Phenology and ecological role of aerobic anoxygenic phototrophs in freshwaters

Villena-Alemany, Cristian ; Mujakić, Izabela ; Fecskeová, Livia K ; Woodhouse, Jason ; Auladell, Adrià ; Dean, Jason ; Hanusová, Martina ; Socha, Magdalena ; Gazulla, Carlota R ; Ruscheweyh, Hans-Joachim ; Sunagawa, Shinichi ; Silva Kavagutti, Vinicius ; Andrei, Adrian-Stefan ; Grossart, Hans-Peter ; Ghai, Rohit ; Koblížek, Michal ; Piwosz, Kasia

Phenology and ecological role of aerobic anoxygenic phototrophs in freshwaters

Cristian Villena-Alemany[1,2*], Izabela Mujakić[1], Livia K. Fecskeová[3], Jason Woodhouse[4], Adrià Auladell[5], Jason Dean[1], Martina Hanusová[1], Magdalena Socha[6], Carlota R. Gazulla[7], Hans-Joachim Ruscheweyh[8], Shinichi Sunagawa[8], Vinicius Silva Kavagutti[2,9,10], Adrian-Ştefan Andrei[11], Hans-Peter Grossart[12,4], Rohit Ghai[9], Michal Koblížek[1,2] and Kasia Piwosz[6*]

Abstract

Background Aerobic anoxygenic phototrophic (AAP) bacteria are heterotrophic bacteria that supply their metabolism with light energy harvested by bacteriochlorophyll-*a*-containing reaction centers. Despite their substantial contribution to bacterial biomass, microbial food webs, and carbon cycle, their phenology in freshwater lakes remains unknown. Hence, we investigated seasonal variations of AAP abundance and community composition biweekly across 3 years in a temperate, meso-oligotrophic freshwater lake.

Results AAP bacteria displayed a clear seasonal trend with a spring maximum following the bloom of phytoplankton and a secondary maximum in autumn. As the AAP bacteria represent a highly diverse assemblage of species, we followed their seasonal succession using the amplicon sequencing of the *puf*M marker gene. To enhance the accuracy of the taxonomic assignment, we developed new *puf*M primers that generate longer amplicons and compiled the currently largest database of *puf*M genes, comprising 3633 reference sequences spanning all phyla known to contain AAP species. With this novel resource, we demonstrated that the majority of the species appeared during specific phases of the seasonal cycle, with less than 2% of AAP species detected during the whole year. AAP community presented an indigenous freshwater nature characterized by high resilience and heterogenic adaptations to varying conditions of the freshwater environment.

Conclusions Our findings highlight the substantial contribution of AAP bacteria to the carbon flow and ecological dynamics of lakes and unveil a recurrent and dynamic seasonal succession of the AAP community. By integrating this information with the indicator of primary production (Chlorophyll-*a*) and existing ecological models, we show that AAP bacteria play a pivotal role in the recycling of dissolved organic matter released during spring phytoplankton bloom. We suggest a potential role of AAP bacteria within the context of the PEG model and their consideration in further ecological models.

Keywords Freshwaters, Aquatic microbial ecology, Microbial seasonal succession, Long-term sampling, Aerobic anoxygenic phototrophs, *puf*M gene, PEG model, Photoheterotrophs

*Correspondence:
Cristian Villena-Alemany
cristian.villena.alemany@hotmail.com
Kasia Piwosz
kpiwosz@mir.gdynia.pl
Full list of author information is available at the end of the article

Villena-Alemany *et al. Microbiome* (2024) 12:65

Page 2 of 14

Introduction

Recurrent seasonal changes of aquatic microbial communities are among the best-studied phenomena in freshwater lakes and reservoirs. The Plankton Ecology Group (PEG) model initially described the dynamic interactions between phytoplankton and zooplankton [1] and was later amended with the eutrophic and oligotrophic scenarios and role description of heterotrophic protists [2]. Subsequently, the importance of bacterioplankton was revealed, especially during the spring phytoplankton bloom [3–5], increasing our understanding of the contribution of microorganisms to the functioning of limnic ecosystems [6]. Bacteria represent an important part of aquatic microbial communities. They generate fresh particulate organic matter by utilizing dissolved organic carbon (DOC) and render it accessible to organisms at upper trophic levels [7]. However, the role of photoheterotrophic bacteria, which present a significant part of bacterial biomass and activity [8], remains overlooked.

Aerobic anoxygenic phototrophic (AAP) bacteria are a functional group of photoheterotrophs that rely upon external sources of organic carbon and supplement their metabolism with energy obtained from light through bacteriochlorophyll-*a* (BChl-*a*) type II reaction centers. The capacity to harvest light energy enables AAP bacteria to reduce their respiration and increase biomass yield [9, 10]. Moreover, the AAP community shows higher growth rates, larger cell sizes, and greater activity than heterotrophic bacteria [11–16]. Photoheterotrophy by AAP bacteria increases carbon transfer efficiency, enlarging the availability of biomass for upper trophic levels and reducing CO_2 emitted to the atmosphere [17]. However, little is known about the phenology of the AAP community and the absence of exhaustive seasonal sampling hampers the understanding of their role in lakes. AAP bacteria peak during spring in lakes, when they may account for up to 22% of bacteria [18, 19]. Their abundances and diversity dynamics correlate with irradiance, temperature, chlorophyll-*a* (Chl-*a*), oxygen, and DOC [18, 20–24].

One of the obstacles in the study of AAP bacteria is the fact that they do not represent a monophyletic group. On the contrary, phototrophic genes have been gained and lost multiple times in closely related species [25, 26]. Therefore, AAP species cannot be identified based on the most common marker used in community studies, the 16S rRNA gene. Instead, the *puf*M gene, which encodes the subunit M of the anoxygenic type-II reaction center, has been widely employed to study AAP communities [15, 27–32]. However, these studies were unsuccessful in providing a taxonomic assignment for abundant *puf*M phylotypes. This is caused by the low taxonomic resolution of the short amplicon sequences and the lack

of a curated reference database. The increased availability of metagenome-assembled and single-cell amplified genomes (MAGs and SAGs) has expanded our knowledge of metabolic potential within multiple bacterial lineages and should allow us to establish a comprehensive *puf*M database for amplicon assignment.

To improve the taxonomic assignment, we designed a novel primer set targeting a larger 450 bp region of the *puf*M gene and compiled an extensive database of 3633 non-redundant *puf*M gene sequences from existing genome and metagenome sequence datasets. We applied this novel metabarcoding assay to 215 samples from 3 years, collected from meso-oligotrophic freshwater Cep lake (Czechia) at biweekly intervals from multiple depths. We hypothesized that the AAP community would show a recurrent seasonal succession. Due to their specific metabolism, this succession would exhibit different abundance patterns than overall heterotrophic bacteria. Specifically, we expected that AAP abundance would peak during the spring phytoplankton bloom. Additionally, since they are a functional, taxonomically diverse group, we surmise that the spring AAP peak is orchestrated by specific phylotypes, rather than the involvement of the entire AAP community.

Materials and methods

Sampling and measuring environmental variables

Samples were collected biweekly from April 2017 to December 2019 from the freshwater Cep lake (48°92′49.24″N, 14°88′68.11″E). This meso-oligotrophic lake is located in the Třeboň Basin Protected Landscape Area, Czechia, and has an area of 130 ha and a maximum depth of 12 m. Five liters of water were collected from 0.5, 2, 5, and 8 m using a 3-L Ruttner water sampler (KC Denmark A/S, Denmark) and transported to the laboratory in closed plastic containers in a cooler box, which were pre-rinsed three times with the sampled water. Temperature and oxygen profiles were taken with an EXO1 multi-parameter probe (YSI Inc., Yellow Springs, USA). Total and AAP bacterial abundances were counted in Zeiss Axio Imager.D2 epifluorescence microscope equipped with Collibri LED module illumination system (Carl Zeiss, Jena, Germany). Microscopic samples were excited at 325–370 nm, 450–490 nm, and 545–565 nm and the AAP cells were detected by their BChl-*a* autofluorescence at wavelengths > 850 nm in epifluorescence. Since Chl-a autofluorescence spectra emission also encompasses the 890 nm wavelength, cells that show fluorescence emission at 690 and 573 nm, corresponding with Chl-*a* and phycoerythrine, respectively, were not counted as AAP bacteria [17, 33, 34]. Concentrations of Chl-*a* and BChl-*a* were determined in organic solvent extracts by reversed-phase high-performance liquid

Villena-Alemany *et al. Microbiome*   (2024) 12:65

Page 3 of 14

chromatography [31]. The quantification of environmental nutrients was performed as described in Procházková, 1959 (nitrate); Murphy and Riley, 1962 (phosphate); Kopáček and Hejzlar, 1993 (total phosphorous); Kopáček and Procházková, 1993 (ammonia) and Shabarova et al. 2021 (DOC) [35–39].

*puf*M gene database

We collected 14,872 *puf*M nucleotide and protein sequences from representative genomes and MAGs available from Genome Taxonomy Database (GTDB) r207 [40], Tara Ocean [41], the LIMNOS database compiled from a set of ~1300 freshwater lake metagenomes (PRJEB47226), and from MAG collections publicly available in the NCBI [3, 42–46]. Bacterial genomes and MAGs taxonomy were determined using GTDB-Tk v2.1.1 [47]. *puf*M sequences in GTDB r207 were found using a hidden Markov model of *puf*M gene (K08929) from the KOFAM database [48] with a score threshold of 394.57, as described in https://github.com/adriaaula/obtain_gene_GTDB. *puf*M-containing non-redundant MAGs from Tara Oceans (https://doi.org/10.6084/m9.figshare.4902923.v1.) were selected using HMMER v3.3.2 (http://hmmer.org/) with a customized *puf*M database from pfam [49]. The *puf*M sequences were confirmed by Diamond v0.9.24 annotation [50]. Nucleotide *puf*M gene sequences from the LIMNOS database were obtained from open reading frames using Prodigal [51] and annotated using a custom pipeline incorporating Diamond v0.9.24 [50] and the KEGG database [52]. *Puf*M genes were compiled alongside the taxonomy of their associated MAG.

All *puf*M sequences were pooled and duplicated sequences were removed. Protein sequences were aligned with MAFFT v7.453 (–maxiterate 1000 –localpair) [53] and a maximum likelihood tree was calculated using iqtree2 [54] with automatic model selection performed by ModelFinder [55], and 1000 iterations of ultrafast bootstrapping with 1000 rounds of SH-aLRT testing (-alrt 1000 -B 1000) [56]. *puf*L sequences were identified as they formed a long branch. Bona-fide *puf*M sequences were retained (Supplementary file S1), and alignment and phylogenetic trees were redone and visualized using iTOL [57]. The environmental origin of each sequence was obtained manually from source databases (Supplementary file S2).

AAP community analysis by *puf*M gene amplicon sequencing

Between 300 and 1460 ml of water was filtered through sterile 0.2 μm Nucleopore Track-Etch Membrane filters (Whatman®, Maidstone, United Kingdom) that were immediately placed inside sterile cryogenic vials

(Biologix Group Limited, Jinan, Shandong China) containing 0.55 g of sterile zirconium beads, flash-frozen in liquid nitrogen and stored at −80 °C until DNA extraction (max. 6 months). Total nucleic acids were chemically extracted according to Griffiths et al. 2000 [58] with modifications [59], re-suspended in 35 μl of DNase and RNase-free water (MP Biomedicals, Solon, OH, USA), and stored at − 20 °C. Concentration and quality of the extracts were checked using NanoDrop (Thermo Fisher Scientific).

To improve the accuracy of the taxonomic assignation and reduce the number of unclassified amplicon sequence variants (ASVs), a new primer pair for *puf*M gene was designed. pufM_uniF primer [27] was used as a reverse (pufM_UniFRC in the current study, 5´-RAANGGR TTRTARWANARRTTNCC-3') and pufM_longF was designed ~450 bp upstream (5'-YGGSCCGWTCTAYST-SGG-3') using a pre-existing database of 1500 sequences [31]. The specificity and coverage of the new primer pair were tested in comparison to the commonly used *puf*M primers [27, 32] against the new *puf*M database. The analysis was done in Geneious Prime (v2023.0.1) with up to three mismatches in the binding region and in both forward and reverse directions. The primers' specificity was also tested separately for Pseudomonadota (formerly known as Proteobacteria), Alphaproteobacteria, Gammaproteobacteria, Gemmatimonadota, Chloroflexota, Myxococcota, and Eremiobacterota based on alignments done in Geneious Prime by MUSCLE alignment (v5.1.).

The PCR conditions were optimized using genomic DNA from *Gemmatimonas phototrophica* (Gemmatimonadota), *Sphingomonas glacialis* (Alphaproteobacteria) and *Congregibacter litoralis* (Gammaproteobacteria), and environmental DNA from the current sampling. The final conditions were as follows: initial denaturation for 3 min at 98 °C, 35 cycles of 98 °C for 15 s, 52 °C for 30 s, 72 °C for 18 s, and final elongation at 72 °C for 5 min. Triplicate PCR reactions (20 μL) using Phusion™ High-Fidelity PCR MasterMix (Thermo Fisher Scientific, USA) were pooled and the amplicons of ~450 bp were purified from 1.5% agarose (MP Roche, Germany) gel using the Wizzard SV Gel and PCR clean system (Promega, USA) and quantified with Qubit dsDNA HS assay (Thermo Fisher Scientific, USA). Samples were randomly distributed within two runs to account for the batch effect and sequenced on Illumina Miseq 2×300 bp PE (Macrogen, South Korea).

Raw reads were quality-checked using FastQC v0.11.7 (Babraham Bioinformatics, Cambridge, UK). The primer sequences were trimmed and read quality filtered using Cutadapt v1.16 maximum error (-e 0.1), quality cut-off (-q 20), and minimum length (-m 250) [60]. Initial number of reads (average ± standard deviation; 81,788 ± 14,590)

Villena-Alemany *et al. Microbiome*   (2024) 12:65

Page 4 of 14

were truncated using *filterAndTrim* (truncLen=c(220, 220), maxEE=c(2,5), truncQ=2) in the R/Bioconductor environment from DADA2 package v1.12.1 [61]. ASVs were constructed and chimeric sequences were removed using the method "consensus" [62]. ASVs present only in one of the runs were removed from downstream analysis using *intersect* and *subset*. Subsequently, ASVs were aligned in Geneious Prime v2019.2.3 using ClustalW v2.1 [63]. Poorly aligned ASVs were confirmed to not be *puf*M with a blast against NCBI non-redundant database [64] and excluded from further analysis. The final dataset consisted of 1588 ASVs (Supplementary file S3, Reference ASV sheet) and 62,729±13,448 reads per sample (76.7% of the total number of initial reads, Supplementary file S3, ASV_table sheet). The sequences were deposited in the NCBI database under Biosamples SAMN38037304-SAMN38037518 as a part of BioProject PRJNA970655.

The taxonomic assignment was done through phylogenetic placement using The Evolutionary Placement Algorithm v0.3.5 [65] that placed the ASVs into the phylogenetic tree calculated from the new reference database sequences that were back-translated from protein alignments using trimAl [66]. The taxonomic assignation was handled according to the ASV phylogenetic position using Gappa [67] (Supplementary file S3, Taxonomy sheet).

Phytoplankton community analysis based on 16S rRNA gene amplicons

The V3-V4 region of the bacterial 16S rRNA gene was amplified using 341F and 785R primer pair [68] as described in Piwosz et al. 2022 [17] The subset of sequences assigned to Chloroplast was extracted and their taxonomy was further affiliated using a curated reference database of the plastidial 16S rRNA gene: PhytoRef [69]. Bar plots were visualized using ggplot v3.4.3 [70].

Data and statistical analysis

Unless stated otherwise, all analyses were done in R v3.6.1 and were visualized using ggplot2 v3.3.6 [70]. Dynamics of environmental and biological variables were interpolated using igraph v1.2.6 and lubridate v1.8.0 [71, 72]. For addressing the compositional bias of amplicon data [73], principal component analysis was done using centered log ratio (CLR) transformation [74] through *transform* from microbiome package v1.17.42. Community composition bar plots and Alphaproteobacteria, Gammaproteobacteria, and Gemmatimonadota bubble plots were done using Phyloseq v1.30.0 [75]. The 100 most abundant ASVs were selected and plotted using *plot_heatmap* [75]. The occurrence of specific ASVs in spring was tested

using analysis of compositions of microbiomes with bias correction in ANCOMBC v2.3.2 [76, 77] and plotted using ggplot2 v3.4.3 [70].

Relationships between environmental data and AAP community were analyzed using distance-based linear models (DistML) [78, 79] in the PERMANOVA+add-on package of the PRIMER7 software [80] (Primer Ltd., Lutton, UK). From strongly correlated environmental variables (correlation coefficient>0.6) only one was selected for further analysis. The model was calculated on the CLR-transformed relative abundance data of AAP bacteria [74], using a stepwise selection procedure. The best model was selected based on statistical significance (9999 permutations) and the value of Akaike's Information Criterion (AICc).

AAP core and network community analyses

ASVs present in more than 80% of the samples from the 3 years and four depths, were considered the AAP lake core microbiome. The percent contribution of each core ASV to their respective maximum percent contribution was calculated and plotted with bubble plots using Phyloseq v1.30.0 [75] and ggplot2 v3.3.6 [70].

SparCC analysis was applied to calculate lake community co-occurrence correlations from the compositional data [81]. Only correlations with pseudo *p* value<0.02 and stronger correlation than ±0.7 were selected. The network was plotted using Cytoscape v3.9.1 [82].

Time series and trend lines

Interannual trend analysis was done in TTR package v0.24.3 (https://github.com/joshuaulrich/TTR) in R v 4.3.0. Raw data on total and AAP bacterial abundances, temperature, and Chl-*a* concentrations were averaged for months and depths and transformed into time series assuming annual frequency. They were decomposed into trend, seasonal, and random components using *decompose* with default settings. Spearman correlation between interannual trends of AAPs abundance and temperature and Chl-*a* concentrations was done for the extracted trend component.

Results

New database and longer amplicons enhance taxonomic assignments of *puf*M gene ASVs

In order to improve the taxonomic assignment of the *puf*M gene amplicons, we constructed a new reference database containing 3633 *puf*M sequences (>646 bp) from Pseudomonadota (synonym for Proteobacteria), Gemmatimonadota, Chloroflexota, Eremiobacteriota, and Myxococcota (Supplementary file S1). The database includes 529 genera, 114 families, 53 orders, and 9

classes (Supplementary Figure S1) from cultured species and MAGs originating from a wide variety of habitats (Supplementary file S2), mostly from freshwater (2140 sequences) and marine environments (381 sequences).

The newly designed primer set covers 80.5% of the new database at maximum of three mismatches (Supplementary file S4). The older, highly degenerated primer pairs, UniF+UniR, pufMF+pufMR and UniF+pufM_WAW, cover 98.9%, 85% and 96.2%, respectively. However, the amplicon length of the novel primers is about two times longer (~450 bp) allowing for the proper taxonomic assignation of more than 95% of the alphaproteobacterial and above 75% of the gammaproteobacterial reads at the order level. Additionally, 38 alphaproteobacterial, 36 gammaproteobacterial, and 6 Gemmatimonadota genera were detected (Supplementary Figures S2, S3, and S4).

Seasonal changes in Cep lake

Environmental conditions in Cep Lake showed seasonal dynamics typical for a temperate freshwater lake (Supplementary file S5). In January and February, the lake was partially frozen and stratified from April/May until September with maximum temperatures around 24 °C in July and August. The metalimnion was located between 5 and 8 m depth in 2017, and between 2 and 5 m in 2018 and 2019. In all 3 years, the autumnal mixing, characterized by higher values of dissolved oxygen and lower

temperatures, was initiated in October (Supplementary Figure S5A–B).

Chlorophyll-*a* measurements varied throughout the year with seasonal maxima representing spring and autumn phytoplankton blooms (Supplementary Figure S5C). The spring phytoplankton bloom terminated at the onset of stratification, and was composed, according to 16S amplicons affiliated to plastids, mostly by Bacillariophyta and Chrysophyceae (Supplementary Figure S6).

Seasonal dynamics of AAP community composition

The maximal AAP abundances ($3.42–5.50 \times 10^5$ cells mL^{-1}) corresponded to 15–20% of the total bacteria (Supplementary Figure S5D–E) closely following the spring phytoplankton blooms. A second AAP bacterial peak occurred towards the end of summer, before the autumn phytoplankton peak. Alpha diversity of the AAP community was lower during the spring abundance peaks and rose during the second part of the year (Supplementary Figure S5F).

The AAP community followed an annual recurrent pattern during the three consecutive years, with a distinction between the epilimnion and hypolimnion communities during the stratified period (Fig. 1A). Samples from the same season in different years were more similar to each other than samples from different seasons in the same year, indicating the persistent temporal succession of the community and similar interannual community
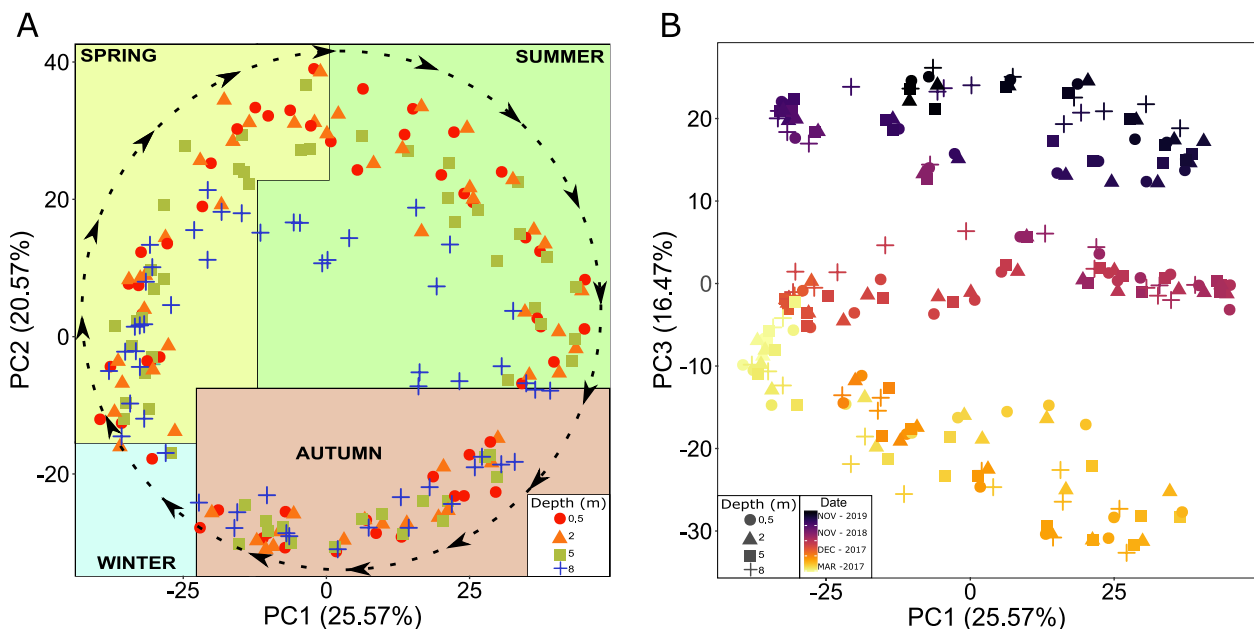
Fig. 1 Development of AAP community structure. Principal component analysis of centered log-ratio transformed AAP community composition. Each point represents a sample with 0.5 m (red circle), 2 m (orange triangle), 5 m (green square) and 8 m (blue cross) in PC1 and PC2 axis (**A**) and 0.5 m (circle), 2 m (triangle), 5 m (square) and 8 m (cross) coloured according to the date of sampling in PC1 and PC3 axis (**B**). Dashed line and arrows in panel A indicate AAP community succession following an annual chronological direction

structure. Distance-based linear models (DistLM) and distance-based redundancy analysis (dbRDA) selected temperature, Chl-*a*, and total, Cyanobacterial, and AAP abundances, to best explain the variability (23.11%) of the AAP community composition (Supplementary file S6). For 2018 and 2019, 2 years for which nutrient data is available (Supplementary file S5), phosphorus and ammonia increased this explanation by 6.75%, up to 29.86%.

Interestingly, in addition to a seasonal cycle, we observed an interannual variation in AAP community composition (Fig. 1B). The decomposition of time series on monthly averaged values for the whole water column showed increasing interannual trends in temperature and AAP abundance during 3 years of sampling, and a decreasing trend in Chl-*a* concentration (Supplementary Figure S7). Trends of temperature and AAP abundance were significantly correlated (Spearman correlation coefficient rho = 0.8, *p* value < 0.0001).

The AAP community was dominated by Gammaproteobacteria over the whole water column, with an average relative contribution exceeding 50% and reaching up to 90% during stratification (Supplementary Figure S8). Alphaproteobacteria was the second most abundant class, showing the maxima contributions in spring and autumn, reaching over 50% in the spring of 2018. Classes Gemmatimonadetes (Gemmatimonadota) and Myxococcia (Myxococcota) made up 5% and 2% of the AAP community, respectively. Unclassified Pseudomonadota and unclassified Myxococcota showed transient contributions of < 9% and < 1% of the AAP community, respectively. Chloroflexota and Eremiobacteriota were not detected.

The 100 most abundant ASVs (based on their average relative abundances) comprised 75% of the reads and exhibited seasonal recurrence, peaking every year at specific times of the year (Fig. 2). The majority of ASVs demonstrated a transient contribution and were generally absent outside their maxima (e.g., ASV28 (Rhizobiales)),
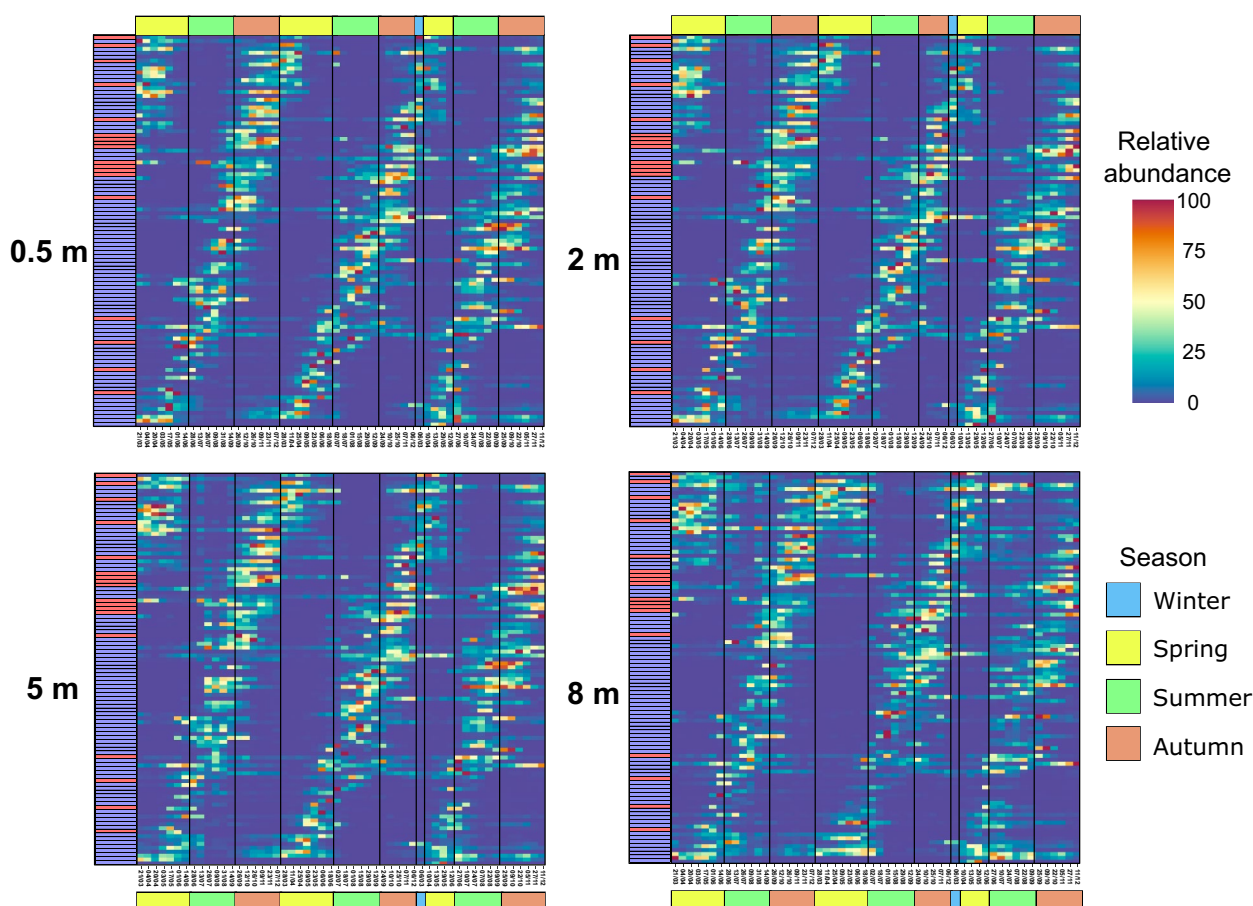

Fig. 2 Recurrence of 100 most abundant ASVs. Relative abundance is individually normalized in 0.5 m, 2 m, 5 m, and 8 m during the different seasons of the 3-year sampling. Coloured box on the left indicate the taxonomic assignation at class level of each ASV, red for Alphaproteobacteria and blue for Gammaproteobacteria. Colour bar on the top and bottom indicate the seasons of the year

Villena-Alemany *et al. Microbiome*   (2024) 12:65

Page 7 of 14

while a few showed pronounced relative abundance throughout the year (e.g., ASV4 (*Rhodoferax*)). Interestingly, ASV67 (*Limnohabitans*), whose relative abundance was on average < 0.36%, was the sole ASV detected in every sample.

Out of 1588 *puf*M ASVs (Supplementary file S3, Reference ASV sheet), the stable part of the AAP community (defined here as ASVs present in > 80% of the samples) consisted of only 22 ASVs (Supplementary Figure S9): 8 *Rhodoferax*, 4 *Limnohabitans*, 2 *Aestuariivirga*, 1 *Methylobacterium*, 1 *Rubrivivax*, and 6 other Burkholderiales. These core ASVs varied largely in their contribution from the most abundant ASV2 (*Aestuariivirga*, with an average relative abundance of 4%) to the least abundant ASV237 (unclassified Burkholderiales, with an average relative abundance of 0.06%). Their seasonal dynamics differed substantially throughout the year and distinct relative abundance patterns were observed even for ASVs from the same genus. For instance, ASV5 and ASV49, both *Rhodoferax*, peaked in autumn and spring, respectively. Similar differences were observed for two *Aestuariivirga*: ASV2 peaked during the spring mixing period (from March to May), while ASV62 showed its highest contribution during the summer stratification. It is noticeable that the core AAP community also included ASVs outside the 100 most abundant, such as ASV115 (*Rhodoferax*) and ASV237 (unclassified Burkholderiales), which had a low but steady contribution during the whole sampling season. Furthermore, the dynamics of some core ASVs, such as ASV31, were different each year.

To identify phylotypes most contributing to the difference in AAP community composition during spring and autumn peaks, we carried out an analysis of compositions of microbiomes with bias correction (Supplementary file S7). Composition of genera and orders contributing to the AAP bacterial peaks was different (Fig. 3). Spring peak consisted of a higher prevalence of Alphaproteobacteria versus Gammaproteobacteria genera (9 vs 5), while during autumn, the community was more diverse and included also Gemmatimonadota and Myxococcota. The highest genera contributors to the spring peak (log fold change > 2) were *RFPW01*, *UBA1936*, *Cypionkella*, and *Rhodoferax*, while *UBA964* and *UBA5518* (Gammaproteobacteria: Steroidobacterales and Pseudomonadales, respectively) contributed most to the late summer peak. The only genera substantially contributing to both peaks was *Aestuariivirga*.

Network analysis

To study possible interactions between AAP bacteria, we performed a network co-occurrence analysis. The network calculated for the entire lake concluded that 99 ASVs presented 139 significant interactions (92% positive; Fig. 4). Thirteen highly connected nodes with more than six co-occurrence correlations were identified as hubs. They belonged to the genus UBA964, Steroidobacterales, *Rubrivivax*, and unclassified Burkholderiaceae from Gammaproteobacteria. The most connected nodes from Alphaproteobacteria belonged to *Aestuariivirga* and Rhodobacteraceae (4 edges each). The majority of correlations (~ 60%) occurred between ASVs from the same genus, family, or order, creating groups of densely connected nodes (e.g. genus UBA964). Furthermore, some ASVs peaking in spring correlated positively with each other (*Aestuariviirga* and unclassified Burkholderiaceae) in contrast to *Rhodoferax* and *Cypionkella*. A similar pattern was observed for ASVs peaking in summer-autumn.

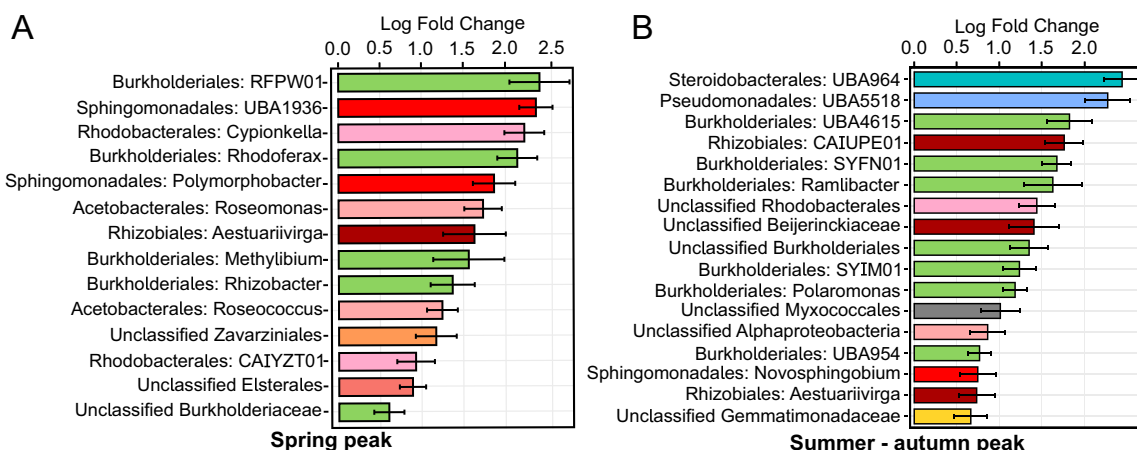**Fig. 3** Community composition of AAP abundance peaks. Analysis of compositions of microbiomes showing log fold change values at order and genus level in Spring (**A**) and in Summer-Autumn peaks (**B**) with green colour scale for Burkholderiales, blue colour scale for other Gammaproteobacteria, red colour scale for Alphaproteobacteria, yellow for Gemmatimonadota and grey for Myxococcota orders
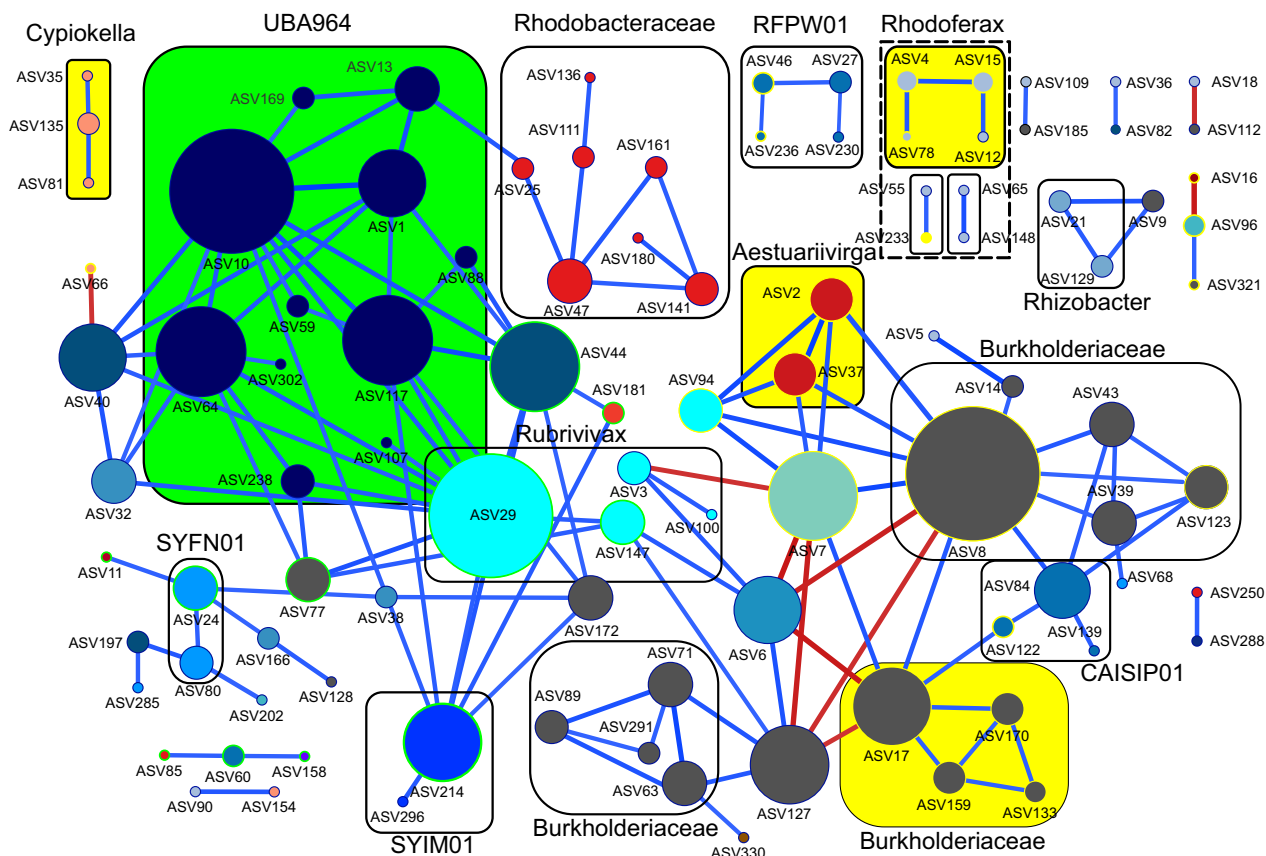
Fig. 4 AAP community co-occurrence correlation network. The radius of the nodes (ASVs) is directly proportional to the number of significant correlations. Colour of the nodes shows the taxonomy at the class level with Alphaproteobacteria in red–orange colour scale, Gammaproteobacteria blue colour scale, unclassified Burkholderiaceae (grey), and Gemmatimonadota (brown). Blue lines indicate positive and red lines have negative correlations. Rectangles with solid lines indicate nodes that interact with other nodes from the same taxonomic rank (genus or family) and positive log fold change for Spring AAP peak are highlighted in yellow rectangles fill and yellow node stroke while green rectangles fill and green node stroke for autumn AAP peak

Discussion

Seasonal succession of planktonic communities in temperate lakes has been intensively studied [2]. The PEG model defines the key phases in the annual development of ecological succession as well as the interactions between different organisms in aquatic ecosystems. The annually recurrent phenomena in freshwater lakes include a spring phytoplankton bloom followed by a zooplankton-induced clear-water phase in early summer, a late-summer phytoplankton bloom, and a period of low productivity in winter. Recently, seasonal dynamics of heterotrophic bacteria have been incorporated into the model [5], as they rapidly respond to the transitions in the lakes' pelagic functioning, especially during the phytoplankton bloom [3, 5, 83]. Whilst AAP bacteria have been shown to substantially contribute to the bacterial abundance, biomass, and activity in

freshwater lakes [13, 17], they are not considered in the PEG model [1, 2].

Dynamics of microbial communities are typically investigated by sequencing of the 16S rRNA gene and such analyses do not allow for disentangling the metabolic functionality of bacteria, especially when traits of interest follow a heterogenic pattern of presence within the same taxonomic ranks. This is the case for AAP bacteria, where members from the same genus might or might not contain the ability to carry out anoxygenic photosynthesis [26]. Amplicon analysis of a functional gene may overcome this hindrance, but it requires a comprehensive database with taxonomically assigned reference sequences. Currently, some phylogenetic clades (A-L) cannot be assigned to the genus or even order level [84, 85]. Moreover, different researchers assemble databases for their environment of interest, with different quality thresholds and criteria [17, 86], which hampers direct comparison between different studies. Finally, short amplicons obtained with the

most commonly used primer combinations, puf_UniF-puf_UniR, puf_UniF−pufM_WAW, and pufMF−pufM_WAW [27, 32] often do not allow for taxonomic assignment below the class or order level, resulting in a substantial number of unclassified reads [18, 29, 86, 87].

Comprehensive database and longer amplicons allow for improved taxonomic assignments of *puf*M ASVs

We constructed the largest curated *puf*M database to date that includes sequences from essentially all environments, with a special high representation of freshwater lakes (Supplementary Figure S1, Supplementary file S1 and S2). Phylogenetic trees based on *puf*M gene do not match 16S rRNA phylogeny due to horizontal gene transfer events [88–90]. In contrast, the taxonomic assignment of genomes and MAGs based on the whole genome or 120 selected marker genes is more accurate and consistent [91]. Thus, we included only sequences originating from taxonomically assigned genomes and MAGs, excluding all environmental *puf*M sequences classified into phylogroups based on phylogenetic analysis [84]. During our quality control, sequences originating from Bdellovibrionota, Verrucomicrobiota, Omnitrophota, Planctomycetota, or Bacteroidota were excluded from the final database. Some members of these phyla were reported to encode *puf*M [87]. However, our manual inspection revealed that their *puf*M gene was present in short contigs and often found as the only phototrophic gene, which does not warrant phototrophic functionality. Furthermore, *puf*M genes from these phyla did not form a monophyletic clade, suggesting dubious multiple and independent events of horizontal gene transfer. Thus, only members of Pseudomonadota, Chloroflexota, Gemmatimonadota, Eremiobacteriota, and Myxococcota were included. Our choice of sequences ensures high quality and allows for future extensions of the database as more metagenomic data is produced and more AAP bacteria are cultured, enhancing its fidelity and functionality. Moreover, as our database contains entire or almost entire *puf*M gene sequences, it can be used for amplicon taxonomy assignment independently of the actual primers used.

The *puf*M gene is one of the most conserved genes from the *puf* operon which codifies the genes for the synthesis of the anoxygenic photosynthesis apparatus [92, 93]. Commonly used puf_UniF-puf_UniR, puf_UniF−pufM_WAW, and pufMF−pufM_WAW primer pairs [27, 32] hybridize on the most conserved regions at the end of the gene, separated by ∼110–160 bp. They show high coverages (Supplementary file S4) but produce short amplicons that hamper taxonomic assignation below the class or order level resulting in a high fraction of unclassified reads [18, 29, 86, 87]. Thus, we designed a new primer set producing longer amplicons to increase the taxonomic resolution, as has been shown for other genes [94]. The new primer pair has lower in silico coverage against our new database than puf_UniF—puf_UniR [27] (80.5 vs 98.9%). Some groups, such as Chloroflexota, *Aquidulcibacter*, and *Polynucleobacter*, were poorly covered, which may explain their absence in our amplicons. Nevertheless, the number of sequences identified at every taxonomic level increased compared to a previous study in the same lake [18]: 95% of the alphaproteobacterial and above 75% of gammaproteobacterial reads were classified at the order level. Additionally, the number of newly detected genera was substantially higher (80 vs 12) and the Shannon index showed a wider range of diversity (Supplementary Figure S5F) since longer amplicons enable us to detect more nucleotide variations, and thus revealing higher diversity, advancing our knowledge on AAP community composition.

Phenology of AAP bacteria and consideration of the PEG model

The seasonal succession of the AAP community revealed differential strategies of adaptation to environmental conditions, unveiling generalist AAP bacteria appearing most of the time, whereas specialists or opportunists showed a transient contribution to the AAP community (Fig. 2). Within the generalists, we identified the core AAP community that consistently contributed throughout the seasons for three consecutive years and across all depths (Supplementary Figure S9). The coexistence of the core AAP community, the dominance of positive correlations in the networks, and the numerous correlations between ASVs of similar taxonomic ranks (Fig. 4) suggest partial metabolic redundancy within some closely related AAP bacteria that maintain functional kinship. In contrast, the complex seasonal succession pattern indicates that the lake's AAP community is extremely diverse, with over 100 reported genera of AAP bacteria (Supplementary Figure S1, S2, and S3). AAP communities represent a large functional repertoire (even within the same genus) allowing for niche speciation via temporal succession, facilitating their geographical coexistence. Finally, the 3-year recurrence of the AAP community (Fig. 1A) documents its indigenous character in this freshwater lake and the higher importance of selection over the environmental drift and dispersal processes at a short temporal scale [95]. Changes in the AAP abundance coincided with shifts in their community composition indicating that abundance peaks were caused by specific phylotypes. These phylotypes differed between both abundance peaks (Fig. 3). Generally, the AAP community was dominated by Gammaproteobacteria except

for spring abundance peaks, when Alphaproteobacteria, which already have shown higher phototrophic activities in spring [31], increased their contribution. Additionally, the directional interannual variation of the AAP community (Fig. 1B) signifies the evolution of AAP populations, potentially influenced by changes in environmental and biological variables such as temperature or Chl-*a* (Supplementary Figure S7).

While the PEG model has enhanced our understanding of seasonal patterns, it still does not encompass all aquatic components, such as viruses or specific functional bacterial groups. This includes AAP bacteria, which are characterized by a heterogenic behavior but still represent an important functional group, fulfilling valuable ecological and biochemical processes in the aquatic environment. For that reason, amending them into the present PEG model will certainly improve our understanding of aquatic community functioning.

Cep Lake is representative of a meso-oligotrophic temperate lake in the northern hemisphere [96], thus our conclusions might be applied to other similar lakes. AAP bacteria played an important role during and shortly after the spring bloom when their abundance and contribution to the total bacterial community were recurrently

the highest. This spring AAP abundance peak preceded that of the overall heterotrophic bacteria (Fig. 5). These results are consistent with the previous study in the same lake [18]. The faster response of AAP bacteria highlights that photoheterotrophy confers a distinct metabolic importance as food for bacterivores in microbial food webs is well documented [97–99] and due to their, on average, larger cell size and higher activity than other heterotrophic bacteria [13, 14], they might contribute disproportionally to the carbon cycling despite their relatively low abundances [19, 99]. Additionally, Chl-*a* concentration has been identified as a variable explaining the dynamics of the AAP bacterial community and it is plausible to assume that the spring abundance peaks of AAP bacteria are triggered by the excess of carbon released by the phytoplankton bloom (mostly diatoms; Supplementary Figure S6) and the lack of grazing pressure after winter. Moreover, AAPs have a highly efficient photoheterotrophic metabolism [17] increasing secondary bacterial production and disposing of more carbon to higher trophic levels via the microbial loop. This emphasizes the urgent need for more quantitative studies to further decipher carbon transfers along microbial and classical food webs. The AAP bacterial peak is terminated by selective
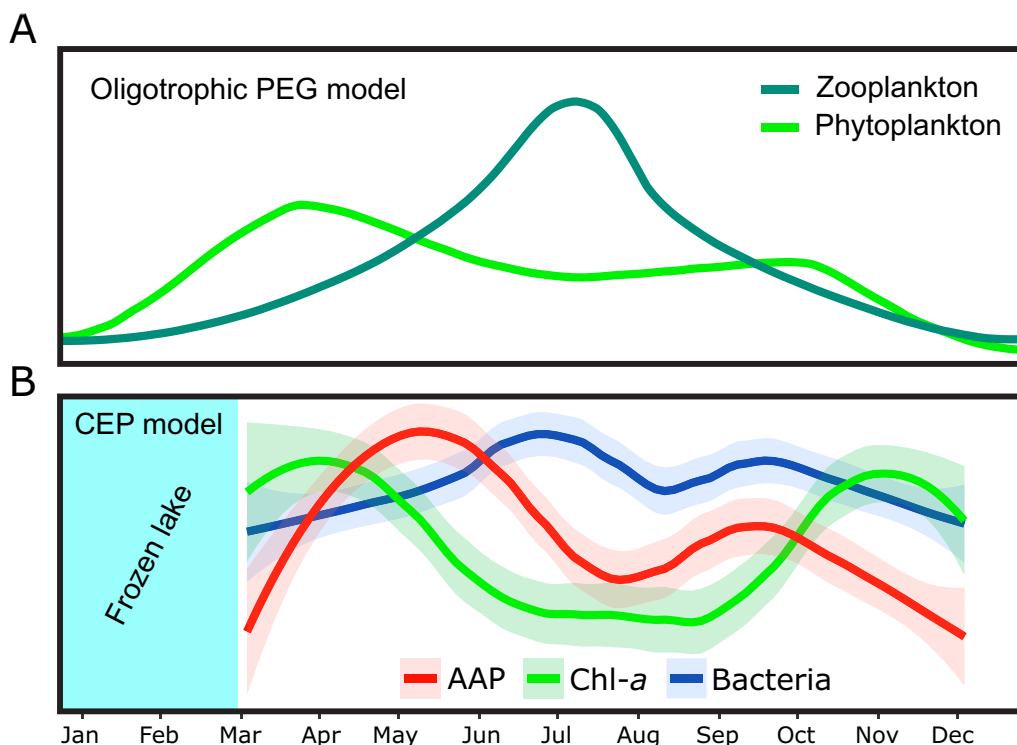

Fig. 5 Proposed inclusion of AAP bacteria in the PEG model. Annual succession patterns of microbial communities for (**A**) Phytoplankton and zooplankton according to original PEG model in oligotrophic scenario (Sommer et al., 1986), and (**B**) monthly averaged annual succession pattern of AAP abundance in CEP lake, phytoplankton dynamics through Chl-*a*, and of bacterial abundance. Trend lines are normalized to maxima and minima values for each variable, light transparent areas indicate 95% confidence interval

Villena-Alemany *et al. Microbiome*   (2024) 12:65

Page 11 of 14

and extensive grazing of bacterivorous protists and meso-zooplankton that are also present in summer [83, 98, 100]. The absence of a pronounced AAPs peak following the phytoplankton bloom in autumn (Fig. 5) could be attributed to the higher grazing pressure, distinct phytoplankton composition (Supplementary Figure S6), decreasing temperatures (Supplementary file S6), and/or the decreasing light availability at shorter day length [2, 18, 30, 101]. Altogether, we propose for the first time the inclusion of the AAP bacteria into the PEG model, encouraging other microbial ecologist to account for their role in other lakes of different trophic status.

Finally, our study provides novel insight into the ecology of phototrophic Myxococcota. While their average contribution was low, they were detected during stratification over three consecutive years (Supplementary Figure S8) and constituted a member of the summer-autumn AAPs peak, emphasizing their potential significance in microbial communities during summer as they showed a potentially predatory and photoheterotrophic metabolism [102].

Conclusions

Our study revealed annual recurrent seasonal patterns of AAP bacteria in a freshwater lake, supporting the potential inclusion of this important functional group into the PEG model. The high abundance of AAP bacteria during the spring phytoplankton bloom highlights their crucial role in recycling phytoplankton-derived dissolved organic matter and their role in aquatic food webs, which needs to be further quantified and better understood. Differential contribution patterns of the core community and temporal succession of the AAP community indicate strong competition within AAP bacteria communities, which forces them to conduct temporal niche partitioning in order to geographically coexist. In contrast, positive co-occurrence correlations between closely related AAP bacteria indicated their functional redundancy. Our findings provide unprecedented insights into the phenology of AAP bacteria in a temperate freshwater lake, blazing a trail for future studies to verify the proposed role in other types of lakes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-01786-0.

Additional file 1: Supplementary Figure S1. Maximum likelihood phylogenetic tree of *puf*M gene sequences of the constructed database. Outer ring represents the environment of origin and the colour of the clades between branches and the outer ring shows the taxonomic classification of the sequences at class level. Colour of the branches refers to the ultra-fast bootstrap values.

Additional file 2: Supplementary Figure S2. Alphaproteobacteria community composition at order and genus level for 3-year sampling at 0.5 (A), 2 (B), 5 (C) and 8 m depth (D). Larger size and brighter colours are directly proportional to the relative contribution of each genus to the total Alphaproteobacteria community.

Additional file 3: Supplementary Figure S3. Gammaproteobacteria community composition at order and genus level for 3-year sampling at 0,5 (A). 2 (B), 5 (C) and 8 m depth (D). Larger size and brighter colours are directly proportional to the relative contribution of each genus to the total Gammaproteobacteria community.

Additional file 4: Supplementary Figure S4. Gemmatimonadota community composition at order and genus level for 3-year sampling at 0.5 (A), 2 (B), 5 (C) and 8 m depth (D). Larger size and brighter colours are directly proportional to the relative contribution of each genus to the total Gemmatimonadota community.

Additional file 5: Supplementary Figure S5. Environmental and biological variables for 8 m' depth profile during 3-year sampling in CEP lake. Temperature (A), AAP abundance (B), dissolved oxygen (C), percentage contribution to total bacterial community (D), Shannon alpha diversity values (E), and Chlorophyll-*a* (F). Light-blue vertical bands represent lack of sampling due to frozen lake surface.

Additional file 6: Supplementary Figure S6. Phytoplankton chloroplast-based community composition at class level for 0.5, 2, 5 and 8 m' depth during 3-years temporal series.

Additional file 7:Supplementary Figure S7. Decomposition of additive time series for bacterial abundance (a), AAP abundance (b), chlorophyll-*a* concentration (c) and temperature (d). Analysis was done in the TTR package version 0.24.3 (R v4.3.0). Spearman correlation of the decomposed trends between AAP abundance and chlorophyll-*a* (e) and AAP abundance and temperature (f). R: spearman's rho value, p: p-value.

Additional file 8: Supplementary Figure S8. AAP bacteria community composition according to *puf*M gene taxonomic assignment at class level for 0.5 (A), 2 (B), 5 (C) and 8 meters' depth (D) during 3-years sampling campaign.

Additional file 9: Supplementary Figure S9. Individually normalized relative abundance of the 22 core AAP ASVs during 3 years in 4 depths. Brighter colours and bigger dots indicate larger contribution to the AAP bacterial community. ASVs are clustered according to taxonomic classification at the maximum possible level (genus, family or order).

Additional file 10: Supplementary file S1. Nucleotide *puf*M gene sequence of the database in fasta format. **Supplementary file S2.** ID of the *puf*M gene sequences from the database and their environment of origin. **Supplementary file S3.** File containing all the information from the amplicon sequence analysis. Reference ASVs (IDs and sequences), ASV table (ID and abundance on each sample) and Taxa (ID and taxonomic assignation of each ASV). **Supplementary file S4.** Primer coverage comparison of the most commonly used *puf*M gene primer pairs and the newly designed one with 0, 1, 2 and 3 mismatches (MM). Numbers represent the percentage of sequences from the *puf*M gene database covered for different phyla and classes. **Supplementary file S5.** Sample identification number (Sample name) and all the environmental and biological variables measured. **Supplementary file S6.** Draftsman plot correlation of the environmental and biological variables, samples removed due to lack of environmental variables, marginal and sequential test for DistLM for 3 years and 2 years (includes also nutrients). **Supplementary file S7.** Positive log fold change (lfc) values, standard error (SE) at genus and ASV levels for the spring and autumn AAP abundance peaks with the p- and q-values (p, q).

Villena-Alemany *et al. Microbiome*   (2024) 12:65

Page 12 of 14

and technical support for the LIMNOS sequencing data set and all related environmental data.

Authors' contributions
CVA, KP and MK did the conceptualization; CVA, IM, KP, JW, AA, CRG, MS, HJR, SS, VK, ASA, HPG and RG curated or provided data for *puf*M database; formal analysis was carried out by CVA, IM, KP, LKF and RG; investigation and experimental processes were done by CVA, IM and KP; methodology of database was developed by RG and CVA; CVA and KP wrote original draft and all the authors helped during the review and editing of the manuscript; JD and MH performed the sampling and measured of environmental variables. All authors have revised and approved the submitted version.

Availability of data and materials
Sequence data that support the findings of this study have been deposited in NCBI database with the primary accession names: Biosamples SAMN38037304—SAMN38037518 as a part of BioProject PRJNA970655.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Laboratory of Anoxygenic Phototrophs, Institute of Microbiology of the Czech Academy of Sciences, Třeboň, Czechia. [2]Department of Ecosystem Biology, Faculty of Science, University of South Bohemia, České Budějovice, Czechia. [3]Associated Tissue Bank, Faculty of Medicine, Pavol Jozef Safarik University in Košice, Košice, Slovakia. [4]Department of Plankton and Microbial Ecology, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Stechlin, Germany. [5]Functional Genomics and Evolution Department, Institute of Evolutionary Biology, CSIC-UPF, Barcelona, Catalonia, Spain. [6]Department of Fisheries Oceanography and Marine Ecology, National Marine Fisheries Research Institute, Gdynia, Poland. [7]Centro Oceanográfico de Málaga, Instituto Español de Oceanografía, IEO-CSIC, 29640 Fuengirola, Málaga, Spain. [8]Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zurich, Zurich, Switzerland. [9]Department of Aquatic Microbial Ecology, Institute of Hydrobiology, Biology Center CAS, České Budějovice, Czech Republic. [10]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France. [11]Limnological Station, Department of Plant and Microbial Biology, University of Zurich, Kilchberg, Switzerland. [12]Institute of Biology and Biochemistry, Potsdam University, Potsdam, Germany.

References
1. Sommer U, Gliwicz ZM, Lampert W, Duncan A. The PEG-model of seasonal succession of planktonic events in fresh waters. Arch für Hydrobiol. 1986;106:433–71.
2. Sommer U, Adrian R, De Senerpont DL, Elser JJ, Gaedke U, Ibelings B, et al. Beyond the plankton ecology group (PEG) model: mechanisms driving plankton succession. Annu Rev Ecol Evol Syst. 2012;43:429–48.
3. Kavagutti VS, Bulzu P-A, Chiriac CM, Salcher MM, Mukherjee I, Shabarova T, et al. High-resolution metagenomic reconstruction of the freshwater spring bloom. Microbiome. 2023;11:15.
4. Zeder M, Peter S, Shabarova T, Pernthaler J. A small population of planktonic Flavobacteria with disproportionally high growth during the spring phytoplankton bloom in a prealpine lake. Environ Microbiol. 2009;11:2676–86.
5. Park H, Shabarova T, Salcher MM, Kosová L, Rychtecký P, Mukherjee I, et al. In the right place, at the right time: the integration of bacteria into the Plankton Ecology Group model. Microbiome. 2023;11:112.
6. Reavie ED, Barbiero RP, Allinger LE, Warren GJ. Phytoplankton trends in the Great Lakes, 2001–2011. J Great Lakes Res. 2014;40:618–39.
7. Pomeroy L, leB. Williams P, Azam F, Hobbie J. The Microbial Loop. Oceanography 2007; 20: 28–33.
8. Koblížek M. Ecology of aerobic anoxygenic phototrophs in aquatic environments. FEMS Microbiol Rev. 2015;39:854–70.
9. Piwosz K, Kaftan D, Dean J, Šetlík J, Koblížek M. Nonlinear effect of irradiance on photoheterotrophic activity and growth of the aerobic anoxygenic phototrophic bacterium Dinoroseobacter shibae. Environ Microbiol. 2018;20:724–33.
10. Koblížek M, Dachev M, Bína D, Nupur, Piwosz K, Kaftan D. Utilization of light energy in phototrophic Gemmatimonadetes. J Photochem Photobiol B Biol. 2020;213:112085.
11. Mašín M, Nedoma J, Pechar L, Koblížek M. Distribution of aerobic anoxygenic phototrophs in temperate freshwater systems. Environ Microbiol. 2008;10:1988–96.
12. Čuperová Z, Holzer E, Salka I, Sommaruga R, Koblížek M. Temporal changes and altitudinal distribution of aerobic anoxygenic phototrophs in mountain lakes. Appl Environ Microbiol. 2013;79:6439–46.
13. Fauteux L, Cottrell MT, Kirchman DL, Borrego CM, Garcia-Chaves MC, del Giorgio PA. Patterns in abundance, cell size and pigment content of aerobic anoxygenic phototrophic bacteria along environmental gradients in northern lakes. PLoS One. 2015;10:1–17.
14. Garcia-Chaves MC, Cottrell MT, Kirchman DL, Ruiz-González C, Del Giorgio PA. Single-cell activity of freshwater aerobic anoxygenic phototrophic bacteria and their contribution to biomass production. ISME J. 2016;10:1579–88.
15. Ferrera I, Sarmento H, Priscu JC, Chiuchiolo A, González JM, Grossart H-P. Diversity and distribution of freshwater aerobic anoxygenic phototrophic bacteria across a wide latitudinal gradient. Front Microbiol. 2017;8:175.
16. Ruiz-González C, Garcia-Chaves MC, Ferrera I, Niño-García JP, Giorgio PA. Taxonomic differences shape the responses of freshwater aerobic anoxygenic phototrophic bacterial communities to light and predation. Mol Ecol. 2020;29:1267–83.
17. Piwosz K, Villena-Alemany C, Mujakić I. Photoheterotrophy by aerobic anoxygenic bacteria modulates carbon fluxes in a freshwater lake. ISME J. 2022;16:1046–54.
18. Villena-Alemany C, Mujakić I, Porcal P, Koblížek M, Piwosz K. Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake. Environ Microbiol Rep. 2023;15:60–71.
19. Kolářová E, Medová H, Piwosz K, Koblížek M. Seasonal dynamics of aerobic anoxygenic phototrophs in freshwater lake Vlkov. Folia Microbiol (Praha). 2019;64:705–10.
20. Kuzyk SB, Ma X, Yurkov V. Seasonal dynamics of Lake Winnipeg's microbial communities reveal aerobic anoxygenic phototrophic populations coincide with sunlight availability. Microorganisms. 2022;10:1690.
21. Ferrera I, Sánchez O, Kolářová E, Koblížek M, Gasol JM. Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. ISME J. 2017;11:2391–3.
22. Tomaš AV, Šantić D, Šolić M, Ordulj M, Jozić S, Šestanović S, et al. Dynamics of Aerobic Anoxygenic Phototrophs along the trophic gradient in the central Adriatic Sea. Deep Sea Res Part II Top Stud Oceanogr. 2019;164:112–21.
23. Szabó-Tugyi N, Vörös L, V.-Balogh K, Botta-Dukát Z, Bernát G, Schmera D, et al. Aerobic anoxygenic phototrophs are highly abundant in hypertrophic and polyhumic waters. FEMS Microbiol Ecol. 2019;95:fiz104.
24. Shi L, Cai Y, Shi X, Zhang M, Zeng Q, Kong F, et al. Community structure of aerobic anoxygenic phototrophic bacteria in algae- and macrophyte-dominated areas in Taihu Lake. China J Oceanol Limnol. 2022;40:1855–67.

25. Kopejtka K, Lin Y, Jakubovičová M, Koblížek M, Tomasch J. Clustered core- and pan-genome content on rhodobacteraceae chromosomes. Genome Biol Evol. 2019;11:2208–17.

26. Kasalický V, Zeng Y, Piwosz K, Šimek K, Kratochvilová H, Koblížek M. Aerobic anoxygenic photosynthesis is commonly present within the genus Limnohabitans. Appl Environ Microbiol. 2018;84:6–17.

27. Yutin N, Suzuki MT, Béjà O. Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. Appl Environ Microbiol. 2005;71:8958–62.

28. Salka I, Čuperová Z, Mašín M, Koblížek M, Grossart H-P. Rhodoferax-related pufM gene cluster dominates the aerobic anoxygenic photo-trophic communities in German freshwater lakes. Environ Microbiol. 2011;13:2865–75.

29. Tang K, Jia L, Yuan B, Yang S, Li H, Meng J, et al. Aerobic anoxygenic phototrophic bacteria promote the development of biological soil crusts. Front Microbiol. 2018;9:2715.

30. Auladell A, Sánchez P, Sánchez O, Gasol JM, Ferrera I. Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. ISME J. 2019;13:1975–87.

31. Fecskeová LK, Piwosz K, Hanusová M, Nedoma J, Znachor P, Koblížek M. Diel changes and diversity of pufM expression in freshwater communities of anoxygenic phototrophic bacteria. Sci Rep. 2019;9:18766.

32. Béjà O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, Hamada T, et al. Unsuspected diversity among marine aerobic anoxygenic phototrophs. Nature. 2002;415:630–3.

33. Cottrell MT, Mannino A, Kirchman DL. Aerobic anoxygenic phototrophic bacteria in the mid-atlantic bight and the north pacific gyre. Appl Environ Microbiol. 2006;72:557–64.

34. Mašín M, Zdun A, Ston-Egiert J, Nausch M, Labrenz M, Moulisová V, et al. Seasonal changes and diversity of aerobic anoxygenic phototrophs in the Baltic Sea. Aquat Microb Ecol. 2006;45:247–54.

35. Shabarova T, Salcher MM, Porcal P, Znachor P, Nedoma J, Grossart H-P, et al. Recovery of freshwater microbial communities after extreme rain events is mediated by cyclic succession. Nat Microbiol. 2021;6:479–88.

36. Murphy J, Riley JP. A modified single solution method for the determination of phosphate in natural waters. Anal Chim Acta. 1962;27:31–6.

37. Kopáček J, Hejzlar J. Semi-micro determination of total phosphorus in fresh waters with perchloric acid digestion. Int J Environ Anal Chem. 1993;53:173–83.

38. Procházková L. Bestimmung der Nitrate im Wasser. Fresenius' Zeitschrift für Anal Chemie. 1959;167:254–60.

39. Kopáček J, Procházková L. Semi-micro determination of ammonia in water by the rubazoic acid method. Int J Environ Anal Chem. 1993;53:243–8.

40. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 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.

41. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat Microbiol. 2018;3:804–13.

42. Kavagutti VS, Andrei AŞ, Mehrshad M, Salcher MM, Ghai R. Phage-centric ecological interactions in aquatic ecosystems revealed through ultra-deep metagenomics. Microbiome. 2019;7:135.

43. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021;39:499–509.

44. Chiriac MC, Bulzu PA, Andrei AS, Okazaki Y, Nakano SI, Haber M, et al. Ecogenomics sheds light on diverse lifestyle strategies in freshwater CPR. Microbiome. 2022;10:84.

45. Buck M, Garcia SL, Fernandez L, Martin G, Martinez-Rodriguez GA, Saarenheimo J, et al. Comprehensive dataset of shotgun metagenomes from oxygen stratified freshwater lakes and ponds. Sci Data. 2021;8:131.

46. Moncadas LS, Shabarova T, Kavagutti VS, Bulzu P, Chiriac M, Park S, Mukherjee I, Ghai R, Andrei A. Rickettsiales' deep evolutionary history sheds light on the emergence of intracellular lifestyles. bioRxiv. 2023. https://doi.org/10.1101/2023.01.31.526412.

47. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022;38:5315–6.

48. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2.

49. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47:427–32.

50. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18:366–8.

51. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.

52. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30.

53. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

54. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.

55. 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.

56. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22.

57. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.

58. Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. Appl Environ Microbiol. 2000;66:5488–91.

59. Nercessian O, Noyes E, Kalyuzhnaya MG, Lidstrom ME, Chistoserdova L. Bacterial Populations Active in Metabolism of C 1 Compounds in the Sediment of Lake Washington, a Freshwater Lake. Appl Environ Microbiol. 2005;71:6885–99.

60. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10.

61. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.

62. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11:2639–43.

63. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947–8.

64. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022;50:D20–6.

65. Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, et al. EPA-ng: massively parallel evolutionary placement of genetic sequences. Syst Biol. 2019;68:365–9.

66. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.

67. Czech L, Barbera P, Stamatakis A. Genesis and Gappa: Processing, analyzing and visualizing phylogenetic (placement) data. Bioinformatics. 2020;36:3263–5.

68. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013;41:e1–e1.

69. Decelle J, Romac S, Stern RF, Bendif EM, Zingone A, Audic S, et al. PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. Mol Ecol Resour. 2015;15:1435–45.

70. Cedric Ginestet, ggplot2: Elegant Graphics for Data Analysis. Journal of the Royal Statistical Society Series A: Statistics in Society. 2011;174(1):245–246. https://doi.org/10.1111/j.1467-985X.2010.00676_9.x.

Villena-Alemany *et al. Microbiome*   (2024) 12:65

Page 14 of 14

71. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal Complex Syst. 2006;1695:1–9.
72. Grolemund G, Wickham H. Dates and times made easy with lubridate. J Stat Softw. 2011;40:1–25.
73. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol. 2017;8:2224.
74. Quinn TP, Erb I, Gloor G, Notredame C, Richardson MF, Crowley TM. A field guide for the compositional analysis of any-omics data. Gigascience. 2019;8:giz107.
75. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8:e61217.
76. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. Nat Commun. 2020;11:3514.
77. Costea PI, Zeller G, Sunagawa S, Bork P. A fair comparison. Nat Methods. 2014;11:359–359.
78. Anderson MJ, Legendre P. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. J Stat Comput Simul. 1999;62:271–303.
79. Legendre P, Andersson MJ. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol Monogr. 1999;69:512–512.
80. Anderson MJ, Gorley RN, Clarke KR. PERMANOVA+ for PRIMER: guide to software and statistical methods. Primer-E, Plymouth, UK. 2008;1–214.
81. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8:e1002687.
82. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
83. Šimek K, Nedoma J, Znachor P, Kasalický V, Jezbera J, Hornňák K, et al. A finely tuned symphony of factors modulates the microbial food web of a freshwater reservoir in spring. Limnol Oceanogr. 2014;59:1477–92.
84. Yutin N, Suzuki MT, Teeling H, Weber M, Venter JC, Rusch DB, et al. Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. Environ Microbiol. 2007;9:1464–75.
85. Lehours A-C, Enault F, Boeuf D, Jeanthon C. Biogeographic patterns of aerobic anoxygenic phototrophic bacteria reveal an ecological consistency of phylogenetic clades in different oceanic biomes. Sci Rep. 2018;8:4105.
86. Gazulla CR, Cabello AM, Sánchez P, Gasol JM, Sánchez O, Ferrera I. A metagenomic and amplicon sequencing combined approach reveals the best primers to study marine aerobic anoxygenic phototrophs. Microb Ecol. 2023;86:2161–72.
87. Galachyants AD, Krasnopeev AY, Podlesnaya GV, Potapov SA, Sukhanova EV, Tikhonova IV, et al. Diversity of aerobic anoxygenic phototrophs and rhodopsin-containing bacteria in the surface microlayer, water column and epilithic biofilms of Lake Baikal. Microorganisms. 2021;9:842.
88. Kopejtka K, Tomasch J, Zeng Y, Tichý M, Sorokin DY, Koblížek M. Genomic Analysis of the Evolution of Phototrophy among Haloalkaliphilic Rhodobacterales. Genome Biol Evol. 2017;9:1950–62.
89. Zeng Y, Feng F, Medová H, Dean J, Koblížek M. Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. Proc Natl Acad Sci. 2014;111:7795–800.
90. Ward LM, Hemp J, Shih PM, McGlynn SE, Fischer WW. Evolution of phototrophy in the Chloroflexi Phylum driven by horizontal gene transfer. Front Microbiol. 2018;9:260.
91. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019;36:1925–7.
92. Nagashima S, Nagashima KVP. Comparison of Photosynthesis Gene Clusters Retrieved from Total Genome Sequences of Purple Bacteria. In Advances in Botanical Research, Volume 66. Amsterdam: Elsevier; 2013. p. 151–178. ISBN 9780123979230.
93. Imhoff JF, Rahn T, Künzel S, Neulinger SC. Photosynthesis is widely distributed among proteobacteria as demonstrated by the phylogeny of PufLM reaction center proteins. Front Microbiol. 2018;8:2679.
94. Jeong J, Yun K, Mun S, Chung W-H, Choi S-Y, Nam Y, et al. The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology. Sci Rep. 2021;11:1727.
95. Huber P, Metz S, Unrein F, Mayora G, Sarmento H, Devercelli M. Environmental heterogeneity determines the ecological processes that govern bacterial metacommunity assembly in a floodplain river system. ISME J. 2020;14:2951–66.
96. Verpoorter C, Kutser T, Seekell DA, Tranvik LJ. A global inventory of lakes based on high-resolution satellite imagery. Geophys Res Lett. 2014;41:6396–402.
97. Koblížek M, Stoń-Egiert J, Sagan S, Kolber ZS. Diel changes in bacteriochlorophyll a concentration suggest rapid bacterioplankton cycling in the Baltic Sea. FEMS Microbiol Ecol. 2005;51:353–61.
98. Garcia-Chaves M, Cottrell M, Kirchman D, Derry A, Bogard M, del Giorgio P. Major contribution of both zooplankton and protists to the top-down regulation of freshwater aerobic anoxygenic phototrophic bacteria. Aquat Microb Ecol. 2015;76:71–83.
99. Cepáková Z, Hrouzek P, Žišková E, Nuyanzina-Boldareva E, Šorf M, Kozlíková-Zapomělová E, et al. High turnover rates of aerobic anoxygenic phototrophs in European freshwater lakes. Environ Microbiol. 2016;18:5063–71.
100. Fecskeová LK, Piwosz K, Šantić D, Šestanović S, Tomaš AV, Hanusová M, et al. Lineage-specific growth curves document large differences in response of individual groups of marine bacteria to the top-down and bottom-up controls. MSystems. 2021;6:e00934-21.
101. Ferrera I, Borrego CM, Salazar G, Gasol JM. Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of pufM gene. Environ Microbiol. 2014;16:2953–65.
102. Li L, Huang D, Hu Y, Rudling NM, Canniffe DP, Wang F, et al. Globally distributed Myxococcota with photosynthesis gene clusters illuminate the origin and evolution of a potentially chimeric lifestyle. Nat Commun. 2023;14:6450.

Publisher's Note