



Metagenomics Analysis to Investigate the Microbial Communities and Their Functional Profile During Cyanobacterial Blooms in Lake Varese

Isabella Sanseverino¹ · Patrizia Pretto² · Diana Conduto António¹ · Armin Lahm³ · Chiara Facca⁴ · Robert Loos¹ · Helle Skejo¹ · Andrea Beghi⁵ · Franca Pandolfi⁵ · Pietro Genoni⁵ · Teresa Lettieri¹ 

Received: 26 April 2021 / Accepted: 26 October 2021 / Published online: 12 November 2021
© The Author(s) 2021

Abstract

Toxic cyanobacterial blooms represent a natural phenomenon caused by a mass proliferation of photosynthetic prokaryotic microorganisms in water environments. Bloom events have been increasingly reported worldwide and their occurrence can pose serious threats to aquatic organisms and human health. In this study, we assessed the microbial composition, with a focus on Cyanobacteria, in Lake Varese, a eutrophic lake located in northern Italy. Water samples were collected and used for obtaining a 16S-based taxonomic profile and performing a shotgun sequencing analysis. The phyla found to exhibit the greatest relative abundance in the lake included Proteobacteria, Cyanobacteria, Actinobacteriota and Bacteroidota. In the epilimnion and at 2.5 × Secchi depth, Cyanobacteria were found to be more abundant compared to the low levels detected at greater depths. The blooms appear to be dominated mainly by the species *Lyngbya robusta*, and a specific functional profile was identified, suggesting that distinct metabolic processes characterized the bacterial population along the water column. Finally, analysis of the shotgun data also indicated the presence of a large and diverse phage population.

Keywords Freshwater · Algal bloom · Microbial populations · Metagenomics · Water quality · Lyngbya

Introduction

Cyanobacteria are photosynthetic bacteria that are mostly found in freshwater systems. Due to their long evolutionary history, they have adapted to climatic, geochemical and anthropogenic changes. They have a key role in geochemical cycles together with picoplankton and other microorganisms

belonging to the microbial loop and are also involved in maintaining the environmental balance and the biodiversity of microorganisms. The increasing anthropogenic eutrophication and climate changes are contributing to the intense proliferation of Cyanobacteria in waterbodies, resulting in bloom formations [1]. Global warming is expected to intensify this phenomenon; indeed, the predicted increase in air and water temperature and the frequent rainfall events alternated with longer periods of drought may favour the water stratification in waterbodies and the dominance of cyanobacterial blooms [2]. The dynamics of bloom events are not yet fully understood; however, it is generally accepted that external factors, such as water temperature, nutrient loading and light intensity, can influence the potential of their occurrence [1]. In addition to environmental elements, biotic interactions between Cyanobacteria and heterotrophic bacteria may also influence the bloom dynamics [3, 4]. These microbial interactions are still little explored but seem to play positive (e.g. nutrient exchange, including vitamins) or negative roles (e.g. cyanolytic properties) in the cyanobacterial proliferation [5–9]. For decades, scientific studies were mainly restricted to addressing the influence of physico-chemical

✉ Teresa Lettieri
teresa.lettieri@ec.europa.eu

¹ European Commission, Joint Research Centre (JRC), Via E. Fermi 2749, 21027 Ispra, VA, Italy

² Biosearch Ambiente Srl, Via Tetti Gai 59, 10091 Alpignano, TO, Italy

³ Bioinformatics Project Support, P.za S.M. Liberatrice 18, 00153 Roma, Italy

⁴ Department of Environmental Science, Informatics and Statistics, University Ca' Foscari Venezia, Via Torino 155, 301702 Mestre, VE, Italy

⁵ ARPA, Agenzia Regionale Per La Protezione Dell'Ambiente Della Lombardia, Via Ippolito Rosellini 17, 20124 Milano, Italy

parameters (e.g. nutrients, pH and temperature) on the blooms without focusing on variations in the composition of the microbial community during a bloom outbreak. Generally, Cyanobacteria grow faster at high temperature determining a cyanobacterial dominance in temperate waterbodies where the vertical stratification, one of the main physical parameters, is intensified, leading to the occurrence of algal bloom events [10–13]. Phosphorus (P) and nitrogen (N) are also considered among the main factors responsible for the cyanobacterial proliferation [14, 15]. Indeed, eutrophication can result in algal bloom formations with a consequent increase in water turbidity and odour problems caused by the decomposition of algae. During a bloom, Cyanobacteria can also produce harmful toxins that can render water unsafe, cause fish mortality and affect human health [1]. Considering the ecological, economical and human health negative impacts of cyanobacterial blooms, their monitoring is crucial for an effective lake management, wherefore several predictive models have been already developed for forecasting cyanobacterial blooms in waterbodies [16]. Advances in sequencing technologies applied to environmental samples have improved our knowledge on the taxonomic composition of cyanobacterial communities, providing information on their relative abundance and their functional profile [17]. To date, many studies have been performed on Cyanobacteria using 16S sequencing with the aim to investigate the microbial community associated with blooms and factors promoting these events [18–23].

In this study, we used a sequencing approach to characterize the microbial community composition, with a focus on Cyanobacteria, in the Lake Varese (Italy) during two bloom events which occurred in two consecutive years. Lake Varese is considered one of the first and most evident examples of eutrophication in Europe [24]. The eutrophication process, in this lake, accelerated in the 1950s and was mainly caused by urban development and fertilization practices in agriculture. Initially classified as hypertrophic lake, following remedial actions aimed at reducing the P loading, Lake Varese is now in eutrophic status with bloom events occurring every year during the summer and early autumn [24–26]. Phytoplankton studies in Lake Varese were carried out only occasionally and first analyses revealed summer blooms associated with many genera of Cyanobacteria such as *Oscillatoria*, *Anabaena*, *Aphanizomenon*, *Gomphosphaeria*, *Leptolyngbya* and *Microcystis* [27–29]. However, a metagenomic approach has been never used before for providing a more comprehensive understanding on the composition of the microbial community in this lake. Information on the vertical distribution of bacteria and their functional profiles along the water column during blooms in Lake Varese is still missing. During summer 1997, a toxic algal bloom reported in this lake was characterized by the presence of a filamentous cyanobacterium, *Planktothrix* spp. FP1 strain, found to

be responsible for the production of potent neurotoxins causing the human syndrome paralytic shellfish poisoning (PSP) [30]. Here, we present the first assessment of the microbial community composition in Lake Varese during cyanobacterial bloom events that occurred in the years 2016 and 2017, using a metagenomic approach. We also looked at the functional profile associated to the bacterial populations detected at the different water depths.

Materials and Methods

Study Area

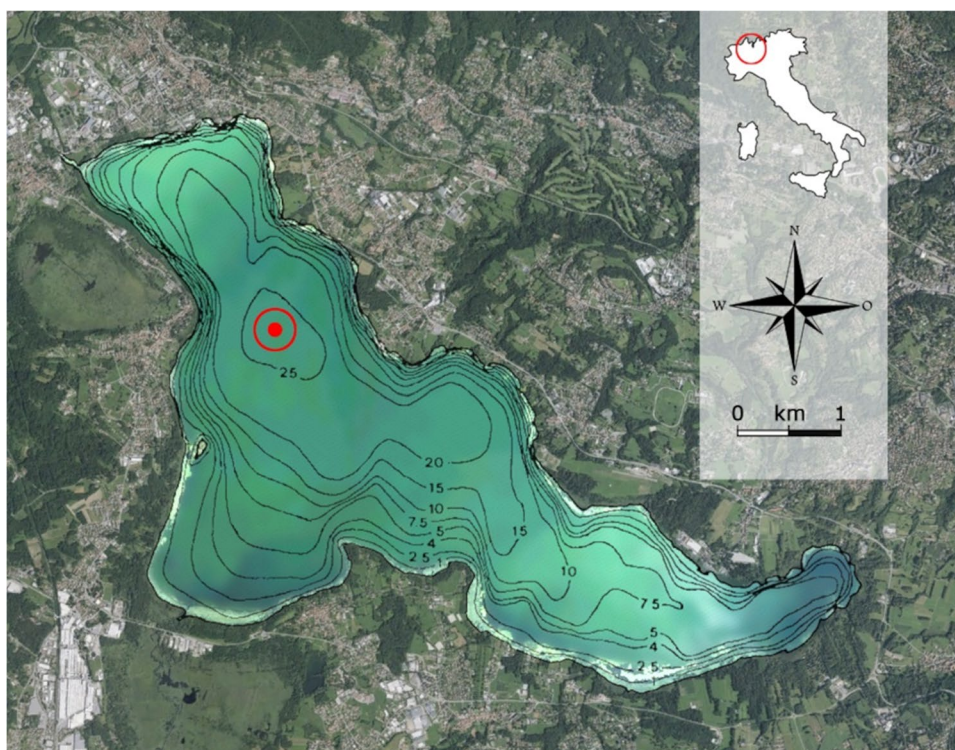
Lake Varese (45°49'N; 8°44'E) is situated in northern Italy at the feet of the Alps mountain range at a mean altitude of 236 m above sea level (Fig. 1). It is a warm monomictic and eutrophic shallow lake, with a mean depth of 11 m, a maximum depth of 26 m, a surface area of 14.8 km² and a theoretical renewal time of 1.7–1.9 years [25, 31].

Sample Collection and Processing

Water collection was performed in Lake Varese at coordinates 45°49'0.738N/008°43'0.190E, corresponding to the deepest region of the lake with a maximum depth of 26 m, and indicated in Fig. 1 by a red dot. The collected samples were stored in thermic containers for the transport to laboratory facilities and immediately prepared for downstream analysis. To evaluate the dynamics of the microbial community in the lake, sampling campaigns were carried out mainly during the summer period and on a weekly basis, during two consecutive years (from 31 August to 5 October in 2016 and from 19 July to 4 October in 2017). Water samples were taken at 3 different water column depths: 0.5 m from surface (Epi depth, E), 13 m (Meso depth, M) and 2.5 times the Secchi disk depth (named in the paper as S). The euphotic depth was determined as 2.5 times the Secchi disk depth or the region where photosynthetically active radiation (PAR) was larger than 1% of the radiation determined immediately below the water surface. E samples represent the region with the highest oxygen and radiation exposure and M samples refer to the near anoxic region, where temperature and radiation are low. Secchi disk depth was determined using a Secchi disk. Part of the campaigns was performed by the department of the Regional Environmental Protection Agency of Lombardia (ARPA) which kindly provided us data and water samples for analysis.

Vertical profiles measuring physico-chemical parameters including pH, conductivity (CD), oxidation–reduction potential (ORP), PAR, dissolved oxygen (DO), oxygen saturation (OS) and water temperature (WT) were collected using a multi-parametric probe (Hydrolab DS5, Corr-Tek).

Fig. 1 Location of the Lake Varese and lake bathymetry. The red dot in the figure indicates the deepest region of the lake where the samples were collected (45°49′.738N/008°43′.190E). The red circle in the upper map indicates the location where the Lake Varese is situated



Chlorophyll a, Cyanotoxins Analysis and Water Chemistry

A volume of 500 mL and 1 L of each water sample was filtered by GF/C membranes (Whatman) for chlorophyll *a* (Chl*a*) analysis. Filter membranes were folded, protected from light and stored at $-20\text{ }^{\circ}\text{C}$ for a maximum of 1 day. For spectrophotometric analysis, 14 mL of methanol was added to each filter and samples were boiled at $70\text{ }^{\circ}\text{C}$ for near 5 min (for the extraction of Chl*a*) and centrifuged for 7 min at 3500 rpm as described in [32], with slight modifications. The optical density (OD) was measured at 665 nm and 750 nm. Chl*a* content was determined by the following equation [32]:

$$\text{Chl}a(\mu\text{g/L}) = [13.9 * (\text{OD}_{665\text{nm}} - \text{OD}_{750\text{nm}}) * v * d] / P * V$$

where v is the added volume of methanol, d is the dilution factor (when applicable), P is the cuvette path in centimetres and V is the filtered water sample in litres.

Cyanotoxins concentrations (Microcystins, MCs; Saxitoxin, SX; and Anatoxin, AX) were quantified by enzyme-linked immunosorbent assay (ELISA) technique. Water samples were stored in glass vials at $-20\text{ }^{\circ}\text{C}$ until analysis. Toxins quantification was performed using commercial kits (Abraxis, USA) according to supplier's instructions.

Nutrients in the JRC laboratory were analysed with a Dionex ion chromatography system consisting of two

pre-column-column-suppressor systems followed by electrochemical detection (NO_3^- limit of detection $\text{LOD}=0.049\text{ mg/L}$, limit of quantification $\text{LOQ}=0.086\text{ mg/L}$; NH_4^+ $\text{LOD}=0.02\text{ mg/L}$, $\text{LOQ}=0.02\text{ mg/L}$; SO_4^{2-} $\text{LOD}=0.002\text{ mg/L}$, $\text{LOQ}=0.004\text{ mg/L}$; total P $\text{LOD}=0.05\text{ mg/L}$) [33]. The analytical columns used were, for the cations, an IonPac CS15 with a guard column IonPac CG15 (column temperature of $45\text{ }^{\circ}\text{C}$) and, for the anions, an IonPac AS9 and a guard column IonPac AG9 (column temperature of $25\text{ }^{\circ}\text{C}$). The ion suppresser after the column increases the sensitivity by lowering the background noise. The solvent eluents (kept under Helium pressure — 6 bars — to eliminate oxygen) were methanesulfonic acid (20 mM) for the cations, and Na_2CO_3 (9 mM) for the anions, and a flow of 0.3 mL/min at approx. 1000 psi. Injection volume was 25 μL or 10 μL . In the BO laboratory, total P was measured according to the internal method MI-122 rev 8 2015. The method is based on the property of orthophosphate ions to react in an acidic environment with molybdate and antimonyl ions to form a phospho-antimonyl-molybdate reducible by ascorbic acid; the reduced heteropolyacid, containing hexa and tetravalent molybdenum, is soluble and shows a characteristic blue color: the final photometric determination is carried out at 800 nm. Organic phosphorus must be previously transformed into orthophosphate by hot oxidation with persulphate. The determination of the P concentration is performed through the use of the LCK 349 cuvette test (Hach Lange).

DNA Extraction

For DNA extraction, aliquots of 250 mL or 500 mL of water were filtered through 0.22 μm nitrocellulose filter membranes (GSWP, Whatman). Filters were stored at $-20\text{ }^{\circ}\text{C}$ (or $-80\text{ }^{\circ}\text{C}$ for longer storage) until use. DNA extraction was performed according to the protocol described in Kisand et al. [34] except for the lyticase incubation. Briefly, each frozen filter was thawed, incubated in 5 mL of 50 mM KH_2PO_4 buffer (pH 7.5) and shaken overnight (160 rpm) at $4\text{ }^{\circ}\text{C}$. The following day, the KH_2PO_4 buffer was recovered and filters were transferred to a tube with 3 mL of fresh 50 mM KH_2PO_4 to be sonicated at $60\text{ }^{\circ}\text{C}$ for a total of 15 min. Filters were discarded, and sonicated and not sonicated buffers were pulled together and then shaken for 3 h at $30\text{ }^{\circ}\text{C}$ with 533 μL lysozyme (100 mg/mL in DEPC water, Sigma Aldrich) and 7.7 μL β -mercaptoethanol (Sigma-Aldrich). Samples were frozen at $-20\text{ }^{\circ}\text{C}$, thawed and centrifuged at $4\text{ }^{\circ}\text{C}$ for 20 min at 14,000 rpm. The DNA was extracted using the DNeasy blood and tissue kit (Qiagen) following the manufacturer's instructions. DNA was quantified by measurements with both Nanodrop and Qubit (ThermoFisher). DNA concentrations measured with the two instruments ranged from 18.6 to 48.6 ng/ μL (Nanodrop) and from 13.7 to 48.8 ng/ μL (Qubit). Purified DNA samples were subjected to 16S sequencing and shotgun analysis.

16S and Shotgun Sample Preparation and Sequencing

16S amplicons and total community genomic DNA were sequenced at Cemet GmbH (Tubingen, Germany). 16S V3–V4 amplicons were generated from 10 ng of DNA using forward primer S-D-Bact-0341-b-S-17 (5'-CCTACGGGNGGCWGCAG-3') and reverse primer S-D-Bact-0785-a-A-21 (5'-GACTACHVGGGTATCTAATCC-3') [35]. Only for the 2016 16S samples, two independent extractions and amplifications were carried out to explore variability of technical replicates. Library generation was performed according to the recommendations given by Illumina. Amplicons were sequenced as 2×250 bp read pairs on an Illumina MiSeq instrument using MiSeq Reagent Kit v2. A minimum of 22,000 read pairs were generated per 16S samples. Shotgun genomic data was generated as 2×100 bp read pairs (2016 dataset) and 2×150 bp read pairs (2017 dataset) on an Illumina HiSeq instrument. An overview on shotgun read numbers and filtering is provided in Supplementary Online resource 1: Supplemental Table S1. Library preparation was performed according to the Illumina Nextera XT protocol.

16S and Shotgun Data Analysis

16S Data For the 16S data, after removal of primers, the read pairs were initially filtered by trimmomatic v0.38 [36] applying a minimum length-cutoff of 245 bp and a minimum average (Avg) quality values of 30. Read pairs were then combined into complete amplicon using FLASH v2.0 [37]. The resulting merged read pairs were then filtered omitting sequences shorter than 350 nucleotides. Average overall length of merged read pairs was 413.5 (median 412), minimum 351 and maximum 480 (Online resource 1: Supplemental Table S2).

16S amplicons were then rarefied to 22,000 merged read pair sequences using usearch [38] and clustered into operational taxonomy unit (OTUs) at 97% sequence identity. Taxonomic assignment was then performed using the usearch SINTAX algorithm [39] against the GTDB [40] release95 ssRNA archaea and bacteria database (<https://data.ace.uq.edu.au/public/gtdb/data/releases/release95/95.0/>). OTUs clustering and heatmaps were generated using the heatmap.2 function in ggplot2 (CRAN repository: <https://cran.r-project.org/web/packages/ggplot2/index.html>) from the R language (<https://www.r-project.org/foundation/>). Clustering was performed with heatmap.2 default parameters (complete agglomeration using the Euclidean measure to obtain a distance matrix). For bubble plots displaying the relative abundance of the 16S amplicons from the 2016 samples, values were averaged across the two replicates.

Shotgun Data Shotgun read pairs were filtered with trimmomatic v0.38 [36] applying a minimum length-cutoff of 90 bp (2016 dataset) and 140 bp (2017 dataset) and a minimum Avg quality value of 28. Reads were filtered by trimmomatic as pairs in order to allow also running Kraken2 [41] for taxonomic binning in paired read mode using a pre-compiled version Kraken2 format of the GTDB database available at https://bridges.monash.edu/projects/Metagenomics_Index_Correction/65534. Assembly of shotgun read pairs into contigs was performed with MegaHit v1.2.9 using default parameters [42]. For each sample, the set of contigs obtained by MegaHit with length ≥ 5000 was confronted with a BLASTN database containing contigs from all other samples. Examination of the BLASTN search results was performed filtering out matches with $\geq 99\%$ sequence identity and with a match length covering at least 50% of the query contig. MegaHit reports a multiplicity value for each contig representing the average coverage by reads for each contig. In order to normalize for the total number of read pairs in each sample, multiplicity values of the BLASTN hits selected with the above criteria were adjusted accordingly. All contigs obtained were also confronted with a BLASTN

database containing all GTDB genomes filtering out matches with at least 90% identity over 10,000 nucleotides.

Functional annotation of shotgun reads was performed, due to limited computing resources, on 2 million randomly selected reads from each sample using DIAMOND [43] and the bacterial subsection of the NCBI nr database. Results from DIAMOND were then imported into MEGAN6 CE [44] and analysed using the SEED database [45]. In particular, variations in relative abundances associated to general pathways or to specific metabolic functions have been investigated using the SEED subsystems level 1 and level 2, respectively. Level 1 subsystems describe a specific biological process (similar to a pathway) or structural complex while level 2 subsystems describe the individual components of the level 1 process.

Reliability of read pair assignment to the *Lyngbya robusta* species by Kraken2 was estimated in the following way: the *Lyngbya robusta* (*Limnoraphis robusta* CS-951) genome assembly GCF_000972705.2 (ASM97270v2) was downloaded from NCBI, and the genome assembly contigs were merged into one contiguous sequence (7,314,117 nucleotides) and converted into a BLASTN database. Sequences of read pairs that had been assigned to *Lyngbya robusta* in the Kraken2 output files were extracted and assembled with MegaHit. The resulting contigs were then confronted with the BLASTN *Lyngbya robusta* database. A custom-written perl script then filtered the BLASTN output accepting, for each match start position, only the hit with the highest percent identity. In addition, the script also collected all regions of the merged *Lyngbya robusta* genome assembly that were covered by hits from the MegaHit contigs.

To prepare contigs for a MetaBat2 [46] analysis, all shotgun read pairs from the 2017 E samples were assembled with MegaHit and read pairs from the individual samples were aligned to the assembled contigs using bowtie2 [47]. Resulting bam files were then processed with MetaBat2 obtaining binned contigs. The same procedure (bowtie2 mapping and MetaBat2 binning) was then applied also to the shotgun data from the 2017 S and M samples and separately, to all 2016 samples. MegaHit analysis of all 2016 and 2017 (or only 2017) shotgun samples together was not possible due to limiting (memory) computing resources. Details of all MetaBat2 are available in Online resource 2: file var-eselake_2.xlsx. N50 values provided indicate the length of that contig within the bin where about 50% of the complete bin length is covered, after ordering the contigs from longest to shortest. Assignment of MetaBat2 bins to known GTDB genomes/genome assemblies was then performed with BLASTN accepting hits of at least 1500 nucleotides with at least 90% identity. Completeness of MetaBat2 bins was performed with checkm [48] and further annotation of

bins was performed with GTDBtk [49]. Estimate of phage content of the MetaBat2 bins was done with virsorter2 [50].

Network co-occurrence was performed with Cytoscape [51] and Conet [52] using the Kraken2 results from the 2017 shotgun data at the genus level.

All processing was performed on a workstation with one Xeon CPU (8 hyper-threads) and 48 Gb RAM.

Redundancy Analysis

The Canoco 5 version [53] was used to run the analysis. Data were compositional with a gradient 2.1 SD units long, so linear method was preferred. Data from 16S sequencing were log-transformed and the taxa which relative abundance was > 1% were selected. Both axes were significant when $p < 0.05$.

Results

Physico-Chemical and Inorganic Chemical Parameters

In 2016 and 2017, six and eight sampling campaigns were performed at Lake Varese, respectively. During the 2016 study period, the $2.5 \times$ Secchi depth (S), which is a good proxy to estimate the euphotic depth, varied between 7.6 and 13 m (see Table 1). Range and Avg of the physico-chemical parameters measured during the sampling campaigns (WT, CD, DO, OS, PAR, ORP and pH) are reported in Table 1.

Generally, a uniform physico-chemical pattern was observed across all three depths during the sampling period. At the greatest depth (M), WT showed lower values compared to an Avg temperature of 22.7 °C recorded at the surface (E) and the CD was minimum at the surface (242 μ S/cm at 25 °C) and maximum at the M depth (323 μ S/cm at 25 °C). In the selected study period, the water column was stratified as evidenced by the DO levels measured at the different depths. The OS was observed to decrease along the water column, with values always equal to 0% at the M depth and PAR showed a fluctuation particularly at E depth (Online resource 1: Supplemental Table S3). The pH was generally alkaline with highest values around 8 measured at E depth. Similar values for physico-chemical parameters were observed in 2017 (see Table 1 and Online resource 1: Supplemental Table S4). No measurements were performed on 19/7/2017 and 26/7/2017 due to technical problems with the probe. Nutrients were also measured in Lake Varese: levels for nitrate were below the LOD (LOD = 0.049 mg/L) in most of the samples analysed, ammonia was less present in the surface but more abundant in the anoxic bottom (0.3–0.8 mg/L) and, in September 2016, total P ranged from

Table 1 Range and average (Avg) of physico-chemical and biological parameters. *E*, Epi depth; *S*, 2.5 × Secchi depth; *M*, Meso depth; *Min*, minimum; *Max*, maximum; *SD*, standard deviation. (*) *S* depth: 13 m on 3/18/2016; 7.6 m on 7/9/2016; 8 m on 14/9/2016; 7.8 m on 21/9/2016; 10.3 m on 28/9/2016 and 13 m on 5/10/2016; 2 m on 2/8/2017; 6 m on 9/8/2017; 5 m on 23/8/2017 and 30/8/2017; 8.75 m on 27/9/2017 and 11 m on 4/10/2017

| Parameters and units of measurement | E (0.5 m) | | | | S (*) | | | | M (13 m) | | | |
|--|-------------------------|-------|-------|-------|-------|-------|-------|-------|----------|------|--------|------|
| | Min | Max | Avg | SD | Min | Max | Avg | SD | Min | Max | Avg | SD |
| 2016 | | | | | | | | | | | | |
| Water temperature (°C) | (WT) 20.4 | 25.2 | 22.7 | 1.9 | 9.2 | 16.8 | 13.2 | 3.0 | 9.2 | 10.3 | 9.8 | 0.3 |
| Conductivity (µS/cm at 25 °C) | (CD) 242 | 252 | 246.5 | 3.7 | 307 | 319 | 312.8 | 4.0 | 312 | 323 | 318.3 | 3.1 |
| Dissolved oxygen (mg/L) | (DO) 8 | 9.7 | 8.6 | 0.5 | 0 | 1.1 | 0.2 | 0.4 | 0 | 0 | 0 | 0 |
| Oxygen saturation (%) | (OS) 91.2 | 120.7 | 102.7 | 8.1 | 0 | 10.8 | 2.1 | 4.1 | 0 | 0 | 0 | 0 |
| Photosynthetically active radiation (µE/m ² /s) | (PAR) 74 | 2444 | 691.4 | 870.1 | 6 | 13 | 8 | 2.7 | 6 | 6 | 6 | 0 |
| Oxidation reduction Potential (mV) | (ORP) -51 | 206 | 130.1 | 100.8 | -154 | 212 | 41.3 | 156.1 | -175 | -136 | -152.7 | 12.5 |
| pH | 8.3 | 8.5 | 8.4 | 0.1 | 7.1 | 7.5 | 7.3 | 0.1 | 7 | 7.2 | 7.1 | 0.1 |
| Chlorophyll <i>a</i> (µg/L) | (Chl _a) 1.5 | 12.8 | 5.8 | 3.9 | 1.6 | 15.4 | 8.3 | 5.9 | 0.8 | 4.7 | 2.9 | 1.6 |
| 2017 | | | | | | | | | | | | |
| Water Temperature (°C) | (WT) 19.5 | 27.2 | 23.5 | 3.1 | 10.8 | 26.4 | 20.1 | 5.8 | 9.2 | 9.8 | 9.3 | 0.4 |
| Conductivity (µS/cm at 25 °C) | (CD) 172 | 206 | 188.8 | 11.3 | 208 | 281 | 245.5 | 27.9 | 281 | 312 | 293.5 | 22.5 |
| Dissolved oxygen (mg/L) | (DO) 8.3 | 11.1 | 9.8 | 1.1 | 0.3 | 11.2 | 3.9 | 3.9 | 0.2 | 0.5 | 0.3 | 0.1 |
| Oxygen saturation (%) | (OS) 102.6 | 136.5 | 116.8 | 11.6 | 2.4 | 139.5 | 47.1 | 47.6 | 1.7 | 4.5 | 2.4 | 0.6 |
| Photosynthetically active radiation (µE/m ² /s) | (PAR) 141 | 443 | 272.6 | 125.0 | 6 | 235.9 | 25.2 | 56.4 | 0 | 6 | 5.6 | 1.5 |
| Oxidation reduction potential (mV) | (ORP) 128 | 284 | 204.9 | 64.9 | -10 | 273 | 172.7 | 110.1 | -165 | -39 | -115.3 | 57.8 |
| pH | 8.6 | 9.1 | 8.6 | 0.5 | 7.3 | 8.9 | 7.5 | 0.8 | 5.6 | 7.5 | 6.9 | 0.7 |
| Chlorophyll <i>a</i> (µg/L) | (Chl _a) 4.1 | 19.2 | 12.1 | 4.8 | 7.1 | 22.4 | 13.6 | 5.5 | 2.4 | 22 | 5.9 | 6.3 |

0.06 to 0.26 $\mu\text{g/L}$ at E depth (Online resource 1: Supplemental Table S5).

Chlorophyll a and Cyanotoxins Analysis

Chl a concentration was found to be generally higher in both E and S than M depth (Fig. 2a, Table 1 and Online resource 1: Supplemental Table S6). In 2016, the highest Chl a levels, exceeding the value of 14 $\mu\text{g/L}$, were observed in S samples, while Chl a analysis at E depth revealed a peak value close to 13 $\mu\text{g/L}$ on the 21st of September (Fig. 2a and Online resource 1: Supplemental Table S6). During the 2017 campaign, Chl a content observed in the epilimnion was alternately higher or lower than the concentrations detected at S depth (Fig. 2a and Online resource 1: Supplemental Table S7). Peak values were reported on 2/8/2017 (M), 9/8/2017 (E) and 30/8/2017 (S) (Fig. 2a and Online resource 1: Supplemental Table S7).

In 2016, levels of the cyanotoxins MCs, SX and AX were measured by ELISA in raw water samples, revealing the intra- and extracellular amount that was present in water at the time of sampling. MCs content exceeded the

value of 0.35 $\mu\text{g/L}$ only on the 21st of September in both E (0.40 $\mu\text{g/L}$) and S samples (0.46 $\mu\text{g/L}$) (Fig. 2b). In all other sampling dates, MCs concentration was higher in E samples than in M and S samples, with values ranging from 0.26 to 0.40 $\mu\text{g/L}$ (Fig. 2b).

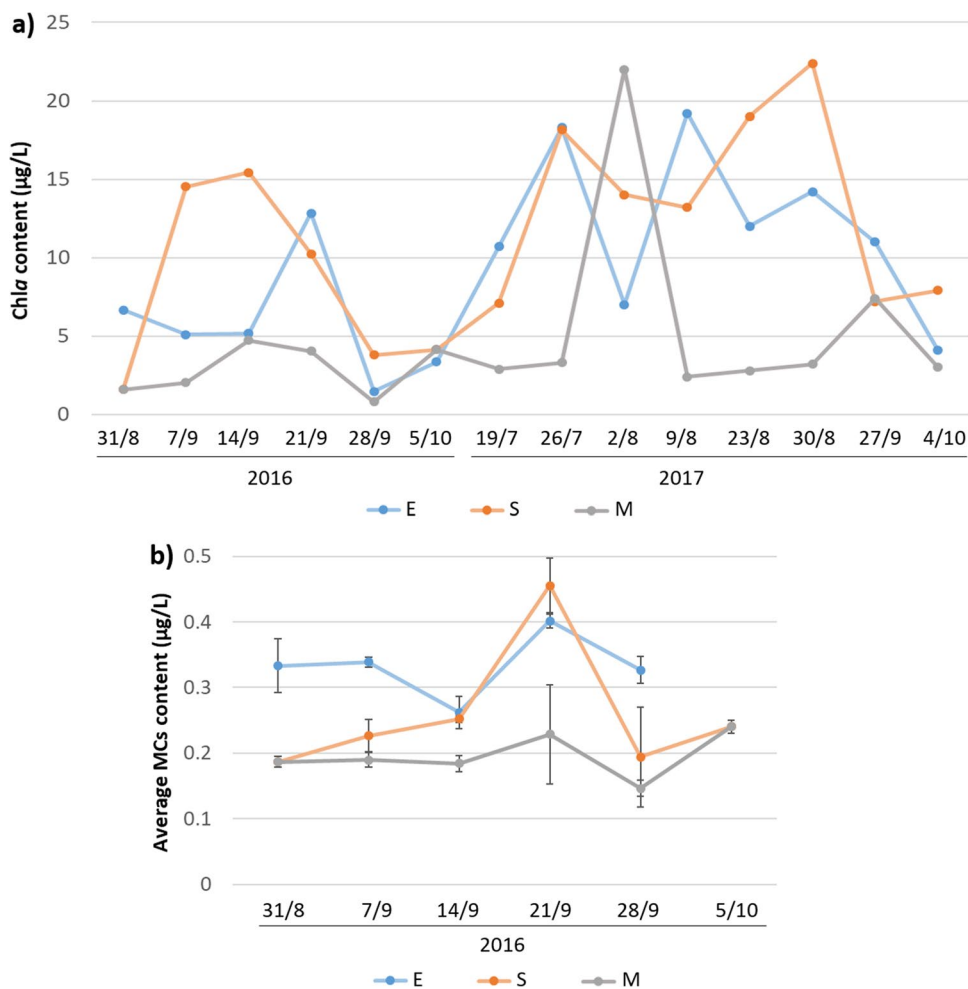
SX concentration was below the detectable level (<0.020 $\mu\text{g/L}$) in most of the samples analysed (data not shown). A low SX content was detected in only two samples found positive and corresponding to the E samples collected on 14th (0.021 $\mu\text{g/L}$) and 28th of September (0.020 $\mu\text{g/L}$). No AX levels were detected during the campaigns (LOD = 0.1 $\mu\text{g/L}$, data not shown).

16S Analysis

Community Composition of Main Phyla

Microbial community composition was analysed by 16S sequencing in water samples collected during the two sampling campaigns (2016 and 2017) at the three different depths: E (0.5 m), S and M (13 m) (the M depth corresponded to the S depth on 31/8/2016 and 5/10/2016). To

Fig. 2 Chlorophyll a (Chl a) and total Microcystins (MCs) content in raw samples collected in Lake Varese. **a** Chl a concentrations, in $\mu\text{g/L}$, measured during the sampling dates (x axis) at the surface (E), 2.5 \times Secchi disk depth (S) and 13 m depth (M). **b** Total MCs content measured by ELISA and detected across the study period (dates reported on the x axis) at the surface (E), 2.5 \times Secchi disk depth (S) and 13 m depth (M). Results are expressed as means of three replicates and standard deviation bars are indicated for each sampling point



assess the variation in the microbial community in Lake Varese, a first global and unbiased view was obtained by clustering OTUs [54] counts (Online resource 1: Supplemental Fig. S1) from samples collected during the two sampling campaigns. Four main clusters were readily visible: cluster 1 present in all samples, clusters 2 and 4 present in M and some S samples and cluster 3 present in E and S samples (Online resource 1: Supplemental Fig. S1). A closer inspection of the taxonomies (phylum level) present in each cluster did however not reveal dominant phyla. In addition, the analysis also indicated the high reproducibility of replicate samples collected during 2016.

Taxonomic binning was then obtained by randomly choosing 22,000 amplicons from each sample, clustering amplicons into OTUs at 97% sequence identity and assigning taxonomies through application of the usearch Sintax algorithm [39]. Rarefaction analysis of OTUs indicated an essentially complete overview on the community complexity for all the samples (Online resource 1: Supplemental Fig. S2). Highest diversity was present in the 2016 M samples, while 2017 E samples showed lowest complexity (Online resource 1: Supplemental Fig. S2).

Microbial Community Structure Based on 16S Amplicons

Between 84 and 95% of OTUs were successfully assigned at the phylum level, revealing an overall relatively constant profile, with some fluctuations across samples and seasons. Taxonomic analysis at phylum level showed that, in most of the samples, Proteobacteria was the predominant taxonomic group followed by Actinobacteriota, Cyanobacteria and Bacteroidota (Online resource 1: Supplemental Fig. S3). A unique peak of Bdellovibrionota, a bacterial predator present in lakes [55], was found in one of the 2016 S samples (Online resource 1: Supplemental Fig. S3). Proteobacteria dominated the profile in all E and M samples collected in both 2016 and 2017. In the dataset 2016, Cyanobacteria showed lower abundance in M samples compared to the other two depths and were found to be the most abundant community in S samples collected during three sampling campaigns in September 2016 (7/9/2016, 14/9/2016 and 28/9/2016) (Online resource 1: Supplemental Fig. S3). In the same month, the cyanobacterial relative abundance at E depth exceeded the value recorded at S depth only on the 21st of September. That abundance was found to be above all the cyanobacterial content observed in 2016 at E depth. Metagenomic data from 2017 suggested that Lake Varese experienced a continuous cyanobacterial bloom during the study period, as also confirmed by the visual inspection. In datasets 2016 and 2017, the phylum Desulfobacterota was identified at percentages above 1% in all 2016 M samples and in three S samples (28/9/2016, 9/8/2017 and 4/10/2017) (Online resource 1: Supplemental Fig. S3).

At the genus level, the fraction of OTUs assigned to a specific taxonomy varied considerably, with values from only 20 up to 60%, thus providing only a partial snapshot of a more detailed taxonomic composition. As expected, fluctuations in the profile were much more pronounced compared to data at phylum level, both between seasons and between samples obtained at different depths. *Planktophila*, *Fonsibacter* and *Nanopelagicus*, three of the most ubiquitous and abundant freshwater bacterial genera, showed great variability at the three different depths with E samples usually showing the highest levels detected among all samples (Fig. 3). Analyses at the genus level allowed to more precisely characterize the unique peak of Bdellovibrionota observed at the phylum level in the dataset 2016 (Supplemental Fig. S3), as belonging to the genera *Silvanigrella* (Fig. 3), according to the GTDB database classification. Microcystis was detected at very low levels in all samples (Fig. 3). In this study, the cyanobacterial phylum was mainly represented by the genus *Limnoraphis* (also known as *Lyngbya*), which is described to form blooms in freshwater environments [56, 57] (Fig. 3). Methanotrophs and methylotrophs like *Methylomonas*, *Methylobacter* and *Methylopumilus* were also detected (Fig. 3).

To verify how environmental variables affect taxa distribution derived from 16S sequencing analysis, a redundancy analysis (RDA) was carried out (Fig. 4). RDA summarizes part of the variation in taxa composition explained by environmental variables. The first two axes in both RDAs were significant ($p < 0.05$) and they explained 80% of cumulative variations and 97% of fitted cumulative variation. It means that the selected environmental variables well explain the species distribution.

Surface water layers (E and S) were characterized by higher temperature, pH, ORP and OS than the M layer, where only CD was high. These parameters all correlated well with the positive side of the first axis in 2016 and with the negative side of the first axis in 2017. Most of the taxa were significantly correlated with the environmental parameters, either directly or indirectly. Only Planctomycetota and Verrucomicrobiota in 2016, and Proteobacteria in 2017 did not have significant correlations.

Shotgun Metagenomics

Microbial Community Structure Based on Shotgun Data

Using the Kraken2 algorithm in combination with the GTDB release89 database, between 14 and 29% of the shotgun reads, analysed as read pairs, could be classified at the phylum level, reproducing relatively well the profile obtained by the 16S data. Supplemental Fig. S4 (Online resource 1) shows the phylum level analysis of the shotgun reads. Shown values represent, for each taxa, the relative

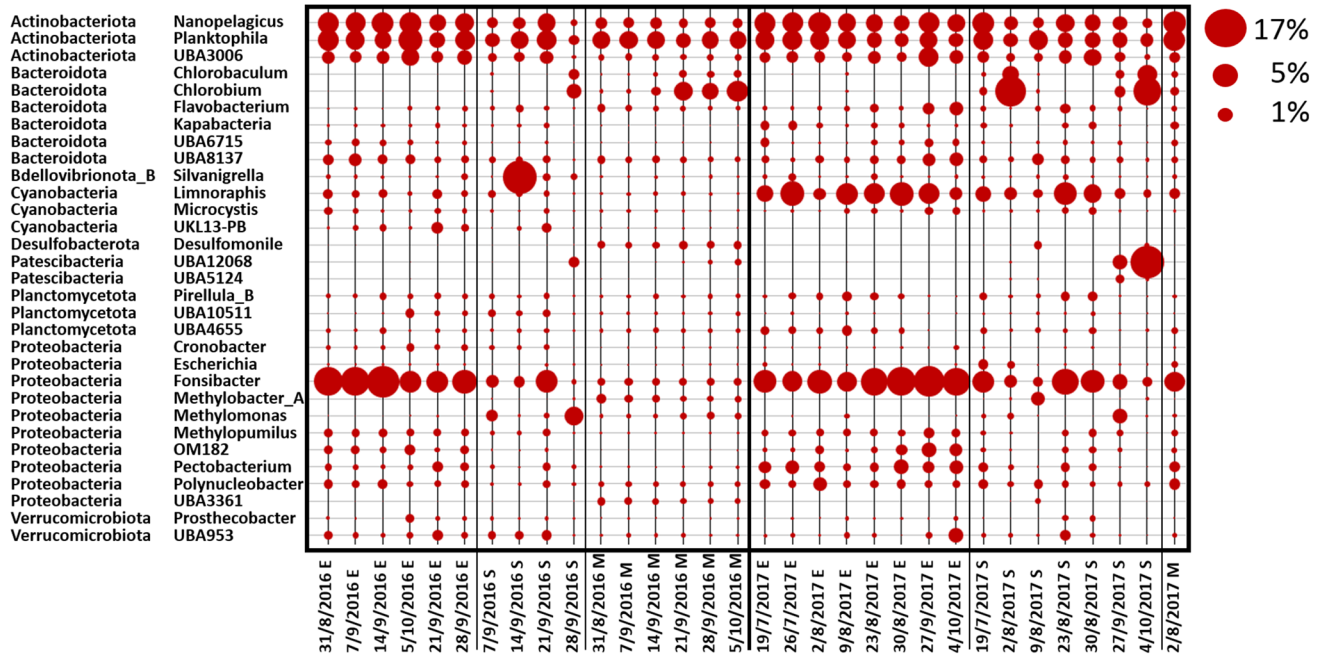


Fig. 3 Community composition at the genus level as determined from the 16S data. The community composition is shown for all the water samples collected during the sampling period in Lake Varese at the three different depths (Epi (E), 0.5 m; 2.5×Secchi (S) and Meso (M), 13 m). Samples were sequenced for 16S rRNA (V3–V4 region) and analysed at the genus level. Secchi depths were measured at each

site using a Secchi disk and values multiplied by 2.5. The M depth corresponded to the S depth on 31/8/2016 and 5/10/2016. For 2016 replicate samples, the average value is shown. Circles shown represent percent values within the fraction of taxonomy assigned shotgun sequences (circles shown at the top right of the figure represent example percent values)

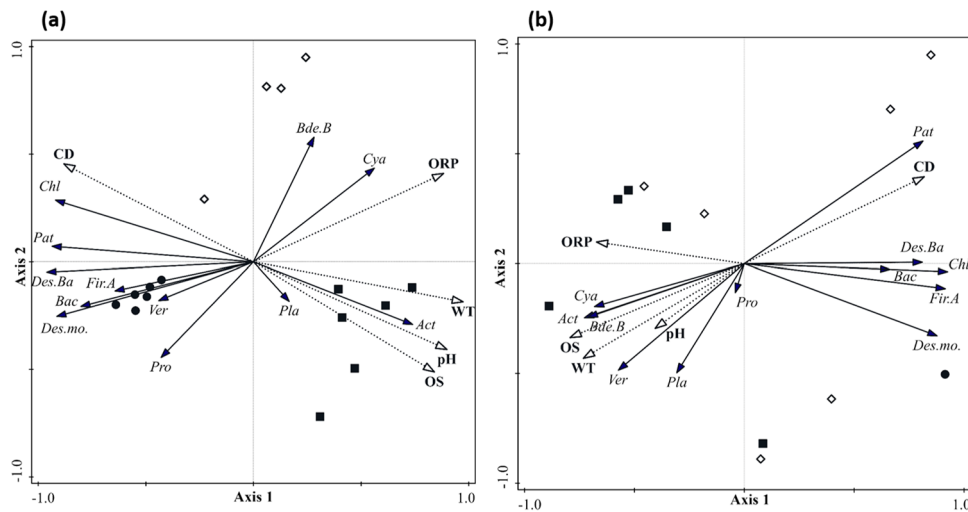


Fig. 4 Redundancy analysis (RDA) of the 16S data 2016 and 2017 at phylum level. RDA triplot with 16S taxa (solid lines) with abundance > 1%, environmental variables (dotted lines) and samples collected from August 31st, 2016, to October 5th, 2016 (a), and from August 2nd, 2017, to October 4th, 2017 (b). Black circles are samples from the Meso layers, diamonds are samples from the 2.5×Secchi layers, black squares are samples from the Epi layers. Abbreviations of environmental variables are as follow: pH, conductivity (CD), oxi-

dation reduction potential (ORP), oxygen saturation (OS) and water temperature (WT). Abbreviations of taxa are as follow: Actinobacteriota (Act), Bacteroidota (Bac), Bdellovibrionota_B (Bde.B), Chloroflexota (Chl), Cyanobacteria (Cya), Desulfobacterota (Des.Ba), Desulfuromonadota (Des.mo.), Firmicutes_A (Fir.A), Patescibacteria (Pat), Planctomycetota (Pla), Proteobacteria (Pro), Verrucomicrobiota (Ver)

abundance (percentage) with respect to all shotgun reads for which a phylum level classification had been obtained. In the E samples, the 2016 results indicate highest levels of Cyanobacteria on 14/9/2016 and 21/9/2016 (in the 16S data, the peak at E depth was on 21/9/2016, Online resource 1: Supplemental Fig. S3), while for S samples, a peak in cyanobacterial content was registered on 14/9/2016 (Online resource 1: Supplemental Fig. S4) and shotgun data confirmed the low levels of Cyanobacteria in the M region. Like for the 16S data, in 2017, the cyanobacterial abundance in the epilimnion suggested a continuous bloom from July to August, showing a drop starting from September as suggested from visual inspection. The low presence of Cyanobacteria at the M depth was also confirmed in 2017. In both the 2016 and 2017 dataset, some variations of the overall microbial composition were also observed for Proteobacteria and Actinobacteriota (Online resource 1: Supplemental Fig. S4). The absence of the *Bdellovibrionota* peak reported in 16S analysis could be due to the fact that a partial *Silvanigrella* genome assembly (GCA_014281055 NCBI Genome database) has only been generated very recently.

At the genus and species level, only a small fraction of shotgun reads, less than 5% could be classified thus allowing only a very limited view into the complexity of the bacterial community at this level. However, at least at the qualitative level, some additional information was extracted. *Lyngbya* was the dominant cyanobacterial

genera with *Lyngbya robusta* (*Limnoraphis robusta*) being the most abundant species (Fig. 5 and Fig. 6). This species was already observed in other lake bloom events [58, 59]. *Microcystis* was generally present at low level in 2017 (Fig. 5) and genera like *Synechococcus* and *Snowella* were also detected in our samples. *Planktophila*, *Fonsibacter* and *Nanopelagicus* were among the most abundant bacterial genera detected at E depth across all samples (Fig. 5). Analyses at species level revealed that the occurrence of these genera was mainly attributed to the species *Planktophila vernalis*, *Fonsibacter ubiquis* and *Nanopelagicus sp001437855* (Fig. 6). Sequences corresponding to *Sulfuritalea hydrogenivorans*, a sulphur-oxidizing species previously isolated from stratified freshwater lakes [60], were found to be generally greater in the mesolimnion compared to the other depths (Fig. 5 and Fig. 6). OTUs belonging to *Thiodictyon syntrophicum* [61], another species involved in sulphur oxidation, were also identified in both M and S samples (Fig. 5 and Fig. 6), as well as *Chlorobium* which is a photoautotrophic sulphur oxidizer (Fig. 5 and Fig. 6). This is consistent with the negative ORP values and anoxia at the bottom of the lake (Online resource 1: Supplemental Tables S3 and S4) which could allow the biological production of reduced species of sulphur from oxidized species as sulphate (Online resource 1: Supplemental Table S5), available for the metabolism of autotrophic sulphur oxidizers.

Fig. 5 Community composition at the genus level as determined from the shotgun data. The community composition is shown for all water samples collected during the sampling period in Lake Varese at the three different depths (Epi (E), 0.5 m; 2.5 × Secchi (S) and Meso (M), 13 m). Samples were analysed for shotgun sequencing. The figure shows variations in the microbial community at the genus level. Secchi depths were measured at each site using a Secchi disk and values multiplied by 2.5. The M depth corresponded to the S depth on 31/8/2016. Circles shown represent percent values within the fraction of taxonomy assigned shotgun sequences (circles shown at the top right of the figure represent example percent values)

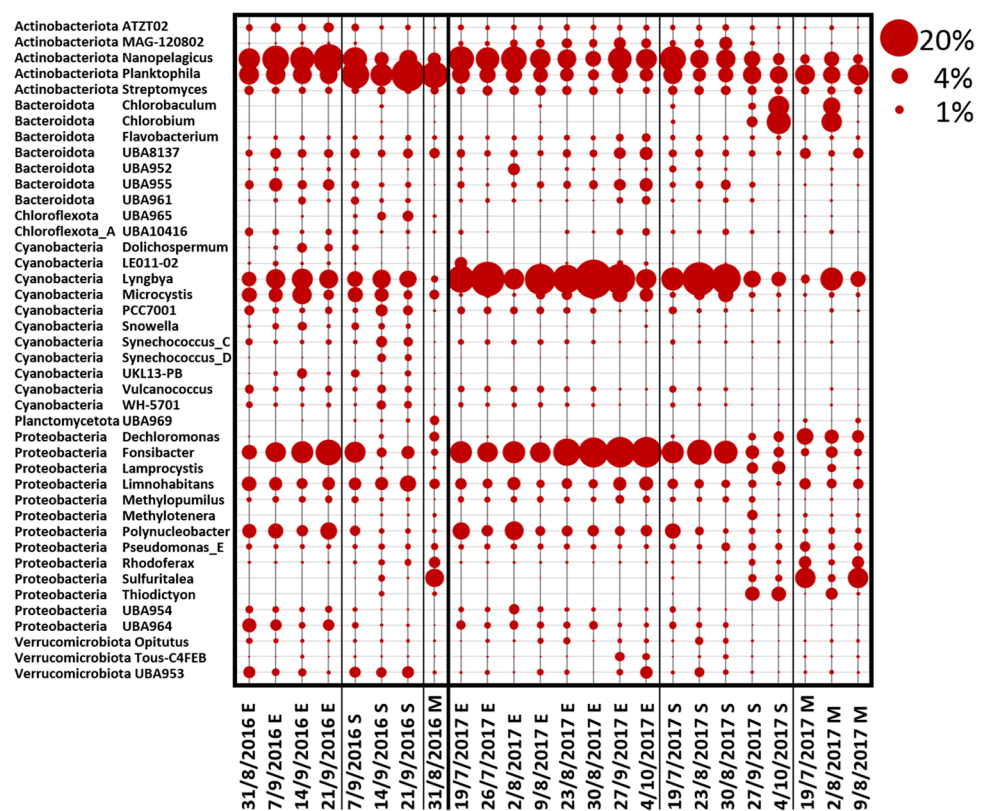
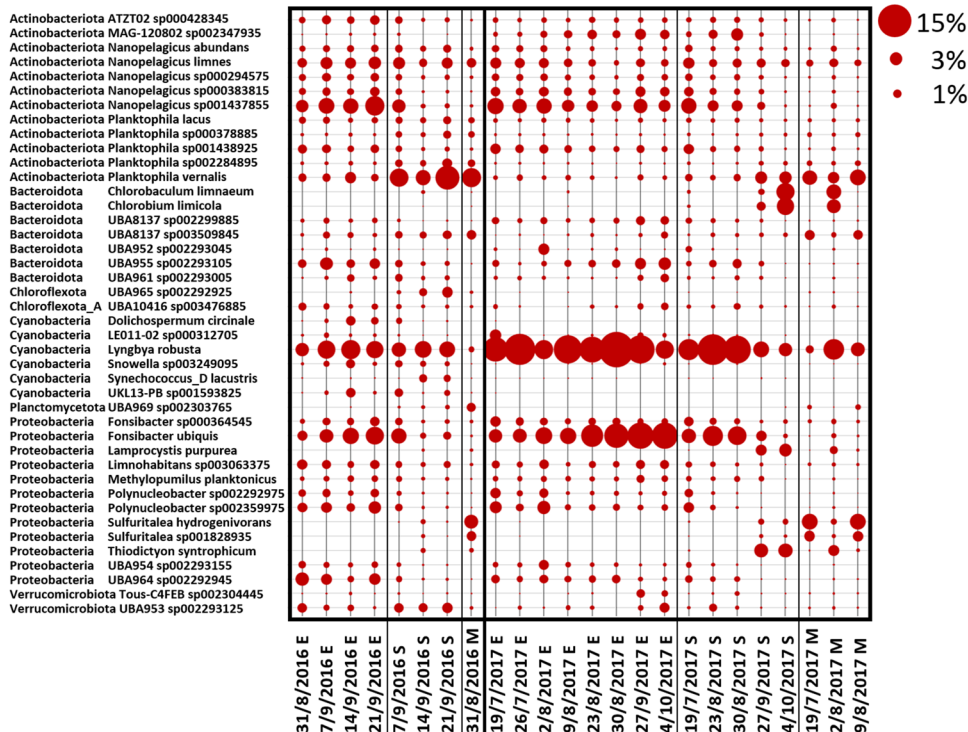


Fig. 6 Community composition at the species level as determined from the shotgun data. The figure shows variations in the microbial community at the species level. For details, refer to the legend of Fig. 5



To complement the taxonomy-focused Kraken2 analysis, shotgun read pairs from all individual samples were submitted to MegaHit for contig assembly. Maximal length and total number of the assembled contigs varied considerably between samples (maximal length from 42,318 to 821,984, number of contigs from 63,170 to 475,171) (Online resource 1: Supplemental Table S8). Presence of highly similar contigs (longer than 5000 nucleotides, 99% sequence identity, match longer than 50% of query length) across all samples was then examined with BLAST searches (Online resource 1: Supplemental Table S9). Putative assignment of contigs to known GTDB genomes was also performed through a BLAST search (Online resource 1: Supplemental Table S10). Contigs matching the *Lyngbya robusta* genome (*Limnoraphis robusta* CS-951; NZ_LATL02000001.1; assembly GCF_000972705.2) were detected in many samples with the highest estimated abundance (MegaHit contig coverage) in the 30/8/2017 E sample, like in the Kraken2 analysis (Online resource 1: Supplemental Table S9). For the 2016 samples, with a lower number of read pairs, no contig with length > 10,000 nucleotides and a sequence similarity higher than 90% to *Lyngbya robusta* was detected.

An additional analysis was then performed with all read pair sequences that were assigned to *Lyngbya robusta* by Kraken2. The read sequences were subjected to a MegaHit assembly run and the resulting contigs (Online resource 1: Supplemental Table S11; 2124 contigs with length 200 to 78,260, N50 14,006 and total length 7,208,187 nucleotides) were confronted with the GTDB genome

database. A large portion of the contigs matched to *Lyngbya robusta* assembly as the best hit, with generally high sequence identity values (Online resource 1: Supplemental Table S11). All these results indicated that the Kraken2 taxonomy assignment, although most likely not perfect and less reliable at the lower ranks, had correctly identified a *Lyngbya* species closely related to *Lyngbya robusta*.

Finally, shotgun read pairs from the 2017 E, 2017 M and S and from the 2016 E, M and S samples were analysed with MetaBat2. Out of the 114 bins obtained for the 2017 E samples, only a small portion (total of 30) could be related to GTDB genomes (Online resource 1: Supplemental Table S12). The best annotated bin corresponded again to *Lyngbya robusta* (*Limnoraphis robusta*) with an estimated completeness of 98% (obtained by checkm) and an overall nucleotide identity of 96.35% determined by GTDBtk (Online resource 2). Instead, for the 2016 samples, *Lyngbya robusta* (*Limnoraphis robusta*) was present only with low relative abundance (Online resource 1: Supplemental Table S13 and Online resource 2).

A final analysis using virsorter2 revealed that a large portion of the sample metagenomes, in particular MetaBat2 bins with a relatively small overall length, are represented by phages (Online resource 1: Supplemental Tables S12 and S13), often corresponding to MetaBat2 bins with the highest relative abundance values. Complete results are available in Online resource 2. However, comparison of the “phage bins” with known phage sequences

using DIAMOND did not allow to more precisely identify their taxonomy (Online resource 2).

Functional Profile

Functional profiling of the shotgun samples, obtained by a DIAMOND blastx search against the bacterial subsection of the NCBI nr database and MEGAN6 mapping to the SEED database [62], showed, at the SEED level 1, an overall rather constant profile during the 2016 and 2017 seasons (Online resource 1: Supplemental Fig. S5).

Instead, changes in the composition of metabolic functions occurred at the SEED level 2 (Fig. 7). In particular, the low cyanobacterial levels observed at the M depths (Online resource 1: Supplemental Fig. S4) appear to be associated with a distinct functional profile compared to samples collected at E and S depths. Only two S samples (27/9/2017 and 4/10/2017) followed a similar pattern as M samples, probably due to the high sampling depth (8.75 and 11 m, respectively). Samples collected at E depth showed no marked differences in their functional profiles compared to most of the S samples (Fig. 7).

Looking at the M pattern, most of the metabolic functions were little represented at this water depth compared to the E and S regions. One abundant function at M depth, less pronounced compared to the other profiles, involves diguanylate cyclase phosphodiesterase (DGC-PDE) domains (GGDEF & EAL) comprising a PAS/PAC sensor. Previous studies have described the role of DGC's in cyclic dimeric GMP (c-di-GMP) synthesis [63], a second messenger implicated

in regulating different processes in bacteria (e.g. biofilm formation and virulence etc.) and its hydrolysis is mediated by phosphodiesterase (PDEs) [63]. In Cyanobacteria, high levels of intracellular c-di-GMP have been associated to a reduced cellular buoyancy [64]. Tuckerman et al. described a DGC protein controlled by oxygen in *Escherichia coli* [65]. Oxygen, as a pivotal ligand, was also found in *Ace-tobacter xylinum* by Chang et al. [66] as a switch off for cellulose synthesis in biofilm production as well as in the pathogen *Bordetella pertussis* with a similar function [67]. It is likely that the modulated expression of the DGC/PDE activity observed at the different water depths influences the cyanobacterial distribution in the water column or the activation of biofilm production in an oxygen scarce environment such as the M depth. Only a few reads matched genes related to the photosystem II process (e.g. photosystem II protein D1; see Fig. 7) and the highest relative abundance was identified in reads assigned to the periplasmic nitrate reductase (Nap) precursor, to the CoB-CoM heterodisulfide reductase subunit A (HdrA) and sulphate permease (SuLP) activities. Nap is a protein expressed only in Proteobacteria [68], the predominant phylum at all depths, and its activity is associated to the reduction of nitrate [69–71], while the modulation in the CoB-CoM HdrA and SuLP may indicate a role in methanogenesis and sulphate reduction activities, respectively [72].

Furthermore, examination of the *Lyngbya robusta* genome annotation revealed an extended region comprising nitrogenase and nitrogen-fixing genes conserved in the

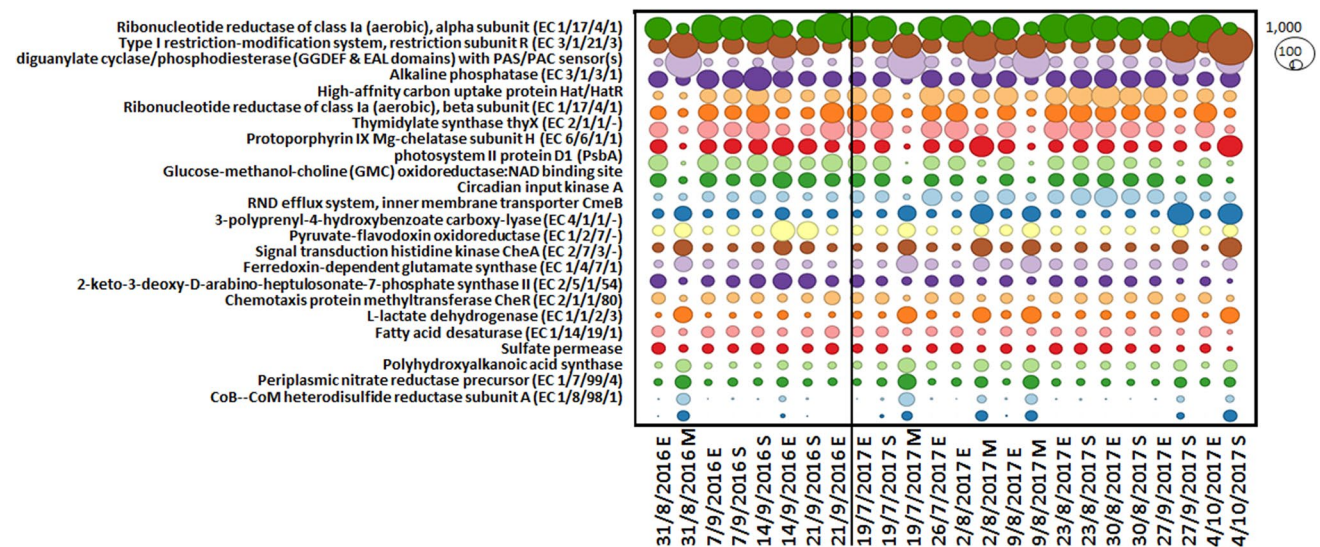


Fig. 7 Functional profile (SEED level 2) of the microbial community (shotgun samples). A representation of the SEED level 2 functional profiles as determined by MEGAN6 from DIAMOND blastx searches. Circles shown are proportional to the number of matches

against the indicated SEED level 2 class. The figure shows the SEED functions with more than 2000 matches across all samples and with an at least threefold difference between the maximum and minimum values observed. E stands for Epi, M for Meso and S for 2.5x Secchi

MetaBat2 bin closely related to *Lyngbya robusta* (Online resource 1: Supplemental Table S14).

Discussion

The aim of this study was to investigate the microbial community composition in Lake Varese and temporal and spatial changes during the bloom events, captured in the sampling campaigns performed in 2016 and 2017. In 2016, metagenomics 16S analysis revealed that major differences in the bacterial community composition during the study period were concentrated more in the S depth area compared to the E and M zones. Cyanobacteria were highly abundant particularly in the surface layers (E and S depths), with low levels recorded in M samples, as confirmed as well by the RDA analysis (Fig. 4 and Online resource 1: Supplemental Fig. S3 and S4). Indeed, the visual inspection of the water during the sampling campaign indicated that the bloom was characterized by a green discoloration of the water due to large growth of cyanobacteria and it was very dense. These data would then suggest that Cyanobacteria were likely responsible for the bloom and the *Chl a* peak observed in September 2016 (Fig. 2a). However, a complementary 18S metagenomics sequencing and a microscopy analysis would be recommended in order to discard a possible contribution of phototrophic eukaryotes during blooms in 2016 and 2017. A complementary analysis of the shotgun data using plant sequence data did not reveal a significant presence of algae (data not shown).

Looking at the community composition, the 16S metagenomic analysis revealed that Actinobacteriota, Proteobacteria, Bacteroidota and Verrucomicrobiota were the four phyla also dominating the community profile at all the three depths (Online resource 1: Supplemental Fig. S3) in accordance with the reported presence of bacteria belonging to these phyla in the bloom community [73–77]. In 2017, oscillations in cyanobacterial abundance and their visual observation in water indicate that Lake Varese experienced a continuous cyanobacterial bloom until a reduction starting from September. The microbial community observed at the three water depths was dominated by the same phyla already detected in 2016 (Proteobacteria, Actinobacteriota, Cyanobacteria and Bacteroidota). 16S data showed that the cyanobacterial abundance in 2017 was confined to the epilimnion (Supplemental Fig. S3) and in the RDA analyses, cyanobacteria were more correlated to surface samples (E depth) showing higher values for the physico-chemical parameters OS, WT and pH in 2017 than in 2016.

Since only small changes in the microbial community population were observed at the M depth during the 2016 campaigns, only one sample was analysed for 16S at this

depth in 2017, while three samples were selected for the shotgun analysis. Phylum-level classification of the shotgun dataset for 2016 and 2017 reproduced relatively well the profile obtained by the 16S data. The overall differences observed in the samples are not surprising since the two sequencing methods are based on different methodological approaches, which can influence the data analysis. Although only a small fraction of shotgun reads was annotated at the genus and species level, *Lyngbya robusta* was identified for the first time in Lake Varese, representing the dominant cyanobacterial species, in particular in the 2017 samples (Fig. 6), by direct taxonomic binning and through metagenomic contig assembly. Other species, mainly detected at E depth, included *Planktophilia vernalis*, *Fonsibacter ubiqvis* and *Nanopelagicus sp001437855* (Fig. 6). *Lyngbya robusta* is a species involved in nitrogen fixation [57, 78–80] and its blooms have been observed in freshwater waterbodies including Lake Atitlan (Guatemala) where concentrations of NO_3^- and NH_4^+ were indicative of a N limitation, as also observed in Lake Varese, where these values were found to be below the LOD in most of the samples analysed (Online resource 1: Supplemental Table S5) [58]. While no reads assigned to the nitrogenase activity were detected by the SEED analysis (Fig. 7), examination of the *Lyngbya robusta* genome annotation revealed an extended region comprising nitrogenase and nitrogen-fixing genes conserved in the MetaBat2 bin closely related to *Lyngbya robusta* (Online resource 1: Supplemental Table S14). The heterotrophic bacteria identified in Lake Atitlan during these bloom phenomena included phyla also detected in Lake Varese such as Actinobacteria, Bacteroides, Proteobacteria and Firmicutes [79] whose putative role was mainly associated with an increased nitrogenase activity. In the same lake [79], analysis of the *nifH* gene profile indicated other potential nitrogen-fixing bacteria to be present, for example genus *Methylomonas*, also reported in Fig. 3. The nutrient content measured in Lake Varese could explain the competitive advantage of *Lyngbya robusta* over other cyanobacteria species such as *Microcystis*, a non-diazotrophic cyanobacteria which use N_2 -fixing cyanobacteria as a source of nitrogen [81]. The same mechanism could explain the presence in 2016/2017 samples of other non-diazotrophic cyanobacteria like *Synechococcus* and *Planktothrix*, which could be also dominant or co-dominant with nitrogen-fixing cyanobacteria [82–85]. Like these genera, *Microcystis* can be just as abundant as N_2 -fixing cyanobacteria under N-limited conditions [86, 87]. In the shotgun analysis, values of relative abundance observed on some dates in 2016 at E depth were similar between *Lyngbya* and *Microcystis*, while in 2017, very low levels of *Microcystis* were detected in most of the samples (Fig. 5). In the literature, there is no evidence showing that *Lyngbya*

robusta is a MCs-producer cyanobacterium. Their production most likely involve the genus *Microcystis* which is a well-known producer of these toxins [1] and has been also detected in the shotgun data (Fig. 5). Other cyanobacteria like *Planktothrix* could contribute to the MCs production [1]. *Planktothrix* was previously identified in Lake Varese during a bloom event [30] and was also found in the shotgun analysis (at the genus and species level), although at very low levels of relative abundance (< 1%).

The functional profile of the microbial community showed that during the two annual campaigns, most of the E and S samples had a similar abundance of reads assigned to the different biological processes (Fig. 7). Lack of marked changes in functional genes was probably due to the limited variations in taxonomic composition of the bacterial community at the higher taxonomic ranks. However, a specific pattern was found to be associated with all M samples analysed and two S samples (27/9/2017 and 4/10/2017) collected at high depth (8.75 and 11 m, respectively). The functional profile identified in these samples may reflect a community composition characterized by low levels of Cyanobacteria. A low relative abundance of reads assigned to the photosystem II process (photosystem II protein D1; see Fig. 7) is notable in M and some S samples and can be explained by the low representativeness of Cyanobacteria at high water depths where light radiation is scarce.

The high relative abundance of genes involved in the Nap precursor activity (Fig. 7), which is associated to the reduction of nitrate during nitrate respiration [69–71], could be due to the selective presence of the Nap protein in Proteobacteria [68], a phylum highly present at M depth, and by the lack of oxygen detected in the mesolimnion (Table 1, Fig. 7, Online resource 1: Supplemental Tables S3 and S4, Supplemental Fig. S4). In addition, the presence of ammonia up to 0.831 mg/L in the M anoxic environment (Online resource 1: Supplemental Table S5) supports nitrate reduction. The relative abundance of reads assigned to the CoB-CoM HdrA activity was observed to be high at M depth (Fig. 7) and may indicate methanogenesis functions even though Archaea have not been identified through shotgun. Although Hdr has been detected only in methanogenes [72], homologous proteins show a widespread distribution and the “bacterial heterosulfide” Dsrc has been detected in *Desulfovibrio* spp. [88]. In this study, *Desulfovibrio* spp. (classified under the phylum Desulfobacterota) were found in high abundance at M depth (Online resource 1: Supplemental Fig. S3) where oxygen was absent, and negative values for the ORP were found (Online resource 1: Supplemental Table S3). Desulfobacterota are anaerobic bacteria which gain energy by sulphate reduction or sulphur disproportionation [89, 90]. Interestingly, an activity which is highly represented in all M samples and can be related to the sulphate reduction process involves the SulP family (Fig. 7), found to be active in

Desulfobacterota [91]. In this functional mechanism, sulphate has to be imported inside the cell and then activated with adenosine triphosphate (ATP) in order to be reduced to sulphide by the sulphate respiratory pathway [92]. In the shotgun datasets, reads corresponding to sulphur-oxidizing bacteria, *Sulfuritalea hydrogenivorans*, *Thiodictyon syntrophicum* and *Chlorobium*, were also detected (Fig. 6). *Thiodictyon* spp. were already found throughout the anoxic zones of water column in a freshwater lake [93], and the strain *Sulfuritalea hydrogenivorans* sk43H was identified in stratified lakes [60]. *Sulfuritalea hydrogenivorans*, the major planktonic sulphur oxidizer detected at M depth, is facultatively anaerobic but an autotrophic growth is also observed under anoxic condition in which hydrogen, elemental sulphur and thiosulphate are used as electron donor, and nitrate as electron acceptor [94]. Moreover, the genus *Sulfuritalea* is known to express the *dsrA* gene encoding dissimilatory sulphite reductase [95]. We also detected methanotrophs and methylotrophs (*Methylomonas*, *Methylobacter* and *Methylopusillus*) in our samples, which are bacteria that use methane and single carbon compounds (i.e. methanol), respectively, as energy source [96].

Overall, the results of this study reveal the potential of high-throughput sequencing (HTS) in obtaining a snapshot of the bacterial community structure, only partially investigated during previous studies performed in Lake Varese, providing more details regarding the temporal and spatial modulations during bloom events. In particular, the HTS approach allowed to identify for the first time *Lyngbya robusta* as the cyanobacterial species mainly responsible for the bloom. Further analyses aimed at correlating the relative abundance of cyanobacteria obtained by HTS with absolute quantitative data from microscopy analysis (or from other techniques like quantitative Polymerase Chain Reaction) will be necessary to confirm *Lyngbya robusta* as the most abundant species in this lake.

In addition, more investigations will be required to elucidate the relationships between *Lyngbya robusta* and cyanobacterial-associated or free-living communities of heterotrophic bacteria in this lake, including the possible contribution of the observed modulations in microbial functional profiles to the *Lyngbya robusta* bloom. In order to better understand the role of heterotrophic bacteria in the dynamics of cyanobacterial blooms, several studies have adopted a co-culture system, which however is not fully representative of the natural environment [7, 8, 97, 98]. The co-occurrence of heterotrophic bacteria and prokaryotic phytoplankton at species level is difficult to assess and compare to other studies since it is well known that only a relatively small amount of 16S or shotgun data at the species level can be taxonomically classified. For what concerns phytoplankton, comparison of shotgun reads (using Kraken2) with the Genbank plant section, diatom genomes and green

algae genomes, allowed only a very low fraction of reads (<0.1%) to be classified as phytoplankton. However, we cannot a priori exclude its presence because the analysis might be biased by a still very sparse level of genome sequence information concerning these organisms. Samples for a complementary microscopy analysis are unfortunately not available but might have indicated how deep the shotgun approach can provide a snapshot of the complete community, with presently available genome sequence information. For bacteria, co-occurrence network analysis (Online resource 1: Supplemental Fig. S6) of the 2017 Epilimnion genus level data (genera with $\geq 1\%$ relative abundance) indicates a single positive interaction between *Lyngbya* and *Limnohabitans* and both positive and negative interactions between *Microcystis* and a number of poorly characterized genera. At $2.5 \times$ Secchi depth (genera with $\geq 1\%$ relative abundance), the interaction network is more complex with both *Lyngbya* and *Microcystis* involved in numerous positive and negative interactions (Online resource 1: Supplemental Fig. S6). Evidence of heterotrophic bacteria degrading microcystin are reported by Christoffersen et al. [99] without presenting analysis of phylogeny, while their role in the production and degradation of *Microcystis* cyanopeptides is reported by Briand et al. [100] including a phylum microbial classification. At the genus level, some heterotrophic bacteria present with low (<1%) relative abundance in our analysis have been shown to enhance (e.g. *Flavobacterium* and *Pseudomonas*) but also inhibit (e.g. *Flavobacterium*) *Microcystis* growth while *Pseudomonas* has been described to have a cyanolytic capacity [101]. While our results have revealed a metagenome with a highly complex composition, including significant amounts of phages, they also show that a large portion of the metagenome still remains unclassified, even at higher taxonomic ranks, due to the presently still limited number of complete (or nearly complete) bacterial genomes and genome assemblies. From these results, it appears that also phages, with a rapidly growing number of known genomes [102], should be included in standard shotgun analysis methods such as Kraken2. We also identified for the first time a specific functional profile at M depth and, less frequently, in S samples suggesting that distinct metabolic processes characterize the bacterial community along the water column in Lake Varese. In particular, we observed that, among the modulated pathways, sulphur and nitrogen cycles are active processes in this lake, particularly in the M zone, characterized by poor or null oxygen content. In conclusion, the genetic and functional profile of the microbial community together with other components of the microbiome, and data on water nutrients and pollutants content could provide a more solid ground to understand the bloom dynamics, and later on, to contribute to the modelling studies for earlier prediction.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00248-021-01914-5>.

Acknowledgements We would like to acknowledge Elena Porcel Rodriguez, who was responsible for collecting and reorganizing all data and Ramón Pérez de Lara for producing Fig. 1.

Author Contribution Conceptualization: Teresa Lettieri; Experiments: Patrizia Pretto, Diana Conduto António and Isabella Sanseverino; Bioinformatics analysis: Armin Lahm; RDA analysis: Chiara Facca; Sampling campaigns: Robert Loos, Helle Skejo, Andrea Beghi, Franca Pandolfi and Pietro Genoni; Methodology: Patrizia Pretto, Diana Conduto António and Isabella Sanseverino; Project administration: Teresa Lettieri; Writing: Isabella Sanseverino, Armin Lahm and Teresa Lettieri. All authors have read and agreed to the published version of the manuscript.

Availability of Data and Material The datasets generated during the current study are available as raw FASTQ files in the NCBI SRA archive under project ID PRJNA694367 (Shotgun data: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA694367>) and PRJNA694444 (16S amplicon data: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA694444>).

Code Availability Not applicable.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Conflict of Interest The authors declare no competing interests.

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/>.

References

1. Sanseverino I, Antonio DC, Pozzoli L, Dobricic S, Lettieri T (2016) Algal bloom and its economic impact. EUR27905EN. <https://doi.org/10.2788/660478>
2. El-Shehawy R, Gorokhova E, Fernández-Piñas F, del Campo FF (2012) Global warming and hepatotoxin production by cyanobacteria: what can we learn from experiments? Water Res 46:1420–1429. <https://doi.org/10.1016/j.watres.2011.11.021>
3. Guedes IA, Rachid CTCC, Rangel LM, Silva LHS, Bisch PM, Azevedo SMFO, Pacheco ABF (2018) Close link between harmful cyanobacterial dominance and associated bacterioplankton in

- a tropical eutrophic reservoir. *Front Microbiol* 9. <https://doi.org/10.3389/fmicb.2018.00424>
4. Steffen MM, Li Z, Effler TC, Hauser LJ, Boyer GL, Wilhelm SW (2012) Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PLoS ONE* 7:e44002. <https://doi.org/10.1371/journal.pone.0044002>
 5. Bagatini IL, Eiler A, Bertilsson S, Klaveness D, Tessarolli LP, Vieira AA (2014) Host-specificity and dynamics in bacterial communities associated with Bloom-forming freshwater phytoplankton. *PLoS ONE* 9:e85950. <https://doi.org/10.1371/journal.pone.0085950>
 6. Osman OA, Beier S, Grabherr M, Bertilsson S (2017) Interactions of freshwater cyanobacteria with bacterial antagonists. *Appl Environ Microbiol* 83. <https://doi.org/10.1128/aem.02634-16>
 7. Xie M, Ren M, Yang C, Yi H, Li Z, Li T, Zhao J (2016) Metagenomic analysis reveals symbiotic relationship among bacteria in microcystis-dominated community. *Frontiers in Microbiology* 7. <https://doi.org/10.3389/fmicb.2016.00056>
 8. Ren M, Zhang G, Ye Z, Qiao Z, Xie M, Lin Y, Li T, Zhao J (2017) Metagenomic analysis reveals potential interactions in an artificial coculture. *AMB Express* 7:193. <https://doi.org/10.1186/s13568-017-0490-2>
 9. Jackrel SL, White JD, Evans JT, Buffin K, Hayden K, Sarnelle O, Deneff VJ (2019) Genome evolution and host-microbiome shifts correspond with intraspecific niche divergence within harmful algal bloom-forming *Microcystis aeruginosa*. *Mol Ecol* 28:3994–4011. <https://doi.org/10.1111/mec.15198>
 10. Wagner C, Adrian R (2009) Cyanobacteria dominance: quantifying the effects of climate change. *Limnol Oceanogr* 54:2460–2468. https://doi.org/10.4319/lo.2009.54.6_part_2.2460
 11. Taranu ZE, Zurawell RW, Pick F, Gregory-Eaves I (2012) Predicting cyanobacterial dynamics in the face of global change: the importance of scale and environmental context. *Glob Change Biol* 18:3477–3490. <https://doi.org/10.1111/gcb.12015>
 12. Kosten S, Huszar VLM, Bécares E, Costa LS, van Donk E, Hansson L-A, Jeppesen E, Kruk C, Lacerot G, Mazzeo N, De Meester L, Moss B, Lürling M, Nöges T, Romo S, Scheffer M (2012) Warmer climates boost cyanobacterial dominance in shallow lakes. *Glob Change Biol* 18:118–126. <https://doi.org/10.1111/j.1365-2486.2011.02488.x>
 13. Wells ML, Trainer VL, Smayda TJ, Karlson BSO, Trick CG, Kudela RM, Ishikawa A, Bernard S, Wulff A, Anderson DM, Cochlan WP (2015) Harmful algal blooms and climate change: learning from the past and present to forecast the future. *Harmful Algae* 49:68–93
 14. Paerl HW (2009) Controlling eutrophication along the freshwater–marine continuum: dual nutrient (N and P) reductions are essential. *Estuaries Coasts* 32:593–601. <https://doi.org/10.1007/s12237-009-9158-8>
 15. Conley DJ, Paerl HW, Howarth RW, Boesch DF, Seitzinger SP, Havens KE, Lancelot C, Likens GE (2009) Controlling eutrophication: nitrogen and phosphorus. *Science* 323:1014–1015. <https://doi.org/10.1126/science.1167755>
 16. Chirico N, António DC, Pozzoli L, Marinov D, Malagó A, Sanseverino I, Beghi A, Genoni P, Dobricic S, Lettieri T (2020) Cyanobacterial blooms in Lake Varese: analysis and characterization over ten years of observations. *Water* 12:675. <https://doi.org/10.3390/w12030675>
 17. Alvarenga DO, Fiore MF, Varani AM (2017) A metagenomic approach to cyanobacterial genomics. *Front Microbiol* 8:809. <https://doi.org/10.3389/fmicb.2017.00809>
 18. Qu J (2018) Dynamics of bacterial community diversity and structure in the terminal reservoir of the south-to-north water diversion project in China. *Water* v. 10: 2018 v.2010 no.2016. <https://doi.org/10.3390/w10060709>
 19. Scherer PI, Millard AD, Miller A, Schoen R, Raeder U, Geist J, Zwirgmaier K (2017) Temporal dynamics of the microbial community composition with a focus on toxic cyanobacteria and toxin presence during harmful algal blooms in two South German Lakes. *Front Microbiol* 8:2387. <https://doi.org/10.3389/fmicb.2017.02387>
 20. Parulekar NN, Kolekar P, Jenkins A, Kleiven S, Utkilen H, Johansen A, Sawant S, Kulkarni-Kale U, Kale M, Sæbø M (2017) Characterization of bacterial community associated with phytoplankton bloom in a eutrophic lake in South Norway using 16S rRNA gene amplicon sequence analysis. *PLoS ONE* 12:e0173408. <https://doi.org/10.1371/journal.pone.0173408>
 21. Zhu B, Cao H, Li G, Du W, Xu G, Domingo JS, Gu H, Xu N, Duan S, Lu J (2019) Biodiversity and dynamics of cyanobacterial communities during blooms in temperate lake (Harsha Lake, Ohio, USA). *Harmful Algae* 82:9–18. <https://doi.org/10.1016/j.hal.2018.12.006>
 22. Kurilkina MI, Zakharova YR, Galachyants YP, Petrova DP, Bukin YS, Domysheva VM, Blinov VV, Likhoshway YV (2016) Bacterial community composition in the water column of the deepest freshwater Lake Baikal as determined by next-generation sequencing. *FEMS Microbiol Ecol* 92. <https://doi.org/10.1093/femsec/fiw094>
 23. Fortin N, Munoz-Ramos V, Bird D, Lévesque B, Whyte LG, Greer CW (2015) Toxic cyanobacterial bloom triggers in Misisquoi Bay, Lake Champlain, as determined by next-generation sequencing and quantitative PCR. *Life* (Basel) 5:1346–1380. <https://doi.org/10.3390/life5021346>
 24. Zaccara S, Canziani A, Roella V, Crosa G (2007) A northern Italian shallow lake as a case study for eutrophication control. *Limnology* 8:155–160. <https://doi.org/10.1007/s10201-007-0209-1>
 25. Premazzi G, Dalmiglio A, Cardoso AC, Chiaudani G (2003) Lake management in Italy: the implications of the Water Framework Directive. *Lakes Reserv Res Manag* 8:41–59. <https://doi.org/10.1046/j.1440-1770.2003.00210.x>
 26. Casamitjana X, Serra T, Colomer J, Baserba C, Pérez-Losada J (2003) Effects of the water withdrawal in the stratification patterns of a reservoir. *Hydrobiologia* 504:21–28. <https://doi.org/10.1023/B:HYDR.0000008504.61773.77>
 27. Morabito G, Hamza W, Ruggiu D (2004) Carbon assimilation and phytoplankton growth rates across the trophic spectrum: an application of the chlorophyll labelling technique. *J Limnol* 63:33–43. <https://doi.org/10.4081/jlimnol.2004.33>
 28. Mosello R, Panzani P, Pugnetti A, Ruggiu D (1991) An assessment of the hydrochemistry of the eutrophic Lake Varese (N. Italy), coincident with the implementation of the first restoration measures. *Memorie dell'Istituto italiano di idrobiologia* 49:99–116
 29. Ruggiu D, Saraceni C, Mosello R (1981) Fitoplancton, produzione primaria e caratteristiche chimiche di un lago fortemente eutrofizzato: il Lago di Varese. *Memorie dell'Istituto Italiano di Idrobiologia* 39:47–64
 30. Pomati F, Sacchi S, Rossetti C, Giovannardi S, Onodera H, Oshima Y, Neilan BA (2000) The freshwater cyanobacterium *Planktothrix* sp. FP1: molecular identification and detection of paralytic shellfish poisoning toxins. *J Phycol* 36:553–562. <https://doi.org/10.1046/j.1529-8817.2000.99181.x>
 31. Ambrosetti W, Barbanti L, Sala N (2003) Residence time and physical processes in lakes. *J Limnol* 62:1–15. <https://doi.org/10.4081/jlimnol.2003.s1.1>
 32. HMSO (1983) Methods for the examination of waters and associated materials. Chapter: the determination of chlorophyll a in aquatic environments 1980. London: HMSO
 33. Nöges P, Nöges T, Ghiani M, Sena F, Fresner R, Friedl M, Mildner J (2011) Increased nutrient loading and rapid

- changes in phytoplankton expected with climate change in stratified South European lakes: sensitivity of lakes with different trophic state and catchment properties. *Hydrobiologia* 667:255–270. <https://doi.org/10.1007/s10750-011-0649-9>
34. Kisand V, Valente A, Lahm A, Tanet G, Lettieri T (2012) Phylogenetic and functional metagenomic profiling for assessing microbial biodiversity in environmental monitoring. *PLoS ONE* 7:e43630. <https://doi.org/10.1371/journal.pone.0043630>
 35. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41:e1–e1. <https://doi.org/10.1093/nar/gks808>
 36. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
 37. Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
 38. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
 39. Edgar RC (2016) SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *bioRxiv*: 074161. <https://doi.org/10.1101/074161>
 40. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004. <https://doi.org/10.1038/nbt.4229>
 41. Wood DE, Lu J, Langmead B (2019) Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
 42. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
 43. Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>
 44. Huson DH, Beier S, Flade I, Górská A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu R (2016) MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLOS Computational Biology* 12:e1004957. <https://doi.org/10.1371/journal.pcbi.1004957>
 45. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goessmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33:5691–5702. <https://doi.org/10.1093/nar/gki866>
 46. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z (2019) MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>
 47. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
 48. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
 49. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH (2020) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
 50. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S (2021) VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9:37. <https://doi.org/10.1186/s40168-020-00990-y>
 51. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>
 52. Faust K, Raes J (2016) CoNet app: inference of biological association networks using Cytoscape. *F1000Res* 5:1519. <https://doi.org/10.12688/f1000research.9050.2>
 53. Ter Braak CJF, Šmilauer P (2012) Canoco reference manual and user's guide: software for ordination, version 5.0. Microcomputer Power, Ithaca USA
 54. Dadheech PK, Glockner G, Casper P, Kotut K, Mazzoni CJ, Mbedi S, Krienitz L (2013) Cyanobacterial diversity in the hot spring, pelagic and benthic habitats of a tropical soda lake. *FEMS Microbiol Ecol* 85:389–401. <https://doi.org/10.1111/1574-6941.12128>
 55. Ezzedine JA, Jacas L, Desdevises Y, Jacquet S (2020) *Bdellovibrio* and like organisms in Lake Geneva: an unseen elephant in the room? *Frontiers in Microbiology* 11. <https://doi.org/10.3389/fmicb.2020.00098>
 56. Willis A, Parks M, Burford MA (2015) Draft genome assembly of filamentous brackish cyanobacterium *Limnoraphis robusta* strain CS-951. *Genome Announc* 3:e00846–e815. <https://doi.org/10.1128/genomeA.00846-15>
 57. Komarek J, Zapomelova E, Smarda J, Kopecky J, Rejmankova E, Woodhouse J, Neilan AB, Komarkova J (2013) Polyphasic evaluation of *Limnoraphis robusta*, a water-bloom forming cyanobacterium from Lake Atitlan, Guatemala, with a description of *Limnoraphis* gen. nov. *Journal of the Czech Phycological Society* 13:39–52. <https://doi.org/10.5507/fot.2013.004>
 58. Rejmánková E, Komárek J, Dix M, Komárková J, Girón N (2011) Cyanobacterial blooms in Lake Atitlan, Guatemala. *Limnologica* 41:296–302. <https://doi.org/10.1016/j.limno.2010.12.003>
 59. Kurobe T, Baxa DV, Mioni CE, Kudela RM, Smythe TR, Waller S, Chapman AD, Teh SJ (2013) Identification of harmful cyanobacteria in the Sacramento-San Joaquin Delta and Clear Lake, California by DNA barcoding. *Springerplus* 491. <https://doi.org/10.1186/2193-1801-2-491>
 60. Kojima H, Watanabe T, Iwata T, Fukui M (2014) Identification of major planktonic sulfur oxidizers in stratified freshwater lake. *PLoS ONE* 9:e93877. <https://doi.org/10.1371/journal.pone.0093877>
 61. Danza F, Ravasi D, Storelli N, Roman S, Lüdin S, Bueche M, Tonolla M (2018) Bacterial diversity in the water column of meromictic Lake Cadagno and evidence for seasonal dynamics. *PLoS ONE* 13:e0209743. <https://doi.org/10.1371/journal.pone.0209743>
 62. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>
 63. Tamayo R, Pratt JT, Camilli A (2007) Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu Rev Microbiol*

- 61:131–148. <https://doi.org/10.1146/annurev.micro.61.080706.093426>
64. Agostoni M, Waters CM, Montgomery BL (2016) Regulation of biofilm formation and cellular buoyancy through modulating intracellular cyclic di-GMP levels in engineered cyanobacteria. *Biotechnol Bioeng* 113:311–319. <https://doi.org/10.1002/bit.25712>
 65. Tuckerman JR, Gonzalez G, Sousa EHS, Wan X, Saito JA, Alam M, Gilles-Gonzalez M-A (2009) An oxygen-sensing diguanylate cyclase and phosphodiesterase couple for c-di-GMP control. *Biochemistry* 48:9764–9774. <https://doi.org/10.1021/bi901409g>
 66. Chang AL, Tuckerman JR, Gonzalez G, Mayer R, Weinhouse H, Volman G, Amikam D, Benziman M, Gilles-Gonzalez M-A (2001) Phosphodiesterase A1, a regulator of cellulose synthesis in acetobacter xylinum, is a heme-based sensor. *Biochemistry* 40:3420–3426. <https://doi.org/10.1021/bi0100236>
 67. Wan X, Tuckerman JR, Saito JA, Freitas TA, Newhouse JS, Denery JR, Galperin MY, Gonzalez G, Gilles-Gonzalez MA, Alam M (2009) Globins synthesize the second messenger bis-(3'-5')-cyclic diguanosine monophosphate in bacteria. *J Mol Biol* 388:262–270. <https://doi.org/10.1016/j.jmb.2009.03.015>
 68. Bru D, Sarr A, Philippot L (2007) Relative abundances of proteobacterial membrane-bound and periplasmic nitrate reductases in selected environments. *Appl Environ Microbiol* 73:5971–5974. <https://doi.org/10.1128/AEM.00643-07>
 69. Rütting T, Boeckx P, Müller C, Klemetsson L (2011) Assessment of the importance of dissimilatory nitrate reduction to ammonium for the terrestrial nitrogen cycle. *Biogeosciences* 8:1779–1791. <https://doi.org/10.5194/bg-8-1779-2011>
 70. Cojean ANY, Zopfi J, Gerster A, Frey C, Lepori F, Lehmann MF (2019) Direct O₂ control on the partitioning between denitrification and dissimilatory nitrate reduction to ammonium in lake sediments. *Biogeosciences* 16:4705–4718. <https://doi.org/10.5194/bg-16-4705-2019>
 71. Burgin AJ, Hamilton SK (2007) Have we overemphasized the role of denitrification in aquatic ecosystems? A review of nitrate removal pathways. *Front Ecol Environ* 5:89–96. [https://doi.org/10.1890/1540-9295\(2007\)5\[89:hwtro\]2.0.co;2](https://doi.org/10.1890/1540-9295(2007)5[89:hwtro]2.0.co;2)
 72. Kaster A-K, Moll J, Parey K, Thauer RK (2011) Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *PNAS* 108:2981–2986. <https://doi.org/10.1073/pnas.1016761108>
 73. Pope PB, Patel BKC (2008) Metagenomic analysis of a freshwater toxic cyanobacteria bloom. *FEMS Microbiol Ecol* 64:9–27. <https://doi.org/10.1111/j.1574-6941.2008.00448.x>
 74. Jones SE, Newton RJ, McMahon KD (2009) Evidence for structuring of bacterial community composition by organic carbon source in temperate lakes. *Environ Microbiol* 11:2463–2472. <https://doi.org/10.1111/j.1462-2920.2009.01977.x>
 75. Dziallas C, Grossart H-P (2011) Temperature and biotic factors influence bacterial communities associated with the cyanobacterium *Microcystis* sp. *Environ Microbiol* 13:1632–1641. <https://doi.org/10.1111/j.1462-2920.2011.02479.x>
 76. Šimek K, Horňák K, Jezbera J, Nedoma J, Vrba J, Straškrábová V, Macek M, Dolan JR, Hahn MW (2006) Maximum growth rates and possible life strategies of different bacterioplankton groups in relation to phosphorus availability in a freshwater reservoir. *Environ Microbiol* 8:1613–1624. <https://doi.org/10.1111/j.1462-2920.2006.01053.x>
 77. Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F (2014) Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. *Mol Ecol* 23:6073–6090. <https://doi.org/10.1111/mec.12985>
 78. Komárková J, Montoya H, Komárek J (2016) Cyanobacterial water bloom of *Limnospira robusta* in the Lago Mayor of Lake Titicaca. Can it develop? *Hydrobiologia* 764:249–258. <https://doi.org/10.1007/s10750-015-2298-x>
 79. Woodhouse JN, Ongley SE, Brown MV, Neilan BA (2013) Microbial diversity and diazotrophy associated with the freshwater non-heterocyst forming cyanobacterium *Lyngbya robusta*. *J Appl Phycol* 25:1039–1045. <https://doi.org/10.1007/s10811-012-9909-y>
 80. Omeregic EO, Crumbliss LL, Bebout BM, Zehr JP (2004) Determination of nitrogen-fixing phylotypes in *Lyngbya* sp. and microcoleus chthonoplastes cyanobacterial mats from Guerrero Negro, Baja California, Mexico. *Applied and Environmental Microbiology* 70:2119. <https://doi.org/10.1128/AEM.70.4.2119-2128.2004>
 81. Beversdorf LJ, Miller TR, McMahon KD (2013) The role of nitrogen fixation in cyanobacterial bloom toxicity in a temperate, eutrophic lake. *PLoS ONE* 8:e56103. <https://doi.org/10.1371/journal.pone.0056103>
 82. Jankowiak J, Hattenrath-Lehmann T, Kramer BJ, Ladds M, Gobler CJ (2019) Deciphering the effects of nitrogen, phosphorus, and temperature on cyanobacterial bloom intensification, diversity, and toxicity in western Lake Erie. *Limnol Oceanogr* 64:1347–1370. <https://doi.org/10.1002/lno.11120>
 83. Whitton BA, Potts M (2002) *The ecology of cyanobacteria*. Springer, Netherlands
 84. Davis TW, Bullerjahn GS, Tuttle T, McKay RM, Watson SB (2015) Effects of increasing nitrogen and phosphorus concentrations on phytoplankton community growth and toxicity during Planktothrix blooms in Sandusky Bay, Lake Erie. *Environ Sci Technol* 49:7197–7207. <https://doi.org/10.1021/acs.est.5b00799>
 85. Agawin NSR, Rabouille S, Veldhuis MJW, Servatius L, Hol S, van Overzee HMJ, Huisman J (2007) Competition and facilitation between unicellular nitrogen-fixing cyanobacteria and non-nitrogen-fixing phytoplankton species. *Limnol Oceanogr* 52:2233–2248. <https://doi.org/10.4319/lno.2007.52.5.2233>
 86. Paerl HW, Paul VJ (2012) Climate change: links to global expansion of harmful cyanobacteria. *Water Res* 46:1349–1363. <https://doi.org/10.1016/j.watres.2011.08.002>
 87. Smith VH (1983) Low nitrogen to phosphorus ratios favor dominance by blue-green algae in lake phytoplankton. *Science* 221:669–671. <https://doi.org/10.1126/science.221.4611.669>
 88. Venceslau SS, Stockdreher Y, Dahl C, Pereira IAC (2014) The “bacterial heterodisulfide” DsrC is a key protein in dissimilatory sulfur metabolism. *Biochem Biophys Acta* 1837:1148–1164. <https://doi.org/10.1016/j.bbabi.2014.03.007>
 89. Ward LM, Bertran E, Johnston DT (2020) Draft genome sequence of *Desulfovibrio sulfodimittens* ThAc01, a heterotrophic sulfur-disproportionating member of the desulfobactera. *Microbiology Resource Announcements* 9:e00202–00220. <https://doi.org/10.1128/MRA.00202-20>
 90. Finster K (2008) Microbiological disproportionation of inorganic sulfur compounds. *J Sulfur Chem* 29:281–292. <https://doi.org/10.1080/17415990802105770>
 91. Marietou A, Røy H, Jørgensen BB, Kjeldsen KU (2018) Sulfate transporters in dissimilatory sulfate reducing microorganisms: a comparative genomics analysis. *Front Microbiol* 9. <https://doi.org/10.3389/fmicb.2018.00309>
 92. Santos AA, Venceslau SS, Grein F, Leavitt WD, Dahl C, Johnston DT, Pereira IAC (2015) A protein trisulfide couples dissimilatory sulfate reduction to energy conservation. *Science* 350:1541. <https://doi.org/10.1126/science.aad3558>
 93. Biderre-Petit C, Boucher D, Kuever J, Alberic P, Jézéquel D, Chebance B, Borrel G, Fonty G, Peyret P (2011) Identification of sulfur-cycle prokaryotes in a low-sulfate lake (Lake Pavin) using aprA and 16S rRNA gene markers. *Microb Ecol* 61:313–327. <https://doi.org/10.1007/s00248-010-9769-4>

94. Kojima H, Fukui M (2011) *Sulfuritalea hydrogenivorans* gen. nov., sp. nov., a facultative autotroph isolated from a freshwater lake. *Int J Syst Evol Microbiol* 61:1651–1655. <https://doi.org/10.1099/ijs.0.024968-0>
95. Luo J, Tan X, Liu K, Lin W (2018) Survey of sulfur-oxidizing bacterial community in the Pearl River water using *soxB*, *sqr*, and *dsrA* as molecular biomarkers. *3 Biotech* 8:73. <https://doi.org/10.1007/s13205-017-1077-y>
96. Salcher MM, Neuenschwander SM, Posch T, Pernthaler J (2015) The ecology of pelagic freshwater methylophils assessed by a high-resolution monitoring and isolation campaign. *ISME J* 9:2442–2453. <https://doi.org/10.1038/ismej.2015.55>
97. Parveen B, Ravet V, Djediat C, Mary I, Quiblier C, Debroas D, Humbert JF (2013) Bacterial communities associated with *Microcystis* colonies differ from free-living communities living in the same ecosystem. *Environ Microbiol Rep* 5:716–724. <https://doi.org/10.1111/1758-2229.12071>
98. Louati I, Pascual N, Debroas D, Bernard C, Humbert J-F, Leloup J (2015) Structural diversity of bacterial communities associated with bloom-forming freshwater cyanobacteria differs according to the cyanobacterial genus. *PLoS ONE* 10:e0140614. <https://doi.org/10.1371/journal.pone.0140614>
99. Christoffersen K, Lyck S, Winding A (2002) Microbial activity and bacterial community structure during degradation of microcystins. *Aquat Microb Ecol* 27:125–136
100. Briand E, Humbert J-F, Tambosco K, Bormans M, Gerwick WH (2016) Role of bacteria in the production and degradation of *Microcystis* cyanopeptides. *MicrobiologyOpen* 5:469–478. <https://doi.org/10.1002/mbo3.343>
101. Berg KA, Lyra C, Sivonen K, Paulin L, Suomalainen S, Tuomi P, Rapala J (2009) High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. *ISME J* 3:314–325. <https://doi.org/10.1038/ismej.2008.110>
102. Tisza MJ, Pastrana DV, Welch NL, Stewart B, Peretti A, Starrett GJ, Pang YS, Krishnamurthy SR, Pesavento PA, McDermott DH, Murphy PM, Whited JL, Miller B, Brenchley J, Rosshart SP, Rehermann B, Doorbar J, Ta'ala BA, Pletnikova O, Troncoso JC, Resnick SM, Bolduc B, Sullivan MB, Varsani A, Segall AM, Buck CB (2020) Discovery of several thousand highly diverse circular DNA viruses. *eLife* 9. <https://doi.org/10.7554/eLife.51971>