MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 6073–6090 doi: 10.1111/mec.12985

Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing

ROHIT GHAI,* CAROLINA MEGUMI MIZUNO,* ANTONIO PICAZO,† ANTONIO CAMACHO† and FRANCISCO RODRIGUEZ-VALERA*

**Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, 03550 Alicante, Spain, †Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Burjassot, E-46100 Valencia, Spain*

Abstract

Freshwater ecosystems are critical but fragile environments directly affecting society and its welfare. However, our understanding of genuinely freshwater microbial communities, constrained by our capacity to manipulate its prokaryotic participants in axenic cultures, remains very rudimentary. Even the most abundant components, freshwater Actinobacteria, remain largely unknown. Here, applying deep metagenomic sequencing to the microbial community of a freshwater reservoir, we were able to circumvent this traditional bottleneck and reconstruct *de novo* seven distinct streamlined actinobacterial genomes. These genomes represent three new groups of photoheterotrophic, planktonic Actinobacteria. We describe for the first time genomes of two novel clades, acMicro (Micrococcineae, related to Luna2,) and acAMD (Actinomycetales, related to acTH1). Besides, an aggregate of contigs belonged to a new branch of the Acidimicrobiales. All are estimated to have small genomes (approximately 1.2 Mb), and their GC content varied from 40 to 61%. One of the Micrococcineae genomes encodes a proteorhodopsin, a rhodopsin type reported for the first time in Actinobacteria. The remarkable potential capacity of some of these genomes to transform recalcitrant plant detrital material, particularly lignin-derived compounds, suggests close linkages between the terrestrial and aquatic realms. Moreover, abundances of Actinobacteria correlate inversely to those of Cyanobacteria that are responsible for prolonged and frequently irretrievable damage to freshwater ecosystems. This suggests that they might serve as sentinels of impending ecological catastrophes.

Keywords: Actinobacteria, Cyanobacteria, freshwater reservoir, lignin degradation, metagenomics, rhodopsins

Received 27 June 2014; revision received 15 October 2014; accepted 17 October 2014

Introduction

The total amount of freshwater available in lakes and rivers is a miniscule amount (<1%) of the total water in the hydrosphere (Shiklomanov & Rodda 2003). It is also unevenly distributed and scarce in arid and semi-arid regions with irregular or seasonal precipitation. Creation of man-made lakes (reservoirs), allowing for accumulation of this resource, is a crucial solution, providing water for human consumption, irrigation and generation of hydroelectric power. At the turn of the last century, the number of reservoirs in the world was approximately 60 000, and hundreds of artificial lakes are added globally every year (Avakyan & Iakovleva 1998). These reservoirs, like natural lakes and rivers, also act as important sinks for organic material originating not only from their natural catchment but also from human activities, and the resident microbiota actually consume major proportions of such inflows (Berggren *et al.* 2010; Pernthaler 2013), making them essential factors directly affecting stored water properties and

Correspondence: Rohit Ghai, Fax: +34 965919576; E-mails: ghai.rohit@gmail.com; rghai@umh.es

The copyright line for this article was changed on 25 November, 2015 after original online publication.

quality. However, and in spite of its obvious ecological and public health implications, freshwater planktonic microbiota have received relatively little attention from microbiologists. Only in the last decade or so, using a combination of both cultivation dependent and independent studies, have we begun to shed some light on these microbes (Glockner *et al.* 2000; Zwart *et al.* 2002; Hahn 2003, 2009; Hahn *et al.* 2003, 2004; Yannarell & Triplett 2004). Remarkably, Actinobacteria, which were considered originally to be high-GC gram-positive typical soil bacteria, have been found to be among the most abundant groups in freshwater habitats (Glockner *et al.* 2000; Hahn *et al.* 2003; Hahn 2009; Ghai *et al.* 2011b, 2012b; Newton *et al.* 2011).

The 16S rRNA sequences obtained using clone libraries revealed that they fall into two groups, the orders Actinomycetales and Acidimicrobiales, both of which are subdivided into smaller subgroups (Glockner *et al.* 2000; Hahn *et al.* 2003; Newton *et al.* 2007, 2011; Hahn 2009). These newly discovered groups are largely uncultured, but there have been successes in enrichment cultures (Hahn 2009; Jezbera *et al.* 2009; Garcia *et al.* 2013). For most of these microbes, the availability of only 16S rRNA sequences made it difficult to infer the characteristics and roles in the natural environment. Microscopic analyses indicated that some of these microbes were very small (Hahn *et al.* 2003), with cell biovolume $<0.1\ \mu m^3$. Another important finding was the discovery of proton-pumping rhodopsins (actinorhodopsins) in freshwater Actinobacteria, revealing a photoheterotrophic lifestyle (Sharma *et al.* 2008, 2009).

Further evidence emerged from next-generation sequencing (NGS) metagenomic data that are not biased by PCR or cloning, providing more reliable quantitative estimations of actinobacterial abundance. For example, they were the most abundant group in the pristine upper stretch of the Amazon River (Ghai *et al.* 2011b). Besides, the unusually low-GC content of these microbes (as low as 42%) (Ghai *et al.* 2011b, 2012b) was revealed. However, genomic information on freshwater Actinobacterial groups has remained scarce. The first genome obtained from the sequencing of a single cell belonged to the acI-B1 lineage (within Actinomycetales) (Garcia *et al.* 2012). This particular genome was very small and streamlined, estimated to be only 1.2 Mb. Two additional genomes, both belonging to the Luna-1 clade (Micrococcineae) have been described. The first, *Ca.* Aquiluna sp. strain IMCC13023, was isolated from an Arctic fjord (Kang *et al.* 2012). The second, *Rhodoluna lacicola* (Hahn *et al.* 2014), isolated from Lake Taihu in China is the only example of a planktonic freshwater actinobacterium in pure culture (Hahn *et al.* 2014). These genomes are slightly larger, although still small (1.43 and 1.35 Mb, respectively), and with higher GC content (51.6 and 51.5%) than the acI-B1 genome. They reinforced the view that freshwater Actinobacteria are quite different from their soil relatives that often possess mycelial structures with complex life cycles and tend to have large genomes (some are among the largest in prokaryotes). Actually, the recently described new subclass of marine Actinobacteria (*Candidatus* Actinomarinidae) is perhaps among the smallest free-living prokaryotes, with a biovolume of only $0.013\ \mu m^3$ (Ghai *et al.* 2013), and also presents a small estimated genome size (<1 Mb). The GC content of this group, only 33%, is the lowest described till date for Actinobacteria, making it clear that high-GC content, a classical feature ascribed to this group, is actually an unsuitable descriptor.

In this work, we have studied the planktonic microbiota of an unstratified, mesotrophic freshwater reservoir, called Amadorio, using deep metagenome sequencing that has already been shown to be an efficient method to recover genomes of microbes. Although not as revealing as a pure culture genome, metagenomic genome reconstruction provides a global view of the genomic properties and potential that becomes a powerful tool to predict ecophysiological properties of the microbes (Albertsen *et al.* 2013; Rinke *et al.* 2013). In particular, we have focused on the abundant actinobacterial component of the microbial community because large genomic fragments of Actinobacteria were assembled from our data sets. We describe the reconstruction and functional capabilities of seven distinct reconstructed genomes of uncultured aquatic Actinobacteria retrieved directly from the metagenome. They reveal how these microbes are responsible for the biogeochemical cycling of important nutrients in relatively pristine freshwater systems.

Materials and methods

Sampling site description

The freshwater water body sampled in this study, Amadorio Reservoir, is a medium-sized (maximum capacity is $15.8 \times 10^6\ m^3$) freshwater man-made lake located near the Mediterranean coast of eastern Spain, approximately 2 Km from the sea, around 125 m.a.s.l. (38°32′06.01″N 0°15′59.51″W; UTM WGS84 30S 738257.91E 4268725.29N). A small stream, the Sella river flows into the reservoir and water flow is highly variable and seasonal. At the time of sampling, its maximum depth of the reservoir was around 35 m. The reservoir water is relatively mineral rich (ranging from 1 to 2 mS/cm depending on the renewal time and evaporation) because of the dominance of a gypsum-rich substrate of the catchment area, sulphate and calcium being the most abundant ions. Due to the high sulphate concentrations and medium primary productivity, an anoxic

hypolimnion with high sulphide concentrations usually develops during the summer stratification period. This commonly lasts from April to October, whereas the rest of the year, the water column remains mixed and waters are well oxygenated. Previous studies have shown that the Amadorio Reservoir commonly displays a mesotrophic to slightly eutrophic status under the normal storage levels, which usually account for less than half of the reservoir's capacity, with average chlorophyll-a concentrations of about 10 μg/L (Dasi *et al.* 1998; Vicente *et al.* 2000). At the time of sampling, the reservoir level was nearly at its maximum capacity, and trophic status was lower and among the mesotrophic values. The water quality of this reservoir is being monitored *in situ* at real time since 2009 using an Automatic Dammed Water Quality Monitoring System (aquaDam®, ADASA Sistemas SL, Spain), including mobile multiprobes and on line data transmission(Camacho *et al.* 2013). Several physicochemical characteristics of the water column are also provided in Fig. S1 and Table S1 (Supporting information).

Prior to sampling, vertical profiles of temperature, electrical conductivity, dissolved oxygen, pH, Eh, density, chlorophyll-a concentration, phycoerythrin and phycocyanin fluorescence, and chromophoric dissolved organic matter (cDOM) were determined *in situ* using a Seabird SBE 19 multiprobe profiler including fluorometers. Secchi disc was also measured *in situ*. Additionally, a number of physical–chemical and biological parameters were analysed on the same sample obtained for the metagenomic analyses. Main chemical variables including inorganic soluble forms of nitrogen (NOx and ammonium) and phosphorus (soluble reactive phosphorus), as well as total nitrogen (TN) and total phosphorus (TP) were determined according to APHA-AWWA-WEF (Greenberg *et al.* 1992). Total organic carbon (TOC) was determined on a Shimadzu TOC-VCSN Analyser. Phytoplankton was counted on an inverted microscope after sedimentation of an integrated sample (0–10 m) using the Utermöhl sedimentation method. Photosynthetic pigments were determined by HPLC after extraction in acetone following Picazo *et al.* (2013).

At the time of sampling, the water column of the Amadorio reservoir was completely mixed, and physical and chemical features were homogeneous through the water column, thus the sampled depth (10 m) can be considered as representative of the reservoir (Fig. S1, Supporting information). According to OECD criteria (Dodds 2002), following chlorophyll-a and total phosphorus concentrations, as well as Secchi disc depth, the reservoir can be assigned to the higher range of the mesotrophic trophic status (Table S1, Supporting information). Well-oxygenated waters and high redox potential through the water column resulted in the dominance of oxidized inorganic nitrogen forms (nitrate), more than five times higher than reduced forms (mainly ammonium), with much lower abundance of biologically available soluble phosphorus compounds. As most phosphorus was mainly found not as soluble compounds but instead as the particulate form, mainly as algal biomass, the molar N/P ratio did not exceed the Redfield ratio (Dodds 2002) by much, which in fact ranged from 14 to 52 through the water column, showing that an evident nutrient limitation can be attributed neither to nitrogen nor phosphorus deficiency exclusively. Bacterioplankton was moderately abundant, around 2 million cell/mL, but most of it was composed by low activity cells (only 19.4% HDNA), which can be related to the relatively low concentrations of organic carbon. A total of 15 species of phytoplankton were determined by microscopic counts, which was typically dominated by centric diatoms (*Cyclotella radiosa* and *C. ocellata*) with two species that jointly accounted for more than 86% of the phytoplankton biovolume, with the presence of subdominant species of dinoflagellates (*Ceratium hirundinella*) and cryptophytes (*Cryptomonas erosa*) that are also typical from such Mediterranean reservoirs. Continuously registered data from the real-time monitoring system during the hydrological cycle 2011–2012 covering several variables (mostly temperature, dissolved oxygen, Eh and chlorophyll-a concentration) are provided by Camacho *et al.* (2013).

Sampling, sequencing, annotation

A single sample was taken on 1 February 2012 from a depth of 10 m. A total of 50 L was sequentially filtered through a series of 20, 5 and a 0.1-μm-pore-size polycarbonate filters (Millipore). DNA was extracted from the 0.1- to the 5-μm filters as previously described (Ghai *et al.* 2010). In brief, filters were treated with 1 mg/mL lysozyme and 0.2 mg/mL proteinase K (final concentrations) and DNA was extracted using phenol/chloroform/isoamyl alcohol and chloroform/isoamyl alcohol. Sequencing was performed using Illumina HiSeq 2000 (GATC, Konstanz, Germany). A total of approximately 200 million sequence reads (PE, 100 bp) representing nearly 20 Gb of sequence data were produced for the free-living fraction (0.1-μm filter). Another 47 million reads were obtained for the large-size fraction (4.7 Gb, 5-μm filter). Both data sets were assembled together using the IDBA assembler (Peng *et al.* 2012). Gene predictions on the assembled contigs were performed using Prodigal in metagenomic mode (Hyatt *et al.* 2010), and tRNAs were predicted using tRNAscan-SE (Lowe & Eddy 1997). Ribosomal rRNA genes were identified using ssu-align (Nawrocki 2009) and meta_rna (Huang *et al.* 2009). Functional annotation

was performed by comparison of predicted protein sequences by comparisons against the NCBI NR, COG (Tatusov *et al.* 2001) and TIGRFAM (Haft *et al.* 2001) databases. Genomes were also annotated using the RAST server (Aziz *et al.* 2008). Additional local BLAST searches against the latest NCBI-NR database were performed whenever necessary.

16S rRNA classification

A non-redundant version of the RDP database was prepared by clustering all available 16S sequences (approximately 2.3 million) into approximately 800 000 sequences at 90% identity level using UCLUST (Edgar 2010). This database was used to identify candidate 16S rRNA sequences in the Illumina data sets. If a sequence hits this database at an *e*-value, <1e-5 was considered a potential 16S sequence. The resulting sequences were aligned to archaeal, bacterial and eukaryal 16S/18S rRNA HMM models using ssu-align to identify true 16S/18S sequences (Nawrocki 2009). The resulting prokaryotic sequences were compared to the entire RDP database and classified into a high level taxon if the sequence identity was ≥80% and the alignment length was ≥90 bp. Sequences failing these thresholds were discarded.

Identification of bona-fide actinobacterial contigs and genome reconstruction

Only contigs that were larger than 10 kb were used in the genome reconstructions. Contigs were considered actinobacterial if a majority of genes gave best BLAST hits to Actinobacteria. The actinobacterial contigs were grouped using taxonomy, principal component analysis of tetranucleotide frequencies, GC% and coverage in both the metagenomes as described previously (Ghai *et al.* 2011a; Albertsen *et al.* 2013; Rinke *et al.* 2013). Tetranucleotide frequencies were computed using the wordfreq program in the EMBOSS package (Rice *et al.* 2000). Principal components analysis was performed using the FACTOMINER package in R (Lê *et al.* 2008).

Genomic phylogenetic trees and genome size estimation

To create the reference protein-concatenate-based phylogenetic tree for Actinobacteria, 150 complete actinobacterial genomes were used. Four complete firmicute genomes were also used as outgroup. A total of 83 conserved genes were found in all (based on the COG database). These 83 genes were used to create the reference phylogeny of Actinobacteria. A list of these genes is provided in the Data S1 (Supporting information). The alignment was performed using KALIGN (Lassmann & Sonnhammer 2005), trimmed using TRIMAL (Capella-Gutiérrez *et al.* 2009). A maximum-likelihood tree was constructed using FASTTREE2 (Price *et al.* 2010) using a JTT+CAT model, a gamma approximation and with 100 bootstrap replicates. To identify the position of each new assembled actinobacterial genome within the actinobacterial phylogeny, each assembled genome was added individually to this reference tree using whatever number of conserved proteins that could be identified. All trees were made using the same method as described above. The same set of 83 conserved genes was also used to estimate genome completeness of the new actinobacterial genomes described in this work.

Single gene trees

All 16S rRNA sequences were trimmed using ssu-align (Nawrocki & Eddy 2010) and aligned using MUSCLE (Edgar 2004), and the alignment was trimmed manually wherever necessary. 16S rRNA trees were constructed using FASTTREE2 (Price *et al.* 2010) with a GTR + CAT model, a gamma approximation and 100 bootstrap replicates. The rhodopsin sequences were aligned using MUSCLE, and the maximum-likelihood tree was constructed using FASTTREE2 (Price *et al.* 2010) using a JTT model, a gamma approximation and 100 rapid bootstraps.

Correlation analysis

The analysis of correlations between the 16S rRNA abundances of taxa across metagenomic data sets found was performed in R (http://r-project.org). The all-vs-all correlations and their significance were computed using the rcorr function in the package HMISC. The correlation matrix was plotted using the corrgram function in the CORRGRAM package. The significant correlations were plotted using SIGMAPLOT.

FISH and bacterioplankton size spectrum determinations

For microscopic counts of heterotrophic bacterioplankton, water samples were fixed *in situ* with a paraformaldehyde:glutaraldehyde solution to a final concentration in the sample of 1%: 0.05% (w/v) (Marie *et al.* 1997). Once in the laboratory, subsamples of 5–10 mL were filtered through 0.2-μm-pore-size black filters (Nuclepore, Whatman). For quantification, a quarter of filter was stained with 4′, 6- diamidino-2-phenylindole (DAPI) (Porter & Feig 1980) (SIGMA) and counted with an inverted Zeiss III RS epifluorescence microscope (1250X, resolution 0.02857 μm/pixel) using a G365 exciting

filter, LP420 suppression filter for blue light and G546 exciting filter, LP590 suppression filter for green light (MacIsaac & Stockner 1993). For FISH detection of the acAMD group and other Actinobacteria, water samples were fixed with paraformaldehyde to 2% final concentration, then filtered within the next 2 h although white polycarbonate filters (0.2 μm pore size) sections, stained with the different oligonucleotide probes and with DAPI, and mounted for microscopic evaluation. We primarily used a previously described general actinobacterial probe (HGC236, Glockner *et al.* 2000). Although HGC236 has commonly been considered as a general FISH probe for Actinobacteria[6], according to our results, this did not appear to be the case for us as it does not match well with some of the freshwater actinobacterial groups we found. Thus, we designed a specific probe for the acAMD using the PRIMER3 tool (Rozen & Skaletsky 2000). This new probe (acAMD829, 'TGCGTT AGCGACGTCGCAGA') was labelled with the indocarbocyanine dye Cy3 (Thermo Scientific, Waltham, MA, USA). These and other probes matching actinobacterial sequences were checked for specificity using the Probe Match RDP (Cole *et al.* 2009), Release 11, Update 2. March 7, 2014 (2 929 433 16S rRNAs database, https://rdp.cme.msu.edu/index.jsp). The FISH protocol was performed as described in Sekar *et al.* (2003). Hybridization conditions for the probe acAMD829 were adjusted by formamide (VWR BDH Prolabo) series applied to different subsamples. Absolute densities of hybridized bacteria were calculated as the product of their relative abundances on filter sections (percentage of DAPI-stained objects) and the DAPI-stained direct cell counts. A minimum of 500 DAPI and probe-stained cells were measured per sample. Images from FISH determinations were analysed using the NIH IMAGEJ Software to determine cell dimensions (http://rsb.info.nih.gov/ij/index.html). The biovolume of acAMD Actinobacteria was calculated assuming a spherical shape.

For cytometric identification, quantification and size structure approximation (Bouvier *et al.* 2001) of the bacterioplankton, a Coulter Cytomics FC500 flow cytometer (Brea, California, USA) equipped with an argon laser (488 excitation), a red emitting diode (635 excitation), and five filters for fluorescent emission (FL1–FL5), was used. Bacterioplankton size structure was determined with argon laser by green fluorescence (Sybr Green I, Sigma-Aldrich, Missouri, USA) using a FL1 detector (525 nm). For size calibration, beads (polystyrene fluorospheres) of different sizes (0.79, 1, 4.9 and 10 mm) were used. The lower and upper size limits of measurement are 0.25–40 μm, respectively. The measured diameter of the smaller Actinobacteria found, the acAMD cells, is close to the lower end of the scale.

Metagenomic recruitment

Recruitments were performed against several publicly available freshwater metagenomes using BLASTN (Altschul *et al.* 1997), and a hit was considered only when it was at least 50 nucleotides long, with a %identity of >95% and with an *e*-value ≤ 1e-5. These hits were used to compute the RPKG (reads recruited per Kb of genome per Gb of metagenome) values that provide a normalized number comparable across various metagenomes.

Metagenomic data sets

All metagenomic data sets used in this work are publicly available, Amazon River (Ghai *et al.* 2011b), Lake Lanier (Oh *et al.* 2011), Lake Gatun (Rusch *et al.* 2007) and Albufera Lagoon (Ghai *et al.* 2012a). Metagenomic data sets from Lake Mendota, Trout Bog Lake, Lake Vattern, Lake Damariscotta, Lake Ekoln, Lake Erken and Sparkling Lake are described in Eiler *et al.* (2013).

Results

Reconstruction of novel actinobacterial genomes

We have studied by deep metagenomic sequencing a freshwater reservoir, known as Amadorio located near the Mediterranean coast of south-eastern Spain, a region with a semi-arid water regime, typical of the Mediterranean climate (Dasí *et al.* 1998; Vicente *et al.* 2000) (see Materials and methods, Fig. S1 and Table S1, Supporting information). Two filter size fractions have been independently sequenced: small (0.1–5 μm) and large (5–20 μm). The broad taxonomic composition, using 16S rRNA sequences, of the prokaryotic community of the reservoir, in comparison to several other freshwater metagenomic data sets is shown in Fig. S2 (Supporting information). These results illustrate that Actinobacteria comprise large proportions of freshwater microbial communities across most data sets. In addition, the largest number of assembled contigs belonged to Actinobacteria and Bacteroidetes. As freshwater Actinobacteria are more uniformly abundant than Bacteroidetes, in this work, we have focussed only on them. To maximize the chances of retrieving long genomic fragments, both the small- and the large-fraction-derived metagenomic data sets were assembled together. After assembly, we have used only long genomic fragments (minimum 10 Kb length) that could be assigned to Actinobacteria based on the majority of genes within these contigs giving highest similarities to actinobacterial genes in databases (see Materials and methods). Further manual examination of these contigs was performed,

and finally, a total of 779 (17.87 Mb) contigs were considered. We classified these contigs into groups using taxonomic information, GC content, principal component analysis of oligonucleotide frequencies and relative coverage in both metagenomic data sets (Ghai *et al.* 2011a; Albertsen *et al.* 2013; Rinke *et al.* 2013). Using these criteria, 327 contigs (8.34 Mb) were binned into seven distinct groups (Table 1). For the sake of simplification, we will refer to each group as a genome even though each is certainly derived from a population of closely related organisms. Among the recovered genomes, two were related to the suborder Micrococcineae (the acMicro group), two to the acI-B lineage and another two to a new group distantly related to the acTH1 lineage (the acAMD group). In addition, a set of 132 contigs, that probably correspond to two different genomes and could not be separated any further, was assigned to the order Acidimicrobiales (the acAcidi group). Three of these genomes were estimated to be approximately 90% complete (Table 1). Apart from these, another 197 contigs (4.6 Mb) were assigned to the acI-A clade, 16 to acI-B clade and 239 remained as unclassified actinobacterial contigs (4.47 Mb). They could not be reliably clustered except for the broad classification mentioned and were used only for limited analyses (e.g. fragment recruitment).

Genome size estimation showed that they are all small sized (approximately 1.2 Mb) (Fig. S3, Supporting information). Moreover, median intergenic distance in planktonic actinobacterial genomes appears to be rather small, with a few exceptions, in the range of approximately 10 bp (Fig. S4, Supporting information). Along with small genomes, densely packed genes are also emerging as a nearly universal feature of planktonic Actinobacteria. Actually both genomes of the acMicro group, and one genome from the acAMD group (acAMD-2), are somewhat exceptional in this regard, in the sense that they have small genomes, with high-GC content, and larger intergenic spacers in comparison to the rest. Interestingly, for all these genomes, although we could not find a correlation between GC content and genome size, there was a remarkably high correlation between GC content and median spacer size ($R = 0.88$), that is densely packed planktonic Actinobacterial genomes tend to have low-GC content (Fig. S4, Supporting information).

To establish the phylogenetic placement of the new genomes, we constructed a concatenate (83 genes) tree of 150 actinobacterial reference genomes with four Firmicutes used as outgroup. We also used the single cell amplified genome (SAG) of acI-B1 described before (Garcia *et al.* 2012) and additional publicly available SAGs of acI-A and acI-B lineages. The resulting tree recapitulates the phylogeny of the entire phylum which was previously based on the 16S ribosomal rRNA alone (Fig. 1). To maximize accuracy of phylogenetic placement using more conserved genes, we created separate phylogenetic trees with closely related references for each of our assembled genomes (Fig. 1). The new genomes could be placed at three different locations within the larger phylogeny of Actinobacteria, six of these within the order Actinomycetales and one within Acidimicrobiales. In addition, the tree provides a more in depth view of actinobacterial phylogeny, suggesting for example that not only do the acI lineages belong to the order Actinomycetales (also suggested previously, Jezbera *et al.* 2009), but also constitute a suborder within the order Actinomycetales.

The two genomes within the suborder Micrococcineae (named acMicro-1 and acMicro-4) had high-GC contents (61 and 59%, respectively). The acMicro-4 genome was nearly complete (92%), while acMicro-1 was 71% complete. Both these genomes appear to be rather small (approximately 1.2 Mb) (Table 1). The average

Table 1 Summary statistics of the reconstructed genomes

	GC	No. of contigs	Total length (kb)	No. of proteins	rRNA length (bp)	%Complete	Estimated genome length (Mb)
acMicro group							
acMicro-1	61.64	36	856	873	—	73.49	1.16
acMicro-4	59.28	15	1083	1114	366	89.15	1.21
acAMD group							
acAMD-2	57.33	37	973	965		73.49	1.32
acAMD-5	44.42	10	1242	1259	982	97.59	1.27
acAcidi group*	52.15	132	2690	2884	—	91.56	1.17
acI-B related							
acIB-AMD-6	42.06	31	523	551	—	37.34	1.4
acIB-AMD-7	40.42	66	974	1004	—	51.8	1.88

*This group contains two genomes.

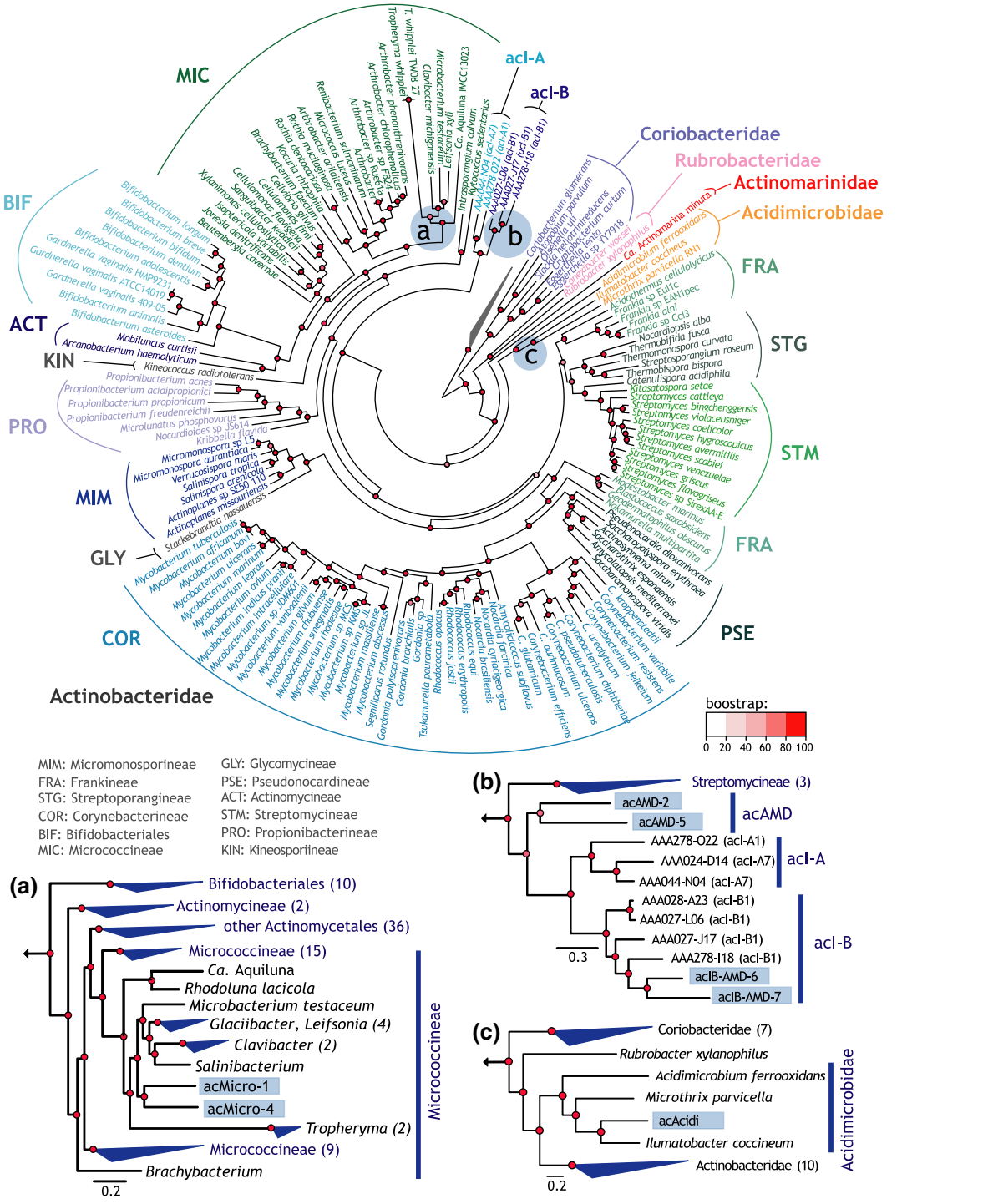

Fig. 1 Phylogenomics of the phylum Actinobacteria. Eighty-three conserved genes were used to generate a maximum-likelihood phylogenetic tree of the entire Actinobacteria phylum. In addition to all the complete genomes, five single cell genomes (prefixed by AAA) and one assembled from metagenomic fosmids (*Candidatus* Actinomarina) were used to create the tree. Four firmicute genomes are used as outgroup. The three locations where the assembled genomes in this work could be placed are indicated by blue-shaded circles and labelled as a, b and c. Bootstrap values are indicated by coloured circles on the nodes, and a colour key is shown below the tree. The high resolution phylogenomic neighbourhoods of the new assembled genomes are shown in (a) *Micrococcineae* genomes acMicro-1 and acMicro-4 (91 genomes, 150 conserved proteins), (b) genomes belonging to the novel acAMD lineage (acAMD-2 and acAMD-5) and the acI-B related genomes (acIB-AMD-6 and acIB-AMD-7) (16 genomes, 55 conserved proteins) and (c) the acAcidi group belonging to the order *Acidimicrobiales* (23 genomes, 280 proteins in tree).

nucleotide identity between them was 71% and conserved DNA (>90% identity) was only 3%, indicating they are likely distinct genera. They appear to be most closely related to each other first and then to *Clavibacter*, *Salinibacterium*, *Leifsonia* none of which are known to be freshwater bacteria (Fig. 1). The closest truly aquatic actinobacterial genome within Micrococcineae is *Rhodoluna lacicola* (Hahn *et al.* 2014) which belongs to the Luna-1 clade (named because of their solenoid morphology) (Hahn 2009) and described first on the grounds of environmental 16S rRNA sequences. To determine whether the acMicro genomes described here have any relationship with the Luna clade, we used the 366-bp 16S rRNA gene fragment from the acMicro-4 genome to recover related sequences. Unfortunately, no hit was obtained among cultured microbes. However, several hits to environmental 16S rRNA were retrieved at 99% identity. These sequences originated from several freshwater lakes, groundwater, sediments, estuaries and even saline lakes. Phylogenetic analysis of these recovered sequences showed that they are closely related to the Luna-2 clade of freshwater Actinobacteria, but belong to an independent clade (Fig. S5, Supporting information). We will refer to this clade as the 'acMicro' clade.

Two genomes (acIB-AMD-6 and acIB-AMD-7) appeared to be related to the acI-B lineage of Actinobacteria using a phylogeny of 55 conserved genes (Fig. 1). Although our contigs did not include any 16S rRNA sequences, the closest genomes are SAGs AAA027-J17 and AAA278-I18 (from Lake Mendota, Wisconsin and Lake Damariscotta, Maine, respectively) (Ghylin *et al.* 2014). These SAGs contain approximately 400-bp 16S rRNA sequences that are nearly 100% identical along their entire lengths to several 16S rRNA clone sequences classified as acIB-1. The genomes of acIB-AMD-6 and acIB-AMD-7 showed average nucleotide identity >70% and low conserved DNA (<3%) both with the acI-B1 SAG genome (Garcia *et al.* 2012) and with each other, showing that they are quite different. Like the published acI-B1 (Garcia *et al.* 2012), the acIB-AMD-6 genome had low GC and small size (approximately 1.4 Mb) but the one of acIB-AMD-7 was predicted to be relatively larger (approximately 1.8 Mb).

Genomes acAMD-2 and acAMD-5 clustered next to the acI group (Fig. 1). These were also estimated to be small (1.36 and 1.25 Mb), however, with low (44%) and high GC (57%), respectively. A large 16S rRNA fragment (982 bp) was recovered in one of the contigs of acAMD-5. We used this sequence to retrieve additional 16S rRNA sequences of this particular group in 16S rRNA clone libraries and found several related sequences (>98% identity along the entire length) in freshwater habitats (e.g. Marathonas reservoir, Greece). Phylogenetic analysis of this group of sequences in the context of known actinobacterial 16S rRNA phylogeny revealed that they formed a separate cluster related to the 16S rRNA defined acTH1 clade (Newton *et al.* 2011) (Fig. S5, Supporting information). We will refer to this group as the acAMD group. No genome sequences are available for acTH1, and so, these new microbes represent the first genomes of freshwater actinomycetales related to the acI-lineage. The 16S rRNA sequence retrieved allowed the development of a specific FISH probe for the newly described acAMD group (see Materials and methods). This probe (acAMD829) was used to quantify these microbes in the same Amadorio sample used for the metagenome. Microscopy and flow cytometry analysis indicated that acAMD cells are not very prevalent but have significant and easily detectable numbers in the water column. 0.4% of the total DAPI-stained bacterioplankton cells (2.01×10^6 cell/mL) hybridized with the lineage-specific probe, while approximately 4.7% hybridized with a more general actinobacterial FISH probe (HGC236[6]). Combined microscopy and flow cytometry analyses (Figs S6 and S7, Supporting information) suggested a small average cell diameter (0.378 ± 0.042 SD μm) for the acAMD group. Assuming a spherical shape, the calculated biovolume is also very small (0.028 ± 0.006 SD μm^3, range 0.021–0.036 μm^3), thus illustrating the discovery of yet another group of small planktonic Actinobacteria.

A hundred and thirty-two (132) contigs (approximately 2.9 Mb) could be affiliated to the order Acidimicrobiales. Several universally conserved actinobacterial genes were found twice in this group of contigs, suggesting that at least two genomes were present in this collection. However, a further separation of both genomes was not possible. The estimated genome length was approximately 1.17 Mb (see Materials and methods). We will refer to these contigs as the acAcidi group. There was no 16S rRNA assembled in these contigs, but several genes gave best hits to *Ilumatobacter coccineum*, isolated from estuarine sediments (Matsumoto *et al.* 2009) and which is also classified within the acIV group of freshwater Actinobacteria (Newton *et al.* 2011). The phylogenetic analysis using 280 conserved genes also suggests that the closest related genome was of *Ilumatobacter coccineum* (Fujinami *et al.* 2013) (Fig. 1). However, *Ilumatobacter* is not a photoheterotroph and remains a very distant relative by sequence analysis (<65% ANI), while one of the acAcidi group contigs encoded for a rhodopsin.

Photoheterotrophic nature of the novel organisms

Planktonic low-GC Actinobacteria, both freshwater and marine, have been shown to have a photoheterotrophic lifestyle (Sharma *et al.* 2006, 2008, 2009; Ghai *et al.* 2013). A total of seven rhodopsin sequences could be identified

in our actinobacterial contigs. Of these, one each was found in the acAMD-5, acMicro-4 and acAcidi genomes and four others in the unclassified actinobacterial contigs. Phylogenetic analysis and examination of specific amino acid residues revealed that most of these rhodopsins, including the ones of acAMD-5 and acAcidi, were 'actinorhodopsins' (Fig. 2, Table S2, Supporting information). Within the actinorhodopsin group, three clusters can also be seen: rhodopsins originating from acI-B, acI-A and those related to *Rhodoluna* and *Ca*. Aquiluna. On the other hand, the acMicro-4 rhodopsin was related to a firmicute proteorhodopsin found in *Exiguobacterium* sp. for which an unusual protein structure has been recently described (Gushchin *et al.* 2013). The acMicro-4 rhodopsin displays characteristic conserved residues indicating that it is a proteorhodopsin and not an actinorhodopsin (Table S2, Supporting information). To our knowledge, this is the first proteorhodopsin described in Actinobacteria, illustrating the extremely promiscuous nature of rhodopsin genes in nature whose distribution is often totally disconnected from the genome phylogeny. This sequence also gave an exact match (107 aa) to a partial rhodopsin sequence originating from the freshwater Lake Kinneret in Israel (Clone identifier Dec08_aaa01c07_1m). All rhodopsins described here absorb in the green part of the spectrum (L variants), which in the ocean are found in relatively shallow environments, for example coastal waters (Table S2, Supporting information) (Sharma *et al.* 2009).

Metabolic overview

A general overview of the metabolic features of these genomes is shown in Fig. 3, and additional features for the most complete representative of each group are shown in Figs S8–S10 (Supporting information). Similar to acI-B

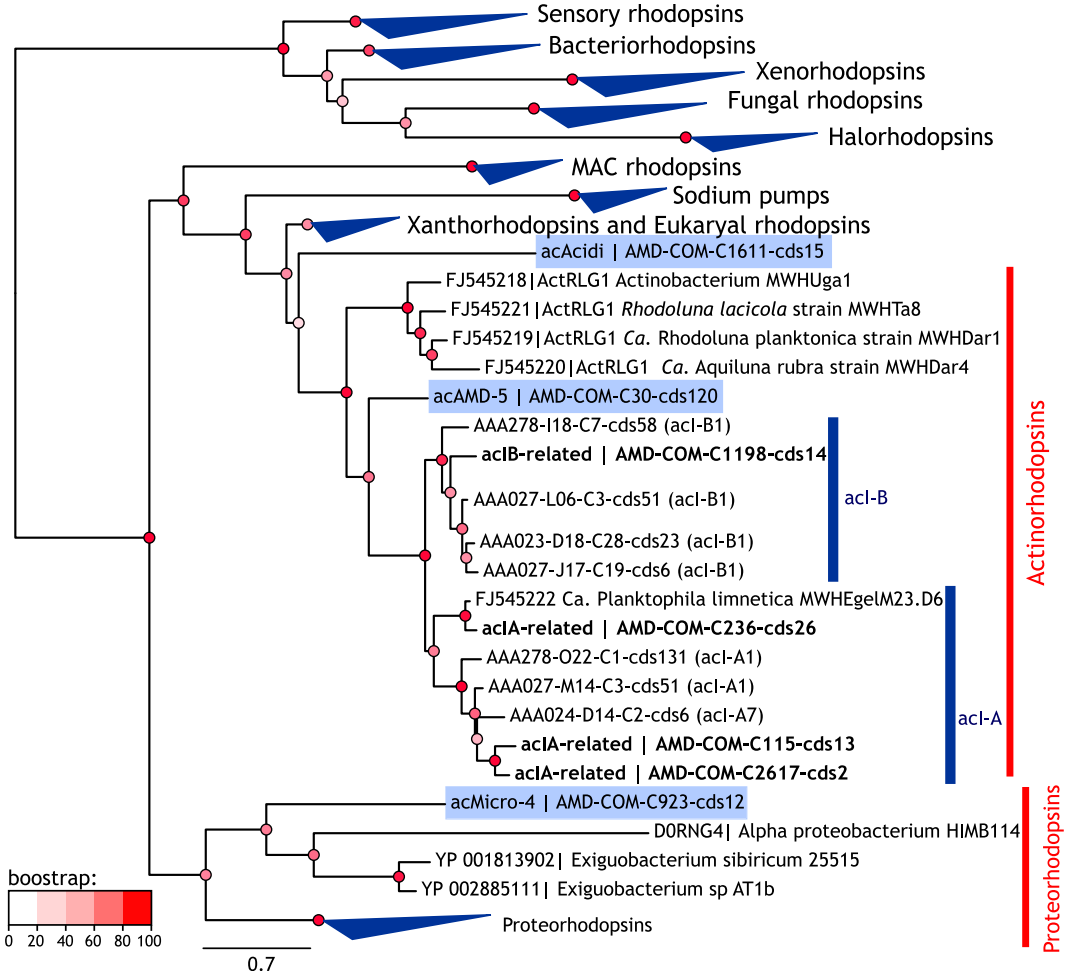

Fig. 2 Phylogenetic analysis of rhodopsins. A maximum-likelihood tree of all actinobacterial rhodopsins identified in this work, along with several reference rhodopsins, is shown. Sequences identified in our actinobacterial contigs are highlighted in boldface. Additionally, the rhodopsins of acAcidi, acAMD and acMicro groups are highlighted with a blue background. Bootstrap values on nodes are indicated by coloured circles (see colour key on bottom right).

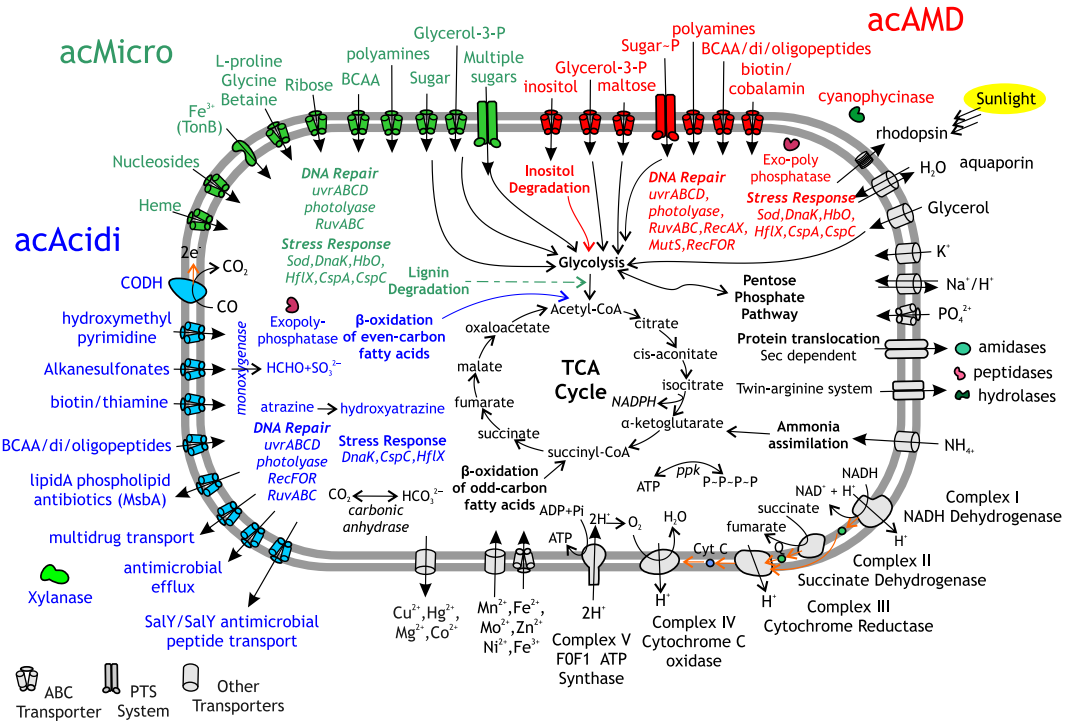

Fig. 3 Cellular overview of freshwater Actinobacteria. Representation of the various physiological processes identified that are common and specific to the freshwater Actinobacterial groups identified in this work. All processes that are shared by the acAcidi, acAMD and the acMicro groups are labelled in black. Processes specific to each group are shown in different colours (acAcidi: blue, acMicro: green, acAMD: red). CODH: Carbon monoxide dehydrogenase; BCAA: Branched chain amino acids.

genome, these organisms respire aerobically, although the presence of a respiratory nitrate reductase in acIB-AMD-7 suggests possibilities for anaerobic metabolism for this organism. Only one genome (acMicro-1) encoded both the oxidative and nonoxidative branches of the pentose phosphate pathway (PPP). In all others, only the reversible nonoxidative branch of PPP was found. However, alternative mechanisms for NADPH production, essential for anabolic pathways, were found in all (e.g. malic enzyme, NADP+ dependent isocitrate dehydrogenase) (Data S1, Supporting information). A major difference with the acI-B1 genome described before (Garcia *et al.* 2012) was that none of these genomes showed any evidence of the Entner–Duodoroff pathway.

All genomes encoded glycerol and glycerol-3-phosphate transporters, thus glycerol appears to be a commonly used carbon source. Glycerol may be used both in aerobic and anaerobic conditions (Murarka *et al.* 2008; Martínez-Gómez *et al.* 2012), what increases the facultative anaerobic capacity to these organisms. Apart from glycerol, a diverse array of transporters to transport organic compounds like branched chain amino acids (BCAA), oligopeptides and dipeptides, diverse sugars, nucleosides and polyamines were found (Fig. 3). For the acI-B genome, it was predicted that xylose, arabinose and ribose could be utilized (Garcia *et al.* 2012). Ribose transporters were also found in both the acMicro-4, acIB-AMD-6 and acIB-AMD-7 genomes. A xylanase, for degradation of xylan, a major component of the hemicellulose fraction of plant cell walls was also found in the acAcidi group. The acAcidi, acAMD-2 and acAMD-5 genomes encoded genes for utilization of both odd and even carbon fatty acids, while the others appeared to be able to use one or the other (Data S1, Supporting information). Use of odd-carbon fatty acids in particular may be important as they are more common in plant material (Lehninger *et al.* 2005), and these Actinobacteria may be involved in recycling of these compounds in the natural environment. We also found a cyanophycinase in both the acAMD group genomes and the acI-A related contigs (similar to the acI-B1 genome, Garcia *et al.* 2012) but not in acAcidi or acMicro groups. This enzyme degrades cyanophycin, a storage polymer of several cyanobacteria that might be a significant source of carbon and nitrogen for these microbes. This also appears to be a common feature of freshwater and marine Actinobacteria (Ghai *et al.* 2013).

Three genes coding for aerobic carbon monoxide dehydrogenases (CODH) were found in the acAcidi group of contigs. Photolysis of dissolved organic matter is a source of carbon monoxide in the aquatic environment, and CODH allow utilization of carbon monoxide as an energy

source (King & Weber 2007). CODH enzymes, based on catalytic site residues and reaction speed, are classified into Form I (fast CO oxidizers) and Form II (10–10 000 times slower than Form I). All sequences found here had the sequence patterns characteristic of Form II enzymes (slow CO oxidizers). Although CODH genes have been widely reported from the marine environment (Moran *et al.* 2004; Tolli & Taylor 2005; Martin-Cuadrado *et al.* 2009), to our knowledge, these are the first to be described from planktonic freshwater microbes.

All genomes encoded exopolyphosphatases, polyphosphate kinases and ABC transporters for phosphate uptake and storage. However, acMicro-4 encoded a polyphosphate glucokinase which phosphorylates glucose using polyphosphate instead of ATP. Such activity has been demonstrated before in other polyphosphate accumulating Actinobacteria, for example *Microlunatus phosphovorus* (Kawakoshi *et al.* 2012). The presence of these phosphate accumulating mechanisms, together with the sheer abundance of these Actinobacteria in their aquatic environments suggests that they might be important in sequestering inorganic phosphorus that might otherwise contribute to eutrophication. This capacity enhances their role as possible indicators of nutrient-depleted conditions, so that the relative abundance of freshwater Actinobacteria could be associated to high water quality standards.

Genes critical for recalcitrant organic matter degradation

Degradation of the lignin heteropolymer is achieved by the action of peroxidases that produce several aromatic compounds, for example benzoate, p-hydroxybenzoate, vanillate, ferulate and coumarate (Buchan *et al.* 2000). Lignin degradation products may also combine with other compounds present in soils, forming humus, a complex mixture of organic compounds (Stevenson 1994). In any case, such compounds may leach into adjoining aquatic systems. Most of the diverse by-products of lignin breakdown (such as those listed above) are ultimately channelled into protocatechuate degradation pathways that may take at least three routes, depending upon the type of protocatechuate dioxygenase employed, which is the key enzyme responsible for the fission of the aromatic ring (Noda *et al.* 1990; Frazee *et al.* 1993; Wolgel *et al.* 1993). In the acMicro-4 genome, we found a cluster of genes coding for a protocatechuate 4, 5 dioxygenase and three additional enzymes in the downstream pathway (Fig. S11, Supporting information). However, the pathway remained incomplete, as the enzymes coding for last two steps, leading towards complete conversion of protocatechuate to pyruvate and oxaloacetate, were not found. This may be because of the incomplete nature of the reconstructed genome (71%). To our knowledge, this is the first report of this key pathway in planktonic freshwater bacteria. However, we did not find any evidence of the existence of lignin peroxidases, the first step in its degradation, in any of our actinobacterial contigs. Therefore, they might rely on other members of the microbial community (planktonic or soil-dwellers) for the production of these enzymes that eventually release soluble compounds. Alternatively, actinobacterial cells could be using the complex and soluble humic substances released from soil. However, it is quite likely that as more genetic diversity of freshwater is uncovered, other organisms that utilize lignin by-products, a biotechnologically relevant activity (Bugg *et al.* 2011), would be discovered.

One genome, acAMD-5, possessed a complete set of genes involved in uptake of inositol and a subsequent complete set genes coding for seven reactions that lead to its conversion to acetyl-CoA and thus entering central metabolism. This suggests that some aquatic Actinobacteria are important natural sinks for inositol. Inositol phosphates (IP) are among the most abundant form of recalcitrant organic phosphorus compounds present in natural environments (Turner *et al.* 2002). These enzymes would also enhance their capacity for phosphate uptake, increasing their performance under oligotrophic conditions. Inositol itself is an essential component of eukaryotic phospholipids (e.g. as phosphatidylinositol) and is also present in high concentrations in fruits and vegetables. Inositol phosphates accumulate in soil and leach into water bodies where they become a potential source of organic phosphorus. Freshwater algae are another source of these compounds in aquatic environments (Eisenreich & Armstrong 1977), and several cyanobacteria have also been shown to be capable of utilizing IP as sole phosphorus sources (Whitton *et al.* 1991).

The acAcidi contigs encoded a sulphonate transporter and an alkanesulphonate monooxygenase which suggests that sulphonates can be used as sulphur sources (Erdlenbruch *et al.* 2001) (Fig. 3). Typically, monooxygenation of alkanesulphonates, producing aldehydes and sulphite, is a common method for their degradation among aerobic bacteria (Kelly & Murrell 1999; Erdlenbruch *et al.* 2001). Sulphonates are widely available compounds in the natural environment (Erdlenbruch *et al.* 2001). A large fraction of organic matter in soils, rivers and marine sediments is sulphonated (Autry & Fitzgerald 1990; Vairavamurthy *et al.* 1994). Methanesulphonate, for example, is one of the primary form of sulphonates in the environment, originating in megatonne amounts from atmospheric dimethylsulfide oxidation (Kelly & Murrell 1999). Another important source of sulphonates is herbicides that can be

widespread and accumulate in freshwater bodies (Boyd 2000). In any case, it certainly appears that some freshwater Actinobacteria are able to capture sulphur from these widely available organosulphur compounds.

Apart from inflows of naturally occurring compounds from the surrounding landscape, runoff from agricultural activity is another important input into freshwater bodies. These may either be additional nutrients (e.g. in the form of fertilizers) or herbicides. Triazine is a widely used chlorinated herbicide, that is a recalcitrant contaminant in both surface and ground water (Buser 1990; Solomon *et al.* 1996). It has been recently banned within the European Union because of its endocrine disruptor and carcinogenic properties (Bethsass & Colangelo 2006). In the acAcidi group, we found two genes encoding for triazine hydrolases, suggesting that these Actinobacteria might be one of the important biotic factors responsible for its degradation.

Distribution of Actinobacteria in freshwater metagenomes

To assess the presence of freshwater Actinobacteria worldwide and to try to correlate their presence to that of other members of the community, we have analysed the 16S rRNA sequences from several publicly available NGS freshwater metagenomic data sets. These data sets have not been subjected to PCR amplification and cloning, and we believe these quantifications are more reliable (Fig. S2, Supporting information). The results

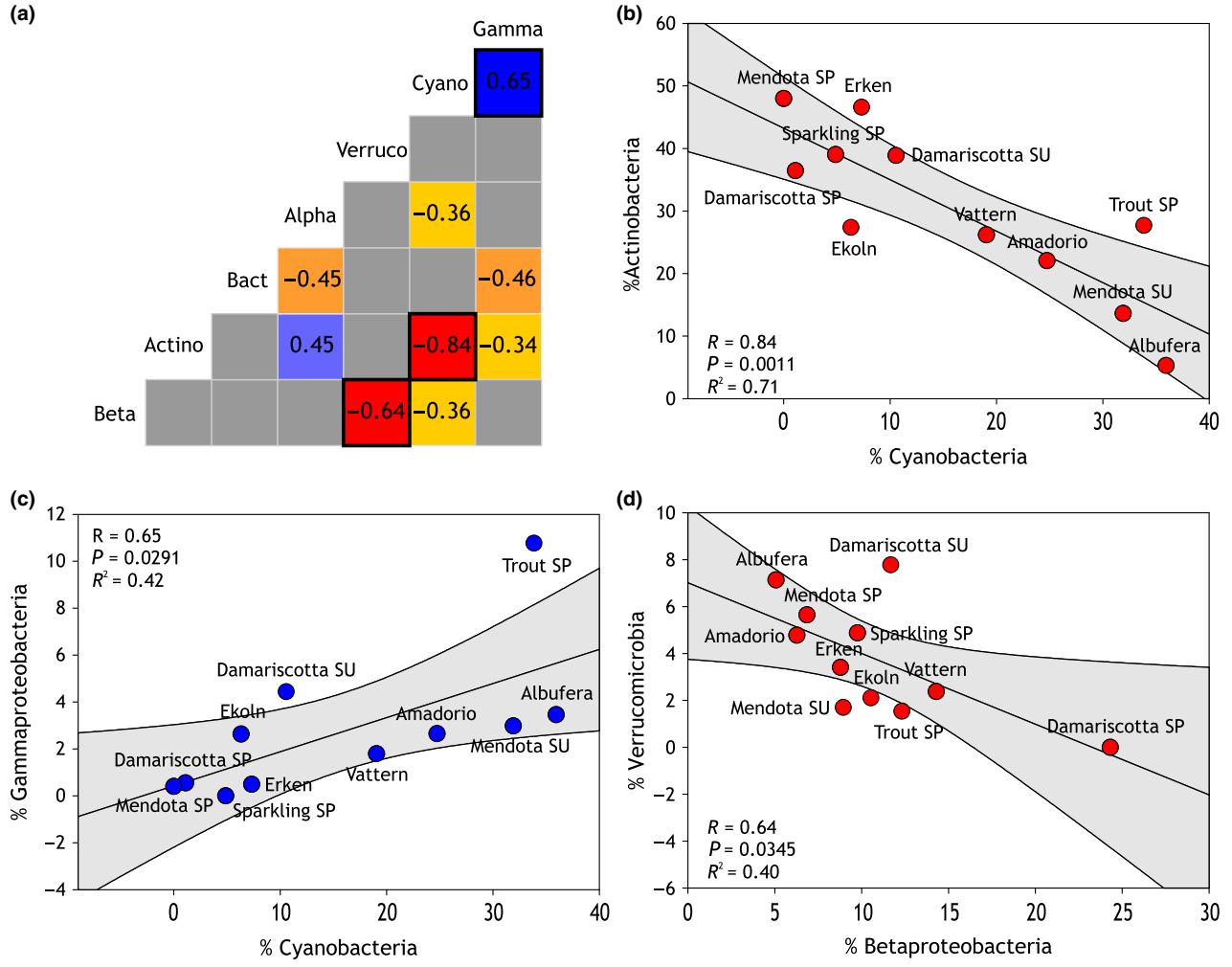

Fig. 4 (a) Correlation matrix of populations of higher taxa across nine metagenomic data sets. Names of taxa are abbreviated (Gamma: Gammaproteobacteria, Cyano: Cyanobacteria, Verruco: Verrucomicrobia, Alpha: Alphaproteobacteria, Bact: Bacteroidetes, Actino:Actinobacteria, Beta: Betaproteobacteria). Pairwise comparisons are colour coded, and only those where $R > \pm 0.3$ are shown. Statistically significant correlations are outlined in bold (b) statistically significantly correlations of % of Cyanobacteria vs. Actinobacteria, (c) % Cyanobacteria vs. %Gammaproteobacteria and (d) %Betaproteobacteria vs. %Verrucomicrobia. The Pearson's correlation coefficient (R), coefficient of determination (R^2) and the P-value of the correlation are provided for b, c and d. The grey area represents the 95% confidence intervals. SU: Summer, SP: Spring.

show that Actinobacteria are among the dominant members of nearly all freshwater data sets, constituting sometimes nearly half of the total community (e.g. see Lake Gatun). In Amadorio, they comprise nearly 35% of the 16S rRNA reads in the small fraction and are underrepresented in the large fraction (approximately 7%), which fits with the general view that freshwater Actinobacteria are mainly small sized. Across all these freshwater data sets, Actinobacteria comprise upwards of 25% of the reads, except for Albufera, which is a highly eutrophic freshwater lagoon, with Cyanobacteria as the dominant microbes (Ghai *et al.* 2012a). We have analysed the correlation between the abundance of the major groups of freshwater microbes in several metagenomic data sets, and the results are shown in Fig. 4. It is important to emphasize that for this analysis, we have compared all prokaryotic groups retrieved by filters so that any microbe larger than 0.1 or 0.22 μm were included (we have either pooled other filter sizes or used metagenomic data obtained without prefilter). This way, we are also counting large cyanobacterial cells or aggregates (filamentous or colonial). There is a strong statistically significant negative correlation between actinobacterial and cyanobacterial populations (P = 0.0011), the latter being indicators of eutrophic conditions. The presence of large numbers of cyanobacterial cells is detrimental to the ecological value of the water body and its use by humans (Paerl & Huisman 2009). The high nutrient loads available in Cyanobacteria dominated waters likely leads to the small actinobacterial cells with more streamlined genomes to be outcompeted. Two more correlations could be detected. A positive correlation between Cyanobacteria and Gammaproteobacteria was found. This is likely because the conditions that positively affect Cyanobacteria affect the frequently opportunistic Gammaproteobacteria as well in a similar fashion. We also found moderate positive correlations between cyanobacterial or gammaproteobacterial populations and temperature (R = 0.58 and R = 0.40, respectively). In the few systems for which data are available (for example Damariscotta, Eiler *et al.* 2013), samples taken in spring have higher density actinobacterial populations that decrease as temperature rises with the onset of summer, coinciding with Cyanobacteria and Gammaproteobacteria becoming dominant community members. In addition, increase in temperature has been shown to have favourable effects on cyanobacterial abundances (Kanoshina *et al.* 2003; Fernald *et al.* 2007; Johnk *et al.* 2008). This together with our data suggests that increasing temperatures, with the concomitant increase in cyanobacterial and gammaproteobacterial populations, has an adverse effect on the actinobacterial community. Another negative correlation between Betaproteobacteria and Verrucomicrobia was also found (Fig. 4), but lack of genomic information regarding these two groups is still an obstacle towards explaining these observations. We also attempted to detect correlations with other physicochemical metadata; however, nothing could be detected. More data sets might be required to shed light on more subtle interactions.

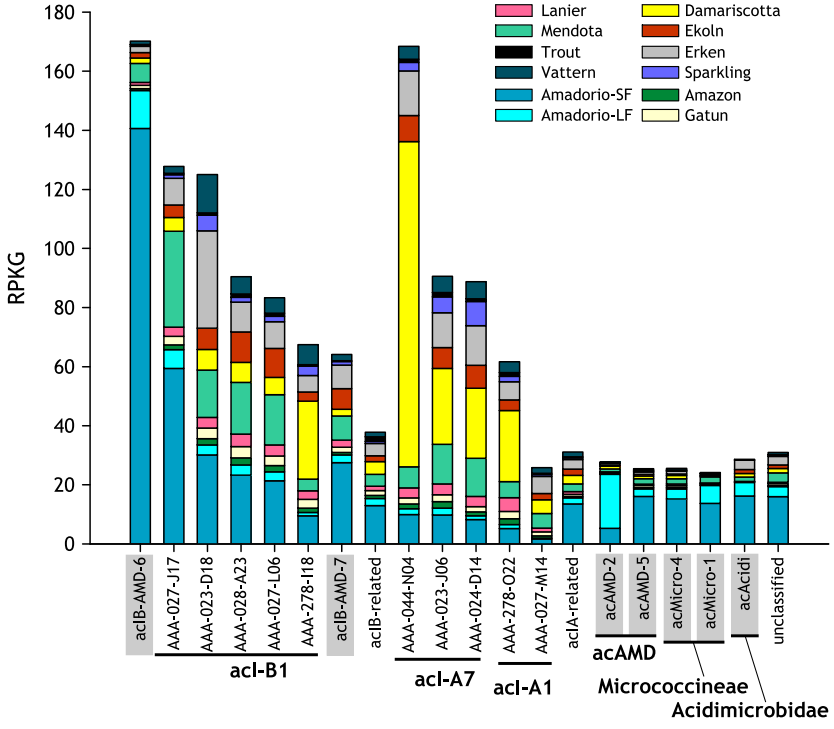

Fig. 5 Fragment recruitment of freshwater Actinobacteria. Abundance of all assembled actinobacterial genomes described in this work (expressed as RPKG, reads per kb of genome per Gb of metagenome), along with all the unclassified actinobacterial contigs described here and publicly available single cell amplified genomes in several freshwater metagenomes. Only hits with ≥95% identity, ≥50-bp alignment length were considered. Genomes are grouped according to phylogeny (acI-B, acI-A, acAMD, Micrococcineae, Acidimicobidae). The genomes assembled in this work are highlighted in grey. SF: small fraction, LF: large fraction.

To examine the distribution of the novel assembled actinobacterial genomes in other freshwater habitats, we performed fragment recruitment with several metagenomes available from lakes, reservoirs and rivers (Rusch *et al.* 2007; Ghai *et al.* 2011b; Oh *et al.* 2011; Eiler *et al.* 2013). We have also included in the analysis all publicly available SAGs of Actinobacteria (Ghylin *et al.* 2014) (Fig. 5). Only genomes obtained from single cell genomes or from metagenomic assembly show comparable abundance levels across metagenomes. The two cultured actinobacterial genomes (*Ca.* Aquiluna and *Rhodoluna*) recruit approximately 10 times less than even the relatively less abundant acMicro clade genomes. The acI-A and acI-B clades seem to be the dominant Actinobacteria in these diverse set of samples (Fig. 5), although all the groups, including the new ones described here for the first time, were in significant numbers at these widespread locations. The fragment recruitment results also indicate some degree of endemism. It appears that similar taxa are present at multiple locations, but the numbers of hits are more at their place of origin. For example, the most abundant acI-B genome in Amadorio (acIB-AMD-6) recruits much more from Amadorio than from any other location. However, the previously described acI-B genome (AAA027-L06), obtained from Lake Mendota, recruits nearly equally from Amadorio and Lake Mendota.

Discussion

By adding more representatives to the available genomes of aquatic Actinobacteria, we are shedding light on their ecological role. Some of these newly described microbes appear to be capable of degrading recalcitrant organic matter in these water bodies. Freshwater is synonymous with continental water, collected by rivers and streams after passing through a filter of vascular plants and soil. In this sense, it can be considered a continuation of soil itself. Furthermore, being at the end of the collecting basin, reservoirs receive an input richer in recalcitrant materials. It is therefore not surprising that (like soil microbes) some aquatic Actinobacteria can degrade the complex mixture of organic matter derived from plant biomass. Their capacity for efficient uptake of nutrients, namely phosphorus (e.g. exopolyphosphatases, inositol acquisition), nitrogen (cyanophycin degradation) and carbon (use of plant xylan, lignin breakdown by-products), enhances the performance of these bacteria under oligotrophic or nutrient-depleted conditions. Additionally, their small size, increasing their surface area-to-volume ratio, allows for efficient nutrient acquisition through high-affinity, broad-specificity uptake systems (Lauro *et al.* 2009).

Lignin is a complex aromatic heteropolymer that forms an important structural component of plant cell walls. After cellulose, it is the second most abundant organic material on the planet (Boerjan *et al.* 2003). It is also among the most recalcitrant natural compounds. Forested catchments, such as those of the Amadorio reservoir, are an important source of these recalcitrant plant compounds that likely are driven to the reservoir through river inlets. Fungi have been considered to be among the most important organisms in the breakdown of lignin, but several soil bacteria including many Actinobacteria have been shown to have the capability to mineralize lignin and lignin-derived fragments (ten Have & Teunissen 2001; Sánchez 2009; Brown & Chang 2014). Apart from soil, lignin degrading organisms have also been identified in cow rumen (Akin & Rigsby 1987) and wood-eating insects (Brune 2014). However, there are no known reports of lignin breakdown from genuinely planktonic freshwater bacteria. The data presented here indicate that at least some freshwater Actinobacteria are likely among the consumers of lignin by-products (e.g. ferulate, vanillate). Interestingly, in a previous metagenomic study on the Amazon river, whose waters are very rich in recalcitrant carbon compounds, some of us discovered that the river microbiota were enriched in pathways for degradation of the same aromatic compounds (Ghai *et al.* 2011b).

The information presented here derived from metagenomes from around the world indicates that Actinobacteria are sensitive to conditions leading to cyanobacterial blooms, usually linked to sustained or pulsed nutrient inputs particularly under high temperatures. It is possible that the highly streamlined actinobacterial cells cannot compete under circumstances of both high organic matter, inorganic nutrient availability and high temperatures under which bloomers such as colonial Cyanobacteria or Gammaproteobacteria flourish. There is a wide range of well-documented effects of increasing numbers of cyanobacteria in freshwater bodies, ranging from oxygen-depletion, fish kills and toxicity responsible for several human health issues (Paerl & Huisman 2009). Freshwater Actinobacteria might serve as sentinel microbes signalling impending ecological damage. Such population shifts could also affect the functional capacity of these ecosystems in dealing with the continuous nutrient input to them. Therefore, these microbes have the potential to become standards of ecological freshwater quality.

Acknowledgements

This work was supported by projects MICROGEN (Programa CONSOLIDER-INGENIO 2010 CSD2009-00006) from the Spanish Ministerio de Ciencia e Innovación, MEDIMAX

BFPU2013-48007-P from the Spanish Ministerio de Economía y Competitividad, MaCuMBA Project 311975 of the European Commission FP7, ACOMP/2014/024, AORG 2014/032 and PROMETEO II/2014/012 from the Generalitat Valenciana. AC and AP are supported by the project CGL2012-38909 (ECO-LAKE) from the Spanish Ministerio de Economia y Competitividad. The authors would like to thank Ana-Belén Martín-Cuadrado for assistance with filtration and DNA extraction.

References

Akin D, Rigsby L (1987) Mixed fungal populations and lignocellulosic tissue degradation in the bovine rumen. *Applied and Environmental Microbiology*, **53**, 1987–1995.

Albertsen M, Hugenholtz P, Skarshewski A *et al.* (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology*, **31**, 533–538.

Altschul SF, Madden TL, Schaffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Autry A, Fitzgerald J (1990) Sulfonate S: a major form of forest soil organic sulfur. *Biology and Fertility of Soils*, **10**, 50–56.

Avakyan A, Iakovleva V (1998) Status of global reservoirs: the position in the late twentieth century. *Lakes & Reservoirs: Research & Management*, **3**, 45–52.

Aziz RK, Bartels D, Best AA *et al.* (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, **9**, 75.

Berggren M, Laudon H, Haei M, Ström L, Jansson M (2010) Efficient aquatic bacterial metabolism of dissolved low-molecular-weight compounds from terrestrial sources. *The ISME Journal*, **4**, 408–416.

Bethsass J, Colangelo A (2006) European Union bans atrazine, while the United States negotiates continued use. *International Journal of Occupational and Environmental Health*, **12**, 260–267.

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. *Annual Review of Plant Biology*, **54**, 519–546.

Bouvier T, Troussellier M, Anzil A, Courties C, Servais P (2001) Using light scatter signal to estimate bacterial biovolume by flow cytometry. *Cytometry*, **44**, 188–194.

Boyd R (2000) Herbicides and herbicide degradates in shallow groundwater and the Cedar River near a municipal well field, Cedar Rapids, Iowa. *Science of the Total Environment*, **248**, 241–253.

Brown ME, Chang MC (2014) Exploring bacterial lignin degradation. *Current Opinion in Chemical Biology*, **19**, 1–7.

Brune A (2014) Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology*, **12**, 168–180.

Buchan A, Collier LS, Neidle EL, Moran MA (2000) Key aromatic-ring-cleaving enzyme, protocatechuate 3, 4-dioxygenase, in the ecologically important marineRoseobacter lineage. *Applied and Environmental Microbiology*, **66**, 4662–4672.

Bugg TDH, Ahmad M, Hardiman EM, Singh R (2011) The emerging role for bacteria in lignin degradation and bio-product formation. *Current Opinion in Biotechnology*, **22**, 394–400.

Buser H-R (1990) Atrazine and other s-triazine herbicides in lakes and in rain in Switzerland. *Environmental Science & Technology*, **24**, 1049–1058.

Camacho A, Battle M, Correcher E *et al.* (2013) *New technologies for Water Quality Monitoring in Dams: Limnological Contrast and Calibration in Mediterranean Reservoirs*. Monograph of the Polish Institute of Meteorology and Water Management, National Research Institute, Warszawa, Poland.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.

Cole JR, Wang Q, Cardenas E *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, **37**, D141–D145.

Dasi MJ, Miracle MA, Camacho A, Soria JM, Vicente E (1998) Summer phytoplankton assemblages across trophic gradients in hard-water reservoirs. *Hydrobiologia*, **369/370**, 27–43.

Dasí M, Miracle M, Camacho A, Soria J, Vicente E (1998) Summer phytoplankton assemblages across trophic gradients in hard-water reservoirs. In: *Phytoplankton and Trophic Gradients*, pp. 27–43. Springer, Belgium.

Dodds WK (2002) *Freshwater Ecology: Concepts and Environmental Applications*. Academic Press, San Diego, California.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

Eiler A, Zaremba-Niedzwiedzka K, Martínez-García M *et al.* (2013) Productivity and salinity structuring of the microplankton revealed by comparative freshwater metagenomics. *Environmental Microbiology*, **16**, 2682–2698.

Eisenreich SJ, Armstrong DE (1977) Chromatographic investigation of inositol phosphate esters in lake waters. *Environmental Science & Technology*, **11**, 497–501.

Erdlenbruch BN, Kelly DP, Murrell CJ (2001) Alkanesulfonate degradation by novel strains of *Achromobacter xylosoxidans*, *Tsukamurella wratislaviensis* and *Rhodococcus* sp., and evidence for an ethanesulfonate monooxygenase in *A. xylosoxidans* strain AE4. *Archives of Microbiology*, **176**, 406–414.

Fernald SH, Caraco NF, Cole JJ (2007) Changes in cyanobacterial dominance following the invasion of the zebra musselDreissena polymorpha: long-term results from the Hudson River estuary. *Estuaries and Coasts*, **30**, 163–170.

Frazee RW, Livingston D, LaPorte D, Lipscomb J (1993) Cloning, sequencing, and expression of the *Pseudomonas putida* protocatechuate 3, 4-dioxygenase genes. *Journal of Bacteriology*, **175**, 6194–6202.

Fujinami S, Takarada H, Kasai H *et al.* (2013) Complete genome sequence of Ilumatobacter coccineum YM16-304 T. *Standards in Genomic Sciences*, **8**, 3.

Garcia SL, McMahon KD, Martinez-Garcia M *et al.* (2012) Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *The ISME Journal*, **7**, 137–147.

Garcia SL, McMahon KD, Grossart HP, Warnecke F (2013) Successful enrichment of the ubiquitous freshwater acI Actinobacteria. *Environmental Microbiology Reports*, **6**, 21–27.

Ghai R, Martin-Cuadrado A-B, Molto AG *et al.* (2010) Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *The ISME Journal*, **4**, 1154–1166.

Ghai R, Pasic L, Fernandez AB *et al.* (2011a) New abundant microbial groups in aquatic hypersaline environments. *Scientific Reports*, **1**, 135.

Ghai R, Rodriguez-Valera F, McMahon KD *et al.* (2011b) Metagenomics of the water column in the pristine upper course of the Amazon river. *PLoS One*, **6**, e23785.

Ghai R, Hernandez CM, Picazo A *et al.* (2012a) Metagenomes of Mediterranean coastal lagoons. *Scientific Reports*, **2**, 490.

Ghai R, McMahon KD, Rodriguez-Valera F (2012b) Breaking a paradigm: cosmopolitan and abundant freshwater actinobacteria are low GC. *Environmental Microbiology Reports*, **4**, 29–35.

Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F (2013) Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Scientific Reports*, **3**, 2471.

Ghylin TW, Garcia SL, Moya F *et al.* (2014) Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. *The ISME Journal*, doi: 10.1038/ismej.2014.135.

Glockner FO, Zaichikov E, Belkova N *et al.* (2000) Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. *Applied and Environment Microbiology*, **66**, 5053–5065.

Greenberg A, Clesceri L, Eaton A (1992) *Standard Methods for the Examination of Water and Wastewater*, **18**. American Public Health Association, Washington, District of Columbia.

Gushchin I, Chervakov P, Kuzmichev P *et al.* (2013) Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. *Proceedings of the National Academy of Sciences*, **110**, 12631–12636.

Haft DH, Loftus BJ, Richardson DL *et al.* (2001) TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Research*, **29**, 41–43.

Hahn MW (2003) Isolation of strains belonging to the cosmopolitan *Polynucleobacter necessarius* cluster from freshwater habitats located in three climatic zones. *Applied and Environment Microbiology*, **69**, 5248–5254.

Hahn MW (2009) Description of seven candidate species affiliated with the phylum Actinobacteria, representing planktonic freshwater bacteria. *International Journal of Systematic and Evolutionary Microbiology*, **59**, 112–117.

Hahn MW, Lunsdorf H, Wu Q *et al.* (2003) Isolation of novel ultramicrobacteria classified as actinobacteria from five freshwater habitats in Europe and Asia. *Applied and Environment Microbiology*, **69**, 1442–1451.

Hahn MW, Stadler P, Wu QL, Pöckl M (2004) The filtration–acclimatization method for isolation of an important fraction of the not readily cultivable bacteria. *Journal of Microbiological Methods*, **57**, 379–390.

Hahn M, Schmidt J, Taipale SJ, Doolittle WF, Koll U (2014) Rhodoluna lacicola gen. nov., sp. nov., a planktonic freshwater bacterium with stream-lined genome. *International Journal of Systematic and Evolutionary Microbiology*, **64**, 3254–3263.

ten Have R, Teunissen PJ (2001) Oxidative mechanisms involved in lignin degradation by white-rot fungi. *Chemical Reviews*, **101**, 3397–3414.

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 *et al.* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**, 119.

Jezbera J, Sharma AK, Brandt U, Doolittle WF, Hahn MW (2009) 'Candidatus Planktophila limnetica', an actinobacterium representing one of the most numerically important taxa in freshwater bacterioplankton. *International Journal of Systematic and Evolutionary Microbiology*, **59**, 2864–2869.

Johnk KD, Huisman JEF, Sharples J *et al.* (2008) Summer heatwaves promote blooms of harmful cyanobacteria. *Global Change Biology*, **14**, 495–512.

Kang I, Lee K, Yang S-J *et al.* (2012) Genome sequence of "Candidatus Aquiluna" sp. strain IMCC13023, a marine member of the actinobacteria isolated from an arctic fjord. *Journal of Bacteriology*, **194**, 3550–3551.

Kanoshina I, Lips U, Leppänen J-M (2003) The influence of weather conditions (temperature and wind) on cyanobacterial bloom development in the Gulf of Finland (Baltic Sea). *Harmful Algae*, **2**, 29–41.

Kawakoshi A, Nakazawa H, Fukada J *et al.* (2012) Deciphering the genome of polyphosphate accumulating actinobacterium *Microlunatus phosphovorus*. *DNA Research*, **19**, 383–394.

Kelly DP, Murrell JC (1999) Microbial metabolism of methanesulfonic acid. *Archives of Microbiology*, **172**, 341–348.

King GM, Weber CF (2007) Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nature Reviews Microbiology*, **5**, 107–118.

Lassmann T, Sonnhammer EL (2005) Kalign–an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics*, **6**, 298.

Lauro FM, McDougald D, Thomas T *et al.* (2009) The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences*, **106**, 15527–15533.

Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software*, **25**, 1–18.

Lehninger A, Nelson DL, Cox MM (2005) *Lehninger's Principles of Biochemistry*. WH Freeman and Company, New York.

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, **25**, 0955–0964.

MacIsaac E, Stockner JG (1993) Enumeration of phototrophic picoplankton by autofluorescence microscopy. In: *Handbook of Methods in Aquatic Microbial Ecology*, pp. 187–197. Lewis Publishers, Boca Raton, Florida.

Marie D, Partensky F, Jacquet S, Vaulot D (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Applied and Environmental Microbiology*, **63**, 186–193.

Martin-Cuadrado A-B, Ghai R, Gonzaga A, Rodriguez-Valera F (2009) CO dehydrogenase genes found in metagenomic fosmid clones from the deep Mediterranean Sea. *Applied and Environmental Microbiology*, **75**, 7436–7444.

Martínez-Gómez K, Flores N, Castañeda HM *et al.* (2012) New insights into Escherichia coli metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. *Microbial Cell Factories*, **11**, 46.

Matsumoto A, Kasai H, Matsuo Y *et al.* (2009) Ilumatobacter fluminis gen. nov., sp. nov., a novel actinobacterium isolated from the sediment of an estuary. *The Journal of General and Applied Microbiology*, **55**, 201–205.

Moran MA, Buchan A, Gonzalez JM *et al.* (2004) Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature*, **432**, 910–913.

Murarka A, Dharmadi Y, Yazdani SS, Gonzalez R (2008) Fermentative utilization of glycerol by *Escherichia coli* and its implications for the production of fuels and chemicals. *Applied and Environmental Microbiology*, **74**, 1124–1135.

Nawrocki EP (2009) *Structural RNA Homology Search and Alignment using Covariance Models*. Washington University, St. Louis, Missouri.

Nawrocki EP, Eddy SR (2010) ssu-align: a tool for structural alignment of SSU rRNA sequences.

Newton RJ, Jones SE, Helmus MR, McMahon KD (2007) Phylogenetic ecology of the freshwater Actinobacteria acI lineage. *Applied and Environment Microbiology*, **73**, 7169–7176.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011) A guide to the natural history of freshwater Lake Bacteria. *Microbiology and Molecular Biology Reviews*, **75**, 14–49.

Noda Y, Nishikawa S, Shiozuka K *et al.* (1990) Molecular cloning of the protocatechuate 4, 5-dioxygenase genes of *Pseudomonas paucimobilis*. *Journal of Bacteriology*, **172**, 2704–2709.

Oh S, Caro-Quintero A, Tsementzi D *et al.* (2011) Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Applied and Environment Microbiology*, **77**, 6000–6011.

Paerl HW, Huisman J (2009) Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environmental Microbiology Reports*, **1**, 27–37.

Peng Y, Leung HC, Yiu S-M, Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, **28**, 1420–1428.

Pernthaler J (2013) Freshwater microbial communities. In: *The Prokaryotes*, pp. 97–112. Springer, Heidelberg.

Picazo A, Rochera C, Vicente E, Miracle MR, Camacho A (2013) 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*, **32**, 139–158.

Porter K, Feig YS (1980) The use of DAPI for identifying and counting aquatic microflora. *Limnology and Oceanography*, **25**, 943–948.

Price MN, Dehal PS, Arkin AP (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. *Trends in Genetics*, **16**, 276–277.

Rinke C, Schwientek P, Sczyrba A *et al.* (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, **499**, 431–437.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*, **132**, 365–386.

Rusch DB, Halpern AL, Sutton G *et al.* (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biology*, **5**, e77.

Sánchez C (2009) Lignocellulosic residues: biodegradation and bioconversion by fungi. *Biotechnology Advances*, **27**, 185–194.

Sekar R, Pernthaler A, Pernthaler J *et al.* (2003) An improved protocol for quantification of freshwater Actinobacteria by fluorescence in situ hybridization. *Applied and Environmental Microbiology*, **69**, 2928–2935.

Sharma AK, Spudich JL, Doolittle WF (2006) Microbial rhodopsins: functional versatility and genetic mobility. *Trends in Microbiology*, **14**, 463–469.

Sharma AK, Zhaxybayeva O, Papke RT, Doolittle WF (2008) Actinorhodopsins: proteorhodopsin-like gene sequences found predominantly in non-marine environments. *Environmental Microbiology*, **10**, 1039–1056.

Sharma AK, Sommerfeld K, Bullerjahn GS *et al.* (2009) Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria. *The ISME Journal*, **3**, 726–737.

Shiklomanov IA, Rodda JC (2003) *World Water Resources at the Beginning of the Twenty-First Century*. Cambridge University Press, Cambridge.

Solomon KR, Baker DB, Richards RP *et al.* (1996) Ecological risk assesment of atrazine in North American surface waters. *Environmental Toxicology and Chemistry*, **15**, 31–76.

Stevenson FJ (1994) *Humus Chemistry: Genesis, Composition, Reactions*. John Wiley & Sons, USA.

Tatusov RL, Natale DA, Garkavtsev IV *et al.* (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research*, **29**, 22–28.

Tolli JD, Taylor CD (2005) Biological CO oxidation in the Sargasso Sea and in Vineyard Sound, Massachusetts.

Turner BL, Paphazy MJ, Haygarth PM, McKelvie ID (2002) Inositol phosphates in the environment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **357**, 449–469.

Vairavamurthy A, Zhou W, Eglinton T, Manowitz B (1994) Sulfonates: a novel class of organic sulfur compounds in marine sediments. *Geochimica et Cosmochimica Acta*, **58**, 4681–4687.

Vicente E, Camacho A, Sendra M *et al.* (2000) Limnological management of the Amadorio Reservoir (Spain) during an extremely dry summer. *Verhandlungen des Internationalen Verein Limnologie*, **27**, 2298–2302.

Whitton BA, Grainger SLJ, Hawley GRW, Simon JW (1991) Cell bound and extracellular phosphatase activities of cyanobacterial isolates. *Microbial Ecology*, **21**, 85–98.

Wolgel S, Dege J, Perkins-Olson P *et al.* (1993) Purification and characterization of protocatechuate 2, 3-dioxygenase from Bacillus macerans: a new extradiol catecholic dioxygenase. *Journal of Bacteriology*, **175**, 4414–4426.

Yannarell AC, Triplett EW (2004) Within-and between-lake variability in the composition of bacterioplankton communities: investigations using multiple spatial scales. *Applied and Environmental Microbiology*, **70**, 214–223.

Zwart G, Crump BC, Agterveld MPKV, Hagen F, Han SK (2002) Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquatic Microbial Ecology*, **28**, 141–155.

R.G. and F.R.V. conceived the study. All authors performed the sample collection and filtration. A.P. and A.C. analysed the sample properties and performed the FISH and flow cytometric analyses. C.M.M. performed the DNA extractions. R.G. and C.M.M. analysed the metagenomic data. R.G., A.C. and F.R.V. wrote the manuscript with feedback from all other authors. All authors read and approved the final manuscript.

Data accessibility

The metagenomic data have been submitted to NCBI SRA and are accessible under the Bioproject identifier PRJNA238866. The assembled genome sequences have been deposited to NCBI GenBank and can be accessed using the accession nos JNSC00000000, JNSD00000000, JNSE00000000, JNSF00000000, JNSG00000000, JNSH00000000, JNSI00000000, JNSJ00000000, JNSK00000000, JNSL00000000. The raw data files for flow cytometry, fluorescence in situ hybridization, phylogenetic trees and 16S rRNA alignments have been submitted to Dryad (http://www.datadryad.org) and are available at http://dx.doi.org/10.5061/dryad.dr216.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Main environmental variables of the Amadorio reservoir corresponding to the sampled depth (10 m) on February 1st, 2012, during the mixing period.

Table S2 Key residues in diverse rhodopsin sequences.

Fig. S1 Vertical profiles of the main environmental variables in the Amadorio reservoir at the date of sampling (February 1st, 2012).

Fig. S2 16S rRNA read classifications of the Amadorio freshwater reservoir and several other freshwater metagenomic datasets.

Fig. S3 Genome size vs. GC content.

Fig. S4 Genomic GC content vs. median intergenic spacer content.

Fig. S5 16S rRNA phylogeny using maximum likelihood.

Fig. S6 Microscopic FISH image of samples from Amadorio reservoir.

Fig. S7 Abundance and size (maximum cell dimension) structure of the bacterioplankton community in the Amadorio reservoir as determined by flow cytometry.

Fig. S8 Metabolic overview of the freshwater Actinobacteria acAcidi group.

Fig. S9 Metabolic overview of the freshwater Actinobacteria acAMD-5.

Fig. S10 Metabolic overview of the freshwater Actinobacteria acMicro-4.

Fig. S11 (a) Gene cluster encoding genes for the protocatechuate degradation pathway in acMicro-1 genome. (b) The biochemical pathway of protocatechuate degradation.

Data S1 (A) Metabolic pathway comparison. (B) Transporter comparison. (C) Carbohydrate-active enzyme comparison. (D) Actinobacterial core.