



ELSEVIER



Flexible genomic islands as drivers of genome evolution

Francisco Rodriguez-Valera, Ana-Belen Martin-Cuadrado and
Mario López-Pérez

Natural prokaryotic populations are composed of multiple clonal lineages that are different in their core genomes in a range that varies typically between 95 and 100% nucleotide identity. Each clonal lineage also carries a complement of not shared flexible genes that can be very large. The compounded flexible genome provides polyclonal populations with enormous gene diversity that can be used to efficiently exploit resources. This has fundamental repercussions for interpreting individual bacterial genomes. They are better understood as parts rather than the whole. Multiple genomes are required to understand how the population interacts with its biotic and abiotic environment.

Address

Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, 03550 Alicante, Spain

Corresponding author: Rodriguez-Valera, Francisco (frvalera@umh.es)

Current Opinion in Microbiology 2016, **31**:154–160

This review comes from a themed issue on **Environmental microbiology**

Edited by **Monica Vasquez** and **Steven Hallam**

<http://dx.doi.org/10.1016/j.mib.2016.03.014>

1369-5274/© 2016 Elsevier Ltd. All rights reserved.

Introduction

In the wild, bacterial and archaeal populations reproduce clonally by simple fragmentation mechanisms (typically bipartition). However, non-reproductive sex is widespread and less confined than in eukaryotes [1]. Not limited by similarity requirements like meiosis prokaryotic cells can incorporate exotic DNA from distant microbes (including eukaryotes). The importance of prokaryotic sex, also known as horizontal gene transfer (HGT), in prokaryotic evolution has been proven so extensively [2], and in such a broad range of microbes, that the possibility of representing their evolution in a tree like form possess serious challenges [3]. However, seen from the other end of the time range, at short evolutionary time spans, the mechanisms and the dynamics of gene acquisition and loss appear more amenable to systematic description. Specifically, genes located in

clusters or genomic islands can be more easily tracked as they vary among closely related genomes. Such genomic islands are essential in providing diversity at the population level but also in maintaining this diversity. In this review we will analyze recent evidence about the stability and dynamics of prokaryotic genomes in the short to medium term. Comparative genomics, metagenomics and single cell genomics have been instrumental in generating this evolving new perspective.

Prokaryotic species definition, a view from metagenomics

The definition of species, the basic units of classification, is a major conundrum of Biology at large. In animal and plants sexual reproduction is a great help since it provides populations with a gene flow that is responsible of coherence within the species and discrimination from other neighboring taxa. This is the biological species concept in which species are considered as evolutionary units [4]. This concept is already problematic for metazoan or plants that reproduce clonally or interbreed with variable frequency, but is largely useless when transferred to clonally reproducing prokaryotes in which exchange of genes has been proven to occur, at variable frequencies, but over vast phylogenetic distances. In prokaryotes even the existence of discontinuities in the diversity space that would justify the existence of discrete units of classification, that is, species, is controversial [5]. In classical definitions, one clone, the type strain, is considered to represent the species, but the existence of many other clones that are closely related is acknowledged. During the pre-genomic era the ‘golden rule’ of 70% DNA–DNA hybridization was considered a formal threshold for two strains to be included in a single species. This value of 70% could be already considered indicative of a singularity of prokaryotes since the values for animal or plant species are much higher [6].

The advent of genomics brought the pan-genome paradigm that refers to all the genes present in a single prokaryotic species. The pan-genome has two components with different evolutionary trajectories and specialization [7]. The core genome comprising genes that are found in all or most strains of the species and the flexible genome (sometimes called adaptive or accessory) made up of genes found only in some strains but with no orthologs in others. This two-tiered nature is radically different from the eukaryotic species genomes in which nearly all genes have orthologs, not only within the species but even in other distant orders or classes [8].

A remarkable feature of the prokaryotic flexible genome is its size. For any given strain only about 50% of the genome is core. Given that each strain contributes hundreds of novel (different) flexible genes, the size of the compounded flexible pool is enormous. In *Escherichia coli* with barely 100 genomes analyzed over 45 000 gene families have been described [9], this is already twice the gene diversity of humans. Even more remarkable is recent evidences that indicate that these large genetic pools coexist within microbial populations (groups of related individuals sharing the same habitat) [10,11].

The most convincing reports of the existence of well-defined species of prokaryotes come from metagenomics. A metagenome has genomic fragments from all the cells present in a microbial habitat (sample) and can be compared to each other or to a reference genome. Genome recruitment is the bioinformatic comparison of a genome (typically from a pure culture) with a metagenome from a sample in which this pure culture (or single cell genome) is expected to be present (e.g. the same sample used for isolation of the culture or the cell). Genome recruitment has been used since the beginning of metagenomics to assess microbial abundance patterns, but it can also be used to analyze the population genomic structure. Individual metagenomic reads form a cloud that is typically located from 95 to 100% identity beyond which the discontinuity or boundary of the species is clearly proven by the empty gap showing no metagenomic fragments [12]. This pattern was taken as a reflection of natural discontinuities in the sequence space, that is, natural species. The threshold of 95% identity coincides with the values established to define species by analysis of genomes of pure culture isolates [13]. Figure 1 displays two recruitment plots generated by mapping metagenomic sequence information from two different environments onto separate reference genomes to illustrate the commonality of recruitment patterns for two very different aquatic habitats and microbes. *Haloquadratum walsbyi* is an archaeon that inhabits NaCl saturated brines while *Alteromonas mediterranea* is a marine bacterium. In spite of the enormous difference between the two reference genomes/habitats, the plots are remarkably similar. This pattern is recurrent in most aquatic (relatively homogeneous) habitats [14], and might be extrapolated to other habitats as well [15].

The local pangenome

These recruitment patterns are actually reflections of the local pangenome of the population. The cloud of fragments recruited over 95% reveals the diversity of core genes of the cells in the population (and the error rate of the sequencing method used), that is, the gene pools shared by most strains within a species that is well conserved in synteny but varies at the sequence level within this range of 95–100%. It is important to emphasize that 5% divergence over a 3 Mb long genome is

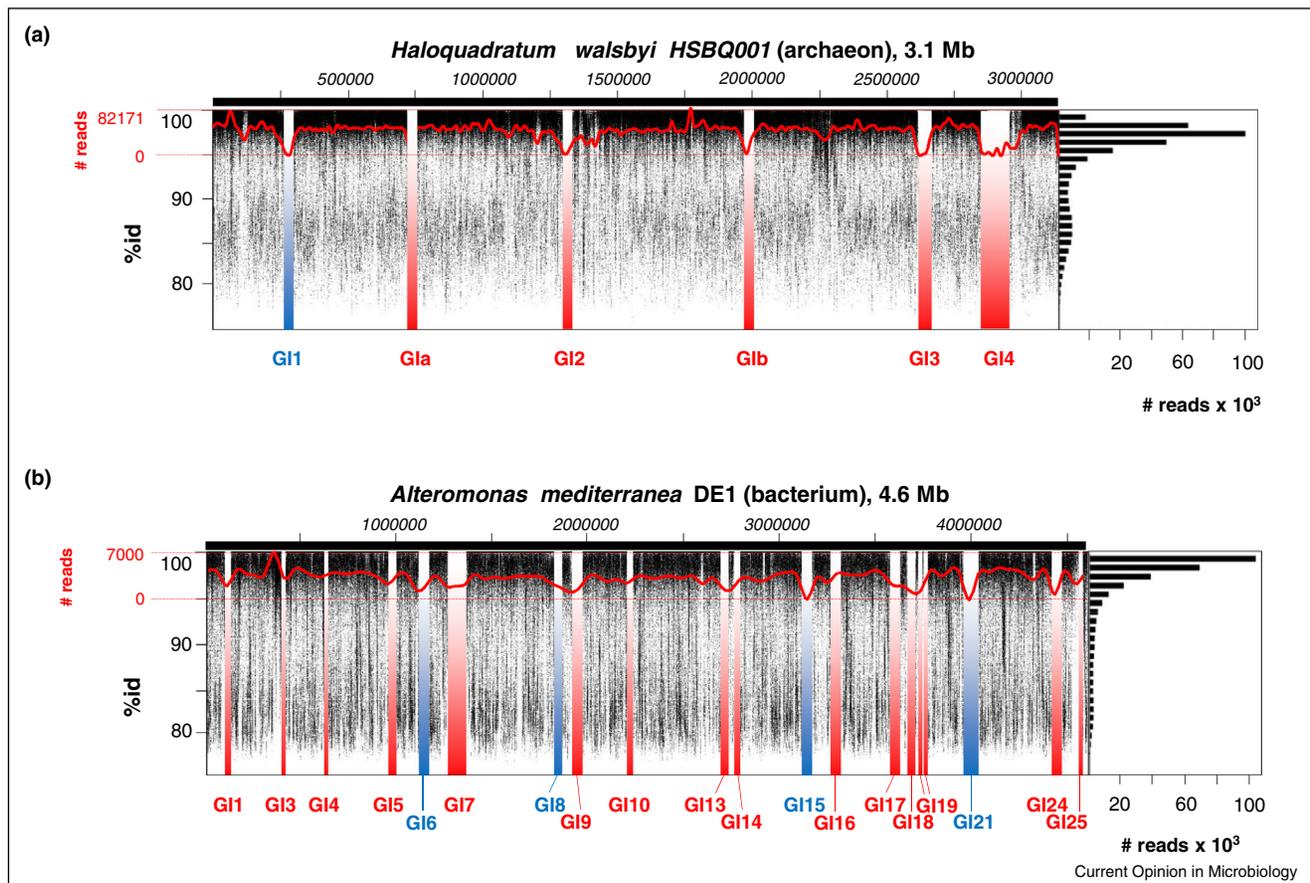
equivalent to 150 000 single nucleotide polymorphisms, that is, the core genomes of the population is already significantly variable at the sequence level. The other significant fact that is revealed in Figure 1 is that the reference genome does not recruit evenly and there are gaps or under-recruiting regions. The name metagenomic island was proposed [16**] to describe these regions that under-recruited from a metagenome in which a large part of the genome recruited over 95% nucleotide identity. Their presence reflects the flexible component of the reference genome. Although flexible genes are often interspersed throughout the core genome, most are concentrated in flexible genomic islands (fGIs). Given that each of the clonal lineages contains different genes at these locations these areas of the reference genome appear as under-recruiting and are largely the explanation for metagenomic islands. There are two different kinds of fGIs with different kinds of biological roles and modes of variation: replacement and additive (Figure 2).

Replacement fGIs are clusters of genes that although present in all clonal lineages at equivalent (syntenic) positions show no sequence homology to each other. However, the biological function for which they code is equivalent. They are often important for basic functions of the cell (if not essential for survival in nature) and all clonal lineages carry a version. The most paradigmatic example would be the gene cluster coding for the O-chain polysaccharide of the lipopolysaccharide of Gram negative bacteria. Additive fGIs on the other hand are a product of illegitimate recombination and it is conceivable that they are in a continuous state of change. They are located in a number of hotspots along the genome, normally associated with a tRNA gene. The tRNA acts as a target for illegitimate recombination in which a variable number of gene cassettes coding for a broad diversity of biological functions are inserted. Integrons are typical examples of additive fGIs with a specific mechanism of variation associated to the integration of gene cassettes [17], but there are many other genome evolution hotspots with different integration mechanisms.

Replacement genomic islands coding for polysaccharides

Most replacement fGIs contain gene clusters involved in the synthesis of polysaccharides. In bacteria polysaccharides synthesized at the external surfaces of the cell have a number of common elements [18]. They require genes involved in the synthesis of sugars that sometimes are enzymatically modified by acylation, methylation or otherwise. Other genes are involved in the specific bonding to UTP to form the UDP-sugar intermediate that carries the energy required for glycosidic linkage. Finally, some specific genes produce specific transporters such as the flippase Wzx that carry the oligosaccharides to their extracellular location [19]. Replacement genomic islands often code for combinations of genes that lead to the

Figure 1

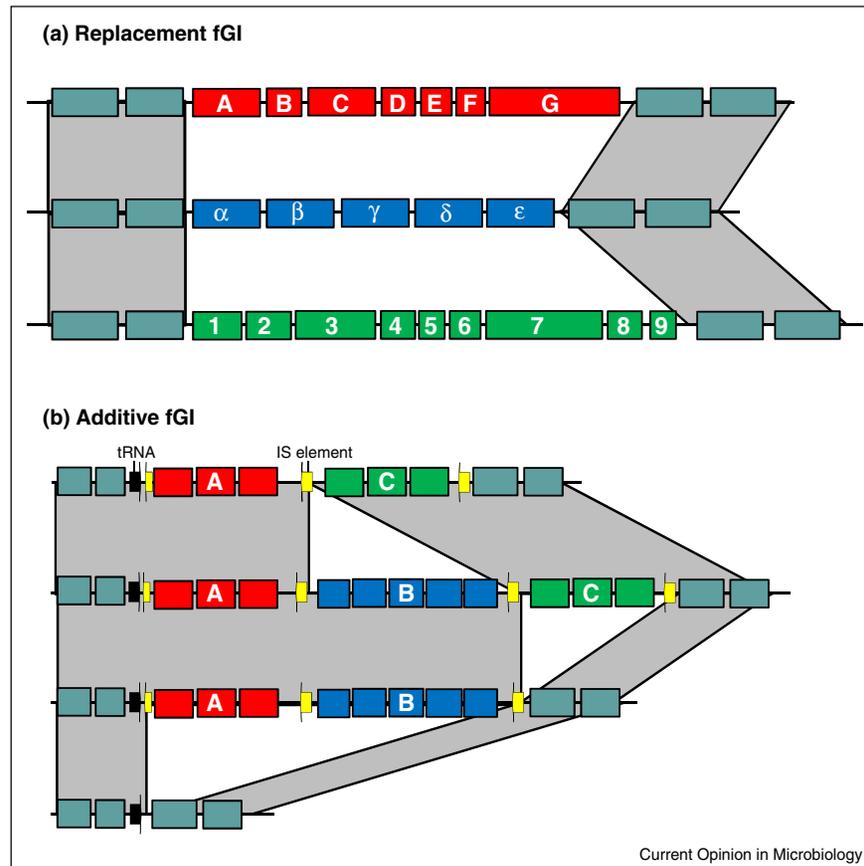


(a) Recruitment of *Haloquadratum walsbyi* HSBQ001 strain from a saltern brine 454 metagenome in which this microbe is predominant (ca. 60% of all the reads). The reads aligned at >75% similarity and over >95% of their length are plotted along the genome at the position with maximum similarity and with their %identity in the y axis. The red line along the recruitment plot shows the number of reads that recruit over 95% of identity (in 95% of their length) every 25 kb. Under-recruiting genomic islands are highlighted in red or blue (replacement islands). Right panel indicates the total number of sequences at each percentage of identity. **(b)** Recruitment of the marine bacterium *Alteromonas mediterranea* DE1 strain from a metagenome from a deep Mediterranean sample, the Matapan-Vavilov basin, 454 metagenome. Representation of recruitment values as in panel a.

production of a specific polysaccharide or the glycosylation of an external feature such as the flagellum. The O-chain of the lipopolysaccharide is a typical example of a character coded in a replacement fGI (Figure 3). All Gram negative bacteria have them. It is a large cluster of genes rich in glycosyltransferases and sugar synthesis enzymes and there are multiple versions in any given species. Each different gene cluster produces a different polysaccharide. They were traditionally described as somatic antigens due to their role in serotyping pathogens, but the diversity is similar in free-living bacteria that never come into contact with an immune system. It would be more appropriate to call them glycotypes. In Gram negative bacteria glycotypes are due to three major components each located in replacement fGIs: the O-chain (O-antigen) of the lipopolysaccharide (LPS), the exopolysaccharide or capsular polysaccharide and the flagellar glycosylation (the two latter are optional but if present in one strain of a

species, most strains within the species carry one version of each). Further, in addition to the O-chain gene cluster, all known exposed structural motifs in other groups of bacteria or archaea reflect similar genomic patterns of variation. For example, capsular or slime layer polysaccharides [20], the teichoic acids of Gram-positives [15], or the S-layer glycoproteins of archaea [21,22] all seem to be located in replacement fGIs. The glycotype of a specific clone will be paramount in determining its sensitivity to their omnipresent viral predators, that is, each glycotype will have a set of viral lineages preying selectively on them. Thus the diversity of glycotypes is mirrored by the diversity of viral predators that act to control their numbers. The more frequent a clone becomes the more it becomes a target for viral predation, leading to equilibrium among carriers of different replacement fGIs. This model is known as kill-the-winner or negative frequency-dependent selection [23,24].

Figure 2



Schematic representation of the two different kinds of flexible genomic islands (fGIs). **(a)** Replacement fGI and **(b)** additive fGI. Capital, small letters or numbers indicate the non-homologous genes present in the replacement fGIs and the syntenic and homologous regions are shaded in grey. The cassettes of additive fGIs are identified by different capital letters.

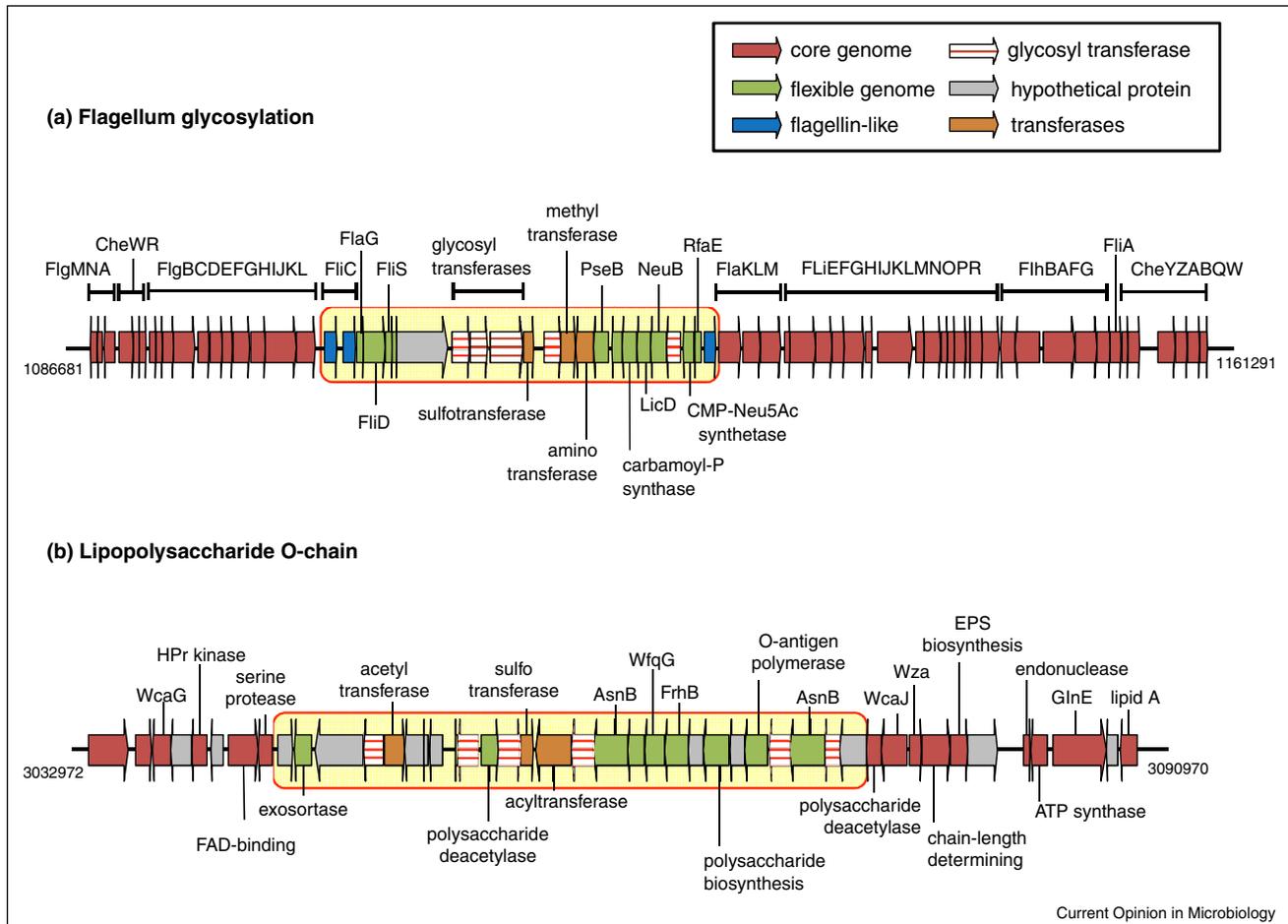
The mechanism of exchange of these fGIs has been known or suspected for a long time [25]. They seem to be always target of double recombination at the ends generating a high frequency of single nucleotide polymorphisms (SNPs) in the neighboring genes [26]. These SNPs are produced by the replacement of genomic segments by functional equivalents from other microbes of the same or different species by the replacement HGT of the complete genomic island [27]. The frequent homologous recombination events affecting these clusters that target specifically these genomic islands are instrumental in maintaining their high diversity. There is evidence that similar (syntenic) clusters are found sometimes in distant relatives at the taxonomic rank of family [25]. Although each different clonal lineage (or strain) carries a different set of replacement fGIs, some examples of conservation in different strains have been described [28]. The mechanism of replacement HGT is less efficient than illegitimate recombination that is characteristic of additive fGIs and therefore their spread throughout the population is not so fast that it would

prevent equalization of the different clones by kill-the-winner [23*].

Additive genomic islands

The variability found in additive fGIs is mechanistically simple, using the diverse and well-characterized mobile genetic elements (including lysogenic phages) as drivers of the diversity generated (Figure 2). They are designated additive because they vary by the presence/absence of a different number of gene cassettes. A typical example of this type of island are integrons, highly variable GIs that are different even within a single clonal frame [28] and allow the acquisition of gene cassettes by site-specific recombination [29]. Additive islands are often associated to tRNA genes that act as insertion sites, the addition of some cassettes is typically hall marked by the tell-tale repeat of part of the tRNA gene downstream of the inserted cassette (direct repeat) [30–32]. The repertoire of accessory genes of these additive fGIs encode functions conferring selective advantages to the host, including different metabolic properties, resistance to heavy

Figure 3



Schematic representation of the gene cluster involved in (a) the biosynthesis and assembly of flagellar components and (b) the lipopolysaccharide O-chain in the marine bacteria *Alteromonas macleodii* ATCC 27128. Regions highlighted by yellow rectangles indicate the variable region detected when comparing closely related strains both are examples of replacement flexible genomic islands.

metals, transporters to interact with the biotic and abiotic components of the environment or biotechnologically relevant features such as polyketide synthesis clusters [33,34]. There is evidence for a faster rate of change in the case of these islands and the presence of identical cassettes in isolates from different species indicates the potential for rapid HGT within microbial communities [35]. Mobile genomic elements (MGE) such as conjugative transposons, plasmids or integrative and conjugative elements (ICEs) can be also considered additive islands but their presence is much more sporadic [36,37].

Origin and consequences

That prokaryotic populations have such a high level of intraspecies diversity is extremely relevant to understand the biology of these microbes [38]. The population has many more options to survive and successfully exploit its environment with a complex and extended gene complement. It can use many more different substrates [39]

(or light intensities and qualities in photosynthetic microbes), it can protect itself better against environmental challenges (Black Queen hypothesis) [40] it can interact more efficiently with the biotic surroundings (virulence and antagonism). But, how can these complex populations endure and maintain their complexity avoiding clonal sweeps (periodic selection)? [41,42]. There is plenty of evidence showing that a part of the pangenome, the viral complement, is (at least partially) responsible for their stability [43,44]. The processes of fluctuating selection or kill-the-winner [45] provide override mechanisms that prevent natural selection from oversimplifying microbial populations [23,46]. Viruses are density dependent predators and hence can prevent any single clone from predominating since it would be preferentially preyed upon [47]. Thus, they act in a manner as tax collectors that prevent any specific individual (i.e. clone) from hoarding too much wealth (i.e. environmental resources). Viruses prevent the extreme simplification

of biodiversity that a pure survival of the (ecologically) fittest might select for.

Presently, bacterial genomes tend to be interpreted as self-standing entities whose content correlates with the physical environment or niche. However, this perception is changing in light of new population level studies pointing to a dynamic network of clonal lineages manifesting extensive flexible gene pools that define adaptation and response patterns in the environment. New tools that combine isolate sequencing, single-cell genomics and metagenomics are needed to deal with the challenge posed by this new level of complexity.

Acknowledgements

Work in FR-V laboratory is supported by projects MEDIMAX BFPU2013-48007-P from the Spanish Ministerio de Economía y Competitividad, MaCuMBA Project 311975 of the European Commission FP7 and PROMETEO II/2014/012 project AQUAMET from the Generalitat Valenciana. We are grateful for the comments and suggestions of one reviewer.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Koonin EV, Wolf YI: **Evolution of microbes and viruses: a paradigm shift in evolutionary biology?** *Front Cell Infect Microbiol* 2012;2.
2. Soucy SM, Huang J, Gogarten JP: **Horizontal gene transfer: building the web of life.** *Nat Rev Genet* 2015, **16**:472-482.
3. O'Malley MA, Koonin EV: **How stands the Tree of Life a century and a half after The Origin?** *Biol Direct* 2011, **6**:32.
4. Ochman H, Lerat E, Daubin V: **Examining bacterial species under the specter of gene transfer and exchange.** *Proc Natl Acad Sci U S A* 2005, **102**(Suppl. 1):6595-6599.
5. Vernikos G, Medini D, Riley DR, Tettelin H: **Ten years of pan-genome analyses.** *Curr Opin Microbiol* 2015, **23**:148-154.
- An updated review about approaches and bioinformatic approaches to study pangenomes focused mostly from a clinical microbiology perspective.
6. Britten RJ: **Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels.** *Proc Natl Acad Sci U S A* 2002, **99**:13633-13635.
7. Tettelin H, Riley D, Cattuto C, Medini D: **Comparative genomics: the bacterial pan-genome.** *Curr Opin Microbiol* 2008, **11**:472-477.
8. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF *et al.*: **The genome sequence of *Drosophila melanogaster*.** *Science* 2000, **287**:2185-2195.
9. Lukjancenko O, Wassenaar TM, Ussery DW: **Comparison of 61 sequenced *Escherichia coli* genomes.** *Microb Ecol* 2010, **60**:708-720.
10. Gonzaga A, Martín-Cuadrado AB, Lopez-Perez M, Megumi Mizuno C, Garcia-Heredia I, Kimes NE, Lopez-Garcia P, Moreira D, Ussery D, Zaballos M *et al.*: **Polyclonality of concurrent natural populations of *Alteromonas macleodii*.** *Genome Biol Evol* 2012, **4**:1360-1374.
11. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, Ding H, Marttinen P, Malmstrom RR, Stocker R *et al.*: **Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*.** *Science* 2014, **344**:416-420.
12. Caro-Quintero A, Konstantinidis KT: **Bacterial species may exist, metagenomics reveal.** *Environ Microbiol* 2012, **14**:347-355.
13. Konstantinidis KT, Tiedje JM: **Genomic insights that advance the species definition for prokaryotes.** *Proc Natl Acad Sci U S A* 2005, **102**:2567-2572.
14. Bendall ML, Stevens SLR, Chan L-K, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B *et al.*: **Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations.** *ISME J* 2016.
15. Belda-Ferre P, Cabrera-Rubio R, Moya A, Mira A: **Mining virulence genes using metagenomics.** *PLoS ONE* 2011, **6**:19.
16. Rodríguez-Valera F, Martín-Cuadrado AB, Rodríguez-Brito B, Pasic L, Thingstad TF, Rohwer F, Mira A: **Explaining microbial population genomics through phage predation.** *Nat Rev Microbiol* 2009, **7**:828-836.
- Proposal of a model to preserve high clonal diversity in prokaryotic populations by the action of phage predation based on metagenomic recruitment data.
17. Engelstädter J, Harms K, Johnsen PJ: **The evolutionary dynamics of integrons in changing environments.** *ISME J* 2016.
18. Kalynych S, Morona R, Cygler M: **Progress in understanding the assembly process of bacterial O-antigen.** *FEMS Microbiol Rev* 2014, **38**:1048-1050.
19. Putker F, Bos MP, Tommassen J: **Transport of lipopolysaccharide to the Gram-negative bacterial cell surface.** *FEMS Microbiol Rev* 2015, **39**:985-990.
20. Schaffer C, Messner P: **Glycobiology of surface layer proteins.** *Biochimie* 2001, **83**:591-599.
21. Cuadros-Orellana S, Martín-Cuadrado AB, Legault B, D'Auria G, Zhaxybayeva O, Papke RT, Rodríguez-Valera F: **Genomic plasticity in prokaryotes: the case of the square haloarchaeon.** *ISME J* 2007, **1**:235-245.
22. Hansen EE, Lozupone CA, Rey FE, Wu M, Guruge JL, Narra A, Goodfellow J, Zaneveld JR, McDonald DT, Goodrich JA *et al.*: **Pan-genome of the dominant human gut-associated archaeon *Methanobrevibacter smithii*, studied in twins.** *Proc Natl Acad Sci U S A* 2011, **1**:4599-4606.
23. Takeuchi N, Cordero O, Koonin E, Kaneko K: **Gene-specific selective sweeps in bacteria and archaea caused by negative frequency-dependent selection.** *BMC Biol* 2015, **13**:20.
- The authors develop a model that proves that complex populations can persist if the phage receptors do not recombine as frequently as genes providing fitness advantage. Evidence from comparative genomics seems to confirm this prediction.
24. Rodríguez-Valera F, Martín-Cuadrado A-B, Rodríguez-Brito B, Pašić L, Thingstad TF, Rohwer F, Mira A: **Explaining microbial population genomics through phage predation.** *Nat Rev Microbiol* 2009, **7**:828-836.
25. Milkman R: **Recombination and population structure in *Escherichia coli*.** *Genetics* 1997, **146**:745.
26. Lopez-Perez M, Martín-Cuadrado AB, Rodríguez-Valera F: **Homologous recombination is involved in the diversity of replacement flexible genomic islands in aquatic prokaryotes.** *Front Genet* 2014, **5**:147.
27. Choi SC, Rasmussen MD, Hubisz MJ, Gronau I, Stanhope MJ, Siepel A: **Replacing and additive horizontal gene transfer in *Streptococcus*.** *Mol Biol Evol* 2012, **29**:3309-3320.
28. Lopez-Perez M, Gonzaga A, Rodríguez-Valera F: **Genomic diversity of "deep ecotype" *Alteromonas macleodii* isolates: evidence for Pan-Mediterranean clonal frames.** *Genome Biol Evol* 2013, **5**:1220-1232.
29. Mazel D: **Integrons: agents of bacterial evolution.** *Nat Rev Microbiol* 2006, **4**:608-620.
30. Hacker J, Kaper JB: **Pathogenicity islands and the evolution of microbes.** *Ann Rev Microbiol* 2000, **54**:641-679.

31. Hou Y-M: **Transfer RNAs and pathogenicity islands.** *Trends Biochem Sci* 1999, **24**:295-298.
32. Williams KP: **Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies.** *Nucleic Acids Res* 2002, **30**:866-875.
33. Dobrindt U, Hochhut B, Hentschel U, Hacker J: **Genomic islands in pathogenic and environmental microorganisms.** *Nat Rev Microbiol* 2004, **2**:414-424.
34. Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW: **Genomic islands: tools of bacterial horizontal gene transfer and evolution.** *FEMS Microbiol Rev* 2009, **33**:376-393.
35. López-Pérez M, Gonzaga A, Ivanova EP, Rodríguez-Valera F: **Genomes of *Alteromonas australica*, a world apart.** *BMC Genomics* 2014, **15**:483.
36. Frost LS, Leplae R, Summers AO, Toussaint A: **Mobile genetic elements: the agents of open source evolution.** *Nat Rev Microbiol* 2005, **3**:722-732.
37. Wozniak RA, Waldor MK: **Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow.** *Nat Rev Microbiol* 2010, **8**:552-563.
38. Rodríguez-Valera F, Ussery DW: **Is the pan-genome also a pan-selectome?** *F1000Res* 2012, **1**:16.
A paper proposing a kin-selection mode for prokaryotes in which the unit of selection encompass multiple different clones and bacteriophages preying in them.
39. Teusink B, Bachmann H, Molenaar D: **Systems biology of lactic acid bacteria: a critical review.** *Microb Cell Fact* 2011, **10**(Suppl. 1):S11.
40. Morris JJ, Papoulis SE, Lenski RE: **Coexistence of evolving bacteria stabilized by a shared Black Queen function.** *Evolution* 2014, **68**:2960-2971.
- Proposal of the Black Queen hypothesis in which some clonal lineages provide service to others (cheaters) simplifying the genetic load of the latter.
41. Cohan FM, Koeppel AF: **The origins of ecological diversity in prokaryotes.** *Curr Biol* 2008, **18**:R1024-R1034.
42. Niehus R, Mitri S, Fletcher AG, Foster KR: **Migration and horizontal gene transfer divide microbial genomes into multiple niches.** *Nat Commun* 2015:6.
This paper provides an interesting alternative to phage predation for keeping a population with a complex genomic make up. The authors develop a model that predicts how the steady input of cells coming from different 'patches', environments that represent slightly different microniches can prevent vertical clonal sweeps and maintain diversity by transferring the same adaptive genes to different genomic backgrounds.
43. Shapiro OH, Kushmaro A, Brenner A: **Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater.** *ISME J* 2010, **4**:327-336.
44. Smid EJ, Erkus O, Spus M, Wolkers-Rooijackers JC, Alexeeva S, Kleerebezem M: **Functional implications of the microbial community structure of undefined mesophilic starter cultures.** *Microb Cell Fact* 2014, **13**(Suppl. 1):S2.
45. Avrani S, Schwartz DA, Lindell D: **Virus-host swinging party in the oceans.** *Mob Genet Elem* 2012, **2**:88-95.
46. Biller SJ, Berube PM, Lindell D, Chisholm SW: **Prochlorococcus: the structure and function of collective diversity.** *Nat Rev Microbiol* 2015, **13**:13-27.
47. Thingstad TF, Pree B, Giske J, Vage S: **What difference does it make if viruses are strain-, rather than species-specific?** *Front Microbiol* 2015, **6**:320.