

ORIGINAL ARTICLE

Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria

Stefan M Neuenschwander^{1,3}, Rohit Ghai², Jakob Pernthaler¹ and Michaela M Salcher^{1,2,3}

¹Limnological Station, Institute of Microbial and Plant Biology, University of Zurich, Zurich, Switzerland and

²Institute of Hydrobiology, Biology Centre CAS, Ceske Budejovice, Czech Republic

Actinobacteria of the acI lineage are the most abundant microbes in freshwater systems, but there are so far no pure living cultures of these organisms, possibly because of metabolic dependencies on other microbes. This, in turn, has hampered an in-depth assessment of the genomic basis for their success in the environment. Here we present genomes from 16 axenic cultures of acI Actinobacteria. The isolates were not only of minute cell size, but also among the most streamlined free-living microbes, with extremely small genome sizes (1.2–1.4 Mbp) and low genomic GC content. Genome reduction in these bacteria might have led to auxotrophy for various vitamins, amino acids and reduced sulphur sources, thus creating dependencies to co-occurring organisms (the ‘Black Queen’ hypothesis). Genome analyses, moreover, revealed a surprising degree of inter- and intraspecific diversity in metabolic pathways, especially of carbohydrate transport and metabolism, and mainly encoded in genomic islands. The striking genotype microdiversification of acI Actinobacteria might explain their global success in highly dynamic freshwater environments with complex seasonal patterns of allochthonous and autochthonous carbon sources. We propose a new order within Actinobacteria (‘*Candidatus Nanopelagicales*’) with two new genera (‘*Candidatus Nanopelagicus*’ and ‘*Candidatus Planktophila*’) and nine new species.

The ISME Journal (2018) 12, 185–198; doi:10.1038/ismej.2017.156; published online 13 October 2017

Introduction

The pelagic zone of oceans and fresh waters is numerically dominated by microbes with cell volumes $<0.1 \mu\text{m}^3$. The genomes of such oligotrophic ‘ultramicrobacteria’ are considered to be ‘streamlined’, that is, <1.5 Mbp, with highly conserved core-genomes, few pseudogenes, low ratios of intergenic spacer DNA to coding DNA, reduced numbers of paralogues and a low genomic GC content (Giovannoni *et al.*, 2014; Luo *et al.*, 2015). Streamlining theory proposed that gene loss reflects an evolutionary adaptation to the competition for limiting resources under oligotrophic conditions (Giovannoni *et al.*, 2014). However, genome reduction might also lead to metabolic dependencies on co-occurring microorganisms, as outlined by the ‘Black Queen Hypothesis’ (Morris *et al.*, 2012), possibly explaining why few taxa with this lifestyle have so far been isolated.

Actinobacteria of the acI lineage are among the most successful bacteria in lacustrine waters where

they may constitute $>50\%$ of all microbes (Newton *et al.*, 2011). They are distributed ubiquitously in a wide range of freshwater habitats of different trophic states and varying pH, oxygen and salinity levels (Newton *et al.*, 2007, 2011). There are no stable pure cultures of acI Actinobacteria that has been explained by a high degree of metabolic interconnectedness and dependencies on co-occurring microbes (Garcia *et al.*, 2015). A strain from this lineage in a mixed enrichment culture has been tentatively described as ‘*Ca. Planktophila limnetica*’ (Jezbera *et al.*, 2009). Cultivation-independent methods and *in situ* experiments indicate slower growth rates and lower competitiveness of acI Actinobacteria than copiotrophic bacteria under elevated nutrient and substrate regimes (Burkert *et al.*, 2003; Neuenschwander *et al.*, 2015), but also proposed a reduced vulnerability to bacterivorous protists (Šimek *et al.*, 2014). Co-occurring phytoplankton blooms (Eckert *et al.*, 2012), and also allochthonous dissolved organic matter from terrestrial sources (Pérez and Sommaruga, 2006), are likely providers of carbohydrates and other carbon sources for acI Actinobacteria (Buck *et al.*, 2009; Salcher *et al.*, 2013).

So far, 4 complete genomes of transient cultures (Kang *et al.*, 2017), 10 single-cell amplified genomes (SAGs; Ghylin *et al.*, 2014) and several metagenome assembled genomes (MAGs; Ghai *et al.*, 2014; Garcia *et al.*, 2015; Bendall *et al.*, 2016) provided first

Correspondence: MM Salcher, Limnological Station, Institute of Plant and Microbial Biology, University of Zurich, Seestrasse 187, Kilchberg, Zurich 8802, Switzerland.
E-mail: michaelasalcher@gmail.com

³These authors contributed equally to this work.

Received 9 January 2017; revised 3 April 2017; accepted 28 June 2017; published online 13 October 2017

insights into the metabolic potential of different acI lineages. However, the low number of complete genomes as well as the intrinsic incompleteness of SAGs and the mixed genotype origin of MAGs limits their usefulness for comparative population genomics and for understanding the metabolic distinctions within sympatric ecotypes.

Here, we report on 16 complete high-quality genomes of freshwater acI Actinobacteria from monoclonal cultures that were obtained from a single site, Lake Zurich, Switzerland. All but one strain were isolated in spring in two consecutive years, thus representing sympatric genotypes. Genome analysis suggests a clear split into two related, but distinct genera within the acI lineage. Our data set allows for a first comparative genomic analysis of representatives from two closely related genera on a (sub)species level and, moreover, provides first insights into diversification and potential niche separation within co-occurring acI populations. We also describe a clear temporal separation between the two genera in Lake Zurich, and we assess the global occurrence of the studied genotypes in all currently available freshwater metagenomes.

Materials and methods

Sampling of Lake Zurich, Switzerland

Lake Zurich is a large, deep (136 m), prealpine, oligomesotrophic lake that is characterized by persistent annual blooms of the toxic cyanobacterium *Planktothrix rubescens* (Posch *et al.*, 2012). Sampling was conducted every second week in 2012 and 2013 ($n=37$) at the deepest part of the lake (47.30N, 8.58E). Vertical profiles of temperature, conductivity, turbidity, oxygen and chlorophyll *a* content were recorded with a YSI multiprobe (Yellow Springs Instruments, model 6600, Yellow Springs, OH, USA) and a bbe fluoroprobe (TS-16-12, bbe Moldaenke GmbH, Schwentimental, Germany) calibrated to distinguish between different phytoplankton groups and *P. rubescens* according to the fluorescent spectra of their pigments (Beutler *et al.*, 2002). Water samples were taken from 0, 5, 10, 20, 30, 40, 60, 80 and 100 m depth, and additionally from the depth of recorded chlorophyll *a* maxima ($n=348$). Then, 40 ml water was fixed with formaldehyde (2% final concentration) for estimating prokaryotic abundances via flow cytometry and stored at 4 °C until measurements. Next, 5–10 ml was fixed with paraformaldehyde (pH 7.4, 2% final concentration) for 1–2 h, filtered onto white polycarbonate filters (0.22 µm pore size, 47 mm diameter, Millipore, Billerica, MA, USA) and stored at –20 °C for CARD-FISH (fluorescence *in situ* hybridization followed by catalysed reporter deposition) analyses. Two litres of water from 5 m depth for the preparation of media and the isolation of Actinobacteria was transported to the laboratory within 15 min. Chemical parameters were determined at monthly intervals by

standard techniques by the Zurich Water Supply Company.

Isolation of planktonic Actinobacteria

Planktonic Actinobacteria were isolated as part of a dilution-to-extinction isolation campaign during 2010–2013 (Salcher *et al.*, 2015). Filtered (0.2 µm pore-sized polysulfone filters, Millipore) and autoclaved lake water amended with different types of carbon and nutrient sources served as medium (Supplementary Table S1) and prefiltered water (0.45 µm pore-sized polycarbonate filters, Millipore) served as inoculum. The final inoculum size ranged between 1 and 10 cells per well of 24-well plates (Thermo Scientific, Waltham, MA, USA) filled with 2 ml of medium. Growth in individual wells was checked microscopically after 2–3 months of incubation at close to *in situ* temperature in a dark/light cycle (16:8 h). Dense cultures (that is, $>10^5$ – 10^6 cells per ml) were split in 500 µl aliquots that were (1) propagated to fresh medium, (2) amended with glycerol (30% final concentration) and stored at –80 °C, and (3) pelleted by centrifugation (30 min, 16 000 g) and used for PCR and 16S rRNA sequencing as well as whole-genome amplification and genome sequencing. For more details on the 46 isolation campaigns see Salcher *et al.* (2015). Sizing of the 16 strains was done with a Zeiss AxioImagerM.1 microscope (Zeiss, Oberkochen, Germany) after 4',6-diamidino-2-phenylindole (DAPI) staining, using the software LUCIA (Laboratory Imaging, Prague, Czech Republic) according to a previously published workflow (Posch *et al.*, 2009).

Whole-genome amplification and sequencing of isolates

The REPLI-g single cell kit (Qiagen, Venlo, The Netherlands) was used for multiple displacement amplification (MDA). All preamplification steps were performed in a particle-free environment dedicated to MDA. Fresh PCR-clean pipet tips were used for each MDA session and reaction tubes and PCR plates were ultraviolet treated before usage. MDA was conducted according to the manufacturer's protocol with the following modifications: lysates for 6 to 8 replicate MDA reactions were produced in a 0.5 ml reaction tube (8 µl samples containing ~4000–40 000 cells, 6 µl each of the reagents D2 and stop) and subsequently distributed (2.5 µl each) among wells of 96-well plates each containing 10 µl MDA reaction mix and SYBR I green (0.2 × final concentration). Amplification was conducted in a CFX96 (Bio-Rad, Hercules, CA, USA) at 30 °C with fluorescence measurements every 6 min until stationary phase was reached, followed by heat inactivation at 60 °C for 3 min.

All MDA products were screened by 16S rDNA PCR amplification and Sanger sequencing. PCR was conducted with GoTaq green Master Mix (Promega,

Madison, WI, USA) and the primers GM1f and GM3r (Muyzer *et al.*, 1993, 1995) according to the manufacturer's protocol. PCR products were purified with a GenElute PCR Cleanup Kit (Sigma-Aldrich, St Louis, MO, USA) and sequenced with BigDye v 3.1 and the primer GM1f on an ABI 3730 Genetic Analyzer (Applied Biosystems, Waltham, MA, USA). MDA products that resulted in clean sequences were used to construct 550-bp libraries with the TruSeq PCR-Free LT Sample Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. Paired-end sequences (2 × 150 bp or 2 × 250 bp) were generated on an Illumina MiSeq instrument (for details see Supplementary Table S2). Raw reads have been deposited to The National Center for Biotechnology Information (NCBI; BioProject PRJNA327031).

Genome assembly and annotation

Adapter sequences and Illumina sequencing artefacts were removed with Trimmomatic (Bolger *et al.*, 2014) and Prinseq-lite (Schmieder and Edwards 2011). MDA-derived coverage peaks were removed with kmernorm (<https://sourceforge.net/projects/kmernorm>). Assembly was conducted with SPAdes (Bankevich *et al.*, 2012) or A5-miseq (Tritt *et al.*, 2012; Coil *et al.*, 2014) and different software combinations were compared for each genome and best assemblies were selected (Supplementary Table S2). Reads were subsequently mapped to the resulting assemblies with Bowtie2 (version 2.1.0; Langmead and Salzberg, 2012) in order to identify potential assembly errors. Regions containing low coverage (≤ 10 fold), ambiguities or anomalies in the mapping were verified by PCR and Sanger sequencing. A total of 350 primers were designed with Primer3 (version 2.3.4; Untergasser *et al.*, 2012) in Geneious 7 and 9 (www.geneious.com) and used for producing 268 additional Sanger sequences as described above.

All genomes were annotated with the NCBI microbial genome annotation pipeline and BlastKOALA (Kanehisa *et al.*, 2016) and inconsistencies between the two methods and hypothetical proteins manually curated with NCBI CDD (Marchler-Bauer *et al.*, 2015). Pathways were reconstructed by using BlastKOALA, the KEGG (Kyoto Encyclopedia of Genes and Genomes) Mapper tool (<http://www.kegg.jp/kegg/mapper.html>) and by manually inspecting KEGG maps. Transporters were identified based on KEGG BlastKOALA annotations.

The average nucleotide identity and average amino acid identity between strains and other Actinobacteria was calculated according to previously published algorithms (Goris *et al.*, 2007; Rodriguez-R and Konstantinidis, 2014), and the percentages of conserved proteins were computed via all-vs-all comparisons ($> 50\%$ identity with a coverage of $> 50\%$) following the method described in Qin *et al.* (2014). The core- and pan-genomes were

computed using all-vs-all comparisons of all proteins for each genome using BLASTP. A cutoff of 50% identity and 50% coverage was used to define an orthologue.

The annotated genome sequences have been deposited in GenBank under accession numbers CP016768–CP016783 (BioProject PRJNA327031).

Genomic and single gene/protein tree reconstructions

Phylogenomic trees of Actinobacteria were constructed by identifying conserved proteins on the basis of COG (Clusters of Orthologous Groups of proteins) annotations (Tatusov *et al.*, 2001). Proteins were concatenated and aligned using Kalign (Lassmann and Sonnhammer, 2005), and maximum likelihood trees (100 bootstraps) were constructed with FastTree2 (Price *et al.*, 2010). Three different trees with a different number of genomes were computed: (1) our 16 genomes and > 100 complete genomes affiliated to all orders of Actinobacteria (Sen *et al.*, 2014), with *Staphylococcus aureus* and *Listeria monocytogenes* used as outgroup (48 concatenated proteins); (2) our genomes, other complete genomes (Kang *et al.*, 2017) as well as SAGs (Ghylin *et al.*, 2014) (only those with an estimated completeness of $> 70\%$) and MAGs (Ghai *et al.*, 2014) affiliated to 'Ca. Nanopelagicales', with *Rhodoluna lacicola* MWH-Ta8 (Hahn *et al.*, 2014) and 'Ca. Aquiluna sp.' IMCC13023 (Kang *et al.*, 2012) used as outgroup (87 concatenated proteins); and (3) for a fine-scale analysis of branching patterns only complete genomes with *R. lacicola* MWH-Ta8 (Hahn *et al.*, 2014) and 'Ca. Aquiluna sp.' IMCC13023 (Kang *et al.*, 2012) used as outgroup (462 concatenated proteins).

Genes encoding 16S and 23S rRNA were aligned with the SINA web aligner (<https://www.arb-silva.de/aligner/>), imported to ARB (Ludwig *et al.*, 2004) using the SILVA database SSU Ref 123 or LSU Ref 123 (Pruesse *et al.*, 2007), manually checked, and bootstrapped maximum likelihood trees (GTR-GAMMA model, 100 bootstraps) were calculated with RAxML (Stamatakis *et al.*, 2005). *R. lacicola* MWH-Ta8 (Hahn *et al.*, 2014) and 'Ca. Aquiluna sp.' IMCC13023 (Kang *et al.*, 2012) served as outgroup. Protein sequences of rhodopsins were aligned with Kalign (Lassmann and Sonnhammer, 2005) and a maximum likelihood tree (100 bootstraps) was constructed with FastTree2 (Price *et al.*, 2010). Xanthorhodopsin sequences of *Salinibacter ruber* and *Thermus aquaticus* were used as outgroup.

Design and application of novel specific 23S rDNA probes for actinobacterial lineages

CARD-FISH with fluorescein-labelled tyramides was conducted as previously described (Sekar *et al.*, 2003) with a probe specific for the whole acI lineage of Actinobacteria (AcI-852; Warnecke *et al.*, 2005) and two novel probes targeting 'Ca. Nanopelagicus'

(sublineages acI-B1) and ‘*Ca. Planktophila vernalis*’ (sublineage acI-A7; see Supplementary Table S3 for details). Probe design was based on 23S rRNA genes extracted from the genomes and additional publically available sequences. Sequences were aligned with the SINA web aligner (<https://www.arb-silva.de/aligner/>), imported to ARB (Ludwig *et al.*, 2004) using the SILVA database LSU Ref 123 (Pruesse *et al.*, 2007) and manually checked. A bootstrapped maximum likelihood tree (GTR-GAMMA model) of 23S rDNA sequences (Supplementary Figure S10) served as backbone for probe design with the ARB tools probe_design and probe_check. The resulting probes Npel-23S-2669 and Pver-23S-1420 with their corresponding competitor and helper oligonucleotides (Supplementary Table S6) were tested with different formamide concentrations to achieve stringent hybridization conditions. CARD-FISH-stained samples were analysed by fully automated high-throughput microscopy (Zeder and Pernthaler, 2009). Images were analysed with the freely available image analysis software ACMEtool (technobiology.ch), and interfering autofluorescent cyanobacteria or debris particle were individually excluded from hybridized cells. At least 10 high-quality images or >1000 DAPI-stained bacteria were analysed per sample. Cell sizes of CARD-FISH-stained ‘*Ca. Nanopelagicus*’ and ‘*Ca. Planktophila vernalis*’ were measured from two different samples (18 July 2012 and 15 May 2013; 5 m depth) with the software LUCIA (Laboratory Imaging) following a previously described workflow (Posch *et al.*, 2009). At least 87 individual cells per sample were subjected to image analysis.

Total numbers of heterotrophic prokaryotes and picocyanobacteria were determined by an inFlux V-GS cell sorter (Becton Dickinson, Franklin Lakes, NJ, USA) equipped with a ultraviolet light (355 nm) laser. Subsamples of 1 ml were stained with DAPI (1 µg ml⁻¹ final concentration), and scatter plots of

DAPI fluorescence vs 90° light scatter were analysed with an in-house software (J Villiger, unpublished).

Recruitment from metagenomes

Publically available metagenomes ($n=126$) were used for fragment recruitment (see Supplementary Table S7 for sampling sites, dates and accession numbers). rRNA sequences were masked in both genomes and metagenomes to avoid biases, and metagenomic reads were queried against the genomes using BLASTN (Altschul *et al.*, 1997) (length ≥ 50 bp, identity ≥ 95%, e -value ≤ 1e-5). These hits were used to compute RPKG values (number of reads recruited per kb of genome per Gb of metagenome) that provide a normalized value that is comparable across different metagenomes. Reads from three different time points (spring, summer, autumn) from Lake Mendota were also mapped on genomes without masking rRNA sequences and with identity values ≥ 80% for a graphical overview of seasonal differences of different genotypes (Supplementary Figure S13).

Comparative genomics and identification of genomic islands (GIs)

Whole-genome alignments were performed by comparison of all proteins from all genomes with each other using BLASTP. Stretches of genes without any hits to other genomes were manually examined to define GIs.

Results and discussion

Isolation and genome analysis of genome-streamlined freshwater ‘*Ca. Nanopelagicales*’

Sixteen strains affiliated to acI Actinobacteria were isolated using a dilution-to-extinction method (Salcher *et al.*, 2015) from the pelagial of Lake Zurich, Switzerland (Table 1 and Supplementary

Table 1 Details of the sequenced strains of planktonic ‘*Ca. Nanopelagicales*’

Taxonomy	Tribe ^a	Strains	Cell volume (µm ³)	Genome size (Mbp)	GC content (%)	No. of CDS	Spacer length (bp)	Coding density (%)
‘ <i>Ca. Planktophila dulcis</i> ’	acI-A1	MMS-IIA-65 MMS-IA-53 MMS-21-155	0.018–0.023	1.35–1.37	47.9–48.0	1344–1361	11–12	95.7
‘ <i>Ca. Planktophila sulfonica</i> ’	acI-A1	MMS-IA-56	0.026	1.34	48.6	1336	12	96.0
‘ <i>Ca. Planktophila versatilis</i> ’	acI-A1	MMS-IIB-76 MMS-IA-79 MMS-IA-105 MMS-IIB-142	0.021–0.027	1.27–1.33	48.2–48.3	1258–1329	14	95.2–95.5
‘ <i>Ca. Planktophila lacus</i> ’	acI-A1	MMS-IIB-106 MMS-IIB-60 MMS-21-148	0.025–0.029	1.38–1.46	47.5–47.8	1368–1438	15–16	95.3–95.6
‘ <i>Ca. Planktophila limnetica</i> ’	Phila	MMS-VB-114	0.016	1.33	45.0	1333	10	96.0
‘ <i>Ca. Planktophila vernalis</i> ’	acI-A7	MMS-IIA-15	0.020	1.36	45.7	1355	10	95.7
‘ <i>Ca. Nanopelagicus limnes</i> ’	acI-B1	MMS-21-122	0.018	1.24	41.5	1216	11	95.7
‘ <i>Ca. Nanopelagicus hibericus</i> ’	acI-B1	MMS-21-160	0.012	1.22	42.4	1211	13	95.4
‘ <i>Ca. Nanopelagicus abundans</i> ’	acI-B1	MMS-IIB-91	0.020	1.16	40.2	1150	14	95.3

Abbreviation: CDS, coding DNA sequence.

^aClassified according to ‘tribes’ proposed by Newton *et al.* (2011).

Tables S1 and S3). All isolates displayed slow growth, but reached relatively high densities (that is, $\sim 10^6$ cells per ml) in the initial cultures; they were of conspicuously small size with cell volumes ranging between 0.012 and 0.029 μm^3 (Supplementary Figure S1 and Supplementary Table S4). This corresponds to the *in situ* cell sizes of freshwater Actinobacteria from Lake Zurich (0.012–0.015 μm^3 , Supplementary Table S3) and other lakes (Posch *et al.*, 2009; Šimek *et al.*, 2014). These microbes are thus among the smallest members of the bacterioplankton, matching the small size of other abundant genome-streamlined bacteria (Supplementary Table S5; Salcher *et al.*, 2011; Ghai *et al.*, 2013; Zhao *et al.*, 2016). Although isolation campaigns were run throughout the growing season (Salcher *et al.*, 2015), only one strain ('*Ca. Planktophila limnetica*' MMS-VB-114) was isolated in summer and all others in spring (Supplementary Tables S1 and S3). Interestingly, each spring isolation campaign yielded a set of different genotypes, with strains affiliated to the same species being isolated multiple times even in different years. This already hints at a high number of sympatric genotypes during spring with remarkably recurrent temporal patterns. All axenic cultures were further inoculated to fresh medium prepared from sterile lake water; however, few strains survived more than

three propagations and all attempts to further maintain them in monocultures failed. Therefore, we used subsamples (0.5 ml) from the initial cultures (concentrated via centrifugation and frozen) for whole-genome sequencing after MDA. Assembly of sequence reads resulted in one–three large contigs that could be ordered and fully closed via PCR and Sanger sequencing. Additional 268 Sanger sequences from PCRs with primers designed for low-coverage regions ($< 10\times$) ensured a very high quality of the final genomes. This clearly contrasts with the fragmented nature of publically available *acI* SAGs consisting of 22 to 91 scaffolds (Ghylin *et al.*, 2014) and confirms that the use of high number of clonal cells (rather than a single cell) adequately compensates for MDA artefacts (Fitzsimons *et al.*, 2013).

The 16S rRNA gene phylogeny (Supplementary Figure S2) suggested a separation into two genera (*acI*-A and *acI*-B; sequence similarity $< 95\%$) and four species-like 'tribes' (*acI*-A1, *acI*-Phila, *acI*-A7, *acI*-B1; sequence similarity $\geq 97\%$; Newton *et al.*, 2011). However, average nucleotide identities ($> 95\%$, Supplementary Figure S3), average amino acid identities ($> 90\%$, Supplementary Figure S4) and phylogenomic trees based on conserved concatenated trees based on conserved concatenated proteins (Figure 1 and Supplementary Figure S5) revealed that the two genera harboured

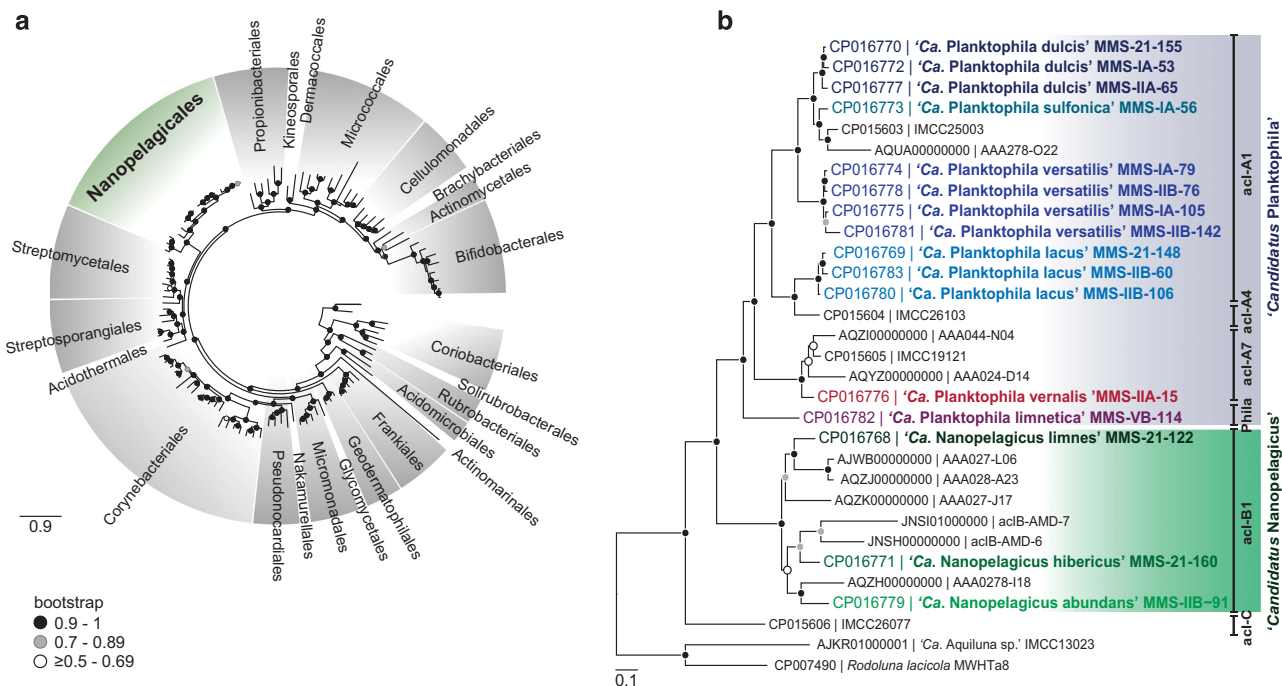


Figure 1 Phylogeny of freshwater '*Ca. Nanopelagiales*'. (a) Phylogenomic tree with > 100 complete genomes of the phylum Actinobacteria. Forty-eight concatenated conserved proteins were used to generate a maximum likelihood phylogenetic tree. The genomes of *Staphylococcus aureus* and *Listeria monocytogenes* were used as outgroup. Bootstrap values are indicated by black (0.9–1), grey (0.7–0.89) or white circles (≥ 0.5 –0.69) on the nodes. The proposed novel order '*Ca. Nanopelagiales*' is highlighted in green. (b) Phylogenomic tree of freshwater '*Ca. Nanopelagiales*'. Eighty-seven concatenated conserved proteins were used to generate a maximum likelihood phylogenetic tree. In addition to 20 complete genomes, seven single amplified genomes (prefixed by AAA) and two genomes assembled from metagenomes (prefixed by acI-B) were used to create the tree. The genomes of '*Ca. Aquiluna* sp.' and *Rhodoluna laticola* were used as outgroup. Bootstrap values are indicated black (0.9–1), grey (0.7–0.89) or white circles (≥ 0.5 –0.69) on the nodes.

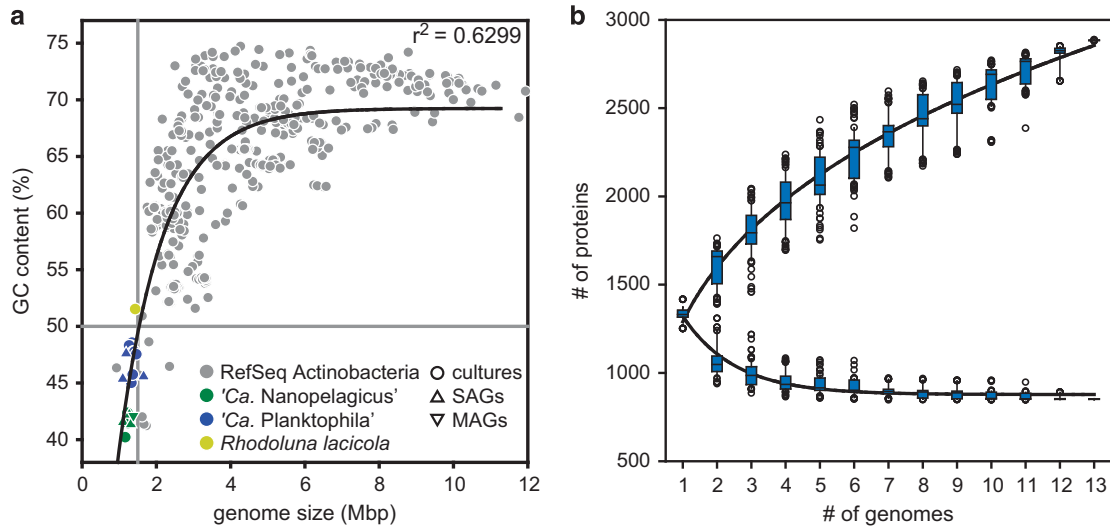


Figure 2 Genome streamlining and core-genome conservation in 'Ca. Nanopelagicales'. (a) Genomic GC content versus genome size for all complete published genomes of Actinobacteria ($n=610$; data taken from RefSeq). 'Ca. Nanopelagicales' and *Rhodoluna ladicola* are marked in different colours. (b) Pan- and core-genome analysis of 'Ca. Planktophila'. Regression lines were fitted with two-parameter power law regression (pan-genome) or exponential decay (core-genome).

nine different species. The majority of strains (13) were affiliated to the acI-A lineage (Newton *et al.*, 2011), one with a 16S rRNA sequence 100% identical to 'Ca. Planktophila limnetica' MWHEgel2-3 (FJ428831, Supplementary Figure S2), the only validly described mixed culture of acI Actinobacteria (Jezbera *et al.*, 2009). Thus, the acI-A lineage strains will be hereafter referred to as 'Ca. Planktophila'. Another strain ('Ca. Planktophila vernalis' MMS-IIA-15) showed high similarity to 'tribe' acI-A7. All other 'Ca. Planktophila' isolates formed a monophyletic branch and were grouped in four putative species containing 1–4 strains each (Figure 1b and Supplementary Figure S5). Three strains were affiliated to lineage acI-B1 (Newton *et al.*, 2011), all of them different species, and we tentatively assigned them to a novel genus, 'Ca. Nanopelagicus'. Together, both genera formed a robust sister lineage to the orders Streptomycetales, Streptosporangiales and Acidothermales (bootstrap value: 0.997) in a phylogenomic tree of >100 complete genomes of Actinobacteria (48 conserved proteins, Figure 1a and Supplementary Figure S6 (Sen *et al.*, 2014). We propose a novel order within the phylum Actinobacteria, 'Ca. Nanopelagicales' (family 'Ca. Nanopelagicaceae'), for the taxonomic placement of these microbes. The four strains published by Kang *et al.* (2017) were also affiliated to 'Ca. Nanopelagicales' representing different species within 'Ca. Planktophila' or a novel genus (strain IMCC26077; Figure 1b and Supplementary Figures S2–S5).

All strains have one circular chromosome of very small size (1.16–1.47 Mbp, Table 1 and Supplementary Table S3) with a low genomic GC content (40.2–48.6%). These are the smallest recorded genome sizes with the lowest GC content of so far

completely sequenced genomes of free-living Actinobacteria, except for the pathogen *Tropheryma whippelii* (Figure 2a). 'Ca. Nanopelagicus' genomes (as also estimated from incomplete SAGs (Ghylin *et al.*, 2014) and MAGs (Ghai *et al.*, 2014)) were smaller with lower GC content than 'Ca. Planktophila'. Summary statistics for all genomes are shown in Table 1 and Supplementary Table S3. As expected, they have very short intergenic spacers (10–16 bp), high coding densities (>95%) and only 2–5 genes encoding sigma factors (Supplementary Figure S7), all typical for streamlined genomes (Giovannoni *et al.*, 2014). The genomes reported here are in the same size range or even smaller than other genome-streamlined freshwater and marine microbes (Supplementary Table S5). All of these organisms are also of very small cell size, have a low genomic GC content and possess rhodopsins (Steindler *et al.*, 2011; Hahn *et al.*, 2014). Rhodopsins in 'Ca. Nanopelagicales' genomes were very similar to previously published actinorhodopsins of the acI lineage (Ghai *et al.*, 2014; Ghylin *et al.*, 2014; Kang *et al.*, 2017) with almost identical protein sequences (>99.3% sequence similarity) within the proposed species (Supplementary Figure S8). The rhodopsins are tuned to green light (L at position 105), like most freshwater and coastal marine rhodopsins. All strains encoded genes for β -carotene biosynthesis; however, five strains lacked the enzyme for the final step of retinal biosynthesis (*blh*; Supplementary Figure S8 and Supplementary Table S8), similar to two 'Ca. Nanopelagicales' genomes published by Kang *et al.* (2017) and *Rhodoluna ladicola*, whose proton-pumping activity was dependent on exogenous retinal (Keffer *et al.*, 2015). It is likely that energy generated by rhodopsins help 'Ca. Nanopelagicales' survive in the

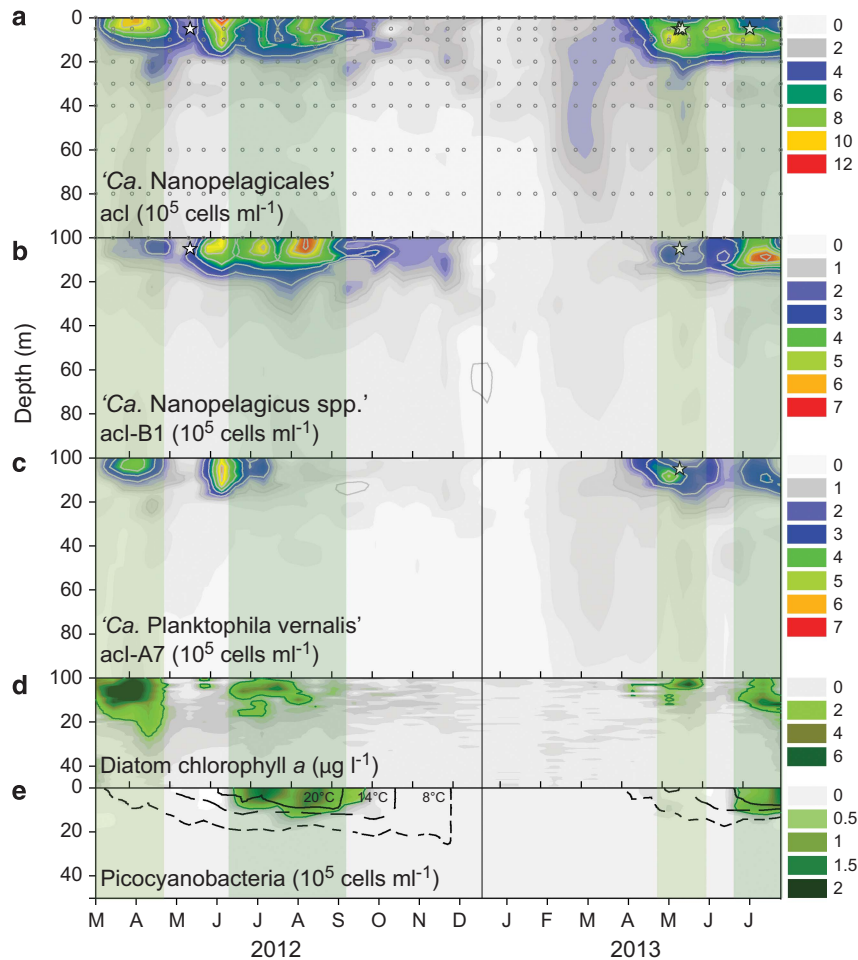


Figure 3 Abundances of ‘Ca. Nanopelagicales’ in Lake Zurich, Switzerland. (a) Abundances of all ‘Ca. Nanopelagicales’ (10^5 cells per ml) in a 2-year high-resolution sampling campaign. Circles in grey indicate individual samples ($n=348$). (b) Abundances of ‘Ca. Nanopelagicus spp.’ and (c) ‘Ca. Planktophila vernalis’. Asterisks indicate date and depth of isolation of strains. (d) Profiles of chlorophyll a derived from Diatoms ($\mu\text{g l}^{-1}$). (e) Profiles of water temperature (lines) and picocyanobacteria (10^5 cells per ml).

oligotrophic conditions of the plankton. The core-genome of the 13 ‘Ca. Planktophila’ strains (Figure 2b) encodes 852 conserved proteins (59–68% of the genome) comparable to that of the SAR11 group (Grote *et al.*, 2012). The size of the pan-genome of ‘Ca. Planktophila’ was estimated to be ca. 2900 genes and power law regression analysis indicated an open pan-genome for this genus, similar to marine SAR11 (Grote *et al.*, 2012).

‘Ca. Nanopelagicales’ are ubiquitous, highly abundant and show seasonal and habitat preferences

As 16S rRNA was insufficient as a taxonomic marker gene to discriminate isolates at strain levels, we used 23S rRNA sequences for designing specific oligonucleotide probes for ‘Ca. Planktophila vernalis’ and the genus ‘Ca. Nanopelagicus’ (Supplementary Figure S9 and Supplementary Table S3). These probes and the general probe for the order ‘Ca. Nanopelagicales’ (probe AcI-852; Warnecke *et al.*, 2005) were used for studying spatiotemporal

occurrence patterns by CARD-FISH in Lake Zurich, the origin of all strains ($n=348$ from two consecutive years). ‘Ca. Nanopelagicales’ were highly abundant and seasonally variable in the epilimnion (0–20 m depth) forming up to one-third of all microbes during the growing season (Figure 3). Distinct peaks of $>10^6$ cells per ml in spring and summer co-occurred with or followed algal and/or picocyanobacterial blooms, in agreement with earlier reports (Allgaier and Grossart, 2006; Salcher, 2014). It has been proposed that these microbes profit from phytoplankton blooms by utilizing carbohydrate-rich algal exudates (Garcia *et al.*, 2013; Salcher *et al.*, 2013; Pérez *et al.*, 2015), besides allochthonous carbon sources (Buck *et al.*, 2009; Pérez and Sommaruga, 2006). Arguably, copiotrophic microbes (for example, *Limnohabitans* spp., *Flavobacterium* spp.) react faster to resource pulses and might thus outcompete oligotrophs such as ‘Ca. Nanopelagicales’ (Simek *et al.*, 2014; Neuenschwander *et al.*, 2015). However, copiotrophs with a medium to large cell size are in the optimal range of size-selective protists and are therefore

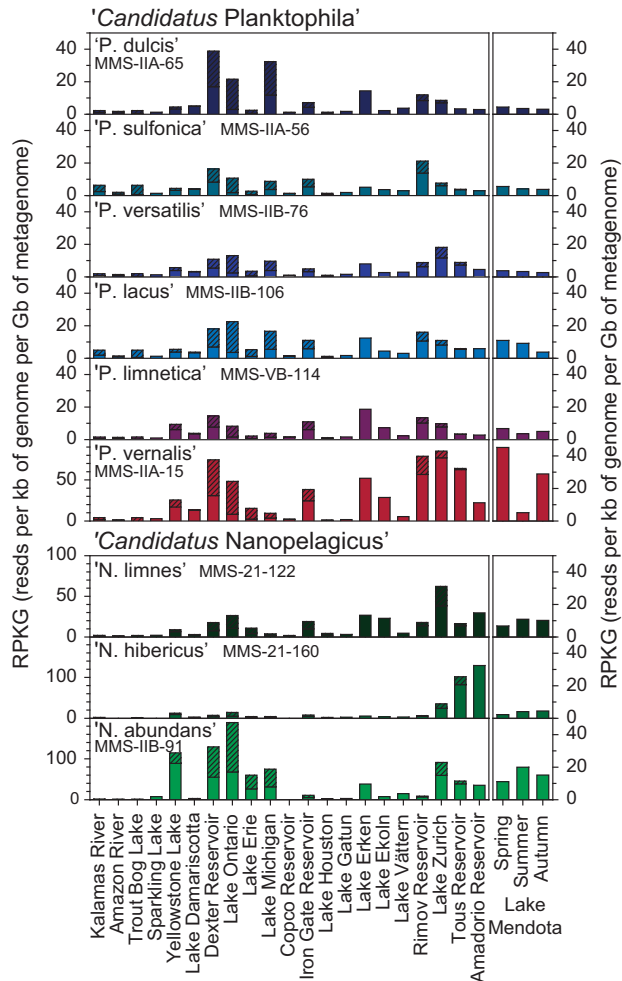


Figure 4 Metagenomic fragment recruitment of ‘*Ca. Nanopelagicales*’ across diverse freshwater ecosystems. Open bars represent the mean RPKG (number of reads recruited per kb of genome per Gb of metagenome) per ecosystem if multiple metagenomes were available, and shaded bars display the maximum RPKG. The data sets used ($n = 130$) are listed in Supplementary Table S7 and sorted according to habitat (rivers and lakes) and latitude (separately for North America and Europe). The right panel gives an example for seasonal differences in recruitment for Lake Mendota, USA.

heavily top-down controlled. Actinobacteria, being of very small size (Supplementary Figure S1 and Supplementary Table S4), are less vulnerable to grazing (Šimek *et al.*, 2013, 2014) and they are less frequently ingested and/or digested by flagellates (Tarao *et al.*, 2009; Šimek *et al.*, 2014). Moreover, ‘*Ca. Nanopelagicales*’ might also indirectly profit from the trophic cascades that typically follow algal blooms (that is, high numbers of copiotrophs followed by a maximum of protists and phages) as they grow on cell material that is released during bacterivory or viral lysis (Eckert *et al.*, 2013).

‘*Ca. Nanopelagicus*’ constituted roughly half of all ‘*Ca. Nanopelagicales*’ in Lake Zurich. They had distinct maxima of $>7 \times 10^5$ cells per ml (ca. 20% of all microbes) in summer, but were not abundant in spring (Figure 3). In contrast, ‘*Ca. P. vernalis*’ were present in high numbers in spring, but less so during

periods when ‘*Ca. Nanopelagicus*’ reached their maxima. Redundancy analysis suggested a clear separation of these two lineages: The abundances of ‘*Ca. Nanopelagicus*’ were positively related to water temperatures and the occurrence of picocyanobacteria, whereas ‘*Ca. P. vernalis*’ were more correlated to typical spring conditions, that is, high concentrations of oxygen, dissolved organic carbon, ammonium and chlorophyll *a* from diatoms and chlorophytes (Supplementary Figures S10 and S11). This niche separation of ‘*Ca. Nanopelagicales*’ is in line with previous observations within single systems and across different lakes (Newton *et al.*, 2007; Buck *et al.*, 2009).

Metagenomic fragment recruitment from freshwater samples ($n = 130$, Supplementary Table S7) showed a ubiquitous distribution for several isolates and very high abundances of all strains in lakes and reservoirs, whereas they were largely absent from the two rivers (Figure 4 and Supplementary Figure S12). ‘*Ca. N. abundans*’ MMS-IIB-91 was the most abundant genotype in all metagenomes and recruited more in summer samples than in spring, following the same trend as detected with CARD-FISH in Lake Zurich (Figure 3 and Supplementary Figure S13). The opposite was observed for ‘*Ca. P. vernalis*’ MMS-IIA-15 that was rare in a metagenome from Lake Mendota in summer but frequent in spring and autumn. This particular strain was highly abundant in most metagenomes, whereas other taxa of ‘*Ca. Planktophila*’ recruited less reads. Generally, the different genotypes were relatively more abundant in lakes of the northern temperate zone than in lakes at lower latitudes (Lake Houston and Lake Gatun). We could also identify a geographic preference of one strain (‘*Ca. N. hibernicus*’ MMS-21-160) that was highly abundant in two Spanish reservoirs. The closest relative of this strain was assembled from a metagenome from Amadorio Reservoir, Spain (acIB-AMD-6; average nucleotide identity: 95%), and displayed a similar recruitment pattern (Supplementary Figure S12; Ghai *et al.*, 2014). The two publically available genomes of Actinobacteria from the luna lineage (Kang *et al.*, 2012; Hahn *et al.*, 2014) recruited <10 times less than any of the ‘*Ca. Nanopelagicales*’ and, thus, seem to be of less relevance in fresh waters. Fragment recruitments of MAGs and SAGs were in the same range or lower than our genomes with the exception of one highly abundant SAG (AAA024-D14; Ghylis *et al.*, 2014) (Supplementary Figure S12).

High metabolic versatility of ‘*Ca. Nanopelagicales*’

Metabolic reconstructions suggested an aerobic chemoheterotrophic lifestyle of ‘*Ca. Nanopelagicales*’ with striking strain-specific differences and multiple auxotrophies. None of the strains encoded genes for assimilatory sulphate reduction (Supplementary Table S8), thus confirming a proposition based on SAGs (Ghylis *et al.*, 2014) and four

complete genomes (Kang *et al.*, 2017). This indicates a dependency on exogenous sources of reduced sulphur for growth, as reported for marine ‘*Ca. Pelagibacter*’ (Tripp *et al.*, 2008). The genomes shared several pathways of the central carbon metabolism (for example, glycolysis, tricarboxylic acid cycle, pentose phosphate pathway; see Supplementary Text and Supplementary Table S7), but differed in gluconeogenesis, carbohydrate metabolism and the biosynthesis of amino acids and cofactors. ‘*Ca. Planktophilia*’ strains were amino acid prototrophs except for ‘*Ca. P. vernalis*’ (serine auxotroph). All ‘*Ca. Nanopelagicus*’ strains were auxotrophs for histidine and ornithine. ‘*Ca. N. abundans*’ MMS-IIB-91 and ‘*Ca. N. limnes*’ MMS-21-122 also had incomplete pathways for methionine and lysine biosynthesis, and proline biosynthesis, respectively. However, most of these incomplete pathways lacked specific aminotransferases whose functions might be replaced by other predicted aminotransferases. ABC (ATP-binding cassette) transporters for amino acids were annotated for all genomes (Table 2), in agreement with their rapid *in situ* uptake (Pérez *et al.*, 2010; Salcher *et al.*, 2010; Salcher *et al.*, 2013). Cyanophycinases were present in all isolates (and in SAGs (Ghylin *et al.*, 2014) and MAGs (Ghai *et al.*, 2014)); this might provide additional amino acid sources in ‘*Ca. Nanopelagicales*’ that tend to co-occur with picocyanobacteria (Figure 3). All strains were prototrophs for vitamins B3 (nicotinamide adenine dinucleotide) and B6 (pyridoxine) and auxotrophs for B1 (thiamine), B5 (pantothenate), B12 (cobalamin; except for ‘*Ca. P. limnetica*’) and B7 (biotin). Two strains (‘*Ca. N. limnes*’ and ‘*Ca. N. hibericus*’) also lacked genes for the biosynthesis of vitamin B2 (riboflavin). This high degree of auxotrophy for various amino acids, vitamins and reduced sulphur sources might explain the unsuccessful long-term cultivation of our isolates. Such a dependence on co-occurring organisms providing essential metabolites seems to be a common feature among genome-streamlined planktonic microbes (Morris *et al.*, 2012; Giovannoni *et al.*, 2014; Garcia *et al.*, 2015).

The most striking differences between the strains were related to membrane transporters for low molecular weight organic compounds (Table 2 and Supplementary Table S8). All isolates shared transporters for spermidine/putrescine and the subsequent degradation pathway to succinate, as well as transporters for branched-chain amino acids, xanthine/uracil, nicotinamide adenine dinucleotide and several ions, in line with other ‘*Ca. Nanopelagicales*’ genomes (Kang *et al.*, 2017). All other membrane transporters ($n=32$) were only present in some strains, mostly in a species-specific manner. ‘*Ca. P. dulcis*’ had the highest number of different carbohydrate transporters (ribose, xylose, cellobiose, raffinose/stachyose/melibiose, α -glucoside) and carbohydrate metabolism in general, whereas the three ‘*Ca. Nanopelagicus*’ strains and ‘*Ca. P. limnetica*’

had the lowest diversity. Strain-specific differences in transporters were most pronounced in ‘*Ca. P. versatilis*’. A specialization on carbohydrates and versatile carbon substrate acquisition has also been hypothesized from the analysis of SAGs (Ghylin *et al.*, 2014). Such a metabolic microdiversification of closely related strains might be one reason for the high numbers of ‘*Ca. Nanopelagicales*’ in fresh waters, where resources are seasonally fluctuating and heterogeneously distributed at the microscale (Stocker, 2012).

High microdiversification of closely related genotypes encoded in GIs

We aligned the genomes of the 11 most closely related strains of ‘*Ca. Planktophilia*’ in a pairwise manner to identify GIs (that is, hypervariable regions; Figure 5 and Supplementary Table S9). Such an analysis can be better conducted with complete genomes and was thus not performed for previously described SAGs and MAGs. Two ‘*Ca. Planktophilia*’ strains (‘*Ca. P. vernalis*’ MMS-IIA-15 and ‘*Ca. P. limnetica*’ MMS-VB-114) had large inversions in some parts of their genomes (verified by PCR) that made it difficult to identify GIs (Supplementary Figure S14). We could recognize 14 distinct GIs that were also largely present in metagenomic recruitments (Supplementary Figure S14; that is, metagenomic islands (Rodriguez-Valera *et al.*, 2009)). Most GIs were additive and encoded genes for membrane transporters and/or carbohydrate metabolism (Figure 5 and Supplementary Table S9). These GIs were typically flanked by a transfer RNA and appear to be of importance for metabolic microdiversification, and might be rapidly evolving via horizontal gene transfer (as deduced from their patchy intra- and interspecific distribution; Vergin *et al.*, 2007; Hoetzing *et al.*, 2017).

One large replacement GI (GI-1) present in all strains featured genes encoding for cell wall biosynthesis and modifications like UDP-glucose 4-epimerase and glycosyltransferases. Moreover, one more additive GI contained genes for cell wall biosynthesis (GI-14), and another one (GI-2) for pilus assembly. All of these extracellular structures are potential phage recognition sites; modifications in the cell surface most likely serves as a defence against phage predation by changing the glycotype that is sensitive to a specific set of viruses (Rodriguez-Valera *et al.*, 2009). The three ‘*Ca. Nanopelagicus*’ strains were relatively distantly related and ‘*Ca. N. limnes*’ had a large inversion in the genome. Still, we identified eight GIs, of which three possessed genes involved in cell wall biosynthesis and/or modification (Supplementary Figure S15 and Supplementary Table S10). This high frequency of GIs encoding for cell surface properties and their high *in situ* abundances hints at high infection rates of phages (Knowles *et al.*, 2016). As abundant freshwater viruses (actinophages) were found to

Table 2 Complete membrane transport modules in 'Ca. Nanopelagiales'

	<i>P. dulcis</i> (n = 3)	<i>P. sulfo-</i> <i>nica</i>	<i>P. versatilis</i> (n = 4)	<i>P. lacus</i> (n = 3)	<i>P. limne-</i> <i>tica</i>	<i>P. ver-</i> <i>nalis</i>	<i>N.</i> <i>limnes</i>	<i>N. hiber-</i> <i>icus</i>	<i>N. abun-</i> <i>dans</i>
<i>Amino acids/peptides</i>									
Branched-chain amino acids	X	X	X	X	X	X	X	X	X
Polar amino acids			X		X	X			
Glycine betaine/proline					X	X			
Glutamate									X
Peptide/Ni ²⁺	X	X		X	X	X	X	X	X
<i>Polyamines</i>									
Spermidine/putrescine	X	X	X	X	X	X	X	X	X
Spermidine, putrescine				1	X	X			X
<i>Carbohydrates and alcohols</i>									
Simple sugar			3		X			X	X
Multiple sugar	X	X	X	X		X			X
Glucose/mannose				X			X	X	
Fructose				X					X
Ribose	X	X	X				X		
Rhamnose							X		
D-xylose	X	X			X	X			
Cellobiose	X		1			X			
<i>N,N</i> -diacetylchitobiose	1								
Raffinose/stachyose/melibiose	X		X	X		X			
Arabinogalactan oligomer/ maltooligosaccharide						X			
Glycerol					X				
Sorbitol/mannitol	2								
α-Glucoside	X	X				X			
<i>Nucleotides and coenzymes</i>									
Xanthine/uracil	X	X	X	X	X	X	X	X	X
Nicotinamide mononucleotide	X	X	X	X	X	X	X	X	X
Biotin	X	X	X	X	X	X			
Riboflavin (putative, riboswitch)							X	X	
Thiamine (putative, riboswitch)							X		X
Cobalamine (putative, riboswitch)	X	X	X	X		X			
<i>Organic acids</i>									
Cyanate			X			X			
Benzoate					X				
Sulfonate		X							
<i>Ions</i>									
PO ₄ ³⁻	X	X	X	X	X	X	X	X	X
H ⁺	X	X	X	X	X	X	X	X	X
NH ₄ ⁺	X	X	X	X	X	X	X	X	X
K ⁺	X	X	X	X	X	X	X	X	X
Fe ²⁺	X	X	X	X	X	X	X	X	X
Ni ²⁺			1		X		X	X	X
Zn ²⁺			1		X		X		X
Metal ions	2	X	3	2				X	
Na ⁺								X	
Co ²⁺ / Ni ²⁺	X	X	3	X		X			
Fe complex		X	X	X	X				
K ⁺					X				
<i>Export</i>									
Osmolytes	X	X	X	X	X	X	X	X	X
Teichoic acid	X	X	X	X	X	X	X	X	X
Lipopolysaccharide					X				X
Drug metabolites	X	X	X						
Lipooligosaccharide						X			

Genomes of the same species were merged for a better overview. Numbers indicate the number of genomes within one species that possess specific transporters.

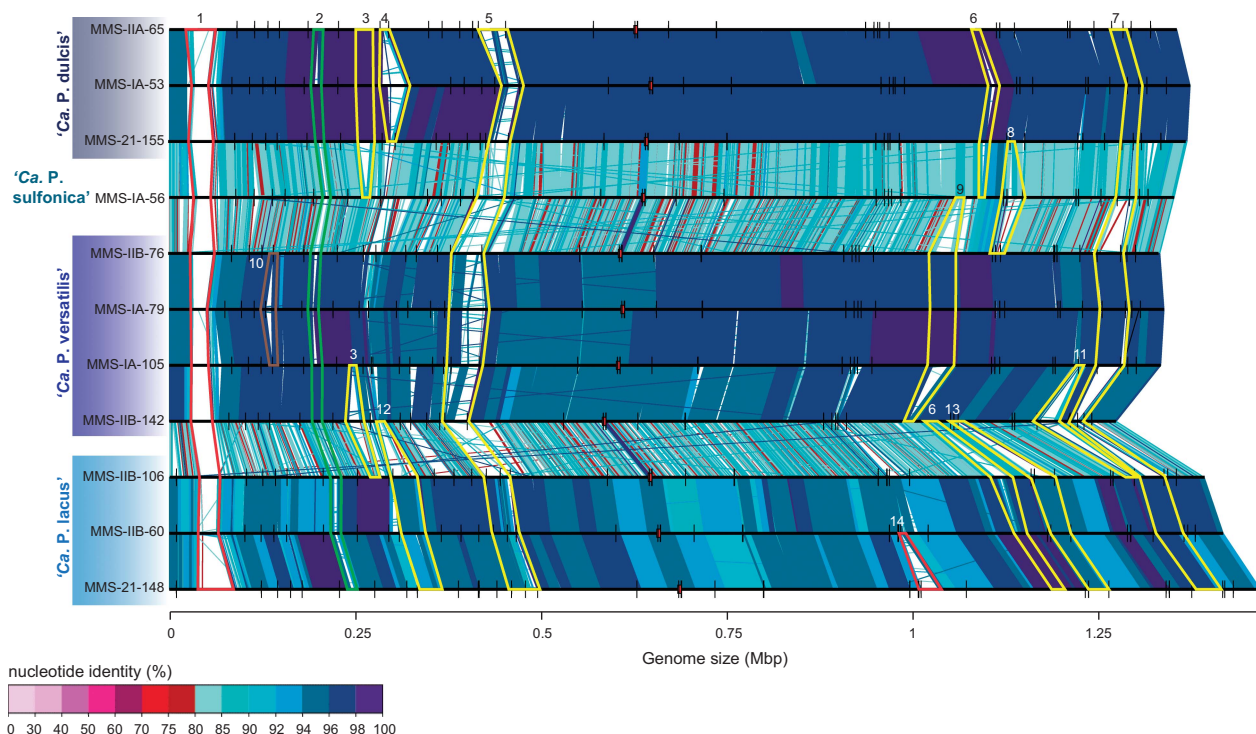


Figure 5 Whole-genome alignment of 11 ‘*Ca. Planktophila*’ strains. The genomes have been linearized for simplicity and are arranged in the same order as in the phylogenomic tree in Figure 1b. Synteny and different degrees of sequence similarity are indicated by vertical lines connecting the genomes. Ribosomal RNA (rRNA) operons in the individual genomes are displayed as red arrows and transfer RNAs (tRNAs) as short vertical lines. Genomic islands (GIs) have been marked in different colours and numbered (see Supplementary Table S9 for genes encoded in each island). Red: genes encoding mainly cell wall biosynthesis and modifications; green: genes encoding pilus assembly; yellow: genes encoding mainly membrane transport and/or carbohydrate metabolism; brown: unknown function.

infect *acI* Actinobacteria (Ghai *et al.*, 2017), these microbes might suffer from similarly high phage mortality rates as marine SAR11 (Zhao *et al.*, 2013). A positive feedback loop of a large number of closely related genotypes with high recombination rates was suggested to explain the large population sizes of SAR11 in the oceans despite being highly infected by viruses (‘king-of-the-mountain hypothesis’; Giovannoni *et al.*, 2013). As ‘*Ca. Nanopelagicales*’ possess multiple GIs with genes encoding potential phage recognition sites (Figure 5, Supplementary Figure S15 and Supplementary Tables S9 and S10), we suggest that a comparable mechanism allows for their high abundances in fresh waters. However, none of the ‘*Ca. Nanopelagicales*’ genomes encoded genes for the CRISPR-Cas system.

Conclusions

We present 16 high-quality genomes from isolates of the most abundant ubiquitously distributed microbes in freshwater systems, the *acI* lineage of Actinobacteria. They form a new order that we putatively named ‘*Ca. Nanopelagicales*’. These isolates are of extremely small cell size and possess the smallest genomes of all fully sequenced free-living Actinobacteria. The process of genome streamlining has

resulted in a high degree of auxotrophy for various vitamins, amino acids and reduced sulphur. We observed a high inter- and intraspecific diversity in metabolic pathways, especially regarding carbohydrate transport and metabolism, and these features were mainly encoded in genomic islands. This microdiversification of different genotypes might explain their global success in heterogeneous and variable environments such as lakes, where short-lived phytoplankton blooms and sporadic terrestrial influx serve as main carbon sources. We further propose that ‘*Ca. Nanopelagicales*’ suffer from high phage mortality rates that they balance off with genotype-specific modifications in their cell envelope, resulting in high total population sizes in the environment.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank Eugen Loher and Thomas Posch for help with sampling of Lake Zurich and Daniel Marty and Katya Pushkareva for help with CARD-FISH. Aharon Oren is acknowledged for etymological advice for correct Latin names for the proposed new taxa. We also thank three anonymous reviewers for supportive and helpful

comments. This work was financed by the Czech Science Foundation (CSF Grant Numbers 13–00243S and 17–04828S), by the Swiss National Science Foundation as part of the European Science Foundation EUROCORES Programme EuroEEFG (SNF grant number 31EE30-132771) and by the University Research Priority Program (URPP) ‘Global Change and Biodiversity’.

References

- Allgaier M, Grossart H. (2006). Diversity and seasonal dynamics of *Actinobacteria* populations in four lakes in northeastern Germany. *Appl Environ Microbiol* **72**: 3489–3497.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Res* **25**: 3389–3402.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.
- Bendall ML, Stevens SLR, Chan L-K, Malfatti S, Schwientek P, Tremblay J *et al.* (2016). Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J* **10**: 1589–1601.
- Beutler M, Wiltshire K, Meyer B, Moldaenke C, Lüring C, Meyerhöfer M *et al.* (2002). A fluorometric method for the differentiation of algal populations *in vivo* and *in situ*. *Photosynth Res* **72**: 39–53.
- Bolger AM, Lohse M, Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Buck U, Grossart HP, Amann R, Pernthaler J. (2009). Substrate incorporation patterns of bacterioplankton populations in stratified and mixed waters of a humic lake. *Environ Microbiol* **11**: 1854–1865.
- Burkert U, Warnecke F, Babenzien D, Zwirnmann E, Pernthaler J. (2003). Members of a readily enriched beta-proteobacterial clade are common in surface waters of a humic lake. *Appl Environ Microbiol* **69**: 6550–6559.
- Coil D, Jospin G, Darling AE. (2014). A5-miseq: an updated pipeline to assemble microbial genomes from illumina miseq data. *arXiv preprint* **1401**: 5130.
- Eckert EM, Salcher MM, Posch T, Eugster B, Pernthaler J. (2012). Rapid successions affect microbial N-acetylglucosamine uptake patterns during a lacustrine spring phytoplankton bloom. *Environ Microbiol* **14**: 794–806.
- Eckert EM, Baumgartner M, Huber IM, Pernthaler J. (2013). Grazing resistant freshwater bacteria profit from chitin and cell-wall derived organic carbon. *Environ Microbiol* **15**: 2019–2030.
- Fitzsimons MS, Novotny M, Lo C-C, Dichosa AEK, Yee-Greenbaum JL, Snook JP *et al.* (2013). Nearly finished genomes produced using gel microdroplet culturing reveal substantial intraspecies genomic diversity within the human microbiome. *Genome Res* **23**: 878–888.
- Garcia SL, McMahon KD, Martinez-Garcia M, Srivastava A, Sczyrba A, Stepanauskas R *et al.* (2013). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *ISME J* **7**: 137–147.
- Garcia SL, Buck M, McMahon KD, Grossart H-P, Eiler A, Warnecke F. (2015). Auxotrophy and intrapopulation complementary in the ‘interactome’ of a cultivated freshwater model community. *Mol Ecol* **24**: 4449–4459.
- 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. *Sci Rep* **3**: 2471.
- Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F. (2014). Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. *Mol Ecol* **23**: 6073–6090.
- Ghai R, Mehrshad M, Megumi Mizuno C, Rodriguez-Valera F. (2017). Metagenomic recovery of phage genomes of uncultured freshwater actinobacteria. *ISME J* **11**: 304–308.
- Ghylin TW, Garcia SL, Moya F, Oyserman BO, Schwientek P, Forest KT *et al.* (2014). Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. *ISME J* **8**: 2503–2516.
- Giovannoni S, Temperton B, Zhao Y. (2013). Giovannoni *et al.* reply. *Nature* **499**: E4–E5.
- Giovannoni SJ, Cameron Thrash J, Temperton B. (2014). Implications of streamlining theory for microbial ecology. *ISME J* **8**: 1553–1565.
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.
- Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ *et al.* (2012). Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *MBio* **3**: pii: e00252-12.
- Hahn MW, Schmidt J, Taipale SJ, Doolittle WF, Koll U. (2014). *Rhodoluna laciola* gen. nov., sp. nov., a planktonic freshwater bacterium with stream-lined genome. *Int J Syst Evol Microbiol* **64**: 3254–3263.
- Hoetzinger M, Schmidt J, Jezberová J, Koll U, Hahn MW. (2017). Microdiversification of a pelagic *Polynucleobacter* species is mainly driven by acquisition of genomic islands from a partially interspecific gene pool. *Appl Environ Microbiol* **83**: pii: e02266-16.
- Jezbera J, Sharma AK, Brandt U, Doolittle WF, Hahn MW. (2009). ‘*Candidatus* Planktophilia limnetica’, an actinobacterium representing one of the most numerically important taxa in freshwater bacterioplankton. *Int J Syst Evol Microbiol* **59**: 2864–2869.
- Kanehisa M, Sato Y, Morishima K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**: 726–731.
- Kang I, Lee K, Yang S-J, Choi A, Kang D, Lee YK *et al.* (2012). Genome sequence of ‘*Candidatus* Aquiluna’ sp. strain IMCC13023, a marine member of the Actinobacteria isolated from an Arctic fjord. *J Bacteriol* **194**: 3550–3551.
- Kang I, Kim S, Islam MR, Cho J-C. (2017). The first complete genome sequences of the acI lineage, the most abundant freshwater Actinobacteria, obtained by whole-genome-amplification of dilution-to-extinction cultures. *Sci Rep* **7**: 42252.
- Keffer JL, Hahn MW, Maresca JA. (2015). Characterization of an unconventional rhodopsin from the freshwater

- Actinobacterium *Rhodoluna ladicola*. *J Bacteriol* **197**: 2704–2712.
- Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG *et al.* (2016). Lytic to temperate switching of viral communities. *Nature* **531**: 466–470.
- Langmead B, Salzberg SL. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–U354.
- Lassmann T, Sonnhammer EL. (2005). Kalign – an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* **6**: 298.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acid Res* **32**: 1363–1371.
- Luo H, Thompson LR, Stingl U, Hughes AL. (2015). Selection maintains low genomic GC content in marine SAR11 lineages. *Mol Biol Evol* **32**: 2738–2748.
- Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY *et al.* (2015). CDD: NCBI's conserved domain database. *Nucleic Acid Res* **43**: D222–D226.
- Morris JJ, Lenski RE, Zinser ER. (2012). The Black Queen hypothesis: evolution of dependencies through adaptive gene loss. *mBio* **3**: pii: e00036-12.
- Muyzer G, Dewaal EC, Uitterlinden AG. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**: 695–700.
- Muyzer G, Teske A, Wirsén CO, Jannasch HW. (1995). Phylogenetic relationships of Thiomicrospira species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol* **164**: 165–172.
- Neuenschwander SM, Pernthaler J, Posch T, Salcher MM. (2015). Seasonal growth potential of rare lake water bacteria suggest their disproportional contribution to carbon fluxes. *Environ Microbiol* **17**: 781–795.
- Newton RJ, Jones SE, Helmus MR, McMahon KD. (2007). Phylogenetic ecology of the freshwater *Actinobacteria* acI lineage. *Appl Environ Microbiol* **73**: 7169–7176.
- Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol R* **75**: 14–49.
- Pérez MT, Sommaruga R. (2006). Differential effect of algal- and soil-derived dissolved organic matter on alpine lake bacterial community composition and activity. *Limnol Oceanogr* **51**: 2527–2537.
- Pérez MT, Hörtnagl P, Sommaruga R. (2010). Contrasting ability to take up leucine and thymidine among freshwater bacterial groups: implications for bacterial production measurements. *Environ Microbiol* **12**: 74–82.
- Pérez MT, Rofner C, Sommaruga R. (2015). Dissolved organic monomer partitioning among bacterial groups in two oligotrophic lakes. *Environ Microbiol Rep* **7**: 265–272.
- Posch T, Franzoi J, Prader M, Salcher MM. (2009). New image analysis tool to study biomass and morphotypes of three major bacterioplankton groups in an alpine lake. *Aquat Microb Ecol* **54**: 113–126.
- Posch T, Koster O, Salcher MM, Pernthaler J. (2012). Harmful filamentous cyanobacteria favoured by reduced water turnover with lake warming. *Nat Clim Change* **2**: 809–813.
- Price MN, Dehal PS, Arkin AP. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J *et al.* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acid Res* **35**: 7188–7196.
- Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, Zhou J *et al.* (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* **196**: 2210–2215.
- Rodríguez-R LM, Konstantinidis KT. (2014). Bypassing cultivation to identify bacterial species. *ASM Microbe Magazine* **9**: 111–118.
- Rodríguez-Valera F, Martín-Cuadrado A-B, Rodríguez-Brito B, Pasic L, Thingstad TF, Rohwer F *et al.* (2009). Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7**: 828–836.
- Salcher MM, Pernthaler J, Posch T. (2010). Spatiotemporal distribution and activity patterns of bacteria from three phylogenetic groups in an oligomesotrophic lake. *Limnol Oceanogr* **55**: 846–856.
- Salcher MM, Pernthaler J, Posch T. (2011). Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria 'that rule the waves' (LD12). *ISME J* **5**: 1242–1252.
- Salcher MM, Posch T, Pernthaler J. (2013). *In situ* substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *ISME J* **7**: 896–907.
- Salcher MM. (2014). Same same but different: ecological niche partitioning of planktonic freshwater prokaryotes. *J Limnol* **73**: 74–87.
- Salcher MM, Neuenschwander SM, Posch T, Pernthaler J. (2015). The ecology of pelagic freshwater methylo-trophs assessed by a high-resolution monitoring and isolation campaign. *ISME J* **9**: 2442–2453.
- Schmieder R, Edwards R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863–864.
- Sekar R, Pernthaler A, Pernthaler J, Warnecke F, Posch T, Amann R. (2003). An improved protocol for quantification of freshwater *Actinobacteria* by fluorescence *in situ* hybridization. *Appl Environ Microbiol* **69**: 2928–2935.
- Sen A, Daubin V, Abrouk D, Gifford I, Berry AM, Normand P. (2014). Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders 'Frankiales' and Micrococcales should be split into coherent entities: proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov. and Nakamurellales ord. nov. *Int J Syst Evol Microbiol* **64**: 3821–3832.
- Šimek K, Kasalický V, Jezbera J, Horňák K, Nedoma J, Hahn MW *et al.* (2013). Differential freshwater flagellate community response to bacterial food quality with a focus on *Limnohabitans* bacteria. *ISME J* **7**: 1519–1530.
- Šimek K, Nedoma J, Znachor P, Kasalický V, Jezbera J, Horňák K *et al.* (2014). A finely tuned symphony of factors modulates the microbial food web of a freshwater reservoir in spring. *Limnol Oceanogr* **59**: 1477–1492.
- Stamatakis A, Ludwig T, Meier H. (2005). RAxML-II: a program for sequential, parallel and distributed inference of large phylogenetic. *Concurr Comput Pract Exp* **17**: 1705–1723.
- Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ. (2011). Energy starved *Candidatus* Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS One* **6**: e19725.

- Stocker R. (2012). Marine microbes see a sea of gradients. *Science* **338**: 628–633.
- Tarao M, Jezbera J, Hahn MW. (2009). Involvement of cell surface structures in size-independent grazing resistance of freshwater *Actinobacteria*. *Appl Environ Microbiol* **75**: 4720–4726.
- Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS *et al*. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acid Res* **29**: 22–28.
- Tripp HJ, Kitner JB, Schwalbach MS, Dacey JWH, Wilhelm LJ, Giovannoni SJ. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452**: 741–744.
- Tritt A, Eisen JA, Facciotti MT, Darling AE. (2012). An integrated pipeline for de novo assembly of microbial genomes. *PLoS One* **7**: e42304.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M *et al*. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Res* **40**: e115.
- Vergin KL, Tripp HJ, Wilhelm LJ, Denver DR, Rappe MS, Giovannoni SJ. (2007). High intraspecific recombination rate in a native population of *Candidatus Pelagibacter ubique* (SAR11). *Environ Microbiol* **9**: 2430–2440.
- Warnecke F, Sommaruga R, Sekar R, Hofer JS, Pernthaler J. (2005). Abundances, identity, and growth state of *Actinobacteria* in mountain lakes of different UV transparency. *Appl Environ Microbiol* **71**: 5551–5559.
- Zeder M, Pernthaler J. (2009). Multispot live-image auto-focusing for high-throughput microscopy of fluorescently stained bacteria. *Cytometry Part A* **75A**: 781–788.
- Zhao X, Schwartz C, Pierson J, Giovannoni SJ, McIntosh JR, Nicastro D. (2016). Three-dimensional structure of the ultra-oligotrophic marine bacterium *Pelagibacter*. *Appl Environ Microbiol* **83**: e02807–e02816.
- Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC *et al*. (2013). Abundant SAR11 viruses in the ocean. *Nature* **494**: 357–360.

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