Chiriac *et al. Microbiome* (2026) 14:27
https://doi.org/10.1186/s40168-025-02272-x

Microbiome

RESEARCH **Open Access**

Ecological success in freshwater lakes: insights from novel cultivated lineages of the abundant *Nanopelagicales* order

Maria-Cecilia Chiriac[1*], Paul Layoun[1,2], Clafy Fernandes[1,2], Tiberiu Szőke-Nagy[1], Vojtech Kasalicky[1], Yusuke Okazaki[3], Jason N. Woodhouse[4], Hans-Peter Grossart[4,5], Kasia Piwosz[6], Petr Znachor[1,2], Bettina Sonntag[7], Cristiana Callieri[8], Sandi Orlić[9,10,11], Ruben Sommaruga[12,13], Cécile Lepère[14], Corinne Biderre-Petit[14], Helen Tammert[15], Daniel P. R. Herlemann[16,17], Mirosław Ślusarczyk[17,18], Anna Bednarska[18], Horia L. Banciu[19,20], Mariusz Zalewski[6], Adam Woźniczka[6], Rohit Ghai[1], Michaela M. Salcher[1] and Markus Haber[1*]

Abstract

Background The order *Nanopelagicales* is the most abundant bacterioplankton lineage in freshwater lakes and exhibits typical streamlined genomic characteristics such as small cell volumes (<0.1 μm^3), reduced genome sizes (<1.5 Mbp), and low GC content. These characteristics reflect adaptations to a free-living life strategy in oligotrophic environments. While many *Nanopelagicales* metagenome-assembled genomes and single-amplified genomes are available in public databases, strain-level microdiversity within this lineage remains poorly understood. This is mainly attributed to the incomplete nature of these genomes and the difficulty in isolating and maintaining pure cultures, with only 20 genome-sequenced cultures available to date.

Results Here, we report the isolation and genome analysis of 72 new *Nanopelagicales* strains, including members of *Planktophila* and a novel, previously uncultured genus, *Aquilimus*. High interspecific diversity and microdiversity were observed in the genus *Planktophila*, which likely facilitates the coexistence of closely related species within the same habitats by allowing fine-scale niche partitioning. The unusually high diversity of transporters for small organic compounds, along with carbohydrate-active enzymes, suggests that *Planktophila* members can degrade plant and algal polymers and import the resulting products to support growth. A notable finding is the repeated, independent loss of the oxidative phase of the pentose phosphate pathway in abundant *Nanopelagicales* species, which may represent an energy-saving adaptation in oligotrophic waters. Two species (*Planktophila vernalis* and *Nanopelagicus abundans*) seem to be equally abundant on a global scale, with water pH likely being the most significant factor influencing the predominance of one group over the other in different water bodies. Additionally, *P. vernalis* may tolerate periods of anoxia due to genomic encoding of respiratory nitrate reductase and nitrate/nitrite antiporters.

Conclusions In conclusion, this work increased to a great degree the cultivated diversity of the abundant *Nanopelagicales* order. Analysis of over 1700 metagenomes showed that only a few cultivated species are globally dominant,

*Correspondence:
Maria-Cecilia Chiriac
cecilia.m.chiriac@gmail.com
Markus Haber
markus.haber@gmail.com
Full list of author information is available at the end of the article

and time-series analyses revealed consistent spring and autumn peaks. Key metabolic adaptations, such as loss of the oxidative phase of the pentose phosphate pathway and a high microdiversity of genes involved in cell surface biosynthesis and modifications, are likely to help these species survive periods of starvation and avoid predation. These findings highlight the ecological importance of *Nanopelagicales* and suggest that microdiversity underpins their adaptability. This work lays a foundation for studying their physiology, ecology, and strain-specific functional variation.

Keywords *Nanopelagicales*, *Planktophila*, *Nanopelagicus*, *Aquilimus*, Genomics, Microdiversity, Seasonal variation, Bacterial cultivation

Background

The pelagic zone of lakes and oceans is often dominated by genome-streamlined microbes adapted to the scarce nutrients available in oligotrophic environments [1]. Their reduced genome size (<1.5 Mbp), low GC content, and small cell sizes (cell volumes <0.1 μm^3) reflect their competitiveness for resources leading to minimal nutrient requirements for survival and reproduction [1–4]. The same evolutionary pressure favors gene loss, a low number of paralogs and pseudogenes, reduced intergenic spacers and a stable core genome [1], as well as an increased dependency on co-occurring microbes [5]. Among such streamlined microbes are members of the order *Nanopelagicales* (also referred to as acI, hgcl, or the ACK-M1 clade [6, 7]) within the phylum *Actinomycetota*. These bacteria have been recovered primarily from freshwater lakes, rivers, and estuaries, with only limited occurrence in brackish waters [6, 8–10], suggesting a specialization to the freshwater environment.

Members of the *Nanopelagicales* order are often the dominant bacterioplankton group in the pelagic zone of freshwaters worldwide [11, 12], where they can make up > 50% of all bacterial cells [11, 13]. Until recently, little was known about their physiology because streamlined microbes are difficult to cultivate and maintain in laboratory settings, largely due to their unknown nutritional requirements, despite significant efforts [3]. Unsurprisingly, most studies explored their ecology and physiology with molecular approaches such as single-cell genomics, metagenomics, metatranscriptomics, fluorescence in situ hybridization (FISH), or microautoradiography-FISH (MAR-FISH) [12, 14–20], limiting an in-depth assessment of their biology. This is very much in contrast to the abundant marine streamlined microbial groups such as *Pelagibacter* (SAR11) and *Prochlorococcus* that have been successfully brought into culture and maintained decades ago [21, 22].

While there is no shortage in the number of metagenome-assembled genomes (MAGs) and single-amplified genomes (SAGs) affiliated to *Nanopelagicales*, since more than 1400 genomes are available in the Genome Taxonomy Database (GTDB) repository (release 220, 18.04.2024), only 20 genomes originate from isolates [3, 23]. Following their initial isolation, genome sequencing, and characterization, these strains were classified as *Planktophila*, *Nanopelagicus*, and the IMCC26077 genus of *Nanopelagicaceae* [3, 23]. Although auxotrophies for essential cofactors and amino acids that were predicted from MAGs and SAGs [3, 23] were taken into account in these cultivation attempts, none of the 20 cultures could be maintained due to the lack of other unknown growth requirements. However, two isolates have been successfully revived from glycerol stocks and kept in stable cultures after the addition of catalase to the growth medium [24]. The addition of catalase also led to the isolation of another 75 *Nanopelagicales* strains from eight tribes (species-like groups) from Lake Soyang, South Korea; however, these isolates were not analyzed further [25]. Later, genome analyses of the 20 genome-sequenced isolates coupled with physiological experiments of the two revived strains revealed that all 20 isolates lacked the gene cluster required for heme biosynthesis and that this auxotrophy rather than a need for (heme-containing) catalase prevented stable cultivation [26].

Given the recently acquired knowledge about their heme auxotrophy, here we report the isolation and genome analysis of 72 new *Nanopelagicales* strains affiliated with the genus *Planktophila* and a novel, previously uncultured genus for which we propose to name *Aquilimus* gen. nov. (AAA044-D11, GTDB r220). The new isolates were obtained from 20 lakes of various trophic levels in Europe and Asia (Japan), as well as from several locations at the coast of the Baltic Sea. The complete, circular genomes obtained in this work analyzed together with 20 previously available complete genomes allowed for an unprecedented survey of microdiversity and interspecies diversity, biogeography, seasonality, and potential niche specialization in one of the most significant freshwater bacterial lineages. This allows us to speculate that the high microdiversity observed within *Planktophila* is indicative of adaptive radiation driven by specialized ecological niches across different freshwater habitats, potentially contributing to the lineage's widespread distribution.

Methods

Isolation of planktonic *Nanopelagicales*

The isolation of novel *Nanopelagicales* strains followed a dilution-to-extinction approach [27] in artificial lake water medium supplemented with various nutrient and carbon sources [28] (Suppl. Table 1). Water samples were collected in several sampling campaigns between 2019 and 2022 from 20 freshwater lakes in Europe (11 countries) and Japan, ranging from ultra-oligotrophic to eutrophic and dystrophic. Additionally, several locations at the Vistula Lagoon, Vistula River, and the coastal Baltic Sea in Poland have been sampled [29] (Suppl. Table 2). Ten milliliters of water were filtered from several sampling depths (Suppl. Table 2) through either 0.4 µm or 20 µm filters and cell abundances in filtrates were determined using either a CytoFlex flow cytometer (Beckman Coulter; Brea, CA, USA) equipped with a blue laser (488 nm, bandpass filters 525/40 and 690/50) after staining of a 190 µl subsample with SYBR Green I (0.5× standardized concentration; Lonza, Rockland, ME, USA) [30] or an epifluorescence microscope (Olympus BX53 or Zeiss Axioscope) after staining of a 1 ml subsample with DAPI (4′,6-diamidino-2-phenylindole, final concentration 0.01 µg/ml [31]). Dilution-to-extinction of the filtrate with approximately 1 cell per well was performed in 96-well plates filled with 1.5 ml of medium [28] and incubated for 6–8 weeks at either 16 °C (samples collected from 0.5 to 5 m) at 12:12 h light dark cycles or 8 °C (samples from deeper water layers) in the dark. Afterwards, growth was evaluated using flow cytometry as described above. Wells with microbial growth were transferred into fresh media in a proportion of 1:10 and were further maintained through subsequent transfers every 6–8 weeks. Screening of 16S rRNA genes and glycerol stocks (20% final concentration) were done as previously described [3, 32]. *Planktophila dulcis* strains IIB-125 and IIB-161, which were previously isolated [3], were successfully revived from glycerol stocks in med2 medium (Suppl. Tables 1, 2).

Whole genome sequencing, assembly, and annotation

A total of 72 *Nanopelagicales* strains were isolated and grown in 400–500 ml culture medium until stationary phase. Cells were filtered onto 0.2 µm PES filters (Merck, Germany) and genomic DNA was purified using the Quick-DNA HMW MagBead Kit (Zymo Research, USA). Paired-end sequencing (PE150) was performed on an Illumina Novaseq 6000 instrument by an external provider (Novogene, HK). Raw reads were quality trimmed using BBMap v36.x (https://sourceforge.net/projects/bbmap/) with default parameters and assembled with SPAdes v3.12.0 (using *k*-mers 29, 49, 59, 69, 79, 89, 99, 109, 119, 127) [33]. Genomes were checked for circularity as described before [28, 29]. Briefly, all single-contig genomes were analyzed for circularity using an in-house perl script that detects direct repeats of at least 30 nucleotides at the contig ends. If such repeats were found, the overlapping region was trimmed from one end, and the genome was classified as circular. Based on this approach, 48 genomes were identified as circular. Genomes assembled into one contig, but without direct repeats at the ends, and assemblies consisting of multiple contigs were manually curated to circular chromosomes through iterative reads mapping to contigs using Geneious Prime 2022.1.1 (default mapper with high sensitivity; www.geneious.com). Contigs were extended bidirectional followed by de novo assembly with Geneious 10 assembler (high sensitivity settings) and the “find repeats” option in Geneious (minimum repeat length 50 nucleotides), until overlapping ends were identified. In six cases, direct repeats were detected at the ends of the contigs, but the sequence following the repeat at one end did not match the opposite end, suggesting a possible assembly artifact. To verify this, we designed primers flanking the repeat regions (Suppl. Table 3), performed PCR amplification, and sequenced the products using Sanger sequencing performed at SEQme s.r.o. (Czech Republic). The PCR products confirmed that the extra sequence was artificial and introduced during assembly. Based on the validated sequence, the genomes were closed. Prokka [34] was used for gene prediction and the annotation was performed with hmmsearch [35] against collections of Interpro [36], COG [37], TIGRFAM [38], Pfam [39], and KEGG [40] hidden Markov models (HMMs). Carbohydrate-active enzymes (CAZy) were searched in the genomes using hmmscan [41] and the dbCAN CAZyme domain HMM database v11 (release date 14.08.2024) [42]. CheckM v1.1.3 [43] was run to estimate completeness, contamination, and strain heterogeneity of the genomes, while taxonomy was assigned using GTDB-Tk [44] v2.3.2 toolkit (https://ecogenomics.github.io/GTDBTk/) based on the GTDB release r220 (April 2024). Twenty previously published genomes of other *Nanopelagicales* isolates [3, 23] were included in the analysis and annotated as described above (Suppl. Table 2). Average nucleotide identity (ANI) and average amino acid identity (AAI) were calculated using established methods [45, 46], applying a 95% ANI threshold to delineate species and a 65% AAI threshold to differentiate genera. The complete set of 92 genomes belonging to *Nanopelagicales* isolates were dereplicated using dRep [47] (default parameters, > 95% ANI), resulting in 41 representative genomes (Suppl. Table 4). In order to determine proteins specific for abundant *Planktophila* species, the genomes were grouped in two sets: abundant (> 1× coverage per Gb in > 10% of freshwater metagenomes from a collection of 1720 publicly available and newly sequenced metagenomes—see below

Metagenomic fragment recruitment) versus rare (> 1× coverage per Gb in < 10% of freshwater metagenomes), and their proteins were clustered using MMseqs2 [48] (0.3 identity, 0.8 coverage).

Phylogeny of *Nanopelagicales* isolates and MAGs

In order to observe the placement of the new isolates, a phylogeny of the entire *Actinomycetes* class was performed with a collection of 1044 high-quality genomes (Suppl. Table 5). For this, all genomes affiliated to the order *Nanopelagicales* with > 70% completeness and < 5% contamination (as determined using CheckM [43]) were queried on GTDB (*n* = 865; GTDB r220 [49], April 2024), retrieved from National Center of Biotechnology Information (NCBI), and dereplicated using the default parameters in dRep [47] to a final set of 212 genomes. Additionally, genomes of all isolates (*Nanopelagicales* *n* = 92, *Actinomycetes* isolates from other orders *n* = 38) were added, as well as the most complete genome (> 90% completeness, < 5% contamination) of each *Actinomycetes* genus (*n* = 702; Suppl. Table 5). A collection of 120 single-copy protein sequences [43] was used to generate the phylogenomic tree using the complete genomes of *Listeria monocytogenes* EGD-e and *Staphylococcus aureus* NCTC-8325 as an outgroup, as previously described [3]. Briefly, individual markers were aligned with PRANK (-protein +F) [50], trimmed with BMGE (-m BLOSUM62 -t AA -g 0.5 -b 5) [51] and concatenated. The best evolutionary model for the dataset was determined using ModelFinder [52] and the maximum likelihood tree was generated with IQ-TREE2 [53] (evolutionary model LG+F+I+G4, 1000 ultrafast bootstraps).

Rhodopsin identification and phylogenetic analysis

Genomes of all isolates were scanned for rhodopsin genes to identify their potential for photoheterotrophic growth. Protein-coding sequences were predicted using Prodigal v2.6.3, and candidate rhodopsins were identified using HMMER (hmmsearch) [54] with Pfam HMMs corresponding to bacteriorhodopsin and heliorhodopsin. Only hits longer than 150 amino acids and with *P* values < 0.01 were kept for further analysis. To classify these proteins, sequences were queried against a curated rhodopsin database (including all UniProt50 and GTDB rhodopsins), dereplicated using MMseqs2 [48] (easy-cluster, 90% identity) [55], and the top 50 matches per query were retrieved. These sequences were aligned using MAFFT [56] (--localpair --maxiterate 1000) to allow the comparison of functionally relevant residues, particularly in the retinal-binding region and transmembrane domains. Transmembrane helix topology was predicted using PolyPhobius [57], and each sequence was examined for the presence of seven transmembrane helices and the conserved retinal-binding motif in TM7 (DxxxK for type I rhodopsins; SxxxK for heliorhodopsins). These helices are a conserved feature of rhodopsins that form the core of the protein and the binding pocket for the retinal chromophore [58]. To classify the rhodopsins phylogenetically, a total of 643 sequences (including the curated rhodopsin database with representative sequences of different rhodopsin subfamilies [55] and rhodopsins predicted in the new *Nanopelagicales* isolates) were aligned with PASTA [59], and a maximum likelihood tree was inferred using IQ-TREE2 [53] (Q.pfam+F+G4 model, with 1000 ultrafast bootstraps and SH-aLRT support). This analysis enabled us to determine whether genomes encoded for actinorhodopsins and heliorhodopsins.

Metagenomic fragment recruitment

The global distribution of the isolates obtained in this study across freshwater and brackish environments was assessed using metagenomic fragment recruitment. For this purpose, we collected 1720 publicly available and newly sequenced freshwater/brackish metagenomes (Suppl. Table 6). All metagenomes were quality trimmed using BBDuk.sh script from the BBTools package (https://archive.jgi.doe.gov/data-and-tools/software-tools/bbtools/) using the following parameters: -Xmx100g qtrim=rl trimq=18 tbo. The adaptors were removed using BBmerge.sh [60] with the -outa flag. To reduce computational cost, individual metagenomes were subsampled to 20 million reads using the reformat.sh script of BBTools. Ribosomal RNA (rRNA) genes in our 41 assembled representative genomes were identified with barrnap version 0.9 (https://github.com/tseemann/barrnap) and masked (their sequences were replaced with "N"s) prior to recruitment analysis to avoid biases. This is necessary because rRNA genes are highly conserved and can have multiple copies in genomes, which can result in disproportionately inflated recruitment signals and misleading abundance estimates [9]. MMseqs2 [48] was used to map metagenomic reads onto the genome representatives and base coverage per gigabase (Gb) was calculated (-minid 0.95 -mincov 0.9 -minlen 50). Metagenomic reads were mapped to the entire genome sequence of each isolate. However, we only considered a genome to be present in a given metagenome if at least 40% of its length was covered by recruited reads. This breadth threshold was chosen because the core genome of *Planktophila* accounts for 40% or more of the total genome. An in-depth analysis of recruitment analysis results in four time-series (Římov Reservoir, Lake Stechlin, Breiter Luzin, and Tiefwaren) was performed for the 33 representative genomes of *Planktophila*. To identify *Planktophila* species that exhibit similar temporal abundance patterns across samples, we calculated pairwise

Spearman correlations between their recruitment values over time. A distance matrix was then generated using $1-\rho$ (Spearman's rho) as a measure of dissimilarity. Next, hierarchical cluster analysis was done using the *hclust* function from the stats package (version 4.3.2) in R with complete agglomeration method and the optimal number of clusters was computed using the Kelley-Gardner-Sutcliffe penalty function [61]. Spearman's correlations were calculated to explore the relationship between the abundance of *Planktophila* strains (expressed as average coverage per Gb) and physicochemical parameters.

Microdiversity of *Planktophila* isolates

To investigate the microdiversity of the *Planktophila* genus, we identified genomic islands in *Planktophila* species groups containing at least two sequenced isolates ($n=69$). Clonal genomes ($n=15$) have been removed for simplicity. Single nucleotide polymorphisms (SNPs) were identified with parsnp [62] by comparing individual genomes to species-representative genomes (Suppl. Table 4), and rates of synonymous substitutions per synonymous site (dS) were computed with the R package orthologr [63]. Genomes were aligned in a pairwise manner following phylogenetic relationships and reciprocal BLASTn analyses were performed to identify contiguous regions exhibiting consistent low sequence similarities between individual genomes and to detect disruptions in genomic synteny. Stretches of genes with consistent low alignment scores (*e*-value threshold of 1e−3) were considered as part of a potential genomic island and selected for manual inspection as performed previously [3, 30]. Specifically, all-versus-all BLASTp comparisons of protein sequences (*e*-value $\leq$1e−3, BLOSUM62 matrix, minimum alignment length 0.5, minimum identity 90%) [64] within each species (ANI > 95%) were conducted to pinpoint the gene content of individual genomic islands. Flex2 (https://github.com/asierzaragoza/flex2) was used to plot the whole genome alignments depicting the synteny and different degrees of sequence identity between genomes. The core and pangenome of *Planktophila* was calculated across all genomes ($n=84$, Suppl. Table 2) following the approach previously used by Neuenschwander et al. [3]. Briefly, a random order of genomes was created. All proteins from the first and second genome were clustered at >50% identity and >50% coverage using usearch6 (https://www.drive5.com/usearch/) and numbers of ortholog clusters (core genes) and unclustered genes (flexible genes) were enumerated. Sequentially, additional genomes were added, clustered, and core and flexible genes enumerated until all genomes were processed, resulting in an estimate of the number of shared genes in all genomes. Ten independent iterations with randomized genome order were done to obtain the final number of core and flexible genes unbiased by genome order.

Scanning electron microscopy

Planktophila alpina GksH-29 and *P. landstejnensis* Bala-9 were cultivated as described above until late stationary phase and fixed overnight at 4 °C with glutaraldehyde (final concentration 2.5%). Four milliliters of culture were filtered onto white polycarbonate filters (0.2 µm pore size, Millipore, Germany) and subsequently dehydrated stepwise in 30%, 50%, 70%, and 100% ethanol, 10 min each. After drying, filters were stored at 4 °C until further processing. For imaging, filters were cut with a blade and mounted on a metal stub using conductive copper tape. Samples were sputter-coated with a 10-nm gold layer in a Polaron Sputter Coater chamber (Polaron Ltd., Watford, UK) and examined using a JEOL 7401-F field emission scanning electron microscope (JEOL Europe, Prague, Czech Republic).

Results

Diversity, genomic characteristics, and distribution of *Nanopelagicales* strains

A total of 72 new *Nanopelagicales* isolates were cultivated and their genome were sequenced in this study. The genomes of these isolates were dereplicated together with the 20 publicly available isolates' genomes [3, 23], resulting in 41 species-level representatives (Suppl. Table 4). Until now, cultured isolates of *Nanopelagicaceae* were affiliated to six *Planktophila* species, three *Nanopelagicus* species, and one species of the IMCC26077 genus [3]. Based on ANI and AAI analyses (Suppl. Tables 7, 8), the 72 isolates represent 24 newly cultured species of *Planktophila* and four species of the newly cultured genus, AAA044-D11 (GTDB v.220, April 2024) (see Fig. 1A; Suppl. Tables 2, 4, 7, 8), for which we propose the name *Aquilimus* gen. nov.. Details on the proposition of the new genus and of novel species names that were registered at SeqCode [65] can be found as Supplementary information. For morphological characterization, SEM imaging of two *Planktophila* strains (*P. alpina* GksH-29 and *P. landstejnensis* Bala-9) showed that the cells are curved rods with very small volumes (<0.1 µm^3), consistent with previous reports [24]. By contrast, DAPI staining of *Aquilimus zluticensis* MZLE-12 revealed even smaller cell sizes in this newly isolated genus (Suppl. Fig. 1).

A phylogenomic tree using 120 marker protein sequences [43] of all *Nanopelagicales* representative genomes included in GTDB (April 2024, dereplicated genomes) together with the new isolates shows that nine other genera in this family remain uncultured (Fig. 1A). While previous isolates were obtained from only two locations, Lake Zurich [3] and Lake Soyang [23–25], the

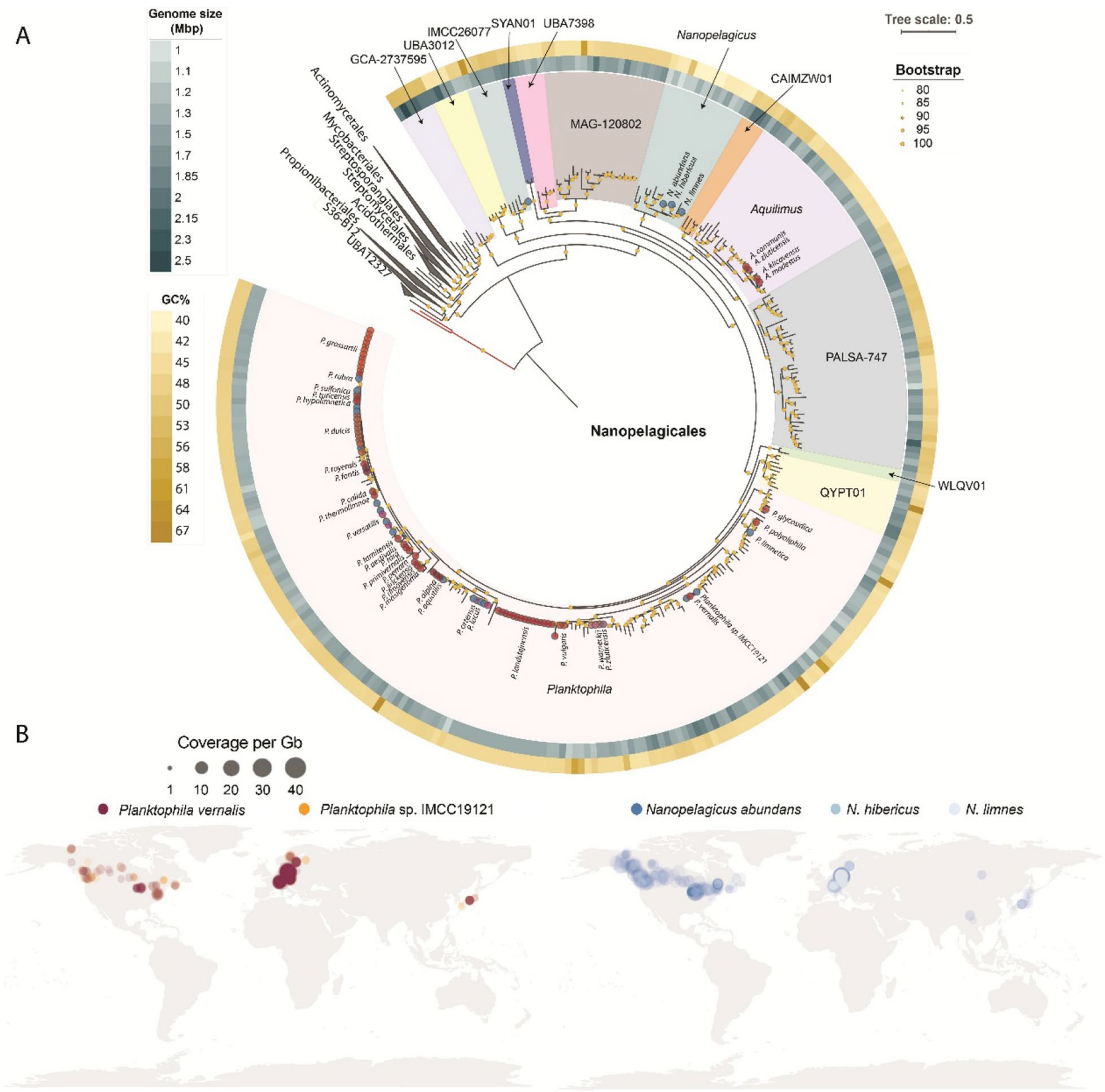

Fig. 1 **A** A maximum likelihood phylogeny of the class *Actinomycetes* was constructed using 120 single-copy marker protein sequences. The 72 new isolates obtained in this study were compared to all available complete genomes of published *Nanopelagicales* isolates ($n = 20$ [3, 23]), as well as the dereplicated *Nanopelagicales* MAGs with >70% completeness from GTDB r220 (April, 2024) ($n = 212$), and the most complete genome (>90% completeness) from each genus within the *Actinomycetes* class ($n = 702$). Genomes from this study are marked with a red dot while previously published culture genomes are marked with a blue dot at the tip of their branch. **B** Global distribution and relative abundance (coverage per Gb) of the most prevalent *Planktophila* and *Nanopelagicus* species

newly isolated strains originate from 20 different locations. These cultures opened the opportunity to study the genome structure of a new genus, as well as the metabolic microdiversity of *Planktophila*.

The genomes sequenced in this study were polished to circular chromosomes with sizes varying from 1.16 to 1.55 Mbp. As expected for streamlined genomes, all four cultivated genera—*Planktophila*, *Nanopelagicus*, *Aquilimus*, and IMCC26077—have a low GC content for *Actinomycetota*, ranging between 40 and 51% (Suppl. Figs. 2, 3 A), and high coding densities (95.2–96.1%) (Suppl. Fig. 2, 3 A; Suppl. Table 9). As anticipated, there is a correlation between the genome size and the GC content with smaller genomes having lower GC content

Fig. 2 The completeness of KEGG modules in the genomes of *Nanopelagicales* isolates. A simplified phylogenetic tree is plotted at the top. Different KEGG categories are indicated by color; if a pathway is complete in at least one isolate of a genus, the corresponding symbol is filled; if the symbol has a black border, the pathway is complete in all genomes of the genus; a symbol filled with white indicates that one enzyme of the pathway or one subunit of a complex is missing in all genomes of that particular genus. If more than one enzyme is missing in all isolates of a genus, no symbol is shown. A manual inspection of all pathways was done for the InterPro, COG, KEGG, TIGR, and Pfam annotations

(Pearson's $R = 0.357$, P value < 0.001) (Suppl. Fig. 2; Suppl. Table 9). The shift toward lower GC content, likely driven by selective pressures, was accompanied by a change in stop codon usage, with TAA being favored over the GC-rich TGA and TAG. Moreover, minimal nitrogen utilization appears to be reflected also in the preference for lysine over arginine and histidine as the positively charged amino acid (Pearson's $R=-0.73$ and -0.85, respectively; P value <0.001 for both) (Suppl. Fig. 2). Another adaptation to the oligotrophic pelagic zone of lakes is the preservation of a minimum number of sensor domains and sigma factors [66], with *Nanopelagicales* members encoding only between two to four sigma factors (Suppl. Table 9). The core genome of the 84 *Planktophila* strains consisted of 736 genes, representing 48.4–62.8% of the entire genome, while the pangenome is open and estimated at >6800 genes (Suppl. Fig. 3B).

Global distribution of *Nanopelagicales* isolates in freshwater lakes

Recruitment results for all dereplicated culture genomes against a collection of 1720 freshwater and brackish metagenomes (samples collected from 230 different lakes, rivers, estuaries, and coastal oceans—Suppl. Tables 4, 6) revealed that only a few cultured species affiliated to *Planktophila*, *Nanopelagicus*, and to some extent, the newly isolated genus *Aquilimus*, were truly ubiquitous and represented major components of the microbial community (Fig. 1B; Suppl. Figs. 4–7). None of the species occurred in dystrophic lakes with low pH (e.g., Grosse Fuchskuhle, Trout Bog, Crystal Bog; $n=554$), in anoxic samples ($n=16$), and in the African Great Lakes Tanganyika, Malawi, and Kivu ($n=42$; Suppl. Table 6). Seven species of *Planktophila* and two of *Nanopelagicus* reached abundances >1× coverage per Gb (Cov/Gb) in ~10% of the analyzed metagenomes, the most abundant and common being *P. vernalis* (maximum recovery 39× Cov/Gb; > 1× in 606 metagenomes) and *N. abundans* (40× Cov/Gb; > 1× in 580 metagenomes). Based on the analysis of the four isolated species, the *Aquilimus* genus was recovered >1× Cov/Gb in 65 metagenomes, mainly in the epilimnion and hypolimnion of Žlutice and Římov Reservoirs (Czech Republic), and sporadically in Lakes Stechlin and Tiefwaren (Germany). While in Lake Zurich and Žlutice Reservoir they appeared simultaneously with *Planktophila*, in Lake Tiefwaren they were found together with both *Planktophila* and *Nanopelagicus*, but in much lower abundances (Suppl. Table 6). *Planktophila vernalis* and *N. abundans* were recovered in high abundances in some instances in the same samples, for example, in the German Lakes Tiefwaren (max. 39× and 17× Cov/Gb for *P. vernalis* and *N. abundans*, respectively), Breiter Luzin (max. 27× and 25× Cov/Gb), and Stechlin (max. 21× and 35× Cov/Gb) as well as in the Klíčava Reservoir, Czech Republic (max. 19× and 7× Cov/Gb). While up to 21 of the isolated species co-occurred in the same samples (>1× Cov/Gb), only one genus and often one species appeared to be clearly dominant in some freshwater lakes (Suppl. Table 6). For example, only *P. vernalis* was recovered from eutrophic Lake Loclat and the eutrophic surface water of River Torrens in Adelaide, South Australia. On the other hand, *N. abundans* dominated oligotrophic boreal and humic lakes (Canada, USA, Finland) reaching its maximum abundance of 40× Cov/Gb in Lake Pine, while being also recovered with high values in Lake Michigan (USA) (27× Cov/Gb), and in the epilimnion of Lake Maggiore (Italy) (23× Cov/Gb). Even though *N. limnes* never reached the abundances of the other two *Nanopelagicus* species, it was much more widespread than *N. hibericus*, having Cov/Gb > 1× in 487 metagenomes. At specific time points, it appeared to be the dominant *Nanopelagicales* species in the epilimnion of mesotrophic Lakes Erken (Sweden) and Biwa (Japan), or in meso-eutrophic Lake Tiefwaren (Germany). It usually co-occurred with *N. abundans* in boreal lakes (e.g., Canada, Northern Europe), and with *P. vernalis* and *N. abundans* in Lake Tiefwaren, Breiter Luzin, and Stechlin (Germany) (Suppl. Table 6). The IMCC26077 genus, with only one isolate (GCA_002284915), was recovered in just seven metagenomes with >1× Cov/Gb, specifically in subtropical locations such as Lake Biwa and Ikeda (Japan) and in River Pearl (China) (Suppl. Fig. 7; Suppl. Table 6).

Time-series patterns in *Planktophila*, *Nanopelagicus*, and *Aquilimus*

To assess how the abundance of *Nanopelagicales* isolates changes over time, metagenomic time-series data were analyzed. These datasets consist of metagenomes collected at regular time intervals for 5–9 years from

(See figure on next page.)

Fig. 3 **A** Schematic representation of the metabolism in *Planktophila*. Characteristics for abundant *Planktophila* species are highlighted with red. Abbreviations: I: NADH:quinone oxidoreductase; II: succinate dehydrogenase; III: cytochrome *bc1* complex; IV: cytochrome *c* oxidase/cytochrome *bd* ubiquinol oxidase; 6PGL: 6-phosphogluconolactone; 6PG: 6-phosphogluconate; CAZy: carbohydrate-active enzymes; ComEF: competence proteins E and F; Cyt bd: cytochrome *bd*; Cyt C: cytochrome *c*; FAD/$FADH_2$: flavin adenine dinucleotide; Formate DH: formate dehydrogenase; GA3P: glyceraldehyde 3-phosphate; NAD/NADH: nicotinamide adenine dinucleotide; NarK: nitrate/nitrite antiporter; S-7P: sedoheptulose 7-phosphate; Sec: general secretion system; Tat: twin-arginine translocation pathway; TCA cycle: tricarboxylic acid cycle; TKT: transketolase. **B** Total number (top panel) and presence/absence of transporters for low molecular weight organic compounds encoded in each genus. **C** Number of structural proteins per genome involved in membrane transport as annotated using KEGG database. **D** Total number of CAZy per genome and number of different CAZys encoded by the isolates of each cultivated *Nanopelagicales* genus

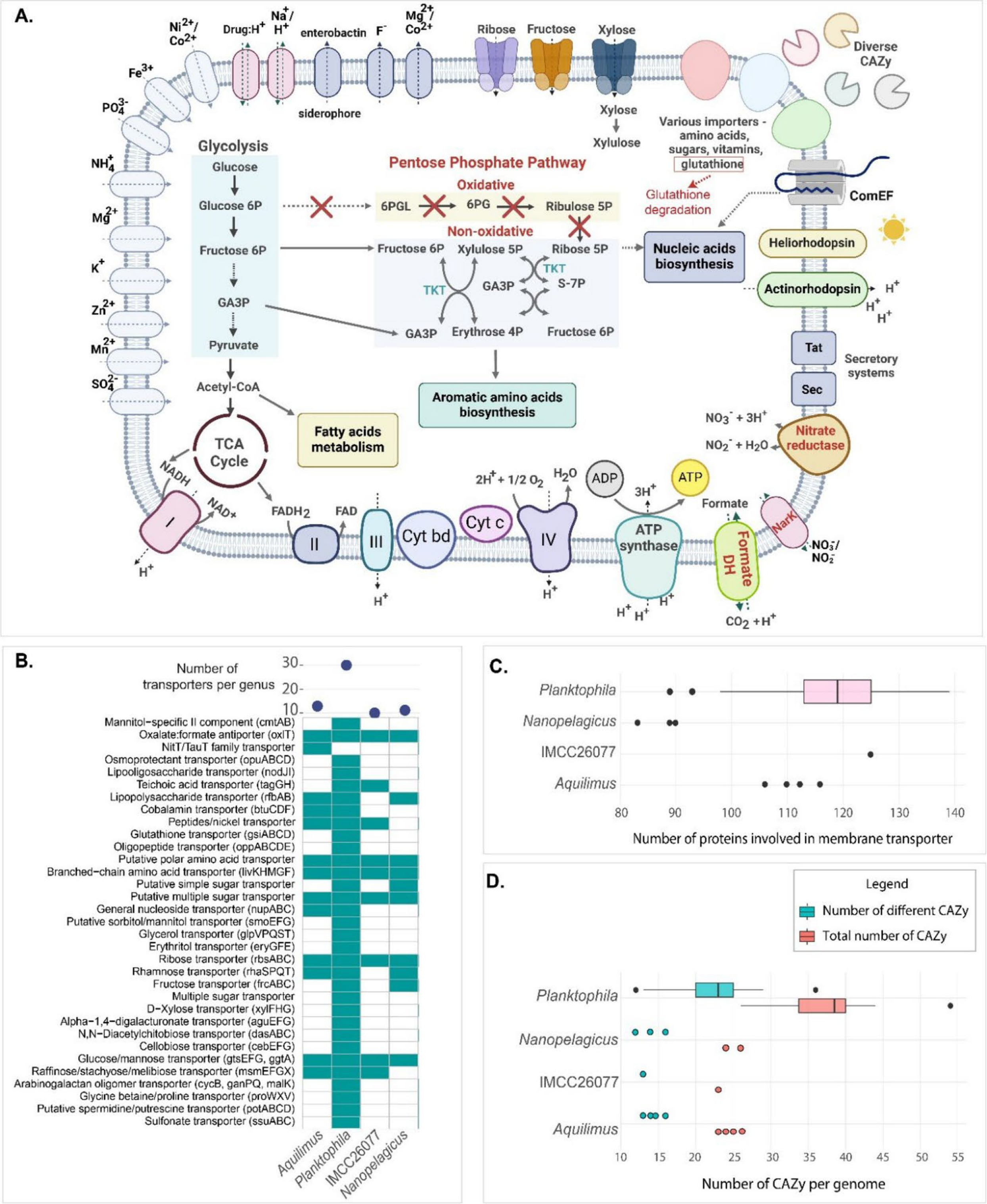

Fig. 3 (See legend on previous page.)

five freshwater lakes from Europe (Římov Reservoir [67–69] in Czech Republic, Lakes Breiter Luzin, Grosse Fuchskuhle, Stechlin, and Tiefwaren in Germany) and Lake Mendota, USA [70]. Recruitment analysis for the genomes of isolates in these samples provided details about seasonal fluctuations in the abundance of each species and their relationships with physicochemical parameters (Suppl. Tables 6, 10; Suppl. Figs. 8–21). As up to 20 *Planktophila* species co-occurred in the same sample (recruitment values > 1× Cov/Gb), cluster analysis was performed for the representative *Planktophila* strains ($n = 33$), which revealed six clusters with different temporal behaviors at the surface (0.5 m depth) and five in the hypolimnion (30 m depth) of Římov Reservoir (Suppl. Table 11; Suppl. Fig. 8). At both depths, genomes affiliated with *P. vernalis* were often the dominating group (cluster 3), although different species belonging to cluster 1 (*P. landstejnensis*, *P. lacus*, and *P. vulgaris*) were sometimes dominant at the surface, as well as cluster 5 (*P. thermolimnae* and *P. calida*), which peaked in summer months in the epilimnion (0.5 m depth). At the surface of the Římov Reservoir, *P. vernalis* reached annual maxima during the colder months (October–March) and faded away by mid-summer (recruitment values ≤ 1× Cov/Gb). The highest abundances were observed in autumn, followed by another bloom in spring. Abundances in the hypolimnion (30 m depth) tended to follow the same pattern as at the surface, but recruitment values remained elevated during the summer months (1–3× Cov/Gb) in the stratified reservoir, suggesting a preference for colder temperatures (Suppl. Fig. 8; Suppl. Table 6). Interestingly, although *N. abundans* was the second most prevalent *Nanopelagicales* isolate in freshwater environments worldwide, it was absent in the Římov Reservoir, the only representative of this genus here being *N. limnes*, which occurred sporadically and in lower abundances compared to *Planktophila* species (Suppl. Table 6). Even though *Aquilimus* species were in general less frequent, they were present with lower recruitment rates than *Planktophila* all year round in the Římov Reservoir, especially *A. communis. Aquilimus* species were positively correlated with diatoms (Spearman's rho = 0.47, *P* value < 0.05) and with green algae (Spearman's rho = 0.35, *P* value < 0.01; Suppl. Table 7). Spearman's correlations between physicochemical parameters and genome recruitment values suggested that in the surface layer (0.5 m depth), most *Planktophila* species are negatively associated with dissolved organic matter and some species associated positively with the presence of diatoms or cryptophytes (Suppl. Fig. 9; Suppl. Table 10). In contrast, a negative association was observed between *N. limnes* and diatoms (Spearman's rho = −0.62, *P* value < 0.01; Suppl. Table 10, Suppl. Fig. 9). In the hypolimnion of Římov Reservoir (30 m depth), some *Planktophila* species and all *Aquilimus* species were associated with lower temperatures and higher dissolved oxygen levels (Suppl. Table 10). *Planktophila vernalis* and *Planktophila* sp. IMCC19121 (GCA_00228495) were notable exceptions as both correlated negatively with dissolved oxygen (Spearman's rho = −0.33 and −0.43, respectively, *P* values < 0.01). In Lake Mendota, *Planktophila* species, including *P. vernalis*, *P. landstejnensis*, and *P. vulgaris*, appeared only occasionally with Cov/Gb > 1. Here, *N. abundans* was the dominant species, with peaks from May to August and to a lower extent, in September–October, having a slight negative association with temperature (Spearman's rho = −0.28, *P* value < 0.05) (Suppl. Tables 6 and 10; Suppl. Fig. 11).

In the lakes from Northeastern Germany, *Planktophila* species formed 4 to 5 clusters with distinct but consistent temporal patterns over the 9-year sampling period (Suppl. Table 11; Suppl. Figs. 12, 16, 19). An exception is Lake Grosse Fuchskuhle, where none of the analyzed *Nanopelagicales* members was present, most probably due to its low pH. Lake Stechlin was dominated by *N. abundans* and *N. limnes*, which seem to preferentially bloom in autumn, while members of *Planktophila* cluster 3 (*P. vernalis*, *Planktophila* sp. IMCC19121) and cluster 4 (*P. ortensis*, *P. lacus*, and *P. calida*) had peaks in spring and summer (Suppl. Fig. 12). In the epilimnion (5 m depth), the abundance of both *Planktophila* and *Nanopelagicus* were predicted to be positively influenced by temperature, and negatively by total phosphorus and nitrate concentrations (Suppl. Table 10; Suppl. Fig. 13). Interestingly, water pH appeared to influence nearly all *Planktophila* species, with acidic conditions being unfavorable (Suppl. Table 10; Suppl. Fig. 13). In general, the abundances of *Nanopelagicales* members in the hypolimnion of Lake Stechlin (40 m depth) tended to fluctuate

(See figure on next page.)

Fig. 4 Microdiversity in *Planktophila* isolates. Genomes have been linearized for simplicity, and the order follows the histogram generated using average nucleotide identity (ANI) values. Species with > 2 previously isolated strains are highlighted with a gray shade, while newly isolated species with > 2 strains are highlighted with a brown shade. Based on fragment recruitment of > 1700 freshwater metagenomes, representative genomes of highly abundant and widespread species (> 1× Cov/Gb in > 20% of metagenomes) are marked with red stars and medium abundant and widespread (> 1× Cov/Gb in ~10% of metagenomes) are marked with blue stars. Nucleotide identity between genomes is represented by a color scale (90–100% identity). Genomic islands (GI) are marked with different colors in the genome alignments and numbered above the alignments. dS profiles of each species with > 3 strains are shown below each alignment group using one genome as reference, tRNAs are shown as dark blue arrows, and rRNA is represented by orange arrows. Clonal genomes have been removed for simplicity. LanE-43 and Bala-9 are representative genomes for *P. landstejnensis* (clonal isolates Bala-138, Bala-106, Bala-103, Bala-114, Bala-156, Bala-19, Bala-220, Bala-242, Bala-34, Bala-43, Bala-69, Bala-80, Bala-132, Bala-209) and MKE-5 also represents the clonal MRE-22 strain

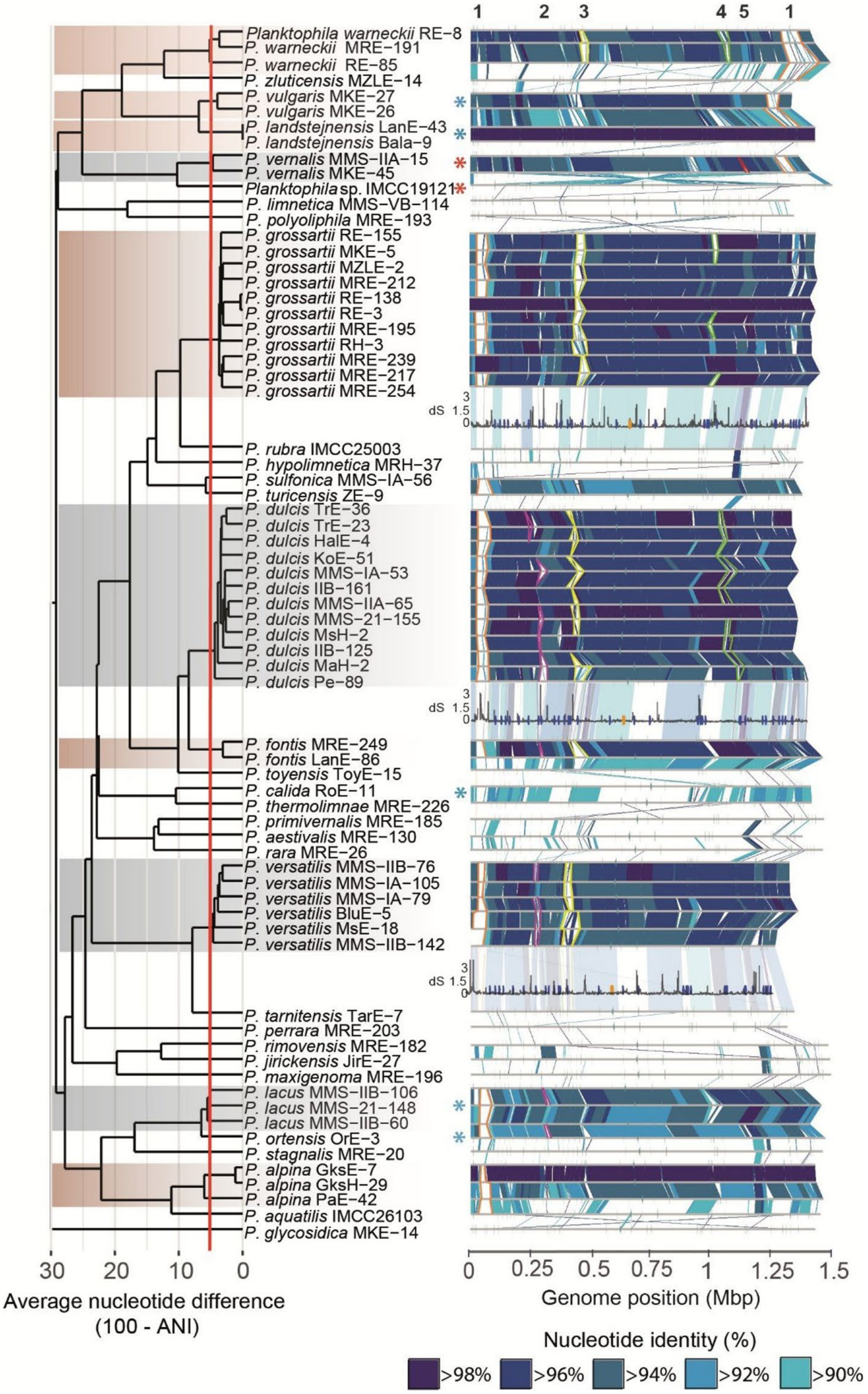

Fig. 4 (See legend on previous page.)

less over the seasons, with one exception in the summer of 2017, when a unique peak of *Nanopelagicus* species was observed. In the hypolimnion, the occurrence of both *Planktophila* and *Nanopelagicus* members was positively associated with dissolved oxygen and pH, and negatively with nitrate, total and soluble reactive phosphorus (Suppl. Table 10; Suppl. Fig. 14). In Lake Breiter Luzin, cluster 3 of *Planktophila* (*P. vernalis*, *Planktophila* sp. IMCC19121) and *N. abundans* and *N. limnes* showed very rhythmic fluctuations in Cov/Gb. A maximum of *Planktophila* cluster 3 appeared in early summer, while *Nanopelagicus* species had a peak in late summer and autumn (Suppl. Fig. 16). As in the epilimnion of Lake Stechlin, most members of both genera were favored by high temperatures. In addition, they showed negative correlations with total chlorophyll *a*, phosphorus, and nitrate concentrations, while *Planktophila* members were negatively impacted by a decrease in pH (Suppl. Fig. 17; Suppl. Table 10). Lake Tiefwaren stood apart from the other investigated German lakes, showing a higher abundance of *Planktophila*, specifically of *P. vernalis* and to a lower extent of *Planktophila* sp. IMCC19121, which are both part of cluster 3, compared to *Nanopelagicus* species (Suppl. Tables 6, 11; Suppl. Fig. 19). Here, the seasonal pattern was not as clear as in the case of Lake Breiter Luzin, as the spring and summer-autumn maxima were sometimes dominated by different groups, although more common was a *Planktophila* peak in spring and a *Nanopelagicus* maximum in summer-autumn (Suppl. Fig. 18). In Lake Tiefwaren, both *N. abundans* and *N. limnes* had a positive association with temperature (Spearman's rho between 0.45 and 0.51, respectively, *P* values < 0.001) and a negative one with dissolved oxygen (Spearman's rho −0.39 and −0.64, respectively, *P* values < 0.001) (Suppl. Fig. 20; Suppl. Table 10). Similar correlations were observed for *P. lacus* and *P. ortensis*, although these taxa were never as abundant. In contrast, *P. vernalis* and *Planktophila* sp. IMCC19121 exhibited opposite patterns (for both species Spearman's rho for temperature −0.34, *P* value < 0.05; for dissolved oxygen Spearman's rho 0.38, *P* value < 0.01).

General metabolic capabilities in *Nanopelagicales* isolates

The genome-based analysis indicated that isolates of all genera had the potential to perform glycolysis (Figs. 2, and 3; Suppl. Tables 12, 13). The complete pyruvate dehydrogenase complex, which is responsible for the oxidation of pyruvate to acetyl-CoA, was present in 48 isolates from all four cultivated genera. The oxidative phase of the pentose phosphate pathway (glucose-6-phosphate to ribulose-5-phosphate) was present only in some *Planktophila* strains (Suppl. Table 13), while the genes involved in the non-oxidative phase (fructose-6-phosphate to ribose-5-phosphate) were present in all groups. Phosphoenolpyruvate carboxylase was detected in all isolates, which is a conserved trait in the *Nanopelagicales* order [3]. Regarding lipid metabolism, the genes responsible for the biosynthesis of fatty acids were present in members of the *Aquilimus*, *Planktophila*, and IMCC26077, while the 3-ketoacyl-ACP (acyl carrier protein) synthase, an enzyme involved in the elongation of fatty acids [71], was not annotated in any *Nanopelagicus* isolates. As *Nanopelagicus* isolates encoded the later steps in the elongation process, they are probably able to perform this reaction using a yet undescribed protein. Members of all groups were able to perform beta-oxidation of fatty acids (Suppl. Tables 12, 13).

Nanopelagicales isolates harbor genes required to synthesize between 17 to 20 amino acids (Fig. 3; Suppl. Table 14); *Planktophila* and *Nanopelagicus* contained genes responsible for cysteine and asparagine biosynthesis, which were lacking in *Aquilimus* and in IMCC26077 genomes. *Nanopelagicus* was auxotrophic for histidine, phenylalanine, and leucine, as opposed to the three other genera (except for *P. perrara* MRE-203 which also lacked a histidine biosynthesis pathway). Members of only six *Planktophila* species were able to de novo synthesize all amino acids. Given the amino acid auxotrophies observed in *Nanopelagicales* members, particularly those other than *Planktophila*, it is likely that they survive by taking up amino acids from their environment. They could achieve this using various ABC transporters, including commonly encoded branched-chain amino acid transporters (*liv*KHMGF), putative polar amino acid transporters, and peptide/nickel transporters. In contrast, *Planktophila* also encodes oligopeptide transporters (*opp*ABCDF) (Fig. 3B; Suppl. Tables 12, 13, 15). Imported peptides can be digested via aminopeptidases encoded by nearly all *Nanopelagicales* isolates (*pep*A, *pep*O, *pep*P, *pep*N, *map*, glutamate carboxypeptidase). Further, all groups, except for *Aquilimus*, encoded for cyanophycinase (Suppl. Table 13), an exopeptidase responsible for the degradation of cyanophycin, which can constitute an important source of carbon and nitrogen in freshwaters [72].

Ten isolates affiliated to *Planktophila*, *Aquilimus*, and IMCC26077 were prototrophs for vitamin B2 (riboflavin). Strains affiliated to different species from all four genera also encoded for the enzymes involved in vitamin B3 (NAD), B6 (pyridoxal-P), and coenzyme A biosynthesis (Suppl. Table 13). Other vitamins and cofactors are probably taken up from the surrounding microenvironment through specific transporters (Suppl. Tables 12, 15). An exception regarding cobalamin (vitamin B12) biosynthesis was identified in five *Planktophila* strains (the three *P. alpina* strains, *P. polyoliphila* MRE-193, and *P. limnetica* MMS-VB-114) that can produce cobalamin from cobinamide through a salvage pathway that was previously reported in both bacteria and archaea [73]. As observed before for the previously available *Nanopelagicales* isolates [3, 23, 24], none of the newly isolated strains, including the newly isolated genus *Aquilimus*, harbored all genes required for assimilatory sulfate reduction. Therefore, *Nanopelagicales* members are probably dependent on exogenous sources of reduced sulfur. Additionally, four *Planktophila* strains (*P. sulfonica* MMS-IA-56, *P. estivalis* MRE-130, *P. grossartii* MRE-217, and *P. grossartii* RE-155) encoded for *ssu*ABC, an ABC transporter that is expressed in sulfate-limiting conditions and can facilitate the uptake of aliphatic sulfonate compounds from the environment [74]. Interestingly, the thiosulfate uptake transporter YeeE [75] was identified in 34 genomes, specifically in all isolates belonging to *P. fontis*, *P. dulcis*, *P. lacus*, and *P. versatilis*, as well as in *A. klicavensis* and ten other *Planktophila* species (Suppl. Table 15). This could imply that thiosulfate might serve as an inorganic sulfur source for their growth, consistent with reports on the freshwater SAR11 genus *Fontibacterium* [29]. An auxotrophy that likely hindered the isolation and stable cultivation of *Nanopelagicus* species, both in this study and in a previous one [3], is their requirement for purine precursors. Our analysis revealed that all *Nanopelagicus* species lack the pathway for de novo purine biosynthesis (Fig. 2; Suppl. Tables 12, 13). However, the artificial lake water medium used here did not contain purine precursors such as inosine 5′-monophosphate (IMP). Future cultivation efforts targeting this genus should therefore supplement the medium with purines or purine precursors.

An electron transport chain composed of NADH: quinone oxidoreductase, succinate dehydrogenase, and cytochrome *c* oxidase was encoded in all isolates, with oxygen as terminal electron acceptor, and an F-type ATP synthase (Fig. 3; Suppl. Table 12). Additionally, some members of *Aquilimus* and *Planktophila* encode the subunits of cytochrome *bd* ubiquinol oxidase (Fig. 3A), which probably provides an increased tolerance for reduced oxygen concentrations [76]. Further, actinorhodopsins and heliorhodopsins were encoded by all isolates (Suppl. Fig. 22), and most *Planktophila* and *Aquilimus* members were likely able to de novo synthesize retinal (Fig. 3, Suppl. Table 16). Therefore, they are probably capable of using light to generate a proton gradient, which can be used either to produce energy or to drive membrane transport processes.

The ecological success of *Planktophila*

The reasons for the dominance of *Planktophila* in freshwater systems are far from being understood. Here, we observed that genomes with high recruitment values (Suppl. Table 6) did not have the oxidative part of the pentose phosphate pathway (ox-PPP) (Fig. 3A), regardless of phylogenetic distance among them (Fig. 4; Suppl. Table 13). The intermediates or products of ox-PPP could be obtained in strains missing this pathway with the use of ABC transporters specific to ribose (*rbs*ABC), fructose (*frc*ABC), and xylose (*xyl*FHG) (Fig. 3A; Suppl. Tables 12, 15). Additionally, cluster analysis of proteins encoded by abundant versus non-abundant *Planktophila* isolates revealed that the dominant strains encode for a distinct and specific RbsK sugar kinase, responsible for the phosphorylation of D-ribose to D-ribose 5-phosphate, and for a xylose isomerase which can produce xylulose from imported xylose. Moreover, although all genomes encode for transketolase, a key enzyme in the non-oxidative part of the PPP (non ox-PPP), clustering analysis showed that some abundant strains encode for a distinct type of transketolase, which is located in a replacement genomic island (GI1—see below) (Suppl. Table 17; Fig. 4). We hypothesize that the transketolase, RbsK sugar kinase, and xylose isomerase enhance fitness through greater efficiency of the non ox-PPP. Additionally, the most abundant *Planktophila* species, *P. vernalis* and *Planktophila* sp. IMCC19121, are able to synthesize all 20 amino acids (Suppl. Tables 6, 12, 14), a capacity that would provide them with a higher level of independence from co-occurring microbes [77].

Besides the aforementioned ABC-type sugar transporters, an unusually high number of transporters for a variety of low molecular weight organic compounds was predicted in the genomes of *Planktophila* (31 different transporters, including some exclusively found in *Planktophila*, such as putative sorbitol/mannitol transporters, glycerol and erythritol transporters, a sulfonate transporter, a cellobiose transporter, and an ABC-type multiple sugar transporter) compared to other *Nanopelagicales* genera with similar genome sizes (eight–13 different transporters) (Fig. 3B; Suppl. Tables 12, 14). Additionally, considering the streamlined genomes of *Planktophila*, they encode not only a high diversity of carbohydrate-active enzymes (CAZy; median = 23 different CAZy per genome), but also a relatively high copy number per genome (median = 38 total CAZy), several times more than any other *Nanopelagicales* group (Fig. 3C; Suppl. Table 18). Although the classification of CAZy is based on sequence similarity and not substrate affinity [78] and therefore the precise enzymatic activity is difficult to predict, some occurrence pattern can be distinguished among genera (Suppl. Table 18). Carbohydrate-binding module (CBM) family proteins were encountered only in *Planktophila*, including CBM41 and CBM91 which are likely to have pullulanase and β-xylosidase activity. As in the case of all CAZy families, carbohydrate esterases (CE) were more diversified in *Planktophila*. Two enzymes that were only found in this genus were CE15, a 4-O-methyl-glucuronoyl esterase that occurs in cellulolytic organisms and cleaves the bond between glucuronoxylan and lignin alcohols [79], and CE7, which probably has acetylxylan esterase activity. Leaving aside the CAZy for which substrate utilization is difficult to predict, those families with specific activity ranges suggest that *Planktophila* may be capable of degrading and using plant polymers such as hemicellulose as a growth substrate. However, considering the diversity of CAZy, their substrate range is very likely much broader. Another advantage of *Planktophila* over the other genera might be given by the capacity for one-carbon metabolism as several of the most abundant *Planktophila* species, including the two *P. vernalis* strains, *Planktophila* sp. IMCC19121, *P. limnetica*, and *P. polyoliphila*, encoded all subunits for formate dehydrogenase and molybdenum cofactor biosynthesis from guanosine 5′-triphosphate (GTP), glycine hydroxymethyltransferase, methylenetetrahydrofolate dehydrogenase, and formate-tetrahydrofolate ligase. Additionally, the topmost abundant *Planktophila* species (*P. vernalis*, *P. vulgaris*, and *P. landstejnensis* among others) encoded for gamma-glutamyl transpeptidases.

Since up to 20 different *Planktophila* species co-occur at the same times of the year, we investigated the microdiversity contained in genomic islands (GIs) within this genus to evaluate the extent of metabolic diversification and niche specialization that might enable the coexistence of various species, as previously explored by Neuenschwander et al. [3]. To predict GIs in the genomes of *Planktophila* strains and identify hypervariable regions, we aligned 69 genomes in a pairwise manner that followed phylogenetic relationships (Fig. 4). Because hypervariable regions can be better predicted in complete genomes, we excluded MAGs or SAGs from the analysis. In some cases, due to a high evolutionary distance and low nucleotide identity, genomes of certain species (e.g., *P. glycosidica*, *P. stagnalis*, *P. hypolimnetica*) were difficult to compare to others. Additionally, we found six large inversion events which made comparing the genome architecture challenging. Nevertheless, 10 species containing between two and 12 closely related genomes allowed the identification of five major GIs (containing ≥10 genes; named GI1–GI5). Four were present across multiple species (GI1–GI4), while one particularly important insertion event was only present in *P. vernalis* MKE-45 (GI5) (Fig. 4; Suppl. Table 17). Reassuringly, our results reproduced and expanded the inter-species diversity and microdiversity previously described for four *Planktophila* species [3] (Fig. 4).

One large replacement GI (GI1, containing between 13 and 50 genes) encoded functions related to synthesis and modification of cell surface structure, as well as for a specific transketolase in strains missing the ox-PPP, as mentioned before. GIs 2 to 4 were additive and flanked by transfer RNAs (tRNAs). GI2 (between 4 and 10 genes) contained genes encoding for the pilus assembly proteins CpaF and TadBCG, but also for the ABC-type transporter AraH which enables the import of L-arabinose and for the broad activity transcriptional regulator IclR [80]. It also contained a high proportion of hypothetical proteins which prevented a detailed analysis of its genomic content (Suppl. Table 17). GI3 (between 3 and 44 genes) encoded for a variety of type I (e.g., HcdS, HsdM) and II (BglII) restriction endonucleases and site-specific DNA methylases (Dcm). Additionally, it encoded for DarT, a single-stranded DNA (ssDNA) thymidine ADP-ribosyltransferase involved in the stress response system [81], and NucS—an endonuclease able to recognize and cleave double-stranded DNA (dsDNA) containing mismatched base pairs [82]. Therefore, this island was predicted to have functions related to phage defense and the maintenance of chromosomal integrity. However, in *P. versatilis* strains IIB-76 and BluE-5, it contained sugar transporters and in the latter case, also sugar isomerases/epimerases. GI4 (between four and 43 genes) had very distinctive gene sets in different species. In *P. dulcis*, it encoded for ABC-type branched-chain amino acid transporters (LivKMH), while in *P. warneckii*, it encoded for transcriptional regulators (e.g., MarT, PurB), response regulators (e.g., OmpR, BaeS), and various ABC-type sugar (UgpABE) and di-/oligopeptide (e.g., DppABE) transporters. In *P. grossartii*, only hypothetical proteins are encoded in GI4 (Suppl. Table 17). Interestingly, in the genomes of the most ubiquitous and abundant isolate (*P. vernalis* MKE-45, Suppl. Table 6), we detected a GI (GI5, 11 genes) that encoded for all subunits of respiratory nitrate reductase NarGIJK, the Nar two-component regulatory system, and two copies of the NarK nitrate/nitrite antiporter (Suppl. Table 17).

Discussion

In the current study, we isolated and genome-sequenced 72 new *Nanopelagicales* strains from 20 locations in Europe and Japan. They were classified into 24 new *Planktophila* species and four novel species of the previously uncultured *Aquilimus* genus (AAA044-D11 in GTDB) (Suppl. Tables 4, 7). Their complete, circular chromosomes allow for a better understanding of genomic organization, especially regarding *Planktophila* members, expanding the general characteristics described previously for the 20 isolates obtained from Lake Zurich and Soyang [3, 23, 24, 26]. All genomes showed typical streamlining traits such as small size (< 2 Mbp), high coding density (> 95%), and low GC content (< 50%) (Suppl. Fig. 2; Suppl. Table 9), all being adaptations to nutrient-limited lake environments [1, 32]. These features align with small, non-motile cells with low replication rates, suited for oligotrophic conditions. Unsurprisingly, only two to four sigma factors were encoded, reflecting minimal regulatory needs under the relatively stable conditions of the pelagic environments. Although these environments are nutrient-limited (oligotrophic), their nutrient scarcity is persistent rather than fluctuating, reducing the need for extensive regulatory flexibility [83].

An additional adaptation to oligotrophic environments is the use of light-driven, proton-pumping rhodopsins for producing energy [84], which are commonly encoded in both marine and freshwater pelagic microbes [66, 85, 86]. One copy of the gene for green light absorbing actinorhodopsins [87] was present in all *Nanopelagicales* isolates (Suppl. Fig. 22). However, only *Planktophila* (with a few exceptions) and the newly isolated genus *Aquilimus* can de novo synthesize retinal, the covalently bound chromophore necessary for rhodopsin activity. As previously observed [3, 23], *Nanopelagicus* members encode for all steps in the biosynthesis of β-carotene, missing only the last gene (*blh*) for its conversion to retinal (Suppl. Table 18). Whether these bacteria rely on exogenous retinal or, as described for *Aurantimicrobium minutum*, encode a hitherto unknown *blh* equivalent is currently unknown [88]. In addition, all our isolates encode another rhodopsin, heliorhodopsin. The function of this protein is not yet completely understood, but it has been associated with a range of regulatory activities, from multidrug resistance and light modulation of ABC transporters to light-induced membrane lipid modifications and protection against oxidative stress [89–91].

Many studies have characterized the abundant freshwater order *Nanopelagicales* and achieved important results regarding seasonality [12–14, 92], diversity, nutrient uptake [15, 93–95], and trophic state preferences of these microorganisms [3, 96]. However, the resolution of these studies was limited due to the use of various FISH probes, some of which failed to detect between nine and 40% of *Nanopelagicales* cells [14] or were unable to distinguish between different clades, such as acI-A and acI-B (now known as *Planktophila* and *Nanopelagicus*), by merging their signals. Only one study used genus- or even species-specific probes [3].

Recent metagenomic studies allowed for species- and strain-level resolution [69, 70, 97, 98]; however, because these studies relied on short reads, they have been unable to link the success of certain groups to specific metabolic adaptations encoded in GIs. Over the years, studies that approached the seasonality of *Nanopelagicales* obtained sometimes conflicting results [11, 14, 15]. In general, two distinct peaks were observed in many freshwater lakes, one during spring and summer, and another one in the autumn [3, 12, 15]. Between 2003 and 2004, Allgaier and Grossart [14] monitored *Planktophila* sp. (acI-A) and *Nanopelagicus* sp. (acI-B) in four northeastern German lakes (Lakes Stechlin, Grosse Fuchskuhle, Breiter Luzin, Tiefwaren) using FISH probes. Distinct peaks occurred during May to June and October to November, showing a seasonal pattern that closely resembles the trends we observed through recruitment analysis from 2012 to 2020 in the same lakes (Suppl. Figs. 12, 16, 19; Suppl. Table 6). One notable exception is the absence of any of the isolated species in dystrophic Lake Grosse Fuchskuhle in the present analysis compared with notable occurrences in the early FISH-based study [14]. This hints at other, low-pH adapted species of *Nanopelagicales* that so far escaped cultivation [18]. *Nanopelagicus abundans*, *N. limnes*, *P. vernalis*, and *Planktophila* sp. IMCC19121 were the dominant actinobacterial species in the other three German lakes (Suppl. Table 6), and *Planktophila* species showed a significant positive correlation with pH (Suppl. Figs. 14, 17, 20). A similar observation was reported in a study that explored the bacterial diversity in 18 lakes over a pH gradient (5.1–8.9), showing that *Planktophila* were restricted to pH ranges between 6.8 and 8.9 and reached the highest abundances between pH 8.6 and 8.9 [11], which is in line with the total absence of any of the *Planktophila* isolates in dystrophic lakes ($n = 554$, Suppl. Table 6). One particularly surprising isolate in this context is *P. jirickensis* JirE-27, which was obtained from the slightly acidic Jiřická Pond in the Czech Republic (pH 6.5 at the time of isolation). However, this strain was successfully isolated and has since been maintained in media with a pH of 7.5. Recruitment analysis showed that this species has a low abundance in freshwater, and in our analyzed dataset it is restricted to Jiřická Pond (Suppl. Table 6). The two seasonal peaks of *Nanopelagicales* were previously observed to co-occur with picocyanobacterial or algal blooms (e.g., *Asterionella formosa*) [3, 14, 16], as it was also the case in our study for *P. rubra*, *P. grossartii*, *P. zluticensis*, and *P. warneckii* (Suppl. Table 10), or they followed protistan grazing events [15, 99] during which *Nanopelagicales* cells might be protected from the top-down control due to their minuscule size [100, 101] and Gram-positive cell walls [102]. At this point, the high number of diverse CAZy and membrane transporters for small molecular weight organic compounds (Fig. 3B, C) could favor the growth of *Planktophila* by breaking down and importing algal exudates or byproducts released through protistan predation or viral lysis [3, 99]. Additionally, because some diatom blooms can surpass very low phosphorus concentrations and might act as a phosphorus sink in oligotrophic environments [103], they might favor the development of *Nanopelagicales* populations, which are particularly well adapted to survive nutrient-limiting conditions [1, 15].

Considering the high global abundance of only a few *Planktophila* species (Fig. 1B), we analyzed their genomes to understand what advantages they might have over other related strains or genera. We could identify some genomic traits present only in globally abundant *Planktophila* strains, while being absent in others. One factor was the repeated loss of the oxidative phase of the pentose phosphate pathway (ox-PPP) during the diversification of *Planktophila* species (Fig. 3; Suppl. Tables 6, 12, 13). This has been previously reported for *Planktophila* species, except for *P. dulcis* and *P. sulfonica*, which were not abundant in the analyzed dataset; likewise, the pathway was absent in all *Nanopelagicus* isolates [3]. As a major consumer of glucose in the cell [104], the loss of ox-PPP suggests the utilization of glucose in these microbes preferentially for energy production through glycolysis, tricarboxylic-acid (TCA) cycle, and electron transport chain (ETC), rather than in anabolic processes like the biosynthesis of ribose-5 phosphate, nucleotides, or erythrose 4-phosphate for the production of various vitamins and aromatic amino acids [104–106]. The absence of this pathway was also reported for other abundant microorganisms in oligotrophic environments, such as marine and freshwater SAR11 [107], implying that it could represent a common adaptation to nutrient limitations. Although the products of ox-PPP could be obtained from the surrounding microenvironment using ABC-type transporters for ribose (RbsABC), fructose (FrcABC), and xylose (XylFHG) (Fig. 3A; Suppl. Tables 12, 15), another concern might be that the loss of ox-PPP could lead to NADPH depletion in the cell. Indeed, some glucose-6-phosphate dehydrogenase-deficient bacterial strains exhibit slower growth rates compared to wild types and an increased sensitivity to oxidative stress [108]. However, adding glucose to the medium redirected the carbon flow to the glycolysis and the non ox-PPP in an *Escherichia coli* ox-PPP knock-out mutant, while also enhancing the activity of the TCA cycle. This redirection resulted in a higher growth rate compared to the wild-type strain [106, 108]. Other enzymes might compensate for NADPH regeneration, such as the malate dehydrogenase and the isocitrate dehydrogenase in the TCA cycle, or the NAD^+ kinase, universally present in *Planktophila*, which catalyzes the conversion of NAD^+ to $NADP^+$, and it is essential for $NADP^+$ and NADPH availability in the cell [108]. The redirection of carbon flow was experimentally proven in different bacteria, and we therefore hypothesize that it is one of the metabolic adaptations in successful *Planktophila* and *Nanopelagicus* species. Nevertheless, because the ox-PPP is also absent in the newly isolated genus *Aquilimus* and in IMCC26077 (Fig. 3), its loss alone might not be the only key to success. Another trait common to *Planktophila* species is the capacity for one-carbon interconversions. Specifically, serine/glycine hydroxymethyltransferase, an enzyme encoded in the genome of all *Nanopelagicales* isolates (Suppl. Table 13), catalyzes the reversible conversion between serine and glycine, generating one-carbon units via tetrahydrofolate [109]. Importantly, the serine produced can be catabolized by serine dehydratase, encoded by several *Planktophila* members and the IMCC26077 isolate, into pyruvate [110], which is then converted to acetyl-CoA and enters the TCA cycle, thereby providing a mechanistic link in which one-carbon metabolism can modulate the flux into central carbon metabolism. Additionally, it can be speculated that glutathione usage, which can be found in nanomolar concentrations in freshwater lakes [111], could act as a source of amino acids and sulfur to the same abundant *Planktophila* species. Although not many specific uptake mechanisms for glutathione have been characterized so far [112], several widespread ABC-type transporters for oligopeptides/dipeptides, similar to those characterized in *E. coli*, might also perform this function [113] and be responsible for the import of glutathione into the *Nanopelagicales* cells (Fig. 3; Suppl. Tables 12, 15). Additionally, genes encoding for gamma-glutamyl transpeptidases, which are typically involved in glutathione cleavage [113] and could provide the cells with an extra source of glycine, glutamate, cysteine, and therefore sulfur, were identified in some abundant *Planktophila* species, specifically *P. vernalis*, *P. vulgaris*, and *P. landstejnensis*. While the presence of these genes does not confirm expression or activity, their occurrence may suggest a potential role in providing additional survival advantages. Nonetheless, the fact that few *Planktophila* strains encode pathways for the biosynthesis of all 20 amino acids could render them less dependent from functional cohorts than previously thought [114].

Core and pangenome analysis of 84 *Planktophila* strains (68 new from this study and 16 previously described ones [3, 23]) revealed a conserved core of 736 genes and an open pangenome that reaches >6800 genes (Suppl. Fig. 3B). The high difference between the core and pangenome suggests the existence of a wide variety of ecotypes, similar to the case of marine *Prochlorococcus* [115, 116]. Also, their high abundance in lacustrine environments and the co-occurrence of different species at the same time and place (Suppl. Table 6) probably creates a very large gene pool available for inclusion by mechanisms such as transfection or horizontal gene transfer, which is reflected in the extended GIs with intra- and interspecific distribution (Fig. 4; Suppl. Table 17) and the almost ubiquitous occurrence of the ComEF operon for natural transformation [117]. The detection of phages, potentially infecting *Nanopelagicales* in metagenomes [68, 118], and alongside *Nanopelagicales* SAGs [119], together with the recent isolation of *Planktophila* phages from the Římov Reservoir (Czech Republic) [120] supports the idea that these strains may have acquired at least some of their genes from past viral infections, even though we did not find any complete prophages or genes for CRISPR-Cas systems in their genomes. These are not unique observations for *Nanopelagicales*, but rather widespread characteristic of oligotrophic, pelagic species (e.g., *Pelagibacter*, *Fontibacterium* [29, 121]). Similar to SAR11, it was previously suggested that the high microdiversity and recombination frequencies account for large *Nanopelagicales* populations in freshwater environments, despite increased levels of viral predation [3, 120, 122]. The large number of *Planktophila* strains ($n = 84$) from 33 species allowed a detailed analysis of metabolic microdiversification found in GIs (Fig. 4). GI4 encoded a variety of ABC-type transporters for sugars, amino acids, and oligopeptides, while GI1 encoded numerous CAZys, especially GTs (glycosyltransferases) (Suppl. Table 17). *Planktophila* members alone also encoded for a vast variety of CAZy outside GIs (Suppl. Table 18). Although it is difficult to assess the substrate specificity for most of the CAZy identified in the *Nanopelagicales* isolates [123], there are at least a few (CE15, CE7, CBM41, CBM91) which suggest that *Planktophila* members are capable to degrade and utilize abundant plant polymers, such as lignin, cellulose, and hemicellulose [118, 124], to support their growth, giving them an advantage over other *Nanopelagicales* lineages that lack this CAZy repertoire. Additionally, the respiratory nitrate reductase (NarGYIJ), the Nar two-component regulatory system, and the NarK nitrate/nitrite antiporters encoded in GI5 in *P. vernalis* (Suppl. Table 11), together with the presence of a formate dehydrogenase, let us believe that the most abundant and ubiquitous *Planktophila* species in freshwater ecosystems can tolerate periods of oxygen limitation by reducing nitrate to nitrite coupled with the oxidation of quinols and the generation of a transmembrane proton gradient [125]. Moreover, GI1 and GI2 encoded for functions related to surface structures, such as the biosynthesis and modification of the cell wall, and for pilus assembly proteins. The evasion of ingestion and/or digestion by flagellates [101], the exposure to high frequency of viral attacks, and the existence of large free-living populations suggest that *Planktophila* species are following a constant-diversity evolutionary model [126]. Therefore, the variable genomic regions play at least two important roles in these pelagic populations: (i) they allow different, concurrent, but suboptimal lineages to exploit all available resources, and (ii) they help the population survive high phage pressure by introducing diversity in phage recognition sites through modifications in the glycosylation patterns of pilus and cell wall surfaces [126].

Conclusion

We cultured and genome-sequenced 72 new *Nanopelagicales* isolates from 20 locations, encompassing 24 new *Planktophila* species and four novel species of the here described *Aquilimus* genus, thereby largely increasing the isolated diversity of this abundant genome-streamlined freshwater group. By analyzing >1700 freshwater metagenomes, we demonstrated that *P. vernalis* and *N. abundans* are the most ubiquitous and abundant cultivated *Nanopelagicales* species in freshwater lakes worldwide. In time-series analyses, both showed consistent temporal patterns, with two annual peaks corresponding with spring and autumn blooms, but the combinations of available nutrients and environmental conditions, particularly pH, could favor the dominance of one over the other. The metabolic adaptations that could contribute to the success of *Nanopelagicales* in oligotrophic conditions include the presence of rhodopsins and the loss of the oxidative phase of the pentose phosphate pathway, indicating a strategic emphasis on energy production rather than anabolic processes. We further identified potential genomic traits that are linked to the ubiquity of some *Planktophila* strains (e.g., the presence of a respiratory nitrate reductase and nitrate/nitrite antiporters, suggesting the ability of *P. vernalis* MKE-45 to survive hypoxic conditions). Generally, genes encoded in genomic islands account for a high metabolic microdiversification that probably enhances resource utilization and survival under high phage predation pressure. We also provided further evidence that the coexistence of *Planktophila* strains is linked to differences in their CAZy repertoire enabling niche separation within an ecosystem and temporal coexistence. Overall, these findings reinforce the ecological significance of *Nanopelagicales* in freshwater ecosystems worldwide and their complex adaptations, contributing to a deeper understanding of their role in aquatic microbial communities. Importantly, while the results presented here are based on genomic analyses, the stable cultivation of these species will enable future studies of their physiology and allow testing of our hypotheses, thereby enhancing our understanding of these key freshwater microbes.

Abbreviations

6PG	6-Phosphogluconate
6PGL	6-Phosphogluconolactone
AAI	Average amino acid identity
ADP	Adenosine diphosphate
ANI	Average nucleotide identity
ATP	Adenosine triphosphate
CAZy	Carbohydrate-active enzymes
CBM	Carbohydrate-binding module
CE	Carbohydrate esterases
ComEF	Competence proteins E and F
Cyt bd	Cytochrome *bd*
Cyt C	Cytochrome *c*
DAPI	4′,6-Diamidino-2-phenylindole

ETC	Electron transport chain
FAD/$FADH_2$	Flavin adenine dinucleotide
Formate DH	Formate dehydrogenase
GA3P	Glyceraldehyde 3-phosphate
GI	Genomic island
GT	Glycosyltransferases
GTP	Guanosine triphosphate
HMMs	Hidden Markov models
IMP	Inosine 5′-monophosphate
MAG	Metagenome-assembled genome
NAD/NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NarK	Nitrate/nitrite antiporter
Non ox-PPP	Non-oxidative branch of the pentose phosphate pathway
Ox-PPP	Oxidative branch of the pentose phosphate pathway
PPP	Pentose phosphate pathway
S-7P	Sedoheptulose 7-phosphate
SAG	Single-amplified genome
Sec	General secretion system
Tat	Twin-arginine translocation pathway
TCA cycle	Tricarboxylic acid cycle
TKT	Transketolase
TM	Transmembrane helix

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-025-02272-x.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgements
We thank Pavel Rychtecký, Patrick Pejsar, the PELAGICS sampling team, local fishermen, captains and technicians for their help during sampling and K. Kabeláčová, F. Kostanjšek, and A. Férová for excellent laboratory support. Sampling in Lake Toya was supported by a joint usage with Toya Lake Station of Field Science Center for Northern Biosphere, Hokkaido University. We thank the LIMNOS and IGB-Dept. 3 team for data analysis and routine sampling of Lake Stechlin, Grosse Fuchskuhle, Breiter Luzin and Tiefwaren.

Authors' contributions

M-CC analyzed the data and wrote the manuscript. MH, MMS and RG commented and helped edit the initial draft. MH conceived the study. PL, MH, MMS and RG helped with data analysis. RG contributed to software development. MH, CF, PL and MMS isolated and maintained the strains. Sampling was performed by M-CC, CF, PL, TSN, VK, YO, JW, H-PG, KP, PZ, BS, CC, SO, RS, CL, CP, HT, DPRH, MŚ, AB, HLB, MZ, AW, RG, MMS, MH. SEM imaging was performed by VK. All authors commented and agreed to the final version.

Funding
This study was supported by Czech Science Foundation (GAČR) grants 20-12496X (RG), 21-21990S (MH, MCC), 22-03662S (CF, MMS), 24-12912M (MCC), 25-15813S (CF, MMS), and 22-33245S (PZ). PL was supported by the grant 24-12912M (GAČR). CF was also supported by the grant 017/2022/P (Grant Agency of the University of South Bohemia in České Budějovice). JW and H-PG were supported by a seed money project from Leibniz-IGB and the DFG-Pycnotrp project (GR1540/37-1). HLB was supported by the Romanian Ministry of Research, Innovation and Digitalization within the PNRR-III-C9-2022 – I5 project code 18/16.11.2022, contract no. 760010/30.12.2022. YO was funded by JST FOREST program (JPMJFR2273) and KAKENHI program (25K18161). SO was partially supported by the project STIM–REI (Contract Number: KK.01.1.1.01.0003), a project funded by the European Union through the European Regional Development Fund—the Operational Programme Competitiveness and by Croatian Science Foundation under the project MALENA (HRZZ IP-2020–02–9021).

Data availability

Genomes from cultures have been submitted to ENA under Study accession numbers PRJEB77526 and PRJEB57446 and metagenomes under ENA Study accession numbers PRJEB86000-PRJEB86004. Novel species were registered at SeqCode.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details
[1]Institute of Hydrobiology, Biology Centre CAS, Ceske Budejovice, Czech Republic. [2]Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic. [3]Institute for Chemical Research, Kyoto University, Kyoto, Japan. [4]Department of Plankton and Microbial Ecology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Zur Alten Fischerhuette 2, Stechlin 16775, Germany. [5]Institute of Biochemistry and Biology, Potsdam University, Maulbeerallee 2, Potsdam 14469, Germany. [6]Department of Fisheries Oceanography and Marine Ecology, National Marine Fisheries Research Institute, Gdynia 81-332, Poland. [7]Research Department for Limnology, Mondsee, University of Innsbruck, Mondsee 5310, Austria. [8]Water Research Institute, National Research Council (IRSA-CNR), Molecular Ecology Group (MEG), Largo Tonolli 50, Verbania 28922, Italy. [9]Division of Materials Chemistry, Ruder Bošković Institute, Bijenička Cesta 54, Zagreb 10000, Croatia. [10]Center of Excellence for Science and Technology-Integration of Mediterranean Region, Zagreb, Croatia. [11]University of Montenegro, Cetinjski Put 2, Podgorica 81000, Montenegro. [12]Universität Innsbruck, Department of Ecology, Innsbruck, Austria. [13]Lake and Glacier Ecology Research Group, Department of Ecology, Universität Innsbruck, Innsbruck 6020, Austria. [14]Laboratoire Microorganismes: Génome Et Environnement, CNRS, Université Clermont Auvergne, Clermont-Ferrand 63000, France. [15]Centre for Limnology, Estonian University of Life Sciences, 6117 Vehendi, Tartu County, Estonia. [16]Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Seestrasse 15, Rostock 18119, Germany. [17]Hydrobiological Station, Faculty of Biology, University of Warsaw, Pilchy 5, Pisz 12-200, Poland. [18]Department of Hydrobiology, Faculty of Biology, Institute of Ecology, Biological and Chemical Research Centre, University of Warsaw, Zwirki I Wigury 101, Warsaw 02-089, Poland. [19]Department of Molecular Biology and Biotechnology, Faculty of Biology and Geology, Babeş-Bolyai University, 5-7 Clinicilor Street, Cluj-Napoca 400006, Romania. [20]Centre for Systems Biology, Biodiversity and Bioresources, Babeş-Bolyai University, 5-7 Clinicilor Street, Cluj-Napoca 400006, Romania.

Received: 25 June 2025 Accepted: 29 September 2025

References

1. Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for microbial ecology. ISME J. 2014;8:1553–65.
2. Luo H, Thompson LR, Stingl U, Hughes AL. Selection maintains low genomic GC content in marine SAR11 lineages. Mol Biol Evol. 2015;32:2738–48.
3. Neuenschwander SM, Ghai R, Pernthaler J, Salcher MM. Microdiversification in genome-streamlined ubiquitous freshwater actinobacteria. ISME J. 2018;12:185–98.
4. Chiriac M-C, Haber M, Salcher MM. Adaptive genetic traits in pelagic freshwater microbes. Environ Microbiol. 2023;25:606–41.
5. Morris JJ, Lenski RE, Zinser ER. The black queen hypothesis: evolution of dependencies through adaptive gene loss. MBio. 2012. https://doi.org/10.1128/mbio.00036-12.
6. Warnecke F, Amann R, Pernthaler J. Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. Environ Microbiol. 2004;6:242–53.
7. Zwart G, et al. Nearly identical 16S rRNA sequences recovered from lakes in North America and Europe indicate the existence of clades

of globally distributed freshwater bacteria. Syst Appl Microbiol. 1998;21:546–56.
8. Hugerth LW, et al. Metagenome-assembled genomes uncover a global brackish microbiome. Genome Biol. 2015;16:279.
9. Mehrshad M, Amoozegar MA, Ghai R, Shahzadeh Fazeli SA, Rodriguez-Valera F. Genome reconstruction from metagenomic data sets reveals novel microbes in the brackish waters of the Caspian Sea. Appl Environ Microbiol. 2016;82:1599–612.
10. Salka I, et al. Distribution of acI-actinorhodopsin genes in Baltic Sea salinity gradients indicates adaptation of facultative freshwater photoheterotrophs to brackish waters. Environ Microbiol. 2014;16:586–97.
11. Newton R, Jones S, Helmus M, Mcmahon K. Phylogenetic ecology of the freshwater actinobacteria *acI* lineage. Appl Environ Microbiol. 2007;73:7169–76.
12. Glöckner FO, et al. Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. Appl Environ Microbiol. 2000;66:5053–65.
13. Warnecke F, Sommaruga R, Sekar R, Hofer JS, Pernthaler J. Abundances, identity, and growth state of actinobacteria in mountain lakes of different UV transparency. Appl Environ Microbiol. 2005;71:5551–9.
14. Allgaier M, Grossart H-P. Diversity and seasonal dynamics of actinobacteria populations in four lakes in northeastern Germany. Appl Environ Microbiol. 2006;72:3489–97.
15. Salcher MM, Pernthaler J, Posch T. Spatiotemporal distribution and activity patterns of bacteria from three phylogenetic groups in an oligomesotrophic lake. Limnol Oceanogr. 2010;55:846–56.
16. Salcher MM, Posch T, Pernthaler J. In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. ISME J. 2013;7:896–907.
17. Ghai R, McMahon KD, Rodriguez-Valera F. Breaking a paradigm: cosmopolitan and abundant freshwater actinobacteria are low GC. Environ Microbiol Rep. 2012;4:29–35.
18. Garcia SL, et al. Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. ISME J. 2013;7:137–47.
19. Hamilton JJ, et al. Metabolic network analysis and metatranscriptomics reveal auxotrophies and nutrient sources of the cosmopolitan freshwater microbial lineage acI. mSystems. 2017;2:e00091–17.
20. Ghai R, et al. Metagenomics of the water column in the pristine upper course of the Amazon River. PLoS ONE. 2011;6:e23785.
21. Partensky F, Hoepffner N, Li WKW, Ulloa O, Vaulot D. Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) strains isolated from the North Atlantic and the Mediterranean Sea. Plant Physiol. 1993;101:285.
22. Rappé MS, Connon SA, Vergin KL, Giovannoni SJ. Cultivation of the ubiquitous *SAR11* marine bacterioplankton clade. Nature. 2002;418:630–3.
23. Kang I, Kim S, Islam MR, Cho J-C. 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. 2017;7:42252.
24. Kim S, Kang I, Seo J-H, Cho J-C. Culturing the ubiquitous freshwater actinobacterial acI lineage by supplying a biochemical 'helper' catalase. ISME J. 2019;13:2252–63.
25. Kim S, Park MS, Song J, Kang I, Cho J-C. High-throughput cultivation based on dilution-to-extinction with catalase supplementation and a case study of cultivating acI bacteria from Lake Soyang. J Microbiol. 2020;58:893–905.
26. Kim S, et al. Heme auxotrophy in abundant aquatic microbial lineages. Proc Natl Acad Sci U S A. 2021;118:e2102750118.
27. Salcher M, Šimek K. Isolation and cultivation of planktonic freshwater microbes is essential for a comprehensive understanding of their ecology. Aquat Microb Ecol. 2016;77:183–96.
28. Salcher MM, et al. Bringing the uncultivated microbial majority of freshwater ecosystems into culture. Nat Commun. 2025;16:7971.
29. Fernandes C, et al. Ecophysiology and global dispersal of the freshwater SAR11-IIIb genus *Fontibacterium*. Nat Microbiol. 2025;10:2194–206.
30. Paul Layoun et al. Flexible genomic island conservation across freshwater and marine Methylophilaceae. TheISME Journal 18, wrad036 (2024).
31. Porter KG, Feig YS. The use of DAPI for identifying and counting aquatic microflora. Limnol Oceanogr. 1980;25:943–8.
32. Salcher MM, Schaefle D, Kaspar M, Neuenschwander SM, Ghai R. Evolution in action: habitat transition from sediment to the pelagial leads to genome streamlining in Methylophilaceae. ISME J. 2019;13:2764–77.
33. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr Protoc Bioinformatics. 2020;70:e102.
34. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
35. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics. 2010;11:431.
36. Blum M, et al. Interpro: the protein sequence classification resource in 2025. Nucleic Acids Res. 2025;53:D444–56.
37. Tatusov RL. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 2001;29:22–8.
38. Haft DH. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res. 2001;29:41–3.
39. Mistry J, Bateman A, Finn RD. Predicting active site residue annotations in the Pfam database. BMC Bioinformatics. 2007;8:298.
40. 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.
41. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:W29.
42. Yin Y, et al. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012;40:W445–51.
43. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043.
44. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics. 2020;36:1925–7.
45. Goris J, et al. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007;57:81–91.
46. Palmer M, Steenkamp ET, Blom J, Hedlund BP, Venter SN. All ANIs are not created equal: implications for prokaryotic species boundaries and integration of ANIs into polyphasic taxonomy. Int J Syst Evol Microbiol. 2020;70:2937–48.
47. Olm MR, Brown CT, Brooks B, Banfield JF. DRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11:2864–8.
48. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35:1026–8.
49. Parks DH, et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022;50:D785–94.
50. Löytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol. 2014;1079:155–70.
51. Criscuolo A, Gribaldo S. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 2010;10:210.
52. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. Modelfinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
53. Minh BQ, et al. Iq-tree 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
54. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:W29-37.
55. Bulzu P-A, Kavagutti VS, Andrei A-S, Ghai R. The evolutionary kaleidoscope of rhodopsins. mSystems. 2022;7:e00405-e422.
56. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
57. Käll L, Krogh A, Sonnhammer ELL. An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics. 2005;21(Suppl 1):i251-257.
58. Ernst OP, et al. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. Chem Rev. 2014;114:126–63.

59. Mirarab S, et al. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. J Comput Biol. 2015;22:377–86.
60. Bushnell B, Rood J, Singer E. Bbmerge – accurate paired shotgun read merging via overlap. PLoS ONE. 2017;12:e0185056.
61. Kelley LA, Gardner SP, Sutcliffe MJ. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng Des Sel. 1996;9:1063–5.
62. Kille B, et al. Parsnp 2.0: scalable core-genome alignment for massive microbial datasets. Bioinformatics. 2024;40:btae311.
63. Drost H-G, Gabel A, Grosse I, Quint M. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. Mol Biol Evol. 2015;32:1221–31.
64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
65. Hedlund BP, et al. SeqCode: a nomenclatural code for prokaryotes described from sequence data. Nat Microbiol. 2022;7:1702–8.
66. Cho BC, Hardies SC, Jang GI, Hwang CY. Complete genome of streamlined marine actinobacterium *Pontimonas salivibrio* strain CL-TW6T adapted to coastal planktonic lifestyle. BMC Genomics. 2018;19:625.
67. Mehrshad M, et al. Hidden in plain sight—highly abundant and diverse planktonic freshwater chloroflexi. Microbiome. 2018;6:176.
68. Kavagutti VS, Andrei A-Ş, Mehrshad M, Salcher MM, Ghai R. Phage-centric ecological interactions in aquatic ecosystems revealed through ultra-deep metagenomics. Microbiome. 2019;7:135.
69. Kavagutti VS, et al. High-resolution metagenomic reconstruction of the freshwater spring bloom. Microbiome. 2023;11:15.
70. Krinos AI, et al. Time-series metagenomics reveals changing protistan ecology of a temperate dimictic lake. Microbiome. 2024;12:133.
71. Cronan JE, Thomas J. Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. Methods Enzymol. 2009;459:395–433.
72. Ghylin TW, et al. Comparative single-cell genomics reveals potential ecological niches for the freshwater acI actinobacteria lineage. ISME J. 2014;8:2503–16.
73. Escalante-Semerena JC. Conversion of cobinamide into adenosylcobamide in bacteria and archaea. J Bacteriol. 2007;189:4555–60.
74. Beale J, Lee SY, Iwata S, Beis K. Structure of the aliphatic sulfonate-binding protein SsuA from *Escherichia coli*. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2010;66:391–6.
75. Tanaka Y, et al. Crystal structure of a YeeE/YedE family protein engaged in thiosulfate uptake. Sci Adv. 2020;6:eaba7637.
76. White D. The physiology and biochemistry of prokaryotes. 3rd ed. New York: Oxford University Press; 2007.
77. Takeuchi N, Fullmer MS, Maddock DJ, Poole AM. The constructive black queen hypothesis: new functions can evolve under conditions favouring gene loss. ISME J. 2024;18:wrae011.
78. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. 1991;280(Pt 2):309–16.
79. Li X-L, Spániková S, de Vries RP, Biely P. Identification of genes encoding microbial glucuronoyl esterases. FEBS Lett. 2007;581:4029–35.
80. Molina-Henares, A. J., Krell, T., Eugenia Guazzaroni, M., Segura, A. & Ramos, J. L. Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. FEMS Microbiol Rev 30, 157–186 (2006).
81. Catara G, Caggiano R, Palazzo L. The DarT/DarG toxin-antitoxin ADP-ribosylation system as a novel target for a rational design of innovative antimicrobial strategies. Pathogens. 2023;12(2):240.
82. Cebrián-Sastre E, Martín-Blecua I, Gullón S, Blázquez J, Castañeda-García A. Control of genome stability by EndoMS/NucS-mediated non-canonical mismatch repair. Cells. 2021;10:1314.
83. Poindexter, J. S. Oligotrophy. in *Advances in microbial ecology* (ed. Alexander, M.) 63–89 (Springer US, Boston, MA, 1981). https://doi.org/10.1007/978-1-4615-8306-6_2.
84. Gómez-Consarnau L, et al. Microbial rhodopsins are major contributors to the solar energy captured in the sea. Sci Adv. 2019;5:eaaw8855.
85. Béjà O, et al. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science. 2000;289:1902–6.
86. Pinhassi J, DeLong EF, Béjà O, González JM, Pedrós-Alió C. Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. Microbiol Mol Biol Rev. 2016;80:929–54.
87. Dwulit-Smith JR, et al. A*cI* actinobacteria assemble a functional actinorhodopsin with natively synthesized retinal. Appl Environ Microbiol. 2018;84:e01678–18.
88. Nakajima Y, *et al.* Bacterium lacking a known gene for retinal biosynthesis constructs functional rhodopsins. Microbes Environ. 2020;35:ME20085.
89. Chazan A, et al. Diverse heliorhodopsins detected via functional metagenomics in freshwater actinobacteria, Chloroflexi and archaea. Environ Microbiol. 2022;24:110–21.
90. Cho S-G, et al. Heliorhodopsin binds and regulates glutamine synthetase activity. PLoS Biol. 2022;20:e3001817.
91. Bulzu, P.-A. *et al.* Heliorhodopsin evolution is driven by photosensory promiscuity in monoderms. mSphere. 2021;6;e0066121.
92. Burkert U, Warnecke F, Babenzien D, Zwirnmann E, Pernthaler J. Members of a readily enriched beta-proteobacterial clade are common in surface waters of a humic lake. Appl Environ Microbiol. 2003;69:6550–9.
93. Eckert EM, Salcher MM, Posch T, Eugster B, Pernthaler J. Rapid successions affect microbial N-acetyl-glucosamine uptake patterns during a lacustrine spring phytoplankton bloom. Environ Microbiol. 2012;14:794–806.
94. Buck U, Grossart H-P, Amann R, Pernthaler J. Substrate incorporation patterns of bacterioplankton populations in stratified and mixed waters of a humic lake. Environ Microbiol. 2009;11:1854–65.
95. Beier S, Bertilsson S. Uncoupling of chitinase activity and uptake of hydrolysis products in freshwater bacterioplankton. Limnol Oceanogr. 2011;56:1179–88.
96. Taipale S, Jones RI, Tiirola M. Vertical diversity of bacteria in an oxygen-stratified humic lake, evaluated using DNA and phospholipid analyses. Aquat Microb Ecol. 2009;55:1–16.
97. Hoetzinger, M. *et al.* Geographic population structure and distinct intra-population dynamics of globally abundant freshwater bacteria. ISME J. 2024;18;wrae113.
98. Okazaki Y, Nishikawa Y, Wagatsuma R, Takeyama H, Nakano S. Contrasting defense strategies of oligotrophs and copiotrophs revealed by single-cell-resolved virus–host pairing of freshwater bacteria. ISME Commun. 2025;5:ycaf086.
99. Eckert EM, Baumgartner M, Huber IM, Pernthaler J. Grazing resistant freshwater bacteria profit from chitin and cell-wall-derived organic carbon. Environ Microbiol. 2013;15:2019–30.
100. Pernthaler J, et al. Predator-specific enrichment of actinobacteria from a cosmopolitan freshwater clade in mixed continuous culture. Appl Environ Microbiol. 2001;67:2145–55.
101. Šimek K, et al. Differential freshwater flagellate community response to bacterial food quality with a focus on *Limnohabitans* bacteria. ISME J. 2013;7:1519–30.
102. Tarao M, Jezbera J, Hahn MW. Involvement of cell surface structures in size-independent grazing resistance of freshwater actinobacteria. Appl Environ Microbiol. 2009;75:4720–6.
103. Ferris JA, Lehman JT. Interannual variation in diatom bloom dynamics: roles of hydrology, nutrient limitation, sinking, and whole lake manipulation. Water Res. 2007;41:2551–62.
104. Kovářová J, Barrett MP. The pentose phosphate pathway in parasitic trypanosomatids. Trends Parasitol. 2016;32:622–34.
105. Lam HM, Winkler ME. Metabolic relationships between pyridoxine (vitamin B6) and serine biosynthesis in *Escherichia coli* K-12. J Bacteriol. 1990;172:6518–28.
106. Zhao J, Baba T, Mori H, Shimizu K. Global metabolic response of *Escherichia coli* to gnd or zwf gene-knockout, based on 13c-labeling experiments and the measurement of enzyme activities. Appl Microbiol Biotechnol. 2004;64:91–8.
107. Schwalbach MS, Tripp HJ, Steindler L, Smith DP, Giovannoni SJ. The presence of the glycolysis operon in SAR11 genomes is positively correlated with ocean productivity. Environ Microbiol. 2010;12:490–500.
108. Spaans SK, Weusthuis RA, Van Der Oost J, Kengen SW. NADPH-generating systems in bacteria and archaea. Front Microbiol. 2015;6:742.
109. Szebenyi DME, Musayev FN, di Salvo ML, Safo MK, Schirch V. Serine hydroxymethyltransferase: role of glu75 and evidence that serine is cleaved by a retroaldol mechanism. Biochemistry. 2004;43:6865–76.
110. Qin Z, Yan Q, Ma Q, Jiang Z. Crystal structure and characterization of a novel L-serine ammonia-lyase from *Rhizomucor miehei*. Biochem Biophys Res Commun. 2015;466:431–7.

111. Hu H, Mylon SE, Benoit G. Distribution of the thiols glutathione and 3-mercaptopropionic acid in Connecticut lakes. Limnol Oceanogr. 2006;51:2763–74.
112. Bachhawat AK, Thakur A, Kaur J, Zulkifli M. Glutathione transporters. Biochimica et Biophysica Acta (BBA) - General Subjects. 2013;1830:3154–64.
113. Suzuki H, Koyanagi T, Izuka S, Onishi A, Kumagai H. The yliA, -B, -C, and -D genes of *Escherichia coli* K-12 encode a novel glutathione importer with an ATP-binding cassette. J Bacteriol. 2005;187:5861–7.
114. Mondav, R. *et al.* Streamlined and abundant bacterioplankton thrive in functional cohorts. mSystems. 2020;5;e00316–20.
115. Kettler GC, et al. Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. PLoS Genet. 2007;3:e231.
116. Rodriguez-Valera F, Molina-Pardines C. On the biological meaning of the population pangenome. Trends Microbiol. 2025. https://doi.org/10.1016/j.tim.2025.07.004.
117. Hahn J, Inamine G, Kozlov Y, Dubnau D. Characterization of comE, a late competence operon of *Bacillus subtilis* required for the binding and uptake of transforming DNA. Mol Microbiol. 1993;10:99–111.
118. Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F. Key roles for freshwater actinobacteria revealed by deep metagenomic sequencing. Mol Ecol. 2014;23:6073–90.
119. Okazaki, Y., Nishikawa, Y., Wagatsuma, R., Takeyama, H. & Nakano, S. Contrasting defense strategies of oligotrophs and copiotrophs revealed by single-cell-resolved virus–host pairing of freshwater bacteria. 2024.07.24.604879 Preprint at https://doi.org/10.1101/2024.07.24.604879 (2024).
120. Kavagutti VS, Chiriac M-C, Ghai R, Salcher MM, Haber M. Isolation of phages infecting the abundant freshwater Actinobacteriota order 'Ca. Nanopelagicales.' ISME J. 2023;17:943–6.
121. Chen L-X, et al. Wide distribution of phage that infect freshwater SAR11 bacteria. mSystems. 2019;4:e00410–19.
122. Giovannoni S, Temperton B, Zhao Y, et al. Giovannoni et al. reply. Nature. 2013;499:E4–5.
123. Drula E, et al. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022;50:D571–7.
124. Tada Y, Grossart H-P. Community shifts of actively growing lake bacteria after N-acetyl-glucosamine addition: improving the BrdU-FACS method. ISME J. 2014;8:441–54.
125. Ceccaldi P, et al. Reductive activation of *E. coli* respiratory nitrate reductase. Biochim Biophys Acta. 2015;1847:1055–63.
126. Rodriguez-Valera F, et al. Explaining microbial population genomics through phage predation. Nat Prec. 2009. https://doi.org/10.1038/npre.2009.3489.1.

Publisher's Note