
arCOG01885	J Tran	anslation, ribosomial structure and biogenesis RPS17A Ribosomal protein S17E	COG01383	pfam00833		I	1	1		1	2	н с і	1		
arCOG01344	J Tran	anslation, ribosomal structure and biogenesis RPS19A Ribosomal protein S19E (S16A)	COG02238	pfam01090			1	2 1	1	1 1	1	1	1 1	1	1
arCOG04186	J Tran	anslation, ribosomal structure and biogenesis RPS1A Ribosomal protein S3AE	COG01890	pfam01015			1				-	1	1 1	£	~
arCOG04182	J Trar	anslation, ribosomal structure and biogenesis RPS24A Ribosomal protein S24E	COG02004	pfam01282					,	-				-	
arCOG04183	Tran	ansiation, ribosofiiai structure and piogenesis - KF327A - Kibosofiiai protein 327E anstation rihosomal structure and hineenesis - R PC37AF - Rihosomal nontain S37AF		/ GOT DUI PID			-1	-	-	-				-	_
arCOG04314	J Tran	anslation, ribosomial structure and biogenesis RP528A Ribosomal protein 528/533	COG02053	pfam01200	cd04457	I		1		1	-	Ę	1	1	
arCOG04093	J Tran	anslation, ribosomal structure and biogenesis RPS4A Ribosomal protein S4E	COG01471	pfam08071, pfam0147	cd00165,cd0608	7	1	2 1		1	1	1	1 1	2	2
arCOG01946	J Tran	anslation, ribosomal structure and biogenesis RPS6A Ribosomal protein S6E/S10	C0G02125	pfam01092			1	1	1	1	1	1	1 1	3	
arCOG04154	J Trar	anslation, ribosomal structure and biogenesis RPS8A Ribosomal protein S8E	COG02007	pfam01201	cd11382	TIGR00307	-1			-		1	1,	2 7	~
arc0604097	L Tran	ansiation, ribosonial structure and plogenesis Rpsb Ribosonial protein 52 anslation ribosomal structure and biogenesis RpsC Ribosomal protein 53	C060092	pram00550.nfam0018	cd02411 1	TIGR01008		- 2	-	-	-			7 0	
arCOG04239	J Tran	anslation, ribosomal structure and biogenesis RpsD Ribosomal protein S4 or related protein	C0G00522	pfam00163, pfam0147	cd00165 1	TIGR01018		4	1	•		4 c a		1	
arCOG04087	J Tran	anslation, ribosomal structure and biogenesis RpsE Ribosomal protein S5	COG00098	pfa m00333, pfam0371	6	TIGR01020	1	2 1		2	1	1	1 1	1	_
arCOG04254	J Trar	anslation, ribosomal structure and biogenesis RpsG Ribosomal protein 57	COG00049	pfam00177	cd14867 1	TIGR01028	ен ,		1			,	1,1		
arcOG04091	J Iran	ansiation, ribosomal structure and biogenesis KpSH Kibosomai protein S8 anslation ribosomal structure and biogenesis Roci Ribosomal protein S9	COG00103	pram00410 nfam00380		TIGR03627		7 T	-						-
arCOG01758	J Tran	anslation, ribosomal structure and biogenesis RosJ Ribosomal protein S10	COG00051	pfam00338		TIGR01046	4	1				1		2	2
arCOG04240	J Tran	anslation, ribosomal structure and biogenesis Rpsk Ribosomal protein S11	COG00100	pfam00411		TIGR03628						Ŧ			
arCOG04255	J Tran	anslation, ribosomal structure and biogenesis RpsL Ribosomal protein 512	COG00048	pfa m00164	cd03367 1	TIGR00982	1	1 1	1	1	1		1 1	1	1
arCOG01722	J Trar	anslation, ribosomal structure and biogenesis RpsM Ribosomal protein S13	COG00099	pfam00416		TIGR03629	، ۱	-	1	-		1	1		_
arCOGUU/82	I Tran	anslation, ribosomal structure and biogenesis _ktpsivkubosomal protein 5.14 ====================================	COG00184	ptamuuz53 nfam08069.nfam0031	-400353	T				1 C	-	-	1		
arcountes	Tran	ansiation, ribosomai structure and progenesis אלאסט אוואסט איז	1005000	pTarriuguos, pranuuus nfamonagg	CGUUSDS	102630	-1	-		- r	-	-1		4 0	~
arcueu40%o	J 11a1	anslation, ribosomai structure and piogenesis kpsQ kiposomai protein >17	ΓΩΩΛΤΟΩ	pramuusoo	-	LIGKU3b3U		T 7		4		1	1	4	~

arCOG04099	-	Translation, ribosomal structure and biogenesis RpsS	Ribosomal protein S19 [COG00	L85 pfam00203		TIGR01025	1	1		1		1	1	1	2
arCOG00678	-	Translation, ribosomal structure and biogenesis Rrp4	Exosome complex RNA-binding protein Rrp4, contains S1 and KH domains COG01	197	cd05789		1 1	1			1		1	1	1
arCOG01574	-	Translation, ribosomal structure and biogenesis Rrp42	Exosome complex RNA-binding protein Rrp42, RNase PH superfamily COG02	L23 pfam01138, pfan	n0372 cd11365	TIG R01966	1 1	1	1		1		1	1	2
arCOG04131		Translation, ribosomal structure and biogenesis RsmA	165 rRNA A1518 and A1519 N6-dimethyltransferase RsmA/KsgA/DIM1 [COG00 are rank corr	330 pfam00398		TIGR00755	1 1		Ŧ		-1 -		1	., .,	ب ا ر
arcog00975	-	Translation, ribosomal structure and biogenesis RSmB Translation. ribosomal structure and biogenesis RSmB	165 rRNA C967 or C1407 C5-methylase, KsmB/RsmF family COG00 165 rRNA C967 or C1407 C5-methylase. RsmB/RsmF family COG00	L44 pram01189 L44 pram01189	cd02440	TIGR00446 TIGR00446	-		-		-			-	η
arCOG01239	, _,	Translation, ribosomal structure and biogenesis RsmE	RNA base methyltransferase family enzyme [COG01	901 pfam04013			1 2		1		1	1		1	1
arCOG04246	-	Translation, ribosomal structure and biogenesis RtcB	RNA 3'-P ligase, RtcB family protein	590 pfam01139		TIGR03073					1		1	1	1
arCOG04187		Translation, ribosomal structure and biogenesis Sdo1	Ribosome maturation protein Sdo1 [COG01 Colonomention consists termination clonention forter of Colon	500 pfam01172, pfan	n09377	TIGR00291	1,	-1		+			1		1 0
arcOG01701	-	Translation, ribosomal structure and biogenesis SEIS Translation ribosomal structure and biogenesis SEN2	Selenocysteine-specific translation elongation ractor of Selb-II domain COG03 tRNA solicing endonuclease COG01	276 nfam02778 nfan	cause96 n01974	TIGR00324						-			n ư
arCOG00403	-	Translation, ribosomal structure and biogenesis SerS	Seryi-tRNA synthetase COG00	L72 pfam02403, pfan	n0058 cd00770	TIGR00414			1	1				1	
arCOG01923	-	Translation, ribosomal structure and biogenesis SIK1	RNA processing factor Prp31, contains Nop domain COG01	198 pfam01798			1 1	1	1	1		1	1	1	2
arCOG01952	-	Translation, ribosomal structure and biogenesis SUA5	tRNA A37 threonylcarbamoyladenosine synthetase subunit TsaC/SUA5/YrdC COG00	009 pfam01300		TIGR00057	1 2	,	1		2	,	.,		
arCOG04223		Translation, ribosomal structure and biogenesis SUI1 Translation ribosomal structure and biogenesis SUI1	Translation initiation factor 1 (elF-1/SUI1) [COG00 Translation initiation factor 2 alpha cubinit (elE-2alpha)	023 pfam01253	cd11567	TIGR01158	1 5		-		-	-			2 6
arcog00056	-	Translation, ribosomial structure and biogenesis 5012 Translation, ribosomal structure and biogenesis Tan1	translation miniation racion 2, alpria suburin (en-zarpria) [COG01 tRNA(Ser.Leu) C12 N-acetvlase TAN1, contains THUMP domain [COG01	318 pfam02926, pfan	n1441 cd11718.cd06	1 TIGR00342	-		-	-			-	-	7
arCOG01630	-	Translation, ribosomal structure and biogenesis TdcF	Translation initiation inhibitor, yigF family COG00	251 pfam01042	cd00448	TIG R00004	2 2	2		1	m	2		2	4
arCOG01561	-	Translation, ribosomal structure and biogenesis TEF1	Translation elongation factor EF-1 alpha, GTPase COG05	256 pfam00009, pfan	n0314 cd01883,cd03	6 TIGR00483	1 1	1	1	1	1		1	1	S
arCOG00989	_	Translation, ribosomal structure and biogenesis Tgt	Queuine/archaeosine tRNA-ribosyltransferase COG00	343 pfam01702		TIGR 00432							1		,
arc0600038		Translation, ribosomal structure and biogenesis Thil	tRNA S(4)U 4-thiouridine synthase	301 ptam02926, ptar	n0256 cd11 / 16,cd01	7 TIGR00342, TIGR	14271			-		-1	7	7 7	7 5
arcog01115		Translation, ribosomal structure and blogenesis Trills	TILI EURIY-TANAA Syntrictase TBNAI(II=2) 7-aematinvlortidine svnthetase: containing Zo-ribhon domain and [COG01	571 Diam08489. Diam	n0728 cd04487	TIGR03280			-	-		-			- t
arCOG04176	, _,	Translation, ribosomal structure and biogenesis TIF6	Translation initiation factor 6 (eIF-6)	976 pfam01912	cd00527	TIGR00323	1 2	1	1	1		4			
arCOG00042	-	Translation, ribosomal structure and biogenesis TilS	tRNA(IIe)-lysidine synthase TilS/MesJ	037 pfam01171	cd01993	TIGR02432	1 1		1				2	1	3
arCOG00985	-	Translation, ribosomal structure and biogenesis Tma 20	Predicted RNA-binding protein, contains PUA domain COG02	016 pfam09183, pfan	n01472	TIGR03684	1 1			1	1	1	1	1	2
arCOG01951		Translation, ribosomal structure and biogenesis TmcA	tRNA(Met) C34 N-acetyltransferase TmcA COG01	444 pfam08351, pfan	n05127, pfam13718	00000	,	•	,	•	d		г,	,	,
arc0601219		Translation, ribosomal structure and biogenesis I KMI	NZ,N 2-GIMETNYIGUADOSINE TKINA METNYITRANSFERASE	56/ pramuzuu5		TIGK00308	7 7	-		-	7	Ę			
arcOG00033	-	Translation, ribosomal structure and biogenesis Irmit I Translation ribosomal structure and biogenesis Trm5	tkiva GLU IN-metriyidse TrmL1 Wubirtosine (vM) hiosunthesis enzume Trm5 methyltransferase COG02	041 pram011/0			- -	-	1 0	-		-1	1 0		7 0
arc0601018	-	Translation, ribosomer structure and biogenesis Trm1	tRNA C32.1132 (rihose-2'-O)-methylase TrmI or a related methyltransferase COG00	565 nfam00588		TIGROOOSO	1 2		1 -	- ~	-		4 -	- ~	2 2
arCOG01887		Translation. ribosomal structure and biogenesis TrpS	Tryptophanyl-tRNA swithetase	L80 pfam00579	cd00806	TIGR 00233	1					4		1	4 m
arCOG04449	-	Translation, ribosomal structure and biogenesis TruA	Pseudouridylate synthase COG00	101 pfam01416, pfan	n0141 cd02866	TIGR00071	1		1					1	2
arCOG00987	-	Translation, ribosomal structure and biogenesis TruB	Pseudouridine synthase COG00	L30 pfam08068, pfan	n0150 cd02572	TIGR00425					1		1	1	1
arCOG04252	-	Translation, ribosomal structure and biogenesis TruD	tRNA(Glu) U13 pseudouridine synthase TruD COG00	585 pfam01142	cd02577	TIGR00094	1 2	1	1	1	2		1	1	1
arCOG00761	-	Translation, ribosomal structure and biogenesis TsaA	tRNA (Thr-GGU) A37 N-methylase COG01	720 pfam01980	cd09281	TIG R00104							1		
arCOG04733	_	Translation, ribosomal structure and biogenesis Tsr3	Ribosome biogenesis protein Tsr3, contains Fer4-like metal-binding domain ol COG02	042 pfam04034	-				1		ц,		ر ا ،		2
arc0G01886		Translation, ribosomal structure and biogenesis Tyrs	Tyrosyi-tRNA synthetase	162 ptam005/9	cd00805	TIGR00234	, ,		1	-			, ۲		2 0
arc061012/	-	Translation, ribosomal structure and piogenesis I TVW1 Translation ribosomal structure and biogenesis TVM2	Wybutosine (yw) piosyntnesis enzyme, re-s oxigoreguctase COGO Wybutosina (yW) hisewathaeis annyma TYM2 transfarasa	To prarm04050, prar	10800 001233	TIG PO1 A AA	7	-1	-		-	-	-		'n
arCOG04156		Translation, ribosomal structure and biogenesis TYW3	Wybutosine (VW) biosynthesis enzyme	590 pfam02676			1						1	4	4
arCOG00808	-	Translation, ribosomal structure and biogenesis ValS	Valyl-tRNA synthetase COG00	525 pfam00133, pfan	n0826 cd00817,cd07	9 TIGR00422	1							2	9
arCOG04225	-	Translation, ribosomal structure and biogenesis YmdB	O-acetyl-ADP-ribose deacetylase (regulator of RNase III), contains Macro dom COG02	L10 pfam01661	cd02907								1	1	1
arCOG00541	-	Translation, ribosomal structure and biogenesis YSH1	Predicted exonuclease of the beta-lactamase fold involved in RNA processing COG01	236 pfam00753, pfan	n10996,pfam07521	TIGR03675	1		1		1	1	1	1	9
arCOG00545	-	Translation, ribosomal structure and biogenesis YSH1	Predicted exonuclease of the beta-lactamase fold involved in RNA processing [COG01	236		TIGR 04 1 2 2	1 1	-1		2					
														l	
CELLULAR PROC	CESSES A	AND SIGNALING			-		•		_			l		-	
arc0611012		Cell cycle control, cell division, chromosome partiti-	Predicted cell division protein, sept-nomolog	150 pram044/2	21000001E oform0041E	oform00.11E of or 00	I I I		-		1			c	ų
arcog04701		Cell cycle control, cell division, chromosome partin CrcB/FX	hipia-tuouni suppressor and related Neet donamicontaining proteins Integral membrane protein nossibly involved in filioride export and chromoso COG00	039 nfam02537	4/074/0711/01/0/1/074/0711	TIGR00494	T IDIN/CT M	4						- -	- m
arCOG02201	0	Cell cycle control, cell division, chromosome partit FtsZ	Cell division GTPase COG00	206 pfam00091, pfan	n1232 cd02201	TIGR 00065	2 3	-1	2	1	1			2	
arCOG05007	٥	Cell cycle control, cell division, chromosome partit Maf	Nucleotide-binding protein implicated in inhibition of septum formation COG00	124 pfam02545	cd00555	TIGR 00172							2	1	3
arCOG03061	<u>م</u>	Cell cycle control, cell division, chromosome partit MreB	Actin-like ATPase involved in cell morphogenesis	077 pfam06723	cd10227	TIGR00904								,	
arc0600585		Cell cycle control, cell division, chromosome partit Mirp	Wrp tamity protein, ALPase, contains Iron-suirur cluster	10.6 pram10609	cd02776 cd02	11GKU1969					-			-1 0	'n
arCOG00586	<u>م</u>	Cell cycle control, cell division, chromosome partit Soi	ATPase involved in chromosome partitioning. ParA family COG01	192 pfam01656	cd02042	TIGR03453					-			n	1
arCOG02416	z	Cell motility -	Pilin/Flagellin, FlaG/FlaF family COG03	130 pfam07790			1				1			3	1
arCOG00434	z	Cell motility -	AAA+ ATPase of MoxR-like family, a component of a putative secretion systen COG00	714 pfam07726	cd0009	TIGR02031	2 1	1	2		2	2	2	2	e
arc0602079	z	Cell motility	S-layer protein, possibly associated with type IV pill like system COG01	361					-						
arc0602911	zz	Cell motility	Predicted allia /flagellia SC 003	13 13			-	-	-	7			- ⁻		
arCOG02382	z	Cell motility CheB	Chemotaxis response regulator containing a CheY-like receiver domain and a ICOG02	201 pfam00072, pfan	n0133 cd00156	TIGR02875								1	
arCOG01819	z	Cell motility CpaF	Pilus assembly protein, ATPase of CpaF family COG04	962 pfam00437	cd01130	TIGR03819	1 2	1	1 1	1		-1		1	1
arCOG01829	z	Cell motility	Archaeal flagellins COG01	581 6 0000	0.00								с с		1
arc06018364	zz	Cell motility FiaU	Archaeilum protein D/E Archaeilum arotein E flarailin of Elac /Elac /Elac family	351 pramU5377, pran	mu4659								7	╈	
arCOG01822	zz	Cell motility Flag	Archaellum protein G. flagellin of Flag/Flaf family	354										1	
arCOG04148	z	Cell motility FlaH	ATPase involved in biogenesis of archaellum	374 pfam06745	cd01394	TIGR03881							1	1	
arCOG01809	z	Cell motility	Archaellum assembly protein J, TadC family [COG01	955									1	1	
arCOG02298	z	Cell motility	Peptidase A24A, prepilin type IV	989 pfam01478, pfan	n06847		1	(, 1		2				
arCOG01808	zz	Cell motility TadC	Pilus assembly protein TadC COG02 Dilus assembly protein TadC	164 ptam00482, ptan	n00482		1 2	7 +	2 2	7 +	2	-	2	2	4
arCOG01812	zz	Cell motility TadC	Pilus assembly protein rade Pilus assembly protein Tade	064			1 r							1	1
arCOG01817	z	Cell motility VirB11	ATPase involved in archaellum/pili biosynthesis COG00	530 pfam00437	cd01130	TIGR03819	1 1	1	1 1	1	1	-1	3	2	2
arCOG02488	2	Cell wall/membrane/envelope biogenesis	Cell surface protein, a component of a putative secretion system SC.002	33			1	t1 -	1			t -	1	,	
arCOG01385	22	Cell wall/membrane/envelope biogenesis	Glycosyl transferase family 2 [COG00 Education contribution which and EN2 domains	163 pfam00535	cd02522	TIGR04283	1	-1	1	+			2 0	2	m
arcog00894	Σ	Cell wall/metituraric/criverope urogenesis	Extracellular protein containing vercin and rivo domains Glycosyl transferase family 2 COG00	463 pfam00535	cd04179	TIGR04182	2 2	2	2 2		7		1 0	2	.0
arCOG02532	Σ	Cell wall/membrane/envelope biogenesis -	Secreted Beta-propeller repeat protein fused to CARDB-like adhesion module COG03	291 pfam07705		TIGR04213	1 11		2 1		2	2	1	2	,
arCOG07560	Σ	Cell wall/membrane/envelope biogenesis	Cellsurface protein					1	_			-1	'	-	1
	1														

arCOG03336	Σ	Cell wall/membrane/envelone biogenesis	Surface protein containing fa	asciclin-like repeats	0602335	nfa m02469			•						~	t.
arCOG09173	Þ	Cell wall/membrane/envelope biogenesis	Cell surface protein			ofam07790.nfam0770.cd(00146.cd001	IGRODR64.TIGR	s						,	
arCOG02086	Σ	Cell wall/membrane/envelope biogenesis	- S-laver domain		COG01361				2							
arCOG03335	Σ	Cell wall/membrane/envelope biogenesis	 Surface protein containing fe 	asciclin-like repeats	:0G02335	pfam02469										1
arCOG08795	Σ	Cell wall/membrane/envelope biogenesis	 Predicted S-layer protein wit 	th Ig-like domain	C.00192	pfam00127 cd:	13921 T	IGR02657	1							
arCOG02482	2	Cell wall/membrane/envelope biogenesis	- WD40/PQQ-like beta propel	ler repeat containing protein	0G01520	pfam13570, pfam1357 cd:	10276 1	IGR03300		1						,
arCOG01383	≥ :	Cell wall/membrane/envelope biogenesis	- Glycosyl transferase family 2		0G01216	pfam00535 cd(04186 T	IGR04017,TIGR01	556			1				2
arC0G00568	2 2	Cell wall/membrane/envelope blogenesis	- Oligosaccharyl transferase S	113 or related protein	0601287											
arc0601391	2 2	Cell wall/membrane/envelope blogenesis	 Glycosyl transferase tamily . 		0601215	pram13641 cd(1064.2.1 I	1GKU3U3U								
arcocoseso	2 2	Cell wall/membrane/envelope blogenesis	- S-layer domain Cell surface protein		0601361				ł			ł				
arCOG02553	2 2	Cell wall/membrane/envelope biogenesis	Cell surface protein		0601572	nfam11824 nfam07705										
arCOG 12808	Σ	Cell wall/membrane/envelope biogenesis	Glycosyltransferase. GT1 far	milv	OG00438		03822		ł			T				
arCOG01389	Σ	Cell wall/membrane/envelope biogenesis	 Glycosyl transferase family 2 		:0G01215	pfam13641 cd(106423 T	1GR03937							1	1
arCOG00895	Σ	Cell wall/membrane/envelope biogenesis	 Glycosyl transferase family 2 	0	:0G00463	pfam00535 cd(06442 T	IGR04182							1	
arCOG02487	Σ	Cell wall/membrane/envelope biogenesis	 Cell surface protein, a compt 	onent of a putative secretion system		pfam02369				_					1	
arCOG02538	Σ	Cell wall/membrane/envelope biogenesis	- Secreted protein, with PKD r	epeat domain	OG03291		E	IGR04213							ب ا	
arCOG00896	2	Cell wall/membrane/envelope biogenesis	- Glycosyl transferase family 2		OG00463	pfam00535 cd(04179 T	IGR04182	1						2	1
arCOG01410	2 2	Cell wall/membrane/envelope biogenesis	AgiA Glycosyltransferase		0G00438	pfam13579, pfam0053 cd0	03794 1	IGR02149	ы ч	m +	., .,	1	4,			Ŧ
arcOGO0800	2 2	Cell wall/membrane/envelope blogenesis	Agib Uligosaccharyitransrerase m Arith2 Dreadicted filmnese	iembrane subunit	0600207	pramuz516, pram13620, pr	Tam13620,pt	IGK04154			-	-	-		7	-
arCOGOD664	2 2	Cell wall/membrane/envelope blogenesis	ABIDZ FIEURCEU IIIJDASE ValE/RfhA Glucose-1-phosphate uridvit	ransfera se	0601209	prarinos/ 00 nfam00483	181 1	IGR03947	-	-						
arCOG03199	Σ	Cell wall/membrane/envelope biogenesis	AgH/Rfe UDP-N-acetylmuramyl penta	apeptide phosphotransferase/UDP-N-acetv/glucosC	0G00472	pfam00953 cd(06856 1	IGR00445	1						•	
arCOG01403	Σ	Cell wall/membrane/envelope biogenesis	AgiL Giycosyltransferase		:0G00438	pfam13439, pfam0053 cd0	03804		3 2	1 2	1	1	m	9	5	6
arCOG00253	Σ	Cell wall/membrane/envelope biogenesis	AgIM/Ugd UDP-glucose 6-dehydrogena	ISE C	COG01004	pfam03721, pfam00984, pt	ofam03720 T	1GR03026	1 1	_		1 1	2		1	
arCOG01381	Σ	Cell wall/membrane/envelope biogenesis	AgIO Glycosyl transferase family 2		OG00463	pfam00535 cd(00761 1	IGR04283							1	1
arCOG02209	2 2	Cell wall/membrane/envelope biogenesis	AgiR MATE family membrane pro	tein, Rfbx family	0G02244	-fe-mootro	13128		3 2	-	1	1	m	2		
arc06.01 3.75	2 2	Cell wall/membrane/envelope blogenesis	BBIC Aryi-phospho-beta-D-glucos	Idase Bglc, GH1 Tamily des ILDD-Gron Ac-invertion / 6-debudratese ElaA 1 -10	0601086	ptam00150	T 1000	10 003 5 90	ł					-		
arCOGOD665	Σ	Cell wall/membrane/envelope blogenesis	Salu UDP-elucose pyrophosihory	des OF - OUIVAC-IIIVEI UIIS -;/O-DEIIYUI BUBSE I IBALI IC	0001210	pfam00483 cd	102541 T	1GR01099	-						÷-	ţ
arCOG00666	Σ	Cell wall/membrane/envelope biogenesis	GCD1 N-acetylglucosamine-1-phos	sphate uridyltransferase	:0G01208	pfam00483, pfam0013 cd0	04181,cd056 T	IGR03992	m • m	m m	2			m	m	
arCOG00663	Σ	Cell wall/membrane/envelope biogenesis	3CD1 Nucleoside-diphosphate-sug	car pyrophosphorylase involved in lipopolysacchar	:0G01208	pfam00483 cd(04181 T	IGR03992				1				
arCOG00668	Σ	Cell wall/membrane/envelope biogenesis	3CD1 Nucleoside-diphosphate-sug	gar pyrophosphorylase involved in lipopolysacchar	:0G01208	pfam00483, pfam0013 cd(104181,cd056 T	IGR03992	1 1	_						
arCOG00057	Σ	Cell wall/membrane/envelope biogenesis	GImS Glucosamine 6-phosphate sy	ynthetase C	0G00449	pfam00310, pfam0138 cd(00714,cd0501	IGR01135	1 1	1		1		1	1	1
arCOG01373	Σ :	Cell wall/membrane/envelope biogenesis	Gmd GDP-D-mannose dehydratas		0G01089	pfam01370, pfam1395 cd0	05260 1	IGR01472						ļ		
arcuguubs	ΣZ	Cell wall/membrane/envelope blogenesis	GUTQ Predicted sugar phosphate I	somerase involved in capsule formation	0600/94	ptam01380 [cdi	1	IGKU312/						-		
arCOG06815	2 2	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	MnaA Murein trinentide amidase N	MaaA	0602866	ptam00246 ptatto2067	06228	7177000								.
arCOG01568	Σ	Cell wall/membrane/envelope biogenesis	MsrS Small-conductance mechanic	Sensitive channel	0600668	pfam00924	04400	I	ر د	-	-	-	-	-	- ~	
arCOG02821	Σ	Cell wall/membrane/envelope biogenesis	MurD UDP-N-acetvimuramovialani	ine-D-glutamate ligase	0600771	pfam08245, pfam02875	F	IGR01082	4			1	•			2
arCOG02820	Σ	Cell wall/membrane/envelope biogenesis	MurE UDP-N-acetylmuramyl triper	otide synthase	:0G00769	pfam08245, pfam02875	T	IGR01082	1 2						1	
arCOG07536	Σ	Cell wall/membrane/envelope biogenesis	NanM N-acetylneuraminic acid mu	tarotase	COG03055										1	
arCOG02005	Σ	Cell wall/membrane/envelope biogenesis	RacX Aspartate racemase	0	:0G01794	pfam01177	F	IGR00035						1		
arCOG01411	Σ :	Cell wall/membrane/envelope biogenesis	RfaG Glycosyltransferase		0G00438	pfam13439, pfam0053 cd0	103801 T	1GR03999	2		1			-		m
arcOG01/17	2 2	Cell wall/membrane/envelope blogenesis	Krad Glycosytransferase		0600438	nfa m0063.4	01625		7			ł		Ţ		
arCOG0141/	Σ	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	Arido Giycosyiu anisterase 3faG Giycosyitransferase		0600438	prami 34 39. nfam1 369 cdf	1 103794	IGR04063						-		
arCOG01409	Σ	Cell wall/membrane/envelope biogenesis	Red Glycosyltransferase		OG00438	pfam13692 cd(03794 T	IGR03087								
arCOG01371	Σ	Cell wall/membrane/envelope biogenesis	3fbB dTDP-D-glucose 4,6-dehydra	atase	:0G01088	pfam01370 cd(05246 7	IGR01181								
arCOG04188	Σ	Cell wall/membrane/envelope biogenesis	RfbC dTDP-4-dehydrorhamnose 3.	5-epimerase or related enzyme	:0G01898	pfam00908	F	IGR01221						1		
arCOG01367	Σ	Cell wall/membrane/envelope biogenesis	RfbD dTDP-4-dehydrorhamnose ru	eductase	0G01091	pfam04321 cd(05254 T	IGR01214		_				1		
arCOG01168	Σ	Cell wall/membrane/envelope biogenesis	RspA L-alanine-DL-glutamate epin	nerase or related enzyme of enolase superfamily C	:0G04948	pfam02746, pfam0118 cd(03316 T	IGR02534						1		
arCOG04827	2	Cell wall/membrane/envelope biogenesis	TagB Glycosyl/glycerophosphate t	transferase involved in teichoic acid biosynthesis TC	0G01887	pfam04464								с і -		,
arc0601222	2 2	Cell wall/membrane/envelope blogenesis	TagD Cytidylyltransferase fused to	conserved domain of DUF357 tamily	0600615	ptam01467cd(02170	IGR02199	7	1					-	m r
arc0601360	2 2	Cell wall/membrane/envelope plogenesis	weag Inucieoside-diphosphate-sug Meag Nucleoside-diphosphate.cug	gar epimerase	0600451	prarmuts/u/pramits99 cut	1 2/200	10KU1179	7 V	-		- T			y	n r
arCOG04826	Σ	Cell wall/membrane/envelope blogenesis	Mcak Polycarcharide pyriate 348	serencease sferase family protein	0602327	pfam04230	- L	IGR03609	7 7	4		4		- r	>	,
arCOG01392	Σ	Cell wall/membrane/envelope biogenesis	WecB UDP-N-acetylglucosamine 2-	epimerase	OG00381	pfam02350 cd0	03786 1	IGR00236				1				1
arCOG00252	Σ	Cell wall/membrane/envelope biogenesis	WecC UDP-N-acetyl-D-mannosami.	nuronate dehydrogenase	:0G00677	pfam03721, pfam0098 cd(05266 1	IGR03026	1 1	1 1				1	1	1
arCOG00118	Σ	Cell wall/membrane/envelope biogenesis	WecE Predicted pyridoxal phospha	ate-dependent enzyme apparently involved in reg C	COG00399	pfam01041 cd(00616 T	IGR03588	3 1	1		2			2	1
arCOG03634	Σ;	Cell wall/membrane/envelope biogenesis	Wech Surface polysaccharide O-ac	Vitransferase, integral membrane enzyme	0603274	ptam01757				, 		τ				
arcOG08906	> >	Defense mechanisms	Predicted antitoving contain	ed protein	0602886	plam03683	l		-	7						
arCOG01467	>	Defense mechanisms	- ABC-type multidrug transpor	rt svstem, permease component	0000842	pfam01061		IGR01247	2 1	-		2	-			
arCOG03966	>	Defense mechanisms	- CopG family DNA-binding pr-	otein									1			
arCOG01463	>	Defense mechanisms	- ABC-type multidrug transpo	rt system, permease component	:0G00842	pfam01061	F	IGR01247	1 1	1		1				
arCOG00722	> :	Defense mechanisms	Predicted antitoxins contain.	ing the HTH domain	0602886	pfam03683			1 2	1		1				Ŧ
arc0600712	> >	Defense mechanisms	DIN domain containing note	ntaining DivA-Dinging protein, orten an antitoxin ic	0601848	nfam01950	00861		-			I			-	-
arCOG01009	>	Defense mechanisms	CopG/MetL, RHH domain cort	ntaining DNA-binding protein, often an antitoxin iC	0003609	0.00010111816	TOOCO		ł							
arCOG01469	>	Defense mechanisms	- ABC-type multidrug transpoi	rt system, permease component	:0G00842				-					1		
arCOG02123	>	Defense mechanisms	- HEPN domain containing pro	otein C	:0G01895	pfam05168				_				1		
arCOG02780	> ;	Defense mechanisms	Endonuclease, HJR/Mrr/Rec.	B family	0,00000	20000 J 07070 J			+					ر ا ,		
arCOG05724	> >	Defense mechanisms	- RecB family restriction endo	nuclease, SegA-like protein	0602810	ptam04313, ptam03925								-1 +		
arc0606948	> >	Defense mechanisms	PIN domain		C.00610											
arCOG06971	>	Defense mechanisms	CopG/RHH family DNA bindi	ng protein	0G03609									1		
arCOG09828	>	Defense mechanisms	RelB family antitoxin	S	.C.00879									1		
arCOG06966	> :	Defense mechanisms	- CopG/RHH family DNA bindi	ng protein	0003609									2 .		Ľ
arCOG03166	> ;	Defense mechanisms	- AAA+ superfamily ATPase fu	IC I	0G01672	0.000.		Ī		+		-		4	·	5
arC0G03521	>	Defense mechanisms	 Type II restriction enzyme, n 	nethylase subunit	COGU1002	pfam12950	-	-	_							

arCOG00815	>	Defense mechanisms Abr Abr	rB family transcriptional regulator	:0G02002	pfam04014		TIGR01439					1		
arCOG01452	>	Defense mechanisms Cas1 CR	ISPR-associated protein Cas1 C	:0G01518	pfam01867 c	d09636	TIGR00287					2		
arCOG02666	>	Defense mechanisms Cas10 CR	ISPR associated protein Cas10, large subunit of Type III system effector con C	:0G01353	pfam01966							1		
arCOG04194	> :	Defense mechanisms CF.	ISPR-associated protein Cas2	0G01343	pfam09827 c	d09725	TIGR01573	, ,		•		н ,	Ŧ	
arcog00790	> >	Defense mechanisms Case Cr	ISPR-associated protein Cas4, RecB family exonuclease	OG01468	D DEETOIIIPIÓ	d09637	TIGR 00372	7	-	4	-		-	
arCOG00794	>	Defense mechanisms CR	ISPR-associated protein Cas4, RecB family nuclease	OG01468	pfam01930 c	d09637	TIGR00372							
arCOG04342	>	Defense mechanisms Case CR	ISPR-Cas system related protein Cas6, RAMP superfamily	:0G01583	pfam01881 c	d09652	TIGR01877					1		
arCOG 00196	>	Defense mechanisms CcmA AB	C-type multidrug transport system, ATPase component	:0G01131	pfam00005 c	d03230	TIGR03740	2 1	1	1	1 1		1	2
arCOG00194	>	Defense mechanisms CcmA AB	C-type multidrug transport system, ATPase component	:0G01131	pfam00005 c	d03230	TIG R01188	7 7	6 2	1	3	S	4	9
arCOG02665	>	Defense mechanisms CBR CB	ISPR-Cas system related protein, RAMP superfamily Cas5 group	0G01769	pfam09700							-1		
arCOG06487	> >	Defense mechanisms CSm255 CF	ISPR/Cas system CSM-associated protein Csm2, small subunit CDP Consistent controls and an antice of the subunity Control and an anti-	0601421	ptam03750 C	d09647	TIGR01870							
arcOG03222	> >	Defense mechanisms Londense mechanisms CR	ISPR-Cas system related protein, KAMP superfamily Cas/ group	0601567	pramus/8/ C	d09663	TIGR01903						T	
arCOG03718	- >	Defense mechanisms CR	ISPR/Cas system CSM-associated protein Csm5. group 7 of RAMP superfamIC	0601332		d09662	TIGR01899					. +		
arCOG01449	>	Defense mechanisms CA	RF domain containing protein	0G01517	>		TIGR01884							
arCOG07641	>	Defense mechanisms CA CA	RF domain containing protein, contains HTH and HEPN domains	:0G01517	pfam09455 c	d09732	TIGR02221					1		
arCOG03847	>	Defense mechanisms CA CA	RF domain containing protein	0G01517								2		
arCOG00373	> :	Defense mechanisms Dr	A sulfur modification protein DndD, ATPase	0G01196	pfam13514		TIGR02169				ti	-1		
arCOG02632	> ;	Defense mechanisms HsdM Ty	oe I restriction-modification system methyltransferase subunit	0G00286	pfam12161, pfam02384	00000	TIGR00497	,	,		,	, I		2
arCOG03779	> ;	Defense mechanisms McrB Gi	Pase subunit of restriction endonuclease	0601401	ptam07728 c	d00009		2 1			7			
arcocone 44	> >	Defense mechanisms Middle IVI	C two multidaria transmost suctam ATDare and normanic component	0604268	pramiuuu/	00000	TICEOUTO	-	-			r		
arc0602777	> >	Defense mechanisms Mrr Re	oc-type intuition og transport system, Arrase and permease component. Construction endomicilease	0601715	pfarm04471	CC7CON		-	4	-		7	v +	n
arCOG01008	- >	Defense mechanisms NikR Tre	anscriptional regulator. CopG/Arc/MetJ family (DNA-binding and a metal-bi C	0600864	pfam08753		TIGR02793					1		1
arCOG02836	>	Defense mechanisms NikR Tra	anscriptional regulator, CopG/Arc/MetJ family (DNA-binding and a metal-bi C	COG00864								1		
arCOG00516	>	Defense mechanisms Nit	roimidazol reductase NimA or a related FMN-containing flavoprotein, pyridC	:0G03467	pfam01243		TIGR04023	1	1					
arCOG00525	>	Defense mechanisms Nin	croimidazol reductase NimA or a related FMN-containing flavoprotein, pyridC	:0G03467	pfam12724, pfam01243			2	1	1				
arCOG01731	>	Defense mechanisms NorM Na	+-driven multidrug efflux pump	OG00534	pfam01554, pfam0155 c	d13137	TIGR00797	1 1	1 1	1		m		
arCOG01663	> :	Defense mechanisms RelE CV	totoxic translational repressor of toxin-antitoxin stability system	0G02026	pfam05016							e	+	
arCOG01665	> ;	Defense mechanisms RelE CY	totoxic translational repressor of toxin-antitoxin stability system	0G02026	pfam05016		710000110	, ,	,		•	m		,
arc0603212	> >	Defense mechanisms 5alX AE	C-type antimicrobial peptide transport system, A I Pase component	0600577	pramuouus c	GC2200	TIGRU26/3	3 1	γ 		7 7	7 C	4 0	, 2
arCOG02957	> =	Deteribe interinations and vasicular tri SRH1 Pre-	C-type antimicioual peptide transport system, permease component.	1/20000	pfam03011			 	-		-	، ۲	4	° °
arc0602673	=	Intracellular trafficking secretion, and vesicular tribuit.	a A/Sno1/YidC translocase suburity securates a A/Sno1/YidC translocase/secretase sec-independent itegration of nascentic	0601422	pfam01956					-	 -		-	1
arCOG04169	=	Intracellular trafficking secretion and vesicular trafficking secretion	annotein translocase subunit SecY	0600201	nfam10559. nfam00344		TIGR 00967	1 2			4 .			2
arCOG01217	=	Intracellular trafficking secretion and vesicular trafficking secretion	nal recognition particle 19 kDa protein	0601400			000000		-	•	4 .			2
arCOG04736	, _	Intracellular trafficking. secretion. and vesicular tri TatC Sec		0000805	pfam00902.pfam00902		TIGR01912.TIGR0	11912	4	-			4	1
arCOG01739		Intracellular trafficking. secretion. and vesicular tri LeoB Sig	nal peptidase 1	0600681	pfam00717	d06530	TIGR02228	1 1	1			Ţ	-	2
arCOG02204		Intracellular trafficking, secretion, and vesicular tri Sss1 Pre	eprotein translocase subunit Sss1	0G02443				1 2			1	. +1		2
arCOG01919	∍	Intracellular trafficking, secretion, and vesicular tri TatC Sec	c-independent protein secretion pathway component TatC	COG00805	pfam00902		TIGR00945	1 2	-1		1		2	ŝ
arCOG01228	0	Intracellular trafficking, secretion, and vesicular tri Ffh Sig	nal recognition particle GTPase C	OG00541	pfam02881, pfam0044 c	d03115	TIG R00959	1	1 1		1	1	1	1
arCOG02694	0	Intracellular trafficking, secretion, and vesicular tri TatA Se	c-independent protein secretion pathway component	:0G01826			TIGR01411	2	1	1	1			1
arCOG04471	þ	Intracellular trafficking, secretion, and vesicular tri-	Sortase	COG04083	pfam09721		TIGR04125				1	1	1	1
arCOG01227	D	Intracellular trafficking, secretion, and vesicular tr FtsY Sig	nal recognition particle GTPase	COG00552	pfam02881, pfam0044 c	d03115	TIG R00064	1				1	1	5
arCOG01997	∍	Intracellular trafficking, secretion, and vesicular tri MarC M.	ultiple antibiotic transporter C	0G02095	pfam01914		TIGR00427					1		
arCOG03382	<u> </u>	Intracellular trafficking, secretion, and vesicular tr TolB	riplasmic component of the Tol biopolymer transport system	0G00823	pfam00930		TIGR02800					1		
arCOG04816	5	Intracellular trafficking, secretion, and vesicular tri-	3G/TraD/VirD4 tamily enzyme, ATPase	0603505	ptam12846									
arCOG02283		Intracellular trafficking, secretion, and vesicular trivinb4 11	de IV secretory patriway, virib4 component Vialsemic component of the Tol higherolymer transport system	1050000	ancanmeta nconnaeta		TIG DO3 BOD					1	f	
arcOG01241	×	Mobilome: pronhazes, transnosons	rD/XerC family integrase	0600582	pfam13495 pfam0058 c	d00796	TIGR02224				-	4	4 -	2
arCOG03473	×	Mobilome: prophages, transposons - Tri	ansposase C	OG05421				2						
arCOG03989	×	Mobilome: prophages, transposons Xer Xe.	rD/XerC family integrase	0G00582	pfam00589 c	d00397						1		
arCOG 10214	×	Mobilome: prophages, transposons - Re	laxase/mobilization nuclease domain-containing protein	1								1		
ar COG 10424	×	Mobilome: prophages, transposons GepA Un	characterized phage-associated protein	:0G03600								1		
arCOG02751	×	Mobilome: prophages, transposons - Tr _i	ansposase, IS5 family CC	COG03039	pfam01609							10	_	
arCOG01915	0	Posttranslational modification, protein turnover, dHflC M.	embrane protease subunit, stomatin/prohibitin homolog	COG00330	pfam01145 c	d08826	TIGR01933	1	1		1	1		1
arCOG01912	00	Posttranslational modification, protein turmover, d-	embrane protein implicated in regulation of membrane protease activity C	0G01585	pfam01957	00000	TIC D04 2 42	, , ,	,			н ,		
arcOG06181	0	Posttranslational modification protein turnover, 4KPL1 41. Desttranslational modification protein turnover 4TrvA Th	P-dependent 205 proteasonne regulatory subunit	0600526	pram00578 c	402966	110KU1242	 	-			-	7	n
arCOG01341	0 0	Posttranslational modification protein turnover dGIM5 Pre-	edicted prefoldin molecular chaperone implicated in de novo protein foldir	0601730	nfam02996	d00584	TIGR00293				4 .	-		
arCOG00609	0	Posttranslational modification, protein turnover, dRseP Mt	embrane-associated protease RseP, regulator of RpoE activity in bacteria	0G00750	pfam02163 c	d06160		1 2	1	1	1			
arCOG02959	0	Posttranslational modification, protein turnover, d ap	 dependent a mino- or carboxypeptidase, M28 family 	:0G02234	pfam04389 c	d05643		1 2			1 1		3	
arCOG00314	0	Posttranslational modification, protein turnover, dTrxA Th	iol-disulfide isomerase or thioredoxin	0G00526	pfam00578 c	d02969		1	1 1	1 1	1 1		1	2
arCOG04463	0	Posttranslational modification, protein turnover, d-	esenilin-like membrane protease, A22 family	:0G03389	pfam06550			1	1	1	1	1	-	1
arCOG01257	0	Posttranslational modification, protein turnover, d GroEL Cr.	aperonin GroEL, HSP60 family	0600459	pfam00118 c	d03343	TIGR02339	·	1			2	m	6
arCOG01833	00	Posttranslational modification, protein turmover, dlbpA M.	blecular chaperone (HSP20 family)	0000071	pfam00011 c	d06464	TIC DO1 202							4
arc0604767	0	Posttranslational modification, protein turmover, 9 IrxB Irr	loredoxin reductase withd archd die tener in omenen die tener in omenen in one	0600657	pramu /992		IIGKU1292		-1 -	7 5		7		4
arCOG06880	0	Posttranslational modification, protein turnover, d Prad	al type Zn finger domain	OG00484			TIGR02349	2 2	7 7	-	 	1	n +	4
arCOG02553	0	Posttranslational modification, protein turnover, dAprE Su	btilisin-like serine protease	OG01404	pfam05048			2	1		1			
arCOG 00980	0	Posttranslational modification, protein turnover, dSlpA FK	BP-type peptidyl-prolyl cis-trans isomerase 2	OG01047	pfam00254						1 1	1	1	2
arCOG01139	0	Posttranslational modification, protein turnover, dRri1 Pr	oteasome lid subunit RPN8/RPN11, contains Jab1/MPN domain metalloenz/C	:0G01310	pfam14464 c	d08072					1 1	1	_	
arCOG03610	0	Posttranslational modification, protein turnover, d AprE Su	btilisin-like serine protease	:0G01404	0	d02619		1		1	1 2			3
arCOG01331	0	Posttranslational modification, protein turnover, d HtpX Zn	dependent protease with chaperone function	0600501	pfam01435	0.000		, 1 1	1	=	1	,	,	,
arC0602436	- c	Posttranslational modification, protein turnover, gPKE1 20. Posttranslational modification protein turnover, d NosY AB	s proteasome, aipna subunit C-type transport system involved in multi-copper enzyme maturation perd C	0601277	ptam10584, ptam0022 [C	dU3 / 5b	11GKU3033	 		_		1	-	-
arCOG05153	0	Posttranslational modification, protein turnover, dGroS	-chaperonin GroES (HSP10)	0G00234	pfam00166 c	d00320		1 1			1			
arCOG05154	0	Posttranslational modification, protein turnover, d GroEL Ch	aperonin GroEL (HSP60 family) C	:0G00459	pfam00118 c	d03344	TIGR02348	1 1			1			IT
arCOG02816	0	Posttranslational modification, protein turnover, d MsrA Pe	ptide methionine sulfoxide reductase	:0G00225	pfa m01625		TIG R00401	1 2	1		1		1	3
arCOG00970	0	Posttranslational modification, protein turnover, d PRE1 20.	S proteasome, beta subunit	:0G00638	pfam00227 c	d03764	TIGR03634	1	1	1	1	1	1	5

arCOG00270	0 Posttranslational modification, protein turnover, d CcmF	Cytochrome c biogenesis factor	0601138	ofa m01578	TIGROOS	353 1			-	1 2	
arCOG00310	O Posttranslational modification, protein turnover, d Bcp	Peroxiredoxin 0	:0G01225	pfam00578 cd0	3018 TIGR031	137 2	2		1	1 2 1:	1
arCOG04560	O Posttranslational modification, protein turnover, discA	Fe-S cluster assembly iron-binding protein lscA	COG00316	pfam01521	TIGR000	049 2	3	1 1	1	1 1	
arCOG03607	O Posttranslational modification, protein turnover, d-	Cysteine protease, C1A family	0G04870	pfam00112 cd0	2619	1				, , ,	
arc0G04236 arC0G04236	 Posttranslational modification, protein turnover, graad Posttranslational modification, protein turnover, dSufC 	Interal-surur cluster plosynchetic enzyme Cysteine desulfurase activator ATPase	0600396	pfam00005 cd0	3217 TIGR019	945				2 T	
arCOG04064	O Posttranslational modification, protein turnover, dRseP	Membrane-associated protease RseP, regulator of RpoE activity in bacteria	0G00750	pfam02163 cd0	6159 TIGR000	054			1	1 2	
arCOG01226	O Posttranslational modification, protein turnover, d ArgK	Putative periplasmic protein kinase ArgK or related GTPase of G3E family 0	:0G01703	pfam03308 cd0	3114 TIGR007	750			11	1 1 2	
arCOG01308	O Posttranslational modification, protein turnover, d Cdc48	ATPase of the AAA+ class , CDC48 family	0G00464	pfam00004, pfam0000 cd0	0009,cd000 TIGR012	243				2 2 4	
arCOG00267	O Posttranslational modification, protein turnover, dCcmC	ABC-type transport system involved in cytochrome c biogenesis, permease col	0600755	pfam01578 cd0	TIGR011	191					
arcOG04957	O Posttranslational modification, protein turnover, 4 Grac O Posttranslational modification protein turnover, 4 CrmF	Cutorhrome c-type hingenesis protein CrmF	0602332	prarriuu4oz cau	1207011 01.62	061					
arCOG02438	O Posttranslational modification, protein turnover, dNosY	ABC-type transport system involved in multi-copper enzyme maturation, perri	0G01277	pfam12679						8	
arCOG03614	O Posttranslational modification, protein turnover, d-	Cysteine protease, C1A family	:0G04870	pfam00112 cd0	2619				1		
arCOG00065	O Posttranslational modification, protein turnover, d csdA	Selenocysteine lyase/Cysteine desulfurase	COG00520	pfam00266 cd0	6453 TIGR019	979 3	2	1 1 1	2 1	4 2	
arCOG00479	O Posttranslational modification, protein turnover, d CyoE	Polyprenyltransferase (cytochrome oxidase assembly factor)	OG00109	pfam01040 cd1	3957 TIGR014	473			2		
arCOG04142	O Posttranslational modification, protein turnover, d DYS1	Deoxynypusine synthase	0601899	ptam01916	1 2 77 CHOMORE	321 2	71 7	2 I I 2	5 c	1 2 4	
arCOGO4772	O Posttranslational modification protein turnover denF	Molecular chanerone Grof (heat shock protein)	0600576	pfam0102500 cut	0446	t -	+ 0	- F	2 C	- -	
arCOG03060	O Posttranslational modification, protein turnover, 4 DnaK	Chaperone DnaK (HSP70)	OG00443	pfam00012 cd1	0234 TIGR023	350 1	7 2		2	1 1 1 1 1 1	
arCOG03103	O Posttranslational modification, protein turnover, d CtaA	Uncharacterized protein required for cytochrome oxidase assembly	:0G01612	pfam02628		2		2	2		
arCOG01972	O Posttranslational modification, protein turnover, dTrxA	Thiol-disulfide isomerase or thioredoxin	:0G00526	pfam00085 cd0	2947 TIGR010	068 2	2	1 1	2	2 3	
arCOG01715	O Posttranslational modification, protein turnover, d SufB	Cysteine desulfurase activator SufB	0G00719	pfam01458	TIGR019	981	2 1	•	2	1 2 2	
arCOG02007	O Posttranslational modification, protein turnover, 4Apre O Posttranslational modification protein turnover 4-	Dranultransferace family nrotain containing thioradoxin domain	0601331	prarmouosz cau nfam03190 cd0	7477 IIGRU33	c 176	0 -	4 4	0 1	n n	
arCOG02834	O Posttranslational modification, protein turnover, d RseP	Membrane-associated protease RseP, regulator of RpoE activity in bacteria	0G00750	pfam02163, pfam1318 cd0	6159,cd009 TIGR000	054			·	1	
arCOG00976	O Posttranslational modification, protein turnover, d Pcm	Protein-L-isoaspartate carboxylmethyltransferase	:0G02518	pfam01135 cd0	2440 TIGR000	080			1	1 1 2	
arCOG03202	O Posttranslational modification, protein turnover, q-	Collagenase family protease	COG00826	pfam01297, pfam01136					н Н		
arCOG06807	O Posttranslational modification, protein turnover, d-	Cytochrome c biogenesis factor	0G01138								
arC0G02160	O Posttranslational modification, protein turnover, d LonB	Predicted ATP-dependent protease	0601067	ptam01078, ptam13654, pt pfam00578, ptam1.041 pt	am05362 TIGR007	764 1				1 1 1	
arCOG01342	O Posttranslational modification, protein turnover, garipo O Posttranslational modification protein turnover d GimC	Anpu/15A tarmity peroxine uoxin Prefoldin chaneronin cofactor	0601382	pramota20 % pramitu41 cd0		1 T T T T T				-	
ar COG 01 3 4 2	 Posttranslational modification; protein turnover, 4 mile D Posttranslational modification. protein turnover. d CcmB 	ABC-type transport system involved in cytochrome c blogenesis. permease col	0602386			1 1				-	
arCOG01846	0 Posttranslational modification, protein turnover, dPaaD	Metal-sulfur cluster biosynthetic enzyme	OG02151	pfam01883, pfam0914 cd0	2037 TIGR025	945,TIGR 1					
arCOG02815	O Posttranslational modification, protein turnover, d-	Conserved domain frequently associated with peptide methionine sulfoxide r	:0G00229	pfam01641	TIG R003	357 1	2				
arCOG00981	O Posttranslational modification, protein turnover, dSlpA	FKBP-type peptidyl-prolyl cis-trans isomerase 2	COG01047	pfam00254	TIGR001	115 1		1		1 1 1	
arCOG03580	O Posttranslational modification, protein turnover, dSTE14	Putative protein-S-isoprenylcysteine methyltransferase	0G02020	pfam04191				1		1	
arCOG06823	O Posttranslational modification, protein turnover, dAprE	Subtilisin-like serine protease	0G01404	pfam00082 cd0	5562 TIGR039	921 1	,				
arc0602173	O Posttranslational modification, protein turmover, d NrdG	Organic radical activating enzyme	0600602	ptam13394 [cd0	1335 TIGR035	963					
arc0602443	O Posttranslational modification, protein turnover, q lap O Posttranslational modification protein turnover d NosY	ZII-dependent armino- or carboxypeptudase, MZ8 farmiy ABC-type transnort system involved in multi-conner enzyme maturation, perio	0601277	pfam12679 cau	0607					-	
arCOG03028	O Posttranslational modification, protein turnover, dNifU	Fe-S cluster biogenesis protein NfuA, 4Fe-4S-binding domain	OG00694	pfam01106	TIGR020	000					Ι
arCOG01297	O Posttranslational modification, protein turnover, d AhpF	Alkyl hydroperoxide reductase, large subunit	:0G03634	pfam07992	TIGR031	140					
arCOG02606	O Posttranslational modification, protein turnover, dGrxC	Glutaredoxin	:0G00695	pfam00462 cd0	2976 TIGR021	196		1			
arCOG 10598	O Posttranslational modification, protein turnover, d-	ATP-dependent protease La PUA-like domain	1	pfam02190				ti i		1	
arCOG00185	O Posttranslational modification, protein turnover, dGsiA	ABC-type glutathione transport system ATPase component, contains duplicated	0G01123	pfam00005, pfam0000 cd0	3225,cd032 TIGR032	269				ц ц	
arCOG05850	0 Posttranslational modification, protein turnover, d Pcp	Pyrrolidone-carboxylate peptidase (N-terminal pyroglutamyl peptidase) [0602039	pfam01470 cd0	0501 TIGR005	504					
	O Posttranslational modification protein turnover 4-	Abc-type surface transport system; permease component. DshA family protein	CCCONDO	100		141					
arCOG00636	O Posttranslational modification, protein turnover, d HypE	Hydrogenase maturation factor	0G00309	pfam00586, pfam0276 cd0	6061 TIGR021	124					
arCOG00946	O Posttranslational modification, protein turnover, d PfIA	Pyruvate-formate lyase-activating enzyme	:0G01180	pfam04055 cd0	1335 TIGR043	337				1	
arCOG01187	O Posttranslational modification, protein turnover, dHypF	Hydrogenase maturation factor	COG00068	pfam00708, pfam07503, pf	am07503,pf TIGR001	143				1	
arCOG01218	O Posttranslational modification, protein turnover, dTrxA	Thiol-disulfide isomerase or thioredoxin	0G00526	pfam08484,pfam1319 cd0	2975,cd029 TIGR021	187				FL.	
arCOG01716	O Posttranslational modification, protein turmover, dSufB	Cysteine desulfurase activator SufB	0600719	pfam01458	TIGR019	980					
arCOG01832	O Posttranslational modification, protein turmover, dlbpA	Molecular chaperone (HSP20 family)	06000/1	ptam00011 [cd0	6464	164					Ι
ar COG02 737	O Postfranslational modification protein turnover, 4 comp	Predicted Fe-Mo cluster-binding protein. NifX family	0601433	pfam02579 crd0	0003,44000 119 000	ŧ0,					
arCOG03689	O Posttranslational modification, protein turnover, d NosY	ABC-type transport system involved in multi-copper enzyme maturation, perrif	OG01277								
arCOG04427	O Posttranslational modification, protein turnover, dHypC	Hydrogenase maturation factor	:0G00298	pfam01455	TIGR000	074	_	_		1	
arCOG04428	O Posttranslational modification, protein turnover, dHypD	Hydrogenase maturation factor	OG00409	pfam01924	TIGR000	075				Ē	
arCOG00322	O Posttranslational modification, protein turnover, dTldE	Inactivated Zn-dependent protease, component of TldD/TldE system	0G00312	pfam01523						2 1	
	O Postfranslational modification, protein turmover, 4 HdD	Ln-dependent protease, component of 11dU/ 11dE system	0600312	ptamU1523							
arCOG00953	O Posttranslational modification, protein turnover, q-	Predicted redox protein, regulator of disulfide pond formation	0601180	pretrinu 2000	TIGR 027	494					
arc0G04933	O Posttranslational modification. protein turnover, 4-	Serine protease inhibitor	0001180	pfam00079 cd0	0172	101				2 2	
arCOG02734	O Posttranslational modification, protein turnover, d-	Predicted Fe-Mo cluster-binding protein, NifX family	:0G01433	pfam02579 cd0	0851					4	
arCOG07441	O Posttranslational modification, protein turnover, dSurA	Parvulin-like peptidyl-prolyl isomerase	:0G00760	pfam00639	TIGR025	933				1 1	
arCOG01929	O Posttranslational modification, protein turnover, dXdhC	Xanthine and CO dehydrogenase maturation factor, XdhC/CoxF family	:0G01975	pfam13478						1 2	
arCOG04648	O Posttranslational modification, protein turnover, d-	Gutaredoxin-related protein	0600278	pfam00462 cd0	3028 TIGR003	365				1 1	
ar COG02 7 846	O Posttranslational modification protein turnover d'Inal	Prutative 303 response-associated pepriudae reun	0600484	nfam00256 rd0	6257 TIGR025	3.49					
arCOG03947	O Posttranslational modification, protein turnover, dcpA	ATP-binding subunits of Clp protease and DnaK/DnaJ chaperones	OG00542	pram02861, pfam0286 cd0	0009,cd000 TIGR033	346					
arCOG04712	O Posttranslational modification, protein turnover, d ECM4	Predicted glutathione S-transferase	:0G00435	pfam13409, pfam1341 cd0	3190					1	
arCOG03026	O Posttranslational modification, protein turnover, d NifU	Fe-S cluster biogenesis protein NfuA, 4Fe-4S-binding domain	COG00694	pfam01106	TIGR020	000				1	
arCOG01153	T Signal transduction mechanisms CpdA	3',5'-cyclic AMP phosphodiesterase CpdA	0G01409	pfam00149, pfam0145 cd0	7400,cd001 TIGR000	040,TIGR02689		•		, , ,	
arCOG04425	T Signal transduction mechanisms - Wzb	Internorarie proteinase or CAAA superramity, regulator or anti-sigma ractor Protein-tyrosine-phosphatase	0G00394	pram1330/ pfam01451 cd0	0115 TIGR026	689	7 7		7	т Т Т Т Т	
arCOG01171	T Signal transduction mechanisms RAD55	RecA-superfamily ATPase implicated in signal transduction	0G00467	pfam06745 cd0	1124 TIGR038	877 2	2		2 2	10 3	
arCOG02967	T Signal transduction mechanisms	NACHT family NTPase fused to HEAT repeats domain	COG05635	pfam13646, pfam13646		1		2	1	1	Π
arCOG02053	T Signal transduction mechanisms UspA	Nucleotide-binding protein, UspA family	:0G00589	pfam00582 cd0	0293	1			-1	1	Π
arCOG05374	T Signal transduction mechanisms	GAF domain-containing protein	:0G01956	pfam13185		1					

arCOG06712	T Signal transduction mechanisms	AtoS	Sensory protein, contains PAS domain	COG02202	pfam13426 cd00130	TIGR00229			1		1		m		
arCOG02333	T Signal transduction mechanisms		Signal transduction histidine kinase, contains REC and PAS domains	COG00642	pfam13492, pfam0051 cd00082, cu	d000 TIG R02966									
arCOG04809	T Signal transduction mechanisms		Signal transduction histidine kinase	COG00642	pfam00512, pfam0251 cd00082, ct	d000 TIGR02956									
arCOG01180	T Signal transduction mechanisms	RIO1	Serine/threonine protein kinase involved in cell cycle control	COG01718	pfam01163 cd05145	TIGR03724	1		1				1	1	2
arCOG04820	T Signal transduction mechanisms		SOUL heme-binding protein		pfam04832	-	-	1	1		1				1
arCOG06193	T Signal transduction mechanisms		Signal transduction histidine kinase and PAS domains	COG00642	pfam13426, pfam1342 cd00130, c	d001 TIGR00229,	TIGR00229, TI								
arCOG04453	T Signal transduction mechanisms	DisA_N	Diadenylate cyclase (c-di-AMP synthetase), DisA_N domain	COG01624	pfam02457								e I	-1	
arc0G03799	T Signal transduction mechanisms	Atos	GAF, PAS/PAC domains containing signal transduction protein	0602202	ptam05763			+						2	-1
arCOG00449	T Signal transduction mechanisms	UspA	Nucleotide-binding protein, UspA tamily	0000589	ptam00582,ptam0058 cd00293,c	d00293		1							
arcoco1172	T Cianal transduction mechanisms	RAU55	Reck-superiamity ATPase implicated in signal transduction	100000	pramoe745	TICECOGOGU					ł				
arCOG01177	T Cimal transduction machanisms	DADEE	DecA-superiarmity ATEase implicated in signal transduction DecA-superfamily ATDase implicated in signal transduction	1040000	pfam06745 c40011333										
arcOG02327	T Signal transduction mechanisms	-	Signal transduction histidine kinase, contains PAS domain	0600642	pfam00130 pfam13426 pfam0251 cd00130 cr	1000 TIGR 07966					T		-		
arCOG02385	T Signal transduction mechanisms	CheY	Rec domain	:0G00784	pfam00072 cd00156	TIG R01818									
arCOG03803	T Signal transduction mechanisms		Membrane assicated inactivated KaiC-like ATPase, DUF835 family	C.00096	pfam05763								1		
arCOG03804	T Signal transduction mechanisms		Membrane assicated inactivated KaiC-like ATPase, DUF835 family	C.00096	pfam05763								1		
arCOG04404	T Signal transduction mechanisms		Cell fate regulator YlbF, YheA/YmcA/DUF963 family (controls sporulation, com	COG03679	pfam06133	_							1		
arCOG06538	T Signal transduction mechanisms		Signal transduction histidine kinase	COG00642		TIGR 002 29,	TIGR02956						1		
arCOG01143	T Signal transduction mechanisms	ApaH	Serine/threonine protein phosphatase PP2A family	COG00639	pfam00149 cd00144	_							2	2	
arCOG01178	T Signal transduction mechanisms	GvpD	GvpD gas vesicle protein, contains RecA/KaiCfamily ATPase domain	COG00467	pfam07088 cd01124	TIGR02655							2		
arCOG02391	T Signal transduction mechanisms	CheY	Rec domain	COG00784	pfam00072 cd00156	TIG R02154							m	1	S
arCOG03413	T Signal transduction mechanisms	CDC14	Protein-tyrosine phosphatase	COG02453	pfam00782 cd00047									1	
arCOG03517	T Signal transduction mechanisms		Formylglycine-generating sulfatase enzyme	COG01262	pfam12867, pfam03781	TIGR03440								1	
arCOG00451	T Signal transduction mechanisms	UspA	Nucleotide-binding protein, UspA family	0000589										2	2
arc0G01992	T Signal transduction mechanisms	SIXA	Phosphohistidine phosphatase SixA	0602062	ptam00300 cd07067	TIGR00249		$\left \right $							
arCOG06801	T Signal transduction mechanisms	SrkA	stress-induced morphogen (activity unknown) Ser/Thr protein kinase RdoA involved in Cox stress response. MazF antagonisti	0602334	pfam01/22 pfam01636 cd05153										- m
															,
METABOLISM															
arCOG01768	F Amino acid transnort and metabolism	GING	Membrane associated serine protease	10600705	hfam01694	-	-	-	-	-	-	-	-	-	
arcocot/00	E Amino acid transport and metabolism		Histidue ammonia-jusce	9800000	pfam01331 2400332	TIGD 01 3 25	4 -			4 -	+ -	• •		4 -	
T/0400000	E Amino acid transportaria metabolism	Veria	To demondant continuoridana AAD familu	00000112	7550000 TZZ0011910	CZZTONOII			-						n r
ar COGO1252	E Amino acid transport and metabolishi E Amino acid transport and metabolism	AbA	Zir-ueperiuerit dai Juoxypepritaase, ivi 22 ratiiriy Girtamata dahvdronaanasa (laricina dahvdronanasa	1100000	pfam02013 pfam020 cd0400					-	7				۲ ۲
ar COULDUE	E Amino acid transport and metabolism		ABC two disortido/olizosontido/sido/disorto terranti rustom normano composi-	000001		TIC DOT 700	4 -		-			• •		-	4
ar COG 00 1 00	E Amino acid transportaria metabolism		AL Dimethyle upe priver oligope priver i Ilocer transport system, permease component	TOPODO	1020000 V2000000	TIC D01070		-			+ +		n +	-	c
	r Amino add transport and metabolism	LING LIN		0000000	pidm022/4	LIGRUTU/8	× (-		nı
	r Amino add transport and metabolism	AIG		0000000	pidifiusz 22, pidifiut 20 cd03304	TOCODYDII	× (nı
arc0609400	E Amino acid transport and metabolism	PIL6	INAD(P) transnyurogenase peta supurit	7070700	pidm02233	our course		-			-			-	n
arCOG00184	E Amino acid transport and metabolism	AppF	ABC-type oligopeptide transport system, ATPase component	0604608	ptam00005 cd03257	TIG R02 769	7 7	~					4		
arCOG01700	E Amino acid transport and metabolism	SpeB	Arginase family enzyme	COG00010	pfam00491 cd11593	TIGR01230	-1	-1			1		e I	-1	m
arCOG05395	E Amino acid transport and metabolism	FtcD	Formiminotetrahydrofolate cyclodeaminase	COG03404			-1	-1					e I	2	5
arCOG00181	E Amino acid transport and metabolism	DppD	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	OG00444	pfam00005 cd03257	TIG R02770	-		1	1			4		
arCOG01430	E Amino acid transport and metabolism	CysK	Cysteine synthase	OG00031	pfam00291 cd01561	TIG R01136							-	1	
arc0600915	E Amino acid transport and metabolism	GabT	4-aminobutyrate aminotransferase or related aminotransferase	0000160	ptam00202 cd00610	TIGR00707		$\left \right $				- 1	2	,	
arc0600035	r Amino add transport and metabolism	916		0000001		TIC DOL 230		•	-				Ŧ		ç
	E Amino acid transport and metabolishi E Amino acid transport and metabolism	Dh+D	Urudijate lijvu ataše Thraonina affluv nrotaln	0001280	C/TTOHIBIC						+ +		-		۲ ۲
arCOG00343	E Amino acid transport and metabolism	1 VSD	Sacriate ettud proteine Sacriate debudromense or related enzyme	8V21000	premozazo	CHECONIDII	4 -				+ -			4 -	4 0
arCOG01158	E Amino acid transport and metabolism	SerB	Phochoserine phosphatese	DGODEGO	nfam00702	TIGR01491	4 -	-	-					-	n c
arCOG01924	F Amino acid transnort and metabolism	AnsR	1-asparadinase/archaeal Gli+tRNAGin amidotrancferase subjunit D	0600252	nfam00710 rd08962	TIG R02153			-	╞				n .	• ~
arCOG01888	F Amino acid transnort and metabolism	AmnS	Leucyl aminonentidase (aminonentidase T)	0602309	nfam02073	00100			-	╞			•		• ~
arCOG04897	E Amino acid transport and metabolism	-	Aspartate /tvrosine/aromatic aminotransferase	0600436	pfam00155 cd00609	TIGR01141	1 -	-			+ -				4
arCOG00924	E Amino acid transport and metabolism	LivF	ABC-twoe branched-chain amino acid transport system. ATPase component	0600410	bfam00005 cd032.24	TIGR03410				-				-	1
arCOG00925	E Amino acid transport and metabolism	LivG	ABC-type branched-chain amino acid transport system. ATPase component	OG00411	pfam00005 cd03219	TIG R03411	1				1 -1				
arCOG01021	E Amino acid transport and metabolism	LivK	ABC-type branched-chain amino acid transport system, periplasmic compone	COG00683	pfam13458 cd06268		1	~.		1	1				1
arCOG01273	E Amino acid transport and metabolism	LivM	ABC-type branched-chain amino acid transport system, permease component	:0G04177	pfam02653 cd06581		1 2	~			1			-1	
arCOG00912	E Amino acid transport and metabolism	ArgF	Ornithine carbamoyltransferase	COG00078	pfa m02729, pfam00185	TIG R00658	1	1	1	1	1 1		1	1	
arCOG04758	E Amino acid transport and metabolism	PepF	Oligoendopeptidase F	COG01164	pfam08439, pfam0143 cd09608	TIGR00181	1	1	1	1	1 1				1
arCOG01646	E Amino acid transport and metabolism	DAP2	Dipeptidyl aminopeptidase/acylaminoacyl-peptidase	COG01506	pfam07859 cd00312		3	3	2		1 1		1	2	5
arCOG00071	E Amino acid transport and metabolism	AsnB	As paragine synthase (glutamine-hydrolyzing)	COG00367	pfam00733 cd01991	TIGR01536			-1	1	1			ر ا ،	2
arcoc01210	E Amino acid transport and metabolism	GCVP GCVP	laycine cleavage system protein P (pyridoxai-pinding), C-terminai domain	OCOPED1	pramu2347 cd00613	TIGROU461			-				-1	-	
arcocon 770	E Amino acid transport and metabolism	Puth	Ind+//promite symptoments	TECONOD	ptdm04/4 cd114/4 cd114/4	CTONNAL	, ,						÷		
arcOG00863	E Amino acid transport and metabolism	ArcC	Isuasparty peptidase or easparaginase, ixurriyu orase superrarmiy Carbamata kinasa	04440000	pfam0696 cd04235	TIG BOU746					+ +			-	
arCOG01534	E Amino acid transport and metabolism	DdnA	ABC-two transport system parinlasmic component	2420000	pfam00030 cd04533	TIGE CD 2 QA					+ +		-	-	Ŧ
arCOG01333	E Amino acid transport and metabolism	unno -	Accorpt transport system; peripresmit component Acnortate /httosine/aromatic aminotra neferace	10000136	bfam00155 cd00600	TIGP/02017		$\left \right $			+ +		2		
arcOG0088	E Amino acid transport and metabolism		Gamma-ditramvi-ramma-aminohithrrate hydrolase Duith (nitrascine derradd	0600011	pfam02203 cd0003	abaUUd DIL		-	-		-	-			
arcOG0060	E Amino acid transnort and metabolism	MatC	Outstathioning heta-lvase/rustathioning gamma-svnthase	10G00676	nfam01053 cd00614	TIGR 02 080			•	4	1	۰ -		·	
arCOG01107	E Amino acid transport and metabolism	ArgE	Acetylomithine deacetylase/SuccinvI-diaminopimelate desuccinvlase or related	:0G00624	pfam01546 cd05650	TIGR01910	- 7				2	1 et	2	4	4
arCOG02706	E Amino acid transport and metabolism	GloA	Lactoviglutathione lyase or related enzyme	COG00346	pfam13669 cd07249	TIG R03081	2			1	1 2	~ ~			1
arCOG00096	E Amino acid transport and metabolism	GItB	Glutamate synthase domain 3	COG00070	pfam01493 cd00981	TIG R03122	1	-	1		1 2				1
arCOG05229	E Amino acid transport and metabolism	PepD	Di- or tripeptidase	COG02195	cd03890	TIG R01893	1	1	1		1 2				
arCOG02297	E Amino acid transport and metabolism	IIvE	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyd	COG00115	pfam01063 cd01558	TIGR01122	2 2		1		2			2	2
arCOG01000	E Amino acid transport and metabolism	РерР	Xaa-Pro aminopeptidase	COG00006	pfa m01321, pfam0055 cd01092	TIG R00500	2 1		1		3	1	33	2	S
arCOG01130	E Amino acid transport and metabolism		Aspartate/tyrosine/aromatic aminotransferase	COG00436	pfam00155 cd00609	TIG R03540	2 3	3 2	1		e		m	2	9
arCOG04490	E Amino acid transport and metabolism	PdaD	Pyruvoyl-dependent arginine decarboxylase (PvlArgDC)	COG01945	pfam01862	TIGR 00 286	1	1	1	1			1		
arC0G01533	E Amino acid transport and metabolism	DdpA	AB C-type transport system, periplasmic component	0000/47	ptam00496 cd08512	01 - FCD017C0			-	7					Ŧ
arCOG00458	E Amino acid transport and metabolism	- 10	Carbamotypriospriate synthase range suburnt Archaemetzincin family protease	0601913	praritiou263/praritio2760/praritio27				-	-				-	-
arCOG02165	E Amino acid transport and metabolism		Transglutaminase-like cysteine protease	:0G01305	pfam01841	╞	1							1	
arCOG00064	E Amino acid transport and metabolism	CarA	Carbamoyiphosphate synthase small subunit	:0G00505	pfam00988, pfam0011 cd01744	TIGR01368	1	1		1		-1		1	1

arCOG01431	E Amino acid transport and metabolism	A T	hreonine dehvdratase	:0G01171	pfam00291 ci	d01562.cd048	TIG R01127	1	1	1		1		
arCOG00923	E Amino acid transport and metabolism GInC	A Dr	BC-type polar amino acid transport system, ATPase component	:0G01126	C.	d03262	TIGR03005			1		1		2
arCOG00771	E Amino acid transport and metabolism -	J	ubicO group peptidase, beta-lactamase class C family [0	:0G01680	pfa m00144							1		
arCOG01798	E Amino acid transport and metabolism HisN	A C	BC-type amino acid transport system, permease component	0600765	pfam00528 ci	d06261	TIGR03003	- -	7 7	1 2		2		m +
arcog01909	E Amino acid transport and metabolism E Amino acid transport and metabolism GInA		ornumine/acetyionnume aminouransierase	0600174	pramouzuz ofam00120	OTODD	TIGR03105		т г т	-				т 7
arCOG00756	E Amino acid transport and metabolism GcvT	VT G	ilycine cleavage system T protein (aminomethyltransferase)	0G00404	pfam01571, pfam08669		TIGR00528	1 1	1					1
arCOG01303	E Amino acid transport and metabolism GcvF	VH 0	ilycine cleavage system H protein (lipoate-binding)	COG00509	pfam01597 CI	d06848	TIGR00527	1 1	1				2	5
arCOG06322	E Amino acid transport and metabolism PutA	tA P	roline dehydrogenase	COG00506	pfa m01619			1 1	1				1	
arCOG00755	E Amino acid transport and metabolism Dadu	add G	ilycine/D-amino acid oxidase (deaminating)	0600665	pfam01266									2
arcoc01110	E Amino acid transport and metabolism Kritte E Amino acid transport and metabolism Arge	9	rreonine emux protein catilornithina dascatilasa Kucrinyi-diaminonimalata dasurcinyiasa or ralatad	000000	prarmutatu	405621	TICD01010	7 1	-	-				-
arCOG02767	E Amino acid transport and metabolism	2	Actionnum concertions / action from the manual from the second	0G01266	pfam02517	TONCON	OTCTOURI	1		1				-
arCOG03600	E Amino acid transport and metabolism -	F	ransglutaminase-like cysteine protease	:0G01305				1	1					
arCOG00177	E Amino acid transport and metabolism PotA	tA A	BC-type spermidine/putrescine transport system, ATPase component [:0G03842	pfa m00005 ci	d03301	TIGR03265	1	1	_				1
arCOG03101	E Amino acid transport and metabolism GloA	DA L	actoylglutathione lyase or related enzyme	:0G00346	pfam00903 CI	d08346		1	1					
arCOG14875	E Amino acid transport and metabolism -	4	spartyl protease	000001	pfam05618	CC CO 14	TIC DO0013	-1 c		-				1 0
arcoconora	E Amino acid transport and metabolism Puth	4 -	Ia+/proline symporter	1600036	pramuu4 / 4 CC	d10322	TIGK00813	7						'n
arCOG07760	F Amino acid transport and metabolism	D	-tyrosine decarooxylase, PLF-dependent protein entidase family C25	0/0000	pram01364 c	d02758							-	
arCOG01512	E Amino acid transport and metabolism HvuE	nB UB	I-methylhydantoinase B/acetone carboxylase, alpha subunit	:0G00146	pfam02538	004400								
arCOG01949	E Amino acid transport and metabolism ArgC	й 60	ysine/arginine efflux permease	:0G01279	pfa m01810		TIGR00948	1						
arCOG05121	E Amino acid transport and metabolism TesA	sA L	ysophospholipase L1 or related esterase	:0G02755				1						
arCOG00121	E Amino acid transport and metabolism AsnE	nB A	s paragine synthase (glutamine-hydrolyzing)	COG00367	pfam00733				1					
arCOG03999	E Amino acid transport and metabolism		lostripain family peptidase, C11	0000110	pfam03415	100100	TIGR 02806		1					Ŧ
arcOG02611	E Amino acid transport and metabolism LIVH	τ Π	ranched-chain amino acid ABC-type transport system, permease component antidace C1A subfamily	0600559	pramu2653 CC	006536 cd142	11GKU3409 54	ł		-			¢	-
arCOG015951	E Amino acid transport and metabolism - CarR	4	eptidase CLA Su Dramily arhamovlinhosnhate svinthase large su hunit	0600458	nfam15632	dU8546,C0142	54 TIGR01205	ł					7	
arCOG00077	E Amino acid transport and metabolism GovP	vP 0	al barroy prospirate syntriase iai ge subdinit. ilycine cleavage system protein P (pyridoxal-binding). N-terminal domain	0000403	pfam02347 C	d00613	TIGR00461	ł		1			-	1
arCOG01434	E Amino acid transport and metabolism ThrC	2	hreonine svnthase and cysteate synthase	OG00498	pfam01155, pfam0029 ci	d01563	TIGR02605, TIGR00	0260						1
arCOG00279	E Amino acid transport and metabolism SpeC	eD S	-adenosylmethionine decarboxylase/arginine decarboxylase	0G01586	pfa m0 2675		TIGR03330							2
arCOG00073	E Amino acid transport and metabolism Cysh	Ρ	APS reductase related enzyme fused to RNA-binding PUA domain and ferred	:0G00175	pfam01507 ci	d01713	TIGR00434							2
arCOG00009	E Amino acid transport and metabolism PotE	itE A	mino acid transporter	COG00531	pfa m00324		TIGR00909							
arCOG00050	E Amino acid transport and metabolism SpeE	eE	permidine synthase	0G00421	pfam01564 ci	d02440	TIG R00417	_						
arCOG00092	E Amino acid transport and metabolism LdcC	P CC	rginine/lysine/ornithine decarboxylase	0G01982	pfam03709, pfam0127 ci	d00615	TIGR04301						_	
arCOG00304	E Amino acid transport and metabolism HIS2	S2	listidinol phosphatase or related hydrolase of the PHP family [0G01387	pfam02811 ct	d12110	TIGR01856							
arCOG01123	E Amino acid transport and metabolism	thA N	Aethylthioribose-1-phosphate isomerase (methionine salvage pathway), a pal	0600182	ptam01008		TIGR00512							
arc0601292	r Amino acid transport and metabolism		ADPH-dependent glutamate syntnase beta chain or related oxidoreductase is	0600493	pramu /992		116KU1316							
arc060164/	r Amino acid transport and metabolism	1	erine protease or the peptidase ramity 59A	0601505	pramuu326									
arcoc0167	E Amino acid transport and metabolism Iviet	- L	ircnaealo-agenosymerinonine syncretase	712020	pramoteat			ł						
arc0602766	E Amino acid transport and metabolism -	- 12	ransgiutaminase-iike cysteme protease Aatal-danandant mambrana nrotaasa CAAY familu	0601366	ptam05517			ł						
	E Amino acid transport and metabolism -		rietai-uepenuent memorane protease, CAAA tamiy	0077090	/TGZOWPid			T						
arcOG03672	F Amino acid transport and metabolism		hermonsin-like nortease	0.00065	pfam05317 nfam06837									
arCOG05065	E Amino acid transport and metabolism CdsB		-costeine desulfidase	0603681	pfam03313			ŀ						
arCOG05363	E Amino acid transport and metabolism -		AAX family protease, possibly associated with type IV pili like system	0G01266	pfam02517									
arCOG05808	E Amino acid transport and metabolism DUR:	IR1 A	Ilophanate hydrolase subunit 2	OG01984	pfam02626		TIGR00724							
arCOG05809	E Amino acid transport and metabolism DUR.	JR1 A	llophanate hydrolase subunit 1	:0G02049	pfa m02682		TIGR 00370							
arCOG07546	E Amino acid transport and metabolism	Z	incin superfamily protease										_	
arCOG 10063	E Amino acid transport and metabolism SdaC	aC T	ryptophan/tyrosine permease family	COG00814	pfa m03222		TIGR00837							
arCOG 13501	E Amino acid transport and metabolism -	S	erine dehydratase beta chain		pfam03315 ci	d04879	TIGR00719						_	
arCOG00070	E Amino acid transport and metabolism GlyA	VA C	ilycine/serine hydroxymethyltransferase	0G00112	pfa m00464 ci	d00378		_					1	1
arCOG01698	E Amino acid transport and metabolism Leuc	2 I	lomoaconitate hydratase/3-isopropylmalate dehydratase large subunit famil	0G00065	pfam00330 ct	d01583	TIGR02086							4
arCOG02230	E Amino acid transport and metabolism LeuL	uD 3	 isopropylmalate dehydratase small subunit 	0G00066	pfam00694 ci	d01577	TIGR02087						-	
arCOG00748	E Amino acid transport and metabolism	pC	BC-type dipeptide/oligopeptide/nickel transport system, permease component	0G01173	pfam12911, pfam0052 c	d06261	TIGR02790	+				- /	,	
arCOG00494	E Amino acid transport and metabolism Asd	d -	spartate-semialdehyde dehydrogenase	0600136	ptam01118, ptam02774		TIGR00978							
arc0602014	r Amino acid transport and metabolism	- GE	intriranijate/ para-aminobenzoate syntnase component i	0600147	pramu4/15, pramuu4/5	1005.00	TIGRU1820							
arCOG04134	F Amino acid transport and metabolism AroA		-uerryurouquinate uerryur atase - enclover ivvishikim ate-3-nhosonhate svothase	00000128	pfam00275 ci	401556	TIGR01356							
arCOG06005	E Amino acid transport and metabolism	- d	redicted metalloprotease. contains C-terminal PDZ domain	00003975	pfa m05299. pfam1318 ci	066000	TIGR02037	ŀ						
arCOG07384	E Amino acid transport and metabolism HisB	H H	listiclinol phosphatase or related phosphatase	0G00241	pfam13242 CI	d01427	TIG R01656						1	1
arCOG01025	E Amino acid transport and metabolism Arok	oK2 A	rchaeal shikimate kinase	:0G01685	pfa m00288		TIGR01920						-	2
arCOG01131	E Amino acid transport and metabolism	A	spartate/tyrosine/aromatic aminotransferase	OG00436	pfa m00155 ci	d00609	TIG R01 265						1	2
arCOG04133	E Amino acid transport and metabolism AroC	oC C	horismate synthase	:0G00082	pfam01264 ci	d07304	TIG R00033						1	2
arCOG04353	E Amino acid transport and metabolism -	m	-dehydroquinate synthase	0G01465	pfa m01959								1	2
arCOG02092	E Amino acid transport and metabolism Leue	al Au	opropylmalate/homocitrate/citramalate synthase	OG00119	pfam00682, pfam0103 ct	d07940	TIGR02146							. 3
arCOG04322	E Amino acid transport and metabolism Pept	DB 	eucyl aminopeptidase	0600260	ptam02789, ptam0088 ci	d00433								
arcoconset	E Amino acid transport and metabolism Ami. E Amino acid transport and metabolism	v L	eucyriaminopepridase (aminopepridase r)	06005303	prarmuzu/ 3			ł						4
arcOG01459	F Amino acid transport and metabolism Tdb	<u>۱</u>	s partoninase hreonine dehvdrogenase or related Zn-denendent dehvdrogenase	0601063	nfa m08240 nfam0010 ci	d08235	TIGRODEG2							
arCOG02759	F Amino acid transport and metabolism	2	Aetal-denendent membrane protease. CAAX family	0601266	nfa m02517		*****	T						
arCOG04044	E Amino acid transport and metabolism -	2	-amino-3,7-dideoxy-D-threo-hept-6-ulosonic acid synthase, DhnA-aldolase fa	OG01830	pfam01791 Ct	d00958	TIGR01949						2	4
arCOG00619	E Amino acid transport and metabolism GltB	1B G	lutamate synthase domain 2 and ferredoxin domain	OG00069	pfam01645 ci	d02808								1
arCOG01033	E Amino acid transport and metabolism AroE	OE S	hikimate 5-dehydrogenase	:0G00169	pfa m08501, pfam0148 c	d01065	TIGR00507							1
arCOG01799	E Amino acid transport and metabolism Hisl	sJ A	BC-type amino acid transport/signal transduction system, periplasmic completed	OG00834	pfam00497 ct	d13624	TIGR01096	+		+				1.
arCOG04273	E Amino acid transport and metabolism HISC	S F	listidinol-phosphate/aromatic aminotransterase or cobyric acid decarpoxyiasi. ۲۰۰۰-۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰	0000079	pfam00155 [ci	d00609	TIGR01140	-		-			+	
arcOG01122	Carhohvdrate transnort and metabolism RpiA	1 1	ilyCine Cleavage system 1 protein (aminomeunymansmase)	0600404	pTdiriut2/Language	401 3 GR	TIGR00021	- -	-	-		5		ч с
arCOG00767	G Carbohydrate transport and metabolism {Mar	tanB} P	hosphomannomutase	0G01109	pfam02878, pfam0287 ct	d03087	TIGR03990		1 1	1		4 	2	, m
arCOG00493	G Carbohydrate transport and metabolism Gap/	pA G	ilyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrd	:0G00057	pfam00044, pfam02800		TIG R01546	1	1		1	1	1	2
arCOG00052	G Carbohydrate transport and metabolism Pgi		ilucose-6-phosphate isomerase	:0G00166	pfam01380, pfam1043 ci	d05017,cd056	TIGR02128	1	1		1	1	1	1

arCOG04170 G	Carbohydrate transport and metabolism	GckA	Glycerate kinase	COG02379	pfam13660, pfam05161					_		1	1	1	_	$\left \right $
arCOG04961 G	Carbohydrate transport and metabolism	GlpX	Fructose-1,6-bis phosphatase/sedohe ptulose 1,7-bis phos phatase or related ph	COG01494	pfam03320 c	d01516	TIG R00330			_		1	1		4	J
arCOG00274 G	Carbohydrate transport and metabolism	RhaT	Permease of the drug/metabolite transporter (DMT) superfamily	COG00697	pfam00892, pfam00892	1010 40	TIGR00950	-	•	τ	τ				C1 C	
arCOG001347 G	Carbohydrate transport and metabolism	Bhck	Atcriadat ituccuose-1,0-105pitatase or relateu erizyrite or iriositor iriorioprio. Sugar kinase iribokinase family	COG00524	nfam00794	401174	TIGR02152	1 T				+ +	n m		4 u	
arCOG01993 G	Carbohydrate transport and metabolism	GpmA	Phosphoglycerate mutase 1	COG00588	pfam00300 c	d07067	TIGR01258	1			2				0	
arCOG04431 G	Carbohydrate transport and metabolism	GIPF	Glycerol uptake facilitator or related permease (Major Intrinsic Protein Family	COG00580	pfam00230 c	d00333	TIGR00861	1 1		1		1			2	
arCOG01087 G	Carbohydrate transport and metabolism	TpiA	Triosephosphate isomerase	COG00149	pfam00121 c	d00311	TIG R00419	1 1				1		1	1	
arCOG05046 G	Carbohydrate transport and metabolism	Rpe	Pentose-5-phosphate-3-epimerase	COG00036	pfam00834 c	d00429	TIGR01163	1	1	1		1		1	m	
arCOG00130 G	Carbohydrate transport and metabolism		MFS family permease	COG00477	pfam07690 c	d06174	0001000	1	-1	, 1	,	Ţ,		4,		T
arcocot 114	Carbohydrate transport and metabolism	Eno	Enolase Dhorahaanalana uuda amthara/muunata ahaanhata diitiinna	COCO0148	pramu3952, pramuu11 C	003313 nfnm02006	TIG RU1060	7	,			-1 (Ŧ		7	T
arcogonade 6	Carbohydrate transport and metabolism	PDAD	Prios prioe noi pyruvate syntnase/ pyruvate priospinate uikinase 3-aborabordioreasta kinasa	10000126	prarrio1320, prarrio0391,	012102090		7		-	-	7 C		, 	-	Γ
arC0601053 G	Carbohydrate transport and metabolism	TktA1	J-priospriugrycerate mirase Transketolase - N-terminal subunit	00603959	pram00456	01000p	TIGR00232	C C		cr	6	7 F	7		-	Γ
arCOG00271 G	Carbohydrate transport and metabolism	RhaT	Permease of the drug/metabolite transporter (DMT) superfamily	COG00697	pfam00892, pfam00892		TIG R00950	4 6	,	1	1		- 6	- 2	, m	Γ
arCOG01420 G	Carbohydrate transport and metabolism	GIgA	Glycogen synthase	COG00297	pfam08323, pfam0053 c	d03791	TIGR02095	1		1			त्न			
arCOG03460 G	Carbohydrate transport and metabolism	LmbE	N-acetylglucosaminyl deacetylase, LmbE family	COG02120	pfam02585		TIGR04001	1 2					1	1		
arCOG07581 G	Carbohydrate transport and metabolism		Glycosyl hydrolase family 18, contains cellulose binding domain	COG03325	pfam02839, pfam0080 c	d12215,cd00	146,cd06548	3	1	2			1	2 4	4	
arCOG02796 G	Carbohydrate transport and metabolism		Glucose/sorbosone dehydrogenase	COG02133	pfam07995		TIG R03606			_	_		1		1	T
arCOG00138 G	Carbohydrate transport and metabolism		MFS family permease	COG00477				1 1	-1	1 1						T
arCOG03641 G	Carbohydrate transport and metabolism	PfkA	6-phosphofructokinase	COG00205	pfam00365 c	d00363	TIGR02483	1,		1	1			+		T
arc0G02876 G	Carbohydrate transport and metabolism	CDA1	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 tamily	COG00726	ptam01522 c	d10941	TIGR03006		-	τ	τ				ſ	T
arcogout4/ 6	Carbohydrate transport and metabolism		Wrts Tamiry permease Dentidorivean/vulan/rehitin dearratvilace Dr.d.A.(CD.A.1. family	00000776	prarriu / 030	#/ TOOD	119500390			-	-			T	7	Γ
arCOG00143 G	Carbohydrate transport and metabolism		MFS family nermease	COG00477	nfam07690	d06174	TIGR00711	+								Γ
arCOG07840 G	Carbohydrate transport and metabolism		Glycosyl hydrolase family 18, contains cellulose binding domain	COG03325	pfam02839, pfam0080 c	d12215,cd00	146,cd06543									
arCOG00754 G	Carbohydrate transport and metabolism	LhgO	L-2-hydroxyglutarate oxidase LhgO	COG00579	pfam01266, pfam04324		TIGR03377				1			1		
arCOG04221 G	Carbohydrate transport and metabolism	NagD	Phosphatase of the HAD superfamily	COG00647	pfam13344, pfam1324 c	d01427	TIGR01457							1	2	
arCOG00135 G	Carbohydrate transport and metabolism		MFS family permease	COG00477	pfam05977 c	d06174	TIGR00900							1	m	T
arc0601051 G	Carbohydrate transport and metabolism	TKtAZ	Transketolase, C-terminal subunit	COG03958	ptam02779, ptam0278 c	d07033	TIGR00204								m	T
arcocoraoa	Carbonydrate transport and metabolism	Galk	Galactokinase Chronid transferror indiana to HDD during and the morferror	COC01810	pramitusus, pramu8544		11GK00131									T
arCOG04120 G	Carbohydrate transport and metabolism	DukE	Diguosyi ti alistelase, i elateu tu uur-guuui ui usyiti alistelase Durivate kinace		nfam00234 nfam0288 c	4002.88	TIG R01 064									Γ
arCOG05061 G	Carbohydrate transport and metabolism	TalA	Transaldolase	COG00176	pfam00923	d00956	TIGR00875			ŀ						
arCOG05412 G	Carbohydrate transport and metabolism	BglB	Beta-glucosidase /6-phospho-beta-glucosidase/beta-galactosidase	COG02723	pfam00232	0	TIGR03356									
arCOG00132 G	Carbohydrate transport and metabolism		MFS family permease	COG00477	pfam07690 c	d06174								1		
arCOG00136 G	Carbohydrate transport and metabolism		MFS family permease	COG00477	pfam13347 c	d06174	TIGR01301,TIGR	300895						1		
arCOG00142 G	Carbohydrate transport and metabolism		MFS family permease	COG00477	pfam07690 c	d06174	TIG R00711							1		
arCOG00144 G	Carbohydrate transport and metabolism		MFS family permease	COG00477	pfam07690 c	d06174	TIG R00711							1	_	
arCOG00272 G	Carbohydrate transport and metabolism	RhaT	Permease of the drug/metabolite transporter (DMT) superfamily	COG00697	pfam00892					_				1	_	
arCOG00760 G	Carbohydrate transport and metabolism	McI/CitE	Beta-methylmalyl-CoA lyase, Citrate lyase beta subunit family	COG02301	pfam03328		TIGR01588			_				1	_	T
arCOG01696 G	Carbohydrate transport and metabolism	ApgM	2, 3-bisphosphoglycerate-independent phosphoglycerate mutase	COG03635	pfam01676		TIGR00306							1		
arCOG02602 G	Carbohydrate transport and metabolism	OxdD	Oxalate decarboxylase/archaeal phosphoglucose isomerase, cupin superfamil.	COG02140	pfam06560									1		T
arc0602682 G	Carbohydrate transport and metabolism		MFS tamily permease	COG00477	ptam00083 c	d06174	TIGR00887									T
arcucida278 G	Carbonydrate transport and metabolism	_	Glycosyl nyarolase tamily 57 Difunctional functors 1.6 himbarahata aldalara/aharahatara EBDA/EBDara	COC01449	ptamU3U65 C	56/ NTD										T
arcogon236 G	Carbohydrate transport and metabolism	- UraD	Biruncuonai muctose-1, p-pispinospilate algoiase/priospilatase FBPA/FBPase Dihulose-5-nhocnhate A-enimensee/Euculose-1-nhocnhate aldolase	COGODIES	prarm0506		TIC DO3 3 20			ł						T
arCOG04443 G	Carbohydrate transport and metabolism	RbcL	Ribulose 1.5-bisphosphate carboxylase. large subunit	COG01850	pfam02788.pfam0001 c	d08213	TIGR03326			ŀ				1		Γ
arCOG05728 G	Carbohydrate transport and metabolism		MFS family permease	COG00477	pfam07690 c	d06174	TIGR00881							1	_	
arCOG00139 G	Carbohydrate transport and metabolism		MFS family permease	COG00477	pfam07690 c	d06174				_				2	_	
arCOG01518 G	Carbohydrate transport and metabolism	FrvX	Cellulase M or related protein	COG01363	pfam05343 c	d05656	TIGR03107							8		
arCOG00210 G	Carbohydrate transport and metabolism	TagH	ABC-type polysaccharide/polyol phosphate transport system, ATPase compon	COG01134	pfam00005 c	d03220	TIGR01188								1	T
arCOG04339 G	Carbohydrate transport and metabolism	TagG	ABC-type polysaccharide/polyol phosphate export systems, permease compo	COG01682	pfam01061					+						T
arc0601967 G	Carbohydrate transport and metabolism	PgK2	Z-phosphoglycerate kinase	00002074		00000	TICDOTOCE								7	T
arcogoui/5	Carbohydrate transport and metabolism	GmbA	ABC-type sugar transport system, ATPase component Dhochhohantoca icomaraca	000003039	connuiteration	003239	110103203			ł					-	Γ
arCOG01490 H	Convolyticate transport and metabolism	Fold	Dihydrofolate reductase	COG00273	nfam00186	4002.09		-		-		-	-		4 4	Γ
arCOG01484 H	Coenzyme transport and metabolism	RibD	Pyrimidine reductase, riboflavin biosynthesis	COG01985	pfam01872		TIGR01508	1		1		1		1		
arCOG04262 H	Coenzyme transport and metabolism		Phosphopantothenate synthetase	COG01701	pfam02006			1 1				1	1	1	1	
arCOG04538 H	Coenzyme transport and metabolism	FolD	5,10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate	COG00190	pfam00763, pfam0288 c	d01080		1 2	1	1	1	1	1	1	1	T
arCOG00476 H	Coenzyme transport and metabolism	UbiA	4-hydroxybenzoate polyprenyltransferase or related prenyltransferase	COG00382	pfam01040 c	d13961	TIGR01476	1,	,	1		.,	е ,		,	T
arcocono34 H	Coenzyme transport and metabolism Coenzyme transport and metabolism		Predicted glutamine amidotransterase involved in pyridoxine plosyncresis Nicotinamida mononiclaotida adamiddtrancfaraca	COGOLOEG	pfam01467 C	d01/49	TIG P01 5 3 7		-	-	-			7 -	7 -	Γ
arcog01223 H	Coenzyme transport and metabolism	CAB4	Phosphopantetheine adenvivitransferase	COG01019	pfam01467 C	d02164	TIGR00125						н с і			
arCOG01940 H	Coenzyme transport and metabolism	BirA	Biotin-(acetyl-CoA carboxylase) ligase	COG00340	pfam03099		TIGR00121	1		1		1	1	1	2	
arCOG00584 H	Coenzyme transport and metabolism	PanB	Ketopantoate hydroxymethyltransferase	COG00413	pfam02548 c	:d06557	TIGR00222	2 1	1	1		1	1		2	
arCOG04139 H	Coenzyme transport and metabolism	ApbA	Ketopantoate reductase	COG01893	pfam02558,pfam08546		TIGR00745	,	,	,	,	Ţ,	-1		5 5	T
arcocoraza H	Coenzyme transport and metabolism	KIDB Pihu	3.4-alinyaroxy-2-butanone 4-pnospnate syntnase Dihoelavia conthree hota chain	COCODE 4	pram00926	11000	TIG ROUSUB								n r	T
arCOG01671 H	Coenzyme transport and metabolism	UbiD	3-polyprenvi-4-hydroxybenzoate decarboxylase or related decarboxylase	COG00043	pfam01977	TTTCOD	TIGR03701								n m	
arCOG04713 H	Coenzyme transport and metabolism	RibC	Riboflavin synthase alpha chain	COG00307	pfam00677, pfam00677		TIGR00187	1		1	1	1			m	
arCOG01703 H	Coenzyme transport and metabolism	UbiX	3-polyprenyi-4-hydroxybenzoate decarboxylase	COG00163	pfam02441		TIGR00421	1 1	1		1	1		1	1	
arCOG00040 H	Coenzyme transport and metabolism	Hpt1	Hypoxanthine phosphoribosyltransferase	COG02236	pfam00156 c	d06223	TIGR01251	1 1		1		1		1		
arCOG00478 H	Coenzyme transport and metabolism	UbiA	4-hydroxybenzoate polyprenyltransferase or related prenyltransferase	COG00382	pfam01040 c	d13956	TIGR01475	1 2	۲I.	1		1		_	'	T
arCOG00113 H	Coenzyme transport and metabolism	BioF	7-keto-8-aminopelargonate synthetase or related enzyme	COG00156	pfam00155 c	d06454	TIGR01825	1 0	,	, ,	τ					
arCOG01321 H	Coenzyme transport and metabolism Coenzyme transport and metabolism	Prinb RihA	rterin-4a-carpinolamine genyaratase GTD rvciobvdrolase II	COG00807	pram0.026 nfam0.02 c	d006.41	TIGRODFOG TIGR	-00505 1		7 7	-				v v	
arCOG01482 H	Coenzyme transport and metabolism	NadC	Nicotinate-nucleotide pyrophosphorylase	COG00157	pfam02749, pfam0172 c	d01572	TIG R00078	1		1 1	1	- ₋ -			2	
arCOG04459 H	Coenzyme transport and metabolism	NadA	Quinolinate synthase	COG00379	pfa m02445		TIGR00550	1				1		1	4	
arCOG04137 H	Coenzyme transport and metabolism	SAM1	S-adenosylhomocysteine hydrolase	COG00499	pfam05221 c	d00401	TIGR00936			$\left \right $		1		1	2	ιT
arCOG07444 H	Coenzyme transport and metabolism	MetK	S-adenosylmethionine synthetase	COG00192	pfam00438, pfam02772,	pfam02773	TIGR01034					1				

1	н		_	-	TIGR01941	cd00322		COG01018)r Flavodoxin reductase (ferredoxin-NAUPH reductase) татну 1	Ч	Energy production and conversion	ر	arcoguzzuu
4 -	•			-	101001011	100 CC CUP-	10000000000000000000000000000000000000			:	Freedom and contraction	, c	
1 1	1	1	_	1	TIGR01043	000 cd01134	pfam02874, pfam0(COG01155	tpA Archaeal/vacuolar-type H+-ATPase subunit A	R	Energy production and conversion	U -	arCOG00868
1 1	1	H		1	TIGR01041	000 cd01135	pfam02874, pfam0(COG01156	pB Archaeal/vacuolar-type H+-ATPase subunit B	ž	Energy production and conversion	U	arCOG00865
1 1	1		1	1		_	pfam01990	COG01436	tpG Archaeal/vacuolar-type H+-ATPase subunit F	ž	Energy production and conversion	U	arCOG04102
1 1	1	-	1	1		_	_	COG01527	tpC Archaeal/vacuolar-type H+-ATPase subunit C	Ĭ	Energy production and conversion	J	arCOG02459
1 1	1 1	1	1	1				COG00636	:pK Archaeal/vacuolar-type Na+/H+-ATPase, subunit K	At	Energy production and conversion	U	arCOG02455
1 1	1	1	1	1				COG01390	pE Archaeal/vacuolar-type H+-ATPase subunit E	Ъ	Energy production and conversion	U	arCOG00869
1 2	1	2 1	1	1		cd00284	pfam13631	COG01290	crB Cytochrome b subunit of the bc complex	ğ	Energy production and conversion	U	arCOG01721
1 2	1		1	1			pfam01496	COG01269	tpl Archaeal/vacuolar-type H+-ATPase subunit I	ž	Energy production and conversion	U	arCOG04138
1 4								COG00644	xC Dehydrogenase (flavoprotein)	Ē	Energy production and conversion	U	arCOG06720
1		2 1	1	1 2 1	5 TIGR01341	069 cd01586,cd01	ofam00330, pfam0(COG01048	inA Aconitase A	Ac	Energy production and conversion	0	arCOG01697
1 5	1	1	1	1 2 1	TIGR 00561	-	pfam12769		INAD(P) transhydrogenase, subunit alpha		Energy production and conversion	U	arCOG09401
2 3	1 2	1	1	1 1	TIGR01108	243 cd07937	pfam00682, pfam02	COG05016	adA1 Pyruvate/oxaloacetate carboxyltransferase	õ	Energy production and conversion	U	arCOG02095
1 3	1	1	1	1 1				COG00838	JoA NADH dehydrogenase subunit A	Ñ	Energy production and conversion	U	arCOG01557
1 3	1	1	1	1 1 1			pfam00146	COG01005	uoH NADH dehydrogenase subunit H	N	Energy production and conversion	U	arCOG01546
1 2	1	1	1	1 1				COG00839	uoJ NADH dehydrogenase subunit J	Ñ	Energy production and conversion	U	arCOG04654
1 2	1	1	1	1 1				COG00713	Jok NADH dehydrogenase subunit 4L (K,kappa)	Ñ	Energy production and conversion	U	arCOG03073
1 2	1	1	1	1 1	TIGR01770		pfa m00361	COG01007	Jon NADH dehydrogenase subunit N	Ñ	Energy production and conversion	U	arCOG01540
1 2	1	1	1	1 1	TIGR01974	361	pfam00662, pfam0(COG01009	Jol NADH dehydrogenase subunit L	Ñ	Energy production and conversion	U	arCOG01539
1 2	1	1	1	1 1	TIGR01972		pfam00361	COG01008	Jom NADH dehydrogenase subunit M	N	Energy production and conversion	U	arCOG01538
1 2	1 2		1	1 1	TIGR01971		pfam12838	COG01143	Joi NADH dehydrogenase subunit I	Ñ	Energy production and conversion	U	arCOG01543
1 3	1	1	1	1 1	TIGR01957		pfam01058	e COG00377	Job F420H2 dehydrogenase subunit, related to NADH:ubiquinone oxidoreductas	N	Energy production and conversion	U	arCOG01554
1 3	1	1	1	1 1	TIGR01961		pfam00329	COG00852	Joc NADH dehydrogenase subunit C	N	Energy production and conversion	U	arCOG01551
1 3	1	1	1	1 1	TIGR01962		pfam00346	COG00649	Job NADH dehydrogenase subunit D	N	Energy production and conversion	U	arCOG01548
					TIGR00212	390 cd13644	pfam01379, pfam03	COG00181	emC Porphobilinogen deaminase	Η	Coenzyme transport and metabolism	т	arCOG04299
2 7					TIGR00423		pfam04055	d COG01060	iii Thiamine biosynthesis enzyme ThiH, FO synthase or related uncharacterize.	부	Coenzyme transport and metabolism	т	arCOG00656
1					TIGR02304			COG01541	tak Coenzyme F390 synthetase	Pa	Coenzyme transport and metabolism	т	arCOG02624
1 4					TIGR00551	2910	pfam00890, pfam02	COG00029	adB Aspartate oxidase	Na	Coenzyme transport and metabolism	т	arCOG00572
1 2					TIGR01475	cd13959	pfam01040	COG00382	oiA 4-hvdroxybenzoate polyprenvitransferase or related prenvitransferase	J	Coenzyme transport and metabolism	т	arCOG00477
1					TIGR01934		pfam08241	COG02226	oiE Ubiquinone/menaguinone biosynthesis C-methylase UbiE	ŋ	Coenzyme transport and metabolism	т	arCOG04348
					TIGR03310	cd04182	pfam12804	ti COG02068	obA IGT-A family glycosyltransferase involved in molvbdopterin guanine dinucleo	ž	Coenzyme transport and metabolism	T	arCOG01873
-	7				TIGD00336	1 07 CdO1 7 04	0meha 29200meha	0000117	In Energy-coupling (actor) transporter transmention are protein curity bD Durimiding daaminase and radurtase		Coenzyme transport and metabolism		ar COGO1 4 85
	4 C				000	10000 0-1-1-1-10000	TCCTOILDIN	00000010	et Enorma complexity structures accords a suburity for the suburity of the former of the second s		Commune transport and metabolism		
	c					C000551 - 100	pram0.25/2	COC01077	UK AI P.COTRINOID AGENOSYITANSTERASE	19	Coenzyme transport and metabolism	т :	arc06046/8
	.				TIGR03964		pfam02633	ir COG01402	1B Creatinine amidohydrolase/Fe(II)-dependent formamide hydrolase involved	A	Coenzyme transport and metabolism	π:	arCOG04536
	1				TIGR00294		pfam02649	COG01469	ptA Fe(2+)-dependent GTP cyclohydrolase	Σ	Coenzyme transport and metabolism	т	arCOG04301
	1							COG00095	Lipoate-protein ligase A associated domain		Coenzyme transport and metabolism	т	arCOG03837
					TIGR00190		pfam01964	COG00422	iC Thiamine biosynthesis protein ThiC	4 H	Coenzyme transport and metabolism	т	arCOG02741
					TIGR00545		nfam03099	COGODO95	II innate-nrotein lipace A	-	Coenzyme transport and metabolism	: I I	arCOG01939
					TIGRUI510	172 cd01571	ptam03740 nfam0	00001/31	004 Arcnaear riporiavin syntriase or Niroctinic acid nhocnhocultrancfarace	and and	Coenzyme transport and metabolism		arc0601481
					TIGR00693	cd00564	ptam02581	C0G00352	NE Thiamine monophosphate synthase	-	Coenzyme transport and metabolism	I :	arCOG01089
	-1				TIGR01379	cd02194	pfam00586	COG00611	iL Thiamine monophosphate kinase	f	Coenzyme transport and metabolism	т	arCOG00638
	1				TIGR02151	cd02811	pfam01070	e COG01304	dD FMN-dependent dehydrogenase, includes L-lactate dehydrogenase and type	Π	Coenzyme transport and metabolism	т	arCOG00613
	1				TIGR00292		pfam01946	COG01635	14 Archaeal ribulose 1,5-bisphosphate synthetase/yeast thiazole synthase	Ŧ	Coenzyme transport and metabolism	т	arCOG00574
					TIGR00176	cd03116	pfam03205	COG01763	obB Molybdopterin-guanine dinucleotide biosynthesis protein	ž	Coenzyme transport and metabolism	т	arCOG00532
	1				TIGR 00638		pfam03459	COG03585	opl Molybdopterin-binding protein	ž	Coenzyme transport and metabolism	т	arCOG00228
				35	TIGR03128, TIGR019:	373 cd04726	pfam00215, pfam05	COG00684	enG Demethylmenaquinone methyltransferase	ž	Coenzyme transport and metabolism	т	arCOG00117
					TIGR00097	cd01169	nfam08543	COG00351	niD Hvdroxvmethvlovrimidine/nhosnhomethvlovrimidine kinase	1	Coenzyme transport and metabolism	: I I	arCOG00020
4 -					TIGR 01 3 2 7	rd05303	pfam00389	COG00111	rrd Dhoshodyrerate dehydrogenase or related dehydrogenase	a S	Coenzyme transport and metabolism	= 1	arCOG01754
7 7						10/0000 070	pfam01512	0000001	IIF DITUDE DUPPERTING ENZYTHE INVOIVED IN THOLYOUDDET IN AND UNDER UNDER DUDE		Coenzyme transport and metabolism	- 1	
, ,				-	TIG BOD 3 EC	C320012	pfam00000 afam0t	COC00476	erri I Protoporpriyrinogen oxidase vic Diaudootido utilizian onumo involuod in molubdootoxin and thiomino hiom	E f	Coenzyme transport and metabolism		arcoco1522
1					TIGR00063	cd00642	ptam01227	C0G00302	DIE GTP cyclohydrolase I	요 :	Coenzyme transport and metabolism	I :	arc0604542
					TIGR02370	231 cd02065	pfam02607, pfam02	COG05012	tbC1 Methanogenic corrinoid protein MtbC1	ž	Coenzyme transport and metabolism	т	arCOG03402
1			_	1	TIGR00223	cd06919	pfam02261	COG00853	InD Aspartate 1-decarboxylase	Pa	Coenzyme transport and metabolism	т	arCOG04813
1 2			1	1	TIGR02144		pfam08443	COG00189	mK Glutathione synthase/glutaminyl transferase/alpha-L-glutamate ligase	Rir	Coenzyme transport and metabolism	т	arCOG01589
1			1	1	TIGR01254	cd13545	pfam13343	COG04143	ppA ABC-type thiamine transport system, periplasmic component	đT	Coenzyme transport and metabolism	т	arCOG00226
1 2				1 2	TIGR00343	569 cd04727	pfam01680, pfam05	COG00214	VZ1 Pvridoxine biosvnthesis enzyme	SN	Coenzyme transport and metabolism	т	arCOG04075
				1			pfam10120	in COG01992	Predicted transcriptional regulator fused phosphomethylpyrimidine kinase.		Coenzyme transport and metabolism	T	arCOG00021
c	n		+	-			nfa m00288	COG01829	u i InAD(F)InFilaVIII i cuuciase Pantoate kinase	×.	Coenzyme transport and metabolism	с I Т	arCOG04263
4						133 (00088/	prarruce455, prarruu	COCO0E 4 2	0eA Ivioiybuopterin piosyntiesis enzyme	NA V	Coommon transport and metabolism		arcoco100
7	•					C0U25U3	pram12804	COG00700	00A INOIYDGODTETIN-guanine dinucleotide plosynthesis protein A	Ň	Coenzyme transport and metabolism	I :	arcoco212/2
1	त्न -				TIGR00581	cd01419	pfam01967	COG00315	oaC Molybdenum cofactor biosynthesis enzyme	ž	Coenzyme transport and metabolism	π:	arCOG01530
1 4	1		_		TIGR02668	546 cd01335	pfam04055, pfam06	COG02896	oaA Molybdenum cofactor biosynthesis enzyme	ž	Coenzyme transport and metabolism	т	arCOG00930
1 4	1		_			cd00756	pfam02391	COG00314	oaE Molybdopterin converting factor, large subunit	ž	Coenzyme transport and metabolism	т	arCOG00534
1 2	1				TIGR02667	cd00886	pfam00994	COG00521	oaB Molybdopterin biosynthesis enzyme	ž	Coenzyme transport and metabolism	т	arCOG00214
1 2	1 1			1	TIGR00521	1127	pfa m02441, pfam0 ^z	COG00452	pp pp pp pp pp pp pp pp pp pp pp pp pp	Df	Coenzyme transport and metabolism	т	arCOG01704
1	1 2	-1	1	96 1 1	TIGR01499, TIGR0149	287 cd00739	pfam08245, pfam02	COG00285	IC Folylpolyglutamate synthase and Dihydropteroate synthase	Fo	Coenzyme transport and metabolism	т	arCOG02817
1 2	1		1	1 1		cd02022		COG00237	a E Dephospho-CoA kinase	S	Coenzyme transport and metabolism	т	arCOG01045
1	•		-	1	TIGR03367	cd00470	pfam01242	COG00720	JeD 6-pvruvovl-tetrahvdropterin svnthase	ō	Coenzyme transport and metabolism	T	arCOG02172
-		-	•	 - -	TIGR00636	10.07.00	pfam01923	00602096	uno Coh(lhalamin ademocultransferace	Pd	Coenzyme transnort and metabolism	: 1	arCOG00489
4	•		-	 - -	TIGR 02 235	rd13962	nfam01040	COC01575	and 1.4-dihydrow-2-nanhthnate octanrenvitrancfera.ce	Ŵ	Contribution transport and metabolism	: 1	arCOG00480
7 T		2 C	-	- - -	OTCOMUDI	1377 nfam01023	pfam12413 nfam1/	00001330	um Lupuate synthase CTD-denendent Dihofisvin kinsee		Contribute transport and metabolism		
, n		7			TIGROUZ 14	10400F-	prarr03039	COC00110					arcococco
, v	-1			7 7 7	TIGKU2 / 48	COUDES	ptam00348	COC001142	DA Geranyigeranyi pyropnospnate synthase	ISP	Coenzyme transport and metabolism	т :	arc0601040
m u					TIGK00552	cdUU53	pramu 2540	COG00117	ade INH-3-dependent NAU+ synthetase	N	Coenzyme transport and metabolism	I 1	arc06001576
	, ,	1	1, 1	1	00000	011001	pfam05402	100000	IqD Coenzyme PQQ synthesis protein D	Pa	Coenzyme transport and metabolism	I :	arCOG03838
										ľ			

-	-				_	_					SC.UUITS	Cytochrome b tamily protein		C Energy production and conversion	arcouveuuz
ſ	4 e				T					-	r 00110	11yuudenaac 4 Incrimitere awameniyie		 Encourter and contention 	
					l					COT CT III III (TC CI III)	0000033	Hudrogense-4 membrane subunit Huff	Huff	C Energy production and conversion	arCOG03676
										fam04432 nfam13183		Coenzyme F420-reducing inyarogenase, dena subumit. Coenzyme E420-reducing hydrogenase, beta subumit (with C-terminal ferredolo	a Ha	C Effection production and conversion	arCOG02653
Ţ					+				TIGR03294	fam01058, pfam12838	COG01941	Coenzyme F420-reducing hydrogenase, gamma subunit	FrhG	C Energy production and conversion	arCOG02473
	۲.								TIGR03294	fam01058	COG01941	Coenzyme F420-reducing hydrogenase, gamma subunit	FrhG	C Energy production and conversion	arCOG02472
	1								TIGR03224	fam13237 cd07030	COG01146	Ferredoxin		C Energy production and conversion	arCOG02461
F					╞	╞			TIGR00314	fam03063 cd01916	COG01152	CO dehvdrogenase/acetyl-CoA synthase alpha subunit	CdhA	C Energy production and conversion	arCOG02428
Ī				+	+	╉		+	TIG R00332	fam00719 cd00412 fam01880 cd03172	COG00221	Inorganic pyrophosphatase	Ppa	C Energy production and conversion	arCOG01711
	, I					+				fam00708	COG01254	Acylphosphatase	AcyP	C Energy production and conversion	arCOG01674
	1					$\left \cdot \right $			20 TIGR03336	fam01855, pfam0277 cd07034, cd0	c0G04231	Indolepyruvate ferredoxin oxidoreductase, alpha and beta subunit		C Energy production and conversion	arCOG01609
F					╞	╞			TIGR03710	fam01855 cd07034	COG00674	Pruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidored	PorA	C Energy production and conversion	arcog01607
T					+	+			TIGR01961	Tam0U374 fam00379		LOEnzyme F42U-feaucing nyarogenase, aipna supunit Ni Fa-hudrozenase III component G	HVCF	 C Energy production and conversion C Energy production and conversion 	arCOG01552
				_	_	+			13 TIGR02026	fam02310, pfam0405 cd02068, cd0	COG01032	Radical SAM superfamily enzyme	-	C Energy production and conversion	arCOG01356
	1					$\left \right $				fam13549	COG01042	Acyl-CoA synthetase, ATP-grasp containing subunit		C Energy production and conversion	arCOG01338
 -					$\left \right $	┢		F		fam02915 cd01041	COG01592	Reterodistinae reacciase; suburin c	, inc	C Energy production and conversion	arCOG01097
T				+	+	+			TIGR02512	C010	COG01146	Ferredoxin	-	C Energy production and conversion	arCOG00959
	1									fam00390, pfam0394 cd05311	COG00281	Malic enzyme	SfcA	C Energy production and conversion	arCOG00853
										fam02730, pfam01314	COG02414	Aldehyde:ferredoxin oxidoreductase	-	C Energy production and conversion	arCOG00709
Ī						╀				tam13459 feam00753 nfam00359	COG01141	Ferredoxin Elsivori-hradovin	Per Nov/	C Energy production and conversion	arcOG00349
	.,					+			TIGR00273	fam02589, pfam13183	COG01139	L-lactate utilization protein LutB, contains ferredoxin domain	LutB	C Energy production and conversion	arCOG00335
	1								TIGR01311	ifa m00370, pfam0278 cd07786	COG00554	Glycerol kinase	GlpK	C Energy production and conversion	arCOG00024
									00400101		COG02811	Archaeal/vacuolar-type H+-ATPase subunit H	NtpF	C Energy production and conversion	arCOG03363
						┼			TIGR04013	fam04055 fam07754 nfam02754	COG01031	Radical SAM supertamily enzyme	- 1	C Energy production and conversion	arCOG01359
2				1	_	-	1		TIGR03555	fam00296 cd01097	COG02141	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin red		C Energy production and conversion	arCOG02410
				4		. m	2	i m		fam12682	COG00716	Flavodoxin	FIdA	C Energy production and conversion	arCOG00519
8	-			2	-		-	IGR 1 IGR 2	TIGR00387,T TIGR03378.T	fa m01565, pfam02913, pfam13183 fa m00890, nfam02910	COG00277 [FAD/FMN-containing dehydrogenase fused to Heterodisulfide reductase, subil Succinate dehydrogenase/fumarate reductase. flavonortein subunit	GlcD	C Energy production and conversion C Energy production and conversion	arCOG00340 arCOG00571
1 1	2			1	1	1	2	1		fam03009 cd08568	COG00584	Glycerophosphoryl diester phosphodiesterase	UgpQ	C Energy production and conversion	arCOG00701
1	1				-	╞			TIGR02177	fam02775, pfam1236 cd03375	COG01013	Prime/cupper-type cycoumonic/quinor oxuase, suburit 3 Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidored	PorB	C Energy production and conversion	ar cod 04 030 ar COG 01 599
,					•		,		TIGR00183	fam00180	COG00538	Isocitrate dehydrogenase	lcd	C Energy production and conversion	arCOG01164
1 2		,				1	1	IGR 1	pf TIGR 00384,T	ifam13085, pfam13183, pfam02754,	COG00479	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit	GIPC	C Energy production and conversion	arCOG00963
m 	-	m n				┼			TIGR01591	fam00384 [cd02766 fam13450 ofam13247	C0G00243	Anaerobic dehydrogenase	BISC	C Energy production and conversion	arcog01491
2 2		2		1	2	2	2	1	TIGR02824	ifa m08240, pfam0010 cd08264	COG00604	NADPH: quinone reductase or related Zn-dependent oxidoreductase	Qor	C Energy production and conversion	arCOG01458
					+	+			TIGR04041	ifam12724, pfam13187	COG01145	Equocimonie u bugenesis procein Flavodoxin fused to ferredoxin domain	NapF	C Energy production and conversion	ar COG 02 4 49
T								IGR04041	TIGR04105,T	fam12838, pfam12838	COG01145	Ferredoxin domain containing protein	NapF	C Energy production and conversion	arCOG02187
1 2	1	4 . .							100-001011	fam13534, pfam02754	c0G00247	Fe-S oxidoreductase	GIPC	C Energy production and conversion	arCOG00333
1 2	1						2		TIGR02476 TIGR02657	fam00881 cd02150 fam00127 cd13021	COG00778	Nitroreductase 0 Discrocuanin	NfnB PetF	C Energy production and conversion	arCOG00288 arCOG0288
		-1		2	1	1		1			COG01290	Cytochrome b subunit of the bc complex	QcrB	C Energy production and conversion	arCOG04594
4 TT	2	4 4	n ∞	2 V	m	0 4	t θ	n 00	TIGR02032	fam13450	COG00644	Latraueriyue veriyurugeriase, sutchirate sermalveriyue veriyurugeriase or utim Dehydrogenase (flavoprotein)	FixC	C Energy production and conversion	ar COG 00570
1			m L	,		,	-	c	001-100-01-	fam02754	COG00247	Membrane associated Fe-S oxidoreductase	GlpC	C Energy production and conversion	arCOG00332
1 1		2	e		1	1	1	IGR 1	03 TIGR02891,T	fam00115,pfam0051 cd01662,cd0	COG00843	Heme/copper-type cytochrome/quinol oxidase, subunit 1 and 3	CyoB	C Energy production and conversion	arCOG01237
1 2 7	1		7 7				7		TIGR01800	fam00285 cd06118	COG00372	Citrate synthase	GItA	C Energy production and conversion	ar COG 04 2 37
, ,			2 6	_		-	-1 -	-	TIGR02512	fam13237 c.do1714	COG01146	Ferredoxin	- Fiv.A	C Energy production and conversion	arCOG04548
1 1	1	त्त	2			1	1	2	TIGR03615	fam01613	COG01853	NADH-FMN oxidoreductase RutF, flavin reductase (DIM6/NTAB) family	RutF	C Energy production and conversion	arCOG02017
1			2 6			-1 -			TIGR 02 866	fam01012, pfam0076 cd01985 fam00116	COG02025	Electron transfer flavoprotein, alpha subunit	FixB	C Energy production and conversion	arCOG00447 arCOG01235
			t I						TIGR02060	fam13187,pfam12139	COG01146	Ferredoxin		C Energy production and conversion	arCOG02618
			1					IGR04105	TIGR01971,T	fam12838, pfam12838, pfam12838	COG01145	Polyferredoxin	NapF	C Energy production and conversion	arCOG02179
	7				+	+			TIGR02061	fam00890, pfam02910	COG01053	Succinate dehvdrogenase/fumarate reductase, flavoprotein subunit	SdhA	C Energy production and conversion	ar COG 00573
Ī	1 0								TIGP 01 060	fam01989 cd01356 cd01356 cd01356	COG01786	Swiveling domain associated with predicted aconitase		C Energy production and conversion	arCOG04279
			1			1	1		TIGR03419	fam01592 cd06664	COG00822	NifU homolog involved in Fe-S cluster formation	IscU	C Energy production and conversion	arCOG02077
7 T	2							7	TIGR01349	ram001982.pfam02852	COG01249	Pyruvate/ z-oxogiutarate denydrogenase complex, uniydrolipoamide acytrana Pyruvate/2-oxogiutarate dehydrogenase complex, dihydrolipoamide dehydrol	Lod	C Energy production and conversion C Energy production and conversion	arCOG01068
1 0 0	1			_		-	-	1 ,	TIGR01104	fam03030	COG03808	Na+ or H+-translocating membrane pyrophosphatase	OVP1	C Energy production and conversion	arCOG04949
1 1	1		1					1	TIGR 00309	fam01813	COG01394 I	Archaeal/vacuolar-type H+-ATPase subunit D	NtpD	C Energy production and conversion	arCOG04101
1 1				4 44	•		²		TIGR03710	fam01558, pfam0185 cd07034	COG00674	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidored	PorA	C Energy production and conversion	arCOG01606
1 2 2				~	-		2 6		TIGR01771 TIGR03102	fam00056, ptam0286 cd00300 fam00127	COG00039	Malate/ lactate denydrogenase	Mdh PetF	C Energy production and conversion	arC0G00246 arC0G0242
1			·			ц.	2	1	TIGR00224	fam01293 cd00484	COG01866	Phosphoenolpyruvate carboxykinase (ATP)	PckA	C Energy production and conversion	arCOG06073
2 5			1				1	1	TIGR03181	fam00676 cd02000	COG01071	Pyruvate/2-oxoglutarate/acetoin dehydrogenase complex, dehydrogenase (E	AcoA	C Energy production and conversion	arCOG01054
1 2	ľ				-				107001011		COG02181	r yravaret z oxog utari ate/ateronii uenyuugenase comprex, uenyur ogenase ter CoBCoM heterodisulfide reductase subunit E	HdrE	C Energy production and conversion	arCOG05014
1 -	FI.								TIGR01019	fam02629, pfam00549 fam02729, pfam02549	COG00074	Succinyl-CoA synthetase, alpha subunit Diviviate/2-2-2001/142-242-242-242-242-242-242-242-242-242-	SucD	C Energy production and conversion	arCOG01339
1 1	1		1		_	1	1	1	TIGR01016	fa m08442, pfam00549	COG00045	Succinyl-CoA synthetase, beta subunit	Succ	C Energy production and conversion	arCOG01337
1 1		2	1	1			1	1	TIGR00979	fa m00206, pfam1041 cd01596	COG00114	Fumarase	FumC	C Energy production and conversion	arCOG01749
	7					╞		IGR04041	TIGR02314,T	fam09383, pfam1283 cd07030	COG01145	Laccate denyarogenase or relateu z-nyaroxyacia denyarogenase Ferredoxin	NapF	C Energy production and conversion C Energy production and conversion	arcog02460
									TIGR01357	fam13685 cd08173	COG00371	Glycerol dehydrogenase or related enzyme	GIdA	C Energy production and conversion	arCOG00982

arCOG04391	C Ener	"gy production and conversion		Rubredoxin COG	01773 pfa	am00301 cd007	730								
arCOG04406	C Ener,	rgy production and conversion	FumA	Tartrate dehydratase beta subunit/Fumarate hydratase class I, C-terminal dor COGC	01838 pfa	am05683	TIC	5R00723		_					
arCOG04407	C Ener	rgy production and conversion	TtdA	Tartrate dehydratase alpha subunit/Fumarate hydratase class I, N-terminal ddCOG	01951 pf	am05681	Ĭ	5R00722							
arcog04475 arcog04475	C Ener	rey production and conversion	OadB	Nu,re-riyurugeriase maturi adoni ractori Na+-transporting methylmalonyi-CoA/oxaloacetate decarboxylase. beta subul COGC	01883 Dfa	am03977 Luuvu	TIC	3R01109							
arCOG04537	C Ener	gy production and conversion	NuoF	NADH:ubiquinone oxidoreductase, NADH-binding 51 kD subunit (chain F) COGC	1894 pfa	am01512, pfam10589	TIC	GR01959							
arCOG04874	C Ener	rgy production and conversion	AIID	Malate/lactate/ureidoglycolate dehydrogenase, LDH2 family COG	12055 pfi	am02615	TIC	GR03175							
arCOG04890	C Ener	rgy production and conversion	NuoE	NADH:ubiquinone oxidoreductase 24 kD subunit	1905 pf	am01257 cd136	537,cd030 TIC	GR01958							
arC0G05128	C Ener	rgy production and conversion	NapF	Ferredoxin Ea-S-cluster-containing hydrogenees commonent 2 COGO	01145 pti 01142 pti	am13746		5R02910							
arCOG05745	C Ener	rey production and conversion	- The	BFD-like [2Fe-25] binding domain COGC	00446 Dfa	am04324	I	GR01372							
arCOG05797	C Ener	*gy production and conversion	OadG	Na+-transporting methylmalonyl-CoA/oxaloacetate decarboxylase, gamma su COGC	3630 pf	am04277	TIC	GR01195							
arCOG05865	C Ener,	rgy production and conversion	PepCK	Phosphoenolpyruvate carboxykinase, GTP-dependent COGC	1274 pfa	am00821 cd008	319								
arCOG06124	C Ener	rgy production and conversion	ACH1	Acetyl-CoA hydrolase COG	0427 pf	am02550, pfam13336	TIC 110	GR03458							
ar COG 13546	C Ener	•ev production and conversion	- -	ACP (acvl carrier) superfamily protein	1007 D16	am06857	TIC	GR01608							
arCOG01605	C Ener	'gy production and conversion	PorD	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidored COGC	1144 pfa	am12838	TIC	5R02179				2	1	1	
arCOG01163	C Ener	rgy production and conversion	LeuB	Isocitrate/isopropyImalate dehydrogenase COG	0473 pfi	am00180	TIC	GR02088				2	1		
arCOG01340	C Ener	rgy production and conversion		Acyl-CoA synthetase (NDP forming)	1042 pf	am13380, pfam13607	Ĭ	5R02717							
arC0601547	C Ener	rgy production and conversion	- HvrF	Ni. Fe-bydrogenase III Jarge subunit and subunit G	11032 ptc 13261 ptc	am00374 nfam00346		5R01962							
arCOG01553	C Ener	*ev production and conversion	- 12	Ni.Fe-hydrogenase III small subunit	13260 Dfa	am01058	I	GR01957							
arCOG01601	C Ener	"gy production and conversion	PorB	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidored COGC	1013 pf	am02775 cd033	376 TIC	5R02176							
arCOG01602	C Ener	rgy production and conversion	PorG	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidored COG	1014 pfi	am01558	TIC	GR03334				2			
arCOG01603	C Ener	rgy production and conversion	PorG	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidored COG	1014 pf	am01558	Ĭ	3R02175							
arcoc01008	C Ener	rgy production and conversion	POLA	Pyruvate:rerredoxin oxidoreductase or related z-oxoacid:rerredoxin oxidored UUGI Haterodiculfida reductase subunit A or related polyferredovin	111.1.0 DT	m07995, ptam02 /8 cd0 /u	01 266 nf TIC	5KUZ1/b 5P02007 TICP02170 T	1011 21 13 100 DI	202170					
arCOG01545	C Ener	rgy production and conversion	HvfC	Formate hydrogen/vase subunit A or related polyterredoxin Formate hydrogen/vase subunit 4	00650 pre	am00146	11 10,002 1.00	1 (272) 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	116KU3143,116	KU21/2					
arCOG00706	C Ener	*gy production and conversion	-	Aldehyde:ferredoxin oxidoreductase COGC	2414 pfa	am02730, pfam01314						1			
arCOG01537	C Ener,	rgy production and conversion	NuoM	NADH:ubiquinone oxidoreductase subunit 4 (chain M) COGC	0651 pfa	am00361	TIC	GR01972							
arCOG00855	C Ener	rgy production and conversion	Pta	Phos photrans acetyla se COG	00280 pfa	am01515	TI	GR02706					-1	1	
arCOG01167	C Ener	rgy production and conversion	CoxL	Aerobic-type carbon monoxide dehydrogenase, large subunit CoxL/CutL hom4COGL	01529 pf	am01315, ptam02738	U L	5R02416							
arc0601925	C Ener	rgy production and conversion	Cox	Aerobic-type carbon monoxide denydrogenase, small subunit Loxs/Luts nom LUGI Aerobic-tune carbon monoxide debudrarenase middle cubunit CoxM/CutM MCOCO	1210 D1210	3 m1 3085, ptam01 /9 cd002	110	5RU3193 5 D03105	ł						
arCOG00962	C Ener	"gv production and conversion	FrdB	Succinate dehydrogenase/fumarate reductase. Fe-S protein subunit	0479 pf	am13085, pfam13183	I	GR00384						5	Ι
arCOG04595	C Ener	"By production and conversion	QcrA	Rieske Fe-S protein COG	0723		TIC	GR01416					1	2	
arCOG00456	C Ener,	rgy production and conversion	GpsA	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase COGC	0240 pfi	am01210, pfam07479	TIC	GR03376		_			1	m.	
arCOG01617	C Ener	rgy production and conversion	Tas	Aryl-alcohol dehydrogenase related enzyme COG(0667 pfa	am00248 cd066	560 TIC	GR01293					1	m	
arCOG04358	C Ener	rgy production and conversion	FdhD	Uncharacterized protein required for formate dehydrogenase activity COG	01526 ptc	am02634	TIC CF	5R00129						m	Τ
arCOG01502	C Eller	rey production and conversion	HvcB	Frenie/copper-cype cytocinonie/quinor oxidase, suorini z COOC Fe-S-cluster-containing hydrogenase component 2	1142 Die	am13247	747	GR07512.TIGR03149							
arCOG02304	C Ener	"gy production and conversion	Mct/CaiB	3 Succinyl-CoA:mesaconate CoA-transferase or predicted acyl-CoA transferase/ICOGC	01804 pf	am02515	Ĭ	GR03253							
arCOG02476	C Ener	'gy production and conversion	HdrA	Heterodisulfide reductase, subunit A; ferredoxin domain	1148 pfi	am07992, pfam13237, pfam	n13454,pf TIC	GR03140,TIGR04105,T	TIG R03385, TIG	R04105			1		
arCOG04522	C Ener,	rgy production and conversion	CcdA	Cytochrome c biogenesis protein COGC	0785								1		
arCOG02189	C Ener	rgy production and conversion	NapF	HTH containing ranscriptional regulator fused to ferredoxin domain COG	01145		TIC	GR02179							
arCOG02842	C Ener	rgy production and conversion	FdX	Ferredoxin Tricture V. transmet curters mombrane comments	00633 pfa	am00111 cd002	207	5R02008	,	7	Ŧ	Ţ	-		
arCOG01959	P Inore	ganic ion transport and metabolism	Trka	Trks-type n+ transport system, memorane component Trka K+ transport system. NAD-binding component	00569 pic	102354 nfam02080 nfam	002254.nfam	02080 I		 		- -	-1		
arCOG04147	P Inorg	ganic ion transport and metabolism	SodA	Superoxide dismutase COGC	0605 pfa	am00081, pfam02777				1					
arCOG04231	P Inor	ganic ion transport and metabolism	CutA	Uncharacterized protein involved in tolerance to divalent cations COGC	01324 pfa	am03091		1	1		1	1	1	3	
arCOG02021	P Inorg	ganic ion transport and metabolism	PspE	Rhodanese-related sulfurtransferase COGC	0607 pfa	am00581 cd001	L58 TIC	GR02981 1		1	1	2	5	9	
arC0G07775	P Inor	ganic ion transport and metabolism	HcaE	Phenylpropionate dioxygenase or related ring-hydroxylating dioxygenase, largCOG	04638 pf	am00355, pfam0084 cd035	535,cd088 TIC	5R03229 1	, н ,	1 1 1				4	
arc0602763		ganic ion transport and metabolism	-	COG(Content and Communication (HMA)	02608 pt	am00403 [cd003	871 TIG	5R02052 2		-		·		~ ~	
arCOG01477	P Inore	game for transport and metabolism	CZCD	Co/Zn/Cd efflux system component	01230 pf	am01545	I	GR01297	7			,	4	,	
arCOG00576	P Inorg	ganic ion transport and metabolism		Predicted divalent heavy-metal cations transporter COG	0428 pfa	am02535				1 1	2	1			
arCOG00238	P Inor	ganic ion transport and metabolism	ArsB	Na+/H+ antiporter NhaD or related arsenite permease COGC	1055 pfa	am03600 cd011	L17 TIG	GR00785 2	1	2 1 1	2	2 3	1	2	
arCOG02569	P Inor	ganic ion transport and metabolism	Eric	Chloride channel protein EriC	0038 pf	am00654, pfam0057 cd004	100,cd045 TIC	GR01302 1	2	1 1 1 1	2				
ar COG 09746	P Inore	game for transport and metabolism	-	Sulfotransferase related protein		am00685	1	1	2		7	e			
arCOG04750	P Inor	ganic ion transport and metabolism		Sirohydrochlorin iron chelatase fused to [2Fe-2S] Ferredoxin COG	12138 pf	a m01903, pfam0190 cd034	116,cd03414,	,cd02980 1		1 1 1		1	1	. 1	
arCOG02499	P Inor	ganic ion transport and metabolism	,	Lipoprotein NosD family, contains CASH domains	03420 pf	am13229, pfam13229	TIC	GR04247 1							
arCOG04205		ganic ion transport and metabolism	Am+B	Mig2+ and Co2+ transporter CO60	70000 Public Pub	3m01544 C0128	278 TIC	a KUU383 a boneae							
arCOG02267	P Inorg	ganicion transport and metabolism	PitA	Phosphate/sulphate permease COGC	0306 pfa	am01384	í.	000000						1	
arCOG02640	P Inor	ganic ion transport and metabolism		Uncharacterized protein YkaA, distantly related to PhoU, UPF0111/DUF47 fan COGC	1392							1			
arCOG02764	P Inor	ganic ion transport and metabolism	CopZ	Copper-ion-binding protein	12608 pfi	am00403 cd003	371 TIG	GR02052				н	-1		
arC06014/4	P Inor	ganic ion transport and metabolism	DerF	Predicted Co/Zn/Cd cation transporter I Incharacterized motein involved in intracellular sulfur reduction	11553 pt	101545 m02635	Ĭ	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	-	-		H			
arCOG00173	P Inore	game for transport and metabolism	PhnE	ABC-type phosphate/phosphonate transport system, permease component COGC	13639 Dfa	am00528 cd062	261 TIC	GR01097 1	-	1			-	e.	
arCOG00206	P Inorg	ganic ion transport and metabolism	PhnC	ABC-type phosphate/phosphonate transport system, ATPase component COGC	13638 pfa	am00005 cd032	256 TIC	GR02315 1		1				m	
arCOG01805	P Inor	ganic ion transport and metabolism	PhnD	ABC-type phosphate/phosphonate transport system, periplasmic component COG(13221 pfa	am12974 cd010	071 TIG	GR01098 1		1 1			1	4	
arCOG04233	P Inor	ganic ion transport and metabolism	FepB	ABC-type Fe3+-hydroxamate transport system, periplasmic component COG	0614 pf	am01497 cd011	143 TIC	5R04281 1		1		·		-	
arCOG01006	P Inore	game for transport and metabolism	Znuc	ABC-type Mn1/zn transport system, zer ase component ABC-type Mn2+/Zn2+ transport system, permease component COGC	01108 pf	am00950 cd065	550 TIC	GR03770 1							
arCOG01005	P Inore	ganic ion transport and metabolism	Lral	ABC-type metal ion transport system, periplasmic component/surface adhesinCOGC	0803 pfa	am01297 cd010	018 TIG	5R03772,TIGR 1		1			1		
arCOG00163	P Inor	ganic ion transport and metabolism	ThiP	ABC-type Fe3+ transport system, permease component COG	01178 pf	a m00528, pfam0052 cd062	261,cd062 TIC	GR03262 1		1			1		
arCOG03127	P Inon	ganic ion transport and metabolism	Kch	Ion channel fused to ion transporting domain Earradovin cubunit of nitrite reductase or ring-budrovulating diovuganses COGO	01226		167								
arc0G08934	P Inorg	ganicion transport and metabolism	-	Alkylhydroperoxidase family enzyme Coord	12128 pf	am02627))	+ FI					$\left \right $		Π
arCOG04487	P Inor	ganic ion transport and metabolism	KatG	Catalase (peroxidase I) COG	0376 pfi	am00141, pfam0014 cd006	549,cd006 TIC	5R00198	1				-		

arCOG02497	P Inorganic ion transport and metabolism		Lipoprotein NosD family, contains CASH domains	COG03420	pfam12708		TIGR04247		1			1	
arCOG01578	P Inorganic ion transport and metabolism	MgtA	Cation transport ATPase	COG00474	pfa m00690, pfam00122	?,pfam00702,pf	TIGR01647					1 1	1
arCOG02050	P Inorganic ion transport and metabolism		Sulfite exporter, TauE/SafE family	COG00730	pfam01925			_				1 1	2
arCOG00202	P Inorganic ion transport and metabolism	EcfA2	Energy-coupling factor transporter ATP-binding protein EcfA2	COG01122	pfam0005	cd03225	TIGR01166					1	
arcoco1040	P Inorganic ion transport and metabolism	CCC1	AdenylyIsuitate kinase of related kinase Dendisted Ecol. (Anol. + ***********************************	CUGUU529	pram01583		11GKU0455					-1 -	
arcocon1090	P Inforgante ion transport and metabolism D Increase from transport and metabolism	LUCAN LUCAN	Predicted Fe2+/Nin2+ (raitsporter, VII 1/OCOLI family ABC-type molyhdate transport system periolosmic component	COG01814	pramu 25 21), pramu 260	-4125A0	11G P/13 7 30						-
arcocona Ea	P Inforgante ion transport and metabolism D Increase from transport and metabolism	Feore	AbC-type morybuate transport system, periplasmic component Ea2±transnort system protain B	-0600370	pramo3331	-d01879	TIGR03/30						
	D Increasifie for transport and metabolism	Mate 2	Dermasse similar to ration transmorter	-0601834	nfam01760	CIOTODO	1000001					4 -	
arCOG01069	P Inorganic ion transport and metabolism	-	CoA-demendent NAD(PIH Sulfur Oxidoreductase	COG01824	nfam07997.nfam02852		TIGR03385					4 -	
arCOG01095	P Inorganic ion transport and metabolism	Etn	Ferritin	20601528	pfam00210	cd01055							
arCOG01475	P Inorganic ion transport and metabolism	MMT1	Predicted Co/Zn/Cd cation transporter	COG00053	pfam01545, pfam0257 c	cd00851	TIGR01297					1	
arCOG01953	P Inorganic ion transport and metabolism	KefB	Kef-type K+ transport system, membrane component	COG00475	pfam00999		TIGR00932					1	
arCOG01954	P Inorganic ion transport and metabolism	KefB	Kef-type K+ transport system, membrane component	COG00475	pfam00999		TIGR00932					1	
arCOG01957	P Inorganic ion transport and metabolism	TrkA	TrkA, K+ transport system, NAD-binding component	COG00569	pfam02254, pfam02080							1	
arCOG02102	P Inorganic ion transport and metabolism	FeoA	Fe2+ transport system protein A	COG01918	pfam04023							1	_
arCOG02248	P Inorganic ion transport and metabolism	CbiM	ABC-type Co2+ transport system, permease component	COG00310	pfam01891		TIGR00123					1	
arCOG03077	P Inorganic ion transport and metabolism	MnhB	Multisubunit Na+/H+ antiporter, MnhB subunit	COG02111								1	
arCOG03159	P Inorganic ion transport and metabolism	CbiM	ABC-type Co2+ transport system, permease component	COG00310	pfam13190			_				1	
arCOG04191	P Inorganic ion transport and metabolism	MET3	ATP sulfurylase (sulfate adenylyltransferase)	COG02046	pfam14306, pfam0174 c	cd00517	TIGR00339					1	
arCOG04355	P Inorganic ion transport and metabolism	TehA	Tellurite resistance protein or related permease	COG01275	pfam03595 c	cd09299	TIG R00816					1	
arCOG05758	P Inorganic ion transport and metabolism		TrkA-C domain contaning protein	COG00490	pfam02080							1	
arCOG05908	P Inorganic ion transport and metabolism		Ferritin-like domain	COG01633	pfam04454							1	
arCOG 10427	P Inorganic ion transport and metabolism		Chromate resistance protein	COG04275	pfam09828							1	
arCOG 10942	P Inorganic ion transport and metabolism		Sulfotransferase family protein									1	
arCOG01103	P Inorganic ion transport and metabolism		Ferritin-like domain	COG01633	pfam02915	cd01045						2	
arCOG01958	P Inorganic ion transport and metabolism	Kch	Kef-type K+ transport system, predicted NAD-binding component	COG01226	pfam07885, pfam02254							2	
arCOG01960	P Inorganic ion transport and metabolism	TrkA	K+ transport system, NAD-binding component fused to lon channel	COG01226	pfam02254, pfam02080	0,pfam02254,pfi	am02080					2	
arCOG01961	P Inorganic ion transport and metabolism	NhaP	NhaP-type Na+/H+ and K+/H+ antiporter	COG00025	pfam00999		TIGR 00831					2	
arCOG02190	P Inorganic ion transport and metabolism	ACR3	Arsenite efflux pump ACR3 or related permease	COG00798	pfam01758		TIGR 00832					2	
arCOG03072	P Inorganic ion transport and metabolism	MnhC	Multisubunit Na+/H+ antiporter, MnhC subunit	COG01006	pfam00420		TIGR 00941					2	
arCOG03078	P Inorganic ion transport and metabolism	,	Predicted subunit of the Multisubunit Na+/H+ antiporter	COG01563	pfam13244							2	
arCOG03079	P Inorganic ion transport and metabolism	MnhB	Multisubunit Na+/H+ antiporter, MnhB subunit	COG02111	pfam04039							2	
arCOG03082	P Inorganic ion transport and metabolism	MnhG	Multisubunit Na+/H+ antiporter, MnhG subunit	COG01320	pfam03334		TIGR01300					2	
arCOG03099	P Inorganic ion transport and metabolism	MnhE	Multisubunit Na+/H+ antiporter, MnhE subunit	COG01863	pfam01899		TIGR 00942					2	
arCOG03121	P Inorganic ion transport and metabolism	MnhF	Multisubunit Na+/H+ antiporter, MnhF subunit	COG02212	pfam04066							2	
arCOG00318	P Inorganic ion transport and metabolism	PhoU	Phosphate uptake regulator	COG00704	pfam04014, pfam01895	5,pfam01895	TIGR02135					1	1
arCOG01576	P Inorganic ion transport and metabolism	ZntA	Cation transport ATPase	C0G02217	pfam00403, pfam0012 t	cd00371,cd014	TIG R00003, TI GR0:	1511				1	1
arCOG02062	P Inorganic ion transport and metabolism	TusA	TusA-related sulfurtransferase	COG00425	pfam01206	cd00291						1	1
arCOG02287	P Inorganic ion transport and metabolism		Putative selenium binding protein, beta/alpha-propeller fold	COG00393	pfam01906							1	1
arCOG05356	P Inorganic ion transport and metabolism	Fes	Enterochelin esterase or related enzyme	COG02382	pfam00756							1	1
arCOG11907	P Inorganic ion transport and metabolism	,	Fe-S metabolizm associated domain		pfam02657		TIGR03391					1	1
arCOG00230	P Inorganic ion transport and metabolism		Periplasmic molybdate-binding protein/domain	COG01910	pfam00126, pfam12727	-	TIGR 00637					1	
arCOG02849	P Inorganic ion transport and metabolism	ArsA	Oxyanion-translocating ATPase	COG00003	pfam02374 (cd02035	TIGR00345					1	
arCOG04559	P Inorganic ion transport and metabolism	EmrE	Membrane transporter of cations and cationic drugs	COG02076	pfam00893							1	
arCOG01964	P Inorganic ion transport and metabolism	Kch	Kef-type K+ transport system, predicted NAD-binding component	COG01226	pfam07885							2	
arCOG06837	P Inorganic ion transport and metabolism		Lipoprotein NosD family, contains CASH domains	COG03420	pfam05048		TIGR04247, TIGR0:	3024				2	
arCOG00167	P Inorganic ion transport and metabolism	PstC	ABC-type phosphate transport system, permease component	COG00573	-	cd06261	TIGR02138						1
arCOG00168	P Inorganic ion transport and metabolism	PstA	ABC-type phosphate transport system, permease component	COG00581	_	cd06261	TIGR00974						1
arCOG00213	P Inorganic ion transport and metabolism	PstS	ABC-type phosphate transport system, periplasmic component	COG00226	pfam12849 (cd13565	TIGR00975						1
arCOG00231	P Inorganic ion transport and metabolism	PstB	ABC-type phosphate transport system, ATPase component	COG01117	pfam00005 (cd03260	TIGR00972						1
arCOG00232	P Inorganic ion transport and metabolism	PhoU	Phosphate uptake regulator	COG00704			TIGR02135						1
arCOG02555	P Inorganic ion transport and metabolism		Lipoprotein NosD family, contains CASH domains	COG03420									1
arCOG02852	P Inorganic ion transport and metabolism	{NirD}	Ferredoxin subunit of nitrite reductase or ring-hydroxylating dioxygenase	COG02146	pfam00355 c	cd03528	TIGR02377						1
arCOG03797	P Inorganic ion transport and metabolism		Ferritin-like domain	COG01633	pfam05763								1
arCOG06728	P Inorganic ion transport and metabolism		Sulfite exporter, TauE/SafE family	COG00730	pfam01925								2
arCOG01709	I Lipid transport and metabolism	CaiA	Acyl-CoA dehydrogenase	COG01960	pfa m02771, pfam0277 c	cd01158	TIGR03207	1 1	1		1		
arCOG04213	I Lipid transport and metabolism	IN01	Myo-inositol-1-phosphate synthase	COG01260	pfam07994		TIGR03450	1 2	1	1	1		
arC0G06112	I Lipid transport and metabolism	ACS	Acyl-coenzyme A synthetase/AMP-(fatty) acid ligase	C0G00365	ptam00501	cd05943	TIGR01217	1 2					,
arC0G02245	I Lipid transport and metabolism		Cytidylyltransferase family enzyme	C0G01836	ptam01940		TIGR00297	-	1	1		, i	2
arc0604106	I Lipid transport and metabolism	LdsA	CUP-digiyceride synthetase	CUGUU5/5	pramu1864	100 111		,	,		-1		
arc0601532	Inpld transport and metabolism	Upps	Undecaprenyi pyrophosphate synthase	C06001020	pram01255	cdUU475	TIG R00055		 -		-1 •		7 (
arc0600340	 Lipid transport and metabolism 	LCB5	Diacyigiycerol Kinase ramily enzyme 2 hidrowiacii CoA dobiidiococacao como fundato Encul CoA biidicatoo	COG01287	TQ/0000190 TQ/0000190	- 40.65 E0	TIGR03/02		- c	-	-	-	7 4
arc0612010	I lipid transport and metabolism	raub	5-riyaroxyacyi-CoA genyarogenase, some rusea to enoyi-CoA nyaratase	CORDIZED	prarruz/3/, prarruu/z	5-10C074	110KU243/,110K	7 7	2 5		7	7	4
arc0613950	I lipid transport and metabolism	-	Priospiratugyigiyeeropriospiratase A	-001000	prarriu4006	T/60002	710-0010-71	 	 	-	-1 -	•	,
	 Lipid transport and metabolism 	-	Isopentenyi prospriate Kinase, enzyme or modineu mevalonate patriway	COG01677	prarrituto90	C004241	TIG BODE 40			-		7 -	7 0
	 Lipid transport and motobolism 	TTOTT			00700111111	I	6+COOND11	+ +		-			n
24000010E40	 Lipid transport and motobolism 		Short shair alsohol dahudisasasas	000000000000000000000000000000000000000	00100muga	-40E307-	TIC D01 7 00	+ +	+			+	
arCOG01085	 Lipid transport and metabolism 	PrrR	751-3-0-crian accuro uciryu ogonace 751-3-0-aeranviakrenvi nhocnhate swithace TIM-harrel fold	_0601646	nfam01884	-d03812	TIGR01768		4	-	4 -		7
arCOG01251	I lipid transport and metabolism		137-3-0-851 attyBet attyBrycet yr Juospitale Synthase, hiwr Dair ei 100 Dradiotad mambrana accordatad llinid hydrolaca nautral caramidaea cunaefam	-0603356	nfam00012	7107007		+ -		4 -	4 -	4	Ŧ
	I Lipid transport and metabolism	- Code	דו בעוגרפע הוופוווטו מודב מסטטנומרט ווטוע הואָטו טומסב, וופענו מו עבו מוווטמסב טעצבוומוו 2- איטראטיס-טו-רידיא למאיטלאממיטינים	COG01250	pfam03727 nfam00735		TIG D() 3 70		-			T	u
arCOG01250	I Lipid transport and metabolism	HMG1	3-LiyuroXyatyi-COA ueriyurogeriase HudrowwathuddiutanaLOAA raductase	COG01257	pfattio2/2/pfattio0/25	-4006/13	TIG DODE 22			-		- 1	0 0
arCOG02936	I lipid transport and metabolism	FRG9	Phytophe/solialayr.com/reduceac	COG01562	nfam00494	-d00683	TIGR03465	- -	4 -		4 -	4	4 17
arCOG01710	I Lipid transport and metabolism	Sbm	Methylmalonyl-CoA mutase. C-terminal domain/subunit (cobalamin-binding)	COG02185	pfam02310 c	cd02071	TIGR00640					1	m
arCOG01590	I Lipid transport and metabolism	AccC	Biotin carboxvlase	COG00439	pfam00289.pfam02786	. ofam02785	TIGR00514	1	1	1	1	- 2	2
arCOG02705	I Lipid transport and metabolism		Acetyl-CoA carboxylase, carboxyltransferase component	COG04799	pfam01039		TIGR01117	2 2		1	2 1	1 4	7
arCOG04232	I Lipid transport and metabolism	Sbm	Methylmalonyl-CoA mutase	COG01884	pfam01642 c	cd03680	TIGR00641	2 2	2		2 1	1 2	2
arCOG01259	I Lipid transport and metabolism	FabG	Short-chain alcohol dehydrogenase	COG01028		cd05233	TIGR01830	6 6	4 1	1	2 2	5	7
arCOG01081	I Lipid transport and metabolism	idi	Isopentenyldiphosphate isomerase	COG01443	pfa m00293 (cd02885	TIGR02150	1 1	1 1	1	2		-
arCOG00773	I Lipid transport and metabolism	<u> </u>	AcvI-CoA hvdrolase	COG01607	pfam03061, pfam0306 t	cd03442,cd0344	42	1	1		2		m

arCOG00670	-	Lipid transport and metabolism	PgsA	Phos phatid viel v cer conhos phate svnt hase	COG00558	pfa m01066		-	2	2 2	1	2		1	2	2
arcocooo		inid transport and metabolism	Cico	Enoul-CoA hudratace/carnithing racemace	COG01034	nfam00278 rd06558	TIG DU3 210			, t		- -				17
ar COC 01 707		tiple support and metabolism	vi v		0000000	ofo.m01771 ofo.m0177 cd006.67	TIC DO 201	, <	•	4 C	-	n n			n <	11
		tipid transport and metabolism	Cd M	Acyr-coA deniyar ugenase Loon nhain and CoA muthatana (AMD familian)	000000000000000000000000000000000000000	pierioz / 1, pierioz / / cuouso/		t +	0 +	0 4		-			t +	11
arc0604133	-	Lipid transport and metabolism	LAAL	Long-criain acyi-coa synthetase (AiviP-forming)	COG0001		TIGR00054	-,	-	7 , T	-				- ,	7
arcugut 13/	-	Lipid transport and metabolism	Lauvi		CUGUU824		TENNNALI	-		-			-		-	-
arCOG04100	-	Lipid transport and metabolism	CaiA	Acyl-CoA dehydrogenase	COG01960											
arCOG01529	-	Lipid transport and metabolism	Acs	Acyl-coenzyme A synthetase/AMP-(fatty) acid ligase	COG00365	pfam00501 cd05966	TIGR 02 188	1	1			1	2			
arCOG01282	-	inid transport and metabolism	Paal	Acetyl-CoA acetyltransferase	COG00183	nfam00108. nfam0280 cd00751	TIGR01930	~	~	1			~			ں د
arCOG01767	-	tipid transport and metabolism	DheG	2-budrow-2-methylalutanyl CoA swithace	COC03475		TIGD 007 AB	-			-	T		÷		, ,
ar COCO1 2 70		tipid transport and metabolism	0.02	Janyai Oxyaamaaniyigada yi cox synaase Acetal CoA acetalteranferance	0000100			+ +		 -	-				n .	4
arcubulz/8	-	Lipid transport and metabolism	Pad.	Acetyl-CoA acetylitaristerase	COCCUSCO	C000029	DEST0451	-,	-1 (-, , -, ,	,			-	-	
arCOG04761	-	Lipid transport and metabolism	UppP	Undecaprenyl pyrophosphate phosphatase	COG01968	ptam02673	TIGR00753	-	2	1	-1	-				
arCOG01879	-	Lipid transport and metabolism		Dolichol kinase family protein	COG00170			1		1					1	
arCOG01880	-	Lipid transport and metabolism	SEC59	Dolichol kinase	COG00170			1						2	_	
arCOG00856	_ _	Lipid transport and metabolism	CaiC	Acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II	COG00318	pfam00501 cd12119	TIGR01923	1								
arCOG09737	-	Lipid transport and metabolism	LicD	Phosphorylcholine metabolism protein LicD	COG03475	pfam04991		1								
arCOG04470	-	Lipid transport and metabolism	Psd	Phosphatidylserine decarboxylase	COG00688	ofa m0 2666	TIG R00164	2	1	1	1				1	
arCOG00671	-	inid transport and metabolism	PscA	Phosnhatidylserine synthase	COG01183	nfam01066	TIGR04217	0	-	-						
arCOG00242	-	inid transport and metabolism	CaiD	Enovl-CoA hvdratase/carnithine raremase	COG01024	nfam00378 rd06558	TIGR02280		-							
arCOC01.6E0		tipid transport and metabolism		Linophorchalitana alaba kata kudralana runarfamilu	1200000	oform13607	TICBOSCOE		-					-		+
		Lipia transport and metabolism	Plab	Lysopriosprioripase, alpha-peta nyurolase superramity	COC00534	1601UP1	110KU3093									-
arcocorate		Lipid transport and metabolism	Pgp4	Memorane-associated phospholipid phosphatase	COG00611	pramU1569 cd03392	0.0000									
arC0G07155	-	Lipid transport and metabolism	IspD	4-diphosphocytidyl-2-methyl-D-erithritol synthase	C0G01211	ptam01128 cd02516	TIGR00453							-1		
arCOG11863	-	Lipid transport and metabolism		Membrane diacylglycerol kinase	COG00818	pfam01219 cd14263								-1		
arCOG02228	-	Lipid transport and metabolism	GtrA	GtrA-like flippase	COG02246	pfam04138								2	_	
arCOG02039	-	Lipid transport and metabolism	cls	Phosphatidylserine/ phosphatidylglycerophosphate/cardiolipin synthase or n	COG01502	pfa m1 309 1, pfa m1 309 cd07 493, cd0	9127,cd09128							3	1	
arCOG00241	_ _	Lipid transport and metabolism	CaiD	Enoyl-CoA hydratase/carnithine racemase	COG01024	pfam00378 cd06558	TIGR 02 2 80								1	1
arCOG01261	-	Lipid transport and metabolism	FabG	Short-chain alcohol dehydrogenase	COG01028	pfam00106 cd05233	TIGR01830		-						2	1
arCOG03951	- -	Lipid transport and metabolism	PgpB	Membrane-associated phospholipid phosphatase	COG00671	pfam01569 cd03386		$\left \right $							2	9
arCOG00674	-	Lipid transport and metabolism	PgsA	Phosphatidylglycerophosphate synthase	COG00558	pfam01066										1
arCOG06122		Lipid transport and metabolism	Acs	AcvI-coenzyme A synthetase/AMP-(fatty) acid ligase	COG00365	pfam00501 cd05959	TIGR02262									1
arCOG01487	L	Nucleotide transport and metabolism	ComEB	Deoxvortidvlate deaminase	COG02131	pfam00383 cd01286	TIGR02571	-		1					Ţ	1
arCOG00067	ц	Nucleotide transport and metabolism	PrcA	Phosnhorihosvinvronhosnhate svnthetase	COGO0462	nfam1 3793 nfam0015 rd062 23	TIGR01251	-	-	-		-			-	-
arCOC0314	- 4	Nucleotide transport and metabolism	Thur	Thumidulate contents	0000001	520002 CT00100 CT00252		+ +		+				4	4 -	
arc0603214		Nucleotide transport and metabolism	H H	Inymidylate synthase	0000010	prarmuusus cauusus	TIGR 00257	-,	-1 (_, ,	- •			- ,	4
arc0602824	<u>د</u> ا	Nucleotide transport and metabolism	Hurh	ALCAK Transformylase/IIMP cyclonydrolase Purh	COGUD138	pramu 2 14 2, pramu 180 cdu 14 21	119K0U355	-	7	 		- ·			-	7
arCOG00689	۷ ۲	Nucleotide transport and metabolism	AIB	Dihydroorotase or related cyclic amidohydrolase	COG00044	pfam13147 cd01318	TIGR00857	1	2	1	1	1	-1	1	1	2
arCOG04276	F	Nucleotide transport and metabolism	NrdA	Ribonucleotide reductase, alpha subunit	COG00209	pfa m03477, pfam0031 cd02888	TIG R02504	1	2	2 1	1	1 1	1	2	1	3
arCOG01039	2	Nucleotide transport and metabolism	AdkA	Archaeal adenylate kinase	COG02019			1	2			1 1	1	-1		
arCOG04184	4	Nucleotide transport and metabolism	RdgB	Inosine/xanthosine triphosphate pvrophosphatase. all-alpha NTP-PPase fami	COG00127	ofam01725 cd00515	TIG R00042	1		1		1	-	1	1	4
arCOG00028	. u	Nucleotide transport and metabolism	-0	Orotate nhoshhorihosvitransferase homolog	COGODREG	nfam00156 cd06223	TIGR02985 TIGR	-	T							
arcOc0050	. u	Nucleotide transport and metabolism	DurtM	Dheraharihaculaminaimidazala (AID) cunt hat aca	00000150	DiamOnt B6 InfamO 76 Cd0 106	TIG DOD 979	4 0	·	+ +	+	-	4 e		4 -	• -
		Nucleotide transport and metabolism	Pur IN	Crotate abovibovibovibovibovitraneforace	DETODOOD	2007 2007 2007 2007 2007 2007 2007 2007		7	4 4	+ +	-					
ar cooooco			L'ALC		T0+00000					+					-	-
arc060092		Nucleotide transport and metabolism	25flA	Cytosine deaminase or related metal-dependent nyurolase	COCO0117	GOSTODO 676TOLUPIO	106202901			-				7 +	Ŧ	
arc0604747		Nucleotide transport and metabolism	Ē		CO 600337	Cautor 2 2000 Cautor 2	00000012		-			- ·		-	- ,	4
arCOG01747	<u>د</u> د ۱	Nucleotide transport and metabolism	PurB	Adenylosuccinate lyase	COG00015	ptam00206, ptam1039 cd01360	TIGR00928	•	-	I		- ,	н ,			7
arc0600102	<u>د</u> ا	Nucleotide transport and metabolism	PurL	Phosphoribosyltormyiglycinamidine (FGAM) synthase, glutamine amidotran:	COG00047	ptam13507 cd01740	TIG R01737						2			2
arCOG04313	<u>د</u>	Nucleotide transport and metabolism	Ndk	Nucleoside diphosphate kinase	COG00105	pfam00334 cd04413		-1	-1			1	2	-1	Ţ	
arCOG00063	<u>ح</u>	Nucleotide transport and metabolism	PyrG	CTP synthase (UTP-ammonia lyase)	COG00504	pfam06418, pfam0011 cd03113, cd0	017 TIG R00337	1	1	1 1	1	1 1		1	1	m
arCOG04462	L L	Nucleotide transport and metabolism	PurS	Phosphoribosylformylglycinamidine (FGAM) synthase, PurS component	COG01828	pfa m02700	TIGR00302	1	1	1 1	1	1		1	1	2
arCOG04421	L L	Nucleotide transport and metabolism	PurC	Phos phoribosylaminoimidazole succinocar boxamide (SAICAR) synthase	COG00152	pfam01259 cd01415	TIG R00081	1	1	1 1	1	1		1	1	
arCOG00641	2	Nucleotide transport and metabolism	PurL	Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain	COG00046	pfa m00586, pfam0276 cd02203, cd0	12.2 TIGR01.736	1	1	1 1	1	-1		1		2
arCOG01565	∠ ۳	Nucleotide transport and metabolism	NrnA	nanoRNase/ pAp phosphatase, hydrolyzes c-di-AMP and oligoRNAs	COG00618	pfam01368		1	1	1	1	1		1	1	1
arCOG00911	2 س	Nucleotide transport and metabolism	PvrB	Aspartate carbamovitransferase, catalytic chain	COG00540	pfam02729, pfam00185	TIGR00670		2	1		1		1	1	2
arCOG01034	4	Nucleotide transport and metabolism	THEP1	Nucleoside-triphosphatase THEP1	COG01618	pfam03266 cd00009		1	2	1		1		1	1	2
arCOG04229		Nucleotide transport and metabolism	Pur	Asnartate carbamovitransferase, regulatory subunit	COG01781	nfam01948. nfam0.2.748	TIG R00240	-	~	-		-				2
arCOGOAOAR	. u	Nucleotide transport and metabolism		Dervertidine deaminase	COG00717	nfam00602 rd07557	TIG R02 2 74	-						-		
arCOG04387	. u	Nicleotide transport and metabolism	Durd	Adenviorinate svnthase	COG00104	nfam00709 rd03108	TIGR00184	•		4 -		-				• ~
arCOG0AA15	. u	Nucleotide transport and metabolism	Durb	Dhornhowlawina-structure linesa Dhornhowlawina-dhurina linesa	COG00151	premo 2844 nfam01071 nfam02843	TIG POD 877	-	+	+ +		-			4 -	1 -
arCOCO1020		Nucleotide transport and metabolism	Fui U	Principitul Principitum - Bigase	TETODOO	pidiliozo44, pidiliozo1. 1, pidiliozo4.3				+ +						
arCOG0018		Nucleotide transport and metabolism	Nnr7	NAD/PIH-bydrate renair enzyme Nnr NAD/PIH-bydrate debydratase domain	COGOD63	pram12230 pram03853 pram0125 rd01171	TIGROD197 TIGRO	10196	4 -	- -					4	4
arCOGOORS	. u	Nicleotide transport and metabolism	2 Internation	GMP conthace Dp_ATPace domain/cubinit	COG00519	nfam02540 nfam005 rd01997	TIGROD884	007700				-				1
arc0600090	. u	Nucleotide transport and metabolism	GuaA	GMP synthase - Glutamine amidotransferase domain	COG00518	nfam00117	TIGROD888			-		+ -			,	• ~
arCOG04346	. u	Nucleotide transport and metabolism		5-formaminoimida zola-4-carboxamide-1-beta-D-rihofi ranosvi 5'-mononhosi	COG01759	nfam06849 nfam06973	TIG ROD8 77					-			۰ د	,
arCOG00603		Nucleotide transport and metabolism	PvrD	Dihvdronrotate dehvdrogenase	COG00167	nfam01180 cd04740	TIGR01037	-	-	-	-	1	-	-		4
arCOG01037	. u	Nucleotide transport and metabolism	a k	Cutidulate kinase	COG01102	nfam13180 cd02020	TIG R02 1 73	-								
arCOG00087	. u	Nicleotide transport and metabolism	Guad	GMP conthace - Glintamine amidotransferase domain	COGOT 18	nfam00117 cd01742	TIGRODR88		1 -	1		1 0				• -
arCOG01891		Nucleotide transport and metabolism	Tmk	Thymidylate kinase	COG00125	pfam02223 cd01672	TIGR00041	- ~		•		1 2	4			2
arCOG00093		Nucleotide transport and metabolism	PurF	Glutamine phosphoribosylpyrophosphate amidotransferase	COG00034	pfa m003 10. pfa m00 15 cd007 15. cd0	062 TIGR01134	m	. m			1	2			2
arCOG02825		Nucleotide transport and metabolism	PurN	Folate-dependent phosphoribosvigivcinamide formvitransferase PurN	COG00299			m	2	1		3	2	1	2	m
arCOG02464	2 س	Nucleotide transport and metabolism	PurE	Phosphoribosylcarboxvaminoimidazole (NCAIR) mutase	COG00041	pfam00731	TIGR01162		1	1				1	1	1
arCOG00612		Nucleotide transport and metabolism	GuaB	IIMP dehvdrogenase/GMP reductase	COG00516	pfam00478 cd00381	TIG R01302	2	2	2					1	m
arCOG02807		Nucleotide transport and metabolism	AzeA	Xanthine/uracil/vitamin C permease. AzeA family	COG02252	pfam00860				1						2
arcocorco;	. u	Nucleotide transport and metabolism	1.Qm.	Dradicted corrected endonuclease distantly related to archaeal Holliday inort-	COG04741	premocococo de m10107		ſ	-	4			4 e	4		,
arCOGODIR		Nucleotide transport and metabolism	Ant	Freducted sect etcd endoructesse disantity related to archidear indillary Junic Adanine/granine nhosphorihosultransferase or related DRPD-binding protein	COG04741	premitoto/ nfam00156 rd06223	TIGR01090	-		-			• •	-		4
		Nucleotide transport and metabolism	Dive	Oratidiae-El-ahorahate decerboxidatselase of related river-binding protein Oratidiae-El-ahorahate decerboxidaee	000000	Main0015 cd04735	DEDTOVIDII		4 6	•			,	4 -		t c
arC0603575		Nucleotide transport and metabolism	- A	Dolymhachate kinase 2	COG02376	pramove 13	TIG R03 708		4	-		+		4	4	4
arCOGOD858	. u	Nicleotide transport and metabolism	DurH	1 Uridvlate kinase	COC02520	nfam0.696 rd0.0253	TIGR02076	-				,		-		1
arCOG04541		Nucleotide transport and metabolism	MIS1	Enuryidee kiildase Formvitetrahvdrofolate svinthetase	COG0759	nfam01268 cd00477		t								- -
arCOG01350		Nucleotide transport and metabolism	-	Predicted inorganic polyphosphate/ATP-NAD kinase	COG03199	nfam01513	ļ	ļ	T			F		• ==	• =	4 m
arCOG02013		Nucleotide transport and metabolism	DeoA	Thymidine phosphorylase	C0G00213	pfam01568.pfam02885,pfam00591	.nf TIGR03327	f	t			F				,
arCOG04311	<u>г</u>	Nucleotide transport and metabolism		5'-deoxynucleotidase, HD superfamily hydrolase	COG01896	pfam13023		F	t			F				
arCOG01075	<u>ح</u>	Nucleotide transport and metabolism	Ļ	NUDIX family hydrolase	COG01051	pfam00293 cd04673									4	∞
arCOG04320	L L	Nucleotide transport and metabolism	DeoC	Deoxyribose-phosphate aldolase	COG00274	pfam01791 cd00959	TIGR00126									
]	Induction a subjet a second subjet	1	CONTRACT Providence and and					-						-	

arCOG01324	u.	Nucleotide transport and metabolism Udp	Uridine phosphorylase	0G02820 pf	am01048	TIG	R01718				-			1		
arCOG01723		Nucleotide transport and metabolism CyaB	Adenylate cyclase, class 2 (thermophilic)	0G01437 pf	am01928 cd07890	0	R00318	+		T	+					
arcogota 173		Nucleotide transport and metabolism Cdd	Triginidate syntrase Ovtidine deaminase	0600295 http://	am00383 cd01283	TIO	R01354			t						
arCOG04298		Nucleotide transport and metabolism	Adenosine/AMP kinase	OG01839 pf	am04008					t				. 4		
arCOG04309	u.	Nucleotide transport and metabolism	S-adenosyl-I-methionine hydroxide adenosyltransferase	OG01912 pf	a m01887									1		
arCOG04889	Ľ	Nucleotide transport and metabolism NrdD	Oxygen-sensitive ribonucleoside-triphosphate reductase	OG01328 pf	am13597 cd01675	5	R02487	_						1		
arCOG05133	щ	Nucleotide transport and metabolism Udk	Uridine kinase	0G00572 pf	a m00485 cd02026	6 TIG	R00235	+						1		
arCOG01327		Nucleotide transport and metabolism Pnp	Purine nucleoside phosphorylase	OG00005 pf	a m01048	Ĕ	R01694	-			+			5		2
arCOG00695		Nucleotide transport and metabolism SsnA	Cytosine deaminase or related metal-dependent hydrolase	0G00402 pt	am01979 cd01298	9 F	R03314	+		t				2	7 7	7 +
arc0603658	- l	Nucleotide transport and metabolism NrdP	Kiponucleotide reductase, beta subunit (rerritin domain)		amuuzb8 cduiu49	5	K041/1	+								
arcog 146	- 4	Nucleotide transport and metabolism AdK	Adenylate kinase or related kinase Draudouridina swithare	UGUU563 DF	amuu4ub cdu14.28	2	R01351	+		t	+					7 6
arCOG01566		Nucleotide transport and metabolism	nanoRNase/bAp phosphatase. hvdrolvzes c-di-AMP and oligoRNAs	OG00618 Df	a m02254. pfam01368. pfam0;	2272	66000	+		t	$\left \right $					4
arCOG06976	σ	Secondary metabolites biosynthesis, transport an HmgA	Homogentisate 1,2-dioxygenase	OG03508				1			1					
arCOG00696	σ	Secondary metabolites biosynthesis, transport an Hutl	Imidazolonepropionase or related amidohydrolase	0G01228 pf	am13147 cd01296	6 TIG	R01224	1		1	1	1		2	1	2
arCOG00235	ď	Secondary metabolites biosynthesis, transport an MhpD	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (c	0G00179 pf	am01557		R02303	2	•	, ,		- •		1	-1	1
arcOG04347	a	Secondary metabolites biosynthesis, transport and- Cerondary metabolites biosynthesis transport and -	SAM-dependent methyltransferase Arcmastic ring-onaning diovunanase catalutic Ling cubunit galated ensume	0G00500 pt	am08241 cd02440	0	R01934	3	-	-	7					
arc0604340	o c	Secondary metabolites biosynchesis, transport and - Secondary metabolites hiosynthesis, transport and -	Alonia uc mig-opennig uroxygenase, catary uc mgo subumit related enzyme	0600500 nf	am13489 cd02440		R02081			t						
arCOG06106	ď	Secondary metabolites biosynthesis, transport an-	Predicted ring-cleavage extradiol dioxygenase	0G02514 pf	am12681, pfam1268 cd07255	5,cd072 TIG	R03211					•				
arCOG03570	σ	Secondary metabolites biosynthesis, transport and -	SAM-dependent methyltransferase	OG00500 pf	a m12847 cd02440	0 TIG	R02021	1 1	1	1	1					1
arCOG01791	σ	Secondary metabolites biosynthesis, transport ani-	SAM-dependent methyltransferase	OG00500 pf	am12847 cd02440	0 TIG	R02021	1		1	1			1		
arCOG01792	a	Secondary metabolites biosynthesis, transport and	SAM-dependent methyltransferase	0G00500 pf	am08241 cd02440		R01934				-				•	ſ
arcogn777	a c	Secondary metabolites biosynthesis, transport and - Cerondary metabolites biosynthesis transport and Baal	SAM-dependent metnyitransrerase HGG motif-containing thioactarase possibly implyed in aromatic compounds		am08241 cd02440	0	RU2U/2		+	+						7 6
arCOG01521		Secondary metabolites biosynchesis, transport and radi	Protoene debydrogenase or related enzyme	0601233 hf	am13450 Luuo440	2 0	COCUUN	7 C	-	-	+					4 6
arCOG03914	ď	Secondary metabolites biosynthesis, transport an Sufl	Multicopper oxidase	OG02132 pf	am07732 cd11024	4 TIG	R02376			T	-					,
arCOG01523	σ	Secondary metabolites biosynthesis, transport and -	Phytoene dehydrogenase or related enzyme	OG01233 pf	a m01593	TIG	R03467	1	1		1				1	1
arCOG01778	σ	Secondary metabolites biosynthesis, transport ani-	SAM-dependent methyltransferase	OG00500 pf	a m08241 cd02440	0 TIG	R02072	1	1	_	1 1					
arCOG01782	σ	Secondary metabolites biosynthesis, transport and-	SAM-dependent methyltransferase	OG00500 pf	a m08241 cd02440	0 TIG	R01934		-1							
arCOG05015	ď	Secondary metabolites biosynthesis, transport and -	SAM-dependent methyltransferase	OG00500 pf	am12847 cd02440	0	R03534	_		+				-		1
arCOG01229	ď	Secondary metabolites biosynthesis, transport and	Cyclic 2, 3-diphosphoglycerate synthetase	0602403	or or o	out root o		+		┥	ł			, 1		
arCOG01400	a	Secondary metabolites biosynthesis, transport and-	SAM-dependent methyltransferase	0000500 pt	am05050 cd0244(0,cd021 TIG	R01444									
arcogo1723	30	Secondary metabolites piosynthesis, transport and - Cocordary metabolites hiserusthesis, transport and	Fromatic ring-opening droxygenase, Ligb subunit		amu2900 cd0/954	7	101034	+		╎	t					
arCOG02703		Secondary metabolites biosynchesis, transport and - Secondary metabolites biosynthesis, transport and -	SAM-dependent methyltransferase SAM-dependent methyltransferase		am08241 rd0244		R02072									
arCOG03688	70	Secondary metabolites biosynthesis. transport and osmC	Organic hydroperoxide reductase	0G01764 pf	am02566	DI DI	R03561				F					
arCOG01943	ď	Secondary metabolites biosynthesis, transport and PncA	Amidase related to nicotinamidase	OG01335 pf	am00857 cd00431	1 TIG	R03614							2	-	
arCOG04786	σ	Secondary metabolites biosynthesis, transport and -	1,2-phenylacetyl-CoA epoxidase, catalytic subunit	OG03396 pf	a m05138	TIG	R02158								1	3
arCOG06169	a	Secondary metabolites biosynthesis, transport an -	Arylsulfotransferase family protein	pt	a m05935			+								
arCOG01402	۵	Secondary metabolites biosynthesis, transport an AglP	SAM-dependent methyltransferase	0G00500 pf	a m05050	<u>T</u>	R01444	_			_				2	2
POORLY CHARAC	TERIZE	ED														
arCOG02177	S	Function unknown	Uncharacterized membrane protein	OG01967 pf	am01889		_	1 1	1	1	1 1	1	1		1	2
arCOG04565	S	Function unknown	Predicted membrane protein, DUF368 family	OG02035 pf	a m04018			1	1	1		1				
arCOG05517	S	Function unknown	Uncharacterized protein			0010			, ,	, - ,		- •		1		.,
arcOGO5368	~ v	Function unknown	Uncharacterized memorane protein, contains prinz (pacterial pieckstrin norm) Thicharacterized protein	C 00138	a musivus, pramusivus, juamu:	3 / 13			-		-					
arCOG04521	n v	Function unknown	Uncharacterized membrane protein. predicted permease	C.00496					t		-	• ••	• ••	1	-	5
arCOG04596	S	Function unknown	Uncharacterized membrane protein					1	1		1	1				
arCOG03678	S	Function unknown	Uncharacterized protein	C.00067				1	1		_	1		1	-1	9
arCOG04308	S	Function unknown	Uncharacterized protein	OG01698 pf	a m03685			1		1				1	-1	1
arCOG03729	s	Function unknown	Metal-binding cluster containing protein					, H	•	t	- -			-		
arCOG11014	~ v	Function unknown	Ununaracterized protein HTH-domain containing transcriptional regulator	C 0033.4 hf	am063.7.4					t	7			-	-	-
arcog 1224	n v	Function unknown	Uncharacterized protein	OG01849 Df	am04010				-	-				-		
arCOG02142	s	Function unknown	Pheromone shutdown protein TraB, contains GTxH motif	OG01916 pf	am01963	DI I	R00261	1 2		1				1		. 4
arCOG03124	S	Function unknown	Pentapeptide repeats containing protein	OG01357 pf	a m00805, pfam00805			1 2	1	1		1			2	e
arCOG03633	S	Function unknown	Uncharacterized membrane protein YckC, RDD family	0G01714 pf	am06271			1		1				1		,
arcog07412	~ ~	Function unknown	Uncharacterized internolane procent, our 2009 rammy Uncharacterized protein	h	6447 T 111 P			7			-			-		-
arCOG05495	s	Function unknown	Uncharacterized protein, contains N-terminal coiled-coil domain	OG04911 pf	am09969					1						
arCOG04076	S	Function unknown	Uncharacterized protein, DUF359 family	OG01909 pf	am04019			1				1		1	1	e
arCOG07813	s	Function unknown	LamG-like jellyroll fold domain	C.00184 pf	am13385			·		+	+				-1	
arc0604370	~	Function unknown	Uncharacterized membrane protein	0G03503 pt	am0//86			-	•					Ŧ	-	
arCOG04412 arCOG08731	~ v	Function unknown	Uncharacterized protein Derinlasmir nrotein with immunoglohin-like fold	0604004	am0.093.7	T				-	+					
arCOG05330	n v	Function unknown	Uncharacterized protein	2	700000		ľ			t	$\left \right $				•	
arCOG02546	s	Function unknown	Secreted protein, contains PKD repeats and vWA domain	OG03291 pf	a m05048, pfam0504 cd00146	6,cd001 TIG	R04247,TIGR0424	1 JL/21		T						2
arCOG03606	S	Function unknown	Uncharacterized protein	C.00401						1		1				
arCOG01917	S	Function unknown	Zn-ribbon domain containing protein		00000		Ī	+				- •		m	•	
arCOG02998	s u	Function unknown	Cupin domain containing protein	0G01917 pt	am07883										-	2
ar COG 02 452 ar COG 02 761	~ ~	Function unknown	Secreted protein, with PAD repeat domain Uncharacterized protein. DUF302 family	0603439 Df	am03625 cd14797	2										
arCOG04321	s	Function unknown	Uncharacterized protein DUF711 family, similar to ribonucleotide reductase a	OG02848 pf	am05167 cd08025					T						
arCOG05852	S	Function unknown	HEAT repeats containing protein	OG01413								1				
arCOG01159	S	Function unknown	Uncharacterized archaeal colled-coll protein	0G01340		Ē	R02168	2 2	7	, ,	1	2 4	r	2	2	5
arcouts14	~ v	Function unknown	Uncharacterized memorane ancnored protein with extractinuar ווא אישיאייאין איז	C:002/2		+		T F			+	7 C	7			
druuuuu a	2	FUNCTION UNKNOWN		-	-	-		-	-1	-		4			-	

arCOG05351	s	Function unknown	- Uncharacterized membrane protein		pfam11433		2				2		1			_
arCOG01907	S	Function unknown	AIM24 Uncharacterized protein, AIM24 family	COG02013	pfam01987 TIGR(00266 3	m	m	e	m	m	m -	1		1	-
arC0G03232	n u	Function unknown	- VanZ like family protein	COG05652	ptam04892		-		-				-1			
arCOG02087	n v	Function unknown	- Predicted membrane protein	COG01470	pfam10633.pfam13620.pfam10633		7		1	-					2	-
arCOG04214	s	Function unknown	- Uncharacterized protein	COG00432	pfam01894 TIGR0	00149 1		-					1	1		
arCOG 10338	S	Function unknown	- Uncharacterized protein	SC.00914	pfam13517, pfam13517, pfam13517, pfam13	517	2	1		1		-1	1			_
arCOG05338	s	Function unknown	- Uncharacterized protein					1		1		-1		1	2	_
arCOG02081	s	Function unknown	- Predicted membrane protein	COG01470	pfa m10633				1					2	2	_
arCOG02206	s	Function unknown	- Uncharacterized membrane protein	00001476			1	1		1					10	
arcOG04350	~ v	Function unknown	- Uncharacterized membrane protein - Drotein pradicted to be involved in DNA renair	COG01602	pramUbb95 hfam04894 hfam04895								7 6		r	
ar COG 02 565	n v	Function unknown	- Fritten, predicted to be involved in privariepan - Secreted uncharacterized protein	COG05276	prarint-403-45 prarint-403.3								7			-
arCOG06429	s	Function unknown	- Uncharacterized protein, contains NRDE domain	COG03332	pfa m05742											
arCOG02508	s	Function unknown	 Secreted protein, with PKD repeat domain 	COG03291	pfa m00801, pfam0080 cd00146,cd001 TIGR0	00864 1	1	2		1		m		1		_
arCOG02884	S	Function unknown	Membrane associated protein with extracellular Ig-like domain, a component	COG04743				-1	, ,	,			2		ç	
arcOG104618 arcOG10444	~ v	Function unknown	- Uncharacterized protein, UUF223/ Tamiiy - Ilincharacterized nrotein	CUGU3651	pramusese										7	
ar COG 12677	n v	Function unknown	- Uncharacterized protein				-		-	1					7	-
arCOG04693	s	Function unknown	- Uncharacterized protein				1								1	
arCOG03888	s	Function unknown	- Uncharacterized membrane protein			1	1									_
arCOG03949	s	Function unknown	- Uncharacterized membrane protein	COG03815	pfa m09858	-	2		1	1			1		2	_
arc061/1680	~ v	Function unknown	- Uncharacterized membrane protein Uncharacterized membrane protein DUEA113 family	CUGU2832	pram04304 Meanut 3430		7 6	-	-	-					2	
arCOG04364	n v	Function unknown	- Uncharacterized protein	COG01772	pfam04407	. [4 m			1 2						
arCOG02491	S	Function unknown	- WD40 repeats containing protein	COG02319	pfam13360	1		1	1	2				1		-
arCOG01119	s	Function unknown	 GYD domain, alpha/beta barrel superfamily 	COG04274	pfa m08734	1		-1	1							_
arCOG08355	ŝ	Function unknown	- Uncharacterized membrane protein					-1	ļ	1						_
arcoconoce	~ u	Function unknown	- Uncharacterized protein LEAT remote contribute motorin	C0001412					-	7						
ar COG 02 3 49	n v	Function unknown	- Incert repeats containing procein	COG01413	nfam0.7884 cd1.2918											
arCOG06533	n v	Function unknown	- Uncharacterized membrane protein	0000	100/01/01/01/01/01/01/01/01/01/01/01/01/	. [- 2						
arCOG05022	s	Function unknown	- Uncharacterized protein, contains PQ loop repeat	COG04095										1	1	
arCOG11882	S	Function unknown	- Uncharacterized membrane protein			1									2	_
arCOG05839	s	Function unknown	- Uncharacterized protein	V	TIGRO	04292 2	1	2		2 2						_
arCOG08211	s	Function unknown	- Uncharacterized membrane protein	SC.00448			1	1		1						_
arCOG08977	s	Function unknown	- Uncharacterized membrane protein	COG04270										-1	2	
arc06050505	~ u	Function unknown	- Uncharacterized protein	50.00862				+					Ŧ		ſ	
arCOG00930	n v	Function unknown	- Unuted exterized protein - HEAT reneats containing motein	COG01413	nfa m13646		4 0						-	-	4 6	
arCOG04579	s	Function unknown	- Uncharacterized protein, DUF2071 family	COG03361	pfam09844			1		1					2	
arCOG06742	s	Function unknown	- NosD-like cell surface protein					1		1						-
arCOG08643	s	Function unknown	 Secreted protein with C-terminal PEFG domain 	Ś				-1								_
arCOG02527	S	Function unknown	 Secreted protein, with PKD repeat domain 	COG03291					1	_			2			_
arCOG04662	ŝ	Function unknown	- Uncharacterized membrane protein	SC.00128		Ì							, ₁	,		_
arcoco4545	~ u	Function unknown	- Uncharacterized protein I Incharacterized protein with In Illo domain				+								Ŧ	
arCOG02170	n v	Function unknown	- Uncharacterized protein with ignite domain - Uncharacterized protein	SC.00300	ofam13559									7		-
arCOG09426	s	Function unknown	- Uncharacterized membrane protein										-1		2	
arCOG01302	S	Function unknown	- Uncharacterized protein	COG01531	pfa m04457								1			_
arCOG01336	s	Function unknown	AMMECR1 Uncharacterized protein	COG02078	pfam01871 TIGR0	04335							1			_
arCOG01472	s	Function unknown	 Hemerythrin HHE cation binding domain containing protein 	COG02461	pfa m04282, pfam01814, pfam 13596								-1			_
arCOG01921	s	Function unknown	- Uncharacterized protein, DUF61 family	COG02083	pfam01886	04100										
arCOG02148 arCOG02150	s u	Function unknown	 Uncharacterized homolog of gamma-carboxymuconolactone decarboxylase su 	COG00599	ptam02627 TIGR(04169	+	+								
ar COG 02 1.23	n v	Function unknown	- Uncharacterized membrane protein	COG04069	ptd://0//36 nfam/04087 TIGR0	00341										
arCOG02566	n v	Function unknown	- Uncharacterized membrane protein	0000000		Theorem										-
arCOG02717	s	Function unknown	 Predicted membrane protein, DUF131 family 	COG02034	pfam01998								r.			
arCOG02991	S	Function unknown	 Uncharacterized protein, DUF1850 family 	COG04729	pfa m08905								1			_
arCOG03107	s	Function unknown	 Uncharacterized protein, YigZ/IMPACT family 	COG01739	pfam01205, pfam09186 TIGR0	00257							1			_
arCOG03128	ŝ	Function unknown	 Pentapeptide repeats containing protein 	COG01357	pfam13599, pfam13599		+	+					., ,			-
arc0603350	~ u	Function unknown	- Uncharacterized membrane protein				+	+								
arc0603379	<i>^ u</i>	Function unknown	- Uncharacterized protein													
arCOG03573	n v	Function unknown	- Uncharacterized protein, DUF1015 family	COG04198	pfam06245											
arCOG03677	S	Function unknown	- Uncharacterized protein	SC.00067									1			_
arCOG03768	s	Function unknown	- Uncharacterized membrane protein, DUF973 family	SC.00418	pfa m06157								-1			_
arc0603776	~ u	Function unknown	- Ln Inger protein Uncharatoriad mambana antaia	76000.75			+									
arCOG04051	n v	Function unknown	- Uncharacterized protein	COG02412	pfam04242											
arCOG04079	s	Function unknown	- Uncharacterized protein										1			-
arCOG04132	S	Function unknown	- Uncharacterized protein	COG04697	pfa m0991.0		H						1			_
arCOG04140	ŝ	Function unknown	- Uncharacterized protein	COG01888	pfam02680	Ì							, ₁			_
arC0G04253	n u	Function unknown	- Uncharacterized protein Developments in actual hinding motoria	COG01415	ptam05559											
arCOG04373	n v	Function unknown	- Prredicted metal-binding protein - Illincharactarized incotein VideV TIDE0045 (DLIE72 familiv	COG00011	premiuuosu nefa.m01910	00106										
arCOG04390	n v	Function unknown	- Uncharacterized protein rugs/, Or 10045/00177 (amily - Uncharacterized protein containing a Zn-ribbon	COG04068	premotato pfem09889	00100										
arCOG04424	S	Function unknown	- Uncharacterized protein	COG03377	pfa m08827								1			-
arCOG04484	S	Function unknown	 Uncharacterized membrane protein, DUF1648 family 	COG05658	pfam07853,pfam13630								1			_
arCOG04555	s	Function unknown	 Uncharacterized membrane protein related to bactofilin 		pfa m04519, pfam04519								ب ا .			_
arCOG04705	S	Function unknown	- Uncharacterized protein	COG02098	pfa m04036,pfam04038					_			1			-

arCOG04938	S Function unknown -	Uncharacterized membrane protein	SC:00516		1
arCOG05238	S Function unknown	Uncharacterized protein			1
arc0605308	5 Function unknown	Uncharacterized membrane protein	Sc.UU13/		
arCOG05323	S Function unknown	Uncharacterized protein Uncharacterized membrane protein			+ +
arc0605352	S Function unknown	Unumaterized incritionarie process Unucharacterized protein			
arcocousts	C Emotion unknown	Uncharacterized protein Uncharacterized protein			H
	5 Fullution unknown	Uncharacterized protein Uncharacterized protein			
arcococasa	C Emotion unknown	Unchara charized protein Unchara charizad Zn-finger protain	COG01336		H
arc0605631	S Function unknown	Unumaterized stranger procent			
arCOG05717	S Function unknown -	Uncharacterized protein	SC:00530		1
arCOG05739	S Function unknown -	Uncharacterized protein with conserved CXXC pairs, DUF1667 family	COG03862 pfam07892 COG03862 pfam07892		1
arCOG05759	S Function unknown -	Uncharacterized protein	pfam12646		1
arCOG05783	S Function unknown	Uncharacterized protein			1
arCOG05800	S Function unknown -	Uncharacterized membrane protein			1
arCOG05803	S Function unknown	Uncharacterized protein			1
arCOG06053	S Function unknown	Uncharacterized protein			1
arCOG06113	S Function unknown	Zn tinger protein, C2C2 type	SC:00488		
arcococeca	5 Function unknown -	Pleckstrin nomology domain containing proteins	Sc. UU612 pram1447U, pramU9851		
arcOG06602	S Full-turk TOWT	Ulturial acterized protein Hincharaztarizad protain			
arcOG06939	S Function unknown	Whore-like protein family	nfam08889		
arCOG07169	S Function unknown	Uncharacterized protein			
arCOG07353	S Function unknown -	Zn-ribbon protein			1
arCOG07356	S Function unknown -	Uncharacterized protein			1
arCOG07411	S Function unknown	Uncharacterized membrane protein	SC:00711 SC:00711		1
arCOG08157	S Function unknown	Uncharacterized membrane protein			1
arCOG08946	S Function unknown -	Uncharacterized protein			1
arCOG09581	S Function unknown -	Uncharacterized membrane protein, DUF1700 family	COG04709 pfam08006		1
arCOG09752	S Function unknown	Uncharacterized membrane protein, DUF2068	C0G03305		г -
arCOG10088	S Function unknown	Uncharacterized protein	C0605504		
arC0G10150	S Function unknown	Uncharacterized protein	SC.00215		
arc061055	5 Function unknown	Uncharacterized protein			
arC0G106/1	S Function unknown	Uncharacterized protein			
arc0610940	5 Function unknown	Uncharacterized memorane protein	St. UU942		
arcoc11014	5 FUNCTION UNKNOWN -	Uncharatecterized protein Haddametecticad mombrane stratein			
arCOC13401	5 FUILUIUI UIIKIUWII 5 Eurotion unknown	Ununaraterized mombrane protein			
arcoc13491	5 FUNCTION UNKNOWN -	Uncharatecterized memorane protein			
arCOG13492	S Function unknown	Unumaracterized protein Il Incharacterized membrane protein			4 -
arCOG13494	S Function unknown -	Uncharacterized protein			
arCOG13495	S Function unknown -	Uncharacterized membrane protein			
arCOG13496	S Function unknown -	Uncharacterized membrane protein			-
arCOG13498	S Function unknown -	Uncharacterized membrane protein			1
arCOG13499	S Function unknown -	Uncharacterized protein, DUF4125 family	pfam13526		1
arCOG13500	S Function unknown -	Uncharacterized protein			1
arCOG 13502	S Function unknown -	Uncharacterized membrane protein			1
arCOG 13503	S Function unknown	Uncharacterized protein			1
arCOG 13507	S Function unknown	Uncharacterized membrane protein			
arCOG13508	5 Function unknown	Uncharacterized membrane protein			
arc0613509	5 Function unknown	Uncnaracterized protein			
arCOG12515	S Full-turk TOWT	Ulturial acterized protein Hinchararterized protein			
arCOG13516	S Function unknown	Uncharacterized protein			4 -
arCOG13517	S Function unknown	Unumeratorized protein Unucherarterized membrane protein			4 -
arCOG13518	S Function unknown	Uncharacterized nrotein			
arCOG13520	S Function unknown -	Uncharacterized protein	pfa m02677		1
arCOG13521	S Function unknown -	Uncharacterized membrane protein	SC.01004		1
arCOG13523	S Function unknown -	Uncharacterized protein			1
arCOG13524	S Function unknown -	Uncharacterized protein	SC.00667		1
arC0G13525	S Function unknown	Uncharacterized protein			
arCOG12527	S FUNCTION UNKNOWN	Uncharacterized membrane protein Hincharacterized membrane protein			
arc0613528	S Function unknown	Unuted actentized membrane protein			
arCOG13529	S Function unknown -	Uncharacterized membrane protein			
arCOG13530	S Function unknown -	Uncharacterized membrane protein			-
arCOG13533	S Function unknown -	Uncharacterized membrane protein			т т
arCOG13534	S Function unknown -	Uncharacterized membrane protein			1
arCOG13535	S Function unknown -	Uncharacterized protein			1
arCOG13539	S Function unknown -	Uncharacterized protein			1
arCOG13540	S Function unknown	Uncharacterized membrane protein			1
arc0613541	S FUNCTION UNKNOWN -	Uncnaracterized protein Uncharacterized membrane protein			
arc0613543	S Function unknown	Unumaterized membrane protein			
arCOG13544	S Function unknown -	Uncharacterized protein			
arCOG13545	S Function unknown -	Uncharacterized protein			1
arCOG13547	S Function unknown -	Uncharacterized protein			1
arCOG13548	S Function unknown	Uncharacterized protein			1
arCOG13549	S Function unknown	Uncharacterized protein		-	1
arCOG13551	S Function unknown	Uncharacterized membrane protein	_		1

arCOG 13552	S	Function unknown		Uncharacterized protein									1		
arCOG13553	S	Function unknown		Uncharacterized membrane protein									-1		
arCOG 13554	~ L	Function unknown	Ĩ	Uncharacterized protein											
arcOG135556	~ ~	Function unknown		Uncharacterized memorane protein Lincharactarized motain											
arCOG 13557	5	Function unknown		Uncharacterized membrane protein											
arCOG 13559	s	Function unknown	Ī	Uncharacterized membrane protein											
arCOG 13560	S	Function unknown		Uncharacterized membrane protein											
arCOG13561	S	Function unknown		Uncharacterized protein											
arCOG 13562	S	Function unknown		Uncharacterized membrane protein											
arCOG 13563	S	Function unknown		Uncharacterized protein											
arCOG13564	SF	Function unknown		Uncharacterized membrane protein	SC.01004			_					1	_	
arCOG 13565	SF	Function unknown		Uncharacterized protein									1	_	
arCOG 13566	SF	Function unknown		Uncharacterized membrane protein					_				1	_	
arCOG 13567	S	Function unknown		Uncharacterized protein					_				1		
arCOG15190	SF	Function unknown		Uncharacterized membrane protein					_				1	_	
arCOG02545	S	Function unknown		Secreted protein, with PKD repeat domain	COG03291	pfam13229, pfam13229	,pfam00801 TI	GR04247					2		
arCOG03338	S	Function unknown		Uncharacterized membrane protein	SC.00033								2		
arCOG03817	S	Function unknown		Uncharacterized protein	SC.00100								2		
arCOG05407	s	Function unknown		Uncharacterized protein									2		
arCOG05805	2	Function unknown		Uncharacterized protein	SC.00141								7		
arCOG08126	~ L	Function unknown	Ĩ	Uncharacterized membrane protein	SC.00775								7		
arc0609431	~ •	Function unknown	Ĩ	Uncharacterized protein	1100000	pram144/8							7		
arcos 12007	~ u	Function unknown	T	Uncharacterized memorane protein Haeu, UUE306 tamiiy	LUGU3241		Ī	I					7 c		
arcog 125.20	n v	Function unknown	T	Ulicital acterized protein Hincharactarized membrane protein	SC 00703								4 C		
arCOG02994	5 00	Function unknown	Ī	Cunin domain containing protein	COG01917	nfam07883							4 m		
arCOG 10113	s	Function unknown	Ī	Uncharacterized protein									m		
arCOG 10161	s	Function unknown		Uncharacterized protein									m		
arCOG03118	SF	Function unknown		Membrane protein DegA family	COG01238	pfam09335								1	1
arCOG07655	SF	Function unknown		Uncharacterized protein										1	1
arCOG01713	s	Function unknown		Uncharacterized protein	COG05440	pfam10061								1	3
arCOG02561	s	Function unknown		WD40 repeats containing protein	COG02319	pfam13191,pfam1160 0	cd00200, cd002 TI	GR02794,TIGR028	300,TIGR02800					1	3
arCOG00620	s	Function unknown		Uncharacterized conserved DUF39 domain fused to CBS domain	COG01900	pfam01837,pfam0057 (cd04605 TI	GR03287,TIGR013	302						
arC0G04574	~ L	Function unknown	MA	Uncharacterized protein	C0G01704	ptam04011									
arc0604811	~ L	Function unknown	T	Uncharacterized membrane protein, Fun14 tamily	CO602323	000000									
arc0606436	~ u	Function unknown	T	Uncharacterized protein, DUFL/22 Tamily	COC003272	ptarriu4403,ptarriu6349	Ī	I							
arcogo6646	n v	Function unknown	T	Ulicital acterized proteiti, 1000/0570001 (attiil) Horbaractarized membrane arotein - VccA (Bav inhikitar family	COG04760	pidii02030									
ar COG00040	n v	Function unknown		Uncharacterized metriolarie procerty rucky bax minuturi rammy Horbaractarized motein	004/100	TTOTTILIPIC									
arC0610710	, v	Function unknown		Uncharacterized protein			T	Ī						- 6	
arCOG01908	5 5	Function unknown AIM	M24	Uncharacterized protein. AIM24 family	COG02013	pfam01987									1
arCOG03873	S	Function unknown		Uncharacterized membrane protein	COG02855	pfam03601	F	GR00698							1
arCOG04987	S	Function unknown		Uncharacterized protein											1
arCOG05100	S	Function unknown		Uncharacterized membrane protein	COG02364										1
arCOG05340	S	Function unknown		Uncharacterized membrane protein											1
arCOG05874	s	Function unknown		Uncharacterized protein					_						1
arCOG06493	S	Function unknown		HEAT repeats containing protein	COG01413	pfam13646, pfam03130									1
arCOG 10568	s	Function unknown		Uncharacterized membrane protein											1
arCOG02539	~ L	Function unknown	Ĩ	Secreted protein, with PKD repeat domain	COG03291										2
arCOG02614	s i	Function unknown		Uncharacterized conserved protein					+						2
arcos 10301	~ "	Function unknown	ſ	Uncharacterized protein	CO603011	pram04134			,	 	,	•			7
arc0602284	× 4	General function prediction only	T	Secreted protein containing C-terminal beta-propeller domain distantly relations	COC0204880	pram09826				-1 (-1 •			
arc0606640	× 4	General function prediction only		Predicted KNA-binding protein containing PUA-domain	COG02224/	pram018/8	14 0000 FF	0001000		7	-	-1 •		-	4
	< 0	General function prediction only biss	INI.	Architemeters of DATI family			11 00000	C D01 E 7E	 	-			Ŧ	-	
arc060035			Ē	Acetylitansierase (GIVAT) Tamity	00000401	5 COCUUIII PID	TO\$+OD	C/CTONS							
arc0602452		General function prediction only	T		COC010407		LF0001-			•	-		n .		ſ
	د م	General function prediction only	T	nu superiaririty priosprioriguroiase Lincharactarizad archaaal Zn-finnar protain	COG01376	00610111910	/ / 0000		7 c	 -	+ -				7
arCOG04227	: 2	General function prediction only		Predicted CoA-hinding protein	COG01832	nfam13380	F	GR02717	1	 -	4	4 -	•		-
arCOG00557	~	General function prediction only [Phr		Lhr-like helicase	COG01201	pfam00270,pfam0027 c	cd00046, cd000 TI	GR04121	1	1		1	2	-	2
arCOG01150	8	General function prediction only -		Predicted phosphohydrolase, MPP superfamily	COG01407		cd07391 TI	GR00024	1	1	1	1	2		
arCOG01857	8	General function prediction only -		SpoU rRNA Methylase family enzyme	COG01303	pfam01994			1	1	1	1	1	1	2
arCOG04151	R	General function prediction only -		ABC-type multidrug transport system, permease component	COG02237	pfam04123			1	1 1	1	1 1	1	1	2
arCOG09205	Я	General function prediction only		Predicted metal-dependent hydrolase	COG01547	pfam03745			1	1	1 1	1 1		_	
arCOG09415	Я	General function prediction only		Membrane associated metal-binding domain fused to Reeler domain		pfam02014 c	cd08544	_	1	1	2	1 1		2	9
arCOG00626	Я	General function prediction only TlyC	ý	Hemolysins or related protein containing CBS domains	COG01253	pfam01595, pfam0057 o	cd04590 TI	GR03520	1	1		1 1		1	2
arCOG02291	Я	General function prediction only -		HAD superfamily hydrolase	COG01011	pfam13419 0	cd01427 TI	GR02252	_			1	2		
arCOG00979	ж	General function prediction only -		Predicted O-methyltransferase YrrM	COG04122	pfam13578 0	cd02440					1			
arCOG01619	2	General function prediction only ARA	8A1	Aldo/keto reductase, related to diketogulonate reductase	COG00656	pfam00248	cd06660 TI	GR01293	1	-	1	1		- 1	1
arCOG04807	~	General function prediction only OP1	F.	Uncharacterized membrane protein, oligopeptide transporter (OPT) family	COG01297	pfam03169	F	GR00733	2	1		1		2	1
arCOG00517	~ "	General function prediction only	Ĩ	IRhodanese Homology Domain fused to Zn-dependent hydrolase of beta-lacts	a COG00491	ptam00753		GR03413	, ₂	,			,	2 -	, 1
arCOG01360	~ 0	General function prediction only		Miab tamily, Radical SAM enzyme	C0G01244	ptam04055 0	cd01335 11	GR01210	 					н,	
arc060015	× 0	General function prediction only	~	Predicted aconitase Drodieted successide utilities commo soluted to moluted activity historythesis	- 100010/9	prarmo441.2		000000	 				-	-	-
	د ۵	General function prediction only Ceneral function prediction only	1	Interfactory independent hydrolace of the heta-last amage superfamily librations in the heta-last amage superfamily librations of the heta-last amage superfamily librations and theta-last amage superfamily librati		priam00753		GR02013	 -	+ -					
arCOG01641	: 2	General function prediction only		Predicted RNA-binding protein. contains TRAM domain	COG03269		-	071-0010	 -						1
arCOG00504	R	General function prediction only	Γ	Metal-dependent hydrolase of the beta-lactamase superfamily II	COG00491	pfam00753	F	GR03413	1	1 1		1		3	1
arCOG00348	Я	General function prediction only	Î	Archaeal enzyme of ATP-grasp superfamily	COG02047	pfam09754	F	GR00162	1 1	1	1 1	1	1	1	2
arCOG04055	ж	General function prediction only		SHS2 domain protein implicated in nucleic acid metabolism	COG01371	pfa m01951			1 1	1		1	1	1	1
arCOG00313	ж	General function prediction only Sco.	01	Cytochrome oxidase Cu insertion factor, SCO1/SenC/PrrC family	COG01999	pfam02630 c	cd02968		1 1		۲ı	1		2	

arCOG00893	R General function prediction only		Predicted metal-dependent hydrolase (urease superfamily)	COG01831 p	fam01026		1	1			1	1	1	2
arCOG00543	R General function prediction only		Predicted metal-dependent RNase, consists of a metallo-beta-lactamase dom	COG01782 F	fam00753, pfam1099 cd02	2410 Th	GR03675 1		1 1	1 1	1	ч	1	ъ
arCOG01093	R General function prediction only		Protein distantly related to bacterial ferritins	COG02406 F	fam00210 cd01	1052	1		1		1			
arCOG00347	R General function prediction only		Archaeal enzyme of ATP-grasp superfamily	COG01938 F	fam09754	Ē	GR00161 2	ļ	1, 1	1 1	1	1	2	-1
arCOG02579	K General function prediction only R General function prediction only		Predicted metal-dependent nydrolase Fe-S-chister containing protein	COG0154/	TamU3/45				T				6	-
arCOG03096	R General function prediction only		NAD dependent epimerase/dehvdratase family enzyme	COG01090	fam01370.pfam0833 cd05	5242 TI	GR01777	2		1	1			4 m
arCOG03991	R General function prediction only		PKD repeats containing protein	0603291	fam00801 cd00	0146 TI	GRODR64 TIGR04213			1	-			, c
arCOG04212	R General function prediction only	,	Predicted DNA-binding protein with PD1-like DNA-binding motif	COG01661	fam03479 cd11	1378					1	1	1	
arCOG06747	R General function prediction only		SIR2 superfamily protein	COG00846	ifam02146, pfam1328 cd03	1406					1			
arCOG06769	R General function prediction only		GTPase SAR1 family domain fused to Leucine-rich repeats domain	COG01100 F	fam13855, pfam1385 cd00	0116,cd099 TI	GR00231				1			
arCOG00499	R General function prediction only		Metal-dependent hydrolase of the beta-lactamase superfamily	COG01235	fam12706	Ē	GR02651 1	1	1 1	1 1	2 1	1	1	1
arCOG01225	R General function prediction only	,	GT Pase SAR1 or related small G protein	COG01100	fam03029 cd02	2027,cd00880	.		1	1	 	LI.	-	¢
arCOG0062	D General function prediction only D General function prediction only		Provinted amidomytrolase	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	fam0.705	71 07 TI	CD02201 1			1	 2	c		7
ar COG02 293	R General function prediction only		Heuticted attriutoriyuri otase HAD superfamily hydrolase	0600637	fam13419 rd0'	1427 TI	GR02009 2	v 6			2 T	4 C		
arCOG02742	R General function prediction only		Uncharacterized membrane anchored protein with extracellular vWF domain	COG01721 r	fam01882		2 2	, t	1 2	4	2 2	1	2	4
arCOG04303	R General function prediction only		Uncharacterized Rossmann fold enzyme	COG01634	fam01973 cd07	7995	1	1	1	1	2	1	1	1
arCOG03450	R General function prediction only		Predicted surface protease of transglutaminase family				-1	2	1 1	1	2	1		
arCOG04466	R General function prediction only		Na+-dependent transporter of the SNF family	COG00733	fam00209 cd10	0336	1	2	1	1	2		1	33
arCOG04950	R General function prediction only	YggS	Uncharacterized pyridoxal phosphate-containing protein, affects Ilv metabolis	COG00325 p	fam01168 cd00	0635 TI	GR00044 1				2	1		
arCOG04624	R General function prediction only	-	Multimeric flavodoxin WrbA	COG00431	fe			•			2	ŗ		
arCOG03682	R General function prediction only	- SPC1	Protein implicated in Kiva metabolism, contains PKC-barrel domain Mambrane associated serine /threonine protein kinase		fam00069 cd12	1014		- "	т т т	т 2	4 T	7		7
arCOG02174	R General function prediction only	5	Predicted exporter of the RND superfamily	COG01033	fam03176.pfam03176	ÍL.	GR00921 4	0 4	4 6	3	+ 12	-	9	11
arCOG01848	R General function prediction only	LddW	Acetyltransferase (isoleucine patch superfamily)	COG00110	ifa m00132, pfam0013 cd03	3358 TI	GR03570 1	2		-1	-		1	1
arCOG06534	R General function prediction only		Cohesin domain containing secreted protein		fam00963 cd08	8547	1		1		1		1	
arCOG07590	R General function prediction only		Predicted acyl esterase	COG02936	fam02129, pfam08530	Ē	GR00976 3	4	1 2	2 2			2	2
arCOG05343	R General function prediction only	,	GTPase SAR1 or related small G protein	COG01100 F	fam00071 cd00	0154 TJ	GR00231			-1			,	,
arCOG03439	R General function prediction only	-	VWFa domain containing protein		fr m0100.4				7			-1 -	7 5	'n
arcoconent	R General function prediction only E Concert function and reliant only	FUCUS	District contributions two CPE domains from a fundation from dealed	COCODE 1 7	framu1984	T DAORS CON	C B 01 3 01 T I C B 01 3 0 1			T	-	-1 -		
arcOG02935	R General function prediction only		Protein containing two cas domains (some rused to C-terminal double-strand	11200000	fam0.2678 nfam05726	4033,00040						-	-	
arCOG02289	R General function prediction only		International Duration Durated Amily	0602043	fam02596			t			1 -		-	
arCOG03994	R General function prediction only		PKD domain containing protein	COG03291	fam05048.pfam0080 cd14	4251.cd001 Th	GR04247.TIGR04275.TI	IGR04275.TIG	304275.TIGR04275.	.TIGR04275.TIGR04	1275.TIGR04 1			
arCOG06897	R General function prediction only		Predicted ATP-grass enzyme	COG03919	fam15632	L L	GR01369 3	3	1	1 1	4			
arCOG03038	R General function prediction only		TPR repeats containing protein	COG00457				2	1 2	1	ى ا	4	2	m
arCOG01285	R General function prediction only		OB-fold domain and Zn-ribbon containing protein, possible acyl-CoA-binding p	COG01545 p	fam12172, pfam01796			1	1	1				
arCOG01651	R General function prediction only		Alpha/beta superfamily hydrolase	COG01073 p	fam08538		1	1	1 2			2		2
arCOG02812	R General function prediction only		Bacteriorhodopsin	COG05524	ifam01036		-1	1	1					1
arCOG00352	R General function prediction only	Nog1	GTP-binding protein, GTP1/Obg family	COG01084 F	fam01926 cd01	1897 TI	GR02729 1	1	1			1	1	1
arCOG01108	R General function prediction only	AbgB	Metal-dependent amidase/aminoacylase/carboxypeptidase	COG01473 F	fam01546 cd03	3886 TI	GR01891 1	1	1			1		1
arCOG07416	R General function prediction only	MhpC	Alpha/beta superfamily hydrolase	COG00596	fam12695	Ē	GR03695 1	1					1	1
arCOG05195	R General function prediction only		TPR repeats containing protein	COG00457	fam13414 cd00	0189 TJ	GR02521 1	1						2
arCOG02889	R General function prediction only		Predicted deacylase	COG03608	fam04952 cd06	6252 TI	GR02994 1	1						
arCOG01425	R General function prediction only		RecB family nuclease with coiled-coil N-terminal domain	SC.00001			- •		1	,				1
arc0602259	R General function prediction only	-	Zn-finger domain containing protein	SC.00304		14 0 14 0	CD0147F 4			1				ç
arCOC01141	R General function prediction only	-	GTPdSE, G3ETdmIIY	0000033	4 main 2494, pramo 700 cuo:	11 7116	GRU24/5 1		-	T		Ŧ		7
arc0601395	R General function prediction only		ILC-like prosprioesterase I i pid-A-disaccharide sunthase related alvcosultrancferase	COG01817	fam04007	11	1 T					-		
arCOG01648	R General function prediction only	MhpC	Alpha/beta superfamily hydrolase	COG00596	fam12697	Ē	GR02427 1							
arCOG02164	R General function prediction only		Predicted transglutaminase-like protease	COG01800	ifam04473		1							
arCOG01622	R General function prediction only	MviM	Predicted dehydrogenase	COG00673	ifam01408, pfam02894		2			1			1	
arCOG03509	R General function prediction only		Cell surface protein, containing PKD repeats	COG03291	fam10102			1	1					
arCOG00614	R General function prediction only	SpolVFB	Zn-dependent protease	COG01994	cd06	6158		1				1	1	2
arCOG03015	R General function prediction only		Uncharacterized protein YbjT, contains NAD(P)-binding and DUF2867 domain	COG00702	fam13460 cd0	5271 TI	GR03466	1						1
arC0G02625	R General function prediction only	-	Predicted metal-dependent hydrolase	COG01451 F	fam01863									
arc0601136	R General function prediction only	-	ABC-2 Tamily transporter protein	000012//	fam126/9	T UUO		-1 (T				Ŧ
arCOG03048	R General function prediction only	-	TPR repeats containing process	COG00457	fam13414. pfam1343 cd00	0189.cd001	GR02917	-	1	1				
arCOG04932	R General function prediction only		Cell surface protein	COG01470					1	1				
arCOG01213	R General function prediction only	Cof	HAD superfamily hydrolase	COG00561 F	ifam08282 cd0:	1427 TI	GR01487		1				1	2
arCOG03850	R General function prediction only		Predicted O-methyltransferase YrrM	COG04122	ifam13578		_			1				
arCOG00302	R General function prediction only	-	Metal-dependent phosphoesterase (PHP family)	COG00613	fam02811,pfam1326 cd07	7432						1	1	1
arCOG00353	R General function prediction only	HfIX	GTP-binding protein protease modulator	COG02262	fam13167, pfam0192 cd0:	1878 TJ	GR03156					1	1	1
arCOG01043	R General function prediction only	, i	Predicted RNA binding protein with dsRBD fold	COG01931 F	fam01877	1						-		-
arc060E137	R General function prediction only E Concert function and reliant only	CIDA	Uncharacterized protein (competence- and mitomycin-induced)	COG01546	13m02464	11 00 r d001	GRUU199			T		-1 -		
arCOGOD327	R General function prediction only		TEN repeats containing protein Incharacterized protein related to pornvate formate-loace activating enzyme	0602108	fam0.4055	1335	1167010							- ~
arCOG04116	R General function prediction only		ATPase (PilT family)	COG01855	fam00437 cd09	9878.cd01130								- 2
arCOG04265	R General function prediction only		C4-type Zn-finger protein	COG01779	fam03367	Ē	GR00340						1	m
arCOG01084	R General function prediction only	MazG	Predicted pyrophosphatase	COG01694	fam03819 cd11	1535						1	1	
arCOG02448	R General function prediction only		Uncharacterized Fe-S center protein	COG02768	ifa m04015, pfam12838	Ē	GR01971					1		2
arCOG00254	R General function prediction only	-	Predicted dinucleotide-utilizing enzyme	COG01712 F	fam03447, pfam01958	Ē	GR03855					1		
arCOG00600	R General function prediction only		CBS domain	COG00517 F	fa m005 71, pfam0057 cd0/	4623,cd046 TI	GR01302,TIGR00393					1		
arCOG00662	R General function prediction only		Biotin synthase-related protein, radical SAM superfamily	COG02516 F	fam04055 cd03	1335 TJ	GR04317,TIGR04043							
arcog00769	R General function prediction only	Thù	Predicted firetar-wepenverit, iryuroiase with the hint-variet row	COG0063	fam01965 cd03	1300 8134 TI	GR01382							
arCOG00828	R General function prediction only	_	Acetvitransferase GNAT superfamily	COG03393 F	fam08445	Ē	GR01575							
arCOG00913	R General function prediction only	-	Predicted methyltransferase	COG01568	fam01861							1		
arCOG00933	R General function prediction only		Radical SAM superfamily enzyme	COG01964	fam04055 cd0:	1335 TI	GR02666					ч		

arCOG00940	R General function prediction	only -		Radical SAM superfamily enzyme	0G00535	pfam04055, pfam1318 cc	01335	TIGR04055			1	
arCOG00969	R General function prediction	only -		Predicted hydrolase (metallo-beta-lactamase superfamily)	0G02248	pfam12706					1	
arCOG01145	R General function prediction	only -		Phosphohydrolase, Icc/MPP superfamily	0G02129	pfa m00149 cd	107392				1	
arCOG01263	R General function prediction	only Dlt	OltE	Short-chain dehydrogenase	0600300	CC meta 2510 meta 200	105233	TIGR01830				
arC0G01295	R General function prediction	only Hcz	fcaD	NAD(FAD)-dependent denydrogenase NAD(FAD)-dependent dehydrogenase	0G00446 0G00446	pfam07992		TIGR01292				
arCOG01377	R General function prediction	only -		Phosphodiesterase of AP superfamily	0G03379	pfam01663					1	
arCOG01378	R General function prediction	only -		Uncharacterized protein of the AP superfamily	0G01524	pfa m01663					1	
arCOG01455	R General function prediction	only Ad	AdhP	Zn-dependent alcohol dehydrogenase	0G01064	pfam08240, pfam0010 cd	108259	TIGR02824				
arCOG01728	R General function prediction	only Mb	Aho1	Predicted ringingupterin-dependent oxigor equicase rigo	0601355	pfam01875	107361	TIGR04336	-			
arCOG01801	R General function prediction	only [m]	du	TRAP-type uncharacterized transport system, periplasmic component	0G02358	pfam12974 cd	113567	TIGR02122			1	
arCOG01849	R General function prediction	only Pai	aaY	Isoleucine patch superfamily protein	0G00663	CC	104650	TIGR02287			1	
arCOG01906	R General function prediction	only -		TRAP-type uncharacterized transport system, fused permease component	0G04666	pfam06808		TIGR02123			₁	
arCOG02008	R General function prediction	only -		Uncharacterized membrane protein, a putative transporter component	0603371	ptam06197	-	TICDOUT				
arC0602317	R General function prediction	only -		Predicted DIVA-Dinding protein Predicted regulator of amino acid metabolism contains ACT domain	0602150	DIGUIDADUT		110KU293/				
arCOG02431	R General function prediction	only -		Predicted Rossmann fold nucleotide-binding protein	0G01611			TIGR00725				
arCOG02444	R General function prediction	only -		Predicted permease	0G03368	pfa m09847						
arCOG02828	R General function prediction	only -		NAD dependent epimerase/dehydratase family	0G03367	pfa m07755					1	
arCOG02900	R General function prediction	only -		vWFA domain containing protein	0G02304	pfam00092 cc	100198				1	
arCOG02902	R General function prediction	only -		vWFA domain containing protein	0G02304	pfam13519 co	101467				, ₁	
arc0603.045	R General function prediction			IPK repeats containing protein	0603510	pram13424 CC	100189,00018	55				
arCOG03639	B General function prediction			Divit superiaring transporter Predicted alutamine amidotransferase	06001210		101908					
arCOG03691	R General function prediction	only -		Cell surface protein								
arCOG04065	R General function prediction	only Pqu	JagL	Zn-dependent peptidase	0G00612	pfa m00675, pfam05193					1	
arCOG04115	R General function prediction	only Sfs	SfsA	DNA-binding protein, stimulates sugar fermentation	0G01489	pfam03749		TIGR00230			1	
arCOG04230	R General function prediction	only -		HD supefamily hydrolase	0G03294						, ₁	
arc0604290	R General function prediction	only -		PIN-domain and Zh ribbon	0601656	pram0192/					-1 -	
arCOG04231	R General function prediction			Jugar isonnerase refateu protein Thioastarase-lika protain	1001000	50	103440					
arCOG04354	R General function prediction	only -		Tripesteresentes procenter (TTT) class transporter	0601906	pfam04165	014001	TIGR00529				
arCOG04359	R General function prediction	only -		Predicted RNA-binding protein containing a C-terminal EMAP domain	0G02517	pfam01588 co	102796	TIGR00472			-1	
arCOG04410	R General function prediction	only -		Predicted ATP-grasp domain fused to redox center	0G01578	pfam01937					1	
arCOG04418	R General function prediction	only -		Predicted HTH domain, homologous to N-terminal domain of RPA1 protein fail	0G03612	pfam09999					1	
arCOG04426	R General function prediction	only Hy	1ybF	Zn finger protein HypA/HybF (possibly regulating hydrogenase expression)	0G00375	pfam01155		TIG R00100		_	1	
arCOG04477	R General function prediction	only -		Predicted metal binding protein, contains two cysteine clusters	0G01860	pfam03684					1	
arCOG05097	R General function prediction	only -		CBS domain	0600517	pfam00571, pfam0057 cc	102205	TIG R01302			₁	
arC0G05366	R General function prediction	only -		Cell surface protein	0601572	ptam07705	10101					
arcodus 10	R General function prediction	only -		Uncharacterized protein, nomolog of lactam utilization protein B	0601540	ptamU3/4b C0	10/8/				-1 -	
arCOG07368	R General function prediction			Cell surface protein	0601572	nfam07205	CONTOR				4 -	
arCOG11082	R General function prediction	only -		Uncharacterized membrane protein. a component of a putative secretion syste								
arCOG00497	R General function prediction	only -		Zn-dependent hydrolase of the beta-lactamase fold	0G02220	pfam13483					2	1
arCOG00606	R General function prediction			CBS domain	0G00517	pfam00478 cc	104623	TIG R01302			2	1
arCOG04469	R General function prediction	only -		Tripartite tricarboxylate transporter (TTT) class transporter	0G01784	pfam01970					2	1
arCOG03032	R General function prediction	only -		TPR repeats containing protein	0G00457	pfam13414, pfam1341 cc	100189	TIGR02917			2	1
arCOG00503	R General function prediction	only -		Metal-dependent hydrolase of the beta-lactamase superfamily I	0601237	pfam00753		TIGR03675			2 2	
arC0601963	R General function prediction	only -		Predicted transcriptional regulator, contains mun and 4vk domain Phol Lilke domain fitsed to TrkA-C domain	0603773	ptam01895 nfam02080					7 6	
arCOG02.292	R General function prediction	only	hh	HAD superfamily hydrolase	0600546	nfam13419	101427	TIG R03351			2 C	
arCOG02603	R General function prediction	only -		Roadblock/LC7 domain	0G02018	22 CT 1010	14-72	1000000			- 2	
arCOG03167	R General function prediction	only -		Predicted ATPase, AAA+ superfamily	0G01373	pfam13173, pfam13635					2	
arCOG03400	R General function prediction	only -		Predicted RNA-binding protein, contains TRAM domain	0G04085	pfam01336 cc	104485				2	
arCOG04409	R General function prediction	only -		Predicted nuclease (RNAse H fold)	0G02410	pfam04250					2	
arCOG07997	R General function prediction	only -		Predicted OB fold RNA-binding domain fused metal-dependent hydrolase	0601988	pfam01336, pfam04307	10050	T/C D04004			7 2	
arC0602175	R General function prediction	only -		Redicted SAWI superiamity enzyme Dredicted transnorter of the RND superfamily	0602409	pram04126 nfam03176	CEETOR	TIGR00833 TIGR0083			n	1
arCOG 10597	R General function prediction	only -		Hemocyanin family protein, binds copper ions	001-0000	pfam00264						1
arCOG 11383	R General function prediction	only -		MOCS domain, sulfur-carrier protein	0G02258	pfa m03473		_				1 1
arCOG00498	R General function prediction	only -		Metal-dependent hydrolase of the beta-lactamase superfamily II	0G00491	pfam00753		TIGR03413				1 2
arC0G00500	R General function prediction	only Elä	ElaC	Metal-dependent hydrolase of the beta-lactamase superfamily	0601234	ptam12/06		TIGR02651				1 2
arCOG02642	R General function prediction	only Der	Per M	Interal-ueperiori rijuruolase or trie beta-laccaritase superrariity Dradictad Durg-ragulatad narmaasa DarM	0600628	nfam01504		TIGR02872				1 c
arCOG02890	R General function prediction	only -		Predicted deacvlase	0G03608	pfam04952 cc	106251	TIGR02994				1 2
arCOG04743	R General function prediction	only -		Helicase associated uncharacterized terminal domain		pfa m05854						1 2
arCOG00654	R General function prediction	only -		Predicted periplasmic solute-binding protein	0G02107	pfam02621 cd	113534					1 3
arCOG00655	R General function prediction	only -		Predicted periplasmic solute-binding protein	0G01427	pfam02621 cd	113634	TIC D01000				1 3
arcubul 189	R General function prediction	Only Md	4arr-	Predicted unusuai protein Kinase المحطنطينية المالية المعالمة المعاصلة المعالمة المعالمة المعالمة المعالمة المعالمة المعالمة المعالمة المعالمة ال	DGUUBB1	ptamusius https://www.com	171500	TIG KU1982	+			гч 1
arcog01850	R General function prediction	only WE	LddV	Predicted Millase related to galactomillase and inevalutate Millase Acetyltransferase (isoleucine patch superfamily)	0000110	pfa m00132 cd	104647	TIGR03532				
arCOG02303	R General function prediction	only Sur	iurE	Predicted acid phosphatase	0G00496	pfam01975, pfam14423		TIG R00087				1
arCOG02839	R General function prediction	only -		Uncharacterized protein related to deoxyribodipyrimidine photolyase	0G03046	pfam04244, pfam03441						1
arCOG02986	R General function prediction	only Bic	BioY	Uncharacterized protein	0G01268	pfam02632						1
arCOG03247	R General function prediction			IPK repeats containing protein ATPase AAA family	0604637	nfam13304.nfam13304						1
arCOG03271	R General function prediction	only Ytfi	ʻtfP	Uncharacterized protein YtfP, gamma-glutamylcyclotransferase (GGCT)/AIG2-I	0G02105	pfam06094 cd	106661					
arCOG05099	R General function prediction	only Ytf	/tfP	Uncharacterized protein YtfP, gamma-glutamylcyclotransferase (GGCT)/AIG2-	0G02105	pfam06094 cd	106661					1,
arCOG08119	R General function prediction	only Pn.	hox	Secreted phosphatase, PhoX family	0603211	pfam05787	44 40EE	T				r
arcoustat	K General Iunction previous	only I-		Secreted protein with beta-properier repeat domain	060337T	DTamU1430, prarriu 1491 uu	1 CC4710			-		2

arCOG07781	R General function prediction only		Cell surface protein					_				2	
arCOG02562	R General function prediction only		Beta-propeller repeat containing protein	COG03391								3	1
arCOG02810	R General function prediction only		Bacteriorhodopsin	COG05524 F	ofam01036								1
arCOG03169	R General function prediction only		AAA+ superfamily ATPase	COG01672 F	ofa m01637, pfam01978			_					1
arCOG07790	R General function prediction only		D-glucuronyl C5-epimerase C-terminal domain related protein	2	ofam06662								1
arCOG02316	R General function prediction only		Predicted regulator of amino acid metabolism, contains ACT domain	COG02150				_					2
arCOG02560	R General function prediction only		Secreted protein with beta-propeller repeat domain	COG03391	0	:d05819	TIG R03866						2
arCOG06256	R General function prediction only		Predicted esterase	COG00400 F	ofam02230								2
arCOG00082	EF	PucG	Serine-pyruvate aminotransferase/archaeal aspartate aminotransferase	COG00075 F	ofam00266	cd06451	TIG R03301	1			1	1	1
arCOG01446	VK	Csa3	CRISPR-Cas assicated transcriptional regulator, contains CARF and HTH doma.	COG00640	0	cd09655	TIGR01884				1		



Supplementary Table 6. Classification of the unique genes based on the arCOG classification.

		arCOG clasification	gene	product	COG classification	pFAM domain	cdc	TIGR classification	EPIPELAGIC	N BATHY1	UMBER OF BE	nes Thalassoar chaea	MG2-GG3
INFORMATIC arCOG00415 arCOG00439		RAGE AND PROCESSING Replication, recombination and repair Replication, recombination and repair	RecA MCM2	RecA/RadA recombinase Predicted ATPase involved in replication control, Cdc46/Mcm family	COG00468 COG01241	pfam14520,pf pfam14551,pf	a cd01123 a cd00009	TIGR02236	1				
arCOG04110 arCOG00787	L	Replication, recombination and repair Replication, recombination and repair	PRI1	Eukaryotic-type DNA primase, catalytic (small) subunit UvrD/Rep family helicase fused to exonuclease family domain	COG01467 COG02887	pfam01896 pfam12705	cd04860 cd09637 cd06137	TIGR00335 TIGR01249,TIGR00372	1			1	1
arCOG01898 arCOG00469	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	Uve HolB	UV damage repair endonuclease ATPase involved in DNA replication HolB, small subunit	COG04294 COG00470	pfam03851 pfam13177,pf	a cd00009	TIGR00629 TIGR02397	1 2				1
arCOG01894 arCOG01073 arCOG01347	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	- CDC9	Endonuclease IV NUDIX family hydrolase ATP-dependent DNA ligase	COG00648 COG00494 COG01793	pfam01261 pfam00293 pfam04675,pf	cd00019 cd03424 a cd07901,cd079	TIGR00587 TIGR00052 TIGR00574	2 2 2 2				
arCOG02840 arCOG01078	L	Replication, recombination and repair Replication, recombination and repair	PhrB -	Deoxyribodipyrimidine photolyase NUDIX family hydrolase	COG00415 COG00494	pfam00875,pf pfam00293	am03441 cd03428	TIGR03556	2	1		1	1
arCOG00464 arCOG03142 arCOG01486	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	- RnmV	3-methyladenine DNA glycosylase/8-oxoguanine DNA glycosylase Nuclease of RNase H fold, RuvC/YqgF family SS rRNA maturation endonuclease (Ribonuclease MS), contains TOPRIM dom	COG00122 COG01658	pfam07934,pf	cd00056	11GR00588	3 3 4	1	1	1	1 1
arCOG08649 arCOG00417	L	Replication, recombination and repair Replication, recombination and repair	- RecA	Topolsomerase IB RecA/RadA recombinase Two IIA topolsomerase (DNA supers (topo II, topolsomerase B(), A subusit	COG03569 COG00468	pfam02919,pf	cd00560,cd006 cd01394	59 TIGR02237	4			2	1
arCOG00872 arCOG00551	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	MPH1 -	ERCC4-like helicase DNA replication initiation complex subunit, GINS15 family	COG01111 COG01711	pfam00270,pf	a cd000187 a cd00046,cd120 cd11714	TIGR00643,TIGR00580,TIGR00596		1	1	2	1
arCOG00427 arCOG02258 arCOG01166	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	RecJ/Cdc - Mutl	Single-stranded DNA-specific exonuclease RecJ RPA family protein, a subunit of RPA complex in P.furiosus DNA mismatch renair enzyme (predicted ATBase)	COG00608 COG03390	ofam13589.of	cd00075.cd007	16900585		1	1	2	
arCOG04121 arCOG00470	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	RnhB HolB	Ribonuclease HII ATPase involved in DNA replication HolB, large subunit	COG00323 COG00164 COG00470	pfam01351 pfam00004	cd07180 cd00009	TIGR00729 TIGR02397				2	1
arCOG04371 arCOG00488 arCOG01527	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	GyrB DnaN TonA	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit DNA polymerase sliding clamp subunit (PCNA homolog) Topoisomerase IA	COG00187 COG00592 COG00550	pfam02518,pf pfam00705,pf ofam01751 pf	a cd00075,cd008 a cd00577 a cd03362 cd001	TIGR01059 TIGR00590 TIGR01057		1		1 2 1	1
arCOG04050 arCOG00558	L	Replication, recombination and repair Replication, recombination and repair	FEN1 SrmB	5'-3' exonuclease Superfamily II DNA and RNA helicase	COG00258 COG00513	pfam00752,pf pfam00270,pf	a cd09867 a cd00268,cd000	TIGR03674 TIGR01389		1		1	2
arCOG00328 arCOG01510 arCOG01072	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	PolB3 RPA1	DNA polymerase PolB3 Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subun NUDIX family hydrolase	COG00417 COG01599 COG00494	pfam03104,pf	a cd05781,cd055 cd04491,cd044 cd03426	TIGR00592 91		2		4 2	1
arCOG03013 arCOG00553	L	Replication, recombination and repair Replication, recombination and repair	PRI2 BRR2	Eukaryotic-type DNA primase, large subunit Replicative superfamily II helicase	COG02219 COG01204	pfam04104 pfam00270,pf	cd06560 cd00046,cd000	TIGR04121		1		3	1
arCOG01526 arCOG00368 arCOG02724	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	TopG2 SbcC Ada	Reverse gyrase ATPase involved in DNA repair, SbcC Methylated DNA-protein cysteine methyltransferase	COG01110 COG00419 COG00350	pfam00270,pf pfam13476,pf pfam01035	a cd00046,cd033 a cd03240,cd001 cd06445	TIGR01054 TIGR00611,TIGR02168 TIGR00589			1	1	
arCOG00397 arCOG00329	L	Replication, recombination and repair Replication, recombination and repair	SbcD PolB2	DNA repair exonuclease, SbcD DNA polymerase PolB2, inactivated	COG00420 COG00417	pfam00149 pfam00136	cd00840 cd05531	TIGR00592			1	1	1
arCOG00802 arCOG02895 arCOG04694	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	RecB MutS2 UvrA	ATP-dependent exoDNAse (exonuclease V) beta subunit (contains helicase an DNA structure-specific ATPase involved in suppression of recombination, Mu Excinuclease ABC subunit A, ATPase	COG01074 COG01193 COG00178	pfam00580,pf pfam00488	cd03243 cd03271,cd032	TIGR02785 TIGR01069 TIGR00630				1 3	1
arCOG00462 arCOG03646	L	Replication, recombination and repair Replication, recombination and repair	MutY XseB	A/G-specific DNA glycosylase Exonuclease VII small subunit	COG01194 COG01722	pfam00730 pfam02609	cd00056	TIGR01084 TIGR01280					1
arCOG04513 arCOG04754 arCOG07300	L	Replication, recombination and repair Replication, recombination and repair Replication, recombination and repair	iseA Lig	NAD-dependent DNA ligase Uncharacterized protein associated with inactivated PolB-like polymerase	COG01570 COG00272	pram13742,pf pfam01653,pf	a cd00114,cd000	TIGR00575					1 1 1
arCOG01082 arCOG01305	L	Replication, recombination and repair Replication, recombination and repair	- -	NUDIX family hydrolase Topoisomerase DNA binding C4 zinc finger fused to uncharacterized N-termin Microsofte sensite ATDate (Multi family)	COG00494 COG01637	pfam00293	cd02885	1/6801070				1	
arCOG01981 arCOG01863	K	Transcription Transcription	SUA7	Mismatch repair Al Pase (Mucs family) Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor Predicted transcription factor, homolog of eukaryotic MBF1	COG01405 COG01813	pfam01624,pf pfam08271,pf pfam01381	a cd003284 a cd00043 cd00093	TIGR00270	5	1	1	1 1	
arCOG02611 arCOG07561	K K	Transcription Transcription Transcription	- -	Predicted transcriptional regulator, containd two HTH domains Predicted membrane-associated trancriptional regulator	COG03398 COG02512	pfam13412,pf	a cd00090,cd000	100	3	1	1	2	1 4
arCOG05671 arCOG01680	K K	Transcription Transcription Transcription	CsgD ArsR	Transcriptional regulator LuxR family Transcriptional regulator containing HTH domain, ArsR family	COG02771 COG00640	pfam01022	cd00090	11GR00004	2		1	2	
arCOG04280 arCOG01753	K K	Transcription Transcription Transcription	NagC Ssh10b	Transcriptional regulator/sugar kinase Archaeal DNA-binding protein DNA directed BNA polymource, subwit M/Transcription elemention factor T	COG01940 COG01581	pfam00480 pfam01918	cd00012	TIGR00744 TIGR00285		1	1	1	
arCOG01684 arCOG00675	K K	Transcription Transcription Transcription	- RPB7	Transcriptional regulator, ArsR family DNA-directed RNA polymerase, subunit E'/Rpb7	COG01777 COG01095	pfam01022 pfam03876,pf	cd00090,cd000 a cd04331,cd044	190 TIGR00448				3	1
arCOG04258 arCOG01580 arCOG05161	K K	Transcription Transcription Transcription	- WecD	DNA-directed RNA polymerase, subunit H, RpoH/RPB5 DNA-binding transcriptional regulator, Lrp family Acetyltransferase (GNAT) family	COG02012 COG01522 COG00454	pfam01191 pfam13412,pf	a cd00090			1		1	1
arCOG04111 arCOG01764	K K	Transcription Transcription	RPB11 SPT15	DNA-directed RNA polymerase, subunit L TATA-box binding protein (TBP), component of TFIID and TFIIIB	COG01761 COG02101	pfam13656 pfam00352,pf	cd06927 cd04518			1			1
arCOG02099 arCOG04241 arCOG02038	K K	Transcription Transcription Transcription	TroR RpoA/Rp -	Mn-dependent transcriptional regulator (DtxR family) DNA-directed RNA polymerase subunit D Sugar-specific transcriptional regulator TrmB	COG01321 COG00202 COG01378	pfam01325,pf pfam01193	cd07030	4023			1	1	1
arCOG02037 arCOG01057	к к	Transcription Transcription	- HxlR	Sugar-specific transcriptional regulator TrmB DNA-binding transcriptional regulator, HxIR family	COG01378 COG01733	pfam01978 pfam01638	cd00090					1	1 3
arCOG00608 arCOG04248 arCOG04377	K K	Transcription Transcription Transcription	- SIR2 -	Predicted transcriptional regulator with C-terminal CBS domains NAD-dependent protein deacetylase, SIR2 family Transcriptional regulator MarR family, contains HTH domain	COG03620 COG00846 COG04738	pfam01381,pf pfam02146	a cd00093,cd046 cd01413	TIGR03070,TIGR01137					1 1 1
arCOG00826 arCOG02644	к к	Transcription Transcription	WecD AcrR	Acetyltransferase (GNAT) family Transcriptional regulator, TetR/AcrR family	COG00454 COG01309	pfam00583 pfam00440	cd04301	TIGR01575 TIGR03613				1	1
arCOG05152 arCOG02271 arCOG02274	K K	Transcription Transcription Transcription	-	Transcriptional regulator, contains HTH domain Transcriptional regulator, contains HTH domain Transcriptional regulator, contains HTH domain	COG03413 COG03413	pfam04967 pfam04967						4	2
arCOG02280 arCOG04818	ĸ	Transcription Transcription	- HepA	Transcriptional regulator, contains HTH domain Superfamily II DNA/RNA helicase, SNF2 family	COG03413 COG00553	pfam04967 pfam13091,pf	a cd09178,cd000	TIGR01587				1 1	
arCOG02197 arCOG04107 arCOG02466	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	SUI2 GAR1	Translation initiation factor 2, alpha subunit (eIF-2alpha) RNA-binding protein involved in rRNA processing	COG01017 COG01093 COG03277	pfam00575,pf	a cd04452	TIGR00717	5 4	1 1			
arCOG02286 arCOG04108	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	Ncl1 RPS27A	Ribosome biogenesis protein, NOL1/NOP2/fmu family Ribosomal protein S27E	COG03270 COG02051	pfam13636 pfam01667		TICBOOAAE	4	1		1	
arCOG00042 arCOG00033	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	TilS Trm5	tRNA(Ile)-lysidine synthase TilS/MesJ Wybutosine (yW) biosynthesis enzyme, Trm5 methyltransferase	COG00037 COG02520	pfam01189 pfam01171 pfam02475	cd01993	TIGR02432	4 4 4	1	1	1 2	1
arCOG04449 arCOG01179 arCOG04150	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	TruA InfA Krr1	Pseudouridylate synthase Translation initiation factor 1 (IF-1) rRNA processing protein Krr1/Ppo1_contains KH domain	COG00101 COG00361 COG01094	pfam01416,pf pfam01176	a cd02866 cd05793	TIGR00071 TIGR00523 TIGR03665	4 3 3			1	
arCOG04149 arCOG00078	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	NMD3 NOP1	NMD protein affecting ribosome stability and mRNA decay Fibrillarin-like rRNA methylase	COG01499 COG01889	pfam04981 pfam01269			3	1		1	1
arCOG00784 arCOG00501 arCOG04070	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	POP4 Rnz RpIC	RNase P/RNase MRP subunit p29 Ribonuclease Z, beta-lactamase superfamily hydrolase Ribosomal protein L3	COG01588 COG01234 COG00087	pfam01868 pfam12706 pfam00297		TIGR02651 TIGR03626	3 3 3		1	2	1
arCOG00985 arCOG00047	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	Tma20 Trm11	Predicted RNA-binding protein, contains PUA domain tRNA G10 N-methylase Trm11	COG02016 COG01041	pfam09183,pf	am01472	TIGR03684 TIGR01177	3	1	1 1	2	1
arCOG00910 arCOG00910 arCOG01695	1	rransiauon, ribosomai structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	11W3 - -	wypurosine (yW) biosyntnesis enzyme Predicted RNA methylase Predicted component of the ribosome quality control (RQC) complex. YloA/T	COG02263 COG01293	ptam02676 pfam13659 pfam05833,pf	am05670		3 2 2	1	1	1 1	1
arCOG00676 arCOG00809	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	Csl4 LeuS	Exosome complex RNA-binding protein Csl4, contains S1 and Zn-ribbon domi Leucyl-tRNA synthetase Acotheteoefferson, Bind from the	COG01096 COG00495	pfam14382,pf	a cd05692 a cd00812,cd008	TIGR00395	2	1	1	1	1
arCOG00842 arCOG04345 arCOG01701	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	RPR2 SEN2	Acetyttransterase, kunt tamity RNase P subunit RPR2 tRNA splicing endonuclease	COG01670 COG02023 COG01676	pfam04032 pfam02778,pf	am01974	TIGR00324	2 2 2	1	1 1 1	1	
arCOG10124 arCOG01358	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	TYW2 MiaB	Wybutosine (yW) biosynthesis enzyme, TYW2 transferase 2-methylthioadenine synthetase	COG02520 COG00621	pfam02475 pfam00919,pf	cd02440 a cd01335	TIGR01444 TIGR00089	2	1		2	1
arCOG00721 arCOG00038 arCOG00990	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	Thil	Enconducese NOD1, COTSISTS of a PTN domain and a 2n-ribbon module tRNA S(4)U 4-thiouridine synthase Queuine tRNA-ribosyltransferase, contain PUA domain	COG01439 COG00301 COG01549	pfam02926,pf pfam14810,pf	cd09876 a cd11716,cd017 am01472	TIGR00342,TIGR04271 TIGR00432	1	1	1	1 1 2	1
arCOG01346 arCOG04130 arCOG01254	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and hingenesis	- - AlaX	RNA-binding protein containing KH domain, possibly ribosomal protein Predicted RNA-binding protein Ser-IRNA(A)a) deacylase AJaX (editing enzyme)	COG01534 COG01491 COG02872	pfam01985 pfam04919 pfam01411 of	am07973	TIGR03683				2	1 1
arCOG00487 arCOG01185	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	ArgS Bud32	Arginyl-tRNA synthetase tRNA A-37 threonylcarbamoyl transferase component Bud32	COG00018 COG03642	pfam00750,pf	a cd00671,cd079 cd05151	TIGR00456 TIGR03724				1 3	1
arCOG04249 arCOG00486 arCOG04161	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	CCA1 CysS DPH5	tRNA nucleotidyltransferase (CCA-adding enzyme) Cysteinyl-tRNA synthetase Diphthamide biosynthesis methyltransferase	COG01746 COG00215 COG01798	pfam01909,pf pfam01406,pf pfam00590	a cd05400 a cd00672,cd079 cd11647	TIGR03671 TIGR00435 TIGR00522		1	1		1
arCOG00035 arCOG04332	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	Dph6 EbsC	Diphthamide synthase (EF-2-diphthineammonia ligase) Cys-tRNA(Pro)/Cys-tRNA(Cys) deacylase, ybaK family	COG02102 COG02606	pfam01902 pfam04073	cd01994 cd04333	TIGR03679 TIGR00011				1 2	
arCOG01988 arCOG04277 arCOG04312	1	rransiation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	EFB1 Efp Fcf1	I ranslation elongation factor EF-1beta Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-SA) rRNA-processing protein FCF1, contains PIN domain	COG02092 COG00231 COG01412	ptam00736 pfam08207,pf	cd00292 a cd04467 cd09879	TIGR00489 TIGR00037				2	1 1 1
arCOG01559 arCOG00978	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	FusA GCD14	Translation elongation factor G, EF-G (GTPase) tRNA(1-methyladenosine) methyltransferase	COG00480 COG02519	pfam00009,pf	a cd01885,cd037 cd02440	TIGR00490 TIGR03534		1	1	1	1
arCOG01616 arCOG00109 arCOG00404	1	rransiauon, ribosomai structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	GEK1 HemK HisS	u-ammoacyi-tKNA deacyiase, involved in ethanol tolerance Methylase of polypeptide chain release factors Histidyl-tRNA synthetase	COG02890 COG00124	ptam04414 pfam13659 pfam13393,pf	cd02440 a cd00773,cd008	TIGR00537 TIGR00442		1	1		1
arCOG00807 arCOG01560	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	lleS InfB	Isoleucyl-tRNA synthetase Translation initiation factor 2 (IF-2; GTPase)	COG00060 COG00532	pfam00133,pf	a cd00818,cd008 a cd01887,cd037	TIGR00392 TIGR00491				1	1
arCOG001736 arCOG00810 arCOG00906	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	MetG Nop10	2 -3 mmw iigd98 Methionyl-tRNA synthetase rRNA maturation protein Nop10, contains Zn-ribbon domain	COG001514 COG00143 COG02260	pfam09334,pf pfam04135	a cd00814,cd079	TIGR00398			1	1	

arCOG01741	J	Translation, ribosomal structure and biogenesis	PelA	Release factor eRF1	COG01537	pfam03463.pfam03464.pfam03	TIGR00111		1	2	1
arCOG00412 arCOG00402	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	PheT ProS	Phenylalanyl-tRNA synthetase beta subunit Prolvi-tRNA synthetase	COG00072 COG00442	pfam03483,pfa cd00769 pfam00587.pfa cd00778.cd008	TIGR00471 TIGR00408			2	1
arCOG04228	J	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	Pth2 Pus10	Peptidyl-tRNA hydrolase HRNA LISA and LISS oseudouridine synthase Pust0	COG01990	pfam01981 cd02430	TIGR00283			2	2
arCOG00039 arCOG05111	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	QueC	7-cyano-7-deazaguaniane synthase (queuosine biosynthesis) 235 r/PM psaudo 11915 N3-methylase BIMH	COG00603	pfam06508 cd01995	TIGR00364				1
arCOG06660	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	RluA	Pseudouridylate synthase, 23S RNA-specific Phocomal exeterior 195	COG00564	pfam00849 cd02869	TIGR00005			1	1
arCOG00780 arCOG04175	1	Translation, ribosoma structure and biogenesis Translation, ribosomal structure and biogenesis	RPL20A	Ribosomal protein L20A (L18A)	COG02157	pfam00828 pfam01775 pfam01346 cd00473	10001071				1
arCOG01752	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	RPL30	Ribosomal protein L30E	COG01911	pfam01248			1	1	
arCOG04049 arCOG04208	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis Translation, discound structure and biogenesis	RPL40A RPL43A	Ribosomal protein Late: Ribosomal protein L37AE/L43A	COG01552 COG01997	pfam01780	TIGR00280			1	
arCOG04288 arCOG04372	1	Translation, ribosomarstructure and biogenesis Translation, ribosomal structure and biogenesis Translation, discount structure and biogenesis	RpIK	Ribosomal protein L11 Discomal protein L11	COG00244 COG00080	pfam03946,pfa cd00349	TIGR01632			2	1
arCOG04094 arCOG04186	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	RPS1A	Ribosomal protein L24 Ribosomal protein S3AE	COG00198 COG01890	pfam01015	TIGRU1080		1	1	
arCOG04182 arCOG04239	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	RPS24A RpsD	Ribosomal protein S24E Ribosomal protein S4 or related protein	COG02004 COG00522	pfam01282 pfam00163,pfa cd00165	TIGR01018		1	1	1
arCOG00678 arCOG04131	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	Rrp4 RsmA	Exosome complex RNA-binding protein Rrp4, contains S1 and KH domains 16S rRNA A1518 and A1519 N6-dimethyltransferase RsmA/KsgA/DIM1	COG01097 COG00030	cd05789 pfam00398	TIGR00755		1	1	1
arCOG04246 arCOG01564	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	RtcB SelB	RNA 3'-P ligase, RtcB family protein Selenocysteine-specific translation elongation factor or SelB-II domain	COG01690 COG03276	cd03696	TIGR03073 TIGR00475		1		
arCOG01923 arCOG01952	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	SIK1 SUA5	RNA processing factor Prp31, contains Nop domain tRNA A37 threonylcarbamoyladenosine synthetase subunit TsaC/SUA5/YrdC	COG01498 COG00009	pfam01798 pfam01300	TIGR00057		1	1	1
arCOG04223 arCOG00056	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	SUI1 Tan1	Translation initiation factor 1 (eIF-1/SUI1) tRNA(Ser,Leu) C12 N-acetylase TAN1, contains THUMP domain	COG00023 COG01818	pfam01253 cd11567 pfam02926,pfa cd11718,cd061	TIGR01158 TIGR00342		1	1	
arCOG01630 arCOG01561	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	TdcF TEF1	Translation initiation inhibitor, yigF family Translation elongation factor EF-1 alpha, GTPase	COG00251 COG05256	pfam01042 cd00448 pfam00009,pfa cd01883,cd036	TIGR00004 TIGR00483			2	1
arCOG00401 arCOG01115	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	ThrS TiaS	Threonyl-tRNA synthetase tRNA(IIe2) 2-agmatinylcytidine synthetase; containing Zn-ribbon domain an	COG00441 COG01571	pfam00587,pfa cd00771,cd008 pfam08489,pfa cd04482	TIGR00418 TIGR03280		1		1
arCOG04176 arCOG01219	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	TIF6 TRM1	Translation initiation factor 6 (eIF-6) N2,N2-dimethylguanosine tRNA methyltransferase	COG01976 COG01867	pfam01912 cd00527 pfam02005	TIGR00323 TIGR00308		1	1	1
arCOG01018 arCOG00987	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	TrmJ TruB	tRNA C32,U32 (ribose-2'-O)-methylase TrmJ or a related methyltransferase Pseudouridine synthase	COG00565 COG00130	pfam00588 pfam08068,pfa cd02572	TIGR00050 TIGR00425		1	1	2
arCOG04252 arCOG04733	1	Translation, ribosomal structure and biogenesis Translation, ribosomal structure and biogenesis	TruD Tsr3	tRNA(Glu) U13 pseudouridine synthase TruD Ribosome biogenesis protein Tsr3, contains Fer4-like metal-binding domain	COG00585 COG02042	pfam01142 cd02577 pfam04034	TIGR00094		1	1	1
arCOG00541	1	Translation, ribosomal structure and biogenesis	YSH1	Predicted exonuclease of the beta-lactamase fold involved in RNA processin	COG01236	pfam00753,pfam10996,pfam0;	TIGR03675			4	
CELLULAR PI arCOG11012	D	ES AND SIGNALING Cell cycle control, cell division, chromosome partitioning	ATS1	Alpha-tubulin suppressor and related RCC1 domain-containing proteins	COG05184	pfam13540.pfam00415.pfam00	415.ofam00415.ofam00415.ofam00415			1	2
arCOG02201 arCOG05007	D	Cell cycle control, cell division, chromosome partitioning Cell cycle control, cell division, chromosome partitioning	FtsZ Maf	Cell division GTPase Nucleotide-binding protein implicated in inhibition of septum formation	COG00206 COG00424	pfam00091,pfa cd02201 pfam02545 cd00555	TIGR00065 TIGR00172	5		1	1
arCOG03061 arCOG00586	D	Cell cycle control, cell division, chromosome partitioning	MreB Soi	Actin-like ATPase involved in cell morphogenesis	COG01077	pfam06723 cd10227 pfam01656 cd02042	TIGR00904 TIGR03453			2	3
arCOG02416 arCOG02070	N	Cell motility Cell motility		Pilin/Flagellin, FlaG/FlaF family S-laver protein. possibly associated with type IV nill like system	COG03430 COG01361	pfam07790		4		\vdash	2
arCOG01829	N	Cell motility Cell motility	FlaB	Archaeal flagellins Archaeal flagellins	COG01681			•		1	1
arCOG01822 arCOG04148	N N N	Cell motility Cell motility Cell motility	FlaH	ATPase involved in biogenesis of archaellum Archaellum assembly protein 1 archaellum	COG02874	pfam06745 cd01394	TIGR03881				1
arCOG02298	N	Cell motility	FlaK/Pul	OPeptidase A24A, prepilin type IV ATDase involved in archaelium fall bissuetosis	COG01933	pfam01478,pfam06847	TIG903810		1		1
arCOG01817 arCOG01410	M	Cell motility Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	AglA	A Pase involved in archaellum/pili biosynthesis Glycosyltransferase Dedited and an algorithm of the second se	COG00630 COG00438	pfam00437 cd01130 pfam13579,pfa cd03794	TIGR03819 TIGR02149	9	2		1
arCOG00118 arCOG00894	M	Cell wair/membrane/envelope biogenesis Cell wair/membrane/envelope biogenesis	-	Glycosyl transferase family 2	COG00399 COG00463	pfam01041 cd00816 pfam00535 cd04179	TIGR04182	5		1	1
arCOG00666 arCOG01568	M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	GCD1 MscS	N-acetylglucosamine-1-phosphate uridyltransterase Small-conductance mechanosensitive channel	COG01208 COG00668	ptam00483,pta cd04181,cd056 pfam00924	TIGR03992	4		1 5	1
arCOG01411 arCOG00252	M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	RfaG WecC	Glycosyltransferase UDP-N-acetyl-D-mannosaminuronate dehydrogenase	COG00438 COG00677	pfam13439,pfa cd03801 pfam03721,pfa cd05266	TIGR03999 TIGR03026	4		3	1
arCOG03336 arCOG01403	M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	- AgiL	Surface protein containing fasciclin-like repeats Glycosyltransferase	COG02335 COG00438	pfam02469 pfam13439,pfa cd03804		3		5	2
arCOG02209 arCOG00899	M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	AgIR AgID2	MATE family membrane protein, Rfbx family Predicted flippase	COG02244 COG00392	cd13128 pfam03706	TIGR00374	3	1		
arCOG00668 arCOG06759	M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	GCD1 RfaG	Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysacch Glycosyltransferase	COG01208 COG00438	pfam00483,pfa cd04181,cd056	TIGR03992	2			
arCOG01222 arCOG03335	M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	TagD -	Cytidylyltransferase fused to conserved domain of DUF357 family Surface protein containing fasciclin-like repeats	COG00615 COG02335	pfam01467 cd02170 pfam02469	TIGR02199	2			
arCOG03199	м	Cell wall/membrane/envelope biogenesis	AgIH/Rfe	UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylgluo	COG00472	pfam00953 cd06856	TIGR00445	1			
aic0002820	M	Cell wall/membrane/envelope biogenesis	MurE	UDP-N-acetylmuramyl tripeptide synthase	COG00769	pfam08245,pfam02875	TIGR01082	1			
arCOG02820 arCOG04827 arCOG01369	M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	MurE TagB WcaG	UDP-N-acetylmuramyl tripeptide synthase Glycosyl/glycerophosphate transferase involved in teichoic acid biosynthesis Nucleoside-diphosphate-sugar epimerase	COG00769 COG01887 COG00451	pfam08245,pfam02875 pfam04464 pfam01370 cd05234	TIGR01082 TIGR01179	1 1 1 1	1	2	4
arCOG02820 arCOG04827 arCOG01369 arCOG01392 arCOG02488	M M M M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	MurE TagB WcaG WecB	UDP-N-acetylmuramyt tripeptide synthase Glycosyl/gyrecophosphate transferase involved in teichoic acid biosynthesis Nucleoside-diphosphate-sugar epimerase UDP-N-acetylgucosamine 2-epimerase Cell surface proteina, a component of a putative secretion system	COG00769 COG01887 COG00451 COG00381 SC.00293	pfam08245,pfam02875 pfam04464 pfam01370 cd05234 pfam02350 cd03786	TIGR01082 TIGR01179 TIGR00236	1 1 1 1 1	1	2	4
arCOG02820 arCOG04827 arCOG01369 arCOG01392 arCOG02488 arCOG01385 arCOG02532	M M M M M M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	MurE TagB WcaG WecB - -	UDP-N-sext/imuramy trippetde synthate Grocov/gl/kyterophosphate transferase modeval in teichoic acid biosynthesis Nucleoside-diplosophate-sugar enjimerase DDP-N-acetr/gl/sucosamine 2-epiimerase GR surface protein, a component of a putative secretion system GR/coxyl transferase family 2 Secreted Beta-propeiler regeat protein fused to CARDB-like adhesion modul	COG00769 COG01887 COG00451 COG00381 SC.00293 COG00463 COG00463	pfam08245,pfam02875 pfam04464 pfam01370 cd05234 pfam02350 cd03786 pfam00535 cd02522 pfam007305	TIGR01082 TIGR01179 TIGR00236 TIGR04283 TIGR04213	1 1 1 1	1	2	4
arCOG02820 arCOG04827 arCOG01369 arCOG01392 arCOG02488 arCOG01385 arCOG02532 arCOG07560 arCOG07560	M M M M M M M M	Cell wal/imembrane/envelope biogenesis Cell wal/imembrane/envelope biogenesis	MurE TagB WcaG WecB - - - - - -	UDP-N-acetylinuramyl tripepted synthae Glycoxyl gykerophosphate transferase micelyde in teichola add biosynthesi Nucleoside dighosphate-sugar egymerase UDP-N-acetyl glucosamie 2-epimerase Glycosy transferase family 2 Secreted Beta-popeller repeat protein frused to CARDB-like adhesion modu Cell surface protein Glycosyl transferase family 2	COG00769 COG01887 COG00451 COG00381 SC.00293 COG00463 COG003291 COG00463	pfam08245,pfam02875 pfam01464 pfam01370 cd05234 pfam02350 cd03786 pfam00535 cd02522 pfam07705	TIGR01082 TIGR0129 TIGR00236 TIGR04283 TIGR04213 TIGR04182	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1	2	4
arCOG02420 arCOG04827 arCOG01369 arCOG01392 arCOG02488 arCOG01385 arCOG02532 arCOG02532 arCOG02538	M M M M M M M M M	Cell wall/membrane/newlope biogenesis Cell wall/membrane/newlope biogenesis	MurE TagB WcaG - - - - - - - - -	UDP-N acetylinuramyl tripeptde synthae Glycoxyl tytergrouppolate transforsa microlwei in teichola acid biosynthesis Okciosyl tytergrouppolate transforsa microlwei in teichola acid biosynthesis Dell surface protein, a component of a putative secretion system Gyrcoyr transforsa family. 2 Secreted Beth-propeler repeat protein fueed to CARDB-like adhesion modul Cell surface protein Gyrcoyr transforsas family. 2 Cell surface protein, a component of a putative secretion system Secreted Dell-propent of a putative secretion system Secreted protein, with PRO repeat Omain	COG00769 COG01887 COG00451 COG00381 SC.00293 COG00463 COG00463 COG00463 COG00463 COG00463	pfam02825,pfam02875 pfam01370 cd05234 pfam01370 cd05234 pfam02350 cd03786 pfam00355 cd02522 pfam07705 pfam07705 pfam00335 cd06442 pfam02369 pfam02369	TIGR01082 TIGR01179 TIGR02286 TIGR04283 TIGR04213 TIGR041182 TIGR041182		1	2	4 2 2 1 1 1 1
arCOG02420 arCOG01369 arCOG01369 arCOG01392 arCOG02488 arCOG02532 arCOG02532 arCOG02535 arCOG02538 arCOG02538 arCOG02538	M M M M M M M M M M M	Cell wal/inverbrane/erwebge biogenesis Cell wal/inverbrane/erwebge biogenesis	MurE TagB WcaG - - - - - - - - - - - - - - - - - - -	UP-N-acetylinoramy trippetde synthase Glycovyl tytergrouppotate transforsa microlwed in teichoic acid biosynthesis Glycovyl tytergrouppotate transforsa microlwed in teichoic acid biosynthesis Coll surface protein a component of a putative secretion system Glycovyl transferase family 2 Socried Bate-propeller repeat protein fused to CA808-like adhesion modul Cell surface protein Glycovyl transferase family 2 Gell surfaces protein a component of a putative secretion system Socreted network PRO repeat domain Glycovyl transferase family 2 Giogaacchavitransferase family 2	COG00769 COG01887 COG00451 COG00451 COG00463 COG00463 COG00463 COG00463 COG00463 COG00463 COG00463 COG00463	pfam08245,pfam02875 pfam04464 pfam01370 cd052314 pfam03505 cd05235 cd05235 cd05252 pfam02535 cd05252 pfam02535 cd05252 pfam02535 cd06442 pfam02536 pfam02535 cd04179	TIGR01082 TIGR01179 TIGR0236 TIGR04283 TIGR04283 TIGR04182 TIGR04182 TIGR04513 TIGR04514	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1	2	4 2 2 1 1 1 1 2
arCOG04227 arCOG04827 arCOG01369 arCOG01392 arCOG01385 arCOG02532 arCOG02532 arCOG02538 arCOG02487 arCOG02538 arCOG02538 arCOG00896 arCOG00565 arCOG00664	M M M M M M M M M M M M	Cell wal/imembrane/envelope biogenesis Cell wal/imembrane/envelope biogenesis	MurE TagB WcaG VecB - - - - - - - - - - - - - - - - - - -	UPP-N accellmoramy trippetde synthase Glocovoly/dycerophonate transformation involve in technic acid biosynthesis Rockostade diplosophate-sugar epimerase Cell surface protein, a component of a putative secretion system Oxyooyt transformate family 2 Secreted Beta-propeller repeat protein fused to CARDB-like adhesion modul Cell surface protein Oxyooyt transformate family 2 Cell surface protein Oxyooyt transformate family 2 Cell surface protein do a putative secretion system Secreted protein, a component of a putative secretion system Secreted protein protein Secretion Secret	COG00769 COG01887 COG00451 SC.00293 COG00463 COG00463 COG00463 COG00463 COG00291 COG00463 COG01287 COG01287 COG01209	plan08285,glam024275 plan01464 glam01470 plan012370 plan012370 glam02385 glam02375 plam02375 plam02375 glam02385 plam02386	TIGR01082 TIGR01082 TIGR0129 TIGR0226 TIGR04283 TIGR04283 TIGR04182 TIGR04182 TIGR04514 TIGR0454 TIGR045 TIGR0454 TIGR045 TIGR045 TIGR045 TIGR04 TIGR045 TIGR045 TIGR045 TIGR0		1	2 1 1 1	4 2 2 1 1 1 1 2 1 1 2 1 1 1
arCOG04827 arCOG04827 arCOG01392 arCOG01392 arCOG01392 arCOG01385 arCOG02532 arCOG02532 arCOG02538 arCOG02538 arCOG025385 arCOG00253 arCOG00253 arCOG0253 arCOG0253	M M M M M M M M M M M M M M M M M M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	MurE TagB WcaG - - - - - - - - - - - - - - - - - - -	UDP-N accell/normally tripeptide synthase Glycoxy0/glycerpolycophate transformation involved in technic acid biosynthesis Roucleoside-diplosophate-sugar epimerase Cell surface protein, a component of a putative secretion system Oryoxy0 transformate family 2 Secreted Beta-propeller repeat protein fused to CARDB-like adhesion modul Cell surface protein Ocycoy) transformate family 2 Generate protein, a component of a putative secretion system Ocycoy transformate family 2 Generate protein, a component of a putative secretion system Secreted protein, a secret protein system Secreted protein, a secret protein system Secret protein system secret protein system Secret protein secret protein system Secret protein secret protei	COG00769 COG01887 COG00451 COG00451 COG00451 COG00463 COG03291 COG00463 COG03291 COG00463 COG01287 COG01287 COG01209 COG01209 COG01208	plan08285,glam02875 plan01464 plan01470 plan01470 plan01470 plan01470 plan01470 plan01470 plan01470 plan01470 plan01570 plan01571 plan01571 plan01571 plan01571 plan01571 plan01571 plan01571 plan01571 plan0258 plan00258 plan00258 <td>TIGR01082 TIGR01082 TIGR01236 TIGR0236 TIGR04283 TIGR04182 TIGR04182 TIGR04182 TIGR04514 TIGR0454 TIGR0454 TIGR0454 TIGR05026 TIGR06283 TIGR05026 TIGR0292</td> <td></td> <td></td> <td></td> <td>4 2 2 1 1 1 1 2 2 1 1 1 1 1 1 1</td>	TIGR01082 TIGR01082 TIGR01236 TIGR0236 TIGR04283 TIGR04182 TIGR04182 TIGR04182 TIGR04514 TIGR0454 TIGR0454 TIGR0454 TIGR05026 TIGR06283 TIGR05026 TIGR0292				4 2 2 1 1 1 1 2 2 1 1 1 1 1 1 1
arCoG04227 arCCG01369 arCCG01392 arCCG01392 arCCG01392 arCCG01385 arCCG02487 arCCG02487 arCCG02487 arCCG02487 arCCG02538 arCCG00253 arCCG005365 arCCG00553 arCCG00251 arCCG00664 arCCG00057 arCCG0075	M M M M M M M M M M M M M M M M M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	MurE TagB WeaG - - - - - - AgIB AgIF/Rft AgIM/U 4 AgIO GCD1 GImS Gmd	UpP -N actifyimuramy't trigetide synthase Grocov/gi/egrocophate transformation involved in technic acid biosynthesis Oncovol/gi/egrocophate transformate involved in technic acid biosynthesis Cell surface grotein, a component of a putative secretion system Grocopit transformate family 2 Secreted Bitta: propeller repeat protein fused to CABDE-like adhesion modul Cell surface protein, a component of a putative secretion system Grocopit transformate family 2 Cell surface protein, a component of a putative secretion system Grocopit transformate family 2 Cell surface protein, a component of a putative secretion system Grocopit transformate family 2 Grocopit transformate family 2 Glocopit-phosphate uridytransformate Glocopit transformate family 2 Nucleoside diplosphate-sugar proposphorylase involved in lipopolysacch Glucopanine 6-phosphate system Glucopitae	COG00769 COG01887 COG00451 COG00451 COG00381 COG003291 COG00463 COG00463 COG00463 COG00291 COG00463 COG00291 COG00463 COG01208 COG00404 COG004208	plano8458,plano242875 plano14644 plano12370 cd95234 plano12370 cd95234 plano12370 cd95234 plano12370 cd95234 plano12370 cd95234 plano12355 cd05422 plano12351 cd06442 plano2535 cd04179 plano12351 cd04181 plano3235 cd07418 plano3235 cd0784, plano335 plano3235 cd0784, plano354, plano1379 plano13351 cd0784, plano354, plano1379	TGR01082 TGR01082 TGR0236 TGR0236 TGR0423 TGR0423 TGR04182 TGR04182 TGR04182 TGR04514 TGR05326 TGR04524 TGR05326 TGR04528 TGR08281 TGR08281 TGR08281 TGR08135 TGR08135 TGR08135 TGR081472		1		4 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
arCoG04227 arCQ601369 arCOG01369 arCOG01389 arCOG01385 arCOG02382 arCOG02380 arCOG02532 arCOG02532 arCOG02538 arCOG02538 arCOG02538 arCOG02538 arCOG02538 arCOG02538 arCOG005365 arCOG0061381 arCOG00365 arCCG00061381 arCOG00373 arCOG01373 arCOG01373	M M M M M M M M M M M M M M M M M M	Cell wall/membrane/meekpe biogenesis Cell wall/membrane/meekpe biogenesis	MurE TagB WecB - - - - - AgIF/Rft AgIF/Rft AgIB/Rft AgIB/U AgIO GCD1 GCD1 GIMS Gmd LoIE MurD	UGP -4 secilymorany fripegible synthase Glycovyl Zyterophysite transformation involved in technic acid biosynthesis Glycovyl Zyterophysite transformation involved in technic acid biosynthesis Cell surface protein, a component of a putative secretion system Glycovyl transformate finity 2 Socreted Bitra-progelier repeat protein fused to CARDB-like adhesion modul Glycovyl transformate finity 2 Glycovyl transformate finity 2 Muclosidie dlycovylate-signer profosphorylase Involved in Ilipopolysacch Glycovyl transformate family 2 Glycovyl transformate family 2 Glycovyl transformate finity 2 Glycovyl transformate finity 2 Glycovyl transformate finity 2 Glycovylate finity 2	COG00769 COG01887 COG00451 COG00451 COG00381 SC.00293 COG00463 COG00463 COG00463 COG01287 COG01287 COG01287 COG01209 COG00463 COG01208 COG00463 COG01208 COG00463 COG01208 COG00463 COG01208	plano8458,plano22875 plano14644 plano12370 cd95234 plano12370 cd95234 plano12370 cd95234 plano12370 cd95234 plano12370 cd95234 plano12375 cd05232 plano12351 cd06422 plano12353 cd064217 plano2353 cd04179 plano13721 plano0383 plano3721 plano0384 plano0335 cd00741 plano0335 cd00784 plano0335 cd00741 plano0335 cd00784 plano0335 cd00784 plano0335 cd00784 plano0335 cd00784 plano0335 cd00784 plano0335 cd00784 plano0336 cd00784 plano1370 plano2586 plano1370 plano2587	TIGR01082 TIGR01082 TIGR0236 TIGR0236 TIGR04213 TIGR04182 TIGR04182 TIGR04182 TIGR0454 TIGR0454 TIGR0454 TIGR0454 TIGR0454 TIGR05026 TIGR04283 TIGR03026 TIGR04283 TIGR0392 TIGR01155 TIGR0145 TIGR0472 TIGR0472				4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
a CCG0.42.7 arCCG01369 arCCG01369 arCCG01369 arCCG01369 arCCG01385 arCCG01385 arCCG01385 arCCG02482 arCCG02487 arCCG02487 arCCG02487 arCCG02487 arCCG02487 arCCG02487 arCCG01381 arCCG00481 arCCG01381 arCCG01381 arCCG01381 arCCG01381 arCCG01373 arCCG02821 arCCG02313 arCCG02531 arCCG02531	M M M M M M M M M M M M M M M M M M M	Cell wall/membrane/envelope biogenesis Cell wall/membrane/envelope biogenesis	MurE TagB WcaG WecB - - - - - - - - - - - - - - - - - - -	UGP -N actifymuramyl trippetde synthase Grocov/givegrouppontate transforma truncked in technic acid biosynthesis Rockov/givegrouppontate transforma truncked in technic acid biosynthesis Grocov/givegrouppontate transforma truncked in technic acid biosynthesis Cell surface protein, a component of a putative secretion system Givernot trunsformate family 2 Cell surface protein, a component of a putative secretion system Givernot trunsformate family 2 Cell surface protein, a component of a putative secretion system Givernot trunsformate family 2 Cell surface protein, a component of a putative secretion system Givernot trunsformate family 2 Cell surface protein, a component of a putative secretion system Givernot trunsformate family 2 Cell cell surface protein a component of a putative secretion system Givernot trunsformate family 2 Cell cell surface protein a cell substitute Givernot trunsformate family 2 Culcosanine 6-phosphate sugar proprophosphorytase involved in lipopotysacch Glucosanine 6-phosphate system Glucosanine 6-phosphate system Gl	Coc00769 COC001887 COC001887 COC001881 COC00381 COC00463 COC00463 COC00463 COC00463 COC001287 COC00463 COC001287 COC00128 COC001208 COC001208 COC001208 COC001208 COC001208 COC001208 COC001208 COC001208 COC004531 COC004531 COC004531 COC004531 COC00755 COC0	plano8455,glm/02875 plano1464 plano12370 c05234 plano12370 c05234 plano12370 c05234 plano02355 c00542 plano0335 c00642 plano0335 c006442 plano2516,glm<3302,opfam1	TGR01082 TGR01082 TGR0236 TGR0236 TGR0238 TGR04213 TGR04213 TGR04182 TGR04182 TGR04182 TGR04514 TGR0455 TGR04524 TGR03026 TGR04283 TGR03026 TGR04283 TGR03026 TGR04283 TGR0305 TGR0472 TGR0472 TGR02212 TGR0475 TGR0222 TGR04063				4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
a CCG02421 arCQ604227 arCQ601369 arCQ601369 arCQ601385 arCQ601385 arCQ601385 arCQ602487 arCQ602538 arCQ602487 arCQ602538 arCQ602487 arCQ60263 arCQ600463 arCQ600463 arCQ600463 arCQ600463 arCQ600453 arCQ600453 arCQ600453 arCQ600453 arCQ600453 arCQ600453 arCQ602451 arCQ604451 arCQ60451 arCQ60451 arCQ60451 arCQ60451 arCQ60451 arCQ60451 arCQ60451 arCQ60451 arCQ60451 arCQ60451 arCQ60451 a	M M M M M M M M M M M M M M M M M M M	Cell wall/membrane/meekpe biogenesis Cell wall/membrane/meekpe biogenesis	Mur/E TagB WcaG - - - - - - - - - - - - - - - - - - -	UGP -N actifymuramy fripegible synthase Glocovol ýkregronychaite trundresa michovile in technic acid biosynthesi Rockovol (Xergerophosite trundresares incolved in technic acid biosynthesi Glocovol Xergerophosite trundresares Cell surface acrossine - acymicrase Glosovol trundresares family 2 Secreted Belta-propeller repeat portein fused to CARDB-like adhesion modul Cell surface acrossine, a component of a putative secretion system Glycovoj trundresares family 2 Cell surface acrossine, a component of a putative secretion system Glycovoj trundresares family 2 Cell surface acrossine, with PKD repeat domain Glycovoj trundresare family 2 Accessite protein, a component of a putative secretion system Glycovoj trundresare family 2 Accessite - phosphate undytrundresare UGP-glucos C- dedyndregenase Glycovoj trundresare family 2 Auccessite - floposphate system prophosphorytase involved in lipopotysacch Glucosamine 6-phosphate system prophosphorytase livolved in lipopotysacch Glucosamine 6-phosphate system prophosphorytase involved in lipopotysacch Glucosamine 6-phosphate system prophosphorytase livolved in lipopotysacch Glucosamine 6-phosphate system prophosphorytase livolved in lipopotysacch Glucosamine 6-phosphate system prophosphorytase livolved in lipopotysacch Glycosystrandresase Glycosystrandresase Glycosystrandresase Glycosystrandresase	CCG00769 CCG001887 CCG001887 CCG00381 CCG00463 CCG00463 CCG00463 CCG00463 CCG00463 CCG00291 CCG00463 CCG00291 CCG00463 CCG01209 CCG01209 CCG01208 CCG00463 CCG0045 CCG005 C	plan08458,plan042875 plan01464 plan012370 c05234 plan012370 c05234 plan012370 c05234 plan02370 c02524 plan02375 c002524 plan02375 c002524 plan02375 c005442 plan02375 c005442 plan02375 c004179 plan02375 c004181 plan03751 c00784, plan01 plan03751 c00784, plan01 plan03751 c00784, plan01 plan03751 c00784, plan01 plan00385 c00784, plan01 plan00385 c00784, plan01 plan00385 c00784, plan01 plan00385 c00784, plan01 plan00385, plan01 c00784, plan01 plan01370, plan02875 plan02875 plan01382, plan02874 c003794 plan13439, plan042875 c003794 plan13439, plan042875 c003794 plan13439, plan04287 c003794 plan03974 c003794 plan013974	TGR01082 TGR01082 TGR0236 TGR0235 TGR0235 TGR04213 TGR04182 TGR04182 TGR04182 TGR04182 TGR04182 TGR0454 TGR05026 TGR04233 TGR0454 TGR05026 TGR04223 TGR01155 TGR0423 TGR04175 TGR04230 TGR0477 TGR04216 TGR04221 TGR04063 TGR0406	9		2 1 1 1 1 1 1 1 1 1 1 1 1 1 1	4 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1
a CCG01421 arCCG01367 arCCG01367 arCCG01367 arCCG01367 arCCG01285 arCCG02532 arCCG02532 arCCG02532 arCCG02532 arCCG02532 arCCG02532 arCCG02533 arCCG02533 arCCG02533 arCCG02533 arCCG02533 arCCG01361 arCCG02531 arCCG02531 arCCG02541 arCCG0140 arCCG02541 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173 arCCG0173	M M M M M M M M M M M M M M M M M M M	Cell wall/membrane/mevlope biogenesis Cell wall/membrane/mevlope biogenesis	Murc TagB WccaG - - - - - - - - - - - - - - - - - - -	UGP -N actifymuramyl trippetde synthase Grocov/gi/ecroposhate transforma tonovled in technica act biosynthesi Nucleosiae diplosphate-sugar epimerase UGP -N actegi/guoamine 2-epimerase Cell surface protein, a component of a putative secretion system Givposit transfersa family 2 Secreted Bita-propeller repeat portain fused to CARDB-like adhesion modul Cell surface protein, a component of a putative secretion system Givposit transfersa family 2 Cell surface protein, a component of a putative secretion system Givposit transfersa family 2 Cell surface protein, a component of a putative secretion system Givposit transfersa family 2 Compacacharythransferase Cell surface protein system Givposit transferase family 2 Compacacharythransferase Givposit transferase family 2 Calcocaentine 6-phosphate synthetase CoP-O-ananose Cellevitates Givposit transferase family 2 Calcocaentine 6-phosphate synthetase Givposit transferase Givpositransferase G	CCG00769 CCG00187 CCG00187 CCG00381 CCG00381 CCG00463 CCG00463 CCG00291 CCG00463 CCG00291 CCG00463 CCG00287 CCG004287 CCG004287 CCG004287 CCG00408 CCG00408 CCG00408 CCG00408 CCG00408 CCG00408 CCG00438 CCG00463 CCG0045 CCG005 CCG005 CCG005 CCG	plano8458,plano22875 plano14644 plano12370 cd05234 plano12370 cd05234 plano12370 cd05234 plano12370 cd05234 plano10335 cd05242 plano0335 cd06422 plano0335 cd06423 plano0335 cd06417 plano0335 cd04175 plano0335 cd07141 plano0335 cd0714 plano0335 cd0714 plano0335 cd0714 plano0335 cd0714 plano0353 cd0714 plano0353 cd0714 plano0353 cd0714 plano0354 cd0734 plano1370 plano254 plano1389 cd03744 <	TGR01082 TGR01179 TGR0236 TGR0236 TGR0236 TGR0238 TGR0231 TGR04213 TGR04182 TGR04182 TGR04182 TGR04184 TGR0454 TGR0454 TGR02026 TGR04233 TGR0454 TGR02026 TGR02232 TGR01155 TGR01155 TGR01155 TGR01252 TGR02212 TGR02188 TGR02037 TGR0212 TGR021 TGR021 TGR021 TGR021 TGR0212 TGR0	9 5 5	1 1 1 1 1 1 1 1 1 1 2 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
arCoG0422 arCOG01369 arCOG01369 arCOG01369 arCOG01369 arCOG0232 arCOG02488 arCOG02532 arCOG02487 arCOG02532 arCOG02487 arCOG02537 arCOG02537 arCOG02537 arCOG02537 arCOG02537 arCOG01373 arCOG01405 ar	M M M M M M M M M M M M M M M M M M M	Cell wall/membrane/meekpe biogenesis Cell wall/m	Murt2 TagB Wcca6 Wcca6 - - - - - - - - - - - - - - - - - - -	UGP -N actifymuramy frippetde synthase Grocov/gi/kergerospicate transforma knowled in technica act biosynthesis Nucleoside diplosphate-sugar epimerase Cell surface protein, a component of a putative secretion system Grocovi transforma family 2 Secreted Bits-propeller repeat portain fused to CARDB-like adhesion modul Cell surface protein, a component of a putative secretion system Givposit transforma family 2 Cell surface protein, a component of a putative secretion system Givposit transforma family 2 Disposacharytransforma family 2 Disposachary	CCG00769 CCC001887 CCG00451 CCG00451 CCG003451 CCG00463 CCG003291 CCG00463 CCG00291 CCG00463 CCG00291 CCG00463 CCG002291 CCG00463 CCG002291 CCG00463 CCG002291 CCG00463 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG0045 CCG005 CCG005 CCG005 CCG005 CCG005 CCG0	#ano8456,g/m/02875 #ano8456,g/m/02875 #ano8230 cd05234 #ano8230 cd05234 #ano8230 cd05234 #ano8230 cd05242 #ano8230 cd05242 #ano8230 cd06422 #ano8230 cd0642 #ano8230 cd0642 #ano8230 cd06412 #ano8230 cd04131 #ano8335 cd00714 #ano8335 cd00714 #ano8335 cd00714 #ano8355 cd0714 #ano8355 cd0714 #ano8355 cd0714 #ano8355 cd0734 #ano8355 cd0734 #ano8355 cd03734 #ano1875 cd03734 #ano1875 cd03734 #ano1875 cd03734 #ano1875 cd03734 #ano1875 cd03734 #ano1875 cd03734 #ano1876 cd03734 #ano1876 cd03734 #ano1876 cd03	TGR01082 TGR01179 TGR0236 TGR0236 TGR0236 TGR0238 TGR04213 TGR04182 TGR04182 TGR04182 TGR04184 TGR04184 TGR04184 TGR0453 TGR0453 TGR0453 TGR0453 TGR0453 TGR0453 TGR0453 TGR04653 TGR0465 TGR04	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 2 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
a (Cool) a (M M M M M M M M M M M M M M M M M M M	Cell wall/membrane/mevbge biogenesis Cell wall/m	Murc TagB WccaG WccaG - - - - - - - - - - - - - - - - - - -	UGP-N-acetylinuramyl trippetde symbase Rockovdylkyterophositae transforma lenkovde in technica cid biosynthesi Nucleoside diphosphate-suger epimerase Cell surface protein, a component of a patative secretion system Grycosyl transforma family 2 Secreted Bits-propeller repeat portein fused to CARDB-like adhesion modul Cell surface protein, a component of a patative secretion system Grycosyl transforma family 2 Cell surface protein, a component of a patative secretion system Grycosyl transformase family 2 Grycosyl transformase family 2 Digosacharbyltransformase tambia Secreted protein, with PKO repeat domain Grycosyl transformase family 2 Digosacharbyltransformase family 2 Digo-D-mannoc dehyltratise Calcosamine 6-phosphate synthetise Coll-O-Mannoc dehyltratise Chycosyltransformase family 2 Chycosyltransformase Chycosyltransformase Chycosyltransformase Chycosyltransformase Chycosyltransformase Chycosyltransformase Chycosyltransformase furthers, family exonuclease Na-driven multifuting transport system, ATPase component CALSHe desiscation related protein ZAL-He desiscation related protein	CCG00769 CCG0187 CCG0187 CCG0187 CCG00451 CCG00351 CCG00351 CCG00463 CCG03291 CCG00463 CCG03291 CCG00463 CCG01292 CCG01267 CCG01209 CCG01209 CCG01209 CCG01208 CCG01208 CCG01208 CCG0451 CCG0551 CCG0551 CCG0551 CCG0551 CCG0552 CCG05	plano8456, plano22875 plano8456 plano8456 plano82370 c05234 plano82370 c0564 plano82351 c006432 plano82351 c006432 plano82351 c006432 plano8353 c0604377 plano8353 c0604377 plano8353 c0604377 plano8353 c0604377 plano8353 c0604377 plano8353 c060761 plano8353 c060761 plano8353 c060761 plano8353 c060761 plano8355 plano8356 plano8355 plano8356	TGR0082 TGR0179 TGR0236 TGR0236 TGR0236 TGR0236 TGR0238 TGR04182 TGR04185 TGR04185 TGR04185 TGR04185 TGR04185 TGR04185 TGR04182 TGR04182 TGR04184 TGR04188 TGR04053 TGR04053 TGR04053 TGR04188 TGR04053 TGR0405 TG	9 5 5 4 4 3 3	1 1 1 1 1 1 1 1 1 1 1 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
acCooka27 arCooka27 arCooka27 arCooka27 arCooka28 arCook	M M M M M M M M M M M M M M M M M M M	Cell wall (membrane) envelope biogenesis Cell wall (membrane) Deferse mechanisms Deferse mechanisms Deferse mechanisms	Murt2 TagB Wcc6 Wcc8 - - - - - - - - - - - - - - - - - -	UGP-N actifymaramy trippetde syntase GrocovlyKreemposhate transformation toweld in technica act biosynthesis Nucleoside diphosphate-sager egimerase Cell surface protein, a component of a patative secretion system Grocovly transformate family 2 Secreted Betts-propeller repeat portein fused to CARDB-like adhesion modul Cell surface protein, a component of a patative secretion system Grocovly transformate family 2 Cell surface protein, a component of a patative secretion system Grocovly transformate family 2 Grocovly transformate family 2 Orgosyntransformate family 2 Digosactrahythrandferses membrans suburit Microsoft - phosphate system functional Calcovanie - phosphate system functional Calcovanie - phosphate system Acteside diphorabate suger prophosphorytase involved in lipopolysacch Glocovanime - phosphate system Diphorational controls and the phosphorabate lipopolysacch Calcovanie - phosphate system functional lipoportein release, permase comp UDP-N-acetyfinansformate lipoportein release, permase comp UDP-N-acetyfinansformate lipoportein release, permase comp UDP-N-acetyfinansformate lipoportein release, permase comp UDP-N-acetyfinansformate lipoportein release, permase component CAC-spen militikug transport system, ATPase component	CC600769 CC600187 CC600451 CC600451 CC600451 CC600451 CC600451 CC600453 CC60055 CC6005 CC60055 CC60055 CC60055 CC60055 CC60055 CC60055 CC60055 CC60055 CC60055 CC6005 CC60055 CC60055 CC60055 CC60055 CC60055 CC6005	afano.8456, plano.82875. afano.8456, plano.82875. afano.82570. cd05234. afano.82570. cd05234. afano.82570. cd05234. afano.82570. cd05234. afano.82570. cd05234. afano.82570. cd05434. afano.82570. cd05434. afano.82570. cd05434. afano.82570. cd05437. afano.82570. cd041377. afano.82570. cd04137. afano.82570. cd0414. afano.82570. cd0414. afano.82570. cd03134. afano.82570. <	TGR0082 TGR0179 TGR0236 TGR0236 TGR0236 TGR0236 TGR0238 TGR04182 TGR04185 TGR04185 TGR04185 TGR04185 TGR04185 TGR04185 TGR04185 TGR04185 TGR04182 TGR04182 TGR0418 TGR04	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1		4 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1
arcD604827 arcD601497 arcD601497 arcD601192 arcD601192 arcD601192 arcD601192 arcD601192 arcD601195 arcD601195 arcD60247 arcD60247 arcD60247 arcD602487 arcD600447 arcD600447 arcD600191 arc	M M M M M M M M M M M M M M M M M M M	Cell wall (membrane) envelope biogenesis Cell wall (membrane) envelope biogenesis Deferes mechanisms Deferes mechanisms Deferes mechanisms Deferes mechanisms Deferes mechanisms Deferes mechanisms Deferes mechanisms Deferes mechanisms Deferes mechanisms Deferes mechanisms	Murt2 TagB WccaG WccaG - - - - - - - - - - - - - - - - - - -	UGP -N acetylinuramyl trippetde synthase GrocovlyKierospotalte transformatione unovel in technica acid biosynthesi Nucleoside diphosphate-sugar epitienses Cell surface protein, a component of a putative secretion system Grocovly transformate family 2 Secreted Betts-propeller repeat portein fused to CARDB-like adhesion modul Cell surface protein, a component of a putative secretion system Grocovly transformate family 2 Cell surface protein, a component of a putative secretion system Grocovly transformate family 2 Cell surface protein, a component of a putative secretion system Corport transformate family 2 Digoscotruhyftransferses membrans subouit Microsoft - Bondente suboutit Digoscotruhyftransferses membrans suboutit Clacoscate - Joposphate urg/https://doi.org/like.involved in lipopolysacch Calcosater - Joposphate sugar pyrophosphorylase involved in lipopolysacch Calcoscater	CCG00769 CCG0187 CCG0451 CCG0451 CCG0451 CCG0451 CCG0451 CCG0451 CCG0451 CCG045291 CCG045291 CCG045291 CCG045291 CCG045291 CCG045291 CCG045291 CCG045291 CCG045291 CCG045291 CCG045291 CCG045291 CCG04531 CCG04547 CCG04547 CCG04547 CCG04547 CCG04547 CCG04547 CCG05547	plano8455,glam02875 plano12855,glam02875 plano12370 c005234 plano12370 c005234 plano12370 c005234 plano12370 c005234 plano12370 c005234 plano02351 c005424 plano12351 c005424 plano12351 c005424 plano12351 c005434 plano13551 c005434 plano03351 c00761 plano03351 c00714,00714,00714,00714,00714,00714,00714,00714,01410 plano1352,014,01700,014,01400 plano1355,014,01700,014,01400 plano1352,014,01700,014,01400 plano1255,0141,01700,014,01400 plano1354,014,012,014,01400,0140,0140,0140,0140,	TGR0082 TGR0119 TGR0236 TGR0236 TGR0236 TGR0236 TGR0238 TGR04182 TGR04182 TGR04182 TGR04182 TGR04182 TGR04182 TGR04182 TGR04184 TGR04182 TGR04182 TGR04182 TGR04182 TGR04185 TGR0418	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
arC0604827 arC0604827 arC0601497 arC0601392 arC0601392 arC0601392 arC0601392 arC0601385 arC0601385 arC0601385 arC0601385 arC0601385 arC0600487 arC0600986 arC0600487 arC0600986 arC06001381 arC0600986 arC06001381 arC0600057 arC06001381 arC0600157 arC06000	M M M M M M M M M M M M M M M M M M M	Cell wall/membrane/mevbge biogenesis Cell wall/m	Murt2 TagB WccaG WccsB - - - - - - - - - - - - - - - - - - -	UGP -N acetylinuramyl trippetde symbare Grocovyl Xyteromychaite transforma toweld in technica add biosynthesi Nucleoside diphosphate-sugar egylinerase Cell surface protein, a component of a putative secretion system Grycoyl transforsa family 2 Secreted Bits-propeller repeat portein fused to CARDB-like adhesion modul Cell surface protein, a component of a putative secretion system Glycoyl transforsa family 2 Cell surface protein, a component of a putative secretion system Glycoyl transforsa family 2 Cell surface protein, a component of a putative secretion system Glycoyl transforsa family 2 Digoscothulyftransforsase membrane subouit Glucose - Iphosphate unifyltransforsase Clocose - Honoposci turily 2 Alcedoside dipolognate sugar pyrophosphorylise involved in lipopolysach Glacose - Iphosphate unifyltransforsase Clocose - Indepolognate sugar pyrophosphorylise involved in lipopolysach Glacose - Honoposci and a subout Glacose - Honoposci and Allore and Allore and Allore and Allore Glacose - Honoposci and Allore and Allore and Allore Glacose - Honoposci and Allore and Allore and Allore Glacose - Honoposci and Allore and Allore and Allore and Allore Glacose - Honoposci and Allore and Allore and Allore and Allore Glacose - Honoposci and Allore and Allore and Allore and Allore and Allore Allor - Marketymine and Allore and Allore and Allore and Allore and Allore and Allore and Allore Allore - Marketymine and Allore a	CCG00769 CCG00187 CCG00451 CCG00451 CCG00451 CCG00453 CCG0055 CCG055 CCG05 CCG055 CCG055 CCG055 CC	plano8455,glm/02875 plano16464 plano12370 c005234 plano12370 c005234 plano02350 c005242 plano02351 c005424 plano02355 c005424 plano02355 c005424 plano02355 c005424 plano02355 c005424 plano02355 c005417 plano02355 c005417 plano02355 c005418 plano02355 c00744 plano02356 c00744 plano02356 c00744 plano0355 c00744 plano0356 c00774 plano0357 c003749 plano1370,plat c003740 plano1370,plat c003740 plano1370,plat c003740 plano1370,plat c003740 plano1370,plat c003740 plano1370,plat c003730 plano1370,plat c003730 plano00005 c013730 plano00005 c013730 plano00005	TicRR0.082 TicRR0.23 TicRR0.23 TicRR0.23 TicRR0.23 TicRR0.23 TicRR0.23 TicRR0.23 TicRR0.182 TicRR0.182 TicRR0.182 TicRR0.184 TicRR0.23 TicRR0.24 T	9 5 5 4 4 3 3 3 2 2 2 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
arC0604827 arC0604827 arC0601392 arC0601392 arC0601392 arC0601392 arC0601392 arC0601385 arC0601385 arC0601385 arC0601385 arC0602387 arC0602487 arC0602487 arC0602487 arC060057 arC060051 arC060051 arC060051 arC060051 arC060051 arC0600151	M M M M M M M M M M M M M M M M M M M	Cell will rembrane (newloge biogenesis Cell w	Mure TragB WecaG WecaG - - - - - - - - - - - - - - - - - - -	UGP N-acetylinouramyl trippetde synthase Occovyl tytercomposite transformate involved in technic acid biosynthesis Nucleoside diplocophate-sugar epilemase Cell surface protein, a component of a putative secretion system Givoyot transfersa family 2 Socreted Bits-propeller repart protein fused to CARDB-like alheidon modu Coll surface protein, a component of a putative secretion system Givoyot transfersa family 2 Gil surface protein, a component of a putative secretion system Givoyot transfersa family 2 Gil surface protein, a component of a putative secretion system Givoyot transfersa family 2 Oligoacchartyl mathemate subunt Givosof transfersa family 2 Oligoacchartyl transfersa family 2 Col-Net sectored family component Col-Net sectored family component Col-N	CC600769 CC600187 CC600451 CC600451 CC600451 CC600451 CC600451 CC600452 CC600547 CC6003467 CC60005 CC6003467 CC60005 CC6003467 CC60005 CC6003467 CC60005 CC	plano8455,plano22875 plano8456,plano22875 plano8230 cd05234 plano8230 cd05234 cd05234 plano8230 cd05234 plano8230 cd05234 cd05234 cd05234 cd05234 plano82305 cd05424 cd05424 cd05424 cd05424 plano82305 cd05432 cd05432 cd05432 cd05433 cd06433 cd064333 cd06433 cd064333 cd064333 cd064333 cd064333 cd064333 cd0643333 cd0643333 cd0643333 cd0643333 cd0643333 cd0643333 cd0643333 cd0643333 cd0643333 cd0663333 cd0663333 cd0663333 cd0663333 cd0663333 cd06633333 cd06633333 cd06633333 cd06633333 cd0663333 cd0663333	TIGR0082 TIGR0179 TIGR0236 TIGR0236 TIGR0236 TIGR0237 TIGR0238 TIGR04182 TIGR04083 TIGR0488 TIGR	9 5 5 4 4 3 3 3 2 2 2 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
arcO604827 arcO601497 arcO601192 arcO601192 arcO601192 arcO601192 arcO601192 arcO601192 arcO601195 arcO601185 arcO601185 arcO602187 arcO602487 arcO602487 arcO602487 arcO602487 arcO600481 arcO600181 arcO600181 arcO600181 arcO600181 arcO600191	M M M M M M M M M M M M M M M M M M M	Cell will rembrane (newloge biogenesis Cell will rembrane	MurrE TragB WrcaG WrecB - - - - - - - - - - - - - - - - - - -	UGP -N acetylinouramyl trippetide synthase Occovyl tytercomposite transforma environment in technic acid biosynthesis Nucleoside diplocophate-sugar enjimense Cell surface protein, a component of a putative secretion system Givpoyl transfersa family 2 Secreted Bits-propeller repeat protein fused to CARDB-like alhesion modu Cell surface protein, a component of a putative secretion system Givpoyl transfersa family 2 Gil surface protein, a component of a putative secretion system Givpoyl transfersa family 2 Gil surface protein, a component of a putative secretion system Givpoyl transfersa family 2 Oligoaccharbyl transfersa Givpoyl transfersa family 2 Cell Acet per transford transfersa Givpoyl transfersa family 2 Cell Scale (Second Second	CCG00769 CCG00187 CCG00451 CCG00451 CCG00451 CCG00451 CCG00453 CCG0055 CCCG055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG0055 CCG005	plano8455,plano22875 plano16464 plano12370 c005234 plano12370 c005234 plano12370 c005234 plano12370 c005234 plano10335 c006422 plano10335 c006422 plano12350 c006423 plano12350 c006423 plano12350 c006437 plano12350 c006417 plano12350 c00714 plano12350 c00714 plano13551 c00714 plano1355 c00714 plano13551 c00714 plano13552 c00714 plano13552 c00714 plano13552 c00714 plano13552 c00714 plano13552 c00714 plano13554 c00714 plano13554 c003744 plano13554 c003749 plano13554 c003749 plano13554 c003730 plano0055 c033230 plano00554 c033255	TIGR0082 TIGR0179 TIGR0236 TIGR0236 TIGR0236 TIGR04213 TIGR04182 TIGR04182 TIGR04182 TIGR04182 TIGR04182 TIGR04182 TIGR04182 TIGR04184 TIGR04182 TIGR04182 TIGR04182 TIGR04083 TIGR0408	9 5 5 4 4 3 3 3 2 2 2 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
anco604827 anco604827 anco601509 anco601192 anco601192 anco601192 anco601192 anco601192 anco601195 anco601195 anco602487 anco602487 anco602487 anco602487 anco602487 anco602487 anco602487 anco602487 anco602487 anco600451 anco600451 anco600191	M M M M M M M M M M M M M M M M M M M	Cell will nembrane (newloge biogenesis Cell will nembrane (newloge) Defense mechanisms Defense mechanis	Murt2 TragB Weca6 - - - - - - - - - - - - - - - - - - -	UGP -N acetylinouramyl trippetide synthase Occovyl tytercomposite transforma environment in technic acid biosynthesis Nucleoside diplocophate-sugar engineerase UGP -N acetylicosomite - Zepimerase UGP - UGP - Zepimerase UGP - Zepimerase	Coccorres Coccor	planol845,glm/m22875 planol845,glm/m22875 planol845,glm/m22875 planol2302 planol2312 planol2312 planol2312 planol2312 planol2312 planol2312 planol2312 planol2312 planol3132 planol3134 planol3143 planol3143 planol3143 planol3143 planol3143 planol3144 planol3144 planol3144 planol3144 planol3144 planol3144 planol3144 planol3144	TIGR0082 TIGR0179 TIGR0236 TIGR0236 TIGR0236 TIGR04213 TIGR04182 TIGR04063 TIGR04064 T	9 5 5 4 4 3 3 2 2 2 2 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	M M M M M M M M M M M M M M M M M M M	Cell wall (membrane (mevolop biogenesis Cell wall (membrane (mevolop biogenes	Mure 2 Tog88 Weca8 - - - - - - - - - - - - - - - - - - -	UP-N acetylinouranyl trippetide synthase Occovylytegroophate transforma envolved in technic acid biopothesi Nucleoside diphosphate-sugar epimerase Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein acomponent of a partative secretion system Secreted grately, with PKD repeat domain Clucyce) transfersa family 2 Cellocose - 3-posphate-sugar proghosphorylase involved in lipopolysacch Clucosen - 4-posphate sign proghosphorylase involved in lipopolysacch Clucosamic 6-polysphate sign proghosphorylase involved in lipopolysacch Clucosamice 6-polysphate synthese Clucosamice 6-polysphate sign proghosphorylase involved in lipopolysacch Clucosamice 6-polysphate sign progenose Clucosamice fore prote	Coc00769 Coc001887 Coc001887 Coc001887 Coc000451 Coc000451 Coc000453 Coc000453 Coc000453 Coc000453 Coc000453 Coc000453 Coc001287 Coc000453 Coc001287 Coc000453 Coc001287 Coc000453 Coc0004	plano3645, plano16454 plano1646, plano12370 c05234 plano12370 c05234 plano12370 c05234 plano2370 c05234 plano2370 c05234 plano0535 c0102420 plano0535 c0104179 plano0535 c0104179 plano0535 c0104179 plano0536 c0104179 plano0537 c0104181 plano0538 c0104181 plano0538 c010714 plano0538 c010714 plano0539 c010714 plano0530 c010714 plano0531 c010714 plano0532 c010714 plano0530 c01310 plano1500 c01310 plano1500 c01310 plano1500 c01310 plano1500 c013130 plano1204 c00009 plano12050 c013130 plano12050 c013130 plano12050 c013130 plano12051 c00009	TIGR0.082 TIGR0.179 TIGR0.253 TIGR0.253 TIGR0.253 TIGR0.253 TIGR0.253 TIGR0.253 TIGR0.213 TIGR0.213 TIGR0.213 TIGR0.213 TIGR0.213 TIGR0.253 TIGR0.253 TIGR0.253 TIGR0.253 TIGR0.253 TIGR0.253 TIGR0.253 TIGR0.254 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
mcOd064827 mcOd064827 mcOd064827 mcOd064827 mcOd07500 mcOd07518	M M M M M M M M M M M M M M M M M M M	Cell wall (membrane (mevelope biogenesis Cell wall (membrane (mevelope	Murc 1 Tag8 WcaG 2 - - - - - - - - - - - - - - - - - - -	UPP-N acetylinouramy trippetide synthase Occovylytgerophytamic starsforma environment in technical coll boyothesis Nucleoside diplocophate sugar engineerase Cell surface protein, a component of a partative secretion system Okycoy transfersa family 2 Secreted fifts-propeller repeat protein fused to CARDB-like adhesion modu Cell surface protein; Okycoy transfersa family 2 Cell surface protein; Okycoy transfersa family 2 Okycoy transfersa family 2 Cell Acky the transford family 2 Cell Sector Acky the transford tytem, ATPase component ACK - type minitidry dransford tytem, permease component ACK - type minitidry dransford tytem, ATPase component ACK - type minitidry dransford transford tytem, ATPase component ACK - type antimicrobial petidet transford tytem, ATPase component ACK - ty	Coccorres Coccor	plano3645, plano3645, plano3645, plano3636, plano3636, plano3253,	TIGR0.082 TIGR0.179 TIGR0.253 TIGR0.254 TIGR0.255 TIGR0.254 TIGR0.255 TIGR0.254 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.255 TIGR0.	9 5 5 4 4 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2 1 1 1 1 1 1 1 1 1 1 1 1 1	4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	M M M M M M M M M M M M M M M M M M M	Cell wall (membrane (mevolop biogenesis Cell wall (membrane (mevolop biogenes	Murc 2 Trag8 WeaG 2 	UP-N acetylinourapyl trippetide synthase Occovyl tytercomponite transforma envolved in technic acid bioporthesi Nucleoside diphosphate-sugar epimerase Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein, a component of a partative secretion system Cell surface protein secretion system Secreted grotein, with PKD repeat domain Clucyce) transfersa family 2 Cell surface protein family 2 Cell surface protein family 2 Cell surface famil	Coc00769 Coc001887 Coc001887 Coc001887 Coc000451 Coc000451 Coc000451 Coc000463 Coc000463 Coc001287 Coc000463 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc000463 Coc001287 Coc000463 Coc001287 Coc000463 Coc000483 Coc0004	plano3645, plano3645, plano3645, plano3230 c05234 plano3230 c05234 plano3230 c05234 plano3230 c05234 plano3235 c05234 plano3235 c02522 plano3235 c02524 plano3235 c02434 plano3235 c024181 plano3236 c024181 plano3235 c020761 plano3236 c020761 plano3236 c020761 plano3236 c020714 plano3236 c020714 plano3236 c020714 plano3236 c020714 plano3236 c03741 plano3236 c03744 plano3236 c03744 plano3236 c03744 plano3256 c03744 plano3257 c030794 plano3258 c03734 plano3256 c03734 plano3257 c03734 plano3258 c03734 plano3256 c03734 plano3257	TIGR0.082 TIGR0.179 TIGR0.26 TIGR0.253 TIGR0.253 TIGR0.223 TIGR0.2	2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
accodeday accodeday <td< td=""><td>M M M M M M M M M M M M M M M M M M M</td><td>Cid will newbane (newbegb biogenesis Cid will newbane newbegb biogenesis Cid will newbane newbane befrese nechanisms Defrese necha</td><td>Murc B TagB Wca3C - - - - - - - - - - - - - - - - - - -</td><td>UpP H-acetylinourapyl trippetide synthate Opcovyl tytercomposite transformate involved in technica add biopathesis Opcovyl transformate ageinarea involved in technica add biopathesis Cell surface proteins, a component of a partative secretion system Opcovpl transformate family 2 Cell surface proteins, a component of a partative secretion system Opcovpl transformate family 2 Cell surface proteins, a component of a partative secretion system Opcovpl transformate family 2 Cell surface proteins, a component of a partative secretion system Opcovpl transformate family 2 Objects of the system of a partative secretion system Secreted grotein, with PKD repeat domain Opcovpl transformate family 2 Objects of the system of a partative secretion system Secreted grotein, with PKD repeat domain Opcovpl transformate family 2 Objects of the system of the system of a partative secretion system Secreted grotein, with PKD repeat domain Opcovpl transformate family 2 Objects of the system of the system of the system DIDP H-acetylinous 6 dehydrates Corp-or transformate family 2 Opcovpl transformate family 2 Opcovpl transformate family 2 DIDP H-acetylinous 6 dehydrates Corp-or transformate family 2 DIDP H-acetylinous 7 DIDP H-acet</td><td>Cc600769 Cc6001887 Cc6001887 Cc600451 Cc600451 Cc600451 Cc600453 Cc600455 Cc60055 Cc60055 Cc60055 Cc60055 Cc60055 Cc60055 Cc60055 Cc60055 Cc60055 C</td><td>plano3654 plano3654 plano12370 c05234 plano12370 c05244 plano12370 c05414 plano12370 c05414 plano12380 c045412 plano23531 c030741 plano03531 c030741 plano03532 c030741 plano03531 c030741 plano03531 c030741 plano0353 c030741 plano0353 c030741 plano0353 c030741 plano1353 c030741 plano1353 c030741 plano1353 c030741 plano1354 c030741 plano1355 c030741 plano1355 c030741 plano1355 c030741 plano1355 c030741 plano1355</td><td>TIGR0082 TIGR01179 TIGR074 TIGR074 TIGR0726 TIGR0723 TIGR0723 TIGR0723 TIGR074182 TIGR07418 TIGR07418 TIGR07418 TIGR07418 TIGR07418 TIGR0741 TIGR074 T</td><td>4 3 2 2 2 2 2 2 2</td><td></td><td>2 1 1 1 1 1 1 1 1 1 1 1 1 1</td><td>4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td></td<>	M M M M M M M M M M M M M M M M M M M	Cid will newbane (newbegb biogenesis Cid will newbane newbegb biogenesis Cid will newbane newbane befrese nechanisms Defrese necha	Murc B TagB Wca3C - - - - - - - - - - - - - - - - - - -	UpP H-acetylinourapyl trippetide synthate Opcovyl tytercomposite transformate involved in technica add biopathesis Opcovyl transformate ageinarea involved in technica add biopathesis Cell surface proteins, a component of a partative secretion system Opcovpl transformate family 2 Cell surface proteins, a component of a partative secretion system Opcovpl transformate family 2 Cell surface proteins, a component of a partative secretion system Opcovpl transformate family 2 Cell surface proteins, a component of a partative secretion system Opcovpl transformate family 2 Objects of the system of a partative secretion system Secreted grotein, with PKD repeat domain Opcovpl transformate family 2 Objects of the system of a partative secretion system Secreted grotein, with PKD repeat domain Opcovpl transformate family 2 Objects of the system of the system of a partative secretion system Secreted grotein, with PKD repeat domain Opcovpl transformate family 2 Objects of the system of the system of the system DIDP H-acetylinous 6 dehydrates Corp-or transformate family 2 Opcovpl transformate family 2 Opcovpl transformate family 2 DIDP H-acetylinous 6 dehydrates Corp-or transformate family 2 DIDP H-acetylinous 7 DIDP H-acet	Cc600769 Cc6001887 Cc6001887 Cc600451 Cc600451 Cc600451 Cc600453 Cc600455 Cc60055 Cc60055 Cc60055 Cc60055 Cc60055 Cc60055 Cc60055 Cc60055 Cc60055 C	plano3654 plano3654 plano12370 c05234 plano12370 c05244 plano12370 c05414 plano12370 c05414 plano12380 c045412 plano23531 c030741 plano03531 c030741 plano03532 c030741 plano03531 c030741 plano03531 c030741 plano0353 c030741 plano0353 c030741 plano0353 c030741 plano1353 c030741 plano1353 c030741 plano1353 c030741 plano1354 c030741 plano1355 c030741 plano1355 c030741 plano1355 c030741 plano1355 c030741 plano1355	TIGR0082 TIGR01179 TIGR074 TIGR074 TIGR0726 TIGR0723 TIGR0723 TIGR0723 TIGR074182 TIGR07418 TIGR07418 TIGR07418 TIGR07418 TIGR07418 TIGR0741 TIGR074 T	4 3 2 2 2 2 2 2 2		2 1 1 1 1 1 1 1 1 1 1 1 1 1	4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	M M M M M M M M M M M M M M M M M M M	Cid will newbare feweloge biogenesis Cid will newbare investore biogenesis Cid will newbare investore biogenesis Cid will newbare investore biogenesis Cid will newbare newbare histore histo	Murc B TragB WcaSC - - - - - - - - - - - - - - - - - - -	UpP-N acetylinourapyl trippetide synthase Ocycolyt/geropylonite transforma envolved in technic acid bioportheli Nucleoside diphosphate-sugar epimerase Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein secretion system, surface component Cell Surface surface and mutarotase Cell surface surface and surface secretion system Cell surface surface and surface secretion system, Strate component CERMs associated protein sack, Ref family econolesie Nach secretion system seclution system section secretion AC system surface and suburits of a related FIM- containing fluxoprotein, ppr Cell suburit of restriction endonuclesie Natomical reductase Nima or a related FIM- containing fluxoprotein, ppr Cell suburit of restriction system methyltransformate suburit Figure in restriction surface protein section pathwy component CalAs ppoint full subactio	Coc00769 Coc00187 Coc00187 Coc00187 Coc00183 Coc000431 Coc000433 Coc000433 Coc001291 Coc000463 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001438 Coc001387 Coc001431 Coc001431 Coc001387 Coc001431 Coc001431 Coc001431 Coc001438 Coc001387 Coc001467 Coc001467 Coc001467 Coc001467 Coc001467 Coc001467 Coc001467 Coc001467 Coc001467 Coc001467 Coc001467 Coc001467 Coc001467 Coc001487 C	plano3645, plano3645, plano3645, plano3230 c05234 plano3230 c05234 plano3230 c05234 plano3230 c05234 plano3230 c05234 plano3230 c05234 plano3230 c05424 plano3230 c05424 plano3230 c05424 plano3230 c05424 plano3230 c05424 plano3230 c05424 plano3230 c034181 plano3230 c030741 plano3231 c030741 plano3231 c030741 plano3231 c03074 plano3231 c03074 plano3231 c03074 plano3231 c03074 plano3231 c03074 plano3245 plano3245 plano325 c03230 plano325 c03230 plano3263 c03230 plano3264 c03230 plano3264 c03230 plano3264 c03230 plano3264 <t< td=""><td>TIGR0082 TIGR01179 TIGR074 TIG</td><td>4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2</td><td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td><td>2 1 1 1 1 1 1 1 1 1 1 1 1 1</td><td>4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td></t<>	TIGR0082 TIGR01179 TIGR074 TIG	4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2 1 1 1 1 1 1 1 1 1 1 1 1 1	4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
accodeday accodeday <td< td=""><td>M M M M M M M M M M M M M M M M M M M</td><td>Cid will newbane (newbegb biogenesis Cid will newbane newbegb biogenesis Cid will newbane newbane befrese nechanisms Defrese necha</td><td>Mure: Tag8 Wcc36 </td><td>UpP H-acetylinouramy trippetide synthate Opcovylytigerophytale sugar epidemiae (Covoylytigerophytale sugar epidemiae Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein secretion system Cell surface surface and mutarotase Cell surface surface and mutarotase Cell surface surface and mutarotase Cell surface surface and surface surface Cell Surface surface and surface surface Cell Surface surface and surface submit face Cell surface surface and surface submit secretion system. ATPsia component CERPs a submit of restriction endouclease Cell Surface and Cell submit of a related PIM- containing flavoprotein, ppr Cell Surface Surface submit Soil Cell Surface Surface submit Soil Cell Surface Surface submit Soil Cell Surface Surface submit Soil Sec - independent protein section pathway component Tacl Sec - independent protein section pathway component tracs Sec - independent model submit</td><td>Coc00769 Coc001887 Coc001887 Coc001887 Coc000451 Coc000451 Coc000453 Coc000453 Coc000453 Coc0014</td><td>plano3654 plano3654 plano1237 c05234 plano1237 c05234 plano1237 c05234 plano1237 c05234 plano1237 c05234 plano2353 c05242 plano2353 c05424 plano2353 c05424 plano2353 c05424 plano2353 c05424 plano2353 c05424 plano2353 c05424 plano2353 c054428 plano2353 c05424 plano353 c05424 plano353 c054428 plano353 c054428 plano353 c05074 plano353 c05074 plano353 c05360 plano353 c05374 plano3545 plano3545 plano355 c03374 plano3550 c03374 plano3550 c03374 plano3550 c03374 plano3550 c03374 plano3550 c03374</td><td>TIGR0032 TIGR0119 TIGR0119 TIGR0236 TIGR0233 TIGR03213 TIGR03243 TIGR03253 TIGR03253 TIGR0326 TIGR0327 TIGR0328 TIGR03292 TIGR03292 TIGR0328 TIGR0328 TIGR03292 TIGR0328 TIGR0328 TIGR03292 TIGR0328 TIGR0328 TIGR03292 TIGR0328 TIGR0372 TIGR0372 TIGR0372 TIGR0373 TIGR0374 TIGR0453 TIGR0454 TIGR0454 TIGR0445 TIGR0445 TIGR0544 TIGR0545 TIGR0545 TIGR0545</td><td>4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2</td><td>1 1 1 1 1 1 1 1 1 1 1 1 1 1</td><td>2 1 1 1 1 1 1 1 1 1 1 1 1 1</td><td>4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td></td<>	M M M M M M M M M M M M M M M M M M M	Cid will newbane (newbegb biogenesis Cid will newbane newbegb biogenesis Cid will newbane newbane befrese nechanisms Defrese necha	Mure: Tag8 Wcc36 	UpP H-acetylinouramy trippetide synthate Opcovylytigerophytale sugar epidemiae (Covoylytigerophytale sugar epidemiae Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein secretion system Cell surface surface and mutarotase Cell surface surface and mutarotase Cell surface surface and mutarotase Cell surface surface and surface surface Cell Surface surface and surface surface Cell Surface surface and surface submit face Cell surface surface and surface submit secretion system. ATPsia component CERPs a submit of restriction endouclease Cell Surface and Cell submit of a related PIM- containing flavoprotein, ppr Cell Surface Surface submit Soil Cell Surface Surface submit Soil Cell Surface Surface submit Soil Cell Surface Surface submit Soil Sec - independent protein section pathway component Tacl Sec - independent protein section pathway component tracs Sec - independent model submit	Coc00769 Coc001887 Coc001887 Coc001887 Coc000451 Coc000451 Coc000453 Coc000453 Coc000453 Coc0014	plano3654 plano3654 plano1237 c05234 plano1237 c05234 plano1237 c05234 plano1237 c05234 plano1237 c05234 plano2353 c05242 plano2353 c05424 plano2353 c05424 plano2353 c05424 plano2353 c05424 plano2353 c05424 plano2353 c05424 plano2353 c054428 plano2353 c05424 plano353 c05424 plano353 c054428 plano353 c054428 plano353 c05074 plano353 c05074 plano353 c05360 plano353 c05374 plano3545 plano3545 plano355 c03374 plano3550 c03374 plano3550 c03374 plano3550 c03374 plano3550 c03374 plano3550 c03374	TIGR0032 TIGR0119 TIGR0119 TIGR0236 TIGR0233 TIGR03213 TIGR03243 TIGR03253 TIGR03253 TIGR0326 TIGR0327 TIGR0328 TIGR03292 TIGR03292 TIGR0328 TIGR0328 TIGR03292 TIGR0328 TIGR0328 TIGR03292 TIGR0328 TIGR0328 TIGR03292 TIGR0328 TIGR0372 TIGR0372 TIGR0372 TIGR0373 TIGR0374 TIGR0453 TIGR0454 TIGR0454 TIGR0445 TIGR0445 TIGR0544 TIGR0545 TIGR0545 TIGR0545	4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1	2 1 1 1 1 1 1 1 1 1 1 1 1 1	4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
accodeday accodeday <td< td=""><td>MA MA MA <</td><td>Cid will (membrane (mevlope biogenesis Cid wi</td><td>Mure: TagB Wcc36 Agib Agib Agib Agib Agib Agib Agib Agib</td><td>UpP H-acetylinouramy trippetide synthate Orcyolytyliceryolophate transforma eniovibed in technica add bioporthelis Nucleoside diphosphate sugar epimerase Cell surface proteins, a component of a pitative secretion system Givpolyt transfersa family 2 Cell surface proteins, a component of a pitative secretion system Givpolyt transfersa family 2 Cell surface proteins, a component of a pitative secretion system Givpolyt transfersa family 2 Cell surface proteins, a component of a pitative secretion system Givpolyt transfersa family 2 Cell surface proteins, a component of a pitative secretion system Givpolyt transfersa family 2 Cell surface proteins, a component of a pitative secretion system Cell surface proteins family 2 Oligoacchurytransfersase family 2 Oligoacchurytransfersase family 2 Celloucose - 3-polytophate sugar prophosphorytase involved in lipopolytack Givpolyt transfersa family 2 Celloucose - 1-polytophate sugar prophosphorytase involved in lipopolytack Givpolyt transfersa family 2 Celloucose - 1-polytophate system, and the protein release, permease comp UpP -N acetylinour and diversate Celycopytransfersase Celycopytran</td><td>Coc00769 Coc001887 Coc001887 Coc001887 Coc000451 Coc000451 Coc000451 Coc000453 Coc000453 Coc000453 Coc000453 Coc001453 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001311 Coc001452 Coc001451 Coc001451 Coc001452 Coc001451 Coc001451 Coc001452 Coc0014</td><td>plano365, plano255, plano2457, plano3646, plano1230 (2023) plano2303 (2023) plano2303 (2023) plano2303 (2023) plano2303 (2024) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2007) plano2303 (2007) plano1203 (20</td><td>TIGR0082 TIGR01179 TIGR074 TIGR074 TIGR074 TIGR074 TIGR07423 TIGR074 T</td><td>2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2</td><td></td><td>2 1 1 1 1 1 1 1 1 1 1 1 1 1</td><td>4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td></td<>	MA MA MA <	Cid will (membrane (mevlope biogenesis Cid wi	Mure: TagB Wcc36 Agib Agib Agib Agib Agib Agib Agib Agib	UpP H-acetylinouramy trippetide synthate Orcyolytyliceryolophate transforma eniovibed in technica add bioporthelis Nucleoside diphosphate sugar epimerase Cell surface proteins, a component of a pitative secretion system Givpolyt transfersa family 2 Cell surface proteins, a component of a pitative secretion system Givpolyt transfersa family 2 Cell surface proteins, a component of a pitative secretion system Givpolyt transfersa family 2 Cell surface proteins, a component of a pitative secretion system Givpolyt transfersa family 2 Cell surface proteins, a component of a pitative secretion system Givpolyt transfersa family 2 Cell surface proteins, a component of a pitative secretion system Cell surface proteins family 2 Oligoacchurytransfersase family 2 Oligoacchurytransfersase family 2 Celloucose - 3-polytophate sugar prophosphorytase involved in lipopolytack Givpolyt transfersa family 2 Celloucose - 1-polytophate sugar prophosphorytase involved in lipopolytack Givpolyt transfersa family 2 Celloucose - 1-polytophate system, and the protein release, permease comp UpP -N acetylinour and diversate Celycopytransfersase Celycopytran	Coc00769 Coc001887 Coc001887 Coc001887 Coc000451 Coc000451 Coc000451 Coc000453 Coc000453 Coc000453 Coc000453 Coc001453 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001311 Coc001452 Coc001451 Coc001451 Coc001452 Coc001451 Coc001451 Coc001452 Coc0014	plano365, plano255, plano2457, plano3646, plano1230 (2023) plano2303 (2023) plano2303 (2023) plano2303 (2023) plano2303 (2024) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2004) plano2303 (2007) plano2303 (2007) plano1203 (20	TIGR0082 TIGR01179 TIGR074 TIGR074 TIGR074 TIGR074 TIGR07423 TIGR074 T	2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2		2 1 1 1 1 1 1 1 1 1 1 1 1 1	4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
acconductor accond	M M M M M M M M M M M M M M M M M M M	Cell wall (membrane (mevlope) biogenesis Cell wall (membrane (mevlope)	MurrE TagB Wcc3G AgB AgB AgB AgB AgB AgB/Rt AgB/C AgB/	UpP Hacetylmuramy trippetide synthase Orcyolytylexerophonia ter surfarerae involved in technica add biopathesis Nucleoside diphosphate sugar epimerase UpP Hacetyliocomine 2-geimerase Cell surface proteins, a component of a partative secretion system Givcoyl transfersa family 2 Cell surface proteins, a component of a partative secretion system Givcoyl transfersa family 2 Cell surface proteins, a component of a partative secretion system Givcoyl transfersa family 2 Cell surface proteins, a component of a partative secretion system Givcoyl transfersa family 2 Cell surface proteins, a component of a partative secretion system Givcoyl transfersa family 2 Cell surface proteins, a component of a partative secretion system Cell surface proteins family 2 Oligoaccharyltransfersa embrane subunit Givcoyl transfersa family 2 Cell surface proteins family 2 Cell surface proteins family 2 Oligoaccharyltransfersase family 2 Cell surface proteins family 2 Cell surface pro	Cc600769 Cc6001887 Cc6001887 Cc600381 Cc600381 Cc600381 Cc6003291 Cc600463 Cc600463 Cc600463 Cc6004291 Cc600463 Cc6004291 Cc600420 Cc60040	plano365, plano2355, plano2457, plano3454, plano1230, e05234, plano2305, e05234, plano2305, e03234, plano2305, e03234, plano2305, e03234, plano2305, e03434, plano2305, e034434, plano2305, e034434, plano2305, e034434, plano2305, pla	TIGR0032 TIGR01179 TIGR0126 TIGR0226 TIGR0223 TIGR0223 TIGR0223 TIGR0223 TIGR0223 TIGR0223 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0257	4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2		2 1 1 1 1 1 1 1 1 1 1 1 1 1	4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
accodday accodday <t< td=""><td>MA MA MA</td><td>Cid will (membrane (mewlope biogenesis Cid wi</td><td>Mure: TagB Wcc30 Agib Agib Agib Agib Agib Agib Agib Agib</td><td>UpP H-acetylinouramy trippetide synthate Orycovylytergrouphate transformate involved in technica add biopathesis Orycovylytergrouphate sugar epidemase Cell surface proteins, a component of a pitative secretion system Orycovyl transformate family 2 Cell surface proteins, a component of a pitative secretion system Orycovyl transformate family 2 Cell surface proteins, a component of a pitative secretion system Orycovyl transformate family 2 Cell surface proteins, a component of a pitative secretion system Orycovyl transformate family 2 Orycovyl transformate orycovyl transformate light Orycovyl transformate regioner (SAR, Ref family econuclease Na-d free multidrug transport system, ATPase component AGC system multidrug transport system, ATPase component Orycovyl transformate regioner or system, Protease component Orycovyl transformate regioner or toxin-mainte fliveprotein, ppt Notroimdia or devictase Nim or a related FMK- containing fliveprotein, ppt Notroimdia or devictase Nim or a related FMK- containing fliveprotein, ppt Notroimdia or devictase Nim or a related FMK- conta</td><td>Coc00769 Coc001887 Coc001887 Coc001887 Coc000451 Coc000451 Coc000451 Coc000453 Coc000453 Coc000453 Coc000453 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001311 Coc001451 Coc004532 Coc004532 Coc00455 Coc004532 Coc00455 Coc004532 Coc004532 Coc004532 Coc004532</td><td>plano365, plano2355, plano2457, plano3454, plano1237, ed05234 plano1230, ed05234 plano2305, ed05234 plano2355, ed05234 plano2355, ed05232 plano2355, ed05234 plano2355, ed05422 plano2355, plano3458, plano2369 plano2358, plano2369 plano2358, plano2369 plano2358, plano2369 plano2358, plano2369 plano2358, plano2369 plano3538, plano2369 plano3538, plano2367 plano358, plano2367 plano358, plano2367 plano358, plano2367 plano358, plano2367 plano358, plano2367 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano238 plano358, plano238 plano359 p</td><td>TIGR0032 TIGR01179 TIGR0126 TIGR0226 TIGR0223 TIGR0223 TIGR0223 TIGR0223 TIGR0223 TIGR0223 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0257</td><td>4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2</td><td></td><td>2 1 1 1 1 1 1 1 1 1 1 1 1 1</td><td>4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td></t<>	MA	Cid will (membrane (mewlope biogenesis Cid wi	Mure: TagB Wcc30 Agib Agib Agib Agib Agib Agib Agib Agib	UpP H-acetylinouramy trippetide synthate Orycovylytergrouphate transformate involved in technica add biopathesis Orycovylytergrouphate sugar epidemase Cell surface proteins, a component of a pitative secretion system Orycovyl transformate family 2 Cell surface proteins, a component of a pitative secretion system Orycovyl transformate family 2 Cell surface proteins, a component of a pitative secretion system Orycovyl transformate family 2 Cell surface proteins, a component of a pitative secretion system Orycovyl transformate family 2 Orycovyl transformate orycovyl transformate light Orycovyl transformate regioner (SAR, Ref family econuclease Na-d free multidrug transport system, ATPase component AGC system multidrug transport system, ATPase component Orycovyl transformate regioner or system, Protease component Orycovyl transformate regioner or toxin-mainte fliveprotein, ppt Notroimdia or devictase Nim or a related FMK- containing fliveprotein, ppt Notroimdia or devictase Nim or a related FMK- containing fliveprotein, ppt Notroimdia or devictase Nim or a related FMK- conta	Coc00769 Coc001887 Coc001887 Coc001887 Coc000451 Coc000451 Coc000451 Coc000453 Coc000453 Coc000453 Coc000453 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001287 Coc001311 Coc001451 Coc004532 Coc004532 Coc00455 Coc004532 Coc00455 Coc004532 Coc004532 Coc004532 Coc004532	plano365, plano2355, plano2457, plano3454, plano1237, ed05234 plano1230, ed05234 plano2305, ed05234 plano2355, ed05234 plano2355, ed05232 plano2355, ed05234 plano2355, ed05422 plano2355, plano3458, plano2369 plano2358, plano2369 plano2358, plano2369 plano2358, plano2369 plano2358, plano2369 plano2358, plano2369 plano3538, plano2369 plano3538, plano2367 plano358, plano2367 plano358, plano2367 plano358, plano2367 plano358, plano2367 plano358, plano2367 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano237 plano358, plano238 plano358, plano238 plano359 p	TIGR0032 TIGR01179 TIGR0126 TIGR0226 TIGR0223 TIGR0223 TIGR0223 TIGR0223 TIGR0223 TIGR0223 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0252 TIGR0257	4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2		2 1 1 1 1 1 1 1 1 1 1 1 1 1	4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
acconductor arCool 1992 arCool 1993 arCool	M M	Cell wall (membrane (mevlope biogenesis Cell wall (membrane (mevlope biogenes	MurrE TagB Wcc3G AgB AgB AgB AgB AgB/Rt AgB/ AgB/ AgB/ AgB/ AgB/ AgB/ AgB/ AgB/	UpP-N acetylinouramy trippedie synthase (Cryocyl/getrophyshate sugar epilierase (Cryocyl/getrophyshate sugar epilierase Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, such PRO repeat domain Cilycopit transfersa family 2 Cellsource - 2 prohyshophate sugar prophosphorylase involved in lipopolytack Cellsource - 3 prohyshophate sugar prophosphorylase involved in lipopolytack Cellsource - 3 prohyshophate sugar prophosphorylase involved in lipopolytack Cellsource - 3 prohyshophate sugar prophosphorylase involved in Cellsource - 4 prohyshophate sugar prophosphorylase involved in Cellsource - 4 prohyshophate sugar prophosphorylase involved in Cellsource - 4 prohyshophate sugar prophosphore - 4 prohyshophate Cellsource - 4 prohyshophate sugar prohysh	Cc600769 Cc6001887 Cc6001887 Cc600381 Cc600381 Cc600343 Cc6003291 Cc600463 Cc600463 Cc600463 Cc6004291 Cc600463 Cc6004291 Cc600463 Cc600429 Cc600429 Cc600429 Cc600429 Cc600429 Cc600429 Cc600429 Cc600429 Cc600429 Cc600453 Cc600429 Cc600453 Cc600455 Cc600452 Cc600452 Cc600452 Cc600455 Cc600452 Cc600453 Cc600455 Cc60045	plano365,plano22875 plano3264,class,plano22875 plano3236 class,22 plano72375 class,22 plano72375 class,22 plano72375 class,22 plano72375 class,22 plano3235 class,22 plano3235 class,22 plano3235 class,22 plano3235 class,22 plano3235 class,22 plano3235 class,22 plano3236 class,22 plano3236 class,22 plano3236 class,22 plano3236 class,22 plano3236 class,22 plano3236 class,22 plano3236 class,22 plano3237 class,22 plano3238 class,23 plano3238 class,23	TIGR00382 TIGR01179 TIGR01179 TIGR0226 TIGR02233 TIGR02132 TIGR02132 TIGR02233 TIGR023423 TIGR02350 TIGR0236 TIGR02372 TIGR02382 TIGR02382 TIGR0236 TIGR02372 TIGR02382 TIGR02382 TIGR02382 TIGR02382 TIGR02382 TIGR0392 TIGR0394 TIGR0394 TIGR0394 TIGR03957 TIGR03957 TIGR03957 TIGR03957 TIGR03957 <	4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1		4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
accondustry arCool 1992 arCool 1993 arCool	M M	Cell wall (membrane (mevlope biogenesis Cell wall (membrane (mevlope biogenes	Mure: TagB Wcc3G AglB AglB AglB AglF/Rt AglA AglO GCD1 LolE MurD GCD1 GCD1 GCD1 GCD1 GCD1 GCD1 GCD1 GCD	UpP-N acetylinouramy trippedie synthase (Cycoyl/gynchyshate-sugar epimerase (Cycoyl/gynchyshate-sugar epimerase Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Cell surface protein, a component of a pitative secretion system Citycoyl transfersa family 2 Cell surface protein, a component of a pitative secretion system Citycoyl transfersa family 2 Cell surface protein, a component of a pitative secretion system Citycoyl transfersa family 2 Cell surface protein cell surface secretion system Citycoyl transfersa family 2 Citycoyl Citycoyl transfersa family 2 Citycoyl Citycoyl Tamber System, ATPae component Citycoyl Citycoyl Citycoyl Citycoyl Citycoyl	Cc600769 Cc6001887 Cc6001887 Cc600381 Cc600381 Cc600343 Cc6003291 Cc600463 Cc60047 C	lano365, μin/22875 jano3646, jano3230 (20376) jano3230 (2037	TIGR02082 TIGR01082 TIGR01179 TIGR0226 TIGR0226 TIGR02213 TIGR02182 TIGR02182 TIGR02182 TIGR02182 TIGR02182 TIGR02182 TIGR02182 TIGR0219 TIGR02192 TIGR02113 TIGR02213 TIGR02154 TIGR0226 TIGR02155 TIGR02063 TIGR0207 TIGR02082 TIGR02083 TIGR02084 TIGR02097 TIGR02082 TIGR02084 TIGR020853 TIGR02084 TIGR02084 TIGR020857 TIGR020857 TIGR020857 TIGR020857 TIGR020857 TIGR020857 TIGR020857 TIGR020857	4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2 1 1 1 1 1 1 1 1 1 1 1 1 1	4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
a COORDAN a COORDAN	M M	Cid wal/membrane/mevkgpb biogenesis Cid wal/mem	MurrE TagB WrcaG AglW AglO AglB AglB AglF/Rt AglM AglO GCD1 AglF/Rt AglM AglO GCD1 Cas4 AglO GCD1 GCD1 GCD1 Cas4 AglO Cas4 Cas4 NorM RfaG Cas4 NorM NorM NorM NorM NorM NorM NorM NorM	UPP N-acetifynuramy trippetke synthase Grocovylytgerophysinke transforma involvel in technica add bioparthesi Nuckosite diphosphate-sugar epimerase UPP N-acetifylocomine 2-gimmerase Cell surface proteins, a component of a partative secretion system Gell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins family 2 Cell surface family 2 Cell surface proteins family 2 Cell surface family 2	Cc600769 Cc6001887 Cc6001887 Cc6001887 Cc6000451 Cc6000451 Cc6000451 Cc6000451 Cc6000453 Cc6001291 Cc6000463 Cc6001291 Cc6000463 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001401 Cc6001451 Cc60014	Janobase, jutro22875, jutro22875, jatro22875, jatro22	TIGR00082 TIGR01082 TIGR01082 TIGR0236 TIGR0236 TIGR02323 TIGR02323 TIGR02323 TIGR02323 TIGR03154 TIGR03092 TIGR0236 TIGR0238 TIGR03092 TIGR0238 TIGR03092 TIGR03092 TIGR0203 TIGR03092 TIGR0203 TIGR03097 TIGR0204 TIGR0204 TIGR0204 TIGR0215 TIGR024 TIGR024 TIGR0257 TIGR024 TIGR0257 TIGR024 TIGR0257 TIGR024 TIGR0257 TIGR024 TIGR0257 TIGR	4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2			4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
a Coordiant and	M M	Cid wal/membrane/mevkgpb biogenesis Cid wal/mem	MurrE TagB WrcaG P AgB AgB AgB AgB AgB AgB AgB AgB AgB AgB	UpP-N actignment tripped synthate OpP-N actignment provides in technical action of the OpP-N actignment of a priority of the opposite Cell surface proteins, a component of a priority execution system Cell surface proteins, a component of a priority execution system Cell surface proteins, a component of a priority execution system Cell surface proteins, a component of a priority execution system Cell surface proteins, a component of a priority execution system Cell surface proteins, a component of a priority execution system Cell surface proteins, a component of a priority execution system Cell surface proteins, a component of a priority execution system Cell surface proteins, a component of a priority execution system Cell surface proteins, a component of a priority execution system Cell surface proteins family 2 Cell surface family proteins cell surface family 2 Cel	Cc600769 Cc6001887 Cc6001887 Cc6001887 Cc6000451 Cc6000451 Cc6000451 Cc6000451 Cc6000451 Cc6000453 Cc6001291 Cc6000463 Cc6001291 Cc6000463 Cc6001291 Cc6000463 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc6001287 Cc600140 Cc6001288 Cc6001311 Cc6001428 Cc6001311 Cc6001428 Cc6001311 Cc6001428 Cc6001311 Cc6001428 Cc6001311 Cc6001428 Cc6001311 Cc6001428 Cc6001311 Cc6001428 Cc6001311 Cc6001428 Cc6001312 Cc600451 Cc600138 Cc600140 Cc600140 Cc600140 Cc60017 Cc600140 Cc600140 Cc60017 Cc600140	Janobase, jutro22875, jutro22875 Janobase, jutro22877 Janobase, jutr	TIGR00082 TIGR01179 TIGR0126 TIGR0226 TIGR0213 TIGR0213 TIGR0213 TIGR02142 TIGR0226 TIGR02276 TIGR0228 TIGR02922 TIGR02922 TIGR02026 TIGR02115 TIGR02115 TIGR02122 TIGR02122 TIGR02037 TIGR02037 TIGR02037 TIGR02042 TIGR02042 TIGR02057 TIGR02042 TIGR02042 TIGR02045 TIGR02	4 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2			4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
accode/action	M W V V V	Cid wal/membrane/mevkgp biogenesis Cid wal/membrane/mevkgp biogen	Мигг В Тадв Уисаб - - - - - - - - - - - - -	UCP-N-acet[incurant] trippetide synthate Orcyoly/gerophytae-sugar epimerase UCP-N-acet[incursoline 1 epimerase Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins, a component of a partative secretion system Cell surface proteins family 2 Cell surface family 2 Cell surface family 2 Cell surface family 2 Cell surface family 2 Cell sur	Cc600769 Cc600187 Cc600187 Cc600187 Cc600187 Cc600451 Cc600451 Cc600451 Cc600451 Cc600451 Cc600451 Cc600452 Cc6	planob365,glm-022875 planob464,glmn02302 cd05214 planob3702 cd05214 planob3702 cd05214 planob3702 cd05214 planob3705 cd05214 planob3705 cd05214 planob3705 cd05412 planob3705 cd05412 planob3705 cd05412 planob380 cd05413 planob380 cd054181 planob380 cd07511 planob380 cd07511 planob380 cd07511 planob380 cd07512 planob380 cd07512 planob380 cd07512 planob380 cd07512 planob380 cd07512 planob380 cd0714.cd0305 planob380 cd03734 planob380	TIGR0082 TIGR0182 TIGR0236 TIGR0236 TIGR0233 TIGR03142 TIGR0326 TIGR0327 TIGR0328 TIGR03292 TIGR03292 TIGR03154 TIGR03292 TIGR03292 TIGR03155 TIGR0326 TIGR0327 TIGR03082 TIGR0318 TIGR0317 TIGR0317 TIGR0317 TIGR0317 TIGR0318 TIGR0317 TIGR0318 TIGR03197 TIGR0310	2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2			4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

arCOG01308	0	Posttranslational modification, protein turnover, chaperones	Cdc48	ATPase of the AAA+ class , CDC48 family	COG00464	pfam00004,pfa	cd00009,cd000	TIGR01243			3	1
arCOG00065 arCOG03103	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	csdA CtaA	Selenocysteine lyase/Cysteine desulturase Uncharacterized protein required for cytochrome oxidase assembly	COG00520 COG01612	pfam00266 pfam02628	cd06453	TIGR01979		1		1
arCOG06880 arCOG02846	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	DnaJ DnaJ	DnaJ type Zn finger domain DnaJ-class molecular chaperone	COG00484 COG00484	pfam00226	cd06257	TIGR02349 TIGR02349			3	1
arCOG04142	0	Posttranslational modification, protein turnover, chaperones	DYS1	Deoxyhypusine synthase Brodicted durathione S. transforme	COG01899	pfam01916	cd02100	TIGR00321		1		1
arCOG01341	0	Posttranslational modification, protein turnover, chaperones	GIM5	Predicted gratatione strainserase Predicted prefoldin, molecular chaperone implicated in de novo protein foldi	COG01730	pfam02996	cd005150	TIGR00293		1		1
arCOG05154 arCOG01257	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	GroEL	Chaperonin GroEL (HSP60 family) Chaperonin GroEL, HSP60 family	COG00459 COG00459	pfam00118 pfam00118	cd03344 cd03343	TIGR02348 TIGR02339	5		7	2
arCOG04772 arCOG02606	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	GrpE GrxC	Molecular chaperone GrpE (heat shock protein) Glutaredoxin	COG00576 COG00695	pfam01025 pfam00462	cd00446 cd02976	TIGR02196	1	1	2	1
arCOG02607	0	Posttranslational modification, protein turnover, chaperones	GrxC	Glutaredoxin Membrane protease subunit, stomatin/probibitin homolog	COG00695	pfam00462	cd02976	TIGR02196		1	1	1
arCOG02959	0	Posttranslational modification, protein turnover, chaperones	lap	Zn-dependent amino- or carboxypeptidase, M28 family	COG02234	pfam04389	cd05643	TICR00040			2	3
arCOG04560 arCOG02443	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	ISCA NosY	Fe-S cluster assembly iron-binding protein ISCA ABC-type transport system involved in multi-copper enzyme maturation, per	COG00316 COG01277	pfamU1521 pfam12679		TIGR00049	1		2	
arCOG02438 arCOG01846	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	NosY PaaD	ABC-type transport system involved in multi-copper enzyme maturation, per Metal-sulfur cluster biosynthetic enzyme	COG01277 COG02151	pfam12679 pfam01883,pfa	cd02037	TIGR02945, TIGR01969	2	1		
arCOG00976 arCOG05850	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	Pcm Pcn	Protein-L-isoaspartate carboxylmethyltransferase Pyrrolidone-carboxylate pentidase (N-terminal pyroglutamyl pentidase)	COG02518	pfam01135 pfam01470	cd02440 cd00501	TIGR00080 TIGR00504			1	1
arCOG04767	0	Posttranslational modification, protein turnover, chaperones	РріВ	Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family	COG00652	-600004		7/0704242			1	1
arCOG01308 arCOG00609	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	RSEP	MiP-dependent 265 processome regulatory subunit Membrane-associated protease RseP, regulator of RpoE activity in bacteria	COG01222 COG00750	pfam02163	cd06160	TIGR01242		1	1	
arCOG04064 arCOG00981	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	RseP SlpA	Membrane-associated protease RseP, regulator of RpoE activity in bacteria FKBP-type peptidyl-prolyl cis-trans isomerase 2	COG00750 COG01047	pfam02163 pfam00254	cd06159	TIGR00054 TIGR00115	2		1	
arCOG00980 arCOG02784	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	SIpA SRAP	FKBP-type peptidyl-prolyl cis-trans isomerase 2 Putative SOS response-associated peptidase YedK	COG01047 COG02135	pfam00254 pfam02586				1		1
arCOG03580	0	Posttranslational modification, protein turnover, chaperones	STE14	Putative protein-S-isoprenylcysteine methyltransferase	COG02020	pfam04191		7/2703003	2	1	1	1
arCOG07441	0	Posttranslational modification, protein turnover, chaperones	SurA	Parvulin-like peptidyl-prolyl isomerase	COG00719	pfam00639		TIGR02933	3	1	1	1
arCOG00321 arCOG00322	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	TIdD TIdE	Zn-dependent protease, component of TIdD/TIdE system Inactivated Zn-dependent protease, component of TIdD/TIdE system	COG00312 COG00312	pfam01523 pfam01523					1	1
arCOG06181 arCOG01972	0	Posttranslational modification, protein turnover, chaperones Posttranslational modification, protein turnover, chaperones	TrxA TrxA	Thiol-disulfide isomerase or thioredoxin Thiol-disulfide isomerase or thioredoxin	COG00526 COG00526	pfam00578 pfam00085	cd02966 cd02947	TIGR01068	3	1		
arCOG01929	0 T	Posttranslational modification, protein turnover, chaperones	XdhC	Xanthine and CO dehydrogenase maturation factor, XdhC/CoxF family	COG01975	pfam13478			3	1	1	1
arCOG02967	T	Signal transduction mechanisms	-	NACHT family NTPase fused to HEAT repeats domain	COG05635	pfam13646,pfa	m13646		3	-		1
arCOG01180 arCOG04820	T	Signal transduction mechanisms	-	Serne/threonine protein kinase involved in cell cycle control SOUL heme-binding protein	0001718	pfam01163 pfam04832	0005145	TIGR03724	3		1	
arCOG06193 arCOG04425	T T	Signal transduction mechanisms Signal transduction mechanisms	- Wzb	Signal transduction histidine kinase and PAS domains Protein-tyrosine-phosphatase	COG00642 COG00394	pfam13426,pfa pfam01451	cd00130,cd001 cd00115	TIGR00229,TIGR00229,TIGR01386 TIGR02689	2			
arCOG05374 arCOG01171	T T	Signal transduction mechanisms Signal transduction mechanisms	- RAD55	GAF domain-containing protein RecA-superfamily ATPase implicated in signal transduction	COG01956 COG00467	pfam13185 pfam06745	cd01124	TIGR03877	1			2
arCOG04453	T	Signal transduction mechanisms	DisA_N	Diadenylate cyclase (c-di-AMP synthetase), DisA_N domain	COG01624	pfam02457					4	1
arCOG01143	T	aignai dansuucuon mechanisms Signal transduction mechanisms	Аран	Serine/threonine protein phosphatase PP2A family	COG02202 COG00639	pram05763 pfam00149	cd00144				1	2
arCOG02391 arCOG03413	T T	Signal transduction mechanisms Signal transduction mechanisms	CheY CDC14	Rec domain Protein-tyrosine phosphatase	COG00784 COG02453	pfam00072 pfam00782	cd00156 cd00047	TIGR02154			3	1
arCOG03517 arCOG01997	T T	Signal transduction mechanisms Signal transduction mechanisms	- SixA	Formylglycine-generating sulfatase enzyme Phosphohistidine phosphatase SixA	COG01262 COG02067	pfam12867,pfa pfam00300	m03781 cd07067	TIGR03440 TIGR00249			1	1
arCOG06801	T	Signal transduction mechanisms	SrkA	Ser/Thr protein kinase RdoA involved in Cpx stress response, MazF antagonis	COG02334	pfam01636	cd05153				3	
METABOLIS	N											
arCOG02767 arCOG03600	E	Amino acid transport and metabolism Amino acid transport and metabolism	-	Metal-dependent membrane protease, CAAX family Transglutaminase-like cysteine protease	COG01266 COG01305	pfam02517			3			
arCOG14875 arCOG03611	F	Amino acid transport and metabolism Amino acid transport and metabolism	-	Aspartyl protease Peotidase C1A subfamily		pfam05618	cd08546 cd142	54	2		1	2
arCOG07760	E	Amino acid transport and metabolism Amino acid transport and metabolism	-	Peptidase family C25		pfam01364	cd02258		1			~
arCOG04333 arCOG01130	E	Amino acid transport and metabolism Amino acid transport and metabolism	-	Aspartate/tyrosine/aromatic aminotransterase Aspartate/tyrosine/aromatic aminotransferase	COG00436 COG00436	pfam00155 pfam00155	cd00609	TIGR03540		1	2	
arCOG02165 arCOG01131	E	Amino acid transport and metabolism Amino acid transport and metabolism	-	Transglutaminase-like cysteine protease Aspartate/tyrosine/aromatic aminotransferase	COG01305 COG00436	pfam01841 pfam00155	cd00609	TIGR01265			1	1
arCOG04044 arCOG01888	E	Amino acid transport and metabolism Amino acid transport and metabolism	- AmpS	2-amino-3,7-dideoxy-D-threo-hept-6-ulosonic acid synthase, DhnA-aldolase f Leucyl aminopeotidase (aminopeotidase T)	COG01830 COG02309	pfam01791 pfam02073	cd00958	TIGR01949			2	1
arCOG01890	E	Amino acid transport and metabolism	AmpS	Leucyl aminopeptidase (aminopeptidase T)	COG02309	pfam02073		7/2003453			2	
arCOG01924 arCOG00863	E	Amino acid transport and metabolism Amino acid transport and metabolism	AnsB ArcC	L-asparaginase/archaeal Glu-titiNAGIn amidotransferase subunit D Carbamate kinase	COG00252 COG00549	pfam00/10 pfam00696	cd08962 cd04235	TIGR00746		1	2	
arCOG01107 arCOG00912	E	Amino acid transport and metabolism Amino acid transport and metabolism	ArgE ArgF	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or relat Ornithine carbamoyltransferase	COG00624 COG00078	pfam01546 pfam02729,pfa	cd05650 m00185	TIGR01910 TIGR00658	3	1		1
arCOG04134 arCOG01033	E	Amino acid transport and metabolism Amino acid transport and metabolism	AroA AroE	5-enolpyruvylshikimate-3-phosphate synthase Shikimate 5-dehydrogenase	COG00128 COG00169	pfam00275 pfam08501,pfa	cd01556 cd01065	TIGR01356 TIGR00507			1	1
arCOG01025	E	Amino acid transport and metabolism	AroK2	Archaeal shikimate kinase	COG01685	pfam00288	m02774	TIGR01920			2	1
arCOG00434	E	Amino acid transport and metabolism	AsnB	Asparate semanderive denyarogenase Asparagine synthase (glutamine-hydrolyzing)	COG00367	pfam00733	1102774	1000978	1		1	1
arCOG00071 arCOG01594	E	Amino acid transport and metabolism Amino acid transport and metabolism	AsnB CarB	Asparagine synthase (glutamine-hydrolyzing) Carbamoylphosphate synthase large subunit	COG00367 COG00458	ptam00733 pfam00289,pfa	cd01991 m02786,pfam02	TIGR01536 TIGR01369	1	1	2	1
arCOG00073 arCOG01430	E	Amino acid transport and metabolism Amino acid transport and metabolism	CysH CysK	PAPS reductase related enzyme fused to RNA-binding PUA domain and ferre Cysteine synthase	COG00175 COG00031	pfam01507 pfam00291	cd01713 cd01561	TIGR00434 TIGR01136		1	2	1
arCOG00755	E	Amino acid transport and metabolism	DadA	Glycine/D-amino acid oxidase (deaminating)	COG00665	pfam01266	cd00312		2	1	2	
arCOG03109	E	Amino acid transport and metabolism Amino acid transport and metabolism	DdaH	N-Dimethylarginine dimethylaminobydrolase	COG01834	pfam02274	000011	TIGR01078	0	1	3	
arCOG05395 arCOG00915	E	Amino acid transport and metabolism Amino acid transport and metabolism	GabT	Formiminotetranydrofolate cyclodeaminase 4-aminobutyrate aminotransferase or related aminotransferase	COG03404 COG00160	pfam00202	cd00610	TIGR00707		1	2	
arCOG01303 arCOG00756	E	Amino acid transport and metabolism Amino acid transport and metabolism	GcvH GcvT	Glycine cleavage system H protein (lipoate-binding) Glycine cleavage system T protein (aminomethyltransferase)	COG00509 COG00404	pfam01597 pfam01571,pfa	cd06848 m08669	TIGR00527 TIGR00528			1	1
arCOG00757 arCOG02706	E	Amino acid transport and metabolism Amino acid transport and metabolism	GcvT GloA	Glycine cleavage system T protein (aminomethyltransferase)	COG00404	pfam01571,pfa nfam13669	m08669	TIGR00528 TIGR03081	1		2	
arCOG00095	E	Amino acid transport and metabolism	GltB	Glutamate synthase domain 1	COG00067	pfam00310	cd01907	TIGR01134		1	1	1
arCOG00096	E	Amino acid transport and metabolism	GltB	Glutamate synthese domain 3 Glutamate synthese domain 2 and ferredoxin domain	COG00069	pfam01645	cd02808			1	1	
arCOG07384 arCOG04273	E	Amino acid transport and metabolism Amino acid transport and metabolism	HisB HisC	Histidinol phosphatase or related phosphatase Histidinol-phosphate/aromatic aminotransferase or cobyric acid decarboxyla	LOG00241 COG00079	ptam13242 pfam00155	cd01427 cd00609	TIGR01656 TIGR01140			1	1
arCOG01799 arCOG01798	E	Amino acid transport and metabolism Amino acid transport and metabolism	HisJ HisM	ABC-type amino acid transport/signal transduction system, periplasmic comp ABC-type amino acid transport system, permease component	COG00834 COG00765	pfam00497 pfam00528	cd13624 cd06261	TIGR01096 TIGR03003			1	
arCOG04671 arCOG04670	E	Amino acid transport and metabolism Amino acid transport and metabolism	HutH HutU	Histidine ammonia-lyase Urocanate hydratase	COG02986 COG07987	pfam00221 pfam01175	cd00332	TIGR01225 TIGR01228	5		2	
arCOG01512	E	Amino acid transport and metabolism	HyuB	N-methylhydantoinase B/acetone carboxylase, alpha subunit	COG00146	pfam02538	cd14050		1			
arCOG04//9	E	Amino acid transport and metabolism	ilvE	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate ly	COG01446	pfam01112	cd01558	TIGR01122	3		1	1
arCOG01698 arCOG01021	E	Amino acid transport and metabolism Amino acid transport and metabolism	LeuC LivK	Homoaconitate hydratase/3-isopropylmalate dehydratase large subunit fami ABC-type branched-chain amino acid transport system, periplasmic compone	COG0065	ptam00330 pfam13458	cd01583 cd06268	IIGKU2086			1	
arCOG00243 arCOG00861	E	Amino acid transport and metabolism Amino acid transport and metabolism	LYS9 LysC	Saccharopine dehydrogenase or related enzyme Aspartokinase	COG01748 COG00527	pfam03435			5	1		1
arCOG00060	E	Amino acid transport and metabolism	MetC MfnA	Cystathionine beta-lyase/cystathionine gamma-synthase	COG00626	pfam01053	cd00614	TIGR02080 TIGR03812	1			1
arCOG04322	E	Amino acid transport and metabolism	PepB	Leucyl aminopeptidase	COG00260	pfam02789,pfa	cd00433	TIC 000404			2	*
arcog04758 arCOG01000	E	Amino acid transport and metabolism Amino acid transport and metabolism	rep⊩ PepP	Congoencopeptidase P Xaa-Pro aminopeptidase	COG01164 COG00006	pramU8439,pfa pfam01321,pfa	cd01092	TIGR00500			1	
arCOG00177 arCOG06322	E	Amino acid transport and metabolism Amino acid transport and metabolism	PotA PutA	ABC-type spermidine/putrescine transport system, ATPase component Proline dehydrogenase	COG03842	pfam00005 pfam01619	cd03301	IIGK03265	3		1	1
arCOG01316 arCOG00088	E	Amino acid transport and metabolism Amino acid transport and metabolism	PutP PuuD	Na+/proline symporter Gamma-glutamyl-gamma-aminobutyrate hydrolase PuuD (putrescine degrad	COG00591 COG02071	pfam00474 pfam07722	cd10322 cd01745	TIGR00813 TIGR00888	1		4	
arCOG01947 arCOG01948	E	Amino acid transport and metabolism Amino acid transport and metabolism	RhtB RhtB	Threonine efflux protein Threonine efflux protein	COG01280 COG01280	pfam01810 pfam01810		TIGR00949	3			1
arCOG01158	E	Amino acid transport and metabolism	SerB	Phosphoserine phosphatase	COG00560	pfam00702	cd08225	TIGR01491	5	1	1	2
arCOG05121	E	Amino acid transport and metabolism Amino acid transport and metabolism	TesA	Lysophospholipase L1 or related esterase	COG01063 COG02755	pramU8240,pfa	cau8235	1 IGNUU092	1			1
arCOG01434 arCOG02014	E	Amino acid transport and metabolism Amino acid transport and metabolism	ThrC TrpE	Threonine synthase and cysteate synthase Anthranilate/para-aminobenzoate synthase component I	COG00498 COG00147	pfam01155,pfa pfam04715,pfa	cd01563 m00425	TIGR02605,TIGR00260 TIGR01820			1	1
arCOG07581 arCOG00138	G	Carbohydrate transport and metabolism Carbohydrate transport and metabolism	-	Glycosyl hydrolase family 18, contains cellulose binding domain MFS family permease	COG03325 COG00477	pfam02839,pfa	cd12215,cd001	46,cd06548	7		3	3
arCOG03641	G	Carbohydrate transport and metabolism	PfkA Glo ^E	6-phospholytotokinase Glucerol untake facilitator or solated parmages (Maria Interior Barris - Barris	COG00205	pfam00365	cd00363	TIGR02483	5	1	1	
arCOG004431	G	Carbohydrate transport and metabolism	-	MFS family permase	COG00477	pfam07690	cd06174	T(CR04224	3	1	-	
arCOG01349 arCOG00271	G	Larbonydrate transport and metabolism Carbohydrate transport and metabolism	SuhB RhaT	Arcnaeai fructose-1,6-bisphosphatase or related enzyme of inositol monopho Permease of the drug/metabolite transporter (DMT) superfamily	LOG00483 COG00697	ptam00459 pfam00892,pfa	cd01642 m00892	тідк01331 ТіGR00950	2	1	1	1
arCOG02876 arCOG01122	G	Carbohydrate transport and metabolism Carbohydrate transport and metabolism	CDA1 RpiA	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family Ribose 5-phosphate isomerase	COG00726 COG00120	pfam01522 pfam06026	cd10941 cd01398	TIGR03006 TIGR00021	2			
arCOG00052	G	Carbohydrate transport and metabolism	Pgi	Glucose-6-phosphate isomerase Phosphoenologyriyate synthase/ovrivate phosphate divinase	COG00166	pfam01380,pfa	cd05017,cd056	TIGR02128 TIGR01418	1	1	1	1
arCOG04934	G	Carbohydrate transport and metabolism	-	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family	COG00726	,			1	-	_	
arcog0/840 arcog00767	G	Carbohydrate transport and metabolism Carbohydrate transport and metabolism	{ManB}	Phosphomannomutase	COG03325 COG01109	pramU2839,pfa pfam02878,pfa	cd12215,cd001 cd03087	40,0000043 TIGR03990	1		2	1
arCOG00274 arCOG00014	G	Carbohydrate transport and metabolism Carbohydrate transport and metabolism	RhaT RbsK	Permease of the drug/metabolite transporter (DMT) superfamily Sugar kinase, ribokinase family	COG00697 COG00524	ptam00892,pfa pfam00294	m00892 cd01174	TIGR00950 TIGR02152		1	2	
arCOG05046 arCOG00496	G	Carbohydrate transport and metabolism Carbohydrate transport and metabolism	Rpe Pgk	Pentose-5-phosphate-3-epimerase 3-phosphoglycerate kinase	COG00036 COG00126	pfam00834 pfam00162	cd00429 cd00318	TIGR01163			1	1
arCOG02796	G	Carbohydrate transport and metabolism	-	Glucose/sorbosone dehydrogenase MFS family nermease	COG02133	pfam07995	cd06174	TIGR03606			1	1
	-	Carbohydrate transport and metabolism	GalK	Galactokinase	COG00153	nfam10509 nfa	m08544	TIGR00131			-	1

arCOG04120	G	Carbohydrate transport and metabolism	PykF	Pyruvate kinase	COG00469	pfam00224,pfa	cd00288	TIGR01064				1
arCOG05061 arCOG05412	G	Carbohydrate transport and metabolism Carbohydrate transport and metabolism	TalA BglB	Transaldolase Beta-glucosidase/6-phospho-beta-glucosidase/beta-galactosidase	COG00176 COG02723	pfam00923 pfam00232	cd00956	TIGR00875 TIGR03356				1
arCOG04339 arCOG04263	G H	Carbohydrate transport and metabolism	TagG	ABC-type polysaccharide/polyol phosphate export systems, permease comp Pantoate kinase	COG01682	pfam01061 nfam00288			4		1	1
arCOG00021	н	Coenzyme transport and metabolism	-	Predicted transcriptional regulator fused phosphomethylpyrimidine kinase,	COG01992	pfam10120		TICBOALOO	4			1
arCOG01484 arCOG01940	н	Coenzyme transport and metabolism Coenzyme transport and metabolism	BirA	Biotin-(acetyl-CoA carboxylase) ligase	COG01985 COG00340	pfam01872 pfam03099		TIGR01208	3	1	2	
arCOG00040 arCOG02172	H	Coenzyme transport and metabolism Coenzyme transport and metabolism	Hpt1 QueD	Hypoxanthine phosphoribosyltransferase 6-pyruvoyl-tetrahydropterin synthase	COG02236 COG00720	pfam00156 pfam01242	cd06223 cd00470	TIGR01251 TIGR03367	3	1		
arCOG01045 arCOG00226	H	Coenzyme transport and metabolism Coenzyme transport and metabolism	CoaE TbpA	Dephospho-CoA kinase ABC-type thiamine transport system, periplasmic component	COG00237 COG04143	pfam13343	cd02022 cd13545	TIGR01254	3		1	1
arCOG01589	Н	Coenzyme transport and metabolism	RimK MthC1	Glutathione synthase/glutaminyl transferase/alpha-L-glutamate ligase	COG00189	pfam08443	cd02065	TIGR02144	2			
arCOG00476	н	Coenzyme transport and metabolism	UbiA	4-hydroxybenzoate polyprenyltransferase or related prenyltransferase	COG00382	pfam01040	cd13961	TIGR01476	1		1	1
arCOG00972 arCOG04139	н	Coenzyme transport and metabolism Coenzyme transport and metabolism	ApbA	Ketopantoate reductase	COG01058 COG01893	pfam01467 pfam02558,pfa	m08546	TIGR00745	1	1	1	1
arCOG01522 arCOG01490	H	Coenzyme transport and metabolism Coenzyme transport and metabolism	HemY FolA	Protoporphyrinogen oxidase Dihydrofolate reductase	COG01232 COG00262	pfam01593 pfam00186	cd00209	TIGR00562	1		1	
arCOG01223 arCOG01320	H	Coenzyme transport and metabolism Coenzyme transport and metabolism	CAB4 RibB	Phosphopantetheine adenylyltransferase 3,4-dihydroxy-2-butanone 4-phosphate synthase	COG01019 COG00108	pfam01467 pfam00926	cd02164	TIGR00125 TIGR00506		1	1	
arCOG01671 arCOG02939	H	Coenzyme transport and metabolism Coenzyme transport and metabolism	UbiD PhhB	3-polyprenyl-4-hydroxybenzoate decarboxylase or related decarboxylase Pterin-da-carbinolamine debydratase	COG00043	pfam01977	cd00488	TIGR03701		1	1	
arCOG01482	н	Coenzyme transport and metabolism	NadC	Nicotinate-nucleotide pyrophosphorylase	COG00157	pfam02749,pfa	cd01572	TIGR00078			1	
arCOG03838	H	Coerzyme transport and metabolism	PqqD	Coencymentacystemie hydrolase	0000433	pfam05402	000401	11000550		1		1
arCOG00069 arCOG01942	н	Coenzyme transport and metabolism Coenzyme transport and metabolism	LipB	NH3-dependent NAD+ synthetase Lipoate-protein ligase B	COG00171 COG00321	pfam02540 pfam03099	C000553	TIGR00552 TIGR00214		1	3	1
arCOG00480 arCOG00489	H	Coenzyme transport and metabolism Coenzyme transport and metabolism	MenA PduO	1,4-dihydroxy-2-naphthoate octaprenyltransferase Cob(I)alamin adenosyltransferase	COG01575 COG02096	pfam01040 pfam01923	cd13962	TIGR02235 TIGR00636				1
arCOG02817 arCOG01704	H	Coenzyme transport and metabolism Coenzyme transport and metabolism	FolC Dfp	Folylpolyglutamate synthase and Dihydropteroate synthase Phosphopantothenoylcysteine synthetase/decarboxylase	COG00285 COG00452	pfam08245,pfa pfam02441,pfa	cd00739 m04127	TIGR01499,TIGR01496 TIGR00521			2	1
arCOG00214 arCOG00534	H	Coenzyme transport and metabolism Coenzyme transport and metabolism	MoaB MoaE	Molybdopterin biosynthesis enzyme Molybdopterin converting factor, large subunit	COG00521 COG00314	pfam00994 pfam02391	cd00886 cd00756	TIGR02667				
arCOG00930 arCOG01530	H	Coenzyme transport and metabolism	MoaA MoaC	Molybdenum cofactor biosynthesis enzyme Molybdenum cofactor biosynthesis enzyme	COG02896	pfam04055,pfa nfam01967	cd01335	TIGR02668 TIGR00581			1	
arCOG01872	Н	Coenzyme transport and metabolism	MobA	Molybdopterin-guanine dinucleotide biosynthesis protein A	COG00746	pfam12804	cd02503	TIC 0004 77			2	
arCOG00217 arCOG02199	н	Coenzyme transport and metabolism Coenzyme transport and metabolism	MCr1	NAD(P)H-flavin reductase	COG00543	pramu3453,pra	000887	11GR00177			3	
arCOG01676 arCOG01873	н	Coenzyme transport and metabolism Coenzyme transport and metabolism	MobA	Dinucleotide-utilizing enzyme involved in molybdopterin and thiamine biosy GT-A family glycosyltransferase involved in molybdopterin guanine dinucleo	t COG02068	pfam00899,pfa pfam12804	cd00757 cd04182	TIGR03310			1	1
arCOG04348 arCOG00572	H	Coenzyme transport and metabolism Coenzyme transport and metabolism	UbiE NadB	Ubiquinone/menaquinone biosynthesis C-methylase UbiE Aspartate oxidase	COG02226 COG00029	pfam08241 pfam00890,pfa	m02910	TIGR01934 TIGR00551			1 1	1
arCOG02624 arCOG00570	H C	Coenzyme transport and metabolism Energy production and conversion	PaaK FixC	Coenzyme F390 synthetase Dehydrogenase (flavoprotein)	COG01541 COG00644	pfam13450		TIGR02304 TIGR02032	6	2	2	1
arCOG02921	c	Energy production and conversion	PetE	Plastocyanin Givreronhoshoryl diester phosphodiesterase	COG03794	pfam00127	cd04220	TIGR03102	4			
arCOG00519	c	Energy production and conversion	FIdA	Plavodoxin NADH-EMM ovidoradurtaca Butt florido roductore (DBAC Artan) (c	COG00716	pfam12682	00000	TIGR03615	3		1	1
arCOG02017 arCOG02929	C	Energy production and conversion	PetE	Plastocyanin	COG01853 COG03794	pfam00127	cd13921	TIGR02657	1		ī	1
arCOG02398 arCOG01599	C C	Energy production and conversion Energy production and conversion	CcdA PorB	Cytochrome c biogenesis protein Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidore	COG01013	ptam02683 pfam02775,pfa	cd03375	TIGR02177	1			
arCOG00571 arCOG01551	C C	Energy production and conversion Energy production and conversion	SdhA NuoC	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit NADH dehydrogenase subunit C	COG01053 COG00852	pfam00890,pfa pfam00329	m02910	TIGR03378,TIGR01812 TIGR01961	1		4	1
arCOG01539 arCOG01540	C C	Energy production and conversion Energy production and conversion	NuoL NuoN	NADH dehydrogenase subunit L NADH dehydrogenase subunit N	COG01009 COG01007	pfam00662,pfa pfam00361	m00361	TIGR01974 TIGR01770		1		1
arCOG02095 arCOG04138	C C	Energy production and conversion Energy production and conversion	OadA1 Ntpl	Pyruvate/oxaloacetate carboxyltransferase Archaeal/vacuolar-type H+-ATPase subunit I	COG05016 COG01269	pfam00682,pfa pfam01496	cd07937	TIGR01108			1	1
arCOG00869	c	Energy production and conversion	NtpE	Archaeal/vacuolar-type H+-ATPase subunit E	COG01390	ofom00676	cd03000	T/CD02494		1	1	
arCOG01034 arCOG04279	c	Energy production and conversion	-	Swiveling domain associated with predicted aconitase	COG01786	pfam01989	cd01356	110103181		1	-	
arCOG00573 arCOG02618	C C	Energy production and conversion Energy production and conversion	SdhA -	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit Ferredoxin	COG01053 COG01146	pfam00890,pfa pfam13187,pfa	m02910 m12139	TIGR02061 TIGR02060		1		
arCOG01235 arCOG00446	C C	Energy production and conversion Energy production and conversion	CyoA FixA	Heme/copper-type cytochrome/quinol oxidase, subunit 2 Electron transfer flavoprotein, beta subunit	COG01622 COG02086	pfam00116 pfam01012	cd13918 cd01714	TIGR02866		1	1	
arCOG01252 arCOG00288	С	Energy production and conversion	PutA NfnB	Lactaldehyde dehydrogenase, Succinate semialdehyde dehydrogenase or ot Nitroreductase	COG01012	pfam00171	cd07088	TIGR01780 TIGR02476		1	3	
arCOG00458	c	Energy production and conversion	Qor	NADPH:quinone reductase or related Zn-dependent oxidoreductase	COG00604	pfam08240,pfa	cd08264	TIGR02824			1	1
arCOG00903	c	Energy production and conversion	CyoC	Heme/copper-type cytochrome/quind oxidase, subunit 3	COG01845	pfam00510	cd00386	TIGR02842			2	2
arCOG00338 arCOG03363	C C	Energy production and conversion Energy production and conversion	SdhC NtpF	Succinate dehydrogenase subunit C Archaeal/vacuolar-type H+-ATPase subunit H	COG02048 COG02811	ptam02754,pta	m02754	TIGR03288			2	1
arCOG00962 arCOG01617	C C	Energy production and conversion Energy production and conversion	FrdB Tas	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit Aryl-alcohol dehydrogenase related enzyme	COG00479 COG00667	pfam13085,pfa pfam00248	m13183 cd06660	TIGR00384 TIGR01293			2	1
arCOG04358 arCOG01236	C C	Energy production and conversion Energy production and conversion	FdhD CyoA	Uncharacterized protein required for formate dehydrogenase activity Heme/copper-type cytochrome/quinol oxidase, subunit 2	COG01526 COG01622	pfam02634 pfam00116	cd13842	TIGR00129 TIGR02866			2	1
arCOG02304 arCOG02476	c	Energy production and conversion	Mct/CaiB	Succinyl-CoA:mesaconate CoA-transferase or predicted acyl-CoA transferase Heterodisulfide reductase subunit A: ferredoxin domain	COG01804	pfam02515 nfam07992 nfa	m13237 nfam13	TIGR03253 TIGR03140 TIGR04105 TIGR03385 TIGR	04105			1
arCOG02842	C	Energy production and conversion	Fdx	Ferredoxin	COG00633	pfam00111	cd00207	TIGR02008			1	
arCOG04750	P	Inorganic ion transport and metabolism	-	Sirohydrochlorin iron chelatase fused to [2Fe-2S] Ferredoxin	COG02138	pfam01903,pfa	cd03416,cd034	14,cd02980	4		1	
arCOG02881 arCOG00206	P	Inorganic ion transport and metabolism Inorganic ion transport and metabolism	PhnC	Ca2+/Na+ antiporter ABC-type phosphate/phosphonate transport system, ATPase component	COG00530 COG03638	pfam01699,pfa pfam00005	cd03256	TIGR02315	3	1	1	
arCOG02021 arCOG04233	P	Inorganic ion transport and metabolism Inorganic ion transport and metabolism	PspE FepB	Rhodanese-related sulfurtransferase ABC-type Fe3+-hydroxamate transport system, periplasmic component	COG00607 COG00614	pfam00581 pfam01497	cd00158 cd01143	TIGR02981 TIGR04281	2	1		3
arCOG00163 arCOG02763	P	Inorganic ion transport and metabolism Inorganic ion transport and metabolism	ThiP -	ABC-type Fe3+ transport system, permease component Heavy-metal-associated domain (HMA)	COG01178 COG02608	pfam00528,pfa pfam00403	cd06261,cd062 cd00371	TIGR03262 TIGR02052	2		1	1
arCOG00238 arCOG01005	P	Inorganic ion transport and metabolism	ArsB Lral	Na+/H+ antiporter NhaD or related arsenite permease ABC-type metal ion transport system, perinlasmic component/surface adher	COG01055	pfam03600 nfam01297	cd01117	TIGR00785 TIGR03772 TIGR03772	1			
arCOG02851	P	Inorganic ion transport and metabolism	{NirD}	Ferredoxin subunit of nitrite reductase or ring-hydroxylating dioxygenase	COG02146	pfam00355	cd03467		1			
arCOG08934 arCOG04487	P	Inorganic ion transport and metabolism	KatG	Catalase (peroxidase I)	COG02128 COG00376	pfam00141,pfa	cd00649,cd006	TIGR00198	1			
arCUG02497 arCOG00318	P	morganic ion transport and metabolism Inorganic ion transport and metabolism	- PhoU	upoprotein NosD tamity, contains CASH domains Phosphate uptake regulator	COG00704	ptam12708 pfam04014,pfa	m01895,pfam01	TIGR02135	1		1	1
arCOG06837 arCOG01959	P	Inorganic ion transport and metabolism Inorganic ion transport and metabolism	- TrkA	Lipoprotein NosD family, contains CASH domains TrkA, K+ transport system, NAD-binding component	COG03420 COG00569	pfam05048 pfam02254,pfa	m02080,pfam02	TIGR04247,TIGR03024 2254,pfam02080	1		1	2
arCOG04231 arCOG02499	P	Inorganic ion transport and metabolism Inorganic ion transport and metabolism	CutA -	Uncharacterized protein involved in tolerance to divalent cations Lipoprotein NosD family, contains CASH domains	COG01324 COG03420	pfam03091 pfam13229,pfa	m13229	TIGR04247			3	1
arCOG02265 arCOG04397	P	Inorganic ion transport and metabolism Inorganic ion transport and metabolism	CorA AmtB	Mg2+ and Co2+ transporter Ammonia permease	COG00598 COG00004	pfam01544 pfam00909	cd12828	TIGR00383 TIGR00836		_		
arCOG02764	P	Inorganic ion transport and metabolism	CopZ PhnD	Copper-ion-binding protein ABC-type phosphate/phosphonate transport system periplasmic composed	COG02608	pfam00403	cd00371 cd01071	TIGR02052 TIGR01098			1	1
arCOG02050	P	Inorganic ion transport and metabolism	-	Suffice exporter, TauE/SafE family	COG00730	pfam01925	-403337	7000000			2	1
arCOG00202 arCOG01040	P	Inorganic Ion transport and metabolism Inorganic ion transport and metabolism	CysC	Adenylylsulfate kinase or related kinase	COG01122 COG00529	pram00005 pfam01583	cd02027	TIGR00455			1	1
arCUG01096 arCOG01576	P	morganic ion transport and metabolism Inorganic ion transport and metabolism	CCC1 ZntA	rreucted Fe2+/Mn2+ transporter, VIT1/CCC1 family Cation transport ATPase	COG02217	ptam02915,pfa pfam00403,pfa	cdU1044,cd024 cd00371,cd014	31 TIGR00003,TIGR01511			1	1
arCOG02062 arCOG05356	P	Inorganic ion transport and metabolism Inorganic ion transport and metabolism	TusA Fes	TusA-related sulfurtransferase Enterochelin esterase or related enzyme	COG00425 COG02382	pfam01206 pfam00756	cd00291				1	1
arCOG11907 arCOG00230	P	Inorganic ion transport and metabolism Inorganic ion transport and metabolism	-	Fe-S metabolizm associated domain Periplasmic molybdate-binding protein/domain	COG01910	pfam02657 pfam00126,pfa	m12727	TIGR03391 TIGR00637			1	1
arCOG02849 arCOG04559	P	Inorganic ion transport and metabolism Inorganic ion transport and metabolism	ArsA EmrF	Oxyanion-translocating ATPase Membrane transporter of cations and rationic drugs	COG00003	pfam02374 pfam00893	cd02035	TIGR00345				1
arCOG00164	P	Inorganic ion transport and metabolism	Kch	Kef-type K+ transport system, predicted NAD-binding component	COG01226	pfam07885	cd06261	TIG902138			1	2
arCOG00168	P	Inorganic ion transport and metabolism	PstA	ABC-type phosphate transport system, permease component ABC-type phosphate transport system, permease component	COG00581	- fa Tan - '	cd06261	TIGR00974			1	
arCOG00231	P	Inorganic Ion transport and metabolism Inorganic ion transport and metabolism	PstB	ABC-type prospriate transport system, periplasmic component ABC-type phosphate transport system, ATPase component	COG00226 COG01117	pram12849 pfam00005	cd03260	TIGR00972			1	
arCOG00232 arCOG02555	P	inorganic ion transport and metabolism Inorganic ion transport and metabolism	PhoU -	Prospnate uptake regulator Lipoprotein NosD family, contains CASH domains	COG00704 COG03420			1/6K02135			1	
arCOG02852 arCOG03797	P	Inorganic ion transport and metabolism Inorganic ion transport and metabolism	{NirD}	Ferredoxin subunit of nitrite reductase or ring-hydroxylating dioxygenase Ferritin-like domain	COG02146 COG01633	pfam00355 pfam05763	cd03528	TIGR02377			1	
arCOG01259 arCOG00670	1	Lipid transport and metabolism Lipid transport and metabolism	FabG PgsA	Short-chain alcohol dehydrogenase Phosphatidylglycerophosphate synthase	COG01028 COG00558	pfam01066	cd05233	TIGR01830	5	1	3	5
arCOG01767 arCOG00671	1	Lipid transport and metabolism Lipid transport and metabolism	PksG PssA	3-hydroxy-3-methylglutaryl CoA synthase Phosphatidylserine synthase	COG03425 COG01183	pfam08545,pfa pfam01066	cd00827	TIGR00748 TIGR04217	5	_	1	2
arCOG13950	1	Lipid transport and metabolism	- F&A1	Phosphatidylglycerophosphatase A	CO601022	pfam04608	cd06971	TIG801923	4			1
arCOG01278	1	Lipid transport and metabolism	Paal	Acetyl-CoA acetyltransferase	COG00183	nfam00301	cd00829	TIGR01930	4	4	_	1
arcuG08932 arCOG00860	1	Lipid transport and metabolism	- -	Isopentenyl phosphate kinase, enzyme of modified mevalonate pathway	COG01597 COG01608	pram00781 pfam00696	cd04241	TIGR02075	3	1	2	1
arCOG01028 arCOG02936	1	Lipid transport and metabolism Lipid transport and metabolism	ERG12 ERG9	Mevalonate kinase Phytoene/squalene synthetase	COG01577 COG01562	ptam00288 pfam00494	cd00683	11GR03465	3	1	3	1
arCOG01137 arCOG01879		Lipid transport and metabolism Lipid transport and metabolism	FadM	Acyl-CoA thioesterase FadM Dolichol kinase family protein	COG00824 COG00170	pfam13279	cd00586	TIGR00051	3			1
arCOG02245 arCOG04351	1	Lipid transport and metabolism Lipid transport and metabolism	-	Cytidylyltransferase family enzyme Predicted membrane associated lipid hydrolase, neutral ceramidase superfa	COG01836 rCOG03356	pfam01940 pfam09843		TIGR00297	2	1	2	1
arCOG00249 arCOG00239	1	Lipid transport and metabolism Lipid transport and metabolism	FadB CaiD	3-hydroxyacyl-CoA dehydrogenase, some fused to Enoyl-CoA hydratase Enoyl-CoA hydratase/carnithine racemase	COG01250 COG01024	pfam02737,pfa pfam00378	cd06558 cd06558	TIGR02437,TIGR02437 TIGR03210	1		1	
arCOG01880	1	Lipid transport and metabolism	SEC59	Dolichol kinase	COG00170	ofom00504	cd12110	TICP01022	1			
arcocoone		epie eensport and metadolism	cail	www.www.syncnecase.pawie-romning//AMIP-acid ligase II	00000318	P1011100501	C015113	1101101223	1			

arCOG04106 arCOG01532		Lipid transport and metabolism Lipid transport and metabolism	CdsA UppS	CDP-diglyceride synthetase Undecaprenyl pyrophosphate synthase	COG00575 COG00020	pfam01864 pfam01255	cd00475	TIGR00055				1	
arCOG01843 arCOG01085 arCOG04260	-	Lipid transport and metabolism Lipid transport and metabolism Lipid transport and metabolism Lipid transport and metabolism	- PcrB HMG1	Steroi carrier protein (S)-3-O-geranylgeranylglyceryl phosphate synthase, TIM-barrel fold Hydroxymethylglutaryl-CoA reductase	COG03255 COG01646 COG01257	pfam02036 pfam01884 nfam00368	cd02812	TIGR01768 TIGR00533		1		2	1
arCOG01710 arCOG02705	-	Lipid transport and metabolism Lipid transport and metabolism Lipid transport and metabolism	Sbm	Methylmalonyl-CoA mutase, C-terminal domain/subunit (cobalamin-binding Acetyl-CoA carboxylase, carboxyltransferase component	COG02185 COG04799	pfam02310 pfam01039	cd02071	TIGR00640 TIGR01117				2	1
arCOG01707 arCOG01282		Lipid transport and metabolism Lipid transport and metabolism	CaiA PaaJ	Acyl-CoA dehydrogenase Acetyl-CoA acetyltransferase	COG01960 COG00183	pfam02771,pfa pfam00108,pfa	a cd00567 a cd00751	TIGR03207 TIGR01930				1	1
arCOG04470 arCOG01650	1	Lipid transport and metabolism Lipid transport and metabolism	Psd PldB	Phosphatidylserine decarboxylase Lysophospholipase, alpha-beta hydrolase superfamily	COG00688 COG02267	pfam02666 pfam12697		TIGR00164 TIGR03695				1	1
arCOG02039 arCOG00241		Lipid transport and metabolism Lipid transport and metabolism Lipid transport and metabolism	Cls CaiD EabG	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase or n Enoyl-CoA hydratase/carnithine racemase Chart chais alcabal debudgegeggs	COG01502 COG01024	pfam13091,pta pfam00378	cd07493,cd091 cd06558 cd05222	27,cd09128 TIGR02280				2	1
arCOG03951 arCOG00674		Lipid transport and metabolism Lipid transport and metabolism	PgpB PgsA	Membrane-associated phospholipid phosphatase Phosphatidylglycerophosphate synthase	COG00671 COG00558	pfam01569 pfam01066	cd03386	10101050				3	-
arCOG06122 arCOG00911	F	Lipid transport and metabolism Nucleotide transport and metabolism	Acs PyrB	Acyl-coenzyme A synthetase/AMP-(fatty) acid ligase Aspartate carbamoyltransferase, catalytic chain	COG00365 COG00540	pfam00501 pfam02729,pfa	cd05959 am00185	TIGR02262 TIGR00670	3			1	
arCOG04415 arCOG01038	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	PurD Fap7	Phosphoribosylamine-glycine ligase Broad-specificity NMP kinase	COG00151 COG01936	pfam02844,pfa pfam13238	am01071,pfam0	TIGR00877	3	1		1	1
arCOG00018 arCOG01891 arCOG00081	F	Nucleotide transport and metabolism Nucleotide transport and metabolism Nucleotide transport and metabolism	Nnr2 Tmk PvrE	NAD(P)H-hydrate repair enzyme Nnr, NAD(P)H-hydrate dehydratase domain Thymidylate kinase Orotidioa-Si-phosphate derarboyulase	COG00063 COG00125	pfam03853,pta pfam02223	cd01171 cd01672 cd04725	TIGR00197,TIGR00196 TIGR00041 TIGR01740	3	1		1	
arCOG03575 arCOG00087	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	- GuaA	Polyphosphate kinase 2 GMP synthase - Glutamine amidotransferase domain	COG02326 COG00518	pfam03976,pfa pfam00117	cd01672 cd01672	TIGR03708 TIGR00888	3				
arCOG00093 arCOG04276	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	PurF NrdA	Glutamine phosphoribosylpyrophosphate amidotransferase Ribonucleotide reductase, alpha subunit	COG00034 COG00209	pfam00310,pfa pfam03477,pfa	cd00715,cd062 cd02888	TIGR01134 TIGR02504	2				
arCOG00090 arCOG02825	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	GuaA PurN	GMP synthase - Glutamine amidotransferase domain Folate-dependent phosphoribosylglycinamide formyltransferase PurN	COG00518 COG00299	pfam00117	cd01741	TIGR00888	1	1		3	2
arCOG04184 arCOG00028	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	RdgB -	Inosine/xanthosine triphosphate pyrophosphatase, all-alpha NTP-PPase fam Orotate phosphoribosyltransferase homolog	COG00127 COG00856	pfam01725 pfam00156	cd00515 cd06223	TIGR00042 TIGR02985,TIGR00336			1	2	1
arCOG00029 arCOG00419 arCOG04462	F	Nucleotide transport and metabolism Nucleotide transport and metabolism Nucleotide transport and metabolism	Hit PurS	Unclate phosphorbosyltransierase HIT family hydrolase Phosphoribosyltramylelycinamidiae (FGAM) synthase. PurS component	COG00481 COG00537 COG01828	pfam01230	cd01275	TIGR00336		1		3	1
arCOG04421 arCOG00641	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	PurC PurL	Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain	COG00152 COG00046	pfam01259 pfam00586,pfa	cd01415 cd02203,cd022	TIGR00081 TIGR01736				1	1
arCOG01565 arCOG04346	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	NmA -	nanoRNase/pAp phosphatase, hydrolyzes c-di-AMP and oligoRNAs 5-formaminoimidazole-4-carboxamide-1-beta-D-ribofuranosyl 5'-monophos	COG00618 COG01759	pfam01368 pfam06849,pfa	am06973	TIGR00877		1			1 2
arCOG00603 arCOG01037	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	PyrD Cmk	Dihydroorotate dehydrogenase Cytidylate kinase	COG00167 COG01102	pfam01180 pfam13189	cd04740 cd02020	TIGR01037 TIGR02173		1	1	2	1
arCOG00612 arCOG05252	F	Nucleotide transport and metabolism Nucleotide transport and metabolism Nucleotide transport and metabolism	GuaB	IMP dehydrogenase/GMP reductase Predicted secreted endonuclease distantly related to archaeal Holliday junct	COG00516 COG04741	pfam00478 pfam10107	cd00381	TIGR01302			1	2	
arCOG00030 arCOG00858 arCOG02013	F	Nucleotide transport and metabolism Nucleotide transport and metabolism Nucleotide transport and metabolism	PyrH DeoA	Ademine/guanine prosphorioosyntransferase or related PRPP-binding protein Uridylate kinase Thymidine phosphorylase	COG00528 COG00213	pfam00156 pfam00696 pfam01568.pfa	cd04253 cd04253 am02885.pfam0	TIGR02076 TIGR03327				1	1
arCOG04311 arCOG01075	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	-	5'-deoxynucleotidase, HD superfamily hydrolase NUDIX family hydrolase	COG01896 COG01051	pfam13023 pfam00293	cd04673					7	1 3
arCOG04320 arCOG01327	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	DeoC Pnp	Deoxyribose-phosphate aldolase Purine nucleoside phosphorylase	COG00274 COG00005	pfam01791 pfam01048	cd00959	TIGR00126 TIGR01694				1	
arCOG00695 arCOG01566	F	Nucleotide transport and metabolism Nucleotide transport and metabolism	SsnA NrnA	Cytosine deaminase or related metal-dependent hydrolase nanoRNase/pAp phosphatase, hydrolyzes c-di-AMP and oligoRNAs	COG00402 COG00618	pfam01979 pfam02254,pfa	cd01298 am01368,pfam0	TIGR03314 2272				2	2
arCOG03570 arCOG03570	Q Q	Secondary metabolites biosynthesis, transport and catabolism Secondary metabolites biosynthesis, transport and catabolism Secondary metabolites biosynthesis, transport and catabolism	raal -	SAM-dependent methyltransferase SAM-dependent methyltransferase	COG02050 COG00500 COG00500	pram03061 pfam12847 pfam08241	cd02440 cd02440	TIGR02021 TIGR01934	4			1	
arCOG01521 arCOG04347	Q	Secondary metabolites biosynthesis, transport and catabolism Secondary metabolites biosynthesis, transport and catabolism	-	SAM-dependent methyltransferase SAM-dependent methyltransferase	COG01233 COG00500	pfam13450 pfam08241	cd02440	TIGR01934	4 3	_			
arCOG03914 arCOG01523	a a	Secondary metabolites biosynthesis, transport and catabolism Secondary metabolites biosynthesis, transport and catabolism	Sufl -	Multicopper oxidase Phytoene dehydrogenase or related enzyme	COG02132 COG01233	pfam07732 pfam01593	cd11024	TIGR02376 TIGR03467	3			1	1
arCOG01778 arCOG01781	Q	Secondary metabolites biosynthesis, transport and catabolism Secondary metabolites biosynthesis, transport and catabolism	-	SAM-dependent methyltransferase SAM-dependent methyltransferase	COG00500 COG00500	pfam08241 pfam08241	cd02440 cd02440	TIGR02072 TIGR02072	3				
arCOG00235 arCOG00696	Q	Secondary metabolites biosynthesis, transport and catabolism Secondary metabolites biosynthesis, transport and catabolism	MhpD Hutl	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (Imidazolonepropionase or related amidohydrolase	COG00179 COG01228	pfam01557 pfam13147	cd01296	TIGR02303 TIGR01224	1			1 2	1
arCOG05015 arCOG01943	Q	Secondary metabolites biosynthesis, transport and catabolism Secondary metabolites biosynthesis, transport and catabolism Secondary metabolites biosynthesis, transport and catabolism	- PncA	Predicted ring-cleavage extradiol dioxygenase SAM-dependent methyltransferase Amidase related to nicrofinamidase	COG02514 COG00500 COG01335	pfam12681,pfa pfam12847 ofam00857	cd02440 cd00431	TIGR03211 TIGR03534 TIGR03614			1	1	1
arCOG04786 arCOG06169	Q	Secondary metabolites biosynthesis, transport and catabolism Secondary metabolites biosynthesis, transport and catabolism	-	1,2-phenylacetyl-CoA epoxidase, catalytic subunit Arylsulfotransferase family protein	COG03396	pfam05138 pfam05935		TIGR02158				1	1
arCOG01402	Q	Secondary metabolites biosynthesis, transport and catabolism	AgIP	SAM-dependent methyltransferase	COG00500	pfam05050		TIGR01444					2
POORLY CH/	RACTE	RIZED											
arCOG04364	S	Function unknown	-	Uncharacterized protein	COG01772	pfam04407			8				
arCOG04364 arCOG05839 arCOG02142	S S S	Function unknown Function Funct	-	Uncharacterized protein Uncharacterized protein Pheromone shutdown protein TraB, contains GTXH motif Anomcrae accession denotein with extra cellular in like demoin a compose	COG01772 COG01916	pfam04407 pfam01963		TIGR04292 TIGR00261	8 6 5	1		1	1
arCOG04364 arCOG05839 arCOG02142 arCOG02884 arCOG10153 arCOG04596	s S S S S	Function unknown	- - - - -	Uncharacterized protein Uncharacterized protein Pheromone shutdowa protein Traß, contains GTM motif Membrane associated protein with extracellular ig-like domain, a componen Uncharacterized membrane protein, DUF4122 family Uncharacterized membrane protein	COG01772 COG01916 COG04743	pfam04407 pfam01963 pfam13430	1107	TIGR04292 TIGR00261	8 6 5 5 5 4	1		1	1
arCOG04364 arCOG05839 arCOG02142 arCOG02884 arCOG10153 arCOG04596 arCOG03124 arCOG02087	s s s s s s s s s	Function unknown FUNCTION FUNCT	- - - - - - - - - - - - - - - - - - - - - - - - - -	Uncharacterized protein Uncharacterized protein Pheromone Shutchen portein Traß, contains GTxH motif Membrane associated protein with extracellular (sike domain, a componen Uncharacterized membrane protein, DUF4112 family Uncharacterized membrane protein Peritapedice repeats containing protein Predicted membrane protein	COG01772 COG01916 COG04743 COG01357 COG01470	pfam04407 pfam01963 pfam13430 pfam00805,pfa pfam10633,pfa	am00805 am13620, pfam1	TIGR04292 TIGR00261	8 6 5 5 4 4 4 4	1	1	1	1
arCOG04364 arCOG05839 arCOG02142 arCOG02884 arCOG10153 arCOG04596 arCOG03124 arCOG02087 arCOG10338 arCOG10344	S S S S S S S S S S	Function unknown Function Function Function Function Func	- - - - - - - - - - - - - - - - - - - - - - - -	Uncharacterized protein Uncharacterized protein Pheromone Shutcher portein Traß, contains GTxH motif Pheromone Shutcher portein DVI-1212 simily Uncharacterized membrane protein DVI-1212 simily Uncharacterized membrane protein Peratopesitie reposits containing protein Predicted membrane protein Uncharacterized protein Uncharacterized protein	COG01772 COG01916 COG04743 COG01357 COG01357 COG01470 SC.00914	pfam04407 pfam01963 pfam13430 pfam00805,pfa pfam10633,pfa pfam13517,pfa	am00805 am13620, pfam1 am13517, pfam1	TIGR04292 TIGR00261 0633 517,pfam13517	8 5 5 4 4 4 4 4 4	1		1 3 2 4	2
arCOG04364 arCOG05839 arCOG02142 arCOG02884 arCOG010153 arCOG0153 arCOG0124 arCOG02087 arCOG10338 arCOG10444 arCOG0533 arCOG04579	s s s s s s s s s s s s s s s	Function unknown	- - - - - - - - - - - -	Bucharacterikei protein Ducharacterikei protein Peromone Buckdown protein Traß, contains GTat motif. Peromone Buckdown protein Derut Buckdown, a componen Uncharacterizet amebraran protein, DP4112 Emaily Bucharacterizet amebraran protein, DP4112 Emaily Bucharacterizet amebraran protein, DP4112 Emaily Ducharacterizet amebraran protein Predicted membrane protein Uncharacterizet protein Uncharacterizet protein Ducharacterizet protein Ducharacterizet protein Ducharacterizet protein	COG01772 COG01916 COG04743 COG01357 COG01470 SC.00914 COG03361	pfam04407 pfam01963 pfam13430 pfam00805,pfa pfam10633,pfi pfam10633,pfi	am00805 am13620,pfam1 am13517,pfam1	TIGR04292 TIGR02261 0633 357.pfam13517	8 5 5 4 4 4 4 4 4 4 4 4 4 4 4	1		1 3 2 4	2
arCOG04364 arCOG02343 arCOG02142 arCOG02142 arCOG0153 arCOG0153 arCOG0124 arCOG01038 arCOG10388 arCOG10388 arCOG10383 arCOG04579 arCOG05495 arCOG04412 arCOG04412 arCOG043232	s s s s s s s s s s s s s s s s s s s	Function unknown	- - - - - - - - - - - - - - - - - - -	Joncharstenikel protein Dacharstenikel protein Pheromone Bulkadown protein Traß, contains GTaH motif Membrane associated protein with extracellular gike domain, a componen Uncharsteniker der meharare protein. DPHI12 Emily Bincharsteniker der meharare protein Perdiactel nembrane protein Perdiactel nembrane protein Uncharsteniker grotein Uncharsteniker grotein Uncharsteniker grotein Uncharsteniker grotein, DUI271 finmit Uncharsteniker grotein, contains N-terminal colled-coll domain Uncharsteniker grotein.	COG01772 COG01916 COG01916 COG01917 COG01357 COG01357 COG01470 SC.00914 COG03361 COG04004 COG04004 COG05652	pfam04407 pfam01963 pfam13430 pfam10633,pfam10633,pf pfam10633,pf pfam13517,pfa pfam09844 pfam09969 pfam04892	am00805 am13620,pfam1 am13517,pfam1	TIGR04292 TIGR04261 5633 5517.pfam13517	8 6 5 5 4 4 4 4 4 4 4 4 3 3 3	1		1	2
arC0G04364 arC0G05839 arC0G02142 arC0G01142 arC0G01153 arC0G01254 arC0G01254 arC0G01267 arC0G0144 arC0G04579 arC0G04579 arC0G04457 arC0G04457 arC0G04258 arC0G04258 arC0G04258	S S S S S S S S S S S S S S S S S S S	Punction unknown	- - - - - - - - - - - - - - - - - - -	Joncharsteried protein Ducharsteried protein Pierconcen glutidown protein YTAB, costains GTM motif Membrane associated protein with extracellular ig-like domain, a componen Uncharsteried membrane protein, DUF412 family Uncharsteried membrane protein Predicted membrane protein Ducharsteried grotein Uncharsteried protein Uncharsteried protein Uncharsteried protein Uncharsteried protein Uncharsteried protein Uncharsteried protein Uncharsteried protein, OUT371 family Uncharsteried protein Uncharsteried protein Uncharsteried protein Uncharsteried protein Uncharsteried protein Secreted protein, WH PKD repeat domain WhOP repeat containing protein	COG01772 COG01916 COG04743 COG04743 COG01357 COG01470 SC.00914 COG03361 COG04911 COG04004 COG05652 COG03291 COG02319	pfam04407 pfam01963 pfam01963 pfam13430 pfam00805,pfam10633,pfi pfam10533,pfi pfam09844 pfam09969 pfam00801,pfi pfam00801,pfi	am00805 am13620, pfam1 am13517, pfam1	TIGR04292 TIGR04292 TIGR04261 5633 5517_pfam13517 TIGR00864	8 6 5 5 4 4 4 4 4 4 4 4 3 3 3 3 3 3	1		1 3 2 4	1 2 1 1
arC0G04364 arC0G05839 arC0G02142 arC0G02142 arC0G02184 arC0G01318 arC0G03124 arC0G0287 arC0G03124 arC0G0287 arC0G038 arC0G04549 arC0G04549 arC0G04529 arC0G0281 arC0G0281 arC0G0281 arC0G0338	S S S S S S S S S S S S S S S S S S S	Punction unknown Function un	- - - - - - - - - - - - - - - - - - -	Jonkharsteriker protein Uncharacteriker protein Personnen aburdonen protein Talk, containa GDH mooff. Amentrare associated protein with extracellular ig-like domain, a Componen Uncharacterized membrane protein. Dentappader eregasis containing protein Predicate membrane protein Uncharacterized protein Uncharacterized protein Uncharacterized protein Uncharacterized protein. Uncharacterized protein. December of the Uncharacterized protein. Uncharacterized protein. Uncharacterized protein. Uncharacterized protein. Uncharacterized protein.	COG01772 COG01916 COG01916 COG04743 COG01357 COG01470 SC.00914 COG03361 COG04911 COG04911 COG04921 COG05652 COG052319	pfam04407 pfam01963 pfam13430 pfam13430 pfam10633,pfi pfam13517,pfi pfam03844 pfam03864 pfam04892 pfam04892 pfam04892 pfam04892	m00805 m13520,pfam1 m13517,pfam1 cd00146,cd001	TIGR04292 TIGR04292 TIGR04261 5613 517_pfsm13517 TIGR00864	8 5 5 4 4 4 4 4 4 3 3 3 3 3 3 2 2			1 3 2 4 2	1 2 1 1 1 1
arC0G04364 arC0G05839 arC0G02142 arC0G02142 arC0G01284 arC0G0153 arC0G04796 arC0G03124 arC0G04703 arC0G04753 arC0G04753 arC0G04753 arC0G04752 arC0G04751 arC0G07538 arC0G04712 arC0G0712	S S S S S S S S S S S S S S S S S S S	Instition unknown Instition unknown Instition unknown Instition unknown Existion unknown Ex	- -	Jocharsterief protein Uncharsterief protein Personnen abudown protein TLB, contained TDB, tenoff Memorane subschengt protein Metarselluhar gelike domain, a Componen Uncharschereid membrane protein Pentgepriste repeats containing protein Pentgepriste repeats containing protein Uncharschereid protein	COG01772 COG01916 COG01916 COG01916 COG04743 COG01377 COG01470 SC00914 COG03361 COG04911 COG04911 COG04091 COG040274	pfam04407 pfam01963 pfam13430 pfam00805,pfi pfam10633,pfi pfam13517,pfi pfam03844 pfam09869 pfam04892 pfam00801,pfi pfam13360	100805 am13620,pfam1 m13517,pfam1 cd00146,cd001	TIGR04292 TIGR04292 TIGR04261 5013 517_pfam13517 TIGR00864	8 6 5 5 4 4 4 4 4 4 3 3 3 3 2 2 2			1 3 2 4	1
arC0G04364 arC0G0245 arC0G02142 arC0G02142 arC0G02142 arC0G0284 arC0G010133 arC0G10133 arC0G103124 arC0G01033 arC0G10443 arC0G014579 arC0G045459 arC0G045459 arC0G045458 arC0G04281 arC0G07412 arC07412 arC074	S S S S S S S S S S S S S S S S S S S	Parktion unknown Function un	- - - - - - - - - - - - - - - - - - -	Juncharatterief portein Uncharatterief portein Personne shutdown portein TLR, contained Utar Market Memorane situscourd portein with extransel Utar guide domain, a component Uncharatterief antembrane portein, DUH112 famity Ducharatterief antembrane portein Petapetischer membrane portein Ducharatterief antembrane portein Ducharatterief antembrane portein Ducharatterief antembrane portein Ducharatterief antembrane portein Ducharatterief antembrane portein Ducharatterief antembrane portein Ducharatterief portein, DUF2071 Emity Ducharatterief portein Ducharatterief portein Ducharatterief portein Ducharatterief portein Ducharatterief portein Ducharatterief portein Ducharatterief portein Ducharatterief membrane portein Ducharatterief membrane portein Ducharatterief membrane portein Ducharatterief membrane portein	COG01772 COG01916 COG01916 COG014743 COG01357 COG01357 COG01370 Sc.00914 COG03361 COG03361 COG056552 COG05291 COG0	pfam04407 pfam01963 pfam01963 pfam10483 pfam10805,pfi, pfam0805,pfi, pfam0381,pfi pfam03844 pfam03892 pfam088734 pfam08734	m00805 m13620.pfam1 m13517.pfam1 cd00146.cd001	TIGR00292 TIGR00261 5613 5517_pfum13517 TIGR00864	8 6 5 5 4 4 4 4 4 4 4 4 4 3 3 3 3 3 2 2 2 2 2 2			1 3 2 4 4 2 2	1 2 1 1 1 1 1
arC0604364 arC06024364 arC0602142 arC0602142 arC0602142 arC0602142 arC0602142 arC0602087 arC060142 arC0604595 arC060459 arC060459 arC0604595 arC060459	S S S S S S S S S S S S S S S S S S S	Anaton unknown	- - - - - - - - - - - - - - - - - - -	Joncharatteried protein Decharatteried protein Personnon shuddown protein TLB, contained/Dat Mither Memorynes shuddown protein TLB, contained/Dat Mither Joncharatteried membrane protein, DUF4112 family Decharatteried membrane protein Pertapealiter prosess containing protein Pertapealiter protein Ducharatteried protein Secreted protein, with PKD repeat Gomain WD40 repeats containing protein Ducharatteried with PKD repeation PTI domain Containing transcriptional regulator Ducharatteried with immungdobten lief fold Purplasnic protein bie fold	COG01772 COG01916 COG014743 COG01357 COG01470 SC.00914 COG01470 SC.00914 COG03361 COG0404911 COG04091 COG04021 COG04021 COG042319 COG02319 COG02319 COG04274 SC.00244 SC.00244 COG04274 SC.00244 COG04774	pfam04407 pfam01963 pfam01963 pfam12430 pfam02805,pfi am02805,pfi am10633,pfi pfam13517,pfi pfam03844 pfam03801,pfi am03860 pfam04822 pfam008734 pfam08734 pfam06224 pfam062714 pfam062714	m00805 m13620.pfam1 m13517.pfam1 cd00146.cd001	TIGR04292 TIGR02261 50502261 50513 50513 53517.pfam13517 TIGR02864	8 6 5 5 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3 2 2 2 2 2 2 1 1			1 3 2 4 4	1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
arC0604364 arC0602142 arC0602142 arC0602142 arC0602142 arC0602184 arC0601284 arC0604596 arC0604596 arC0601284 arC0604291 arC06044 arC0604291 arC0604291 arC0604291 arC0604291 arC0604291 ar	S S S S S S S S S S S S S S S S S S S	Function unknown Function	- - - - - - - - - - - - - - - - - - -	JondwarterBiele protein JondwarterBiele protein Personnos Bluddown protein TAB, contains GIIa's Mont Membrane stockown protein TAB, contains GIIa's Side domain, a componen JondwarterBiel protein with estudiality Side domain, a componen Uncharactered membrane protein Perstapeptide repeats containing protein Uncharactered protein Uncharactered protein Uncharactered protein, DUTADT family Uncharactered protein, DUTADT family Uncharactered protein Uncharactered membrane prot	COG01772 COG01916 COG04743 COG01357 COG01470 SC.00914 COG03361 COG04911 COG04911 COG042319 COG02319 COG02319 COG04274 SC.00448 COG01714 COG01714	pfam04407 pfam01963 pfam13430 pfam10305,pfi am03085,pfi am03085,pfi am0385,pfi am0385,pfi am03854 pfam03844 pfam03959 pfam04892 pfam04892 pfam04892 pfam068734 pfam06224 pfam06232 pfam063 pfam16 pfa	m00805 m13620,pfam1 m13517,pfam1 cd00146,cd001	TIGR00292 TIGR00261 D633 D633 D533 TIGR00864 TIGR00864	8 6 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3			1 3 2 4 	
arC0604364 arC0605393 arC0602142 arC0602284 arC0602884 arC0601287 arC0601287 arC0604596 arC0604596 arC0604287 arC0604291 arC0604291 arC0604291 arC0604291 arC0604291 arC0604291 arC0604291 arC060410 arC060410 arC060410 arC060410 arC0604210 arC0604210 arC0604210 arC0604210 arC0604210 arC0604210 arC0604210 arC0604210 arC0604210 arC0604200 arC0604200 arC0604200 arC0604200 arC0604200 arC0604200 arC0604200 arC0604200 arC0604200 arC0604200	S S S S S S S S S S S S S S S S S S S	Instition unknown Enstition un	- - - - - - - - - - - - - - - - - - -	Joncharsteried protein Ducharsteried protein Pieronnen shutdown protein Traß, contains GDM motif Memdane issociated protein with extraction GDM motif Uncharsteried membarae protein, DUF412 family Uncharsteried membarae protein Prediced membarae protein Uncharsteried protein Uncharsteri	COG01772 COG01916 COG04743 COG01357 COG01357 COG01357 COG01357 COG04170 SC.00914 COG04091 COG04091 COG04091 COG04274 SC.002319 COG04274 SC.002319 COG04274 SC.002319 COG04274 SC.002319 COG04274 SC.002319 COG04274 COG04274 COG04274 COG04274 COG04274 COG01714	pfam04407 pfam01963 pfam01963 pfam13430 pfam10805,pfi, pfam0683,pfi, pfam03517,pfi qfam13517,pfi qfam034892 pfam06801,pfi pfam03892 pfam080734 pfam062734 pfam062734 pfam062734 pfam062734	m00805 m13620,pfam1 m13517,pfam1 cd00146,cd001	TIGR00222 TIGR02261 0633 5517_pfam13517 TIGR00864	8 6 5 5 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3 3 3 2 2 2 2			1 3 2 4 4 2 2 2 2	1 2 2 1 1 1 1 1 1 1 2 1 1
arcC6014864 arcC605189 arcC6072142 arcC6072142 arcC6072142 arcC6072142 arcC6072142 arcC6072142 arcC6072142 arcC6072017 arcC607217 arcC607217 arcC607212 ar	S S S S S S S S S S S S S S S S S S S	Punction unknown Punction		Joncharsterikei protein Uncharsterikei protein Pieromos Bundown protein Talk, costalaci GDH motif. Membrane associated protein with estancia GDH motif. Membrane associated protein UO4132 family Uncharacterized membrane protein Predicated membrane protein Uncharacterized membrane protein Uncharacterized membrane protein Uncharacterized methrane protein Uncharacterized protein Unchara	COG01772 COG01916 COG04743 COG01745 COG01770 SC.00914 COG01357 COG01371 COG003361 COG004911 COG00365 COG03291 COG005291 COG02319 COG01714 COG01714 COG01714	pfam04407 pfam01963 pfam13430 pfam13430 pfam10805,pfa pfam10517,pfa pfam03821,pfam03824 pfam03801,pfa pfam03801,pfa pfam03824 pfam068734 pfam068734 pfam06734	m00805 m13620,pfam1 m13517,pfam1 m13517,pfam1 cd00146,cd001	TIGR00222 TIGR02261 0603 517_pfam13517 TIGR00864	8 6 6 5 5 5 4 4 4 4 4 4 4 4 4 4 3 3 3 3 3 3			1 3 2 4 4 2 2 2 2 2 1 1 1 1 2	1 2 1 1 1 1 1 1 2 2 1 1
arCG024364 arCG02539 arCG02142 arCG02142 arCG02142 arCG02142 arCG02142 arCG02142 arCG02142 arCG02142 arCG02153 arCG02124 arCG02124 arCG02124 arCG02427 arCG02427 arCG02427 arCG02421 arCG02421 arCG02421 arCG02421 arCG0211 arCG0211 arCG0211 arCG0211 arCG0211 arCG0211 arCG0211 arCG0211 arCG0211 arCG0211 arCG0211 arCG0221 arCG0221 arCG02227 arCG02227 arCG02227	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	Panction unknown Panction unknown Pancton unknown	- - - - - - - - - - - - - - - - - - -	bicharatterief protein bicharatterief protein Personne abudown protein Talk, contains GDM modif Membrane sociated protein Me detains GDM modif Membrane sociated protein Me detains GDM modif Dicharacteried membrane protein Predicate membrane protein Dicharacteried	COG01772 COG01916 COG04743 COG01357 COG01357 COG01357 COG01351 COG04274 COG04274 COG04274 COG04274 SC.00448 COG04274 COG04274 COG01714 COG01714 COG01774	pfam04407 pfam01963 pfam13430 pfam10805,pfi pfam10803,pfi pfam15317,pfi pfam0884 pfam08801,pfi pfam08802 pfam08734 pfam08224 pfam06224 pfam06224 pfam06224 pfam06234 pfam0633 pfam0633 pfam10633	m00805 m013620.pfam1 m13517.pfam1 cd00146.cd001	TIGR04292 TIGR04292 TIGR02261 5633 5517_pfsm13517 TIGR02864	8 6 6 5 5 5 4 4 4 4 4 4 3 3 3 3 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1			1 3 2 4 4 2 2 2 1 1 1 1 1 2 2	1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1
arCG024364 arCG02539 arCG02242 arCG02242 arCG02242 arCG02244 arCG02142 arCG02142 arCG02142 arCG02153 arCG02153 arCG02124 arCG02131 arCG02427 arCG0247 ar	S S S S S S S S S S S S S S S S S S S	Punction unknown Punction		Joncharatteriler protein Joncharatteriler protein Personnen abudown protein TLBR, contains GTM mooff Memorane abudown protein TLBR, contains GTM mooff Joncharacteried membrane protein, DU4112 family Ducharacteried membrane protein Ducharacteried membrane protein Characteried protein Ducharacteried protein Du	COG01772 COG01916 COG04743 COG04743 COG01357 COG01357 COG01470 SC.00914 COG04274 SC.0064420 COG04274 SC.0064274 SC.0064274 SC.0064274 COG04275 COG04274 COG04275 COG0	pfam04407 pfam01963 pfam13430 pfam10805,pfi pfam1083,pfi pfam0882,pfi pfam0882 pfam0882 pfam0882 pfam08224 pfam06224 pfam06224 pfam06224 pfam06233 pfam0633 pfam0633	ecd00146,cd001	TIGR04292 TIGR04292 TIGR04261 5013 517_pfam13517 TIGR00864	8 6 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4			1 3 2 4 4 	1 2 1 1 1 1 1 1 2 1 1 1
arCG024364 arCG02542 arCG02542 arCG02242 arCG02242 arCG02244 arCG02142 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0274 arCG0277 arCG	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	Punction unknown Punction	- -	uncharacterized protein Uncharacterized protein Determines abudiown protein TLAB, contained Dath mooff Meromone abudiown protein TLAB, contained Dath are soft and Meromane abudiown protein TLAB, contained Dath are soft and Uncharacterized membrane protein Determines abudiown protein Determines abudiown abudi	COG01772 COG01916 COG04743 COG04743 COG04743 COG01357 COG01357 COG01357 COG01357 COG01351 COG04075 COG01351 COG040911 COG040911 COG04274 SC.00243 COG04270 SC.00243 COG04770 COG04095 COG04095 COG04095	pfam04407 pfam01963 pfam13430 pfam10805,pfi pfam10805,pfi pfam0805,pfi pfam080801,pfi pfam08802 pfam08802,pfi pfam08734 pfam08734 pfam06224 pfam06224 pfam0633 pfam0633 pfam01889 pfam01889 pfam01889	m00805 m13520.pfam1 ecd00146,cd001	TIGR04292 TIGR04292 TIGR04261 Si33 Si37_fam13517 TIGR00864	8 6 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4		1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 3 2 4 4 2 2 2 1 1 1 2 2 1 1 2 2 1 1 1 2 2	1
arc0604364 arc060539 arc06052142 arc06052142 arc0605284 arc060153 arc0602484 arc0601053 arc0604284 arc0601038 arc0600429 arc0600428 arc0600429 arc060040040 arc060040040 arc0600400000000000000000000000000000000	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	Particion univono		Juncharatterief portein Uncharatterief portein Personne shurdown portein TLR, cotaine GLM end Memorizen sistochem portein TLR, cotaine GLM end Uncharatterief membrane portein, DU1112 famity Uncharatterief membrane portein Personne structure of the structure of the structure of the structure Personne structure of the structure of the structure of the structure Ducharatterief membrane portein Uncharatterief portein Uncharatter	COG01772 COG01916 COG01916 COG01743 COG01357 COG01357 COG01357 COG01357 COG01357 COG0175 COG0175 COG01714 COG01714 COG01714 COG01714 COG01775 COG01775 COG0175 COG0175 COG01775 COG01775 COG01775 COG01775 COG01775 COG0175	pfam04407 pfam01953 pfam01953 pfam13430 pfam0285,pf pfam0285,pf pfam0285,pf pfam0385,pf pfam0385,pf pfam03854 pfam03892 pfam03824 pfam038734 pfam038734 pfam038734 pfam038734 pfam038734 pfam038734 pfam038734 pfam03838 pfam03838 pfam03838 pfam03838 pfam03838 pfam03838 pfam03838 pfam13888 pfam13885	ed00146,cd001	TIGR00261	8 8 6 5 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 1 3 3 3 3 3 3 3			1 3 2 2 2 4 4 4 2 2 2 2 2 2 2 2 1 1 1 2 2 2 1 1 1 2	
arc0604364 arc060539 arc0602142 arc0602142 arc0602142 arc0602142 arc0602142 arc0602142 arc0602142 arc0602151 arc0602113 arc0602113 arc0600429 arc060040040 arc060040040 arc060040040 arc060040040 arc0	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	Instruction unitation	- -	Juncharatterief portein Uncharatterief portein Personne shutdown portein TLR, cottarell GLM et al. Memory and social portein with extra GLM et al. Uncharatterief emembrane portein Uncharatterief emembrane portein Decharatterief emethod (Unceared emembrane) Decharatterief emethod (Unceared eme	COG01772 COG01916 COG01916 COG01743 COG01357 COG01357 COG01357 COG01357 COG01357 COG01351 COG01351 COG01914 COG01351 COG01911 COG01219 COG01714 COG01274 SC0048 COG01714 COG01777 COG01774 COG01777 COG01	pfam04407 pfam01953 pfam01953 pfam13430 pfam0285,pf pfam0285,pf pfam0285,pf pfam0385,pf pfam0385,pf pfam0382 pfam03824 pfam08734 pfam06274 pfam06274 pfam06274 pfam06274 pfam06274 pfam0633 pfam0401889 pfam0401889 pfam040389 pfam040389 pfam040389 pfam04330,pf pfam13485 pfam13485 pfam13485 pfam13485	m00805 m13620,pfam1 m13517,pfam1 cd00146,cd001 cd00146,cd001 cd00146,cd001	TIGR04292 TIGR02261 TIGR02261 5613 3517_pfam13517 TIGR00864 TIGR00864 TIGR00864 TIGR02627_TIGR04247_TIGR00864, TIGR	8 8 6 6 5 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3			1 3 2 2 2 4 4 4 2 2 2 2 2 2 2 1 1 1 1 2 2 2 1 1 1 1	
arc0604364 arcC005187 arC005287 arC0	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	Anatation unknown	- -	Juncharaterized protein Decharaterized protein Personnos ibudidown protein TLB, contains GIDH mooff Memorynos ibudidown protein TLB, contains GIDH wordf Juncharaterized membrane protein Decharaterized membrane protein Perstapesitier prosess containing protein Perstapesitier protein Ducharaterized protein Duc	COG01772 COG01772 COG01916 COG04743 COG04743 COG04743 COG04743 COG04743 COG0474 COG04774 COG04775 COG04775 COG04775 COG04775 COG04775 COG04775 COG04775 COG04775 COG04775 COG04775 COG04775 COG04775 COG04775 COG0475 COG0475 COG04	pfam04407 pfam01953 pfam13430 pfam13430 pfam13430 pfam10353,pf pfam10353,pf pfam0385,pf pfam0385,pf pfam0385,pf pfam03874 pfam08734 pfam0874	m00805 m01820.pfam1m1820.pfam1 m13517.pfam1 cd00146.cd001 cd00146.cd001 cd00146.cd001 cd00146.cd001 cd14797 cd100145.cd001	TIGR04292 TIGR02201 TIGR02201 S31720fam13517 TIGR02864 TIGR02864	8 8 6 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3			1 3 3 2 2 4 4 4 	
arc0604364 arcC005437 arC005242 arcC005242 arcC005242 arcC005242 arcC005242 arcC005242 arcC005242 arcC005312 arcC005042 arcC005042 arcC005042 arcC005242 a		Particito unitonom	- -	Jondmartsreeting protein Jondmartsreeting protein Personnon shuddown protein TLB, contains GDat mordf Membrane stack-den protein TLB, contains GDat Medda Jondmartsreeting denothance protein Jondmartsreeting membrane protein Jondmartsreeting prote	COG01772 COG01772 COG01772 COG017916 COG01743 COG01377 COG01377 COG01377 COG01377 COG01376 COG01376 COG01376 COG01376 COG01376 COG01370 COG01372 COG01714 COG01372 COG01714 COG01372 COG017 COG07 COG0 COG07 COG0 COG07 COG07 COG07 COG07 COG07 COG	pfsm04407 pfam01963 pfam13430 pfam13430 pfam10835,pfi pfam10835,pfi pfam10835,pfi pfam0385,pfi pfam0385,pfi pfam03874 pfam038734 pfam0874 pfam0874 pfam	m00805 m01820.pfam1 m13517.pfam1 cd00146.cd001 cd00146.cd001 cd00146.cd001 cd00146.cd001 cd00146.cd001	TIGR04292 TIGR02261 TIGR02261 S317,0fam13517 TIGR02664 TIGR02664 TIGR02664 TIGR02664 TIGR02664 TIGR02664 TIGR02668	8 8 6 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 3 3 3 3 3			1 3 2 4 4 	
arc0604364 arc060539 arC060242 arc060242 arc060242 arc0602482 arc0602482 arc0601313 arc0604294 arc0604314 arc0604314 arc0604314 arc060429 arc0604214 arc060429 arc0604214 arc060429 arc0604214 arc0604		Pandtan unknown	- - - - - - - - - - - - - - - - - - -	Incharacterized protein Decharacterized protein in Call, contains GPat model Meromone situationen protein Tall, contains GPat model Meromane situationen protein Tall, contains GPat model Mucharacterized membrane protein Decharacterized membrane protein Decharacterized methrane protein Uncharacterized protein Ducharacterized protein Uncharacterized methrane protein Uncharacterized methrane protein Uncharacterized methrane protein Uncharacterized protein Uncharacterized protein Uncharacterized protein Uncharacterized methrane protein Uncharacterized methrane protein Uncharacterized methrane protein Uncharacterized protein	COG01772 COG01772 COG017916 COG0174 COG0174 COG0174 COG01737 COG01737 COG01737 COG0173 COG01714 COG0172 COG072 COG072 COG072 COG072 COG072 COG072 COG	pfsm04407 pfam01963 pfam01963 pfam10805,pfi pfam10805,pfi pfam08085,pfi pfam08085,pfi pfam08085,pfi pfam08085,pfi pfam0882 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam086221 pfam086221 pfam08621 pfam08633 pfam08635 pfam04889 pfam04889 pfam04889 pfam048621 pfam04889 pfam048621 pfam04889 pfam04889 pfam048621 pfam04889 pfam0	ecd00146,cd001 ecd00146,cd001 cd00146,cd001 cd00146,cd001 cd00146,cd001	TIGR02268	8 6 5 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4			1 3 3 4 4 2 2 2 2 2 2 2 2 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 2 2 4 4 4 4	
arCG024364 arCG02539 arCG02242 arCG02242 arCG02242 arCG02242 arCG02242 arCG02242 arCG02242 arCG02242 arCG02242 arCG02242 arCG02427 arCG0247 a	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	Particion univono	- - - - - - - - - - - - - - - - - - -		COG01772 COG01772 COG01772 COG01772 COG01774 COG0174 COG074	pfam04407 pfam01963 pfam13430 pfam10805,pfi pfam1063,pfi pfam08846 pfam08846 pfam08846 pfam08846 pfam08826 pfam08827 pfam08224 pfam08224 pfam06224 pfam06224 pfam06224 pfam0633 pfam0633 pfam0633 pfam040188 pfam03038,pfi pfam03889 pfam030418 pfam03048,pfi pfam03889 pfam0548,pfi pfam054	m00805 m013520_pfam13520_pfam1 m13517_pfam1 cd00146_cd001 cd00146_cd001 cd00146_cd001 cd14797 cd008025	TIGR00292 TOGR00261 TOGR00261 TOGR00261 TIGR00266 TIGR00864 TIGR00864 TIGR00266 TIGR00247,TIGR00247,TIGR00864,TIGR	8 8 6 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 4 4			1 3 2 4 4 	
arCG024364 arCG02542 arCG02542 arCG02542 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG02742 arCG0274 arC	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	Pandia unknown	- - - - - - - - - - - - - - - - - - -	Jucharaterizeria protein Jucharaterizeria protein Meromone Buddown protein TLB, cotains GDH mooff Meromone Buddown protein TLB, cotains GDH mooff Meromane Buddown protein TLB, cotains GDH mooff Jucharaterizeria denobrane protein Ducharaterizeria denobrane protein Merofischer Stender Stender Merofischer Stender Stender Merofischer Stender Stender Merofischer Stender Merofisc	COG01772 COG01916 COG01916 COG01473 COG01357 COG01473 COG01357 COG01357 COG01351 COG01351 COG01351 COG01351 COG01351 COG01210 COG0120 COG020 COG020 COG020 COG020 COG020 COG020 COG020 COG020 COG020 COG020 COG02	pfam04407 pfam01963 pfam13430 pfam10805,pfi pfam10813,pfi pfam0805,pfi pfam0805,pfi pfam08081,pfi pfam08244 pfam084844 pfam08484 pfa	m00805 m013520_pfam1 m13517_pfam1 ecd00146_cd001 cd00146_cd001 cd00146_cd001 cd00146_cd001 cd14797 cd08025	TIGR04292 70600261 70600261 5013 5017_pfam13517 TIGR00864 70600864 70600266 70600268 70600268	8 6 5 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4			1 3 2 4 4 2 2 2 2 1 1 1 2 2 1 1 1 2 2 1 1 1 2 2 1 1 1 2 2 1 1 1 2 2 1 1 2 2 2 2 1 2	1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
arCG024364 arCG02542 arCG02542 arCG02542 arCG02242 arCG02242 arCG0274 arCG0274 a	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	PactSon unknown	- - - - - - - - - - - - - - - - - - -	uncharacterized protein Uncharacterized protein in Taß, costains GTat mooff Meromone shundown protein Taß, costains GTat mooff Meromane shundown protein Taß, costains GTat mooff Uncharacterized membrane protein Ducharacterized membrane protein Ducharacterized methrase protein Ducharacterized methrase protein Ducharacterized methrase protein Ducharacterized membrane protein Ducharacterized pr	COG01772 COG01772 COG017916 COG01743 COG01747 SC.0001743 COG01357 COG01757 COG01757 COG01757 COG01751 COG01714 COG01774 SC.0001714 COG01775 COG01775 COG07775 COG0775	pfam04407 pfam01963 pfam13430 pfam10805,pfi pfam10805,pfi pfam08085,pfi pfam08085,pfi pfam080801,pfi pfam08081,pfi pfam080734 pfam08224 pfam08224 pfam06224 pfam0632 pfam0632 pfam01889 pfam01889 pfam01889 pfam01889 pfam010188 pfam01889 pfam03032 pfam03185 pfam03187 pfam01894 pfam03187 pfam01894 pfam03187 pfam01894 pfam03894,pfi pfam01894 pfam04894,pfi pfam04894,pfi pfam04894,pfi pfam04894,pfi pfam05742	m00805 m013520_fam m13517_pfam1 m13517_pfam1 m03701_pfam0 cd00146_cd001 cd100146_cd001 cd10146_cd001 cd10146_cd001 cd10146_cd001 cd14797 cd008025	TIGR04292 70600261 70600261 5013 5017_pfam13517 TIGR00864 70600864 70600266 TIGR00266 TIGR00266 TIGR00269	8 6 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4			1 3 2 2 2 4 4 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 1 1 1 2	
arc0004364 arcC005436 arcC005242 arcC005242 arcC005242 arcC005245 arcC005245 arcC005245 arcC00513 arcC00545 arcC001312 arcC001312 arcC00142 arcC00		Instruction unitshown	- - - - - - - - - - - - - - - - - - -	Jucharaterief portein Decharaterief portein in Taß, contain GDate moof Meromone shundown portein Taß, contain GDate moof Meromane shundown portein Taß, contain GDate Moof Locharaterief embersare portein Decharaterief embers. DUTOST Ismiy Decharaterief portein Subdaraterief protein Decharaterief embers. Decharaterief embers. DUTOST Ismiy Decharaterief protein Decharaterief protein Decharaterief protein Decharaterief protein Decharaterief protein Decharaterief protein Decharaterief protein Decharaterief protein Decharaterief embers upper tamiy Decharaterief protein Decharaterief embers upper tamiy Decharaterief embers upper taming Decharaterief embers protein Decharaterief embers protein Dech	COG01772 COG01916 COG01916 COG0174 COG01357 COG01357 COG01357 COG01357 COG01357 COG0175 COG0175 COG0175 COG0171 COG0171 COG0171 COG01714 COG0172 COG0172 COG0172 COG0172 COG0172 COG0172 COG0172 COG0172 COG0172 COG0172 COG0172 COG0172 COG0772 COG0772 COG0777 C	pfam04407 pfam01953 pfam01953 pfam13430 pfam10825,pf pfam13517,pf pfam0825,pf pfam0825,pf pfam0825,pf pfam0382,pf pfam08734 pfam08734 pfam08734 pfam08224 pfam08224 pfam08224 pfam08224 pfam08234 pfam08234 pfam03224 pfam03224 pfam03224 pfam03022 pfam13489 pfam13489 pfam13489 pfam13489 pfam13489 pfam03625 pfam0365 pfam0365 pfam0365 pfam0365 pfam0365 pfam0365 pfam0465 pfam0465 pfam0465 pfam0465 pfam0465 pfam0465 pfam0465	ed00146,cd001 ed00146,cd001 ed00146,cd001 ed00146,cd001 ed00146,cd001 ed00146,cd001 ed00146,cd001	TIGR02261	8 8 6 5 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 3 3 3 3			1 3 2 2 2 2 2 2 2 2 2 2 2 2 2	
arc0604364 arcC00539 arc06052142 arc06052142 arc06052142 arc06052142 arc0605115 arc06052142 arc0605115 arc0605124 arc0605421 arc0605		AnataGan unknown AnataGan	- - - - - - - - - - - - - - - - - - -	uncharacteries ¹ portein Decharacteries ¹ portein Decharacteries ¹ portein Decharacteries ¹ portein Decharacteries ¹ membrase protein Decharacteries ¹ membrase pro	COG01772 COG01916 COG01916 COG01916 COG01743 COG01357 COG01357 COG01357 COG01357 COG01357 COG01470 COG01357 COG01911 COG040911 COG040211 COG01219 COG01274 COG01274 COG01274 COG01274 COG01274 COG01715 COG01715 COG0175 COG07	pfam04407 pfam01953 pfam13430 pfam10825,pf pfam10835,pf pfam10835,pf pfam10835,pf pfam13517,pf pfam08824 pfam088734 pfam088734 pfam088734 pfam088734 pfam088734 pfam088734 pfam088734 pfam08822 pfam08838 pfam0888 pfam0888 pfam0888 pfam0888 pfam0888 pfam0888 pfam08	en04895	TIGR00261	8 8 6 5 5 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4			1 1 3 2 2 2 2 2 2 2 2 2 2 2 2 2	
nrc0604364 nrc0605436 nrc06052142 nrc06052142 nrc06052142 nrc06052142 nrc06052142 nrc06052142 nrc0605313 nrc0605512 nrc060552 nrc0605552 nrc0605552 nrc0605552 nrc0605552 nrc0605552 nrc0605552	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	Anancon unknown		uncharacterized protein Decharacterized protein Networks builds on protein TLB, contain GDH modif Memory and protein protein TLB, contain GDH modif Memory and Strand protein UH112 family Uncharacterized membrane protein Pentapetiter reposits containing protein Pentapetiter proteins containing protein Pentapetiter proteins containing protein Decharacterized protein Uncharacterized protein Uncharacterized protein Concharacterized protein Uncharacterized protein Concharacterized protein Uncharacterized p	COG01772 COG01772 COG01772 COG01773 COG0173 COG0173 COG0173 COG01357 COG0173 COG0173 COG0173 COG0173 COG0173 COG0173 COG01714 COG0172 COG072 COG0	pfam04407 pfam01953 pfam01953 pfam13430 pfam03085,pf pfam10353,pf pfam03085,pf pfam0385,pf pfam0385,pf pfam03874 pfam08734 pfam03703,pf pfam04389 pfam03889 pfam03889 pfam04305 pfam03645 pfam05685 pfam03872 pfam085685 pfam04304 pfam04304	m00805 m00805 m13620,pfam1 m13517,pfam1 cd00146,cd001 cd00146,cd001 cd00146,cd001 cd00146,cd001 cd14797 cd08025 cd00146,cd001 cd14797 cd08025	TIGR04292 TIGR02261 TIGR02261 TIGR02261 TIGR02261 TIGR02264 TIGR02664 TIGR02664 TIGR02267,TIGR0247,TIGR02664,TIGR TIGR02268 TIGR02268 TIGR02268	8 8 6 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 3 3 3 3 3			1 3 2 2 2 2 2 2 2 2 2 2 2 2 2	
arc0604364 arcC005436 arcC005242 arcC005252		Anström unknown		borcharacterized protein borcharacterized protein beromon shuddown protein TLB, contains GDH mord beroman shuddown protein TLB, contains GDH mord borcharacterized membrane protein borcharacterized membrane protein borcharacterized membrane protein borcharacterized metharase protein borcharacterized protein, DUF2071 fmilly borcharacterized protein, DUF2071 fmilly borcharacterized protein borcharacterized metharase protein borcharacterized metharase protein borcharacterized protein borcharact	CGG01772 CGG01772 CGG01772 CGG01772 CGG0174 CGG0174 CGG01377 CGG01377 CGG01377 CGG01377 CGG01377 CGG01370 CGG01371 CGG0171 CGG012319 CGG01714 CGG07174 CGG074 CGG07174 CGG074 CGG07	pfam04407 pfam01963 pfam13430 pfam10835,pfi pfam13430 pfam0805,pfi pfam0805,pfi pfam0385,pfi pfam0387,4 pfam03969 pfam04892 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam0374 pfam0374 pfam0374 pfam0374 pfam0374 pfam0374 pfa	m00805 m01820.pfam1 m13512.pfam1 cd00146.cd00146.cd001 cd00146.cd00146.cd001 cd00146.cd00146.cd0010000000000000000000000000000000000	TIGR02265	8 8 6 5 5 5 5 5 4 4 4 4 4 4 4 4 3 3 3 3 3 2 2 2 2 2 2 2			1 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2	
arc0604364 arcC005430 arC0002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002142 arcC002121	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	Anatotion unitoxion Anatoti	- - - - - - - - - - - - - - - - - - -	Indexatterilier protein Contrast Control (Control (Contro	COG01772 COG01772 COG01772 COG01772 COG01773 COG01773 COG01773 COG01773 COG01773 COG01773 COG01774 COG01775 COG0175	pfsm04407 pfam01963 pfam10395,pfi pfam10387,pfi pfam10387,pfi pfam03986,pfi pfam0387,pfi pfam0387,pfi pfam03874 pfam038734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam08734 pfam03872 pfam03838 pfam03838 pfam03303,pfi pfam03625 pfam03625 pfam03625 pfam03625 pfam03625 pfam03625 pfam03888 pfam03625 pfam03888 pfam03625 pfam03888 pfam03625 pfam03888 pfam03625 pfam03828 pfam03828 pfam03828 pfam03828 pfam03828 pfam03835 pfam03334 pfam0335 pfam0335 pfam0335 p	m00805 m01862,0fam113512,0fam1 m13512,0fam1 cd00146,cd001 cd00146,cd001 cd00146,cd001 cd14997 cd00146,cd001 cd14995	TIGR02264 TIGR02264 TIGR02264 TIGR02264 TIGR02864 TIGR02	8 8 6 5 5 5 5 5 4 4 4 4 4 4 4 4 3 3 3 3 3 2 2 2 2 2 2 2			1 3 3 2 4 4 - - - - - - - - - - - - -	
arcCoddAide arcCoddCaide arcCod	S S	Anström unknown	- - - - - - - - - - - - - - - - - - -		COG01772 COG01772 COG01772 COG01772 COG01773 COG01773 COG01773 COG01773 COG01773 COG01773 COG01774 COG07774 COG07774 COG07774 COG01775 COG0175 COG0175 COG	pfsm04407 pfam01963 pfam10387, pf pfam13430 pfam10387, pf pfam10387, pf pfam0387, pf pfam0383, pf pfam0383, pf pfam03889, pf pfam0383, pf pfam0333, pf pfam04031, pf pfam04033, pf pfam04041, pf pfam04041, pf pfam04041, pf pfam04041, pf pf pfam04041, pf pf pf pf pf pf pf pf pf pf	m00805 m01820,dfam11312,pfam1 m13512,pfam1 cd00146,cd001 cd00146,cd001 cd00146,cd001 cd14797 cd08025	TIGR02266 TIGR02260, TIGR02800, TIGR02800 TIGR02266, TIGR02800, TIGR02800 TIGR02266, TIGR02800, TIGR02800	8 6 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 4 4			1 3 3 2 4 4 - - - - - - - - - - - - -	

arCOG06646	S	Function unknown	-	Uncharacterized membrane protein, YccA/Bax inhibitor family	COG04760	pfam12811						1
arCOG14289	S	Function unknown	-	Uncharacterized protein								1
arCOG10710	S	Function unknown	-	Uncharacterized protein								2
arCOG01908	S	Function unknown	AIM24	Uncharacterized protein, AIM24 family	COG02013	pfam01987	TIC 200202				1	
arCOG05340	5	Function unknown	-	Uncharacterized membrane protein	0002855	pramosous	11GR00098				1	
arCOG05874	s	Function unknown	-	Uncharacterized protein							1	
arCOG06493	S	Function unknown	-	HEAT repeats containing protein	COG01413	pfam13646,pfam03130					1	
arCOG10568	S	Function unknown	-	Uncharacterized membrane protein							1	
arCOG02539	S	Function unknown	-	Secreted protein, with PKD repeat domain	COG03291	(2	
arCOG10301	5	Function unknown	-	Uncharacterized protein Membrane accorded corine (threeping protein kings)	COG03011	ptam04134		6	4		2	1
arCOG06897	R	General function prediction only	-	Predicted ATP-grasp enzyme	COG03919	pfam15632	TIGR01369	6	-	1		-
arCOG00347	R	General function prediction only	-	Archaeal enzyme of ATP-grasp superfamily	COG01938	pfam09754	TIGR00161	5	1			1
arCOG04444	R	General function prediction only	-	ACT domain containing protein	COG04747	pfam01842,pfa cd04908,cd049	08	5	2	1		
arCOG02293	R	General function prediction only	-	HAD superfamily hydrolase	COG00637	pfam13419 cd01427	TIGR02009	5		1		
arCOG04303	R	General function prediction only	-	Uncharacterized Rossmann fold enzyme	COG01634	pfam01973 cd07995	TIC 000034	4	1		7	1
arCOG02174 arCOG01285	R	General function prediction only	-	OB-fold domain and Zn-ribbon containing protein, possible acvI-CoA-binding	COG01033 COG01545	pfam03176,pfam03176 pfam12172.pfam01796	11GR00921	4			/	2
arCOG00517	R	General function prediction only	-	Rhodanese Homology Domain fused to Zn-dependent hydrolase of beta-lacta	COG00491	pfam00753	TIGR03413	3			1	2
arCOG00313	R	General function prediction only	Sco1	Cytochrome oxidase Cu insertion factor, SCO1/SenC/PrrC family	COG01999	pfam02630 cd02968		3	1			2
arCOG03096	R	General function prediction only	-	NAD dependent epimerase/dehydratase family enzyme	COG01090	pfam01370,pfa cd05242	TIGR01777	3	1			
arCOG02812	R	General function prediction only	-	Bacteriorhodopsin	COG05524	pfam01036	7/0003730	3				
arCOG00352	R	General function prediction only	NOgi	Cir-binding protein, Ciri/Obg family	SC 00304	piam01926 cd01897	11GR02729	3				
arCOG03439	R	General function prediction only	-	vWFa domain containing protein	50.00504			2		1	3	2
arCOG07416	R	General function prediction only	MhpC	Alpha/beta superfamily hydrolase	COG00596	pfam12695	TIGR03695	2			1	1
arCOG02889	R	General function prediction only	-	Predicted deacylase	COG03608	pfam04952 cd06252	TIGR02994	2				
arCOG01622	R	General function prediction only	MviM	Predicted dehydrogenase	COG00673	ptam01408,pfam02894		2				1
arCOG03991 arCOG01651	R	General function prediction only	-	Alpha/beta superfamily hydrolase	COG01073	prarn00801 cd00146	HGR00864, HGR04213	1	1		1	
arCOG01425	R	General function prediction only	-	RecB family nuclease with coiled-coil N-terminal domain	SC.00001			1	1		1	
arCOG01141	R	General function prediction only	-	ICC-like phosphoesterase	COG00622	pfam12850 cd00841	TIGR00040	1				
arCOG01395	R	General function prediction only	-	Lipid-A-disaccharide synthase related glycosyltransferase	COG01817	pfam04007		1				
arCOG01648	R	General function prediction only	MhpC	Alpha/beta superfamily hydrolase	COG00596	ptam12697	TIGR02427	1				
arCOG00514	R	General function prediction only	- SnolVEP	rreuicieu transgiutaminase-like protease	COG01800	platfl04473		1				
arCOG03015	R	General function prediction only	- PLANDER	Uncharacterized protein YbiT, contains NAD(P)-binding and DUF2867 domain	COG001394	pfam13460 cd05271	TIGR03466	1			1	
arCOG02625	R	General function prediction only	-	Predicted metal-dependent hydrolase	COG01451	pfam01863		1			-	
arCOG06095	R	General function prediction only	-	ABC-2 family transporter protein	COG01277	pfam12679		1				
arCOG01136	R	General function prediction only	ARC1	EMAP domain RNA-binding protein	COG00073	pfam01588 cd02800	TIGR00399	1			1	_
arCOG01213	R	General function prediction only	Cof	HAD superfamily hydrolase	COG00561	pfam08282 cd01427	TIGR01487	1			2	1
arCOG00833	R	General function prediction only	Riml	Acetyltransferase (GNAT) family	COG00456	pfam00583 cd04301	TIGR01575		1			1
arCOG02432	R	General function prediction only	-	HD superfamily phosphohydrolase	COG00407	ofam01966 cd00077			1		2	1
arCOG02680	R	General function prediction only	-	Uncharacterized archaeal Zn-finger protein	COG01326				1	1		1
arCOG09205	R	General function prediction only	-	Predicted metal-dependent hydrolase	COG01547	pfam03745				1		
arCOG09415	R	General function prediction only	-	Membrane associated metal-binding domain fused to Reeler domain		pfam02014 cd08544				1	2	2
arCOG00626	R	General function prediction only	TIYC	Hemolysins or related protein containing CBS domains	COG01253	pfam01595,pfa cd04590	TIGR03520			<u> </u>	1	
arCOG02291	R	General function prediction only	-	HAD superfamily hydrolase Bradicted O methyltraesforace YerM	COG01011	pfam13419 cd01427	TIGR02252		1	1		
arCOG00979	R	General function prediction only	- ARA1	Aldo/keto reductase related to diketogulonate reductase	COG04122 COG00656	nfam00248 cd06660	TIG801293		1	1		
arCOG01641	R	General function prediction only	-	Predicted RNA-binding protein, contains TRAM domain	COG03269						1	
arCOG00504	R	General function prediction only	-	Metal-dependent hydrolase of the beta-lactamase superfamily II	COG00491	pfam00753	TIGR03413					1
arCOG00348	R	General function prediction only	-	Archaeal enzyme of ATP-grasp superfamily	COG02047	pfam09754	TIGR00162				1	
arCOG04055	R	General function prediction only	-	SHS2 domain protein implicated in nucleic acid metabolism	COG01371	pfam01951					1	1
arCOG02579	R	General function prediction only	-	Predicted metal-dependent hydrolase (urease supertamily) Ee-S-cluster containining protein	COG01831	ptam01026					2	1
arCOG04212	R	General function prediction only	-	Predicted DNA-binding protein with PD1-like DNA-binding motif	COG01661	pfam03479 cd11378			1			
arCOG06747	R	General function prediction only	-	SIR2 superfamily protein	COG00846	pfam02146,pfa cd01406			1			
arCOG06769	R	General function prediction only	-	GTPase SAR1 family domain fused to Leucine-rich repeats domain	COG01100	pfam13855,pfa cd00116,cd099	TIGR00231		1			
arCOG00499	R	General function prediction only		Metal-dependent hydrolase of the beta-lactamase superfamily	COG01235	pfam12706	TIGR02651		1		1	
arCOG01225	R	General function prediction only	-	GTPase SAR1 or related small G protein Predicted surface protease of transglutaminase family	COG01100	ptam03029 cd02027,cd008	80		1			1
arCOG04624	R	General function prediction only	-	Multimeric flavodoxin WrbA	COG00431	-			-		1	1
arCOG02155	R	General function prediction only	-	Protein implicated in RNA metabolism, contains PRC-barrel domain	COG01873	pfam05239			1			
arCOG01848	R	General function prediction only	WbbJ	Acetyltransferase (isoleucine patch superfamily)	COG00110	pfam00132,pfa cd03358	TIGR03570					1
arCOG06534	R	General function prediction only	-	Cohesin domain containing secreted protein		pfam00963 cd08547				1		
arc0G04179	R	General function prediction only	PDCD5	UNA-binding IFAK19-related protein, PDSD5 family	COG02118	ptam01984	TIG801202 TIG801202			1		1
arCOG00001 arCOG02289	R	General function prediction only	-	Uncharacterized protein, DUF169 family	COG02043	pramo0571,pra c004633,00046	110h01302,110R01302			1		
arCOG03994	R	General function prediction only	-	PKD domain containing protein	COG03291	pfam05048,pfa cd14251,cd001	TIGR04247,TIGR04275,TIGR04275,TIGR	04275,TIGR042	275,TIGR042	1		
arCOG03038	R	General function prediction only	-	TPR repeats containing protein	COG00457					1	1	1
arCOG05195	R	General function prediction only	-	TPR repeats containing protein	COG00457	pfam13414 cd00189	TIGR02521				2	
arCOG00353	R	General function prediction only	HtiX	G IP-binding protein protease modulator	COG02262	ptam13167,pta cd01878	ныки3156				1	1
arCOG01043 arCOG04589	R	General function prediction only	CinA	Uncharacterized protein (competence- and mitomycin-induced)	COG01931 COG01546	prartitu1877	TIGR00199				1	1
arCOG05137	R	General function prediction only		TPR repeats containing protein	COG00457	pfam13414,pfa cd00189,cd001	TIGR02917				1	
arCOG01084	R	General function prediction only	MazG	Predicted pyrophosphatase	COG01694	pfam03819 cd11535						1
arCOG00497	R	General function prediction only	-	Zn-dependent hydrolase of the beta-lactamase fold	COG02220	pfam13483						1
arCOG00606	R	General function prediction only	-	LBS domain	COG01784	ptam00478 cd04623	ныки1302					1
arCOG03032	R	General function prediction only	-	TPR reneats containing protein	COG001/84	prartitu1970	TIG802917				1	T
arCOG10597	R	General function prediction only	-	Hemocyanin family protein, binds copper ions	-5000+37	pfam00264					1	
arCOG00498	R	General function prediction only	-	Metal-dependent hydrolase of the beta-lactamase superfamily II	COG00491	pfam00753	TIGR03413				1	1
arCOG00500	R	General function prediction only	ElaC	Metal-dependent hydrolase of the beta-lactamase superfamily	COG01234	pfam12706	TIGR02651				1	
arCOG02642	R	General function prediction only	PerM	Predicted PurR-regulated permease PerM	COG00628	ptam01594	TIGR02872				2	1
arCOG01850	R	General function prediction only	- Whbi	Preukteu kinase related to galactokinase and mevalonate kinase Acetyltransferase (isoleucine natch superfamily)	COG00110	nfam00132 cd04647	TIGR03532				5	1
arCOG02303	R	General function prediction only	SurE	Predicted acid phosphatase	COG00496	pfam01975,pfam14423	TIGR00087					1
arCOG02839	R	General function prediction only	-	Uncharacterized protein related to deoxyribodipyrimidine photolyase	COG03046	pfam04244,pfam03441						1
arCOG02986	R	General function prediction only	BioY	Uncharacterized protein	COG01268	pfam02632						1
arCOG03047	R	General function prediction only	-	TPR repeats containing protein	COG00457	-60004 -40000						1
arCOG05000	ri p	General function prediction only	YtfP	Uncharacterized protein ttir, gamma-glutamylcyclotransferase (GGCT)/AIG2	COG02105	nfam06094 cd06661						1
arCOG08119	R	General function prediction only	PhoX	Secreted phosphatase, PhoX family	COG03211	pfam05787						1
arCOG03561	R	General function prediction only	-	Secreted protein with beta-propeller repeat domain	COG03391	pfam01436,pfa cd14955						2
arCOG07781	R	General function prediction only	-	Cell surface protein								2
arCOG02562	R	General function prediction only	-	Beta-propeller repeat containing protein	COG03391	-604627-6					1	3
arCOG07790	R	General function prediction only	-	/www-super/amily ATPase D-eliucuronyl C5-enimerase C-terminal domain related protein	CUGU1672	prarn01637,ptam01978					1	
arCOG02560	R	General function prediction only	-	Secreted protein with beta-propeller repeat domain	COG03391	cd05819	TIGR03866				1	
arCOG06256	R	General function prediction only	-	Predicted esterase	COG00400	pfam02230					2	
240000082	FF		PucG	Serine-minutate aminotransferase/archaeal aspartate aminotransferase	0000075	nfam00266 cd06451	TIG803301	1			1	1

Supplementary Table S7. Genes found in MG-III Metabolic Pathways

		Epipelagic MG-III	Bathy1	Bathy2
Glycolysis				
	hexokinase (glk)			
	phosphoglucoisomerase(pgi)	Х	Х	Х
	phosphofructokinase (pfkA)	Х		
	aldolase (fba/dhnA)	Х	Х	Х
	triosephosphate isomerase(tpi)	Х	Х	
	glyceraldehyde 3-phosphate dehydrogenase (gapA)	Х	X	X
	3-phosphoglycerate kinase (pgk)	X	X	X
	phosphoglyceromutase (pgm/yibO)	X	X	
	enolase(eno)	X	X	
	pyruvate killase (pykA)			
Gluconeogenesis				1
andoncogenesis	phosphoenolpyruvate synthase (ppsA)	х	х	x
	enolase (eno)	х	х	
	phosphoglyceromutase (pgm)	х		
	3-phosphoglycerate kinase (pgk)	х	Х	Х
	glyceraldehyde 3-phosphate dehydrogenase (gapA)	х	Х	Х
	triosephosphate isomerase(tpi)	Х	Х	
	aldolase (fba/dhnA)	Х	Х	Х
	fructose bisphosphatase (suhB)	Х	Х	Х
	phosphoglucoisomerase (pgi)	Х	Х	Х
		ļ,		1
Pentose phosphate shunt and pentos	se biosynthesis			
	glucose-6-phosphate dehydrogenase (zwf)			
	6-phosphogluconate dehydrogenase (gnd)			
	transketolase (tktA)	X	Χ	Χ
	u ansanuOldSE (ldIA) nentose-5-nhosnhate-3-animerase (uhfO)	×	×	
	ribose 5-phosphate isomerase (miA)	x	x	x
	deoxyribose-phosphate aldolase (deoC)			
		1000	a series of	<u>,</u>
Entner–Doudoroff pathway				
	glucose-6-phosphate dehydrogenase (zwf)			
	6-phosphogluconate dehydratase (edd)			
	2-keto-3-deoxy-6-phosphogluconate aldolase (eda)			
TCA cycle				
	citrate synthase (gltA)	Х	X	
	aconitase(acnA)	X	Х	X
	Isocitrate denydrogenase (ico)	X		
	a-recognitiance denythogenase (such, such)	×	 X	
	fumarate reductase (frdA_frdB)	×	~	
	fumarase (fumA)	x	х	x
	malate dehydrogenase (mdh)	X	X	
	, , , , , , , , , , , , , , , , , , , ,	<u> </u>		,
Purine biosynthesis				
	phosphosphoribosylpyrophosphate synthase (prsA)	х	Х	Х
	amidophosphoribosyltransferase (purF)	Х	Х	Х
	GAR synthase (purD)	Х	Х	
	GAR transformylase(purN/purT)	Х	Х	Х
	FGAM synthase (purL)	Х	X	Х
	AIR synthase (purM)	Х	Х	X
	NCAIR synthase (purK)			
	NCAIR mutase (purE)	X	~	X
	SAICAK Synthase (purc)	×	X V	 v
	AUCAR transformulace (purB)	×	×	×
	IMP cyclobydrolace (purH1)	x	×	x
	adenvlosuccinate synthase (purta)	x	X	
	IMP dehvdrogenase (guaB)	x		х
	GMP synthase (guad)	X	х	
		'		·
Pyrimidine biosynthesis				
	carbamoylphosphate synthase(carA, carB)	Х		Х
	aspartate carbamoyltransferase (pyrB)	Х	Х	
	dihydroorotase (pyrC/ygeZ)			
	dihydroorotate dehydrogenase(pyrD)	X	X	X
	orotate phosphoribosyl-transferase (pyrE)	X	Х	X
	orotidine-5 -phosphate decarboxylase (pyrF)	X		
	UMP kinase (pyrH)	I		

	NDP kinase (ndk)	х	х	х
	CTP synthase (pyrG)	Х	Х	
istidine hiosynthesis				
istume biosynthesis	phosphosphoribosylpyrophosphate synthase (prsA)	х	x	х
	ATP-phosphoribosyltransferase (hisG)			
	phosphoribosyl-ATP pyrophosphatase (hisl2)			
	phosphoribosyl-AMP cyclohydrolase(hisl1)			
	58-ProFAR isomerase (hisA)			
	imidazoleglycerol phosphate synthase (hisH, hisE)			
	imidazolegiycerol phosphate dehydratase (hish) hish)			
	histidinoll phosphate aminotransferase (hisC)	х	x	
	histidinol phosphate and phosphatese (hisb)			
	histidinol dehydrogenase (hisD)			
ranched chain amino acids biosynth	esis			
	threonine deaminase (ilvA)	Х		Х
	acetohydroxyacid synthase (ilvB, ilvN)		Х	
	acetohydroxyacid isomeroreductase (ilvC)			
	dihydroxyacid dehydratase (ilvD)			
	2-isopropylmalate synthese (leuA)			
	isopropulmalate isoperase (leuc)			
	isopropylmalate isomerase (ieuC, ieuD)			
	3-isopropyi-maiate dehydrogenase (leuB)			
	giutamate transaminase (ilvE)	X	X	
romatic amino acids biosynthesis				
a smalle annio acius biosynthesis	3-denxyhentulosonate 7-nhosnhate synthase (aroc/kdcA)			
	2 dobudroguinate synthase (arou)			
	S-uenyuroquinate synthase (arOB)			
	s-denydroquinate denydratase (aroD)			
	snikimate denydrogenase (aroE)			
	shikimate kinase (aroK)			
	5-enolpyruvoylshikimate 3-phosphate synthase (aroA)			
	chorismate synthase (aroC)			
	chorismate mutase (pheA1)			
	prephenate dehydratase (pheA2)			
	prephenate dehydrogenase (tyrA2)			
	tyrosine aminotransferase (tyrB)			
	antranilate synthase (trpD1, trpE)			
	antranilate phosphoribosyl-transferase (trpD2)			
	nhosphorihosylantranilate isomerase (trpC2)	and the second		
	indole-glycerol phosphate synthese (trpC1)			
	tryptophan synthese (trpA trpB)	x		
		X		
hreonine biosynthesis				
	aspartokinase (thrA1)			
	aspartate semialdehyde dehydrogenase (asd)			
	homoserine debydrogenase (thrA2)			
	homoserine kinase (thrB)			
	threeping synthese (thrC)			
Nethionine biosynthesis				
	aspartokinase (metL1)			
	aspartate semialdehyde dehydrogenase (asd)			
	homoserine dehvdrogenase (met 2)			
	homoserine transsuccinvlase (metA)			
	cystathioning g_synthese (motP)	Y	¥	Y
	b system of mile g-synthase (meter)			
	methionine synthese (metE/motH)			
			I	
rginine biosynthesis				
	acetylglutamate synthase (argA2)			
	acetylglutamate synthase (argA2)	 X		
	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductare (argC)	 X		
	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylgrithing amigntransformer (argC)	 X 		
	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD)	 X X		
. 9	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD) acetylornithinase (argE)	 X X X	 	
	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD) acetylornithinase (argE) ornithine carbamoyltransferase (argF)	 X X X	 X	
	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD) acetylornithinase (argE) ornithine carbamoyltransferase (argF) argininosuccinate synthase (argG)	 X X X X	 X	
	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD) acetylornithinase (argE) ornithine carbamoyltransferase (argF) argininosuccinate synthase (argG) argininosuccinate lyase (argH)	 X X X X 	 X 	
	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD) acetylornithinase (argE) ornithine carbamoyltransferase (argF) argininosuccinate synthase (argG) argininosuccinate lyase (argH)	 X X X X X 	 X 	
IAD biosynthesis	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD) acetylornithinase (argE) ornithine carbamoyltransferase (argF) argininosuccinate synthase (argG) argininosuccinate lyase (argH)	 X X X X X 	 X 	
NAD biosynthesis	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD) acetylornithine carbamoyltransferase (argF) argininosuccinate synthase (argG) argininosuccinate lyase (argH) argininosuccinate lyase (argH)	 X X X X X 	 X 	
IAD biosynthesis	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD) acetylornithine carbamoyltransferase (argF) argininosuccinate synthase (argG) argininosuccinate lyase (argH) aspartate oxidase (nadB) quinolinate synthase (nadA)	 X X X X X 	 X	
VAD biosynthesis	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD) acetylornithine carbamoyltransferase (argF) ornithine carbamoyltransferase (argG) argininosuccinate synthase (argG) argininosuccinate lyase (argH) aspartate oxidase (nadB) quinolinate synthase (nadA)	 X X X X X X X X	 X X	
IAD biosynthesis	acetylglutamate synthase (argA2) acetylglutamate kinase (argB) acetylglutamate phosphate reductase (argC) acetylornithine aminotransferase (argD) acetylornithine carbamoyltransferase (argF) ornithine carbamoyltransferase (argF) argininosuccinate synthase (argG) argininosuccinate lyase (argH) aspartate oxidase (nadB) quinolinate synthase (nadA) quinolinate phosphoribosyltransferase (nadC) nicotinic acid mononucleotide adenylyltransferase (nadD)	 X X X X X X X X 	 X X X 	
Riboflavin biosynthesis				
--	----------------	-------------	---	
GTP cyclohydrolase II (ribA)	Х	Х		
pyrimidine deaminase (ribD1)				
pyrimidine reductase (ribD2)	Х	Х	Х	
3,4-dihydroxybutanone-4-phosphate synthase (ribB)	Х	Х		
6,7-dimethyl-8-ribityllumazine synthase (ribE)	Х	Х		
riboflavin synthase (ribC)	Х	Х		
Siroheme biosynthesis				
Glutamyl-tRNA reductase (hemA)				
glutamate 1-semialdehyde aminotransferase (hemL)				
probilinogen III synthase (hemB)				
hydroxymethylbilane synthase (hemC)				
uroporphyrinogen III synthase (hemD)				
uroporphirinogen methyltransferase (cysG2)				
dimethyluroporphirinogen III dehydrogenase (cysG1)				
Cobalamin biosynthesis				
uroporphyrinogen III methylase (cysG2)				
precorrin-2 methylase (cbiL)				
precorrin-3B methylase (cbiH)				
precorrin-4 methylase (cbiF)				
precorrin-6A reductase (cbiJ)				
precorrin 6B methylase (cbiE)				
precorrin 6B decarboxylase (cbiT)				
precorrin-8x isomerase (cbiC)				
cobyrinic acid a,c-diamide synthase (cbiA)				
cobalt insertion protein (cobN)				
cob(I)alamin adenosyltransferase (cobA)				
cobyric acid synthase (cbiP)				
cobyric acid aminotransferase (cobD)				
, cobinamide synthase (cbiB)				
nicotinate-nucleotide:dimethylbenzimidazole phosphoribosyltransferase (cobT)				
cobalamin synthase (cobS)				
	-		•	
Biotin biosynthesis	a second and	a series of		
pimeloyl-CoA synthetase (bioW)	-			
7-keto-8-aminopelargonate synthetase (bioF)	Х	Х		
7,8-diaminopelargonate aminotransferase (bioA)				
dethiobiotin synthetase (bioD)	and the second			
biotin synthetase (bioB)				
biotin-[acetyl-CoA carboxylase] holoenzyme synthetase (birA)	Х	Х	х	

Cells marked with a "X" means that the protein was found.





AGRADECIMIENTOS



Quiero expresar mi más sentido agradecimiento a todas aquellas personas que me han apoyado durante el desarrollo de esta Tesis doctoral. Han sido 5 años de duro trabajo y sin su ayuda no hubiera sido posible.

En primer lugar, quiero agradecer a mi familia, tanto por el apoyo económico como emocional. En especial a mis padres José y Manoli, que me apoyaron durante mis estudios de Química, Bioquímica y el posterior Máster en Bioingeniería, así como mi estancia de 6 meses en el frio Estocolmo. Sin vuestro cariño y apoyo hoy día no sería quien soy. Mil gracias.

En segundo lugar, quiero agradecer a mis dos directores de tesis, Mario y Paco, por haberme enseñado tanto. He aprendido y he disfrutado aprendiendo, y eso lo he conseguido porque habéis sido unos magníficos directores y habéis resuelto tantas dudas tuviera. Sobre todo, mil gracias Paco por concederme la oportunidad de trabajar en tu grupo y de intentar sacar de mí el mejor científico posible.

También quiero agradecer a Ana Belén y a Rohit, que aunque coincidimos poco tiempo, me ayudaron enormemente a integrarme en el laboratorio. En especial a Ana Belén, ya que un tercio de esta tesis también es tuya. Gracias. Por supuesto, no me puedo olvidar de aquellas personas con las que he trabajado codo con codo. Muchas gracias Pedro J, mi más fiel compañero de piso y de trabajo, por esas largas charlas sobre ciencia y futbol. Gracias Carol, Nieves, Rafa, Ricardo, Felipe, Asier, Cesar, Raquel, Eva, Aurelia, Juanjo y Elena por vuestro apoyo y por las risas. Por supuesto, quiero agradecer a Antonio, por hacer amena la hora de la comida, y a Encarni, por el muy buen tiempo que he pasado con ella en el coche.

No me puedo olvidar de aquellas personas que han conseguido que sea feliz y que desconectara del trabajo. A Salva y María, por apoyarme desde que tenía 15 años y estar siempre ahí, aún con mi tosco comportamiento. A Sergio, por ser mi amigo dentro y fuera de la universidad. A Carlos, Edu, Carol, Tamara, Susana y María José, por el gran apoyo durante la licenciatura y después de ésta y por esos partidos de futbol amateur. A Ángel, Berna, Fran, Abraham, Violeta, Rocío, Gemma, Mariasun y Bea, porque no he conocido mejores amigos que vosotros. A todos vosotros, ¡muchas gracias!

Por último, pero no menos importante, quiero agradecer a Sara, por todo el apoyo y cariño que me ha dado. Mil gracias por comprenderme, apoyarme y quererme. Incluso en esos momentos en los que me atasco y me encierro emocionalmente. Gracias por entender todos esos sábados, domingos y días de vacaciones que esta Tesis te ha robado. Sin ti, no sé si hubiera sido capaz de salir ileso.

