View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ZORA

University of Zurich[UZH]

Zurich Open Repository and Archive
University of Zurich
Main Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2019

Niche-directed evolution modulates genome architecture in freshwater Planctomycetes

Andrei, Adrian-Ştefan; Salcher, Michaela M; Mehrshad, Maliheh; Rychtecký, Pavel; Znachor, Petr; Ghai, Rohit

Abstract: Freshwater environments teem with microbes that do not have counterparts in culture collections or genetic data available in genomic repositories. Currently, our apprehension of evolutionary ecology of freshwater bacteria is hampered by the difficulty to establish organism models for the most representative clades. To circumvent the bottlenecks inherent to the cultivation-based techniques, we applied ecogenomics approaches in order to unravel the evolutionary history and the processes that drive genome architecture in hallmark freshwater lineages from the phylum Planctomycetes. The evolutionary history inferences showed that sediment/soil Planctomycetes transitioned to aquatic environments, where they gave rise to new freshwater-specific clades. The most abundant lineage was found to have the most specialised lifestyle (increased regulatory genetic circuits, metabolism tuned for mineralization of proteinaceous sinking aggregates, psychrotrophic behaviour) within the analysed clades and to harbour the smallest freshwater Planctomycetes genomes, highlighting a genomic architecture shaped by niche-directed evolution (through loss of functions and pathways not needed in the newly acquired freshwater niche).

DOI: https://doi.org/10.1038/s41396-018-0332-5

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-163221
Journal Article
Published Version

The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:
Andrei, Adrian-Ştefan; Salcher, Michaela M; Mehrshad, Maliheh; Rychtecký, Pavel; Znachor, Petr; Ghai, Rohit (2019). Niche-directed evolution modulates genome architecture in freshwater Planctomycetes. The ISME journal, 13(4):1056-1071.
DOI: https://doi.org/10.1038/s41396-018-0332-5

The ISME Journal
https://doi.org/10.1038/s41396-018-0332-5

ARTICLE

Niche-directed evolution modulates genome architecture in freshwater Planctomycetes

Adrian-Ştefan Andrei[1] · Michaela M. Salcher[1,2] · Maliheh Mehrshad[1] · Pavel Rychtecký[1] · Petr Znachor[1] · Rohit Ghai[1]

Received: 15 May 2018 / Revised: 22 November 2018 / Accepted: 29 November 2018
© The Author(s) 2019. This article is published with open access

Abstract

Freshwater environments teem with microbes that do not have counterparts in culture collections or genetic data available in genomic repositories. Currently, our apprehension of evolutionary ecology of freshwater bacteria is hampered by the difficulty to establish organism models for the most representative clades. To circumvent the bottlenecks inherent to the cultivation-based techniques, we applied ecogenomics approaches in order to unravel the evolutionary history and the processes that drive genome architecture in hallmark freshwater lineages from the phylum Planctomycetes. The evolutionary history inferences showed that sediment/soil Planctomycetes transitioned to aquatic environments, where they gave rise to new freshwater-specific clades. The most abundant lineage was found to have the most specialised lifestyle (increased regulatory genetic circuits, metabolism tuned for mineralization of proteinaceous sinking aggregates, psychrotrophic behaviour) within the analysed clades and to harbour the smallest freshwater Planctomycetes genomes, highlighting a genomic architecture shaped by niche-directed evolution (through loss of functions and pathways not needed in the newly acquired freshwater niche).

Introduction

Planctomyces bacteria (*sensu* Woese et al.) [1] encompass one of the most enigmatic branches of the prokaryotic tree of life that have been brought into axenic culture [2]. This division, envisioned as a phylum [3], was thought to accommodate members that either rooted deeply in the bacterial line of descent [4] or paved the way to eukaryality [5, 6]. The obscurity surrounding the phylum arose decades ago when Nándor Gimesi described what he considered an unusual planktonic fungus (i.e., *Planctomyces bekefii*) in the eutrophic waters of Lake Langymanyos (Budapest, Hungary) [7]. However, this microbe was later acknowledged to be of bacterial origin [8] and used to denominate the phylum Planctomycetes (Gr. adj. planktos wandering, floating; Gr. masc. n. mukês fungus; N.L. masc. n. Planctomyces floating fungus). The atypical morphology (e.g., microcolonial rosettes of cells joined together at the tips of their stalks) that misled Gimesi was found to be the norm for a phylum that accommodates bacteria with a vast array of shapes (from spherical and ellipsoidal to bulbiform), appendages (from spikes and bristles to stalks) [9–12] and outer membrane crateriform complexes [10]. Moreover, their puzzling appearance was found to be accompanied by a cell plan that seemed to diverge from the classical bacterial 'Gram-negative' one due to the following: (i) apparent cytosolic compartmentalization [13], (ii) lack of peptidoglycan (i.e., a hallmark of free-living bacteria) [14] and (iii) presence of an endocytosis-like macromolecular uptake mechanism (a process universal among eukaryotes) [15]. The phylum's peculiarities generally withstood genome-centric analyses, that in a way further deepened the knowledge gap

Supplementary information The online version of this article (https://doi.org/10.1038/s41396-018-0332-5) contains supplementary material, which is available to authorized users.

✉ Adrian-Ştefan Andrei
adrian.stefan.andrei@hbu.cas.cz

✉ Rohit Ghai
ghai.rohit@gmail.com

[1] Department of Aquatic Microbial Ecology, Institute of Hydrobiology, Biology Centre CAS, Na Sádkách 702/7, 370 05 České Budějovice, Czech Republic

[2] Limnological Station, Institute of Plant and Microbial Biology, University of Zurich, Seestrasse 187, 8802 Kilchberg, Switzerland

Published online: 04 January 2019

by revealing the presence of a large 'ORFan black hole' (functional prediction for only 32–54% of ORFs) [16–18] and 'giant genes' [19] harbored by huge genomes (median genome size of sequenced Planctomycetes is 7.4 Mb in comparison to the more typical 3-4 Mb of other sequenced genomes). In light of recent research not only is the supposed 'link' to eukaryotes a product of convergent evolution [20], the endocytosis-like macromolecule uptake questionable [21] and cell plan an altered 'Gram-negative' one [22], the integration of genomic and structural data into an ecological framework also lags behind.

Consisting of two classes (i.e., Phycisphaerae and Planctomycetacia) that exhibit global ubiquity [23–27], the Planctomycetes phylum evaded extensive ecological characterization as a result of the inability to bring environmentally abundant representatives into axenic culture, or to access their genomic information [28]. In spite of their initial description in freshwater environments [2, 7], the majority of ecological and genomic studies were performed on marine ecosystems and seawater isolates [16, 18, 23, 29]. Although they represent one of the major prokaryotic groups in freshwater (with highly variable abundances from <1 up to 22%) [27, 30–32] and have been shown to have major roles in dissolved organic matter fractionation [33], our understanding of Planctomycetes is based on data derived largely from culture-based approaches [2, 34, 35], short reads analyses or/and hybridization-based techniques [27, 31, 32, 36]. While prone to primer coverage biases [37], the 16 S rRNA gene-based studies pointed out that the abundant freshwater ribotypes do not have counterparts in culture and that their genomic diversity and ecological significance remains elusive [27, 32, 38]. Although in the light of recent research, Planctomycetes groups have been defined based on 16 S rRNA gene relatedness (i.e. CL500-3, CL500-15 and CL500-37) and some are considered to be abundant in lakes and envisioned as hypolimnion specific [27], our apprehension of their ecology remains dim.

Here we use ecosystem-scale taxonomic profiling (based on 298 metagenomic data sets), genome-resolved metagenomics (60 Planctomycetes genomes recovered from ten large metagenomic data sets) and spatio-temporal abundance patterns (using CARD-FISH) to elucidate the evolutionary history of lacustrine Planctomycetes, and to link their genome evolution patterns to their lifestyle strategies. In doing so, we not only characterized some of the most iconic freshwater bacterial lineages from an ecologic, genomic and metabolic perspective, but also broadened our view on their evolution at large.

Results and discussion

An aquatic Planctomycetes census based on short-read technology

To explore the taxonomic extent of aquatic Planctomycetes and to assess their contribution to prokaryotic community structure, we taxonomically profiled 298 metagenomic data sets derived from lacustrine (64 data sets), fluvial (36 data sets), freshwater sediments (40 data sets) and marine (158 data sets) habitats (see Extended Data for a complete list). By making use of high spatial scale data (spread over four continents and along the Global Ocean) we show that Planctomycetes are ubiquitously present in aquatic habitats and sediments, where their contribution to prokaryotic assemblages varies (from absence to 13.13%) by environmental spatial heterogeneity (e.g., intralake; Fig. 1) and to a lesser extent, habitat (with higher absolute abundances registered in freshwater habitats) (Fig. 1, Supplementary Figure 1). For instance, the fluctuation in abundance (as assessed by the percentage of 16S rRNA gene reads), within prokaryotic community structure (e.g., Lake Zurich, samples collected on 13th of May 2013; Fig. 1), from scarcely present (0.1% rRNA gene reads in the epilimnion) to highly abundant (13.1% rRNA gene reads in the hypolimnion) pointed towards a niche, rather than habitat, preference. We observed that the taxonomic categories of aquatic Planctomycetes have a tendency to be more uniform within- than between-habitats (e.g. freshwater vs marine) (Fig. 1), and that representatives of class Phycisphaerae form a major constituent of prokaryotic communities (up to 11.8% of total 16S rRNA reads) in lakes and reservoirs (e.g., Lake Zurich; Fig. 1).

In spite of the broad intra-phylum diversity (as represented by 16S rRNA genes), the taxonomic breakdown revealed the presence of a freshwater Planctomycetes blueprint (Fig. 1) largely characterized by the dominance of clade CL500-3 (Phycisphaerae) and a collection of 'uncultured' groups of the Planctomycetaceae (Planctomycetacia). Regardless of their wide environmental distribution, the dominant taxa were found not to relate to the phylum's cultured diversity, or even to appertain to groups underrepresented in sequence databases (i.e. CL500-3 group consists of 40 sequences in SILVA's SSU Ref NR 99 128 dataset, where it represents only ca. 1.6% of the Phycisphaerae sequences). While 'Planctomycetaceae uncultured' represents an umbrella taxonomic category (composed of multiple polyphyletic clusters without any cultured representative), the CL500-3 forms a cohesive phylogenetic clade, described initially in the deep water column of the ultra-oligotrophic Crater Lake (hence the name of the group) [38].

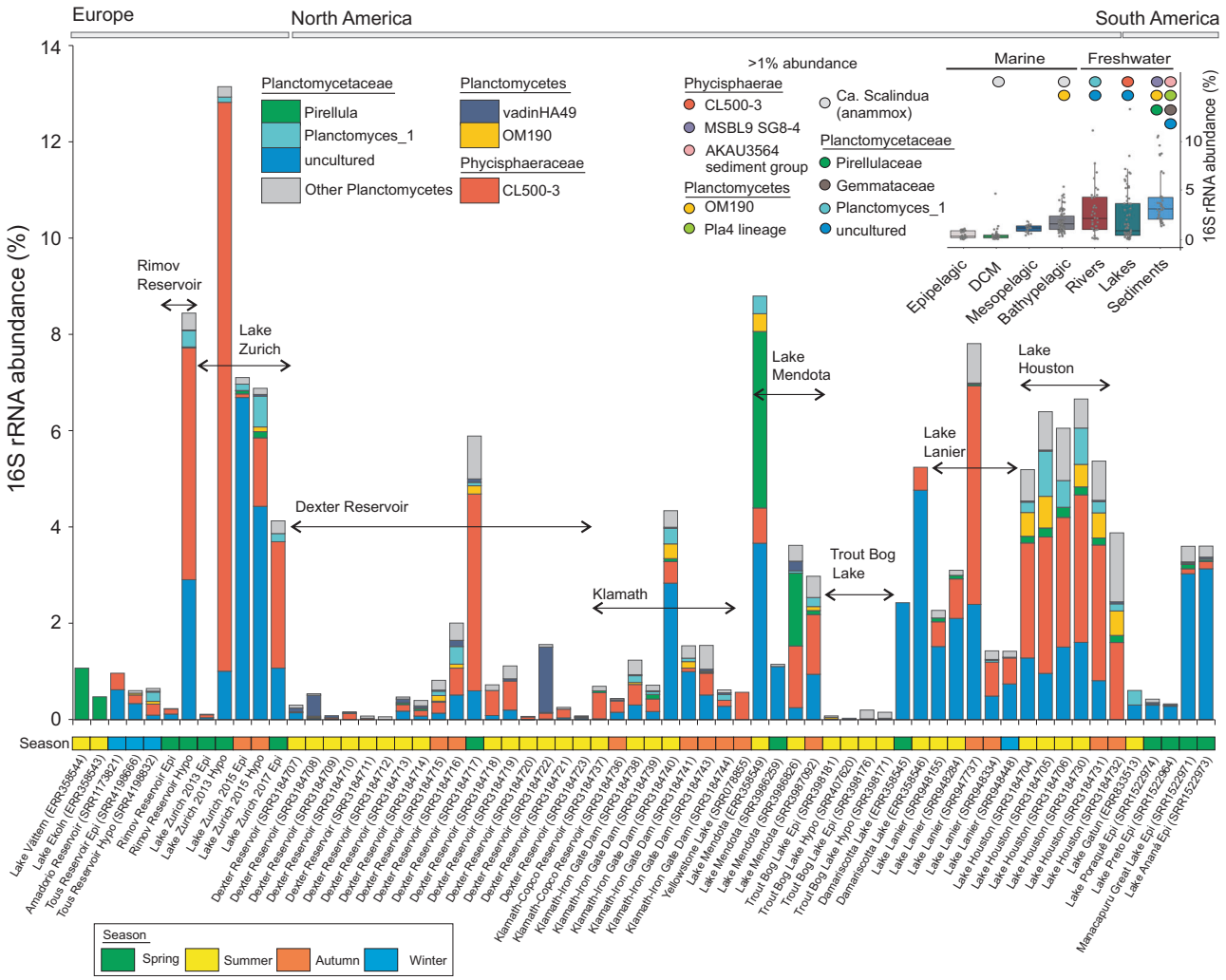

Fig. 1 Taxonomic milieu of Planctomycetes phylum in worldwide lacustrine habitats. The figure depicts the SILVA SSU (Ref NR 99 128) classification of 16S rRNA gene fragments (as unassembled shotgun reads) retrieved from 64 freshwater data sets. The X-axis shows the taxonomic ranks and the geographic distribution of the sample collection sites, while the Y-axis indicates the percentage of Planctomycetes within the prokaryotic communities (as assessed by 16S rRNA genes abundance). The sample collection time, following a four-seasons breakdown, is indicated by colored boxes arranged along the X-axis. The SRA identifier for each metagenome is indicated in the parentheses that follow the habitat name. The figure's inset (upper right panel) shows the contribution of Planctomycetes (as assessed by 16S rRNA gene abundance in 298 metagenomic data sets) to the prokaryotic communities present in aquatic and freshwater sediments (64 lacustrine, 36 fluvial, 34 epipelagic, 46 deep chlorophyll maxima, 16 mesopelagic, 62 bathypelagic and 40 sediments). The colored circles highlight taxa that reached more than 1% abundance within prokaryotic communities. DCM: deep chlorophyll maxima

A fine-scale phylogenomic picture of freshwater Planctomycetes

The applied hybrid binning strategy (taxonomy dependent, using homology searches and taxonomy independent, using tetra-nucleotide frequencies and mean base coverages) allowed the recovery of high-confidence Planctomycetes-affiliated contigs, and their segregation into individual metagenome-assembled genomes (MAGs). The obtained MAGs were further assessed for completeness and redundancy based on the presence of ubiquitous single-copy genes (360 Planctomycetes-specific genes) and amino acid identity between multicopy ones (Supplementary Figure 2). After performing additional data curation we obtained 60 MAGs (9 548 contigs; total length 123.7 Mb; average contig length 12.9 Kb) that simultaneously met our quality criteria (completeness ≥10%, contamination ≤10%, number of contigs ≤500), and had an average coverage depth higher than 5-fold over 90% of the nucleotides (ensuring for high-confidence base identification) (Extended Data, Supplementary Figure 2). To the best of our knowledge, the present dataset of 60 MAGs encompasses by far the largest compilation of genomic information available for freshwater Planctomycetes (in contrast the 7 903 UBA genomes dataset contains only six freshwater Planctomycetes MAGs) [39].

We emphasize that the obtained 60 MAGs represent 'genomic pools' of Planctomycetes populations that share high sequence identity, and that they do not accurately reflect the genomic make up of specific clonal lineages. The alignment of short metagenomic reads to the MAGs showed that freshwater Planctomycetes typically consist of ecologically coherent and sequence-discrete populations (characterized by 98.5 – 100% sequence identity), that exhibit both panmictic and clonal lifestyles. For instance, we observed that the population represented by the MAG TH-plancto1 was undergoing a selective event (at the time of sampling), which was on the way of producing a (nearly) clonal population (Supplementary Figure 3). On the other side of the spectrum, the ZH-13MAY13-plancto44 population was found to harbor highly panmictic gene pools (Supplementary Figure 3).

The evolutionary relations and the taxonomic ranks of the 60 Planctomycetes MAGs were investigated through gene- and genome-focused phylogenies. The topological backbone of the phylogenomic tree was supported by the phylogenetic one (i.e. using 16 S rRNA – the most-adopted phylogenetic marker), and both methods reinforced a three-clade branching pattern comprising anammox planctomycetes and the two classes Planctomycetacia and Phycisphaerae (Fig. 2, Supplementary Figure 4). All our 60 freshwater MAGs branched within these two existing classes, where they formed monophyletic groups that were usually divergent from the cultured and metagenomics-recovered representatives (Fig. 2, Supplementary Figure 4).

We found that 22 MAGs share a common evolutionary lineage within class Phycisphaerae (red box, Fig. 2), which (at the time of writing) comprises only three cultured non-freshwater species (i.e. *Phycisphaera mikurensis*, *Algisphaera agarilytica* - both isolated from a marine alga and *Tepidisphaera mucosa* isolated from a terrestrial hot spring) from which one genome is publicly available (that of *Phycisphaera mikurensis*). This phylogenetic cluster seems to form an ecologically coherent aquatic group together with the marine *P. mikurensis* that shares a common ancestry with the deeper branching sediment-dwelling representatives of the class (branches represented by the MAGs: Phycisphaerae bacterium SM23_33 and Uncultured Phycisphaerae; Fig. 2). Hereinafter, we made use of 16 S rRNA genes as 4 MAGs from the 22 were found to have 16S rRNA genes (Supplementary Figure 4) to anchor the phylogenomic Phycisphaerae clade (comprised of 22 MAGs, red box) into the larger gene-based bacterial taxonomy, and show that the MAGs fall within the CL500-3 clade (Supplementary Figure 4), the hallmark taxonomic group of lacustrine habitats (Fig. 1). Thus, in this study, we managed to recover not just one of the largest number of Phycisphaerae MAGs ($n = 22$), but also the most extensive genomic repertoire of an ecologically relevant and abundant freshwater bacterial lineage that so far completely resisted cultivation-dependent and –independent analyses. As indicated by the clade topology within the phylogenomic tree (e.g., statistically supported monophyletic lineage; Fig. 2) and its congruence within 16 S rRNA phylogeny (Supplementary Figure 4), we propose to designate a taxonomic category, to encompass this uncultured group, in accordance with the guidelines of Konstantinidis et al. [40]. Based on average amino-acid identities between the 22 MAGs (that registered values lower than 65%; Supplementary Figure 5) [41], we suggest the creation of the family-rank Nemodlikiaceae (fam. nov.; Slavic, fem. n. p., named after Nemodliki, tutelary deities of water in Bohemian and Moravian mythology), to formally denominate the taxonomic group previously known from 16S rRNA data as the CL500-3 clade.

The remaining 38 MAGs expanded the genomic representation of Planctomycetacia-the class that contains the bulk of cultured species (14 described genera at the time of writing) and considered (from a historical perspective) to encompass the planctomycetes *par excellence*. From them, 14 MAGs were found to form clusters affiliated to cultivated representatives (e.g. *Pirellula* spp., *Schlesneria paludicola*, *Planctomyces* spp., etc.), while the remaining 24 MAGs segregated in two coherent and divergent (from the other genomes and MAGs) groups within the class (Fig. 2). The first one comprises 9 MAGs (green box, Fig. 2) and branches in the proximity of *Gemmata/Zavarzinella* group, while the second (19 MAGs, blue box, Fig. 2) shares an evolutionary ancestry with *Blastopirellula/Pirellula/Rhodopirellula* clade and appears to be phylogenetically more related to a sediment-derived MAG cluster (Fig. 2). The 16 S rRNA phylogeny showed that both of these clusters (green and blue boxes, Fig. 2) fall under the umbrella rank "Planctomycetaceae uncultured" (i.e. Planctomycetaceae_uncultured G1 and Planctomycetaceae_uncultured G2; Supplementary Figure 4). As a consequence, based on within-group average amino-acid identity values (Supplementary Figure 5) we propose the creation of the families Nixeaceae (fam. Nov.; Germanic, fem. n., named after Nixe, aquatic being in Germanic folklore) (green box, Fig. 2) and Vodnikaceae (fam. nov.; Slavic, masc. n., named after Vodník, mythical Slavic water spirit) (blue box, Fig. 2) to accommodate the members of the 16 S rRNA groups Planctomycetaceae_uncultured G1 and Planctomycetaceae_uncultured G2, respectively.

Past and present of freshwater Planctomycetes explored by evolutionary genomics

The pattern of ancestry, divergence and descent (as shown by the phylogenomic trees, Figs. 2 and 3) indicated that Phycisphaerae and Planctomycetacia are sister lineages of a

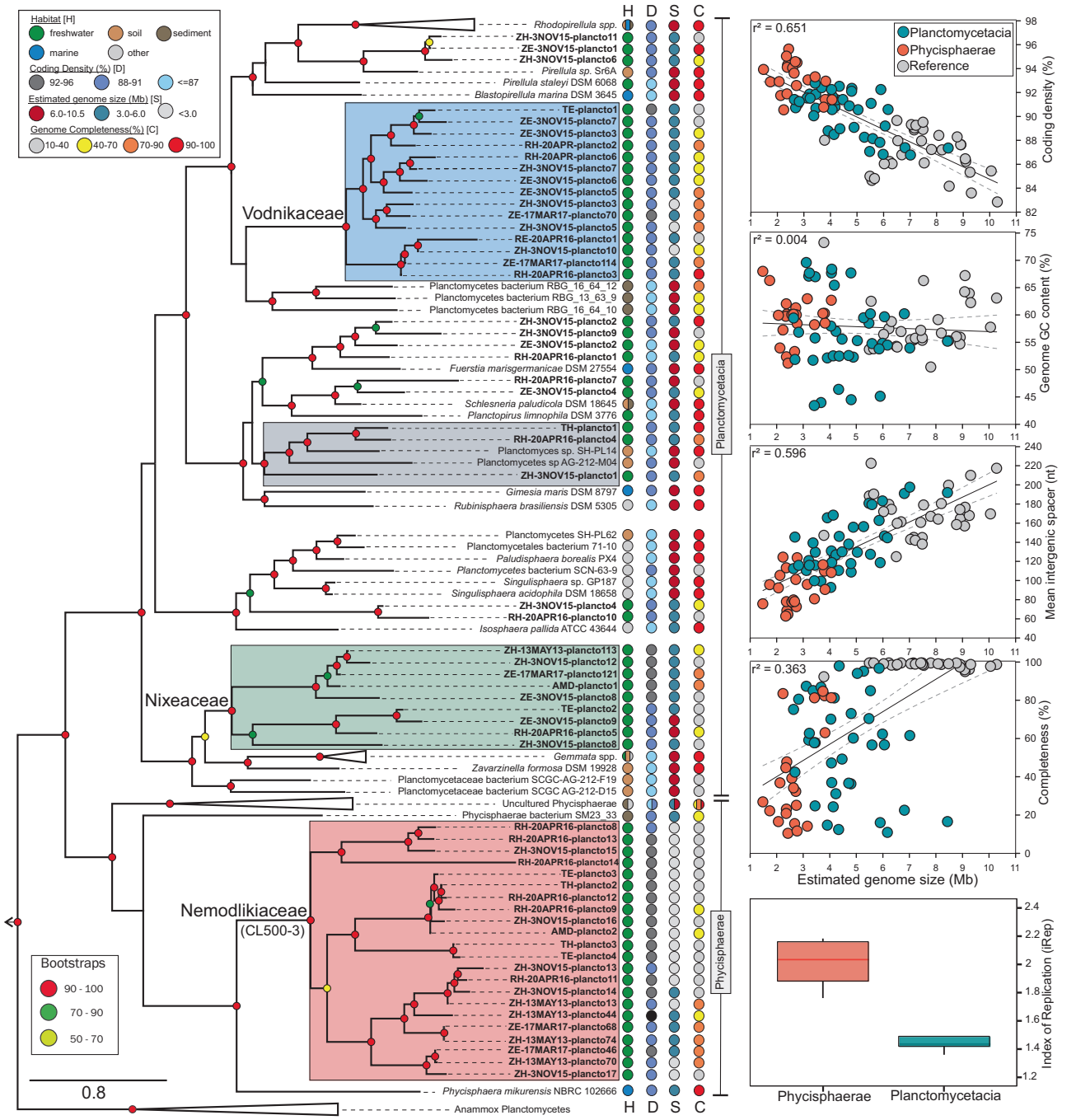

Fig. 2 Phylogenomics of Planctomycetes phyla. The left panel shows accurate whole-genome phylogenies through a maximum likelihood (phylogenomic) tree inferred from 138 genomes (complete and partial). The topology of the tree emphasizes the major phylogenomic groups found in lacustrine habitats (for details regarding tree inference see Methods). The names of the 60 metagenome-assembled genomes (MAGs), obtained in this study, are highlighted in boldface, while the culture-derived genomes (references) and other available MAGs are depicted in italic and roman type, respectively. The strength of support for internal nodes was assessed by performing bootstrap replicates, with the obtained values shown as colored circles (left legend). Ecological data (i.e., habitat of origin = H) and genomic characteristics (coding density = D, genome size = S, and completeness = C) are indicated by colored circles for each branch in the tree (top left legend). The relations between the genomic characteristics (i.e. estimated genome size, coding density, GC content, mean intergenic spacer length, genome completeness) of MAGs (Phycisphaerae and Planctomycetacia MAGs; see vertical taxonomic delineators) and reference Planctomycetes (31 culture-derived genomes) are shown by linear regressions in the 4 insets present in the right part of the figure. The lowermost insert (right side) shows the iRep values for Phycisphaerae ($n = 4$) and Planctomycetacia ($n = 9$) MAGs

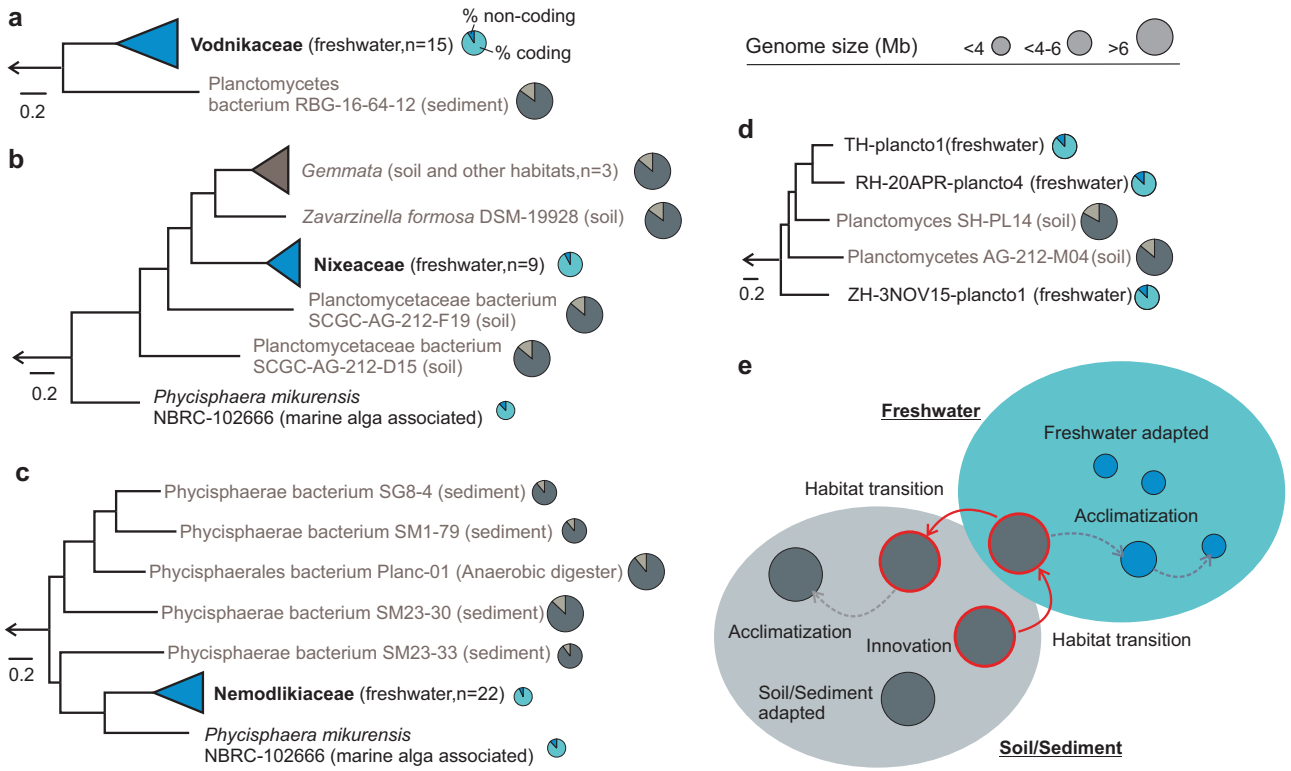

Fig. 3 Phylogenomic subtrees **a**–**d** generated using maximum-likelihood methods and alignments of concatenated conserved proteins (54, 20, 206 and 315 proteins). The black colored branches designate aquatic groups, while the grey ones their closest relatives (found in soil/sediments). The circular symbols, situated at the tips of the branches, are proportional with genome size and depict gene densities (within genomes). The number of genomes present in the collapsed groups is specified in parenthesis. **e** Putative model of niche-directed genome evolution in freshwater Planctomycetes

common ancestor which shared evolutionary relatedness with anammox planctomycetes (median genome size (MGS) 3.9 Mb, median intergenic spacer (MIS) 85 nt, median coding density (MCD) 86), bacteria that thrive at aerobic-anaerobic interfaces of sediment and water bodies [42].

We observed that the deep evolutionary history of Phycisphaerae is intrinsically linked to a sediment-specific lifestyle, as the basal branch of the class was found to accommodate bacteria (MGS 6.1 Mb, MIS 76.5, MCD 89%) that live in estuarine sediments ('Uncultured Phycisphaerae' group comprising 4 MAGs; Fig. 2, Fig. 3). Both Nemodlikiaceae (MGS 2.6 Mb, MIS 43.5 nt, MCD 92.9%) and its sister lineage, typified by *P. mikurensis* (GS 3.8, MIS 97 nt, CD 88%) appeared to be the descendants of an ancestor which underwent a habitat transition from sediments to an aquatic lifestyle (the node of the leaf Phycisphaerae bacterium SM23_33; GS 4.7 Mb, MIS 78, MCD 90%). Noteworthy, we observed that the adaptation to freshwater appears to be accompanied by a reduction in genome size (i.e. from 6.13 Mb in the deep branching sediment clade to 4.7 Mb in the sediment sister lineage of the aquatic branch and to 2.6 Mb in the freshwater Nemodlikiaceae) (Fig. 3). The Vodnikaceae (MGS 4.1 Mb, MIS 73 nt, MCD 91.2%) and Nixeaceae families (MGS 4.7 Mb, MIS 59 nt, MCD 92.2%) were found to be related to lineages comprising soil/sediment planctomycetes that (compared to them) harbor considerably larger genomes (7.1 Mb MGS for Vodnikaceae sister lineage and 9.2 Mb for Nixeaceae sister linage) (Fig. 3).

Taken together, these observations point to the fact that freshwater Planctomycetes (as typified by the families Nemodlikiaceae, Vodnikaceae and Nixeaceae) may possess a sediment/soil ancestry, and that during adaptation to the freshwater environment underwent substantial genome downsizing. On the assumption that this hypothesis is accurate, we would expect that (within a 'lower-rank' taxonomic category) a transition from sediment/soil environments into freshwater will be accompanied by a decrease in genome size and vice versa: a freshwater-sediment/soil transition will be accompanied by an increase in genome size. Furthermore, the freshwater-specific foliage of such a phylogenetic cluster (depicting transitions from freshwater to sediment/soil and vice versa) would be characterized by larger genomes (as compared to Nemodlikiaceae, Vodnikaceae and Nixeaceae), since they typify more recent habitat transitions. In line with the hypothesis, we identified in the phylogenomic tree a family-level clade (amino-acid

identity within group 52.2 -66.8 %) (grey box Figs. 2 and 3), in which the basal freshwater branch (ZH-3NOV15-plancto1, GS 5.3 Mb, MIS 82 nt, MCD 91.0%) is succeeded by a habitat transition to soil (Planctomyces sp. SCGC AG-212-M04: GS 6.9 Mb, MIS 107 nt, MCD 88%; Planctomyces sp. SH-PL14: GS 8.2 Mb, MIS 151 nt, MCD 83%) which is followed by a reversion to freshwater (RH-20APR-plancto4: GS 5.4 Mb, MIS 125 nt, MCD 87.0%; TH-plancto1: GS 5.2 Mb, MIS 107 nt, MCD 88.2%) (Fig. 3). Comparative functional genome analyses of this family-level clade revealed that all freshwater lineages have genes encoding for transporters involved in nitrogenous nutrients uptake (i.e. NitT/TauT transport system in ZH-3NOV15-plancto1, spermidine/putrescine transport system in RH-20APR16-plancto4 and Nitrate/Nitrite transport system in TH-plancto1). We argue that the presence of these uptake systems in the freshwater-recovered genomes and their absence in those from soil is linked to the different nitrogen acquisition strategies necessary for survival in freshwater ecosystems. Moreover, we observed that the soil representatives of the clade (i.e. Planctomyces sp. SCGCAG-212-M04 and Planctomyces sp. SH-PL14) have genes involved in cell-surface interactions (WspA, WspB/D, WspC, WspE, WspF and WspR), chemotaxis (MCP, CheW, CheA, CheR and CheB) and flagellar apparatus (29 genes involved in flagellar assembly) that are absent in genomes recovered from freshwater. Furthermore, we observed that habitat transitions (from sediment/soil to aquatic environments and *vice versa*) are scattered throughout the evolutionary history of Planctomycetacia (Fig. 2).

As the fitness of a prokaryotic cell (and its success in a heterogeneous environment) is generally considered to be dependent by its ability to modulate the gene expression patterns in response to fluctuating environmental stimuli (temperature, pH, ionic strength, light, etc.), we investigated the distribution of signal transduction systems (STS) across the recovered Planctomycetes genomes. The 60 MAGs were grouped in a phylogenetic fashion (Nemodlikiaceae, Nixeaceae and Vodnikaceae) with the exception of 14 MAGs that did not generate discriminable freshwater-specific clusters (i.e the 14 MAGs clustered together with cultivated representatives were grouped in 'Planctomycetacia_diverse'). The inventory of sigma factors, signal transduction domains (histidine kinase A, Per-Arnt-Sim and GGDEF domains) and PP2C-phosphatases, revealed that Nemodlikiaceae (Phycisphaerae) harbored a higher number of signal transduction pathways and genetic regulatory circuits per Mb of genome (in comparison to Nixeaceae, Vodnikaceae and Planctomycetacia_diverse) (Supplementary Figure 6). This inverse relation, in Nemodlikiaceae, (between STS/Mb and genome size) is unexpected since signal transduction systems and genome size are reported to positively correlate [43, 44]. Furthermore, in spite of harboring the largest genomes (MGS 5.3 Mb, MIS 109.5 nt, MCD 88.1%) within the 4 groups, Planctomycetacia_diverse was found to rank the lowest for GGDEF domains, and to respectively lack the histidine kinase A ones and PP2C phosphatases (Supplementary Figure 6). On the other hand, Planctomycetacia_diverse was found to contain the largest number of transposases (i.e. mobile genetic elements), which suggests an increased potential for genome plasticity and accelerated diversification through horizontal gene transfers and genomic rearrangements [45]. Taken together, the above observations suggest that signal transduction systems are critical components in the repertoire of freshwater Planctomycetes (that are retained in spite of genome shrinkage, increasing their genomic density) and may represent prerequisites for their survival and thriving in the lacustrine ecosystems. Moreover, we consider that the higher number of transposases found in Planctomycetacia_diverse may represent a genomic reminiscence that aided in habitat adaptation (Fig. 3), and that their low numbers of signal transduction systems (together with their genome size and phylogenomic position) may be an indication of a more recent transition to freshwater environments in comparison to Nemodlikiaceae, Nixeaceae and Vodnikaceae. The higher density of sigma factors observed in Nemodlikiaceae (Supplementary Figure 6, Supplementary Figure 7) could be the consequence of adaptation to a fundamentally heterogeneous niche.

Freshwater Planctomycetes across space and time

Differential genome coverage was used to estimate the fraction of the Planctomycetes populations undergoing active DNA replication. By taking advantage of the coverage bias in actively replicating populations (as more sequences are recovered from the regions proximal to the origin, rather than the terminus of replication) and single time-point metagenomic sequences, we used the iRep algorithm [46] to infer in situ replication rates. We stress that in a population in which the majority of the Planctomycetes are replicating the iRep value would be equal to 2. From the analyzed MAGs (13 MAGs that meet the iRep requirements: >= 75% complete, <= 175 fragments/Mbp sequence, and <= 4% contamination) we inferred that on average 44% of Planctomycetacia and all the Phycisphaerae (Nemodlikiaceae) cells were undergoing replication at the time of sampling (Fig. 2). Remarkably, the highest iRep values were registered for ZE-17MAR17-plancto46 (iRep = 2.1) and ZH-13MAY13-plancto70 (iRep = 2.1), MAGs belonging to the same species (Supplementary Figure 5) that were recovered at a four year-interval (from epilimnion and hypolimnion of Lake Zurich, respectively). The fact that the two MAGs had similarly high replication indexes,

at different time points, suggests they represent a fast-growing genotype that is persistent and successful in the lacustrine habitats. Although, the low number of observations (4 for Phycisphaerae and 9 for Planctomycetacia) precludes generalization, it seems (from the available data) that the Phycisphaerae MAGs (i.e. Nemodlikiaceae) have higher rates of replication in the freshwater environments (within the analysed freshwater Planctomycetes clades).

The biogeographic distribution of the 60 Planctomycetes MAGs was assessed in 64 lacustrine freshwater habitats scattered over three continents (Supplementary Figure 8). The results corroborated well with the 16S rRNA short-read taxonomic profiles and highlighted that, in general, the MAGs achieve higher 'abundances' in the habitat of origin, and scarcely few of them (e.g., AMD-plancto2, RH-20APR16-plancto14, ZH-3NOV15-plancto16, RE-20APR16-plancto1, TE-plancto2, ZH-3NOV15-plancto11) were well-represented in other European lakes. Considering that the majority of MAGs show a restricted geographic dispersal indicates that (in this case) the lakes' low habitat connectivity supported a distributional pattern governed by a distance-decay relationship.

A quantitative dimension of freshwater Planctomycetes revealed by CARD-FISH imaging

We made use of the CARD-FISH technique to monitor the yearlong spatio-temporal distribution of Planctomycetes in Lake Zurich and Římov Reservoir throughout 2015. Hence, ten CARD-FISH probes were designed using the 16S rRNA gene sequences recovered from MAGs and additional publicly available sequences. Seven probes were constructed to target groups from which MAGs were available and another three were designed to quantify Planctomycetes groups that were found to be abundant in the metagenomic 16 S rRNA gene pool but from which MAGs were not recovered (Supplementary Figure 4, Extended Data).

Nemodlikiaceae (class Phycisphaerae) numerically surpassed the other detected Planctomycetes with the exception of thermal stratification events when, in the warmer epilimnion, members of class Planctomycetacia prevailed (Supplementary Figures 9 and 10). We observed that the uniform abundance patterns of Nemodlikiaceae (Phycisphaerae) that were displayed within the water column during mixing in Lake Zurich, became skewed during stratification (Summer and Autumn distributions), when the group's numbers declined in the epilimnion (Fig. 4). A similar trend in spatial and temporal distribution was also detected in Římov Reservoir, where Nemodlikiaceae's contribution to prokaryotic communities was at its lowest in the strata above the thermocline (Fig. 4). Furthermore, Nemodlikiaceae (i) maintained its high numbers in the surface strata long after the end of mixing events (6.6-7% in April, Lake Zurich; 2.7-2.8% in April, Římov Reservoir), (ii) reached higher abundances in strata below the thermocline during stratification periods (Fig. 4) and (iii) registered seasonal peaks in abundances at low water temperatures (median temperature for seasonal peak in prokaryotic communities is 5.3 °C). By taking into account the above-mentioned effect of lake stratification-mixes cycles, we consider that Nemodlikiaceae are composed of habitat specialists (recording significant abundances in freshwaters, Fig. 1) that show a trend towards psychrotrophic behavior (*sensu* Gounot) [47]. This is in line with previous studies that reported high abundances of this lineage in the deep and cold hypolimnion of several Japanese lakes [27, 31]. Nemodlikiaceae were found both free-living and attached to lake snow particles and had the smallest cell sizes of all analyzed Planctomycetes (ovoid shape, length 0.4 µm, width 0.3 µm; Fig. 4, Supplementary Figure 11 and Extended Data).

The overall contribution of Planctomycetacia to prokaryotic communities in Zurich Lake and Římov Reservoir was generally low (Fig. 4). The most abundant group detected was Pirellula-like, which mostly maintained subunitary contributions in the prokaryotic assemblages (Supplementary Figures 9 and 10, probe pir-663). In both lakes, the contribution of Pirellula-like microbes peaked in the surface strata during stratification (1.4%, 5 m, April, Římov Reservoir; 1.2%, 0.1 m, June, Lake Zurich), concomitantly with phytoplankton blooms (mostly green algae) as inferred by high chlorophyll *a* values. These microbes were also found to colonize lake snow particles and were slightly larger than Nemodlikiaceae (rod-shaped, length 0.51 µm, width 0.39 µm, Supplementary Figure 11 and Extended Data).

Life in the lacustrine realm

Here, we explore the nature of Planctomycetes-environment interactions in a reductionist fashion centered on survival-reproduction strategies. Thus, our niche inferences stem from the means employed by bacteria to probe the physico-chemical landscape (e.g., respond to chemical gradients and uptake nutrients), since they typify ecological strategies (for increasing fitness) and allow general behavioral predictions.

We reason that Planctomycetes in lacustrine environments may adopt dual lifestyles (free-living and surface attached) since some lineages were microscopically observed to colonize particles (Fig. 4) and they possess the capacity for both motility and adherence encoded in their genomic repertoire (Extended Data). Thus, while the presence of WspE-WspRF (all groups) and FlrB-FlrC (only Nemodlikiaceae) two-component systems may regulate surface affinities [48, 49], the flagellar apparatus (present in Nemodlikiaceae and 3 MAGs from

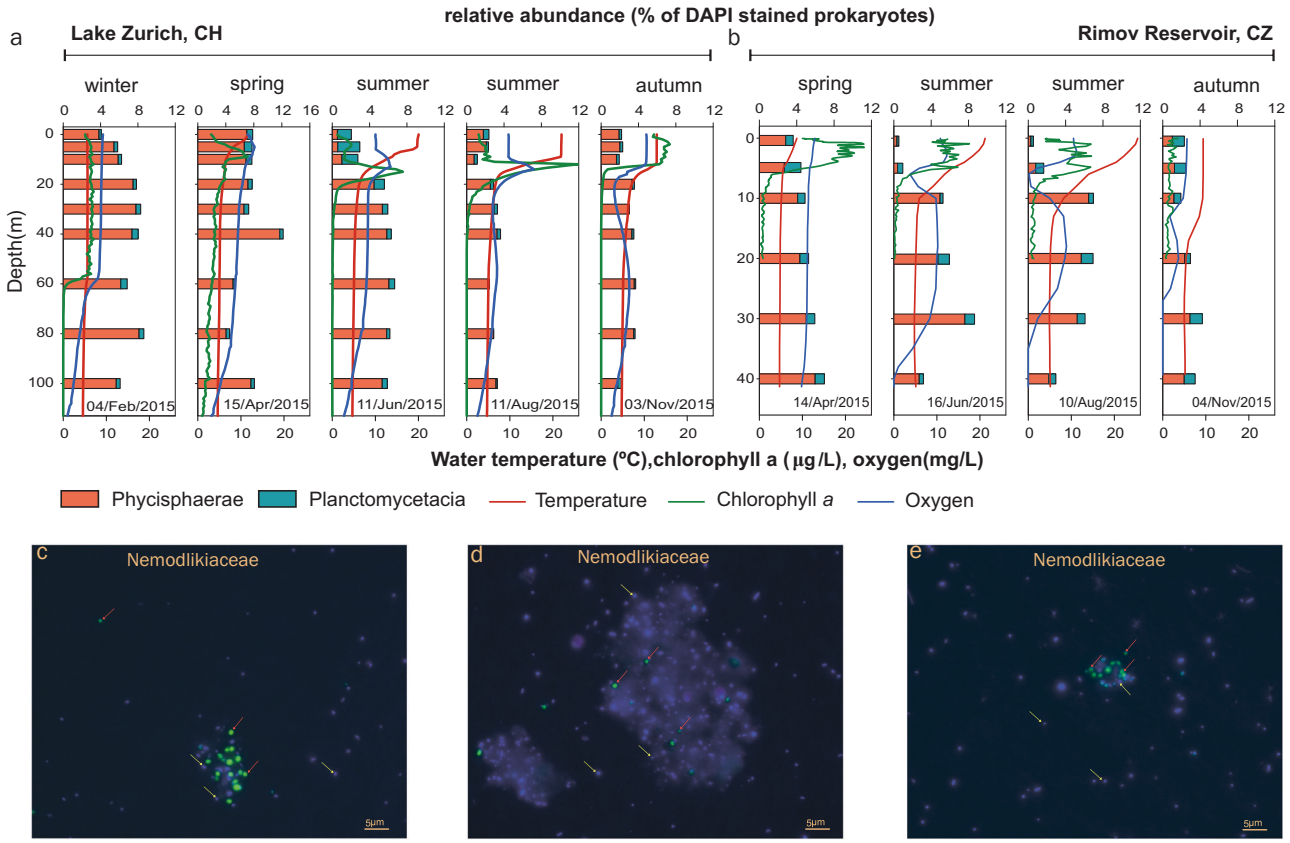

Fig. 4 Spatio-temporal profiles of Planctomycetes relative abundance (horizontal bars), temperature (red line), chlorophyll *a* (green line) and oxygen (blue line) in Lake Zurich (**a**) and Rimov Reservoir (**b**) during 2015. The vertical axis shows the depth (m), within the water column, from which the samples were collected (9 for Lake Zurich and 6 for Rimov Reservoir). The upper X-axis shows the percentage of Phycisphaerae (red bars) and Planctomycetacia (dark cyan) within the prokaryotic communities (estimated as the total sum of DAPI-positive cells), while the lower one displays the values for temperature, chlorophyll *a* and oxygen. The sampling date is shown above the lower X-axis. **c**–**e** Display superimposed images of CARD-FISH-stained Planctomycetes (class Phycisphaerae, family Nemodlikiaceae) and DAPI-stained prokaryotes. The red arrows point towards free-living and particle-associated Planctomycetes, while the yellow ones designate unhybridized prokaryotic cells. The scale bar is 5 µm

Planctomycetacia_diverse) suggests directional swimming (Extended Data). Noteworthy, some of the genomes of Planctomycetacia (Planctomycetacia_diverse) were found to encode additional genes involved in adherence/surface colonization (tad cluster and type IV pili).

The genome-scale metabolic reconstructions, performed on the 60 Planctomycetes MAGs, revealed a typical heterotrophic metabolism in which beta-Oxidation, the hexose monophosphate shunt and glycolysis (incomplete in Nemodlikiaceae) fuel the tricarboxylic acid cycle and oxidative phosphorylation (Extended Data). We observed that while the core metabolism was highly similar between Phycisphaerae and Planctomycetacia the substrate uptake capacity showed phylogenetic segregation. Thus, albeit glucose (through porin OprB), ribose, nucleosides and 3-phenylpropionic acid uptake was inferred to be common in both Planctomycetes classes, the preferences towards monosaccharides and organic acids showed group specificity. Accordingly, we found that the uptake of hexoses (i.e., L-rhamnose, L-fucose, D-glucose/D-mannose, D-gluconate), modified monosaccharides (i.e., glycoside/pentoside/hexuronide and L-arabinose) and organic acids (glucarate, hexuronate, lactate, oxalate) was favored in Planctomycetacia, while D-fructose was preferred in Phycisphaerae (Nemodlikiaceae). Moreover, even though the uptake systems for amino acids (polar, basic and branched-chain) and oligopeptides were found to be common across both lacustrine classes, they were more abundant in Phycisphaerae (Nemodlikiaceae) (4.5 vs. 2.0 transporter components/MB). Although the presence of ammonium/ammonia (the preferred nitrogen source for microbial growth) transport channels (i.e., AmtB) was a common feature within Planctomycetacia, they were not detected in Phycisphaerae, thus, Nemodlikiaceae may lack ammonium/ammonia uptake capacity. Additionally, the enzymatic repertoire necessary for pyrimidines and amino acids (i.e., methionine, leucine, tryptophan and histidine) degradation was present exclusively in Nemodlikiaceae, implying an important role

of these compounds in fueling their metabolic machinery. We detected that the amino acid biosynthetic pathways were also distributed unequally among the phylogenetic groups and that while both classes were auxotrophic for methionine, phenylalanine and tyrosine, Nemodlikiaceae suffered additional impairments in the synthesis of threonine, valine/isoleucine, leucine and proline (Extended Data). Evidence for sulfate transport (through ABC transporters and SulP permease family) was found to be present only in Planctomycetaceae, where the assimilatory reduction pathway was inferred to be complete. The widespread capacity to regulate (through PhoR-PhoB two-component system) the high-affinity acquisition of inorganic phosphate (through phosphate-selective porins OprO and OprP) pointed towards a phylogenetically conserved strategy among all lacustrine Planctomycetes (Extended Data).

Inter alia, we inferred that Planctomycetes cellular membranes are dotted by mechanosensitive channels (both large- and small-conductance) that could jettison cytoplasmic solutes during hypo-osmotic conditions, and that Nemodlikiaceae intriguingly decorate their external surfaces with sialic acids. Noteworthy, we detected the presence of five green-light rhodopsins (one in Phycisphaerae and 4 in Planctomycetacia, Supplementary Figure 12) and CO dehydrogenases (form II; Planctomycetacia) that may be involved in energy conservation through generation of proton motive force. We found that members of Planctomycetacia (Vodnikaceae, Nixeaceae and Planctomycetacia_diverse) have the capacity to enhance their fitness and increase their niche persistence by antagonizing their (non-self) neighbors with lethal, toxin injecting devices (i.e. type VI secretion systems) and bacteriocins (Planctomycetacia_diverse) (Extended Data).

Surprisingly, we found that Nemodlikiaceae genomes encode cohesin/dockerin modules (signature-domains of cellulosome, Supplementary Figure 13) that did not fit in the established cellulosome model [50] since we found no evidence for their involvement in cellulose degradation. Thus, we hypothesize that Nemodlikiaceae may use instead a non-canonical cellulosome-like machinery to degrade polypeptides, for which we tentatively propose the term "planctosome" (Fig. 5). This putative structure resembles in its complexity the mesophile's simple cellulosome systems [51] and supports a new non-cellulosomal function [51, 52] for the high-affinity cohesin-dockerin interactions. By combining homology-, motif- and structure-based methods with protein domain co-occurrence (Supplementary Figure 14), we consider that the lamin tail domain-containing protein facilitates peptidases anchoring on the outer membrane and activation through the proprotein convertase P-domains, while the cohesin/dockerin containing one facilitates substrate binding through a hyaline repeat domain (Fig. 5).

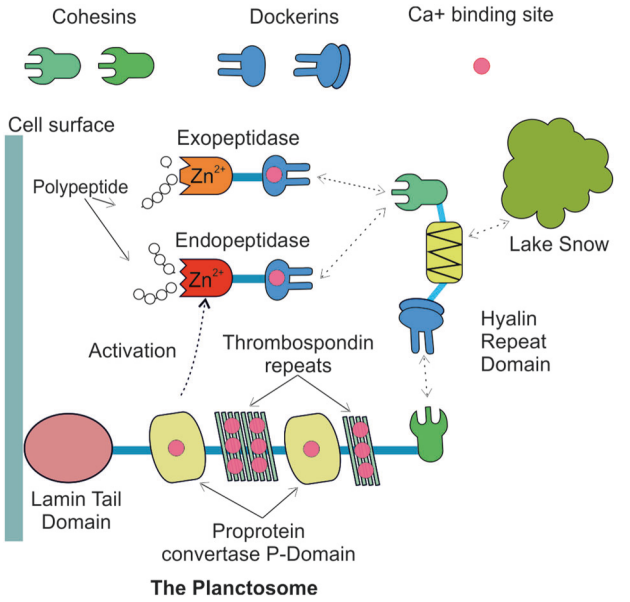

Fig. 5 Hypothetical model of multiprotein complex (planctosome) involved in peptide degradation. The complex is tethered to extracellular membrane through a lamin A/C globular tail domain (LTD). The "anchoring" protein (2,210 aa) consists of a N-terminus signal peptide (26 aa) followed by the LTD, multiple proprotein convertase P-domains (PCD) divided by thrombospondin type 3 repeats (TSP), and a cohesin domain (CD). The "adaptor" protein contains a N-terminus signal peptide (26 aa), a dockerin domain (DD), a hyaline repeat domain (HYRD) and a cohesin (CD). The "adapter" binds Zn^{2+}-dependent endo- (M12B Reprolysin4-like) and exopeptidases (M14 carboxypeptidase subfamily A) through Ca^{2+}-dependent cohesin-dockerin interactions

Conclusion

While the performed large-scale taxonomic profiling (based on 298 metagenomic data sets; Fig. 1 and Supplementary Figure 1) showed the existence of a lacustrine Planctomycetes blueprint, the in situ spatio-temporal abundance patterns and metabolic reconstructions pointed towards lineage-specific lifestyles. Thus, we observed that members of the Nemodlikiaceae (i.e. the hallmark lacustrine Planctomycetes lineage) exhibit psychrotrophic tendencies as they prefer to colonize the deeper and/or colder water strata, where they locate (by using signal transduction systems and flagella) and mineralize (through a highly-tuned metabolism) the nitrogen-rich sinking aggregates (lake snow, Fig. 4). By contrast, Planctomycetacia (e.g., Vodnikaceae and Nixeaceae) showed preferences towards shallower and warmer water layers, where their versatile heterotrophic metabolism is fueled by phytoplankton-derived dissolved organic matter.

Genomic shrinkage, once considered a characteristic of symbiotic microorganisms [53], was found to be widespread in specific environmental niches [54] and phylogenetic lineages of free-living bacteria [55–58]. While it is

generally assumed that bacteria with reduced genomes evolved from lineages with larger ones [58], the evolutionary routes to minimalism appeared to be lifestyle-dependent (i.e. host-associated and free-living). While genome reduction in host-associated bacteria is generally complemented by massive gene losses, low-coding densities and even impairment of basic metabolism [53], in the free-living ones it is typified by high-coding densities and preservation of pathways involved in cellular growth and replication [59]. Bacterial lineages with reduced genomes dominate planktonic communities in both freshwater and marine ecosystems [54, 55]. Some of the typical features of these abundant groups are small genomes (1.1–1.6 Mb) with low GC content (29–35%GC) that are characterised by high coding densities and short intergenic spacers [56]. The streamlining selection hypothesis that emerged in order to explain these observations is based upon the idea that selection favours genome reduction in lineages with large population sizes that thrive in nutrient limited environments [56]. Thus, the primary mechanism invoked for genome reduction is metabolic efficiency which could be achieved through resource management (selection acts to reduce the amount of nitrogen required for cell replication) or efficient nutrient uptake (selection acts to reduce cell size and increases the surface-to-volume ratio) [56]. While this hypothesis seems to fit well to abundant marine planktonic microbes, it falls short in explaining the 'streamlining' of prokaryotic lineages in environments where nitrogen is not a limiting nutrient (e.g. freshwater environments, where similarly large population sizes are observed for streamlined microbes, e.g. acI Actinobacteria [57], *Methylopumilus* [58]). In contrast to the positive selection model favoured by the streamlining hypothesis, genome-wide analyses have also shown the importance of genetic drift in bacterial genome size reduction [59, 60]. In spite of the fact that the freshwater Planctomycetes lineages have smaller genome sizes, shorter intergenic spacers and higher coding density that their soil/sediment relatives (Fig. 3), their dimensions and genomic characteristics (e.g., higher cell sizes, genome length and GC content) (Extended Data) render them unfit in the current streaming theory of genome reduction. Although freshwater Planctomycetes genomes do not appear to be streamlined, they seem to share analogous life histories with some streamlined bacterial lineages [56, 57, 60]. For instance, parallels could be drawn between the habitat transition of Planctomycetes (from sediment/soil to freshwater) and the niche partitioning of *Prochlorococcus* genotypes in the tropical and subtropical ocean [56, 57]. While in *Prochlorococcus* a niche transition from the lower (i.e. low-light IV clade) to the upper part of the euphotic zone (i.e. low-light 1) is accompanied by a reduction in genome size (approx. 35.6%) [56], in Planctomycetes a habitat transition from soil to freshwater (Fig. 3d) is associated with a similar size decrease (approx. 30%).

Remarkably, the most abundant lacustrine-specific Planctomycetes lineage (i.e. Nemodlikiaceae) had simultaneously the smallest genome sizes with highest coding densities and the most specialized lifestyle, suggesting niche-directed genome evolution. Thus, we consider that in Nemodlikiaceae genetic drift may have fine-tuned their metabolic circuitry and decreased their genome size towards the minimum needed for efficient niche exploitation (selection of features necessary to colonize and utilize sinking aggregates; loss of biosynthetic pathways for molecules available in the niche). In line with the evolutionary history inference obtained by phylogenetic reconstruction, we suggest a scenario in which sediment/soil Planctomycetes transitioned to aquatic environments where they give rise to new habitat-specific lineages (e.g., lacustrine-specific). By corroborating our results with recent phylogenetic reconstructions of abundant freshwater bacterial lineages (i.e., Betaproteobacteria and Verrucomicrobia) [61, 62], we consider that the above-mentioned evolutionary path in which ancient soil/sediment transitions are steered by the niche towards genome reduction may be wide-spread in freshwater ecosystems.

Materials and Methods

Sampling and Sequencing

The meso-eutrophic Římov Reservoir (470 m a.s.l, 48° 50'N, 14°29′E, Czech Republic) is a canyon-shaped dimictic water body with an area of 2.0 km^2 (length 13.5 km, volume of 34.5×10^6 m^3, mean retention time 77 days, maximum depth 43 m), that was built during 1974–1979 by damming a 13.5 km long section of the River Malše [63]. The sampling was performed during Spring 2016 (20 April), above the deepest point of the reservoir by using a Friedinger sampler. Two multi-parametric probes were deployed in order to profile the physicochemical characteristics of the water column (temperature, pH, oxygen; GRYF XBQ4, Havlíčkův Broc, CZ) and chlorophyll *a* (FluoroProbe TS-16-12, bbe Moldaenke, Kiel, Germany). 10 L of water were collected from 0.5 and 30 m depths and subjected to sequential peristaltic filtration through a series of 20, 5 and 0.2-µm-pore-size polycarbonate membrane filters (⌀ 142 mm) (Sterlitech Corporation, USA). The DNA was extracted from the 0.2 to 5-µm fraction, as described elsewhere [64] and subjected to deep shotgun sequencing (paired end, 150 bp) on Illumina's HiSeq 4 000 platform (BGI, Hong Kong).

The oligomesotrophic Lake Zurich (406 m a.s.l, 47° 18'N, 8°34'E, Switzerland) is a perialpine, monomictic

water body, with an area of 67.3 km^2 (length 40 km, volume 3.3 km^3, mean retention time 1.4 years, maximum depth 136 m). The sampling was conducted during an ongoing fortnightly monitoring program at the deepest point of the lake [65]. Vertical profiles of temperature, conductivity, turbidity, and oxygen were recorded with a YSI multiprobe (Yellow Springs Instruments, model 6 600) and the chlorophyll *a* concentration was measured with a submersible fluorescence probe (FluoroProbe TS-16-12, bbe Moldaenke, Kiel, Germany). Water samples from the following depths were collected with a Friedinger sampler and processed for sequencing: 5 and 80 m (13th May 2013), 5 and 80 m (3rd November 2015) and 2 m (17 March 2017), respectively. Approx. 1–2 L of water was sequentially filtered onto 5 and 0.2-µm-pore-size filters, and the genomic DNA was extracted from the 0.2 µm filter one by using the PowerBiofilm DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA). Library preparation of 550-bp fragments was done with a KAPA Hyper Prep Kit (Kapa Biosystems, Wilmington, MA, USA) and deep metagenomic sequencing (paired-end, 150 bp) was carried out on a HiSeq 2 000 instrument at the Functional Genomics Center Zurich.

Classification of shotgun 16S rRNA gene fragments

FASTQ files (recovered from 298 environmental metagenomes: 64 lacustrine, 36 fluvial, 158 marine and 40 freshwater sediments, Extended Data) containing aquatic/sediment-derived raw shotgun reads, produced by second-generation sequencing platforms, were quality-filtered by a combination of bbduk.sh (adapter trimming and contaminant filtering) [66], bbmerge.sh (*de novo* adapter identification) [67] and sickle (quality trimming) [68]. Subsequently, they were converted to FASTA format and subsampled to 10 million sequences using reformat.sh [69]. These subsets (containing 10 million sequences each) were screened to identify RNA-like sequences by using UBLAST [70] against a non-redundant version of RDP database [71], which was previously clustered at 85% sequence identity by UCLUST [70] and contained 7 552 sequences with a length ≥ 800 bp. The sequences that matched the RDP database at an *E* value < 1e-5 were considered candidate 16S rRNA gene sequences and screened using SSU-ALIGN [72]. The *bona fide* 16S rRNA gene sequences (as identified by SSU-ALIGN) were further compared by BLAST [73], in nucleotide space (using as cutoff the E-value 1e-5), against a curated SILVA SSU database [74] that contained 447 012 sequences, and classified if the sequence identity was ≥80% and the alignment length was ≥ 90 bp (sequences failing these thresholds were not used for downstream analyses).

Assembly and binning

Ten lacustrine shotgun metagenomic datasets generated from lakes with contrasting trophic states (i.e. Lake Zurich and Římov, Tous and Amadorio reservoirs) were used for in-depth analyses. The metagenomic datasets derived from the Spanish freshwater reservoirs (i.e. Tous and Amadorio) were recovered from NCBI's SRA database, under the accession numbers SRR1173821 (Amadorio), SRR4198666 and SRR4198832 (Tous). Seven shotgun metagenomic libraries, generated from Římov Reservoir ($n = 2$) and Lake Zurich ($n = 5$), were sequenced during this study. All raw metagenomic sequences were filtered to remove low quality bases/reads as mentioned above, by using a combination of bbduk.sh [66], bbmerge.sh [67] and sickle [68] (Římov Reservoir data sets) or trimmomatic [75] (Lake Zurich data sets). The obtained high quality sequences were then assembled independently with MEGAHIT v1.1.1 [76] using the parameters: --min-count 2 and --k-step 10 (k-mer range was 31–99 for the Tous and Amadorio data sets, and 31-149 for the Římov data sets) or metaSPAdes [77] (k-mer range 21–127) for the Lake Zurich data sets.

Phylogenomics

In order to investigate if the obtained 60 Planctomycetes MAGs were identical to previously described ones, we performed genome distance estimations, using Mash software [78] (with the parameters k-mer 25 and sketch size 5 000), against the Planctomycetes genomes publicly available in NCBI Genome database (102 entries in May 2017).

The average MAG coverage depth (defined as the average number of reads covering a base pair in the reference MAG) was computed by using BBMap version 36.19 (with default settings) [79] and quality-trimmed metagenomic reads. In order to estimate the abundance of each MAG within and between metagenomes, we calculated RPKG values (i.e., the number of reads recruited per kilobase of genome per gigabase of metagenome) using an in-house pipeline. Briefly, in order to avoid analysis bias, we concatenated the contigs belonging to each MAG and masked all the rRNA gene sequences present. Subsequently, BLASTN [73] (with the cutoffs: alignment length ≥ 50 nt, identity > 95%, *E* value < = 1e-5) was used in order to align the quality-filtered shotgun reads (20 million reads each from 64 freshwater metagenomic data sets) against the 60 Planctomycetes MAGs. The obtained BLAST best-hits results were further used to compute RPKG values. In order to assess the genetic diversity of the Planctomycetes populations, we used blast-tools [46] to plot

the best-hits results generated by performing BLASTN (with 180 million sequences equally sub-sampled from 10 metagenomes generated from Lake Zurich and, Římov, Tous and Amadorio reservoirs; alignment length ≥ 100 nt, identity > 90%, *E* value < = 1e-5) against the 60 MAGs.

In order to establish the evolutionary relationships among the 60 MAGs (with variable degrees of genome completeness) and previously available Planctomycetes genomes (in NCBI Genome repository), we carried out a phylogenomic analysis using PhyloPhlAn [80]. Briefly, the CDSs predicted in Prodigal's metagenomic mode [81] were translated to protein sequences and screened for the presence of 400 universally conserved and phylogenetically discriminating proteins (found in PhyloPhlAn database) by USEARCH [70] (E-value < 1e-40). The minimum number of proteins used was 35 (for ZH-3NOV15-plancto17), the maximum 319 (RH-20APR16-plancto1) and the median 170. The homologs of these proteins were independently aligned by MUSCLE [82], concatenated and further used in generating a maximum likelihood tree with FastTree software (JTT + CAT model) [83]. Subtrees were constructed by concatenating and aligning conserved proteins as described elsewhere [84].

The average amino acid identity (AAI) within coherent phylogenomic groups was determined by performing whole-genome pairwise CDSs comparisons, using BLAST, as previously described by Konstantinidis and Tiedje [85]. Taxonomic categories for the MAGs were defined using the standards suggested by Konstantinidis et al. [40].

Planctomycetes in situ replication rates were determined based on measuring the rate of the decrease in average sequence coverage across all genomic fragments (present in one MAG), by using iRep [46]. Briefly, quality-filtered shotgun reads were mapped against the MAGs (>= 75% complete, <= 175 fragments/Mbp sequence, and < = 4% contamination) recovered from the same metagenome by Bowtie 2 (version 2.3.4) [86] with --very-sensitive option. The obtained mapping files, in SAM format, were used for calculating an index of replication (iRep) based on the sequencing coverage trend that results from bi-directional genome replication from a single point of origin as described by Brown et al. [46].

Genome annotation

MAGs *de novo* gene predictions were performed by Prokka [87]. BlastKOALA [88] was used to assign KO identifiers (K numbers) to orthologous genes present in the 60 MAGs. The K numbers were further mapped to KEGG pathways, BRITE hierarchies, and KEGG modules for inferring the systemic functions of individual MAGS. The annotations were further refined by using the standard operation procedures from the Rapid Annotations using Subsystems Technology server [89]. Additional gene annotations were performed by protein sequence searches (using hmmscan with E-value 1e-5) [90] against the HMM databases COG [91] and TIGRfam [92]. The carbohydrate-active enzymes were annotated using the dbCAN-seq database [93]. Several protein sequences were further analyzed using jackhmmer [94] and Phyre2 [95].

Phylogenetics

The 16S rRNA gene sequences present in the MAGs were identified by SSU-ALIGN, aligned by SINA (https://www.arb-silva.de/aligner/), imported in ARB software [96] using the SILVA SSU Ref 123 database and manual refinements of alignments, and used for the construction of a RAxML [97] maximum likelihood tree (100 bootstraps, GTRGAMMA model). The rhodopsin sequences identified by HMMER [90] were aligned with MAFFT under L-INS-i model [98], and used for a maximum likelihood tree construction (100 bootstraps) with FastTree2 [83].

Probe design and CARD-FISH

The 16S rRNA gene sequences present in MAGs as well as 16S rRNA sequences extracted from the raw metagenomics reads were used for probe design for fluorescence in situ hybridization followed by catalyzed reporter deposition (CARD-FISH) (see Supplementary information).

Accession numbers

All sequence data produced during the study is deposited in the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) and could be found linked to the Bioprojects PRJNA429141 (Římov Reservoir) and PRJNA428721 (Lake Zurich). All MAGs used in this study can be accessed under the Bioproject PRJNA449258 (accession numbers: QWOG00000000-QWQN00000000).

Acknowledgements We thank E. Loher and T. Posch for help with sampling of Lake Zurich and S. Neuenschwander for help with metagenomic library preparation for Lake Zurich samples. A-Ş.A was supported by the research grants: 17-04828S (Grant Agency of the Czech Republic) and MSM200961801 (Academy of Sciences of the Czech Republic). RG was supported by the research grant 17-04828S (Grant Agency of the Czech Republic). MM was supported by the Postdoctoral program PPPLZ (application number L200961651) provided by the Academy of Sciences of the Czech Republic.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Woese CR, Stackebrandt E, Macke TJ, Fox GE. A phylogenetic definition of the major eubacterial taxa. Syst Appl Microbiol. 1985;6:143–51.
2. Staley JT. Budding bacteria of the Pasteuria–Blastobacter group. Can J Microbiol. 1973;19:609–14.
3. Garrity GM, Holt JG. The road map to the manual. Bergey's manual® of systemic bacteriology. p. 119–66, (Springer, New York, NY, 2001).
4. Erko S, Ludvig W, Schubert W, Klink F, Schlesner H, et al. Molecular genetic evidence for early evolutionary origin of budding peptidoglycan-less Eubacteria. Nature. 1984;307:735–7357.
5. Devos DP, Reynaud EG. Intermediate steps. Science (80-). 2010;330:1187–8.
6. Fuerst JA, Sagulenko E. Beyond the bacterium: Planctomycetes challenge our concepts of microbial structure and function. Nat Rev Microbiol. 2011;9:403–13.
7. Gimesi. I Planctomyces Bekefii Gim. nov. gen. et sp. (Ein neues Glied des Phytoplanktons.). Hydrobiologiai Tanulmiinyok (Hydrobiologische Studien). Budapest: Kiadja a Magyar Ciszterci Rend; 1924.
8. Hirsch P. Two identical genera of budding and stalked bacteria: Planctomyces Gimesi 1924 and Blastocaulis Henrici and Johnson 1935. Int J Syst Bacteriol. 1972;22:107–11.
9. Bauld J, Staley TJ. Planctomyces maris sp. nov.: a marine isolate of the planctomyces-blastocaulis group of budding bacteria. Microbiology. 1976;97:45–55.
10. Schmidt JM, Starr MP. Morphological diversity of freshwater bacteria belonging to the Blastocaulis-planctomyces group as observed in natural populations and enrichments. Curr Microbiol. 1978;1:325–30.
11. Schlesner H, Rensmann C, Tindall BJ, Gade D, Rabus R, Pfeiffer S, et al. Taxonomic heterogeneity within the Planctomycetales as derived by DNA-DNA hybridization, description of Rhodopirellula baltica gen. nov., sp. nov., transfer of Perillula marina to the genus Blastopirellula gen. nov. as Blastopirellula marina comb. nov. an. Int J Syst Evol Microbiol. 2004;54:1567–80.
12. Fukunaga Y, Kurahashi M, Sakiyama Y, Ohuchi M, Yokota A, Harayama S. Phycisphaera mikurensis gen. nov., sp. nov., isolated from a marine alga, and proposal of Phycisphaeraceae fam. nov., Phycisphaerales ord. nov. and Phycisphaerae classis nov. in the phylum Planctomycetes. J Gen Appl Microbiol. 2009;55:267–75.
13. Lindsay MR, Webb RI, Fuerst JA. Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus pirellula. Microbiology. 1997;143:739–48.
14. König E, Schlesner H, Hirsch P. Cell wall studies on budding bacteria of the Planctomyces/Pasteuria group and on a Prosthecomicrobium sp. Arch Microbiol. 1984;138:200–5.
15. Lonhienne TGA, Sagulenko E, Webb RI, Lee K-C, Franke J, Devos DP, et al. Endocytosis-like protein uptake in the bacterium Gemmata obscuriglobus. Proc Natl Acad Sci. 2010;107:12883–8.
16. Glockner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, et al. Complete genome sequence of the marine planctomycete Pirellula sp. strain 1. Proc Natl Acad Sci. 2003;100:8298–303.
17. Guo M, Zhou Q, Zhou Y, Yang L, Liu T, Yang J, et al. Genomic evolution of 11 type strains within family Planctomycetaceae. PLoS ONE. 2014;9:e86752.
18. Delmont TO, Quince C, Shaiber A, Esen OC, Lee STM, Lucker S, et al. Nitrogen-fixing populations of planctomycetes and Proteobacteria are abundant in the surface ocean. bioRxiv. 2017;3:129791.
19. Reva O, Tümmler B. Think big - Giant genes in bacteria. Environ Microbiol. 2008;10:768–77.
20. Mcinerney JO, Martin WF, Koonin EV, Allen JF, Galperin MY, Lane N, et al. Planctomycetes and eukaryotes: a case of analogy not homology. Bioessays. 2011;33:810–7.
21. Boedeker C, Schüler M, Reintjes G, Jeske O, Van Teeseling MCF, Jogler M, et al. Determining the bacterial cell biology of Planctomycetes. Nat Commun. 2017;8:14853.
22. Jeske O, Schüler M, Schumann P, Schneider A, Boedeker C, Jogler M, et al. Planctomycetes do possess a peptidoglycan cell wall. Nat Commun. 2015;6:7116.
23. Neef A, Amann R, Schlesner H, Schleifer KH. Monitoring a widespread bacterial group: in situ detection of planctomycetes with 16S rRNA-targeted probes. Microbiology. 1998;144:3257–66.
24. Gade D, Schlesner H, Glockner FO, Amann R, Pfeiffer S, Thomm M. Identification of Planctomycetes with order-, genus-, and strain-specific 16S rRNA-targeted probes. Microb Ecol. 2004;47:243–51.
25. Buckley DH, Huangyutitham V, Nelson TA, Rumberger A, Thies JE. Diversity of Planctomycetes in soil in relation to soil history and environmental heterogeneity. Appl Environ Microbiol. 2006;72:4522–31.
26. Ivanova AA, Kulichevskaya IS, Merkel AY, Toshchakov SV, Dedysh SN. High diversity of planctomycetes in soils of two lichen-dominated sub-arctic ecosystems of Northwestern Siberia. Front Microbiol. 2016;7:1–13.
27. Okazaki Y, Fujinaga S, Tanaka A, Kohzu A, Oyagi H, Nakano SI. Ubiquity and quantitative significance of bacterioplankton lineages inhabiting the oxygenated hypolimnion of deep freshwater lakes. ISME J. 2017;11:2279–93.
28. Lage OM, Bondoso J. Bringing Planctomycetes into pure culture. Front Microbiol. 2012;3:1–6.
29. Reintjes G, Arnosti C, Fuchs BM, Amann R. An alternative polysaccharide uptake mechanism of marine bacteria. ISME J. 2017;11:1640–50.
30. Ntougias S, Polkowska Ż, Nikolaki S, Dionyssopoulou E, Stathopoulou P, Doudoumis V, et al. Bacterial community structures in freshwater polar environments of Svalbard. Microbes Environ. 2016;31:401–9.
31. Okazaki Y, Nakano SI. Vertical partitioning of freshwater bacterioplankton community in a deep mesotrophic lake with a fully oxygenated hypolimnion (Lake Biwa, Japan). Environ Microbiol Rep. 2016;8:780–8.
32. Karlov DS, Marie D, Sumbatyan DA, Chuvochina MS, Kulichevskaya IS, Alekhina IA, et al. Microbial communities within the water column of freshwater Lake Radok, East Antarctica: predominant 16S rDNA phylotypes and bacterial cultures. Polar Biol. 2017;40:823–36.
33. Tadonléké RD. Strong coupling between natural Planctomycetes and changes in the quality of dissolved organic matter in freshwater samples. FEMS Microbiol Ecol. 2007;59:543–55.

34. Hirsch P, Müller M. Planctomyces limnophilus sp. nov., a stalked and budding bacterium from freshwater. Syst Appl Microbiol. 1985;6:276–80.
35. Labutti K, Sikorski J, Schneider S, Nolan M, Lucas S, del Rio TG, et al. Complete genome sequence of planctomyces limnophilus type strain (mü 290 T). Stand Genom Sci. 2010;3:47–56.
36. Zwart G, Crump BC, Kamst-van Agterveld MP, Hagen F, Han SK. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquat Microb Ecol. 2002;28:141–55.
37. Mao DP, Zhou Q, Chen CY, Quan ZX. Coverage evaluation of universal bacterial primers using the metagenomic datasets. BMC Microbiol. 2012;12:66.
38. Urbach E, Vergin KL, Young L, Morse A, Larson GL, Giovannoni SJ. Unusual bacterioplankton community structure in ultra-oligotrophic Crater Lake. Limnol Oceanogr. 2001;46:557–72.
39. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Author Correction: Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2:1.
40. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. ISME J. 2017;11: 2399–406.
41. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017; 35:725–31.
42. Kuenen JG. Anammox bacteria: from discovery to application. Nat Rev Microbiol. 2008;6:320–6. 2008
43. Galperin MY. A census of membrane-bound and intracellular signal transduction proteins in bacteria: Bacterial IQ, extroverts and introverts. BMC Microbiol. 2005;5:1–19.
44. Ulrich LE, Koonin EV, Zhulin IB. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. 2005;13:52–56.
45. Vigil-Stenman T, Ininbergs K, Bergman B, Ekman M. High abundance and expression of transposases in bacteria from the Baltic Sea. ISME J. 2017;11:2611–23.
46. Brown CT, Olm MR, Thomas BC, Banfield JF. Measurement of bacterial replication rates in microbial communities. Nat Biotechnol. 2016;35:725–31.
47. Gounot A-M. Psychrophilic and psychrotrophie microorganisms. Experientia. 1986;42:1192–7.
48. Hickman JW, Tifrea DF, Harwood CS. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. Proc Natl Acad Sci. 2005;102: 14422–7.
49. Luo G, Huang L, Su Y, Qin Y, Xu X, Zhao L, et al. FlrA, flrB and flrC regulate adhesion by controlling the expression of critical virulence genes in Vibrio alginolyticus. Emerg Microbes Infect. 2016;5:e85–11.
50. Bayer EA, Shimon LJW, Shoham Y, Lamed R. Cellulosomes - Structure and ultrastructure. J Struct Biol. 1998;124:221–34.
51. Artzi L, Bayer EA, Moraïs S. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol. 2017;15:83–95.
52. Peer A, Smith SP, Bayer EA, Lamed R, Borovok I. Non-cellulosomal cohesin- and dockin-like modules in the three domains of life. Mol Microbiol. 2011;291:1–16.
53. Moran NA. Microbial minimalism: genome reduction in bacterial pathogens. Cell. 2002;108:583–6.
54. Sorensen JW, Dunivin TK, Tobin TC, Shade A. Ecological selection for small microbial genomes along a temperate-to-thermal soil gradient. Nat Microbiol. 2019;4:55–61.
55. Dufresne A, Garczarek L, Partensky F. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biol. 2005;6:R14.
56. Luo H, Huang Y, Stepanauskas R, Tang J. Excess of non-conservative amino acid changes in marine bacterioplankton lineages with reduced genomes. Nat Microbiol. 2017;2:17091.
57. Biller SJ, Berube PM, Lindell D, Chisholm SW. Prochlorococcus: the structure and function of collective diversity. Nat Rev Microbiol. 2014;13:13.
58. Luo H, Moran MA. How do divergent ecological strategies emerge among marine bacterioplankton lineages? Trends Microbiol. 2015;23:577–84.
59. Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for microbial ecology. ISME J. 2014; 8:1553.
60. Getz EW, Tithi SS, Zhang L, Aylward FO. Parallel evolution of genome streamlining and cellular bioenergetics across the marine radiation of a bacterial phylum. mBio. 2018;9:e01089-18.
61. Salcher MM, Neuenschwander SM, Posch T, Pernthaler J. The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. ISME J. 2015;9: 2442–53.
62. Cabello-Yeves PJ, Ghai R, Mehrshad M, Picazo A, Camacho A, Rodriguez-Valera F. Reconstruction of diverse verrucomicrobial genomes from metagenome datasets of freshwater reservoirs. Front Microbiol. 2017; 8:2131.
63. Znachor P, Nedoma J, Hejzlar J, Seďa J, Kopáček J, Boukal D, et al. Multiple long-term trends and trend reversals dominate environmental conditions in a man-made freshwater reservoir. Sci Total Environ. 2018;624:24–33.
64. Martín-Cuadrado A-B, López-García P, Alba J-C, Moreira D, Monticelli L, Strittmatter A, et al. Metagenomics of the Deep Mediterranean, a Warm Bathypelagic Habitat. PLoS ONE. 2007; 2:e914.
65. Salcher MM, Pernthaler J, Posch T. Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria that rule the waves (LD12). ISME J. 2011;5:1242–52.
66. Bushnell B. BBDuk. 2016. https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh.
67. Bushnell B, Rood J, Singer E. BBMerge – accurate paired shotgun read merging via overlap. PLoS ONE. 2017;12:1–15.
68. Joshi NA, Fass J. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. 2011. https://github.com/najoshi/sickle.
69. Bushnell B. Reformat. 2016. https://github.com/BioInfoTools/BBMap/blob/master/sh/reformat.sh.
70. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.
71. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The ribosomal database project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009;37:141–5.
72. Nawrocki E. Structural RNA homology search and alignment using covariance models. Washington: Washington University in Saint Louis, School of Medicine; 2009.
73. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST:a new generation of protein database search programs. Nucleic Acids Res. 1997;25: 3389–402.
74. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 2013;41:590–6.
75. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–20.

76. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHITv1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016;102:3–11.
77. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–34.
78. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: Fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17:1–14.
79. Bushnell B. BBMap. 2015.
80. Segata N, Börnigen D, Morgan XC, Huttenhower C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun. 2013;4:2304.
81. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
82. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
83. Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5:e9490.
84. Mehrshad M, Rodriguez-Valera F, Amoozegar MA, López-García P, Ghai R. The enigmatic SAR202 cluster up close: Shedding light on a globally distributed dark ocean lineage involved in sulfur cycling. ISME J. 2018;12:655–68.
85. Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. J Bacteriol. 2005;187:6258–64.
86. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
87. Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
88. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428:726–31.
89. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genom. 2008;9:1–15.
90. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics. 2010;11:431.
91. Tatusov RL. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28:33–36.
92. Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. Nucleic Acids Res. 2003;31:371–3.
93. Huang L, Zhang H, Wu P, Entwistle S, Li X, Yohe T, et al. DbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Res. 2018;46:D516–D521.
94. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 Update. Nucleic Acids Res. 2015;43:W30–8.
95. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10:845.
96. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar A, et al. ARB: A software environment for sequence data. Nucleic Acids Res. 2004;32:1363–71.
97. Stamatakis A, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics. 2005;21:456–63.
98. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.