

RESEARCH

Open Access



High-resolution metagenomic reconstruction of the freshwater spring bloom

Vinicius S. Kavagutti^{1,2*}, Paul-Adrian Bulzu¹, Cecilia M. Chiriac¹, Michaela M. Salcher¹, Indranil Mukherjee¹, Tanja Shabarova¹, Vesna Grujčić^{1,3}, Maliheh Mehrshad^{1,4}, Vojtěch Kasalický¹, Adrian-Stefan Andrei⁵, Jitka Jezberová¹, Jaromir Sedá¹, Pavel Rychtecký¹, Petr Znachor^{1,2}, Karel Šimek¹ and Rohit Ghai^{1*}

Abstract

Background The phytoplankton spring bloom in freshwater habitats is a complex, recurring, and dynamic ecological spectacle that unfolds at multiple biological scales. Although enormous taxonomic shifts in microbial assemblages during and after the bloom have been reported, genomic information on the microbial community of the spring bloom remains scarce.

Results We performed a high-resolution spatio-temporal sampling of the spring bloom in a freshwater reservoir and describe a multitude of previously unknown taxa using metagenome-assembled genomes of eukaryotes, prokaryotes, and viruses in combination with a broad array of methodologies. The recovered genomes reveal multiple distributional dynamics for several bacterial groups with progressively increasing stratification. Analyses of abundances of metagenome-assembled genomes in concert with CARD-FISH revealed remarkably similar in situ doubling time estimates for dominant genome-streamlined microbial lineages. Discordance between quantitations of cryptophytes arising from sequence data and microscopic identification suggested the presence of hidden, yet extremely abundant aplastidic cryptophytes that were confirmed by CARD-FISH analyses. Aplastidic cryptophytes are prevalent throughout the water column but have never been considered in prior models of plankton dynamics. We also recovered the first metagenomic-assembled genomes of freshwater protists (a diatom and a haptophyte) along with thousands of giant viral genomic contigs, some of which appeared similar to viruses infecting haptophytes but owing to lack of known representatives, most remained without any indication of their hosts. The contrasting distribution of giant viruses that are present in the entire water column to that of parasitic perkinsids residing largely in deeper waters allows us to propose giant viruses as the biological agents of top-down control and bloom collapse, likely in combination with bottom-up factors like a nutrient limitation.

Conclusion We reconstructed thousands of genomes of microbes and viruses from a freshwater spring bloom and show that such large-scale genome recovery allows tracking of planktonic succession in great detail. However, integration of metagenomic information with other methodologies (e.g., microscopy, CARD-FISH) remains critical to reveal diverse phenomena (e.g., distributional patterns, in situ doubling times) and novel participants (e.g., aplastidic cryptophytes) and to further refine existing ecological models (e.g., factors affecting bloom collapse). This work provides a genomic foundation for future approaches towards a fine-scale characterization of the organisms in relation

*Correspondence:

Vinicius S. Kavagutti
viniciuskavagutti@gmail.com

Rohit Ghai
ghai.rohit@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

to the rapidly changing environment during the course of the freshwater spring bloom.

Introduction

The spring bloom in freshwaters is a fascinating and dynamic phenomenon that has captivated the attention of microbiologists for decades [1–3]. Typically, all water bodies in temperate regions experience low temperatures in winter and react to the onset of spring when a combined effect of multiple physicochemical parameters initiates a cascade of events leading to a drastic change in the resident microbial community [1, 2]. A classic spring bloom scenario unfolds by a mixing of the water column, after which increasing light, temperature, and nutrients allow for the development of a phytoplankton bloom (often dominated by cryptophytes and diatoms), characterized by a peak of chlorophyll-*a*. [1, 4–6]. Multiple factors like enhanced grazing by protists [7], mortality by viruses/parasites [8, 9], nutrient exhaustion [10], and increasing zooplankton predation [1] lead to the collapse of the bloom, giving way to the clear water phase when phytoplankton decline dramatically. These changes in plankton assemblages have been encapsulated into generalized models that incorporate multiple factors affecting the initiation, expansion, and collapse of the spring bloom and the development of the clear water phase, e.g., the Plankton Ecology Group (PEG) model [1, 2].

It has been argued that capturing such events with a fast turnover of the community in a matter of a few weeks necessitates a high-temporal sampling approach, i.e., sampling every 2/3 days to identify short-lived peaks of rapidly growing microbes (generation times in hours to days) as seasonal (e.g., spring, summer, winter) or monthly samplings are insufficient [6, 11]. While long-term time series exploration of freshwater habitats is becoming increasingly common, and approaches ranging from amplicon analyses [12–14], fluorescence in situ hybridization and catalyzed reporter deposition (CARD-FISH) [15–17], and cultivation [18] to metagenomic sequencing [19, 20] have been applied, studies with high-temporal sampling, particularly in freshwaters are limited [4–6, 21, 22]. Even so, many approaches have been applied to disentangle this complex phenomenon. Spring blooms in freshwaters have been investigated by using CARD-FISH probes to detect short-lived peaks in multiple bacterial groups to couple their abundances to dynamics of phytoplankton and zooplankton [6], to follow specific bacterial groups (e.g., Flavobacteria) [4, 5, 18], cryptophytes [23], or ciliates [24, 25]. Other studies have addressed the uptake of specific metabolites (Chitin, NAG) [5, 26]. Similar algal-microbial blooms were also investigated in marine habitats, which has brought new important insights into factors affecting succession,

e.g., availability of diverse carbohydrates or phosphorus [27, 28] and promoted cultivation efforts targeting fast-responding heterotrophs like Bacteroidota, Alphaproteobacteria, Gammaproteobacteria, and Verrucomicrobiota [29, 30].

However, there are methodological limitations of many approaches that have been previously applied that preclude the discovery of novel participants of the spring bloom across the entire community. For instance, CARD-FISH analyses require the design of probes that target specific lineages, which may not be possible for many taxonomic groups of protists [31]. Similarly, amplicon-based studies provide no information on viruses, and culture-based approaches are biased towards relatively easily cultivable copiotrophs (e.g., Flavobacteria), overlooking many dominant, genome-streamlined oligotrophic microbes that remain hard to culture [32, 33]. Finally, microscopic analyses alone are insufficient to reliably distinguish many heterotrophic nanoflagellates [34]. In this regard, a metagenomic approach towards the de novo recovery of genomes of novel or understudied groups has been singularly lacking. Such an approach can potentially reveal genomes of prokaryotes, eukaryotes, and viruses, providing a much higher resolution of the entire community. This work presents a high-resolution temporal dissection of the annual spring bloom in a freshwater reservoir over 37 days of intensive sampling (every 2–3 days, a total of 57 samples), combining classical microbiological methods, CARD-FISH analyses, and metagenomic sequencing of three different size fractions. This allowed the recovery of metagenome-assembled genomes of prokaryotes, eukaryotes, and their viruses that constitute the overall microbial community. Moreover, sampling the hypolimnion that is typically ignored provides a more complete perspective of the entire water column. The application of a high-frequency sampling coupled with a metagenomic approach allowed us to add significant detail to the overall tapestry of planktonic succession in this remarkable event.

Results and discussion

Initial characterization of the spring bloom

To capture the dynamic events of the spring bloom, a high-frequency sampling approach was implemented, sampling the epilimnion every 2–3 days (from Apr 03 to May 09, 2018) and the hypolimnion every week (see the “Methods” section for details). The chief features are summarized in Fig. 1 (epilimnion) and Additional file 1: Figure S1 (hypolimnion). The present study time frame can be divided into two distinct phases based on the water mixing patterns, chlorophyll-*a* dynamics, and the composition of the planktonic community. The first stage is the mixis (days 1–4), when the entire water

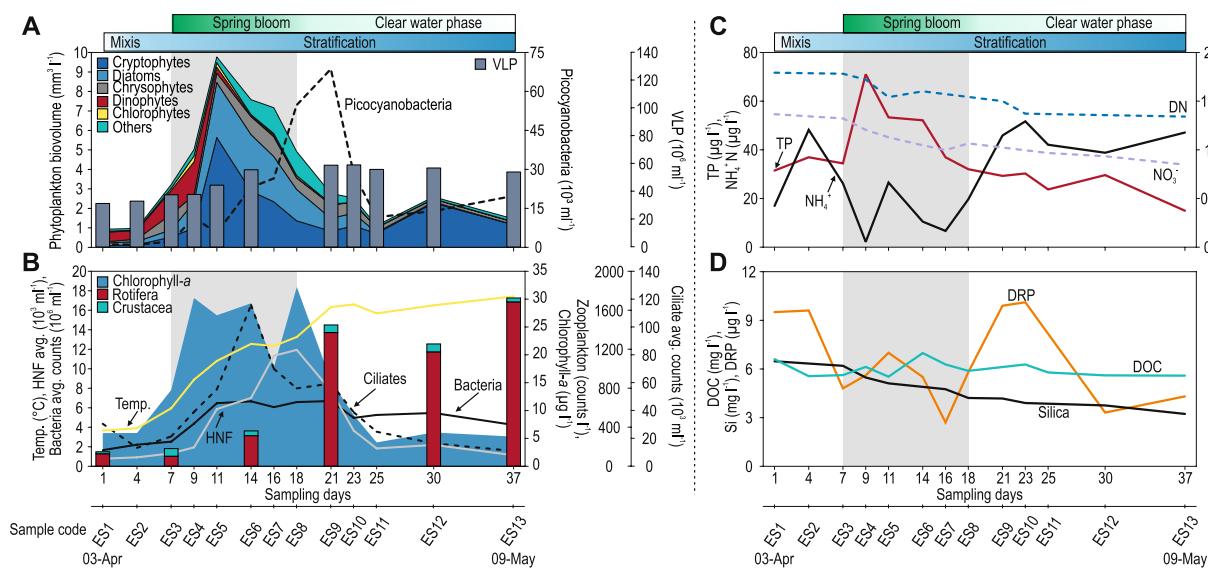


Fig. 1 Time course of different features of the spring bloom in the epilimnion. **A** Phytoplankton biovolume, picocyanobacterial abundances, and viral-like particle counts (VLP). **B** The concentration of chlorophyll-*a*, temperature, counts of total heterotrophic bacteria, heterotrophic nanoflagellates (HNF), ciliates, rotifers, and crustaceans. **C** Total phosphorus (TP), dissolved nitrogen (DN), nitrate and ammonium concentrations, and **D** dissolved reactive phosphorus (DRP), dissolved organic carbon (DOC), and silica concentrations. The gray rectangle in the background shows the duration of the spring bloom

column is vertically isothermal (4 °C), with low chlorophyll-*a* ($5 \mu\text{g l}^{-1}$), and high nutrient (dissolved inorganic nitrogen and phosphorus) concentrations. The second is characterized by the progressive stratification induced by increased temperature (days 7–37). Two major events mark the second stage. The spring bloom becomes evident with high measurements of chlorophyll-*a* (days 7–18). Phytoplankton started to increase between days 4 and 7, peaking at day 18 (chlorophyll-*a* concentration of $30 \mu\text{g l}^{-1}$). Finally, the clear water phase (days 30–37) becomes distinguishable by a well-established water density gradient, low chlorophyll values, and nutrient depletion, i.e., phosphorus. According to microscopic observations, diatoms and photosynthetic cryptophytes accounted for most phytoplankton biomass during the bloom period (days 7–18). A number of diverse diatom genera, e.g., *Asterionella*, *Fragilaria*, *Nitzschia*, *Synedra*, and centric diatoms (potentially *Cyclotella*) prevailed, in addition to the mixotrophic haptophyte *Chrysochromulina* (Additional file 2: Table S1). A small but discernible maximum in cryptophytes was also observed around day 30 (Fig. 1).

The highest concentration of total phosphorus (TP) was paralleled with the lowest availability of DRP (dissolved reactive phosphorus) during the bloom, reflecting a fast turnover of dissolved inorganic phosphorus and its accumulation by phytoplankton. Ammonium was rapidly depleted owing to uptake by phytoplankton as reduced nitrogen forms are preferentially taken

up, and if not available, phytoplankton utilizes nitrate as an alternative nitrogen source [35]. Nitrate also decreased gradually during the study period. Upon bloom-collapse, ammonium and DRP were released from phytoplankton by excretion and sloppy feeding of zooplankton [36–38]. Total heterotrophic prokaryotic numbers rapidly increased from 2 to 6×10^6 cells ml^{-1} during the phytoplankton bloom and remained relatively stable thereafter. Picocyanobacteria reached their highest levels after the first bloom event (day 21). However, they accounted for only a negligible portion of the total phytoplankton biomass (maximum 7×10^4 cells ml^{-1}). Ciliates reached their maximum on day 14, preceding a peak of HNF (heterotrophic nanoflagellates) by 4 days. The abundance of zooplankton (rotifers and crustaceans) continuously increased and reached the highest levels in the clear water phase (Fig. 1, Additional file 1: Figure S1).

Despite ciliates and HNF being important consumers of prokaryotes in aquatic habitats, their temporal dynamics are hard to capture since they often form short-lived peaks, necessitating a high-frequency sampling approach [4–6]. In this study, we detected peaks for both groups during the spring bloom (Fig. 1, Additional file 1: Figure S2). Estimates of total grazing rates suggest that HNF are more important bacterivores than ciliates during this period (Additional file 1: Figure S2), which has been observed before [6]. There are also clear shifts in the ciliate populations, from those that feed largely on

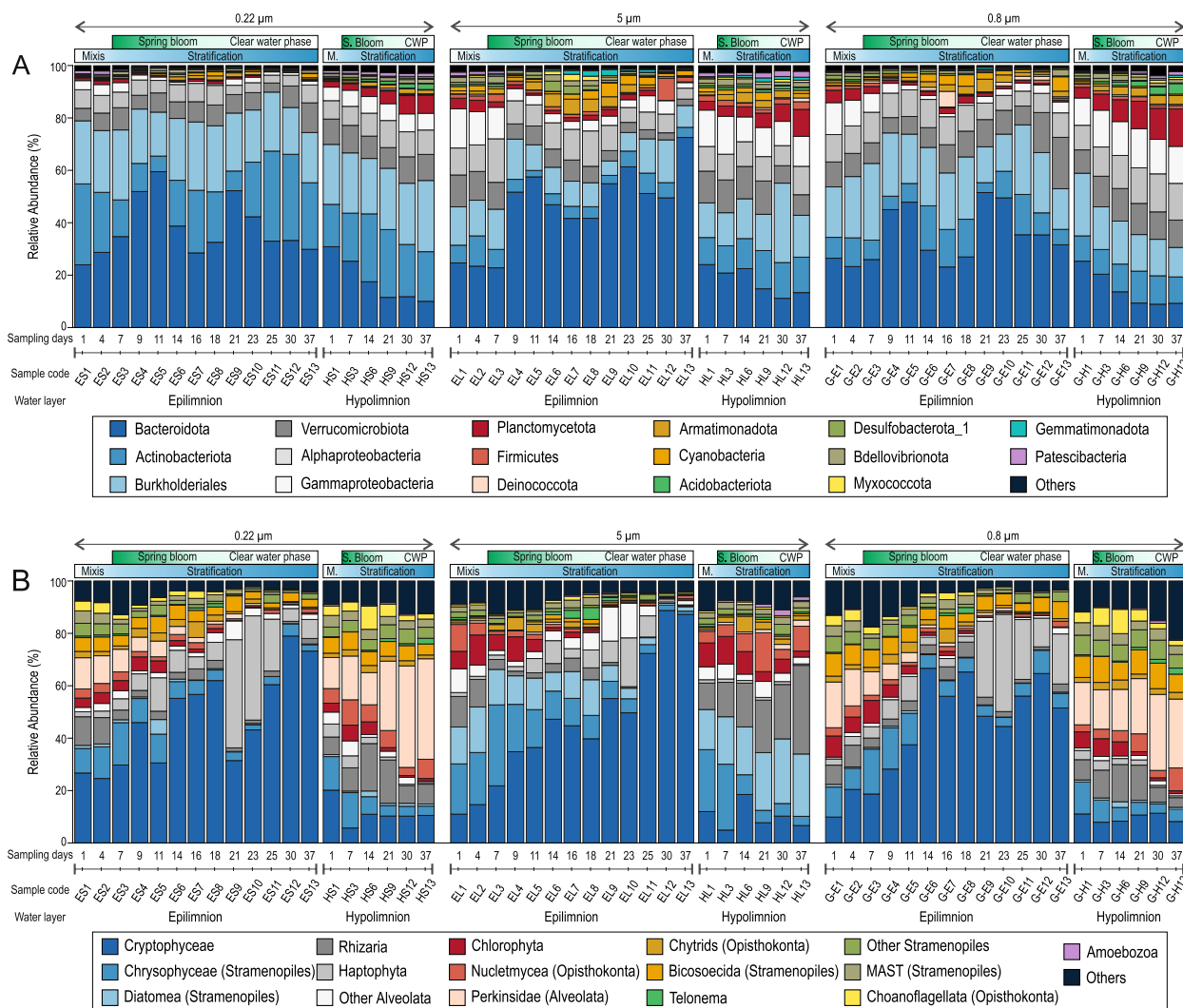


Fig. 2 Relative abundances of different taxonomic groups (expressed as a percentage) in the spring bloom as assessed by 16S rRNA (top, prokaryotes) and 18S rRNA (below, eukaryotes) gene read sequences. The results from different filters are shown (left, 0.22 μm; middle, 5 μm and right, 0.8 μm). Epilimnion and hypolimnion samples are also indicated

algae, HNF, and cryptophytes (e.g., *Balanion*, *Rimostrombidium*, *Urotricha*) [24, 39], towards filter-feeding omnivores (e.g., *Halteria*) capable of feeding on multiple food sources in a size range from small bacteria to small algae and HNF [24, 40].

A metagenomic perspective of the spring bloom

We analyzed 57 metagenomic samples from two different depths over 37 days (a total of 1.96 billion reads, ca. 830 Gb), using multiple filter sizes (5 μm, 0.8 μm, and 0.22 μm) and different filtration methods (see Methods) providing an unprecedented view of the changes in the microbial community across time (Fig. 2). 16S rRNA screening of the 0.22-μm filters indicated that

Actinobacteriota, Bacteroidota, and Burkholderiales (formerly Betaproteobacteria) are the most dominant in the epilimnion during the entire period, followed by Alphaproteobacteria and Verrucomicrobiota (Fig. 2). However, even at this broad taxonomic level, distinct responses of these taxa are clearly discernible. Bacteroidota rapidly increased to >50% of 16S rRNA reads at the onset of the phytoplankton bloom (days 9–11), slightly decreased thereafter, and showed a second maximum at day 21 when the bloom started to decay. These bacteria are likely relatively large or associated with protists or aggregates [41], as is evident from their higher recovery from the 5-μm filter. The maximum abundance of Bacteroidota happened in parallel with two sharp decreases

in the relative abundance of Actinobacteriota (belonging mostly to the order 'Ca. Nanopelagicales') detected in the epilimnion on days 11 and 21, followed by their recovery to previous levels (Fig. 1). Freshwater Actinobacteriota are repeatedly reported to be among the most dominant microbial groups across various freshwater habitats, including this particular site [20, 32, 42, 43]. Opposed to Bacteroidota, they were much less recovered from 5- μm and 0.8- μm filters because of their small size. On the other hand, sequences affiliated with Burkholderiales appeared to be relatively stable throughout the sampling period. A continuous increase and a preference for the hypolimnion in certain taxa is visible for Gammaproteobacteria and Planctomycetota (especially on the 0.8- μm filter). Planctomycetota have been previously shown to prefer hypolimnetic waters and appear in the upper layers mainly when the stratification is eroded [44]. A similar trend was observed here, with their numbers continuously decreasing in the epilimnion with rising temperature and gradually increasing in the hypolimnion. Surprisingly, we also observed an increase in Armatimonadota in the hypolimnion, a group scarcely reported from freshwaters. All other microbial groups remain at lower levels except for small increases in Deinococcota captured on the 0.8- μm gravity filters in the epilimnion on day 16.

Similar to the trends observed for (plastidic) cryptophytes in microscopy (Fig. 1), cryptophyte rRNA gene sequences represent the major part of all eukaryotic sequences recovered from the epilimnion (Fig. 2). At the same time, their abundances are considerably lower in the hypolimnion. Both cryptophyte peaks that were observed microscopically (Fig. 1) were also detected in the rRNA gene abundance analysis (Fig. 2). Across all filter sizes, and in both epilimnion and hypolimnion, Chrysophyceae and Chlorophyta sequences appeared to decline continuously, and Rhizaria (SAR group) sequences were more abundant in the hypolimnion.

Short-lived peaks in other groups were also observed. For example, two peaks of haptophytes were visible in the epilimnion on days 11 and 21–25 (Fig. 2). These blooms were mostly composed of the mixotroph *Chrysochromulina* (Prymnesiophyceae), as we also recovered 19 metagenomic bins that could be ascribed to *Chrysochromulina* (see below). *Chrysochromulina* was also observed in the samples accounting for 1–15% of the total phytoplankton biovolume in the later stages of the bloom. Another consistent feature of the bloom was the appearance of diatoms (confirmed by microscopy) in the first phase of the bloom (Fig. 1, Additional file 2: Table S1). The rRNA sequences of diatoms could be mostly ascribed to *Thalassiosira* and corroborated by the recovery of multiple partial genomes of *Thalassiosira*

from the metagenomic data (see below). Other photosynthetic protists, e.g., Chlorophyta, continuously decrease in abundance as the bloom progresses (Fig. 2).

Most studies on the freshwater spring bloom have focussed primarily on the epilimnion [4–6, 26] because photosynthetic phytoplankton has been considered the major players. The changes in the microbial community in deeper water layers have remained relatively unknown. Remarkably, the abundance of sequences ascribed to Perkinsozoa (a poorly understood group of parasitic protozoa) increased in the hypolimnion and decreased in the epilimnion. Conceivably, a low-temperature preference for Perkinsozoa might exist, although their extremely high relative abundances (ca. 40%) in the hypolimnion (Fig. 2) suggest there may be other unknown driving factors as well. It remains unclear if these Perkinsozoa are parasitic or free-living. Parasitic Perkinsozoa are better known for marine habitats and infect mollusks [45], dinoflagellates [46], chlorophytes [47], and even fish [48]. They have also been reported from surface freshwaters [15, 49, 50] and deeper water strata [51, 52]. Some perkinsids have even been described to infect cryptophytes (e.g., the perkinsid *Cryptophagus* infecting the cryptophyte *Chilomonas*) [53]. Given their broad host range in the freshwaters and the eukaryotic community's complexity, their preferred freshwater hosts remain obscure. In case they infect cryptophytes, which represent a prominent segment of the spring phytoplankton bloom, the abundant host population would also provide a niche for the rapid proliferation of the parasites. However, as perkinsids are far more abundant in the hypolimnion than in the epilimnion, it is unlikely that they affect bloom collapse significantly in the epilimnion where the bloom takes place. Moreover, the viral collapse of marine algal blooms is well-known [54]. In support of this hypothesis, the metagenomic data suggest a greater abundance of perkinsids in the hypolimnion but not in the epilimnion. On the other hand, eukaryotic viruses are prevalent throughout the water column and likely contribute more to the bloom collapse in the epilimnion in concert with other bottom-up controls e.g., nutrient limitations (see also below).

Recovery of bacterial and phage genomes

A total of 2214 bacterial metagenome-assembled genomes (MAGs) ($\geq 40\%$ completeness and $\leq 5\%$ contamination) were reconstructed from 57 metagenomic datasets generated from three different filter-pore sizes (0.22 μm , 0.8 μm , 5 μm) and two different depths. See Additional file 3: Table S2 and Additional file 1: Figure S3 for more details. The number recovered from each filter varies greatly between fractions, 1288 from 0.22 μm , 808 from 0.8 μm , and only 112 from the 5 μm filter, i.e.,

increasing filter-pore size reduces prokaryotic genome recovery dramatically as they likely capture more eukaryotic genomic material. The most recovered bacterial MAGs were from the phyla Bacteroidota ($n=690$), Proteobacteriota ($n=677$: 493 Gammaproteobacteria and 184 Alphaproteobacteria), Actinobacteriota ($n=453$), Verrucomicrobiota ($n=196$), and Planctomycetota ($n=92$) (See also Additional file 3: Table S2). De-replication of these MAGs resulted in 855 genomes representing prokaryotic bacterial diversity along the spring bloom (Additional file 3: Table S2). We did not recover any archaeal MAGs from these datasets.

It is worth noting that 615 MAGs encode putative rhodopsins distributed across 326 clusters of very diverse phyla, i.e., Actinobacteriota, Bacteroidota, Chloroflexota, Proteobacteria, and Verrucomicrobiota. Most encode type-1 rhodopsin ($n=508$, 223 clusters) and a smaller number of heliorhodopsin ($n=214$, 103 clusters). Moreover, 203 of those MAGs encode both types of rhodopsin (48 clusters). In sum, nearly 38% of the microbial community (326 out of 855 dereplicated MAGs) responds to light in a rhodopsin-dependent fashion (Additional file 4: Table S3). Almost all these type-1 rhodopsins across different phyla display the DTE motif in transmembrane helix 3, which suggests all these are proteorhodopsins [55], enabling outward proton translocation to generate the proton-motive force to produce ATP. Only a few exceptions to this were found, with less than ten sequences with DTG and ATI motifs that also indicate outward H^+ pumping activity. Moreover, all detected rhodopsins were found to be green-absorbing rhodopsins (leucine as residue 93, bacteriorhodopsin numbering) [56], except for two sequences originating from freshwater '*Ca. Nanopelagicales*' MAGs appeared to be blue-light absorbing (glutamine as residue 93). It has been suggested before that as blue light penetrates deeper into the water column, organisms encoding blue-light-absorbing rhodopsins might have an advantage in deeper waters [57]. These two MAGs belong to two clades that are basal to '*Ca. Nanopelagicus*' and '*Ca. Planktophila*' and are found at low abundances at the mixis phase and later transition to deeper waters where the possession of blue-light absorbing rhodopsin may provide a selective advantage. Other MAGs related to these possess green-light-absorbing rhodopsins.

While viruses are recognized as important participants in the spring bloom, no freshwater viruses have been isolated, nor have any viral genomes recovered from this specific period. We recovered 679 complete phage genomes (see the "Methods" section) from the three different filter types that were dereplicated to 175 complete phage genomes. Previously, we have published a large dataset of complete freshwater phage genomes [20], and

1398 genomes originated from the same site (Římov phages). De-replication of all phage genomes from the same site in these two datasets together revealed that of the 679 phage genomes recovered in this study, 129 had already been seen before and may be considered 'persistent' dwellers in the environment for at least 2 years. The remaining 462 (dereplicated 162) are novel phage genomes identified in this work. We identified 35 genomes (out of 679) that were predicted to infect freshwater Actinobacteriota. Actinophages can be recognized because they encode an actinobacteriota-specific transcription factor, *whiB* [58]. Only three of these 35 phages were described before [20]. We also recovered six complete phage genomes (30–31 Kb, three dereplicated clusters) that are predicted to infect freshwater cyanobacteria. All of these encodes a cyanobacterial-specific RNA polymerase sigma factor (TIGR02997). This gene is largely specific to cyanobacteria (99% of all hits in GTDB are cyanobacterial) [59], strongly suggesting these are freshwater cyanophages. It has been shown before that similar sigma factors carried by cyanophages strongly repress cyanobacterial growth [60], and given that no photosystem genes were found in these phage genomes, this gene is likely important for the phage infection cycle. In further support of these being cyanophages, the metagenomic abundances of two of these predicted representative cyanophage genomes (HRS-EL5-C113 and HRSG-E8-C80, Additional file 1: Figure S4) also appeared to match the peak of picocyanobacteria by microscopy (time points ES9 and ES10, Fig. 1). Most freshwater cyanophage genomes from cultured isolates do not encode photosystem genes. Of the 18 currently available genomes, only two encode photosystems (both infecting *Synechococcus*), indicating that capturing photosystems for the phage infection cycle is not very common for freshwater cyanophages and other strategies may be more prevalent (Additional file 5: Table S4). Another seven phage genomes were predicted to infect Burkholderiales genomes using matches from tRNA genes (see methods) (Additional file 5: Table S4).

Dynamics of selected, abundant prokaryotic groups

Above all, the spring bloom is a transitory period for microbial populations that are directly affected by the increase in light availability and temperature and the onset of stratification coupled with rapid changes in the availability of nutrients [2]. This results in distinct microbial communities at both ends of the spectrum. The availability of MAGs from the entire spring period allows us to distinguish distinct modes of succession in the spring bloom at a genome level (Fig. 3 and Additional file 1: Figure S5~S7). Temperature dependence is visible for several bacterial populations and phages (Fig. 3, Additional

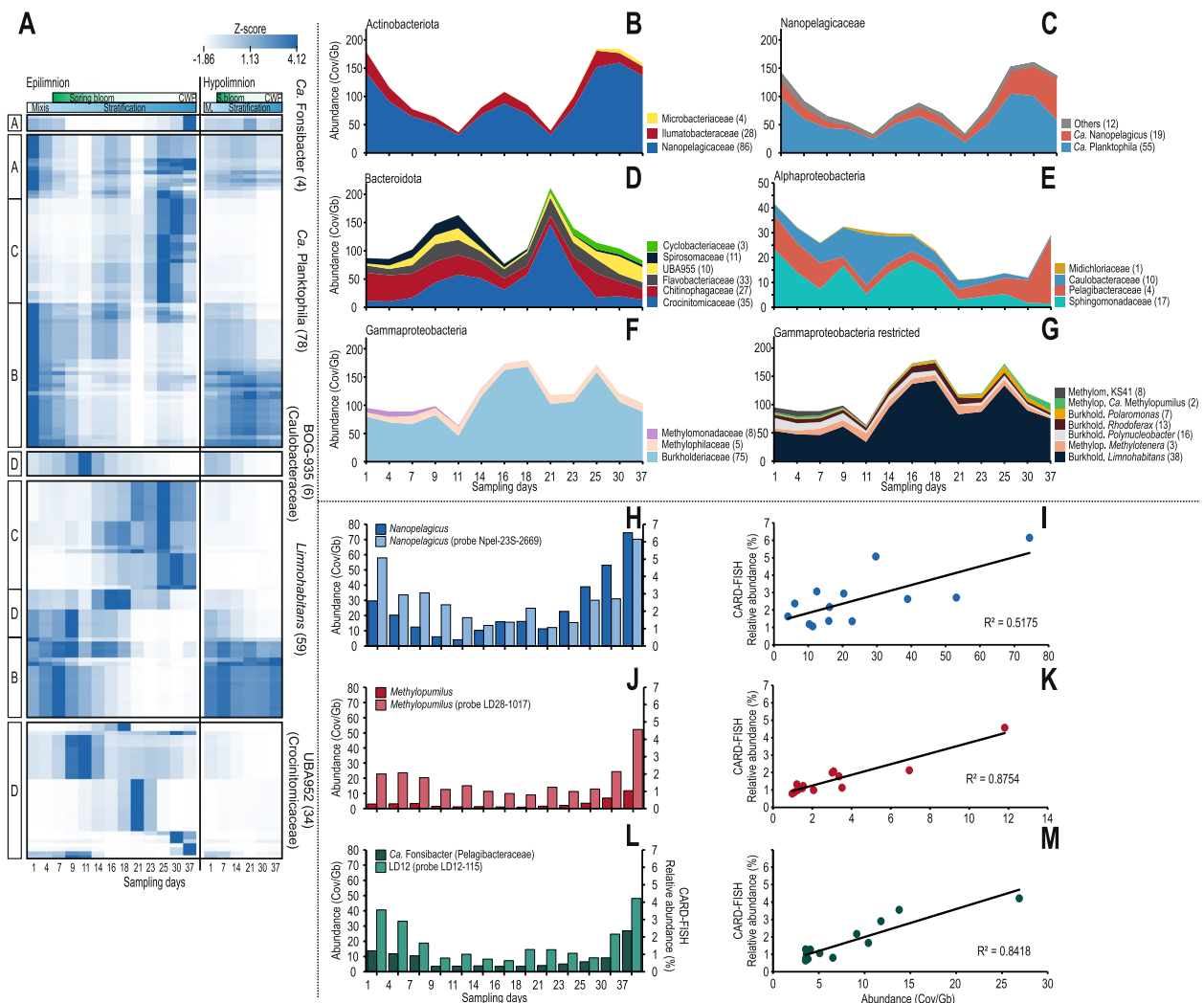


Fig. 3 **A** Relative abundance of selected dereplicated MAGs during the spring bloom in the small filter (0.22 μ m). Each row was normalized by Z score and clustered by average linkage (with Spearman’s rank correlation method). Different letters represent different patterns of abundance common within groups. The numbers of genomes used for each group are mentioned in parentheses. For more details about other groups, check Additional file 1: Figure S4, S5, and S6. **B–G** Abundances of selected microbial genomes/groups during the spring bloom: The X-axes show the progression of the bloom (in days), and Y-axes show the abundance of the MAGs. All Y-axes range from 0 to 200 except for Alphaproteobacteria (0–50). The numbers of dereplicated genomes used for each group are given in parentheses. Methylomonadaceae, Methylophilaceae, and Burkholderiaceae are abbreviated as ‘Methylom,’ ‘Methylop,’ and ‘Burkhold.’ respectively, in the figure. **H–M** Metagenomic abundance vs CARD-FISH relative abundance. **H, J, and I** Coverage per Gb on the left Y-axis and CARD-FISH relative abundance % on right Y-axis. **I, K, and M** Correlation between metagenomic abundance vs CARD-FISH relative abundance

file 1: Figure S4, S8~S12). At least four broad categories of populations are detectable during the spring period, (A) temperature generalists that are abundant both at the beginning and the end of the study period, perhaps with some fluctuations during the bloom, likely well adapted to both cold and warm temperatures (e.g., ‘*Ca. Fonsibacter*’, some ‘*Ca. Planktophila*’); (B) populations preferring colder temperatures that are primarily abundant in the epilimnion in the beginning but transition to the hypolimnion as surface temperature rises (some

‘*Ca. Planktophila*’ and *Limnohabitans*); (C) populations preferring warmer temperature, gradually reaching maxima in the clear water phase and not abundant in the hypolimnion (some ‘*Ca. Planktophila*’ and *Limnohabitans*); and finally (D) populations that also prefer warmer temperatures, but show sporadic, short-lived peaks in the epilimnion and similarly to (C) are not abundant in the hypolimnion (BOG935, UBA952, and some *Limnohabitans*). Populations D and A can broadly be categorized as bloom specialists and bloom avoiders, respectively.

It is also apparent that phylogenetically closely related genomes may have different ecological behavior, suggesting that minor genomic alterations may have significant ecological consequences (Additional file 1: Figure S5–S7). The genome-streamlined Actinobacteriota represented by ‘*Ca. Planktophila*’ (‘*Ca. Nanopelagicales*’) appear to consist of very diverse populations displaying different dynamics during the spring period. Moreover, the populations in the epilimnion at the beginning of the spring bloom are quite different from those at the end, and groups that show similar abundance profiles are not necessarily phylogenetically related (Fig. 3, Additional file 1: Figure S6). Remarkably similar population replacement patterns in the epilimnion and transitions to the hypolimnion can also be seen for the faster-growing copiotroph *Limnohabitans* (Additional file 1: Figure S7). A high (micro-)diversity with different habitat preferences and organic matter utilization has been proposed for both taxa [32, 61, 62]. The opposite holds true for genome-streamlined Alphaproteobacteria represented by ‘*Ca. Fonsibacter*’ (four de-replicated genomes), where the abundances of genomes across the timeline are very similar with only minor differences. Their abundance declines with increasing chlorophyll and total phosphorus and remains low until the same populations recover during the clear water phase. The same clades also appear in the hypolimnion, and a clear-cut temperature dependence is not discernable. This suggests an intriguing reason for the success of ‘*Ca. Fonsibacter*’ in freshwaters in a low-chlorophyll regime: there is low overall genomic diversity (Additional file 1: Figure S4), but this low diversity is well adapted to variations in temperature [14, 63, 64]. This contrasts with most ‘*Ca. Planktophila*’, where genomic diversity is high (Additional file 1: Figure S5), and separate clades usually have different temperature preferences. At the same time, only a few populations are equally abundant at the beginning and the end of the spring period.

Fine-scale dynamics of different taxa have been examined during the spring bloom using CARD-FISH, e.g., ‘*Ca. Nanopelagicales*’ (formerly acI Actinobacteria), *Limnohabitans*, and some lineages of Bacteroidota [4–6], but never coupled with a metagenomic assessment (i.e., MAGs) what allows a much finer distinction even between closely related organisms than is possible by CARD-FISH alone. For Actinobacteriota, as has been stated before, most MAGs were affiliated to ‘*Ca. Nanopelagicales*’ and, to a lesser extent Illumatobacteriaceae (formerly acIV Actinobacteria) and Microbacteriaceae (Fig. 3). In the epilimnion, there is a pronounced minimum in the abundance of actinobacterial MAGs on day 11, a recovery until day 16, and another decline on day 21, followed by a recovery until day 25. The population

decreases in ‘*Ca. Nanopelagicales*’ can be more specifically attributed to ‘*Ca. Planktophila*’, while the related genus, ‘*Ca. Nanopelagicus*’, which is generally more stable throughout, starts increasing in the clear water phase (days 30–37). A similar trend with sharp declines and rapid growth thereafter is also visible for Burkholderiaceae (mainly *Limnohabitans*, day 11) and Sphingomonadaceae (days 7 and 11, Fig. 3). However, it is unclear if these population collapses were caused by phage infection, grazing by protists, or competitive disadvantages to fast-growing copiotrophs. At the same time, other groups dramatically increase in numbers, e.g., Caulobacteraceae (at day 11), Crocinibacteraceae (days 11 and 23), and to a lesser extent, Flavobacteriaceae (day 9). Caulobacteraceae (genus BOG935), Crocinibacteraceae (genus UBA952), and Flavobacteriaceae appear to be bloom specialists, dominating only for a short period during favorable conditions. Similar typical algal bloom specialists have been observed before (e.g., *Fluviicola*, *Flavobacterium*) [4]; however, BOG935 and UBA952 represent new additions to this category of fast-growing copiotrophs, but no cultured representatives are as yet available. Also apparent is the relatively fast growth of *Limnohabitans* after a minimum at day 11, after which their abundance appears to stabilize. Other characteristic patterns are also visible, e.g., ‘*Ca. Fonsibacter*’ (shown as Pelagibacteriaceae in Fig. 3) has a pronounced minimum during the bloom; it is well known that these oligotrophs start growing in the clear water phase and reach their annual maximum in summer-autumn [16].

These relatively fast changes in multiple populations prompted us to examine the possibility of estimating in-situ growth rates for at least some taxa using metagenomic abundance data, especially for genome-streamlined microbes (‘*Ca. Nanopelagicus*’, ‘*Ca. Fonsibacter*’, and ‘*Ca. Methylophilus*’). While representatives of these taxa are available in culture [17, 65, 66], their doubling times in the natural environment remain unknown, except for a doubling time estimate made collectively for ‘*Ca. Nanopelagicales*’ (formerly acI Actinobacteria) [32], ca. 66 h based on CARD-FISH abundances [6]. Moreover, while some of these microbes have been obtained in culture, doubling time estimates from cultures might not be similar to those in real environmental conditions [67] where there is significant competition for resources, mortality by grazers (e.g., heterotrophic nanoflagellates) and viral infections. We also performed CARD-FISH on these selected genome-streamlined microbes throughout the spring bloom, using general probes for ‘*Ca. Nanopelagicales*’ and Methylophilaceae, and specific probes for the genera ‘*Ca. Nanopelagicus*’, ‘*Ca. Fonsibacter*’, and ‘*Ca. Methylophilus*’ (See Additional file 6: Table S5).

Table 1 Observed and estimated doubling times in hours (h) (\pm SD) for genome-streamlined, oligotrophic taxa obtained using different approaches. The number of MAGs used for each group is indicated inside the brackets

Group	Doubling times				
	CARD-FISH		Metagenomic abundance		gRodon
	Observed, in situ ^a (h)	Estimated ^b (h)	Observed, in situ ^a (h)	Estimated ^b (h)	Inferred from genome (h)
<i>Ca. Nanopelagicales</i> (41)	72.12 \pm 16.68	36.67 \pm 4.15	44.44 \pm 7.69	27.86 \pm 3.18	10.67 \pm 7.15
<i>Ca. Nanopelagicus</i> (17)	106.72 \pm 72.72	56.06 \pm 40.81	46.85 \pm 3.16	28.94 \pm 1.31	9.21 \pm 2.43
Methylophilaceae (5)	87.85 \pm 35.01	43.43 \pm 14.03	80.1 \pm 17.17	44.2 \pm 9.34	11.2 \pm 11.21
<i>Ca. Methylopusillus</i> (2)	134.74 \pm 63.17	70.39 \pm 33.3	61.66 \pm 1.67	34 \pm 0.5	22.98 \pm 4.26
<i>Ca. Fonsibacter</i> (3)	114.66 \pm 113.2	62.99 \pm 51.01	106.78 \pm 10.58	76.07 \pm 5.47	31.12 \pm 15.61

^a Observed, in situ: average of in situ doubling times (in hours) obtained using observed abundances (assuming mortality)

^b Estimated: average of estimated doubling times (in hours) obtained after correcting for 55% mortality

A high concordance was observed between the relative abundances in metagenomes, and CARD-FISH for all three genome-streamlined taxa examined (Fig. 3). Because mortality rate due to protists and viral infections is hard to predict without predation experiments, we used a value of 55% mortality rate for all bacterial groups (on average, 24% from grazing and 31% from viral infections) obtained from 19 temperate lakes [68]. Lastly, we also applied a codon-bias-based doubling time estimation from genomes using gRodon [69].

In the case of '*Ca. Nanopelagicales*', Methylophilaceae, and '*Ca. Fonsibacter*', there is broad agreement between doubling time estimates from recruitment and CARD-FISH (Table 1, Fig. 3), while results for '*Ca. Nanopelagicus*' and '*Ca. Methylopusillus*' differed by about two-fold. Generally, there appears more concordance at the higher taxonomic level than at the genus level. For all microbes, gRodon predicted a doubling time of > 5 h, which only suggests that the microbe is slow-growing, and the results are considered to be an underestimation. The observed doubling times for '*Ca. Nanopelagicales*' (72 h) are remarkably similar to those previously reported (66 h) [6]. Upon correction, for mortality, this value was reduced to 37 h. Estimated doubling times for '*Ca. Nanopelagicales*' and Methylophilaceae (using CARD-FISH or metagenomic abundance) fall in a similar range of 28–44 h. Surprisingly, similar estimates have been obtained for cultures of '*Ca. Methylopusillus*' (0.4 divisions per day) [17] and '*Ca. Fonsibacter*' (0.52 divisions per day) [65], and even for marine *Pelagibacter ubique* (0.4–0.58 divisions per day) [70], which translates to a doubling time of roughly 2 days, not very different from the values obtained here in the presence of mortality. This suggests that existing culture media for genome-streamlined microbes are already well-optimized. As is evident from the

variations in doubling times computed at different sampling times for all microbial groups, the accuracy of such estimates may be influenced by environmental conditions and even the mode of observation. Still, they do provide a more realistic appreciation of the growth of highly abundant and cosmopolitan microbes in the natural environment.

Recovery of eukaryotic and giant viral genomes

Owing to the genomic complexity and large genome sizes, eukaryotic genomes have rarely been recovered from metagenomes, and the only recent report is from the marine habitat [71]. Upon examining the recovered MAGs after binning, we noticed that 25 appeared to originate from eukaryotes. Of these, 19 were highly related to *Chrysochromulina* (a Haptophyta) and four to *Thalassiosira* (a centric diatom). Genome completeness estimates using eukaryotic or lineage-specific markers (wherever identifiable) show completeness in the range of 6–37% (See Additional file 7: Table S6). Apart from *Chrysochromulina* and *Thalassiosira*, two other bins appeared to be affiliated to stramenopiles and cryptophyta, but as their genome completeness was only ca. 10%, they were not examined further. The recovered MAGs of *Chrysochromulina* and *Thalassiosira* represent the first metagenome-assembled genomes from freshwater eukaryotes.

The genus *Chrysochromulina* is broadly distributed in marine and freshwater ecosystems [72, 73] and has been documented as an under-ice bloomer in nutrient-poor conditions in habitats influenced by freshwater [74]. These mixotrophic algae play crucial ecological roles in global carbon sequestration and bloom formation and are a rich food source for an extensive range of grazers [75–77]. Average nucleotide identity comparisons between all bins of *Chrysochromulina* suggested the presence of two distinct clusters with 11 (group 1) and 8 bins (group 2)

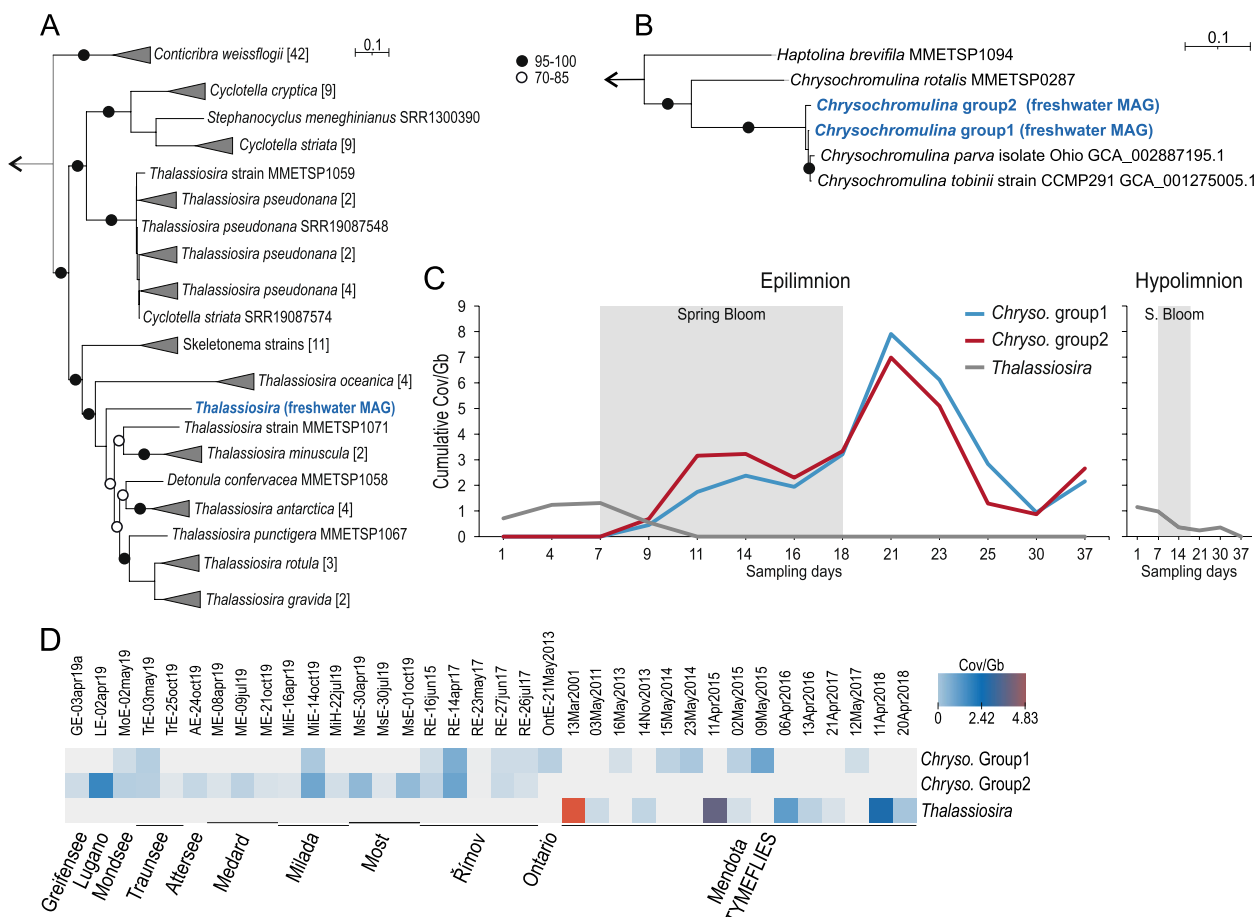


Fig. 4 **A** Phylogenomic trees of *Thalassiosira*. **B** *Chrysochromulina* metagenome-assembled genomes. Ultrafast bootstrap values (UFB) are defined by full (95–100) and empty (70–85) circles. Collapsed clades (triangles) followed by square brackets containing the total number of members within the clade. **C** Metagenomic abundance of MAGs of *Chrysochromulina* and *Thalassiosira* along the entire spring bloom timeline (epilimnion and hypolimnion). **D** Heatmap of metagenomic abundances of MAGs of *Chrysochromulina* and *Thalassiosira* across several freshwater metagenomic datasets. A scale is shown to the right. *Chrysochromulina* is abbreviated as “Chryso.” in **C** and **D**

each (Additional file 1: Figure S13). As most of these bins had genome completeness estimates of less than 40%, we collapsed both groups [78] to improve genome recovery and obtained much-improved genome completeness estimates for both groups (51% for group 1 and 44% for group 2) and genome size approached 59 Mb for both groups (group 1, 59.7 Mb and group 2, 59.97 Mb). These genome sizes are comparable to the available genomes of freshwater *Chrysochromulina* isolates (*C. tobini*: 59 MB, 58% completeness; *C. parva*: 65 MB, 68% completeness) [79, 80]. Comparisons of these recovered bins to the two available *Chrysochromulina* genomes indicated extremely high genomic identities, particularly to group 1 MAGs (Additional file 1: Figure S13). This was also supported by phylogenomic analyses using conserved eukaryotic marker genes (Fig. 4). As mentioned before, a haptophyte bloom was visible after the main phytoplankton bloom (predominantly cryptophytes) on days 21 and 23,

accounting for nearly 20% of all rRNA sequences (Fig. 2), which is also in line with a maximum in metagenomic fragment recruitment of the recovered genomes (Fig. 4). Similarly, for *Thalassiosira* (a centric diatom), we obtained four metagenomic bins that were ca. 50% complete (Additional file 7: Table S6). These MAGs had nearly 100% average nucleotide identity to each other but very low ANI to available *Thalassiosira* genomes (Additional file 1: Figure S13). We merged these MAGs into a consensus genome that was ca. 31 MB in size and estimated to be 63% complete. Genome sizes from cultures of *Thalassiosira* appear highly variable, i.e., *T. pseudonana* (freshwater and brackish) [81] and *T. oceanica* (marine) [82] have quite different genome sizes ranging from 29 to 90 MB, respectively. Phylogenomic analysis using all available genomes and transcriptomes from centric diatoms (*Coscinodiscophyceae*) also placed this MAG within the genus *Thalassiosira* (Fig. 4). This is surprising

as *Thalassiosira* are usually considered marine species and *Cyclotella* is understood to be prevalent in freshwater lakes [83]. Our microscopic identification, in sharp contrast to the phylogenomic analyses also potentially identified this organism as a *Cyclotella*. Phylogenomic analyses, however, suggest (Fig. 4) that the genus *Cyclotella* is perhaps polyphyletic. In any case, the organismal genome recovered here, according to genomic analyses appears better placed to be classified as a *Thalassiosira* and not *Cyclotella*.

The freshwater *Thalassiosira* appears to be present only at the beginning of the spring bloom and declines thereafter, in line with 18S rRNA screening (Fig. 2), and its profile is similar in both epi- and hypolimnion. We could also detect the *Chrysochromulina* and *Thalassiosira* MAGs in metagenomes from different locations in Europe and North America (Fig. 4), indicating that the organisms represented by these MAGs are widely distributed across large geographical distances.

Since the discovery of Nucleo-Cytoplasmic Large DNA Viruses (NCLDV, phylum Nucleocytoviricota) [84, 85], giant viruses continue to be isolated from diverse protists, e.g., the kinetoplastid *Bodo saltans* [86], the bicoecid *Cafeteria roenbergensis* [87], and the marine coccolithophore *Emiliana huxleyi* that forms vast phytoplankton blooms that can be visualized from space [88]. Giant viruses are not only unique in that they have extremely large genomes (in excess of 1 MB), but also that they encode a nearly complete protein translation apparatus hitherto unknown for any other viruses. Metagenomic identification of giant viral genomes from diverse habitats and as endogenous viruses in protist genomes has recently exposed a vast diversity [89–91]. In the marine habitat, it has been recognized that giant viruses infecting *E. huxleyi* are important contributors to bloom collapse [92, 93], supporting aggregation of particles, higher zooplankton grazing, and increased transfer of carbon to deeper water levels [88]. However, abundant giant viruses from freshwater habitats, particularly their dynamics in a high-frequency sampling regime, remain largely unknown. We recovered 3309 contigs from nucleocytoplasmic large DNA viruses infecting microbial eukaryotes using ViralRecall (see the “Methods” section). These contigs ranged from 10 to 266 kb and were dereplicated to 1721 clusters.

Phylogenetic analysis using concatenated markers and protein sequence comparisons with a curated set of 939 reference NCLDV genomes [89–91] revealed that most of these NCLDVs belong to the order Imitervirales (Additional file 8: Table S7, Additional file 1: Figure S14). Additional comparisons to known NCLDV viruses with CheckV [94] suggested similarities to viruses infecting haptophytes (*Chrysochromulina ercinia* virus,

99 contigs; *Phaeocystis globosa* virus 163 contigs) and the chlorophyte *Tetraselmis* (*Tetraselmis* virus, 41 contigs). In total, tentative host assignment was possible for only 323 NCLDV contigs (Additional file 8: Table S7). This relatively low prediction rate is largely owing to the absence of NCLDV genomes/isolates from dominant spring bloom participants. Fragment recruitment of the NCLDV contigs across the entire time revealed short-lived peaks that suggest a near-continuous turnover in the epilimnion in contrast to deeper waters where their abundances appeared more stable (Additional file 1: Figure S15). The role of viruses in the top-down control of unicellular eukaryotes (the viral shunt) [95, 96] has long been understood to be vital in the release of organic and inorganic nutrients. Additionally, NCLDV viruses have been implicated in the collapse of the vast blooms of the marine haptophyte *Emiliana huxleyi* [54, 97]. Taken together, it is not inconceivable that NCLDVs may have a similarly important role in bloom collapse in the epilimnion.

Dynamics of plastidic and aplastidic cryptophytes

As cryptophytes are major players in the spring bloom (Fig. 1), we attempted to assess the abundance of different cryptophyte groups. As noted before, there appear to be two peaks of cryptophytes highlighted by microscopic observations and rRNA gene sequence abundance information (Figs. 1 and 2). However, microscopic observations suggest the first peak being larger than the second (Fig. 1), while rRNA gene sequence abundance peaks suggest two nearly equal maxima (Fig. 2). Additionally, most cryptophyte 18S rRNA gene reads were associated with *Cryptomonas* or *Teleaulax/Plagioselmis* (Fig. 5). *Teleaulax* has been reported before in arctic under-ice spring blooms [98] and in annual blooms in the Columbia estuary [23]. Consistent with this, we also recovered three complete and circular mitochondrial genomes that appeared to originate from a *Teleaulax*-related microbe. These genomes are considerably distinct from the published *Teleaulax* mitochondrial genome [99] (ca. 80% nucleotide identity) but, nevertheless, syntenic to the latter apart from the lack of introns from the *Cox1* gene in the assembled mitochondrial genomes (Fig. 5). A comparison of these mitochondrial genomes to all other available mitochondrial genomes also indicated *Teleaulax* as the closest neighbor (Fig. 5). Given the relatively low nucleotide identity, it is more likely that they originate from an uncultured cryptophyte.

Apart from *Cryptomonas* and *Teleaulax*-like sequences, which both belong to the order Cryptomonadales, the as-yet uncultured aplastidic CRY1 group (which does not belong to Cryptomonadales) appears also to be dominant (Fig. 5). The uncultured CRY1

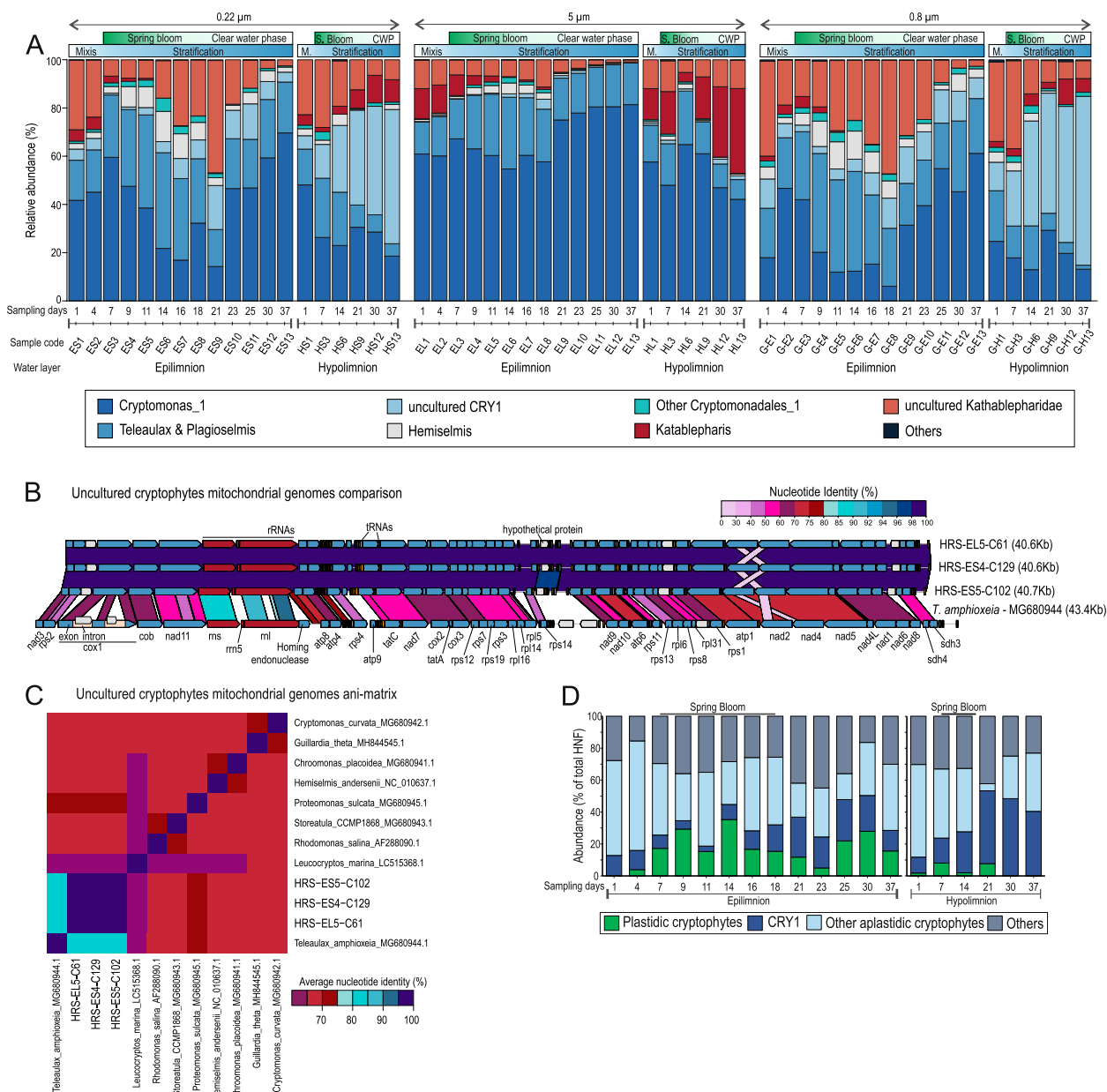


Fig. 5 **A** Dynamics of cryptophytes and closely related group Katablepharids during the spring bloom assessed by 18S rRNA gene read sequences. Results are shown as the percentage of total cryptophyte reads recovered. **B** Genomic comparison (nucleotide-nucleotide) of three complete uncultured cryptophytes mitochondrial genomes recovered here with the only available reference mitochondrial genome for *Teleaulax amphioxeia*. A scale is shown at the top right. **C** Average nucleotide identity (ANI) of uncultured cryptophyte mitochondrial genomes in comparison to all known cryptophyte mitochondrial genomes. **D** CARD-FISH abundance estimates for plastidic and aplastidic cryptophytes are shown as % of total eukaryotes. Sampling days are indicated along the X-axis

group has been described as among the major bacterivorous flagellates in freshwaters [100, 101] (Fig. 5). While CRY1 has been shown to be potentially a major bacterivore in the epilimnion [31], where a peak is also evident (Fig. 5), its abundances in deeper water strata have remained less examined. With increasing stratification, the abundance of CRY1 rises gradually

in the hypolimnion (Fig. 5). However, the abundance of the uncultured CRY1 alone does not explain the Cryptomonadales associated 18S rRNA gene abundances in the epilimnion (Fig. 1). Such discrepancies between microscopic counts and sequencing have been described before, specifically for cryptophytes, including the CRY1 lineage that may be under-represented

[102] and other protists, e.g., diplomonads that may be over-represented [103]. This inconsistency between sequence-based abundances and microscopic counts for cryptophytes may either be explained by non-photosynthetic or aplastidic cryptophytes within the order Cryptomonadales that are closely related to *Cryptomonas* (e.g., *Chilomonas*). This inference is supported by the observation that another known aplastidic lineage (apart from CRY1) represented by *Goniomonas* [104] was practically absent in this time period and most 18S rRNA gene reads could clearly be classified as Cryptomonadales. Hence, we suspect that these aplastidic cryptophytes (non-CRY1) are likely related to photosynthetic cryptophytes and are responsible for the second peak, while the first peak is primarily caused by photosynthetic cryptophytes as is also evidenced by elevated chlorophyll-*a* concentrations (Fig. 1).

To ascertain the prevalence of non-photosynthetic cryptophytes in this time period, we performed CARD-FISH analyses on the entire timeline in both the epi- and hypolimnion for cryptophytes using the general cryptophyte probe CryptoB (targeting all plastidic and aplastidic cryptophytes) and a specific probe for the aplastidic CRY1 lineage [100]. In addition, we also enumerated plastidic (photosynthetic, CryptoB positive) and aplastidic cryptophytes (non-photosynthetic, CryptoB positive). CARD-FISH results revealed an even higher abundance of aplastidic cryptophytes than hinted at by sequence data (Fig. 5).

The aplastidic CRY1 lineage is well-represented pre-bloom (ca. 10%) and increases gradually, reaching an abundance of 20–30% of total eukaryotes in the epilimnion. More pronounced increases for CRY1 were observed in the hypolimnion, in line with 18S rRNA sequences, where they constituted ca. 40–50% of total eukaryotes (Fig. 5). Plastidic cryptophytes, on the other hand, are practically absent before the bloom, start to increase only around day 4 and reach max. 40% of total eukaryotes at day 14. In contrast, aplastidic cryptophytes are already the most prominent HNF, accounting for ca. 70% of total eukaryotes at day 1, and were highly abundant throughout the entire spring bloom period. Moreover, during plastidic cryptophytes' bloom, only small abundance reductions are visible for aplastidic cryptophytes, which nearly always constitute the most dominant eukaryotes. This also suggests that the population dynamics of aplastidic cryptophytes are relatively independent of plastidic cryptophytes (or other algae) as they are bacterivores [100, 101]. The identity of these aplastidic cryptophytes remains a mystery as they are indistinguishable from their plastidic counterparts by 18S rRNA gene sequences. Their dominance during the algal spring bloom, which has so far been perceived as a primarily

phototrophic event, is unexpected and might cause a paradigm shift in understanding the role of heterotrophic nanoflagellates during the spring bloom.

Concluding remarks

The multiple data sources at hand, the combination of different methods, and the identification of major participants allowed us to reconstruct the sequence of events that play out in the theater of the spring bloom in far greater detail than was possible before (Fig. 6). The integration of microscopic and metagenomic data facilitated the discovery of the most abundant players in the spring bloom of the Řimov reservoir. We identified multiple successional patterns across diverse and abundant prokaryotic groups. While Caulobacteraceae and some genera of Crocinitomicaceae and Flavobacteriaceae seem to be bloom specialists, '*Ca. Fonsibacter*' and '*Ca. Nanopelagicus*' seem to avoid the phytoplankton spring bloom (Figs. 5 and 6). Other genera like *Limnohabitans* or '*Ca. Planktophila*' show a more complex pattern with multiple (sometimes even closely related) populations with striking differences in abundance. The high-resolution sampling data also allowed us to compute mortality-corrected doubling times for several abundant planktonic taxa. For some bacteria, their transitions between the epi- and hypolimnion could be linked to temperature, but additional evidence linking metabolically active genes was not generated in this work. Also, in the hypolimnion, some phages reach very high abundances, indicative of a massive phage infection (Fig. 6) in the clear water phase.

The dominance of aplastidic cryptophytes in the spring bloom and their persistent abundances suggest decoupling from their plastidic counterparts and also suggest that there are abundant aplastidic 'cryptic' cryptophytes that comprise a large fraction of unicellular eukaryotes and whose identities are as yet unknown. Parasites and viruses have been considered as potential biological drivers of algal bloom collapse additionally to zooplankton grazing [2]. However, Perkinsozoa appears far more abundant before the bloom and in the hypolimnion, whereas NCLDV can be found at high abundances in both epi- and hypolimnion suggesting that top-down control by NCLDV, in tandem with bottom-up controls of nutrient limitations may be more relevant for bloom collapse than perkinsids. Perkinsids on the other hand show profile abundances similar to aplastidic cryptophytes (including CRY1) (Fig. 6) and may be parasitizing aplastidic cryptophytes as shown before for *Cryptophagus* infecting *Chilomonas* [53].

Applying high-temporal resolution sampling using multiple methods to a single site increased our ability to discern several biological patterns. However, many aspects remain unclear, e.g., specific interactions

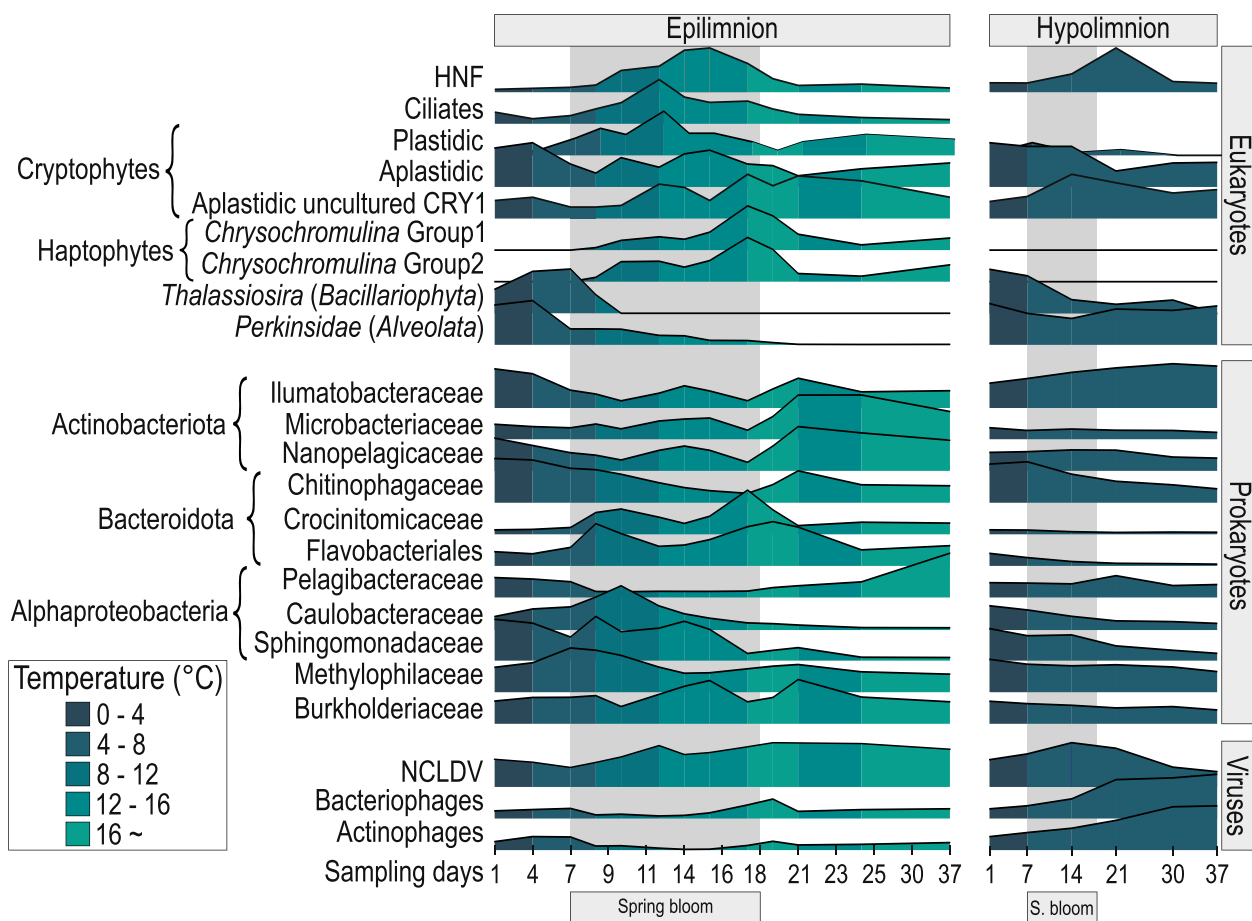


Fig. 6 Schematic overview of the spring bloom in Římov reservoir. Both Epilimnion and Hypolimnion are shown. Abundances (minimum to maximum, scaled from 0 to 1) from multiple methods are shown (microscopy, CARD-FISH, 18S rRNA abundances, and cumulative MAG abundances from all three filter sizes). Microscopy: HNF and Ciliates, CARD-FISH: Plastidic, Aplastidic, and uncultured CRY1 cryptophytes; 18S rRNA abundances: Perkinsidae; MAG abundances: *Chrysochromulina* and *Thalassiosira*; all prokaryotes and viruses. All prokaryotic MAGs and viral/NCLDV contigs were dereplicated. A temperature scale is shown at the bottom left, and the spring bloom period is highlighted in a gray box

between eukaryotes and prokaryotes, and even between eukaryotes themselves, the nature of metabolites supplied by the phytoplankton bloom to the copiotrophic bloom specialists, and specific genomic adaptations that distribute prokaryotes to specific niches or enable them to cooperate, e.g., Black Queen hypothesis [105]. Moreover, the adaptations of viruses (both phages and NCLDV) and the identity of their hosts remain mostly unknown. The annual spring bloom, with its dynamic nature and compressed view of planktonic succession, still withholds many hidden ecological mysteries that will only be revealed by future work.

Material and methods

Study site and sampling

This study was conducted in the canyon-shaped meso-eutrophic Římov Reservoir (Czech Republic; 48° 50' 56" N, 14° 29' 26" E; 470 m a.s.l.; area 2.06 km²; volume

34.5 × 10⁶ m³; length 13.5 km; max. depth 43 m; mean depth 16.5 m; mean retention time 77 days; dimictic) [106]. The reservoir was built as a drinking water reservoir by impounding River Malše, the main reservoir tributary, accounting for 90% of the water inflow. Římov Reservoir is also part of the Czech Long-Term Ecological Research network [107]. Water samples were collected from April 3 to May 9, 2018, 3 times a week except for the last week of the study period covered with a weekly sampling. Zooplankton composition was analyzed at weekly intervals.

Vertical profile of different physicochemical factors

A multi-wavelength submersible fluorescence probe (FluoroProbe, bbe-Moldaenke, Kiel, Germany) was employed to measure chlorophyll-*a* concentrations in a discrete vertical profile at ~0.2 m intervals down to the bottom. According to the specific fluorescence spectra

of distinct phytoplankton groups, the probe permits differentiation of cyanobacteria, chromophytes + dinoflagellates (a mixed group with diatoms frequently most important), cryptophytes, and Chlorophyta in mixed natural populations [108]. A submersible multiparametric probe (YSI EXO 2, Yellow Springs Instruments, Yellow Springs, USA) was deployed to measure detailed vertical profiles of temperature, pH, oxygen concentrations, and conductivity (Additional file 1: Figure S16).

Fixed depths of 0.5 m and 30 m were chosen for further analyses and measurements. Samples (20 l) were collected above the deepest point at the dam using a Friedinger sampler, transported to the laboratory within 30 min, and split into subsamples for microbial enumeration, CARD-FISH, metagenomic filtration, and chemical analyses. Samples for dissolved silica (DSi) and dissolved reactive phosphorus (DRP) assessments were filtered through filters of 0.4- μm porosity (Macherey Nagel GF-5, Macherey Nagel, Düren, Germany) in the laboratory. DSi and SRP were determined spectrophotometrically according to [109, 110], respectively. Total phosphorus (TP) and concentrations of NO_3^- -N were analyzed as described before [111, 112].

Phytoplankton samples were preserved with Lugol's solution and stored in the dark. Species were enumerated employing the Utermöhl method and an inverted microscope (Olympus IX 71) [113]. The mean algal cell dimensions were obtained for biovolume calculation using the approximation of cell morphology to regular geometric shapes [114].

Bacterial, protistan, and VLP counts; protistan bacterivory rate; and CARD-FISH analyses

Subsamples of 20 ml were fixed with formaldehyde (2% final concentration), stored at 4 °C, and processed within 24 h after sampling. Total bacterial abundance was quantified via flow cytometry in samples stained with the fluorochrome SYBRgreen (Molecular Probes) using a CytoFLEX S flow cytometer (Beckman Coulter), equipped with a blue laser and bandpass filters 525/40 and 690/50. Cell sizing (>300 cells per sample) based on measuring cell width and length parameters in 4',6-diamidino-2-phenylindole (DAPI)-stained preparations [115] was conducted by using a semiautomatic image analysis system (NIS-Elements 3.0, Laboratory Imaging, Prague). Bacterial biomass was calculated based on volumetric formulas and volume-to-carbon conversion factors [115]. Details of the groups examined are provided in Additional file 2: Table S1.

Duplicates of Lugol–formaldehyde–thiosulfate fixed subsamples [31] of 5–10 ml (HNF) and 15–45 ml (ciliates) were stained with DAPI, filtered onto a black 1.0- μm pore-size filter (Osmonic, Livermore, CA), and HNF

and ciliate abundances were determined via epifluorescence microscopy as described elsewhere [6]. Protistan bacterivory was estimated using fluorescently labeled bacteria (FLB) [116] prepared from a mixture of two *Limnohabitans* strains and one strain of the PnecC lineage of *Polynucleobacter* as detailed before [117]. The sizes of FLB reassembled the typical size class distribution of the reservoir bacterioplankton with a mean cell size of around 0.065 μm^3 . HNF and ciliate FLB uptake rates were determined in short-term FLB direct-uptake experiments by inspecting protist cells [118], with ciliates being determined to species, genus, or morphotype level (for details, see [24, 117]). To estimate total protistan grazing, we multiplied the average uptake rates of HNF and ciliates by their in situ abundance.

To quantify virus-like particles, 1 ml from each water sample was fixed with glutaraldehyde (1% final concentration) for 10 min, shock-frozen in liquid nitrogen, and stored at –80 °C. Enumerations of VLP were done using an Influx V-GS cell sorter (Becton Dickinson) as previously described [119].

CARD-FISH for prokaryotes

Formaldehyde-fixed subsamples (5 ml) for CARD-FISH were filtered onto 0.2- μm pore-sized polycarbonate filters (Millipore, Merck, Darmstadt, DE) and stored at –20 °C. CARD-FISH was conducted as previously described [120] using fluorescein-labeled tyramides and the following HRP-labeled oligonucleotide probes: AcI-852 [121] targeting the order 'Ca. Nanopelagicales' (acI Actinobacteria), Npel-2669 [32] targeting the genus 'Ca. Nanopelagicus' (acI-B1), MET1217 [122] targeting the family Methylophilaceae, LD28-1017 [123] targeting the genus 'Ca. Methylopumilus', and LD12-115 [124] targeting the genus 'Ca. Fonsibacter' (LD12). Filters were counterstained with DAPI and analyzed by epifluorescence microscopy (Zeiss Imager.Z2, Carl Zeiss, Oberkochen, DE) with a colibri LED light system. Images were recorded with an Axiocam 506 (Carl Zeiss, Oberkochen, DE) and analyzed with the software ACME-tool (www.technobiology.ch) as outlined before [16].

CARD-FISH for eukaryotes

Water samples fixed with Lugol-Formaldehyde-Thiosulfate [125] were filtered on polycarbonate filters (pore size 0.8 μm , Millipore) within 24 h of fixation, and filters were stored at –20 °C until further processing. Two oligonucleotide probes were used in the present study targeting all cryptophytes (CryptoB) [126] and their monophyletic CRY1 lineage (Cry1-652) [100] with the hybridization conditions described before [127]. CARD-FISH was performed with tyramides [118] labeled with fluorescein according to the method described recently [31].

CARD-FISH preparations were analyzed and enumerated with an epifluorescence microscope (Olympus BX 53) at 1000× magnification under blue/UV excitation. Total microbial eukaryotes were counted simultaneously from the same preparations with the DAPI counter-stain under UV excitation. Images of CARD-FISH targeted cells were captured and processed using a semi-automatic image analysis system (NIS-Elements 3.0, Laboratory Imaging, Prague, Czech Republic).

Zooplankton abundance and community composition

Crustaceans were sampled weekly by vertical hauls using an Apstein plankton net (200 µm mesh size). Two vertical hauls were taken from 5 m to the surface (755 l in total), representing the depth of epilimnion, where usually 90% of the crustaceans from the whole water column are accumulated [128]. Rotifers were sampled from the uppermost 5 m of the water column using a plastic tube of the appropriate length. Forty liters of the sampled water were subsequently concentrated using a 30-µm net. Zooplankton samples were preserved in 4% formaldehyde, and the abundance of the main zooplankton species was determined microscopically [129]. The final counts were normalized by the total amount of filtered water (Additional file 1: Figure S17).

Metagenomic filtration and DNA extraction

In total, fifty-seven DNA samples were collected from epilimnion (0.5 m, $n=39$) and hypolimnion (30 m, $n=18$) using three different filter-pore sizes (5 µm, 0.8 µm, and 0.22 µm) and two types of filtration procedures (positive pressure and gravity filtration). More details about the samples are available in Additional file 2, Table S1. Two of these metagenomic datasets have been published previously [20]. The rest were generated in this study.

Positive pressure filtration

Water samples (ca. 20 L) from both depths were sequentially filtered through a 20-µm mesh plankton net to remove larger organisms, followed by 5-µm and 0.22-µm polycarbonate membrane filters using a peristaltic pump until filters were clogged. A 5-µm and 0.22-µm fractions were collected and stored at -80°C until DNA extraction.

Gravity filtration

Water samples (ca. 20 L) from both depths were gravity filtered through a 20-µm-mesh plankton net, followed by 5-µm and 0.8-µm filters, as described before [100]. DNA was extracted only from the 0.8-µm filters.

All filters were cut into small pieces ($\cong 3\text{--}5$ mm) using sterile scissors and processed for DNA extraction using the ZR Soil Microbe DNA MiniPrep kit (Zymo Research, Irvine, CA, USA), according to the manufacturer's instructions.

Preprocessing and assembly of metagenomic datasets

DNA samples were sequenced using Illumina Novaseq 6000 platform (Novogene, Hong Kong, China). Low-quality bases, reads, and adaptors were trimmed using the bbmap package (<https://sourceforge.net/projects/bbmap/>). Raw reads were interleaved, and quality trimmed by reformat.sh, followed by bbdutk.sh (Phred score = 18). Additionally, bbdutk.sh was used to remove any adapter or PhiX and p-Fosil2 contamination. A final check using bbmerge.sh was done to ensure the quality threshold for assembly of the contigs. MEGAHIT (v1.1.5) was used with default settings to assemble the preprocessed reads (k-mer sizes: 49,69,89,109,129,149) [130].

16S and 18S rRNA gene abundance-based taxonomic classification

Each metagenome was randomly subsampled (20 million reads) and compared to SILVA v138 database [131] to detect candidates of 16S and 18S rRNA gene reads (e value $1e^{-5}$) using MMSeqs2 [132]. The putative rRNA reads were scanned with ssu-align (<http://eddylab.org/software/ssu-align/>) to identify 16S rRNA and 18S rRNA sequences. Lastly, the taxonomy classification was attributed using blastn [133] against the SILVA database v138. For 18S rRNA classification, sequences originating from organisms known to have thousands of rRNA operons, e.g., Dinoflagellates, Ciliophora, and those from multicellular organisms, i.e., Metazoa, were ignored. For both 18S and 16S rRNA, sequences originating from Chloroplasts, Mitochondria, and Nucleomorphs of Cryptophyceae were excluded from the analysis. The sequences were classified based on 95% identity to the databases.

Recovery of prokaryotic metagenomic assembled genomes

Preprocessed metagenomic datasets were mapped using bbwrap.sh (kfilter = 31 subfilter = 15 maxindel = 80) against the assembled contigs (≥ 3 kb). The contig abundance files were obtained using jgi_summarize_bam_contig_depths [134] and used for the binning process with MetaBAT2 with default parameters [134]. Contigs suspected to be of viral origin by scanning with VIBRANT [135] or ViralRecall [136] were excluded. An additional cleaning step, considering the

homogeneity of the contigs in each bin, was performed in order to assure high-quality bins: Open reading frames from each contig inside the bins were predicted using PRODIGAL v2.6.3 [137], and taxonomy of each gene was assigned using MMseqs2 [132] with Genome Taxonomy DataBase (version r95) (GTDB) [138]. Contigs with more than 30% of genes with no hits or hits to eukaryotes or viruses as well as contigs in which the taxonomy disagreed with the consensus class were removed.

The taxonomy of each bin was compared to the taxonomy of 16S and 23S rRNA gene sequences within the bin. Briefly, rRNA sequences were identified using barrnap with default parameters (<https://github.com/tseemann/barrnap>), extracted, and compared to the SILVA v138 database [131]. Simultaneously, the taxonomy of the whole bin was assigned using the GTDB-Tk [139] toolkit based on the GTDB taxonomy (version r95). Contigs in which 16S or 23S rRNA taxonomic classification did not match with GTDB taxonomy were removed. The resulting bins were evaluated with CheckM v1.0.18 [140], and only those with >40% completeness and <5% contamination were retained for further analysis ($n=2214$). These bins were further dereplicated using dRep (-comp 40 -con 5) [141], resulting in 855 representative prokaryotic bins.

Recovery of phage genomes, dereplication, and host prediction

In order to obtain complete phage genomes, assembled circular contigs >10 kb ($n=1225$) were used for the downstream analysis as described before [20, 58]. The selected contigs were scanned using VIBRANT—default settings [135]. Lastly, an additional step of manual curation using the NCBI Batch CDD server [142] was performed to enhance the accuracy of phage recognition. We recovered 679 bonafide complete phage genomes with this protocol.

We clustered phages by an all-vs-all BLASTN approach, retaining only significant matches ($1e-3$). If, in comparison, two phages had genome coverage of >95% and >95% nucleotide identity, they were considered part of a cluster. Next, we used a single linkage to merge all potential clusters. Finally, we obtained 175 distinct phage genome clusters. Multiple approaches were used to link the recovered phages to a putative host; for example, we searched for photosystem genes in phage genomes that are a hallmark for a cyanobacterial host [143]. Similarly, the *whiB* transcriptional regulator is a marker gene for phages infecting Actinobacteria, and *whiB* for Actinobacteria [58]. Additionally, phages frequently insert in tRNA loci (termed *attB*) in the host genome, and the presence of such a site in the phage genome may imply

a specific association [144]. We used BLASTN [133] for examining such identities between putative *attB* sites and phage genomes as described before (alignment length ≥ 30 bp, $\geq 97\%$ nucleotide identity, $\geq 97\%$ query coverage, $\leq 1e-5$) [144]. We also used CRISPR spacers (detected using minced, available from <https://github.com/ctSkennerton/minced>) from the recovered microbial genomes to query phage genomes using BLASTN with stringent cutoffs (alignment length ≥ 30 bp, $\geq 97\%$ nucleotide identity, $\geq 97\%$ query coverage, $\leq 1e-5$). Finally, we made direct comparisons of phage genomes to host genomes using BLASTN to identify shared nucleotide sequences (alignment length ≥ 30 bp, $\geq 97\%$ nucleotide identity, $\geq 97\%$ query coverage, $\leq 1e-5$) [145].

Recovery of NCLDV genomes and phylogenomics

All contigs longer than 10 kb were scanned with Viral-Recall [136] to identify signatures of putative nucleocytoplasmic large DNA viruses (NCLDVs). For capturing bonafide NCLDVs, we applied a strict threshold of at least five viral hits, at least one marker gene, and a score ≥ 1 . Following this, we manually curated these contigs by examining hits to multiple databases, some contigs were removed. Finally, 3309 were finally retained as bonafide NCLDV contigs. (results in Additional file 8, Table S7).

A collection of previously published NCLDV genomes ($n=1447$) was generated from (1) annotated genomes in GenBank, (2) giant virus MAGs published by Schulz et al. [91], and (3) a highly curated and taxonomically coherent collection of giant viruses compiled by Aylward et al. [90]. All viruses were dereplicated together with dRep [141] (default parameters; completeness and contamination data not used for scoring) resulting in 1245 unique genomes. Dereplication was performed in the same manner for NCLDV contigs retrieved in this study ($n=3309$) resulting in $n=1721$ representative sequences. Information regarding viral clusters and the number of members within each cluster are included in Additional file 8: Table S7. All dereplicated genomes (references and the ones retrieved in this study) were scanned with *ncldv_markersearch* [146] for the presence of 7 giant virus marker genes (GVOGm) previously selected due to their strong phylogenetic signal [90]. We discarded all instances with less than 2 identifiable markers resulting in $n=1240$ reference genomes and $n=153$ viral contigs from this study. Protein markers retrieved from all genomes were individually aligned using Clustal Omega [147], trimmed with BMGE (-t AA -g 0.5 -b 3 -m BLOSUM30 -g 0.5) [148] and concatenated. An initial maximum likelihood (ML) phylogenomic tree was constructed using IQ-TREE v2.1.3 [149] (1000 iterations for ultrafast bootstrapping [150] and SH testing, respectively; best model chosen automatically by ModelFinder

[151]: LG + F + I + G4. Upon inspection of the initial tree followed by manual curation (removal of references from overrepresented groups) a final phylogenomic tree was generated including $n = 940$ reference genomes and without changes to our own set (Additional file 8: Table S7). The final tree was annotated to include labels containing the original genome/contig name and taxonomic classification derived from the study of Aylward et al. [90]. All data associated with the published phylogenomic tree is made available in FigShare (<https://figshare.com/s/5150884f3fbb534c0302>).

Recovery of organellar genomes

Circular contigs (> 30 bp overlap at the end) were compared using MMSeqs2 to a database containing all GTDB proteins (r95) and all UniProt eukaryotic and viral proteins. Additionally, we used metaeuk [152] in easy-predict mode vs the UniProtKB database. The results were manually examined, and those that appeared to be of mitochondrial or chloroplast origin were segregated. Details of organellar genomes recovered in this work are provided in Additional file 9: Table S8.

Recovery of eukaryotic metagenome-assembled genomes

All bins that could not be classified as prokaryotic by GTDB-Tk [139] and CheckM [140] were input to BUSCO [153] in `-auto-lineage` mode to identify candidate eukaryotic MAGs. Once a candidate MAG was identified, we used metaeuk [152] in easy-predict mode vs the UniProtKB [154] and MMETSP [155] databases to identify the taxonomic origin of contigs. Contigs that did not conform to the majority taxonomic consensus within a MAG were discarded. After cleaning the bin, we estimated genome completeness using BUSCO [153] again with specific lineage datasets wherever possible.

Prokaryotic and eukaryotic phylogenomic analyses

Maximum-likelihood (ML) phylogenomic trees were constructed for major bacterial groups using IQ-TREE2 (v.2.1.2) with 1000 iterations of ultrafast bootstrapping [150] and SH testing [149]. The best fitting evolutionary models were automatically determined for each group by activating the ModelFinder [151] option of IQ-TREE. Reference genomes and outgroups were selected from a curated collection of ~100,000 uniformly classified genomes (using gtdb-tk, r95) which were scanned with `hmmsearch` together with the ones recovered in this study for the presence of 120 conserved protein HMM markers [138]. Amino-acid sequences for each of the 120 makers were aligned using PASTA [156] and trimmed by BMGE (parameters: `-t AA -g 0.5 -b 3 -m BLOSUM30`) [148]. Individually trimmed alignments were concatenated using `catfasta2phym.pl` (<https://github.com/>

`nylander/catfasta2phym.pl`) and finally used as input for IQ-TREE. All details about selected models and alignment statistics are available in Additional file 10: Table S9.

BUSCO [153] was used to identify single-copy genes in recovered eukaryotic MAGs, available transcriptomes, or genomes. For *Thalassiosira* and *Chrysochromulina* phylogenomic trees, 71 and 68 conserved single copy markers were used, respectively. Only those genomes/transcriptomes with at least 70% of these markers were used. Alignments, trimming, concatenation of aligned markers, and construction of phylogenomic trees were performed as described above for prokaryotic genomes.

Fragment recruitment, estimation of replication, and duplication times

From each metagenome, 20 million quality-filtered reads were mapped against the recovered genomes using RazerS 3 (`-no-gaps, -max-hits 1,000,000`) [157]. All alignments were filtered to at least 50 bp and 95% nucleotide sequence identity. The obtained number of hits were used to compute coverage per Gbp values, offering normalized abundances comparable for different MAGs and metagenomes. Accession numbers of the publicly available metagenomic datasets used for the recruitment shown in Fig. 4D are listed in Additional file 11: Table S10.

We used GRiD for estimating the rate of bacterial replication (GRiD multiplex module with default options) [158]. According to the recommendation, we used only forward reads from each metagenome.

For computing doubling times using gRodon, only dereplicated MAGs were used. Additionally, only those MAGs that had a minimum of 15 ribosomal genes and 500 predicted genes were retained. The doubling times for each genome were estimated with the default parameters using gRodon with the appropriate sample temperature as recommended [69]. The same MAGs (considered high-quality MAGs for doubling time estimations) were further used for computing doubling times using metagenomic abundances.

Estimation of doubling times as performed here (both using CARD-FISH or metagenomic abundances) was done according to the following assumptions: Firstly, estimates of average mortality by grazing (24%) and by viruses (31%) were obtained from prior observations from a relatively large sample of 19 lakes [68]. Taken together, a total of 55% mortality was assumed as a combined effect. Additionally, for all groups examined, only those adjacent time points that showed a change of at least 25% were considered to identify the fastest doubling times which may seem at odds with the constant mortality assumption but can also be considered as ignoring the initial lag phase during exponential growth [4]. As there is usually a strong dependence of sampling site and

season upon these values, not all assumptions may be always equally valid. Hence, the doubling times obtained here must be considered rough estimates. The observed in situ doubling time (in hours) was computed as described before [4] for each pair of adjacent time points that showed at least a 25% increase. Briefly, growth rate $r = 1/(t_1 - t_0) * \ln(N_2/N_1)$, where t_0 and t_1 are initial time points (hours), and N_1 and N_2 are abundance values at t_0 and t_1 , respectively. The doubling time may be computed as $\ln(2)/r$. The final observed in situ doubling times as shown in Table 1 are expressed as an average of all doubling times. The doubling times in absence of mortality (i.e., estimated doubling time) was obtained by correcting for the mortality estimate (in this case 55%) and averaged as shown in Table 1.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-022-01451-4>.

Additional file 1: Figure S1. Time course of different features of the spring bloom in the hypolimnion. (A) Chlorophyll a concentrations, temperature, heterotrophic nanoflagellates (HNF), viral like particles (VPL), (B) total phosphorus (TP), NH₄-N and dissolved nitrogen (DN), (C) silica, dissolved organic carbon (DOC) and dissolved reactive phosphorus (DRP). **Figure S2.** Heterotrophic nanoflagellates and ciliate counts, bacterial ingestion, grazing rate estimations and turnover of observed ciliates in epilimnion. (A) HNF (10³ ml), Bacteria (10⁶ ml), and ciliate (per ml) abundances, (B) Total grazing rates of heterotrophic nanoflagellates (HNF) and ciliates, expressed as number of ingested bacteria/day, (C) Bacteria ingested by HNF (per hour) and ciliates (per hour). A temperature scale is also shown (D) Counts (per ml) of four ciliates identifiable by morphology. The grey region indicates the spring bloom as identified by chlorophyll content. **Figure S3.** (A) Completeness and contamination estimates for 2214 genomic bins recovered in this study. Three categories of genomic bins were defined based on completeness estimated by CheckM: high quality genomic bins (completeness $\geq 90\%$) representing 22.6% ($n = 502$) of total, medium quality genomic bins (completeness $\geq 70-89\%$) amounting to 26.8% ($n = 594$), and partial genomes (completeness 40-70%) the remaining 50% (1118). Bins with contamination $> 5\%$ and completeness $< 40\%$ were not analyzed. The histogram parallel to the x axis shows the percentage distribution of bins according to level of completeness while the one parallel to the y axis indicates the percentage distribution of genomes by contamination. (B) Percent distribution of bins according to tRNA gene copy content. (C) Relationship between gene coding density and estimated genome length. (D) Number of MAGs recovered from each dataset ($n=57$). (E) MAG percentage distribution across phyla. **Figure S4.** Dereplicated bacteriophages genome abundances of 0.22 μ m filter (Cov/Gb) normalized by Z-score in epi and hypolimnion ($n=175$). The rows are clustered by relative abundance by average linkage (with Spearman Rank correlation method). Inferred host taxonomy is shown at the right wherever applicable. **Figure S5.** Pruned phylogenomic tree of *Ca. Fonsibacter* (Pelagibacteriaceae) metagenome assembled genomes. Relative abundance of selected MAGs during the spring bloom in the small filter (0.22 μ m) are shown on the right side. Each row was normalized by Z-score and clustered by average linkage (with Spearman Rank correlation method). Ultrafast bootstrap values are shown at each node. **Figure S6.** Pruned phylogenomic tree of *Ca. Planktophila* (Actinobacteriota) metagenome assembled genomes. Relative abundance of selected MAGs during the spring bloom in the small filter (0.22 μ m) are shown on the right side. Each row was normalized by Z-score and clustered

by average linkage (with Spearman Rank correlation method). Ultrafast bootstrap values are shown at each node. **Figure S7.** Pruned phylogenomic tree of Limnohabitans (Gammaproteobacteria) metagenome assembled genomes. Relative abundance of selected MAGs during the spring bloom in the small filter (0.22 μ m) are shown on the right side. Each row was normalized by Z-score and clustered by average linkage (with Spearman Rank correlation method). Ultrafast bootstrap values are shown at each node. **Figure S8.** Dereplicated Alphaproteobacteria (Proteobacteria) MAG abundances of 0.22 μ m filter (Cov/Gb) normalized by Z-score in epi and hypolimnion ($n=67$). The rows are clustered by taxonomy and relative abundance by average linkage (with Spearman Rank correlation method). **Figure S9.** Dereplicated Bacteroidota MAG abundances of 0.22 μ m filter (Cov/Gb) normalized by Z-score in epi and hypolimnion ($n=238$). The rows are clustered by taxonomy and relative abundance by average linkage (with Spearman Rank correlation method). **Figure S10.** Dereplicated Gammaproteobacteria (Proteobacteria) MAG abundances of 0.22 μ m filter (Cov/Gb) normalized by Z-score in epi and hypolimnion ($n=198$). The rows are clustered by taxonomy and relative abundance by average linkage (with Spearman Rank correlation method). **Figure S11.** Dereplicated Planctomycetota and Chloroflexota MAG abundances of 0.22 μ m filter (Cov/Gb) normalized by Z-score in epi and hypolimnion ($n=28$ and 7, respectively). The rows are clustered by taxonomy and relative abundance by average linkage (with Spearman Rank correlation method). **Figure S12.** Dereplicated Verrucomicrobiota MAG abundances of 0.22 μ m filter (Cov/Gb) normalized by Z-score in epi and hypolimnion ($n=52$). The rows are clustered by taxonomy and relative abundance by average linkage (with Spearman Rank correlation method). **Figure S13.** Average Nucleotide Identity (ANI) matrix of (A) Chrysochromulina bins, (B) Thalassiosira bins. The color key and a histogram of frequency of identity is shown on the left side. Black squares inside the matrices highlight the groups formed on each matrix. **Figure S14.** Phylogenomic tree of Nucleocytoviricota based on 7 conserved protein markers. Clades were collapsed at Family level while colored rectangles highlight affiliations to different Orders. Family identifiers are provided on the right of each collapsed clade (triangles) followed by square brackets containing the total number of members within the clade (references + genomes from this study) and by the number of dereplicated genomes recovered here from Epilimnion (E) or Hypolimnion (H) if they are present at all. Ultrafast bootstrap values (UBF) are defined by full (90-100) and empty (80-89) circles. UBF values below 80 are not represented. The root was defined between classes Pokkesviricetes and Megaviricetes as in a previous study [90]. **Figure S15.** Dereplicated NCLDV contigs cumulative abundances of 0.22, 0.8 and 5 μ m filters (Cov/Gb) normalized by Z-score in epi and hypolimnion ($n=1721$). The rows are clustered relative abundance by average linkage (with Spearman Rank correlation method). **Figure S16.** Limnological parameters in Rimov reservoir along the spring bloom (03 April ~ 09 May 2018). **Figure S17.** Total counts of zooplankton during the spring bloom. Rotifera (top) and Crustacea (bottom) abundance. For more details of the sampling scheme please check material and methods. Gray square shown in the background mark the Spring bloom during the time period.

Additional file 2: Table S1. Microscopy data - Phytoplankton counts.

Additional file 3: Table S2. General statistics and information about the bacterial MAGs.

Additional file 4: Table S3. MAGs containing rhodopsins.

Additional file 5: Table S4. Bacteriophage general information.

Additional file 6: Table S5. FISH counts in epilimnion and hypolimnion.

Additional file 7: Table S6. Eukaryotic MAGs information.

Additional file 8: Table S7. NCLDV abundances and general information.

Additional file 9: Table S8. Details of organellar genomes recovered.

Additional file 10: Table S9. Characteristics of phylogenomic trees constructed using IQTREE2.

Additional file 11: Table S10. Public datasets used in Fig. 4, pane D.

Acknowledgements

We would like to thank Radka Malá for excellent technical assistance.

Authors' contributions

R.G., V.S.K., and K.S. conceived the work. V.S.K., P.-A.B., C.M.C., and R.G. performed the metagenomic data analyses. M.M.S. and I.M. performed the CARD-FISH analyses for prokaryotes and eukaryotes, respectively. P.R. and P.Z. performed the fieldwork, sampling, and phytoplankton biomass analyses. J.S. enumerated all zooplankton. K.S. quantified the microbial eukaryotes and performed the FLB tracer uptake experiments. All authors assisted in sampling and sample processing during the entire study period. V.S.K. and R.G. wrote the initial draft, and the authors read and approved the final manuscript.

Funding

V.S.K. was supported by the research grant 116/2019/P (Grant Agency of the University of South Bohemia in České Budějovice, 2019–2021), 21-21990S and 19-23261S (Grant Agency of the Czech Republic). P.-A.B., I.M., and R.G. were supported by the research grant 20-12496X (Grant Agency of the Czech Republic). M.M.S. was supported by research grants 19-23469S and 22-03662S (Grant Agency of the Czech Republic). C.M.C. was supported by research grants 21-21990S and 22-03662S (Grant Agency of the Czech Republic). P.Z. and P.R. were supported by the grant CZ.02.1.01/0.0/0.0/16_025/0007417 (ERDF/ESF). A.-S.A. was supported by the Ambizione grant PZ00P3_193240 (Swiss National Science Foundation). J.J. was supported by the research grant 19-23261S. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

All sequence data generated in this work has been deposited to EBI ENA and is available with the bioproject identifier PRJEB52406. Additionally, bacterial and eukaryotic metagenome-assembled genomes, complete phage genomes, NCLDV contigs, alignment files, and trees are also available in figshare (<https://figshare.com/s/5150884f3fbb534c0302>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute of Hydrobiology, Biology Centre CAS, Na Sádkách 7, 370 05 České Budějovice, Czech Republic. ²Faculty of Science, University of South Bohemia, Braníšovská 31, 370 05 České Budějovice, Czech Republic. ³Present address: Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden. ⁴Present address: Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala 750 07, Sweden. ⁵Limnological Station, Microbial Evogenomics Lab (MIEL), University of Zurich, Kilchberg, Switzerland.

Received: 23 August 2022 Accepted: 16 December 2022

Published online: 26 January 2023

References

- Sommer U, Gliwicz ZM, Lampert W, Duncan A. The PEG-model of seasonal succession of planktonic events in fresh waters. *Arch Hydrobiol.* 1986;106:433–71.
- Sommer U, Adrian R, De Senerpont Domis L, Elser JJ, Gaedke U, Ibelings B, et al. Beyond the Plankton Ecology Group (PEG) model: mechanisms driving plankton succession. *Annu Rev Ecol Syst.* 2012;43:429–48.
- 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. American Society of Limnology and Oceanography Inc.
- Zeder M, Peter S, Shabarova T, Pernthaler J. A small population of planktonic Flavobacteria with disproportionately high growth during the spring phytoplankton bloom in a prealpine lake. *Environ Microbiol.* 2009;11:2676–86.
- 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 Wiley.* 2012;14:794–806.
- 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.
- Tirok K, Gaedke U. Spring weather determines the relative importance of ciliates, rotifers and crustaceans for the initiation of the clear-water phase in a large, deep lake. *J Plankton Res Oxford Academic.* 2006;28:361–73.
- Frenken T, Velthuis M, de Senerpont Domis LN, Stephan S, Aben R, Kosten S, et al. Warming accelerates termination of a phytoplankton spring bloom by fungal parasites. *Glob Chang Biol.* 2016;22:299–309.
- Pradeep Ram AS, Mauduit M-E, Colombet J, Perriere F, Thouvenot A, Sime-Ngando T. Top-down controls of bacterial metabolism: a case study from a temperate freshwater lake ecosystem. *Microorganisms.* 2022;10. <https://doi.org/10.3390/microorganisms10040715>.
- Huppert A, Blasius B, Stone L. A model of phytoplankton blooms. *Am Nat.* 2002;159:156–71.
- Moreira D, López-García P. Time series are critical to understand microbial plankton diversity and ecology. *Mol Ecol.* 2019;28:920–2.
- Linz AM, Cray BC, Shade A, Owens S, Gilbert JA, Knight R, et al. Bacterial community composition and dynamics spanning five years in freshwater bog lakes. *mSphere.* 2017;2. <https://doi.org/10.1128/mSphere.00169-17>.
- Tromas N, Fortin N, Bedrani L, Terrat Y, Cardoso P, Bird D, et al. Characterising and predicting cyanobacterial blooms in an 8-year amplicon sequencing time course. *ISME J.* 2017;11:1746–63.
- Eiler A, Heinrich F, Bertilsson S. Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J nature.com.* 2012;6:330–42 nature.com.
- Mangot J-F, Domaizon I, Taib N, Marouni N, Duffaud E, Bronner G, et al. Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. *Environ Microbiol.* 2013;15:1745–58.
- Salcher MM, Pernthaler J, Posch T. Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria "that rule the waves" (LD12). *ISME J. nature.com;* 2011; Available from: <https://www.nature.com/articles/ismej20118>
- Salcher MM, Neuenschwander SM, Posch T, Pernthaler J. The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. *ISME J.* 2015;9:2442–53.
- Neuenschwander SM, Pernthaler J, Posch T, Salcher MM. Seasonal growth potential of rare lake water bacteria suggest their disproportional contribution to carbon fluxes. *Environ Microbiol.* 2015;17:781–95.
- Linz AM, He S, Stevens SLR, Anantharaman K, Rohwer RR, Malmstrom RR, et al. Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. *PeerJ.* 2018;6:e6075.
- 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.
- 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; <https://doi.org/10.1038/s41564-020-00852-1>.
- Lopez-Garcia P, Reboul G, David G, Jardillier L, Annenkova N, Bertolino P, et al. Environmental drivers of plankton and sediment microbial communities along latitudinal and vertical gradients in the deepest freshwater lake (Baikal, Southern Siberia). 2021. p. EGU21-8920.
- Hamilton M, Hennon GMM, Morales R, Needoba J, Peterson TD, Schatz M, et al. Dynamics of Teleaulax-like cryptophytes during the decline of a red water bloom in the Columbia River Estuary. *J Plankton Res Oxford Academic.* 2017;39:589–99.

24. Posch T, Eugster B, Pomati F, Pernthaler J, Pitsch G, Eckert EM. Network of interactions between ciliates and phytoplankton during spring. *Front Microbiol.* 2015;6:1289.
25. Pitsch G, Bruni EP, Forster D, Qu Z, Sonntag B, Stoeck T, et al. Seasonality of planktonic freshwater ciliates: are analyses based on V9 regions of the 18S rRNA gene correlated with morphospecies counts? *Front Microbiol.* 2019;10:248.
26. Eckert EM, Baumgartner M, Huber IM, Pernthaler J. Grazing resistant freshwater bacteria profit from chitin and cell-wall-derived organic carbon. *Environ Microbiol Wiley.* 2013;15:2019–30.
27. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science science. sciencemag.org.* 2012;336:608–11.
28. Francis TB, Bartosik D, Sura T, Sichert A, Hehemann JH, Markert S, et al. Changing expression patterns of TonB-dependent transporters suggest shifts in polysaccharide consumption over the course of a spring phytoplankton bloom. *ISME J nature.com.* 2021;15:2336–50 nature.com.
29. Hahnke RL, Bennke CM, Fuchs BM, Mann AJ, Rhie E, Teeling H, et al. Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. *Environ Microbiol Wiley.* 2015;17:3515–26.
30. Alejandre-Colomo C, Harder J, Fuchs BM, Rosselló-Móra R, Amann R. High-throughput cultivation of heterotrophic bacteria during a spring phytoplankton bloom in the North Sea. *Syst Appl Microbiol.* 2020;43:126066 Elsevier.
31. Piwosz K, Mukherjee I, Salcher MM, Grujić V, Šimek K. CARD-FISH in the sequencing Era: opening a new universe of protistan ecology. *Front Microbiol.* 2021;12:640066.
32. Neuenschwander SM, Ghai R, Pernthaler J, Salcher MM. Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J.* 2018;12:185–98.
33. Salcher MM, Schaeffle 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.
34. Šimek K, Mukherjee I, Szöke-Nagy T, Haber M, Salcher MM, Ghai R. Cryptic and ubiquitous aplastidic cryptophytes are key freshwater flagellated bacterivores. *ISME J.* 2022; <https://doi.org/10.1038/s41396-022-01326-4>.
35. Raven JA, Wollenweber B, Handley LL. A comparison of ammonium and nitrate as nitrogen sources for photolithotrophs. *New Phytol Wiley.* 1992;121:19–32.
36. Glibert PM. Interactions of top-down and bottom-up control in planktonic nitrogen cycling. In: Tamminen T, Kuosa H, editors. *Eutrophication in Planktonic Ecosystems: Food Web Dynamics and Elemental Cycling: Proceedings of the Fourth International PELAG Symposium, held in Helsinki, Finland, 26–30 August 1996.* Dordrecht: Springer Netherlands; 1998. p. 1–12.
37. Saba GK, Steinberg DK, Bronk DA. The relative importance of sloppy feeding, excretion, and fecal pellet leaching in the release of dissolved carbon and nitrogen by *Acartia tonsa* copepods. *J Exp Mar Bio Ecol.* 2011;404:47–56.
38. Valdés VP, Fernandez C, Molina V, Escribano R, Joux F. Dissolved compounds excreted by copepods reshape the active marine bacterioplankton community composition. *Front Mar Sci.* 2017. <https://doi.org/10.3389/fmars.2017.00343>.
39. Müller H, Schöne A, Pinto-Coelho RM, Schweizer A, Weisse T. Seasonal succession of ciliates in lake constance. *Microb Ecol.* 1991;21:119–38.
40. Šimek K, Jürgens K, Nedoma J, Comerma M, Armengol J. Ecological role and bacterial grazing of *Halteria* spp.: small freshwater oligotrichs as dominant pelagic ciliate bacterivores. *Aquat Microb Ecol Inter-Research Science Center.* 2000;22:43–56.
41. 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.
42. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev.* 2011;75:14–49.
43. 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.
44. Andrei A-Ş, Salcher MM, Mehrshad M, Rychtecký P, Znachor P, Ghai R. Niche-directed evolution modulates genome architecture in freshwater Planctomycetes. *ISME J.* 2019;13:1056–71.
45. Moss JA, Xiao J, Dungan CF, Reece KS. Description of *Perkinsus beihaiensis* n. sp., a new Perkinsus sp. parasite in oysters of Southern China. *J Eukaryot Microbiol.* 2008;55:117–30.
46. Jeon BS, Park MG. *Tuberlatum coatsi* gen. n., sp. n. (Alveolata, Perkinsozoa), a new parasitoid with short germ tubes infecting marine dinoflagellates. *Protist.* 2019;82–103. <https://doi.org/10.1016/j.protis.2018.12.003>
47. Jobard M, Wawrzyniak I, Bronner G, Marie D, Vellet A, Sime-Ngando T, et al. Freshwater Perkinsea: diversity, ecology and genomic information. *J Plankton Res Oxford Academic.* 2019;42:3–17.
48. Karlsbakk E, Nystøyl CF, Plarre H, Nyland A. A novel protist parasite, *Salmoxcellia vastator* n. gen., n. sp. (Xcellidae, Perkinsozoa), infecting farmed salmonids in Norway. *Parasit Vectors.* 2021;14:431.
49. Mangot J-F, Debroas D, Domaizon I. Perkinsozoa, a well-known marine protozoan flagellate parasite group, newly identified in lacustrine systems: a review. *Hydrobiologia.* 2011;659:37–48.
50. Lepère C, Domaizon I, Debroas D. Unexpected importance of potential parasites in the composition of the freshwater small-eukaryote community. *Appl Environ Microbiol American Society for Microbiology.* 2008;74:2940–9.
51. Mukherjee I, Hodoki Y, Nakano S. Seasonal dynamics of heterotrophic and plastidic protists in the water column of Lake Biwa, Japan. *Aquat Microb Ecol. int-res.com;* 2017; Available from: <https://www.int-res.com/abstracts/ame/v80/n2/p123-137/>
52. Lepère C, Masquelier S, Mangot JF, Debroas D, Domaizon I. Vertical structure of small eukaryotes in three lakes that differ by their trophic status: a quantitative approach. *ISME J.* 2010;4:1509–19. Springer Science and Business Media LLC.
53. Brugerolle G. *Cryptophagus subtilis*: a new parasite of cryptophytes affiliated with the Perkinsozoa lineage. *Eur J Protistol Elsevier.* 2002;37:379–90.
54. Vincent F, Sheyn U, Porat Z, Schatz D, Vardi A. Visualizing active viral infection reveals diverse cell fates in synchronized algal bloom demise. *Proc Natl Acad Sci U S A.* 2021;118. <https://doi.org/10.1073/pnas.2021586118>
55. Nagata T, Inoue K. Rhodopsins at a glance. *J Cell Sci.* 2021;134. <https://doi.org/10.1242/jcs.258989>
56. Fuhrman JA, Schwalbach MS, Stingl U. Proteorhodopsins: an array of physiological roles? *Nat Rev Microbiol.* 2008;6:488–94.
57. Man D, Wang W, Sabehi G, Aravind L, Post AF, Massana R, et al. Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* 2003;22:1725–31.
58. Ghai R, Mehrshad M, Mizuno CM, Rodriguez-Valera F. Metagenomic recovery of phage genomes of uncultured freshwater actinobacteria. *ISME J.* 2017;11:304–8.
59. Mendl K, Chen H, Parks DH, Lobb B, Hug LA, Doxey AC. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res Oxford Academic.* 2019;47:4442–8.
60. Sawa N, Tatsuke T, Ogawa A, Hirokawa Y, Osanai T, Hanai T. Modification of carbon metabolism in *Synechococcus elongatus* PCC 7942 by cyanophage-derived sigma factors for bioproduction improvement. *J Biosci Bioeng.* 2019;127:256–64.
61. Kasalický V, Jezbera J, Hahn MW, Šimek K. The diversity of the Limnohabitans genus, an important group of freshwater bacterioplankton, by characterization of 35 isolated strains. *PLoS ONE.* 2013;8: e58209.
62. Jezberová J, Jezbera J, Znachor P, Nedoma J, Kasalický V, Šimek K. The limnohabitans genus harbors generalistic and opportunistic subtypes: evidence from spatiotemporal succession in a canyon-shaped reservoir. *Appl Environ Microbiol.* 2017;83. <https://doi.org/10.1128/AEM.01530-17>
63. Salcher MM, Pernthaler J, Posch T. Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria “that rule the waves” (LD12). *ISME J Nature Publishing Group.* 2011;5:1242–52.
64. Heinrich F, Eiler A, Bertilsson S. Seasonality and environmental control of freshwater SAR11 (LD12) in a temperate lake (Lake Erken, Sweden). *Aquat Microb Ecol Inter-Research Science Center.* 2013;70:33–44.

65. Henson MW, Lanclus VC, Faircloth BC, Thrash JC. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *ISME J*. 2018;12:1846–60.
66. Kim S, Kang I, Lee J-W, Jeon CO, Giovannoni SJ, Cho J-C. Heme auxotrophy in abundant aquatic microbial lineages. *Proc Natl Acad Sci U S A*. 2021;118. <https://doi.org/10.1073/pnas.2102750118>
67. Kirchman DL. Growth rates of microbes in the oceans. *Ann Rev Mar Sci*. 2016;8:285–309.
68. Ram ASP, Palesse S, Colombet J, Thouvenot A, Sime-Ngando T. The relative importance of viral lysis and nanoflagellate grazing for prokaryote mortality in temperate lakes. *Freshw Biol Wiley*. 2014;59:300–11.
69. Weissman JL, Hou S, Fuhrman JA. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc Natl Acad Sci U S A National Academy of Sciences*; 2021;118. Available from: <https://www.pnas.org/content/118/12/e2016810118/tab-article-info>. [Cited 2021 May 28].
70. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome streamlining in a cosmopolitan oceanic bacterium. *Science*. 2005;309:1242–5.
71. Delmont TO, Gaia M, Hinsinger DD, Frémont P, Vanni C, Fernandez-Guerra A, et al. Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genomics*. 2022;2: 100123.
72. Dahl E, Bagøien E, Edvardsen B, Stenseth NC. The dynamics of *Chrysochromulina* species in the Skagerrak in relation to environmental conditions. *J Sea Res*. 2005;54:15–24.
73. Seoane S, Eikrem W, Pienaar R, Edvardsen B. *Chrysochromulina palpebralis* sp. nov. (Prymnesiophyceae): a haptophyte, possessing two alternative morphologies. *Phycologia*. 2009;48:165–76. Taylor & Francis.
74. Søgaard DH, Sorrell BK, Sejr MK, Andersen P, Rysgaard S, Hansen PJ, et al. An under-ice bloom of mixotrophic haptophytes in low nutrient and freshwater-influenced Arctic waters. *Sci Rep*. 2021;11:2915.
75. Hansen PJ, Nielsen TG, Kaas H. Distribution and growth of protists and mesozooplankton during a bloom of *Chrysochromulina* spp. (Prymnesiophyceae, Prymnesiales). *Phycologia*. 1995:409–16. <https://doi.org/10.2216/10031-8884-34-5-409.1>.
76. Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou S, Frada M, et al. Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc Natl Acad Sci U S A*. 2009;106:12803–8.
77. Nicholls KH, Beaver JL, Estabrook RH. Lakewide odours in Ontario and New Hampshire caused by *Chrysochromulina breviturrita* Nich. (Prymnesiophyceae). *Hydrobiologia*. 1982:91–5. <https://doi.org/10.1007/bf00006281>.
78. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37:540–6.
79. Hovde BT, Deodato CR, Hunsperger HM, Ryken SA, Yost W, Jha RK, et al. Genome sequence and transcriptome analyses of *chrysochromulina* tobini: metabolic tools for enhanced algal fitness in the prominent order prymnesiales (Haptophyceae). *PLoS Genet*. 2015;11: e1005469.
80. Hovde BT, Deodato CR, Andersen RA, Starckenburg SR, Barlow SB, Cattolico RA. *Chrysochromulina*: genomic assessment and taxonomic diagnosis of the type species for an oleaginous algal clade. *Algal Research Elsevier*. 2019;37:307–19.
81. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, et al. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*. 2004;306:79–86.
82. Lommer M, Specht M, Roy A-S, Kraemer L, Andreson R, Gutowska MA, et al. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol*. 2012;13:R66.
83. Saros JE, Anderson NJ. The ecology of the planktonic diatom *Cyclotella* and its implications for global environmental change studies. *Biol Rev Camb Philos Soc*. 2015;90:522–41.
84. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, et al. A giant virus in amoebae. *Science*. 2003;299:2033.
85. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. *Science*. 2004;306:1344–50.
86. Deeg CM, Chow C-ET, Suttle CA. The kinetoplastid-infecting *Bodo saltans* virus (BsV), a window into the most abundant giant viruses in the sea. *Elife*. 2018;7. <https://doi.org/10.7554/eLife.33014>
87. Fischer MG, Allen MJ, Wilson WH, Suttle CA. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc Natl Acad Sci U S A*. 2010;107:19508–13.
88. Laber CP, Hunter JE, Carvalho F, Collins JR, Hunter EJ, Schieler BM, et al. Coccolithovirus facilitation of carbon export in the North Atlantic. *Nat Microbiol*. 2018;3:537–47.
89. Moniruzzaman M, Weinheimer AR, Martinez-Gutierrez CA, Aylward FO. Widespread endogenization of giant viruses shapes genomes of green algae. *Nature*. 2020;588:141–5.
90. Aylward FO, Moniruzzaman M, Ha AD, Koonin EV. A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biol*. 2021;19: e3001430.
91. Schulz F, Roux S, Paez-Espino D, Jungbluth S, Walsh DA, Denef VJ, et al. Giant virus diversity and host interactions through global metagenomics. *Nature Nature Publishing Group*. 2020;578:432–6.
92. Bratbak G, Egge JK, Heldal M. Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Mar Ecol Prog Ser*. 1993;93:39–48. Inter-Research Science Center.
93. Rosenwasser S, Ziv C, Van Creveld SG, Vardi A. Virocell metabolism: metabolic innovations during host–virus interactions in the ocean. *Trends Microbiol*. Elsevier; 2016; Available from: <https://www.sciencedirect.com/science/article/pii/S0966842X16300695>
94. Nayfach S, Camargo AP, Schulz F, Eloë-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol*. 2021;39:578–85.
95. Thingstad TF, Lignell R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microb Ecol Inter-Research*. 1997;13:19–27.
96. Wilhelm SW, Suttle CA. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *Bioscience*. 1999;49:781–8.
97. Wilson WH, Tarran GA, Schroeder D, Cox M, Oke J, Malin G. Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. *J Mar Biol Assoc U K*. 2002;82:369–77. Cambridge University Press.
98. Meshram AR, Vader A, Kristiansen S, Gabrielsen TM. Microbial eukaryotes in an arctic under-ice spring bloom north of Svalbard. *Front Microbiol*. 2017. <https://doi.org/10.3389/fmicb.2017.01099>
99. Kim JI, Yoon HS, Yi G, Shin W, Archibald JM. Comparative mitochondrial genomics of cryptophyte algae: gene shuffling and dynamic mobile genetic elements. *BMC Genomics BioMed Central*. 2018;19:1–14.
100. Grujčić V, Nuy JK, Salcher MM, Shabarova T, Kasalický V, Boenigk J, et al. Cryptophyta as major bacterivores in freshwater summer plankton. *ISME J*. 2018;12:1668–81.
101. Šimek K, Mukherjee I, Nedoma J, de Paula CCP, Jezberová J, Sirová D, et al. CARD-FISH and prey tracer techniques reveal the role of overlooked flagellate groups as major bacterivores in freshwater hypertrophic shallow lakes. *Environ Microbiol*. Wiley; 2021; Available from: <https://onlinelibrary.wiley.com/doi/10.1111/1462-2920.15846>.
102. Piwosz K, Shabarova T, Pernthaler J, Posch T, Šimek K, Porcal P, et al. Bacterial and eukaryotic small-subunit amplicon data do not provide a quantitative picture of microbial communities, but they are reliable in the context of ecological interpretations. *mSphere*. 2020;5. <https://doi.org/10.1128/mSphere.00052-20>
103. Mukherjee I, Salcher MM, Andrei A-Ş, Kavagutti VS, Shabarova T, Grujčić V, et al. A freshwater radiation of diplomonads. *Environ Microbiol Wiley*. 2020;22:4658–68.
104. Cenci U, Sibbald SJ, Curtis BA, Kamikawa R, Eme L, Moog D, et al. Nuclear genome sequence of the plastid-lacking cryptomonad *Goniomonas avonlea* provides insights into the evolution of secondary plastids. *BMC Biol*. 2018;16:137.
105. Morris JJ. Black Queen evolution: the role of leakiness in structuring microbial communities. *Trends Genet*. 2015;31:475–82.
106. Znachor P, Nedoma J, Hejzlar J, Seda J, Kopáček J, Boukal D, et al. Multiple long-term trends and trend reversals dominate environmental conditions in a man-made freshwater reservoir. *Sci Total Environ*. 2018;624:24–33.
107. Znachor P, Hejzlar J, Vrbá J, Nedoma J, Seda J, Šimek K, et al. Brief history of long-term ecological research into aquatic ecosystems and their catchments in the Czech Republic. Part I: Manmade reservoirs. *researchgate.net*; 2016;1. Available from: <https://www.researchgate.net>.

- net/profile/Petr-Znachor/publication/311416951_Brief_history_of_long-term_ecological_research_into_aquatic_ecosystems_and_their_catchments_in_the_Czech_Republic_Part_I_Manmade_reservoirs/links/5853827808ae0c0f322282e6/Brief-history-of-long-term-ecological-research-into-aquatic-ecosystems-and-their-catchments-in-the-Czech-Republic-Part-I-Manmade-reservoirs.pdf
108. Beutler M, Wiltshire KH, Meyer B, Moldaenke C, Lüring C, Meyerhöfer M, et al. A fluorometric method for the differentiation of algal populations in vivo and in situ. *Photosynth Res.* 2002;72:39–53.
 109. Mackereth F, Heron J, Talling J. *Water analysis: some revised methods form limnologists.* Kendal: Titus Wilson and Son Limited; 1989.
 110. Murphy J, Riley JP. A modified single solution method for the determination of phosphate in natural waters. *Anal Chim Acta Elsevier.* 1962;27:31–6.
 111. 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. Taylor & Francis.
 112. Procházková L. Bestimmung der Nitrate im Wasser. *Fresenius' Zeitschrift für Analytische Chemie.* 1959;167:254–60. Springer Nature.
 113. Lund JWG, Kipling C, Le Cren ED. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. *Hydrobiologia Springer Nature.* 1958;11:143–70.
 114. Hillebrand H, Dürselen C-D, Kirschtel D, Pollinger U, Zohary T. Biovolume calculation for pelagic and benthic microalgae. *J Phycol Wiley.* 1999;35:403–24.
 115. Posch T, Loferer-Krößbacher M, Gao G, Alfreider A, Pernthaler J, Psenner R. Precision of bacterioplankton biomass determination: a comparison of two fluorescent dyes, and of allometric and linear volume-to-carbon conversion factors. *Aquat Microb Ecol.* 2001;25:55–63.
 116. Sherr BF, Sherr EB, Fallon RD. Use of monodispersed, fluorescently labeled bacteria to estimate in situ protozoan bacterivory. *Appl Environ Microbiol.* 1987;53:958–65.
 117. Šimek K, Grujić V, Nedoma J, Jezberová J, Šorf M, Matoušů A, et al. Microbial food webs in hypertrophic fishponds: Omnivorous ciliate taxa are major protistan bacterivores. *Limnol Oceanogr Wiley.* 2019;64:2295–309.
 118. Jezbera J, Hornák K, Šimek K. Prey selectivity of bacterivorous protists in different size fractions of reservoir water amended with nutrients. *Environ Microbiol Wiley Online Library.* 2006;8:1330–9.
 119. Brussaard CPD. Optimization of procedures for counting viruses by flow cytometry. *Appl Environ Microbiol.* 2004;70:1506–13.
 120. Sekar R, Pernthaler A, Pernthaler J, Warnecke F, Posch T, Amann R. An improved protocol for quantification of freshwater *Actinobacteria* by fluorescence in situ hybridization. *Appl Environ Microbiol.* 2003;2928–35. <https://doi.org/10.1128/aem.69.5.2928-2935.2003>
 121. 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.
 122. Friedrich U, Van Langenhove H, Altendorf K, Lipski A. Microbial community and physicochemical analysis of an industrial waste gas biofilter and design of 16S rRNA-targeting oligonucleotide probes. *Environ Microbiol.* 2003;439–439. <https://doi.org/10.1046/j.1462-2920.2001.00169.x-i1>
 123. Salcher MM, Pernthaler J, Frater N, Posch T. Vertical and longitudinal distribution patterns of different bacterioplankton populations in a canyon-shaped, deep prealpine lake. *Limnol Oceanogr.* 2011;2027–39. <https://doi.org/10.4319/lo.2011.56.6.2027>
 124. Neuenschwander SM, Salcher MM, Pernthaler J. Fluorescence in situ hybridization and sequential catalyzed reporter deposition (2C-FISH) for the flow cytometric sorting of freshwater ultramicrobacteria. *Front Microbiol.* 2015;6:247.
 125. Kemp PF, Cole JJ, Sherr BF, Sherr EB. *Handbook of methods in aquatic microbial ecology.* CRC Press; 1993.
 126. Metfies K, Medlin LK. Refining cryptophyte identification with DNA-microarrays. *J Plankton Res Oxford University Press.* 2007;29:1071–5.
 127. Šimek K, Grujić V, Mukherjee I, Kasalický V, Nedoma J, Posch T, et al. Cascading effects in freshwater microbial food webs by predatory Cerozoa, Katablepharidacea and ciliates feeding on aplastidic bacterivorous cryptophytes. *FEMS Microbiol Ecol. academic.oup.com;* 2020;96. <https://doi.org/10.1093/femsec/fiaa121>
 128. Seda J, Kolarova K, Petrusek A, Machacek J. *Daphnia galeata* in the deep hypolimnion: Spatial differentiation of a “typical epilimnetic” species. *Hydrobiologia.* 2007. p. 47–57.
 129. McCauley E, Downing JA, Rigler FH. *A manual on methods for the assessment of secondary productivity in fresh waters.* Oxford: Blackwell Scientific UK; 1984. p. 228–65.
 130. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods.* 2016;102:3–11.
 131. Quast C, Priesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013;41:D590-6. academic.oup.com.
 132. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol.* 2017;35:1026–8.
 133. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402. academic.oup.com.
 134. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ.* 2015;3:e1165. peerj.com
 135. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome.* 2020;8:90. microbiomejournal.biomedcentral ...
 136. Aylward FO, Moniruzzaman M. ViralRecall-A flexible command-line tool for the detection of giant virus signatures in omic data. *Viruses.* 2021;13. Available from: <https://doi.org/10.3390/v13020150>
 137. 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.
 138. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol.* 2018;36:996–1004.
 139. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2019. academic.oup.com; <https://doi.org/10.1093/bioinformatics/btz848>.
 140. 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;25:1043–55.
 141. 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. nature.com.
 142. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 2020;48:D265–8.
 143. Mann NH, Cook A, Millard A, Bailey S, Clokie M. Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature.* 2003;424:741.
 144. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using metagenomics. *PLoS Genet.* 2013;9: e1003987.
 145. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol Rev.* 2016;40:258–72.
 146. Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO. Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat Commun.* 2020;11:1710.
 147. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 2018;27:135–45.
 148. Crisculo 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.
 149. 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.
 150. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 2018;35:518–22.

151. 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.
152. Levy Karin E, Mirdita M, Söding J. MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome*. 2020;8:48. Springer.
153. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol Springer*. 2019;1962:227–45.
154. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49:D480–9.
155. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol*. 2014;12:e1001889.
156. Mirarab S, Nguyen N, Guo S, Wang L-S, Kim J, Warnow T. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *J Comput Biol*. 2015;22:377–86.
157. Weese D, Holtgrewe M, Reinert K. RazerS 3: faster, fully sensitive read mapping. *Bioinformatics*. 2012;28:2592–9.
158. Emiola A, Oh J. High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. *Nat Commun*. 2018;9:4956 nature.com.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

