

RESEARCH ARTICLE OPEN ACCESS

Phylogenomic and Pangenomic Assessment of a Mediterranean Strain of *Raphidiopsis raciborskii* Extends Knowledge of the Global Distribution and Characteristics of a Potentially Toxigenic Cyanobacterium

Nico Salmaso^{1,2}  | Leonardo Cerasino¹ | Margherita Di Brizio³ | Massimo Pindo¹ | Adriano Boscaiini¹

¹Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Italy | ²NBFC, National Biodiversity Future Center, Palermo, Italy | ³UOLM Sito Perugia, ARPA Umbria, Perugia, Italy

Correspondence: Nico Salmaso (nico.salmaso@fmach.it)

Received: 13 February 2025 | **Revised:** 8 April 2025 | **Accepted:** 16 April 2025

Funding: This work was supported by the National Biodiversity Future Center and received funding from the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR)-Missione 4 Componente 2, Investimento 1.4-D.D. 1034 17/06/2022, CN00000033, CUPD43C22001280006).

Keywords: cyanotoxins | *Cylindrospermopsis* | genome mining | metabolic reconstruction | pangenomics | phylogenomics | *Raphidiopsis*

ABSTRACT

Among potentially toxigenic cyanobacteria, *Raphidiopsis raciborskii* has attracted considerable attention due to its ability to produce massive blooms and its recent spread to temperate regions. In this work, we reported for the first time a taxonomic and functional assessment of a *R. raciborskii* strain isolated from the Mediterranean region, contributing to filling a gap in the global distribution and characteristics of this species. The strain LT_0923 was isolated from Lake Trasimeno, a large and shallow lake in central Italy. The phylogenomic analyses based on selected marker genes and the core genome obtained from a pangenomic analysis based on a selection of available high-quality genomes showed a strong correspondence of the Lake Trasimeno strain with the North American and, at a lower average nucleotide identity, with the South American genomes. The LT_0923 genome did not show the presence of gene clusters encoding legacy cyanotoxins or emerging toxigenic compounds. The open pangenome and the large fraction of distinct gene families identified in the cloud and partly shell genome, enriched with genes specialised in environmental-specific functions and defence mechanisms, are consistent with the development of *Raphidiopsis* in geographically distinct regions.

1 | Introduction

Among the potentially toxigenic cyanobacteria, *Raphidiopsis raciborskii* (Woloszynska), Aguilera et al. has received rapid and increasing attention due to its recent spread to temperate regions and ability to develop massive blooms and toxic strains (Padisák 1997; Antunes et al. 2015).

This cyanobacterium was originally included in the genus *Cylindrospermopsis* and was distinguished from *Raphidiopsis* by its ability to produce heterocytes. Nevertheless, as this species is

characterised by great morphological diversity and phenotypic plasticity, in the absence of heterocytes it did not show any other clear diacritical features to distinguish it from *Raphidiopsis* (Komárek and Mareš 2012). In addition, several genetic and phylogenetic studies showed that species of *Raphidiopsis* and *Cylindrospermopsis* formed a polyphyletic group, suggesting the evolution of populations that have lost the ability to produce heterocytes (Moustaka-Gouni et al. 2009). Based on these and new evidence, Aguilera et al. (2018) proposed the unification of these two genera under the name *Raphidiopsis*, respecting the principle of priority (Komárek and Mareš 2012).

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Environmental Microbiology Reports* published by John Wiley & Sons Ltd.

Considered to be of tropical origin, the rapid spread of *R. raciborskii* to temperate regions has been documented in several countries (Padisák 1997). The first record of *R. raciborskii* in Europe was documented in Lake Kastoria, Greece (Skuja 1937) and then in Hungary. Since then, the cosmopolitan presence of this species has been increasingly reported in several countries of temperate regions (Antunes et al. 2015; Kokociński et al. 2017). In Italy, *R. raciborskii* was reported for the first time in 1995 in Lake Trasimeno (Manti et al. 2005), and then in several other lakes throughout Italy (Del Pasqua et al. 2024; Austoni et al. 2025).

Several phylogenetic studies based on individual marker genes highlighted the existence of different genetic populations among *R. raciborskii* from various geographical regions (Haande et al. 2008). Among the most recent investigations, Moreira et al. (2015) showed a congruence between strains from Europe, Asia and Australia, separated from the American and North African clusters, suggesting that the strains from Europe originated from Asia and Australia. After adding to the analyses other strains from southern Europe, Cirés et al. (2014) showed a strong link between the populations of Central and North America and Tunisia with strains isolated from Spain, all more distantly related to other strains from Portugal and Northern Europe, as well as other continents. These results, which demonstrated a genetic inhomogeneity of *R. raciborskii* populations in Europe, were further confirmed by Panou et al. (2018), who demonstrated that several Greek strains of this species formed a compact cluster with strains from Spain, Tunisia and America.

The success of *R. raciborskii* is attributed to its phenotypic plasticity and to the development of ecotypes adapted to local environmental constraints, which gave rise to the existence of populations with defined biogeographic patterns (Vico et al. 2021). One of the most striking differences between populations is the ability to evolve diverse strains capable of producing different types of cyanotoxins on different continents. Cylindrospermopsin (CYN) producing strains have been isolated from Australia (Saker and Griffiths 2000), New Zealand (Wood and Stirling 2003) and Asia (Li et al. 2001; Zarenezhad et al. 2012; Rzymiski and Poniedziałek 2014; Nguyen et al. 2017), while saxitoxins (STXs) producing strains have been isolated from South America (Lagos 2003; da Costa et al. 2013; Mesquita et al. 2019). Conversely, only very rare STX or CYN-producing strains were identified in Europe (Sha et al. 2025).

The full range of results reported above was based on target analyses performed on selected gene markers and analytical determinations on numerous isolated strains from different continents. As a downside, target approaches are necessarily restricted to a limited number of marker genes, with functional analyses limited to only a few selected metabolites.

Owing to a much more limited available number of strains, the relationships between *R. raciborskii* populations analysed using high-throughput sequencing and genomic approaches were restricted to selected regions, particularly Australia, Asia and South America (Stucken et al. 2010; Sinha et al. 2014; Abreu et al. 2018; Willis et al. 2018; Willis and Woodhouse 2020). Information from other American regions and Europe was limited to a few genomes, respectively from North America and Central Europe

(Vico et al. 2020; Laux et al. 2023; Tokodi et al. 2025), while no genomes are available for the Mediterranean regions. Given the strong relationships between *R. raciborskii* strains from North America and some Mediterranean regions based on the use of single marker genes, this still represents an important knowledge gap in the results obtained from phylogenomic and pangenomic analyses.

In this work, we report a taxonomic and functional assessment of a *R. raciborskii* strain isolated from Lake Trasimeno, a large and shallow lake in central Italy. The characteristics of this Mediterranean strain will be compared with an extended set of high-quality genomes including recently published strains from Central Europe and North America. Specific objectives of this work include: (i) the taxonomic and functional genomic analysis of the Lake Trasimeno strain; (ii) the phylogenomic characterisation of this Mediterranean strain in relation to representative high-quality genomes isolated worldwide; (iii) through a pangenomic analysis to evaluate the main differences in the entire set of encoded proteins among *R. raciborskii* strains; (iv) further attention will be given to the potential ability to synthesise cyanotoxins and other secondary metabolites.

2 | Materials and Methods

2.1 | Study Site, Sampling Stations and Field and Laboratory Measurements

Lake Trasimeno is a large ($\approx 124 \text{ km}^2$ and $586 \times 10^6 \text{ m}^3$) and shallow ($z_{\text{max}} \approx 6 \text{ m}$) lake located at 258 m a.s.l. in central Italy. The lake level can vary considerably in response to atmospheric precipitation, which influences the hydrodynamic conditions and the size of the lake. Lake Trasimeno is endorheic with no natural outlet, and an artificial canal flowing into the Tiber River. Trophic status is between mesotrophy and eutrophy (Ludovisi et al. 2021). Based on satellite data, since the 1980s the lake has shown a significant warming at a rate of around $0.017^\circ\text{C yr}^{-1}$ (Pareeth et al. 2017).

Sampling was carried out on September 12, 2023 at the boat pier located at Castiglione del Lago (CL) ($43^\circ 07' 23.6'' \text{ N}$, $12^\circ 03' 26.5'' \text{ E}$; Figure S1A). Samples were collected at the surface using a sterilised bucket and nets (mesh size $80 \mu\text{m}$). Subsamples were stored in refrigerated conditions until the isolation of *Raphidiopsis*. A subsample (250 mL) for phytoplankton analyses was fixed soon after collection with Lugol's solution and successively stored in the laboratory at 4°C . At the sampling point, water temperatures were measured using a handheld thermometer and transparency using a Secchi disk. More complete physical and chemical analyses, representative of the general conditions of the lake, were performed the day before (September 11, 2023) in two pelagic stations by the laboratories of the ARPA Umbria. The analyses and methods are publicly available at <https://www.arpa.umbria.it> (for details, see Figure S1A).

2.2 | Cyanotoxins Analyses

After filtration on GF/C filters of 0.4 L from water samples and 0.3 L from net samples, cyanotoxins were quantified as described

by Cerasino and Salmaso (2020) and Salmaso et al. (2024). Analytical determinations did include several microcystin (MCs) variants, anatoxins (ATXs; ATX-a and homoATX-a), cylindrospermopsin (CYN) and saxitoxins (STXs; STX, dc-STX, NeoSTX, GTX1, GTX4, GTX5, C1 and C2).

2.3 | Microscopical Observations

Field and culture samples were observed under an invertoscope Zeiss Axiovert 135 and an upright microscope Leica DM 2500. The counting and measurement of the *Raphidiopsis* cells from the field sample were performed using the Utermöhl method (Lund et al. 1958), as described in detail in Rott et al. (2007).

2.4 | Isolation, DNA Extraction and Sequencing

Single filaments of *Raphidiopsis* were isolated using micropipettes under stereomicroscopes within 1 day after the sampling, and isolates were incubated in Z8 medium (Kotai 1972) at 20°C under a 12h:12h light–dark photoperiod ($70\mu\text{mol m}^{-2}\text{s}^{-1}$) in climatic simulation chambers (ProClimatic, Italy); complete general isolation procedures are described in Salmaso et al. (2016).

The non-axenic culture of the *R. raciborskii* strain isolated from Lake Trasimeno was filtered through 8µm cellulose nitrate membranes (Cytiva) and DNA was extracted using the PowerWater Kit (Qiagen). The library was prepared using the Kapa HyperPlus Kit (Roche) and processed by 150-bp paired-end sequencing on an Illumina NovaSeq 6000 (Illumina Inc., San Diego, CA, USA) following the procedures described in Salmaso et al. (2024).

An aliquot of the extracted DNA from the isolate was used to amplify the 16S rRNA gene using the primers pA and B23S, successively sequenced using the internal primers 16S544R, 16S1092R and 16S979F, as described in Rajaniemi-Wacklin et al. (2005) and Salmaso et al. (2015) (GenBank accession number, PV054108).

2.5 | Draft Bacterial Genomes: Binning and Assembly

Assembling and binning followed the procedures described in detail in Salmaso et al. (2024), using the most up-to-date package versions. In short, raw reads were checked with FastQC 0.12.1 (github.com/s-andrews/FastQC). Removal of residual adapters and trimming were performed using bbduk 39.05 (<https://jgi.doe.gov>), keeping only reads with quality scores greater than 20 (trimq=20 and minavgquality=20). After removal of human reads contaminants (mapping of reads on the human reference genome GRCh38.p14 using Bowtie 2.5.2 and filtering with SAMtools 1.19) (Langmead and Salzberg 2012; Danecek et al. 2021), the successive analyses were performed on 10% of the quality checked and processed paired-end reads. Before assembling, a preliminary assessment of the taxonomic composition and relative abundance of the microbial community was carried out using MetaPhlAn 4.0.6 with the --unclassified_estimation parameter (Blanco-Míguez et al. 2023).

Paired reads were corrected with metaSPAdes 4.0.0 (--only-error-correction) (Nurk et al. 2017) and assembled with megahit 1.2.9 (--presets meta-sensitive) (Li et al. 2015). Before binning, contigs shorter than 1000bp were discarded and the names of the remaining contigs were simplified with anvio 8 (Eren et al. 2021). The resulting contigs were binned using CONCOCT 1.1.0 (Alneberg et al. 2014), MetaBAT 2.17–21 (Kang et al. 2019) and SemiBin2 2.1.0 (Pan et al. 2022). The resulting bins were then combined using DAS Tool 1.1.6 (Sieber et al. 2018), using default options, and assessed for the presence of chimerism using GUNC 1.0.6 (Orakov et al. 2021) and other sources of contamination using MDMcleaner 0.8.7 (Vollmers et al. 2022). The final *R. raciborskii* bin was checked and confirmed with anvio 8. Completeness and redundancy of the assemblies were estimated using CheckM2 1.0.2 (Chklovski et al. 2023) and coverages using CoverM 0.6.1 (github.com/wwood/CoverM).

The Genome Shotgun project obtained from the isolate has been deposited at DDBJ/ENA/GenBank under the project number PRJNA1215010.

2.6 | Phylogenomic, Functional and Pangenomic Analyses

Genomes to be compared with the *R. raciborskii* strain isolated from Lake Trasimeno were selected to cover all the *Raphidiopsis* species available in the Genome Taxonomy Database (GTDB; release 09-RS220) (Parks et al. 2022), integrated with assemblies from NCBI (Schoch et al. 2020). From this initial set, a subset of genomes with completeness greater than 94% and/or contamination less than 4% (as determined by CheckM2) and a number of contigs less than 1000 were retained for subsequent analyses (Table S1). Other redundant/duplicated genomes (GCA_000175835.1), genomes with length too small (less than 3 Mbp) and genomes with ambiguity codes were excluded from the analysis.

The *R. raciborskii* genome from Lake Trasimeno was compared with the selected *Raphidiopsis* spp. genomes using the average nucleotide identity (ANI) computed using average_nucleotide_identity.py and the ANI_b option from the package pyani 0.3.0-alpha (Pritchard et al. 2016); the reciprocal ANI_b comparisons between a couple of genomes were then averaged to obtain one unique value. ANI values between 0.95 and 0.96 have been proposed as a boundary to identify genomes belonging to the same species (Rosselló-Móra and Whitman 2019), while genomes of different species generally have ANI < 0.90 and ANI values in the range 0.90–0.95 are comparatively rare (Rodríguez-R et al. 2024). Phylogenomic analyses were carried out using GTDB-Tk 2.4.0 (identify, align, infer commands) (Chaumeil et al. 2022), based on 120 markers (Hidden Markov Models, HMMs) specifically designed for the analysis of bacteria, alignment using hmmalign (Eddy 2011) and tree building using FastTree (Price et al. 2010). The tree topology was confirmed using GToTree 1.8.8 based on pre-packaged HMM single-copy genes set specific for Cyanobacteria (251 genes) (Lee 2019), alignment using muscle 5.0 (Edgar 2022), and tree building using IQTree (-m MFP -B 1000) (Nguyen et al. 2015). In both cases, the trees were rooted using the python package ETE3

(Huerta-Cepas et al. 2016) and *Sphaerospermopsis kisseleviana* CS-549 (GCA_028329225.1) as an outgroup. The final trees were built and annotated using the R package ggtree 3.10.1 (Yu 2020).

Functional annotation of the *R. raciborskii* draft genome was performed using the NCBI stand-alone software package Prokaryotic Genome Annotation Pipeline (PGAP) version 2024-07-18.build7555 (Li et al. 2020) and finally confirmed by annotation using the PGAP service in NCBI (<https://www.ncbi.nlm.nih.gov/>). Annotations were complemented using bakta 1.9.4 (Schwengers et al. 2021). Target sequences for the corresponding intervals defined in the PGAP/bakta feature tables were extracted from the *R. raciborskii* genome using bedtools 2.31.1 (Quinlan and Hall 2010). Identification of ribosomal rRNA genes was further evaluated using Barrnap 0.9 (github.com/tseemann/barrnap). ARGs in all the *Raphidiopsis* genomes were predicted using AMRFinderPlus 4.0.3 with database version 2024-12-18.1 (Feldgarden et al. 2021) and ABRicate (version 1.0.1), incorporating the NCBI AMRFinder, ARG-ANNOT, ResFinder and Card databases (github.com/tseemann/abricate) and using minimum DNA identity and coverage values of 80% and 50%, respectively.

Metabolism and phenotypic features of the *Raphidiopsis* genomes were identified using the metabolic reconstruction set of programs in *anvi'o* 8 (Eren et al. 2021; Watson et al. 2023), with the options -E 1e-15 and -H 0.95 in *anvi-run-kegg-kofams*. Functional annotations were based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2014). K numbers were assigned to sequence data by HMMER/HMMSEARCH against Kofam, a customised HMM database of KEGG Orthologs (KOs) (Aramaki et al. 2020). KOs functional annotations were then matched to metabolic pathways (modules) (Kanehisa and Sato 2020) and their completeness was estimated using the program *anvi-estimate-metabolism*. Functional enrichment was finally computed to check if specific modules were significantly enriched in contrasting groups of genomes (Shaiber et al. 2020).

Pangenomic analysis was performed including all the *Raphidiopsis* genomes following the *anvi'o* 8 pangenomic workflow (Delmont and Eren 2018; Eren et al. 2021). After associating genes in the respective *anvi'o* contigs database with COGs (COG20) functions (Tatusov et al. 1997), a pangenome was generated using *anvi-pan-genome* with the option --mcl-inflation 7 (van Dongen and Abreu-Goodger 2012) to identify gene clusters. Evolutionary associations in the pangenome were assessed by computing a phylogenomic analysis using 1526 single-copy core genes (SCGs) across all genomes. After removing gap characters in more than 50% of the sequences using trimAl v1.4.rev15, a tree was built using IQTree (-m MFP) and midpoint rooted using the ETE3 package. The phylogenomic tree based on the SCGs was compared with the corresponding GTDB-Tk tree by means of a co-phylogenetic tree using Phytools 2.4.4 (Revell 2024). A parallel pangenomic analysis was further performed using panaroo 1.5.1 (Tonkin-Hill et al. 2020). A heatmap showing the distribution of the main COGs categories was performed using the R package heatmaply 1.5.0. To estimate if the *Raphidiopsis* pangenome was open or closed, the Heaps law was applied to the results obtained with *anvi'o* and panaroo; in the Heaps model, the relationship between new genes discovered with the addition of new genome sequences is modelled by a power-law

function, $n = kN^\gamma$, where n = pan-genome size, N = number of genomes and k and γ are fitting parameters (Tettelin et al. 2008). The pangenome is closed and approaches a constant size as more genomes are used when $\gamma < 0$, whereas when $\gamma > 0$, the pangenome is open and increases in size as more genomes are included. Heaps parameters were estimated with 1000 genome random samples using the *specaccum* function in the R package *vegan* 2.6.10 (Oksanen et al. 2020), and a least-squares fit of the power law to the medians calculated for each distribution, following the approach of Tettelin et al. (2008).

The presence of secondary metabolite biosynthetic gene clusters (BGCs) in all the *Raphidiopsis* genomes was estimated using antiSMASH 7.1.0 (default mode) (Blin et al. 2023). The presence of selected genes encoding MC, ATX-a, CYN and STX, as well as odorous compounds (geosmin) was tested using ISeqDb 3.0 (github.com/hts-tools/iseqdb).

3 | Results

3.1 | Environmental Data

In the littoral sampling site, surface water temperatures and transparency were 24.6°C and 0.3m, respectively. The general condition of the lake during sampling is exemplified by the data collected by ARPA Umbria in the two pelagic stations, which showed equivalent temperature and transparency values to those measured on the littoral shore (Table 1; Figure S1). The high trophic status of the lake is confirmed by the high concentrations of SRP and TP (17–18 and 70–80 µg P L⁻¹, respectively), chlorophyll a (44–50 µg L⁻¹), pH (9) and dissolved oxygen (110%–126%). Conversely, nitrogen compounds always showed concentrations < 0.1 mg L⁻¹ (NO₃-N and NH₄-N) and < 2 mg L⁻¹ (total nitrogen).

3.2 | Cyanotoxins

The LC-MS analyses did not show any quantifiable presence of cyanotoxins in all the water and net samples analysed.

3.3 | Light Microscopy

The filaments observed in the environmental samples collected in the littoral station and the isolated strain (Figure S1B) showed the diacritical characters typical of *R. raciborskii* (Manti et al. 2005; Mugnai et al. 2008; Karadžić et al. 2013; Komárek 2013). In the lake littoral sample, the length of the *Raphidiopsis* filaments was generally between 40 and 200 µm. The density was high (164,000 cells mL⁻¹), with akinetes present in around 10% of the filaments and with length 8–15 µm and width 3–4 µm. Heterocytes were always terminal and present in around 7% of the filaments.

3.4 | Lake Trasimeno *Raphidiopsis* Genome

Novaseq sequencing of the genomic DNA extracted from the *Raphidiopsis* culture generated 89,169,269 paired-end reads

TABLE 1 | Physical and chemical characteristics of surface samples collected on September 11, 2023 in the two pelagic stations of Lake Trasimeno (Figure S1).

Sampling station	Passignano sul Trasimeno	
	Magione	
Latitude	43°09'17.1" N	43°06'18.0" N
Longitude	12°06'46.6" E	12°09'34.0" E
Temperature (°C)	25.0	25.0
Conductivity (µS cm ⁻¹) 20°C	1826	1841
pH	9.0	9.0
Oxygen (mg L ⁻¹)	10.1	8.8
Nitrate nitrogen, NO ₃ -N (mg L ⁻¹)	<0.10	<0.10
Ammonium nitrogen, NH ₄ -N (mg L ⁻¹)	<0.04	<0.04
Total nitrogen, TN (mg L ⁻¹)	1.3	1.7
Soluble reactive phosphorus, SRP (µg L ⁻¹)	17	18
Total phosphorus, TP (µg L ⁻¹)	80	70
Chlorophyll a (µg L ⁻¹)	43.9	50.0
Secchi disk depth (m)	0.4	0.3

with 97.3% and 92.4% of bases with average quality scores greater than 20 and 30, respectively. Raw data quality processing removed 5% of the raw reads. The preliminary assessment of the community represented in the non-axenic culture by metaphlan showed a large majority of quality processed reads belonging to *Cylindrospermopsis (Raphidiopsis) raciborskii* (91%), followed by unclassified sequences (8%). The other few reads were mostly classified in the genera *Brevundimonas* (0.7%) and *Flavobacterium* and *Sandarakinorhabdus* (0.01%–0.02%).

After resampling, the successive analyses were performed on 8,434,616 quality-checked and processed paired-end reads, of which 92.3% was attributed by metaphlan to *Cylindrospermopsis*. Following correction with metaSPAdes, assembly with Megahit yielded 1250 contigs larger than 1000 bp, with a total size of 5.83 Mbp and N50 of 29,680 bp. The Lake Trasimeno *Raphidiopsis* strain was 3.46 Mbp assembled into 87 contigs, with N50 139,068 bp, GC content 39.94% and coverage 655x; based on CheckM2, completeness and contamination estimates were 99.8% and 0.04%, respectively (Table 2).

Following the PGAP analysis, besides 3095 protein-coding genes, the Lake Trasimeno strain included 38 tRNAs, complete sequences of 5S rRNA and 23S rRNA and incomplete sequences

TABLE 2 | Summary of statistics from the Lake Trasimeno *Raphidiopsis raciborskii* genome assembly.

Variable	
Total length (bp)	3,455,381
No. of contigs	87
GC content (%)	39.94
Mean coverage (x)	655
Size of longest contig (bp)	350,162
N50 (bp)	139,068
No. of protein-coding genes	3095
No. of tRNA	38
Complete/partial 5S rRNA	1/1
Complete/partial 16S rRNA	0/1
Complete/partial 23S rRNA	1/0
Completeness (CheckM) (%)	99.93
Contamination (CheckM) (%)	0.44
Strain heterogeneity (CheckM) (%)	0.0
Completeness (CheckM 2) (%)	99.81
Contamination (CheckM 2) (%)	0.04

of 5S rRNA and 16S rRNA (Table 2). After blastn analyses, all these sequences showed a high percentage identity to different *Cylindrospermopsis* and *Raphidiopsis* sequences belonging to strains isolated globally (Table S2). In addition to the short 5S rRNA (114–118 bp) and the partial 16S rRNA (120 bp) sequences, the results obtained by querying the core_nt database were strongly biased by both the incompleteness of some sequences and the incomplete geographical coverage, especially in the markers that were not the object of phylogenetic studies. The results were much more consistent and comparable when querying the Whole Genome shotgun contigs (wgs) database (Table S2); excluding the shortest 5S rRNA and the incomplete 16S rRNA sequences, all the markers showed perfect or high correspondence with strains from several North American water bodies.

The long 16S rRNA sequence (1485 bp) identified in the un-binned contigs (Table S2C) was 100% identical to the 16S rRNA sequence determined by Sanger sequencing of both the strain 1LT32S01 isolated from Lake Trasimeno in 2002 by Mugnai et al. (2008) and the strain isolated from Lake Trasimeno in this work (LT_0923; accession number PV054108).

Based on the GTDB taxonomy and GTDB-Tk results, the closest genome reference of the Lake Trasimeno strain was *R. brookii* D9 (ANI screen, 0.952). Considering the ANI_b computations performed on the whole set of genomes analysed in this work, the closest genomes (ANI_b > 0.99) to the Lake Trasimeno isolate were the two genomes from North America, that is, *Cylindrospermopsis raciborskii* KL1 and Lake Mendota metagenome assembled genome (MAG) *R. brookii* (GCA_044352285.1). All the other ANI_b to the other genomes were between 0.922 and 0.960.

3.5 | Metabolomic Functions of the Lake Trasimeno *Raphidiopsis* Strain

The complete and incomplete pathway modules reconstructed from the KO numbers identified in the Lake Trasimeno LT_0923 strain did include several categories represented by the metabolism of cofactors and vitamins (42), amino acid metabolism (41), energy metabolism (33) and carbohydrate metabolism (31), whereas the remaining pathways included from 2 to 12 modules (Table S3). Reactions required for the central metabolisms of photosynthetic cyanobacteria were represented by oxygenic photosynthesis (photosystems II and I; modules M00161 and M00163), including phycobilisomes (allophycocyanin and phycocyanin/phycoerythrocyanin), beta-carotene biosynthesis (M00097), the reductive pentose phosphate cycle (Calvin cycle) (M00165), the tricarboxylic acid (Krebs) cycle (M00009) and glycolysis (M00001, M00002).

Nitrogen metabolism was sustained by assimilatory nitrate reduction (M00531) and nitrogen fixation (M00175). Specifically, besides the *nifH* gene, which is commonly used as a marker of taxonomic relevance, the ability to fix nitrogen was confirmed by the presence of 11 different *nif* genes identified by PGAP and part of the core and auxiliary set of genes involved in the fixation of atmospheric nitrogen. Nitrogen assimilation was further supported by potential ammonium uptake (K03320; *amt* gene) and by genes encoding nitrate/nitrite transporter (*nrtABD*), part of the ATP-binding cassette (ABC) membrane transporters. No other reactions associated with dissimilatory N-reduction, denitrification, nitrification and anammox were identified.

ABC transporters were also involved in the transport of a broad range of different microelements, nutrients and organic molecules; most of them were classified as complete by KEGG Mapper. Within the C-metabolism, the bicarbonate transporter *CmpABCD* was involved in the cell carbon concentration (CCM). The P-metabolism was supported by several genes involved in the active transport of phosphate (*PstSCAB*) and organophosphorus compounds (phosphonate) (*PhnDEC*), completed by *Pho* regulon components (*PhoHURB*; K06217, K02039, K07636 and K07657) involved in the regulation of P-uptake. Besides the ABC transporter for sulfate/thiosulfate (*Cys/Sbp*), a gene encoding a non-ABC lower-affinity transport sulfate permease was also detected (K03321, *SulP* family). The availability of molybdate, the main bioavailable form of molybdenum, required for many enzymatic reactions (such as N- and S-metabolism), was sustained by the ModABC high-affinity transport system. Another set of genes was involved in the selective transport of organic molecules, including oligosaccharides (chitobiose, *ChiEFG*), phospholipids (*mldDEF*), amino acids (General L-Aminoacid, Branched-chain amino acid, Neutral amino acid/Histidine), the rich-N urea (*UrtABCDE*) and lipopolysaccharides (*LptFGB*). The availability of several microelements was potentially supported by selected ABC transporters (complete or incomplete) for zinc (*ZnuAB*), cobalt (*CbiNMQO*) and nickel (*CbiN/LMQO*). The potential uptake of osmoprotectants was indicated by the presence of the *OpuBCBBBA* ABC transport system.

A primary role in the maintenance of cellular functionality was potentially played by several modules in the biosynthesis

of cofactors and vitamins involved in various biochemical reactions, such as several vitamins with complete or almost complete modules (≥ 0.75) (B12, cobalamin; K1, phyloquinone; and E, Tocopherol/tocotorienol).

The removal of reactive oxygen species (ROS) superoxide (O_2^-) and hydrogen peroxide (H_2O_2) produced as byproducts by photosynthesis was directly or indirectly supported by a set of enzymes encoding superoxide dismutase, catalase, catalase-peroxidase, glutathione peroxidase, peroxiredoxins and other enzymes involved in the regulation of oxidative stress response (K04564, K24157, K24158, K00799, K01920, K00383, K09825).

The ability to control vertical migration was confirmed by the presence of genes encoding proteins forming gas vesicles (PGAP analysis: *GvpA*, C, F, G, J, L, N).

The analysis of the presence of gene clusters encoding secondary metabolites confirmed the absence of operons encoding legacy cyanotoxins. Furthermore, the Lake Trasimeno *Raphidiopsis* genome did not show the presence of any *mcvBDE*, *anaCF*, *cyrABJ* and *sxtAI* residual sequence. In addition to heterocyst glycolipids, which are associated with heterocyst functionality and N-fixation, and a very low similarity (8%) with arylpolyene, the only genomic region showing similarity to a known cluster was associated with the production of hassallidin, while other regions were connected to unknown lanthipeptides and terpenes and one NRPS.

No ARGs were identified in the genome of the *R. raciborskii* isolated from Lake Trasimeno.

3.6 | Phylogenomic Analyses and Metabolomic Potential of *Raphidiopsis* Strains

The phylogenomic tree obtained from the GTDB-Tk analysis clearly showed a separation of the two blocks of species identified as *R. raciborskii* and *R. brookii* in the GTDB taxonomy (R1 and R2 clades; Figure 1). The computations of the ANI_b between the genomes included in the clades R1 and R2 provided values between 0.916 and 0.936 (average 0.928). In the R1 clade, *R. curvata* and *R. curvispora* (*R. curvispora_A* in the GTDB taxonomy) formed a separate cluster (R1B; $ANI_b = 0.959$) from the main group of *R. raciborskii* species isolated from Australia, Asia and Central Europe (R1A; ANI_b between 0.976 and 0.999); the ANI_b values between the genomes included in the R1A and the R1B clades were between 0.946 and 0.959. The Lake Trasimeno strain formed a compact cluster with the two genomes identified in the USA (R2B; ANI_b between 0.995 and 0.997). At lower ANI_b values (R2A vs. R2B, 0.953–0.961), this group was part of the larger group of the GTDB *R. brookii* species isolated from South America (R2A; ANI_b between 0.964 and 0.999). The topology of the phylogenomic tree obtained with GTDB-Tk was confirmed by that obtained using GtoTree (figure not shown).

The main separation in the *Raphidiopsis* tree (R1 and R2) showed correspondence in the presence of two different gene clusters encoding CYN and STX in a selected number of species in the R1A and R2A groups, respectively (Figure 1). No operons encoding legacy cyanotoxins were identified in

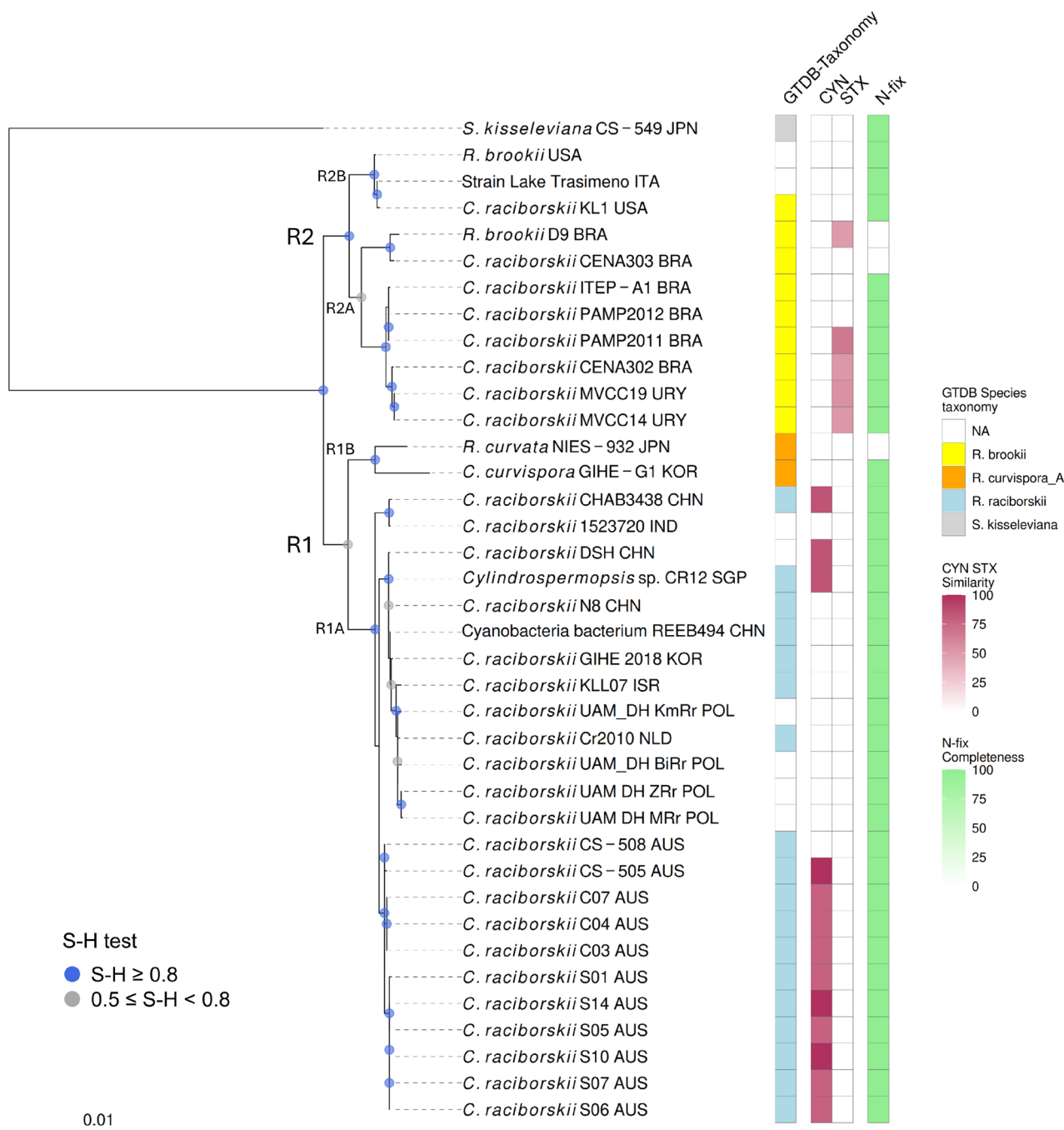


FIGURE 1 | Phylogenomic tree of *Raphidiopsis*, including the strain isolated from Lake Trasimeno and a selection of genomes from the Genome Taxonomy Database (GTDB) and NCBI GenBank. Species names and strain identifiers in the tree tips are from the NCBI taxonomy; the last three letters in the taxa names indicate the geographic origin of the genomes (Table S1). The main clades in the tree are indicated with R1A-B and R2A-B. The GTDB taxonomy classification is given in the GTDB- Taxonomy colour code column. The colour code columns CYN and STX indicate the presence in the genomes of biosynthetic gene clusters (BGCs) encoding cylindrospermopsins and saxitoxins, respectively; colour gradients refer to the similarity of the identified BGCs to known clusters of genes encoding the two toxins estimated by antimash. The column N-fix reports the presence of N-fixing genes identified with the KEGG module analysis. The tree was rooted with *Sphaerospermopsis kisseleviana* CS-549 as outgroup. S-H, Shimodaira-Hasegawa test computed by FastTree; high SH-like local support values (e.g., ≥ 0.8) indicate strong statistical support for the branch. The scale bar indicates the number of substitutions per site.

the *R. curvispora*/*R. curvata* group and (besides the Lake Trasimeno strain), the two genomes from North America. In the R1A group, the presence of gene clusters encoding CYN was found only in the Australian and Asian isolates and not in the European and Israel strains. Nevertheless, the presence of short fragments (355–481 bp) identical (100%) to *cyrJ* and *cyrA* sequences was identified in the UAM/DH-ZRr (POL) strains, while short fragments of sequences equivalent to the *cyrJ* gene (>98%) were identified in the other three Polish genomes, and in the strains from Israel, Netherlands and in *C. curvispora* GIHE-G1 (KOR). In the South America group (R2A), among the strains lacking the STX encoding operon, the strain ITEP-A1 showed the presence of *sxtA* and *sxtI* genes. Among secondary metabolites, a gene cluster encoding hassallidin (antifungal lipopeptide) was identified in almost all the species, with exclusion of a few strains in China (*C. raciborskii* CHAB3438_CHN) and South America (5 strains). Furthermore, an operon encoding rhizomide A/B/C (lipopeptides with antimicrobial and siderophore-like activity) was identified in R1A strains (6 Australian and 1 Chinese) and in one R2A South American strain (*C. raciborskii* CENA303). Save for a few other rare gene clusters present in single species, the arylpolyene (antioxidants) operon was identified (with a very low similarity, 8%–12%) in several strains in all the phylogenetic clades. Excluding an operon encoding puwainaphycin F/minutesamide A/B/C/D (30% similarity) in *R. raciborskii* KL1, the three R2B genomes shared a comparable secondary metabolite profile.

With the exclusion of *R. brookii* D9 and *C. raciborskii* CENA303 (R2A), and *R. curvata* NIES-932 (R1B), the KEGG analysis allowed identification of the *nif* genes (KEGG module M00175) in all the analysed strains (Figure 1). With a very few, minor differences, the other pathway modules identified by the KEGG annotation of the Lake Trasimeno genome were the same as those obtained with the annotation of the other two R2B genomes, that is, *C. raciborskii* KL1 and *R. brookii* (Table S3). Overall, differences among all the strains were mostly limited to a few modules, generally with low completeness, such as Lipopolysaccharide metabolism, Biosynthesis of terpenoids and polyketides and Biosynthesis of other secondary metabolites (Table S3). Nevertheless, functional enrichment analyses contrasting the R1 and R2, and the R1A and R2A groups did not show any significant geographical pattern in the presence of the functional modules (adjusted *q* value > 0.4).

Similarly to the Lake Trasimeno strain, no ARGs were identified in any of the other *Raphidiopsis* strains considered in this work, with the sole exception of *R. brookii* D9, which showed the presence of a type A-1 chloramphenicol O-acetyltransferase.

3.7 | Pangenomic Analysis

The pangenomic analysis of the 37 *Raphidiopsis* genomes identified a total of 126,318 genes grouped in 6066 gene clusters. The single-copy core gene clusters (SCGs) and the singletons (genes present in a single genome) represented 25.2% and 24.7% of the total gene clusters, respectively (Figure 2). Overall, 60% of all clusters contained genes that were annotated with COG functions (COG20_FUNCTION). The functional annotation rate

was 84% for the SCGs and 27% for the singletons. The analysis by panaroo identified 6800 ‘total genes’, with a distribution in the corresponding classes defined by panaroo (Core, Soft, Shell and Cloud) comparable with the results obtained by Anvi’o (Figure S2). The distribution of the gene families classified using the COG20 categories showed a prevalence of mobilome (prophages transposons), defence mechanisms, secondary metabolites biosynthesis, carbohydrate and amino acid transport metabolisms in the cloud and, partly, in the shell pangenome (Figure 3).

The phylogenomic analysis based on the 1526 SCGs (Figure 2, top right) grouped the 37 genomes consistently with the ANI_b values and the topology of the phylogenomic analysis by GTDB-Tk based on a selection of 120 gene markers, as well as with the geographic distribution of *Raphidiopsis* strains (Figure 2; Figure S3). Furthermore, the grouping was also consistent with the differential pattern (genomic islands *sensu* Coleman et al. 2006) of shell and cloud gene clusters in the upper left of Figure 2.

Enrichment analysis allowed the identification of functions unique to the two main geographical groups R1 (R1A and R1B) and R2 (R2A and R2B), corresponding to genomes from Australia, Asia and Central Europe, and South/North America and central Italy, respectively (Table S4). Given that the test is not reliable when some groups have very few genomes (Shaiber et al. 2020), other comparisons (e.g., contrasting R2B, including the Lake Trasimeno genome) were not possible. Enrichment analysis revealed several differences, some within common categories with different specific functions in the two groups (defence mechanisms; translation, ribosomal structure and biogenesis), but the majority within different categories. Specifically, major differences between the two groups included functions in carbohydrate metabolism, metal resistance, lipid metabolism and secondary metabolite pathways (see Discussion).

The Heaps model applied to the whole pangenome gave gamma values higher than zero considering the gene clusters from both Anvi’o ($\gamma=0.20$) and panaroo ($\gamma=0.21$). Similar results were obtained considering the fitting of the model to the Anvi’o and panaroo outputs separately for the R1 ($\gamma=0.14$ and $\gamma=0.16$) and R2 clades ($\gamma=0.18$ and $\gamma=0.14$), respectively.

4 | Discussion

The genomic analysis of a strain of *R. raciborskii* isolated from a surface sample collected in Lake Trasimeno allowed us to clarify for the first time the taxonomic position and functional characteristics of a species collected in a southern European region.

4.1 | Main Features of the Lake Trasimeno Strain

After 30 years since the identification of *R. raciborskii* in Lake Trasimeno in 1995, the presence of this species has been progressively documented in many other Italian water bodies, representing an important constitutive species in meso-eutrophic lakes (Austoni and Stefanelli 2024; Del Pasqua et al. 2024; Austoni et al. 2025).

Raphidiopsis pangenome

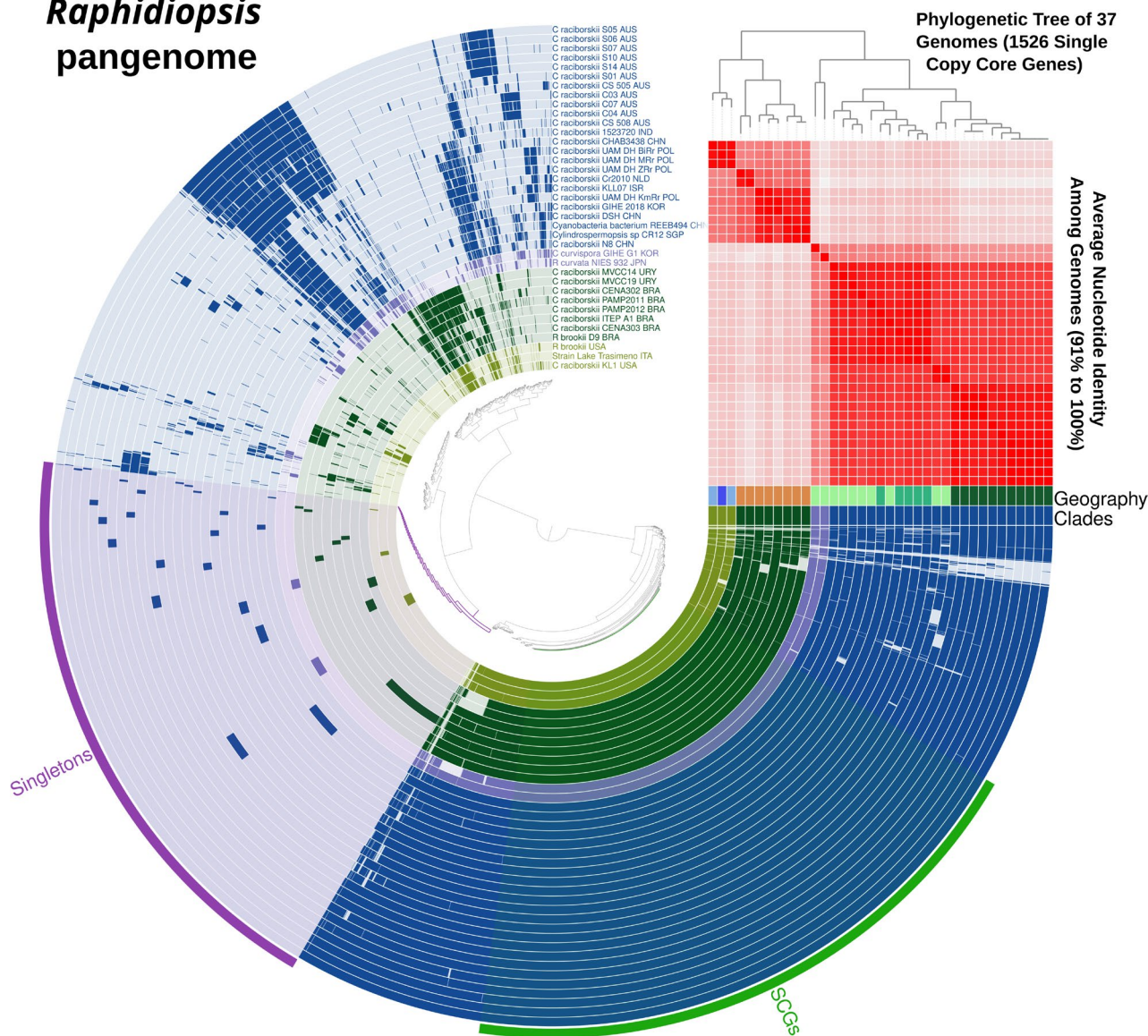


FIGURE 2 | The pangenome of *Raphidiopsis*. Each one of the 6066 gene clusters is represented by a radius that contains one or more amino acid sequences that originate from one or more genomes, which are represented by the 37 circles. Gene clusters are organised according to their distribution across the genomes, so that gene clusters that occur together in the same group of genomes are closer together. Genomes are ordered consistently with the phylogenomic tree on the top right side of the figure, based on the single copy core genes (SCGs), as well as the symmetrical matrix reporting the ANI_b values computed between all 37 genomes. The outer circle highlights the two regions of the pangenome including the 1526 SCGs and the 1500 singletons. On the right-hand side of the figure, Clades and Geography show the four main clades identified in the phylogenomic tree (Figure 1) and the indication of the geographical distribution of the genomes (as reported in detail by the last three code letters in the names of the taxa) respectively.

From a water management perspective, the absence of legacy cyanotoxins in the environmental samples and the lack of operons encoding cyanotoxins further confirm the general absence of toxigenic strains in the European continent. The detection of CYN in central Italy (Manti et al. 2005; Messineo et al. 2010) or STX in Lake Comabbio (northern Italy) (Austoni and Stefanelli 2024) during the development of *R. raciborskii* was based on the analysis of environmental samples, which cannot exclude the presence of other CYN/STX-producing species, highlighting the need to examine isolated strains or genome/MAGs for a consistent evaluation of (potential) toxin producers. However, these observations call for a more comprehensive survey focused on studying the cyanobacterial community in

different seasons. This is important also considering the huge abundance that the populations of *R. raciborskii* can develop in Lake Trasimeno, that is well over 100,000 cell mL⁻¹ (this work, and <https://apps.arpa.umbria.it/>). It is worth highlighting that the absence of legacy cyanotoxins does not necessarily imply an absence of toxicity (Fastner et al. 2003).

Genome annotation by KEGG identified the main modules associated with the physiology of photosynthetic cyanobacteria, and several other modules encoding specialised functions relevant to bloom-forming nitrogen-fixing heterocytous Nostocales (Cao et al. 2020). Besides N-fixation, the LT_0923 genome included several modules involved in the

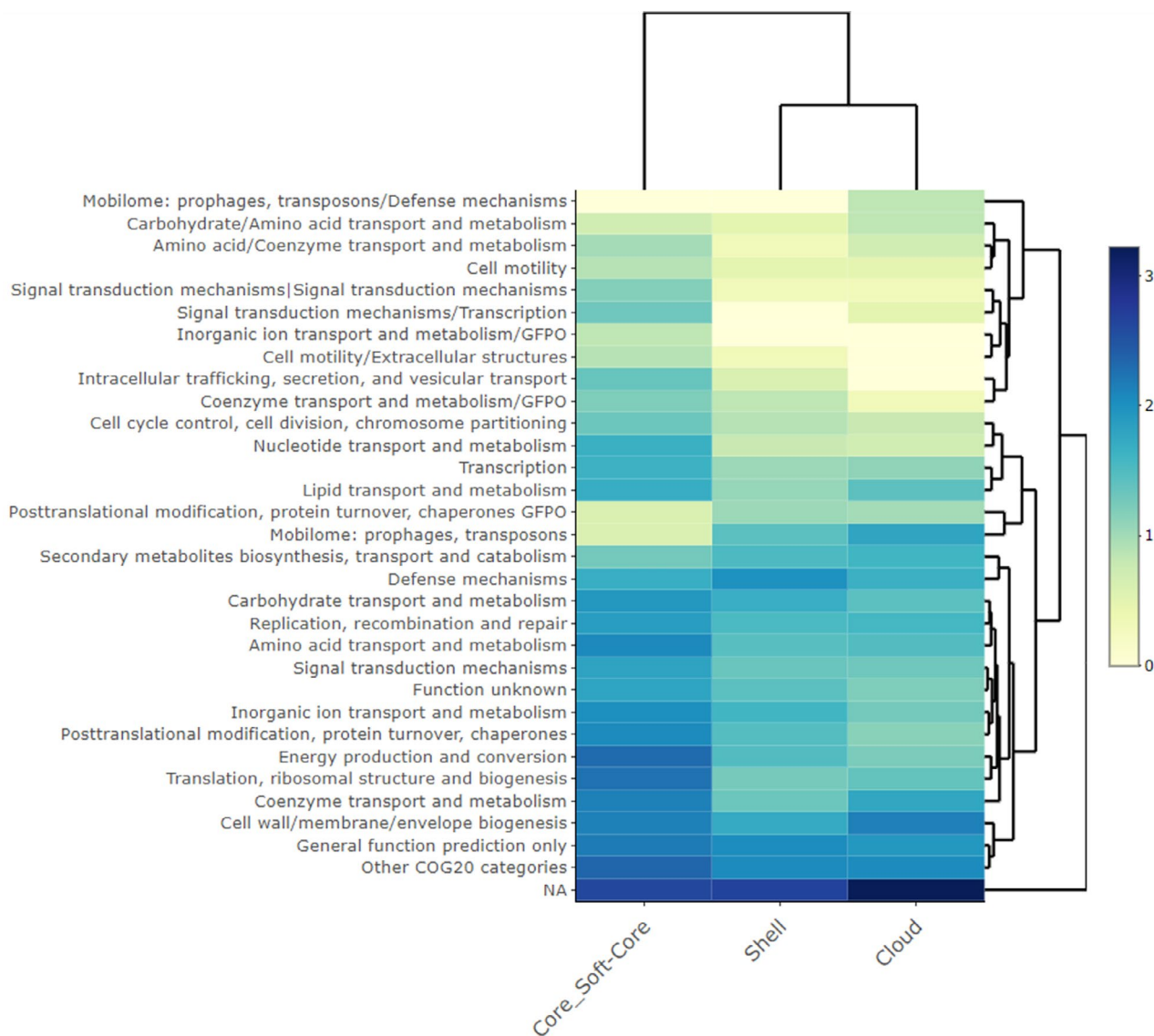


FIGURE 3 | Heatmap showing the most abundant COG20 categories in the main pangenome groups, namely Core_Soft-Core (grouping the Core and Soft-Core), Shell and Cloud genomes. Less represented categories in genomes (<6) were grouped under 'Other COG20 categories'. GFPO, General Function Prediction Only. Before plotting, the data were transformed by $\log_{10}(x+1)$. Unclassified gene families were grouped under 'NA'.

scavenging and transport of macro- and micronutrients. In Lake Trasimeno, P, N and C can reach low critical concentrations during the bloom periods. This was apparent in the low concentrations of N and the high values of pH detected during the sampling. Besides potential N-fixation (Natwora and Sheik 2021; Ehrenfels et al. 2023), the presence of several ABC cassette transporters for several organic compounds suggested an active role in the transport of N and other nutrients from the environment (Shvarev and Maldener 2018). At pH around 9, CO₂ is almost absent (Stumm and Morgan 1996), and cells require specific transporters for carbon concentration (Maeda et al. 2000; Koropatkin et al. 2007).

The regulation of oxidative stress response from the production of ROS as a byproduct of photosynthesis was associated with a specific set of genes. Nevertheless, compared to cyanobacteria forming surface blooms and scums (Salmaso et al. 2024; Buzzi et al. 2025), this specific stress may be mitigated in species developing massively in a turbid water column.

4.2 | Taxonomic Assessment

The ANI_b values (>0.99) and the phylogenomic analysis based on the 120 GTDB-Tk markers associated the Lake Trasimeno LT_0923 strain with the two published North American genomes. In turn, at lower values (ANI_b between 0.95 and 0.96), these three assemblies were associated with all the other genomes isolated from South America. Within the South American group, the strong association between *R. brookii* D9 and *R. raciborskii* CENA303 was also confirmed by the common inability to fix atmospheric nitrogen. In algal cultures, these two species did not produce heterocytes (Hill 1972; Abreu et al. 2018), showing attenuated (CENA303) and continuously attenuated towards ends (D9) trichomes. The presence of very attenuated apices and the absence of heterocytes were considered two essential diacritical features characterising the original description of the *R. brookii* species by Hill (1972); see also Komárek (2013). However, the presence of heterocytous strains (including Lake Trasimeno strain) within the R2A/R2 clade challenges the validity of these

two original diacritical characters in supporting the definition of different species.

Besides the American and the unique Southern Europe genome (LT_0923), all the other genomes in the R1 group formed a separate cluster including strains from Australia, Asia and Central Europe. Besides the separation of the *R. curvata/R. curvispora* isolates (R1B), the other genomes in R1A showed a compact cluster ($ANI_b > 0.976$), but with a further secondary separation of the Australian clade from the Asian and Central European strains.

Overall, the ability to fix atmospheric nitrogen and the presence of heterocysts can be considered an ancestral and not distinguishing character of *Raphidiopsis*. As outlined by Aguilera et al. (2018), the lack of heterocysts would be a consequence of independent events of secondary loss occurring in different geographical areas. Conversely, the production of CYN and STX was circumscribed to the R1A and R2A clades, respectively, showing a strong geographical pattern. The presence of vestiges of *cyr* genes in non-toxicogenic producing European strains is consistent with the observation that CYN production is an ancestral trait (Hoff-Rissetti et al. 2013; Piccini et al. 2013).

The distinction of different *Raphidiopsis* populations and their correspondence with specific geographical locations is an intriguing observation, also considering that most of the culture isolates showed morphological diacritical features which were considered to be essentially consistent with the description of the species *R. raciborskii* as classified and/or described, for example, for the Australian (Fuentes-Valdés et al. 2018; Willis et al. 2018), Asian (Chen et al. 2022), European (Tokodi et al. 2025), South American (Abreu et al. 2018) and North American (Martin et al. 2021) strains. Excluding the *R. curvispora*_A clade (R1B), while the main clusters R1A and R2 showed ANI_b values each congruent or partially congruent with the attribution of genomes to one unique species ($ANI_b > 0.98$ and $ANI_b > 0.95$, respectively), the ANI_b computed between the genomes in the two main clades R1 and R2 showed values between 0.92 and 0.94, that is, in the limbo of values between 0.90 and 0.95 defined by Rodríguez-R et al. (2024) as representing a discontinuity/gap between the thresholds separating species.

A comparison of the main KEGG pathway modules allowed us to identify, with very few minor exceptions, a close correspondence of functions in the two North American strains and the Lake Trasimeno strain (R2B). On the other hand, an enrichment analysis contrasting the two main clades R1 and R2 did not show any significant difference in the modules identified in the two groups. This result is not surprising considering that the KEGG modules mostly include important conserved and general pathways, often representing core metabolic functions that are only differentiated in separate evolutionary groups. Considering the pangenomic analysis (next section), less than 10% and 33% of the gene families were annotated with the KEGG modules and Kofam codes, respectively.

4.3 | Pangenomic Assessment

The number of gene families identified by the pangenomic analyses must be treated with caution, as they depend not only

on the number and origin of the genomes analysed but also on the software and on the thresholds and coverage parameters for similarity sequences used to predict gene families (Tettelin et al. 2008; Vernikos et al. 2015; Costa et al. 2020). For example, by analysing a different set of 31 *Raphidiopsis* genomes, Laux et al. (2023) obtained a pangenome including 7943 protein clusters, with indications about the presence of an open pangenome.

The geographic and genomic groups defined by the phylogenomic analyses showed a close correspondence with the presence of groups of genes belonging to subgroups of genomes in the shell and cloud gene clusters, which further helped to distinguish genomes within the major groups. While the core genome generally includes genes necessary for the basic metabolic activities, the shell and cloud pangenomes include genes involved in the adaptation to specific environments, providing insights into the mechanisms and causes driving differentiation.

The open pangenome in *Raphidiopsis* is indicative of the existence of populations developing in diverse regions. Compared to closed pangenomes, open pangenomes are larger, have a smaller proportion of core genes, and have higher rates of gene acquisition by horizontal gene transfer (HGT) (Brockhurst et al. 2019). These characteristics align with populations developing in distant geographical regions, acquiring genes from other species or developing adaptations necessary for survival in different habitats. Conversely, closed pangenomes are typical of species growing in homogeneous environments or extreme habitats, which require highly specialised adaptations and conserved genetic characteristics such as in pathogenic and symbiotic species (Park et al. 2019).

The differential presence of COGs in the two main R1 and R2 clades may be indicative of different functional adaptations to local environmental pressures. The list of differential COGs and their discussion is reported in Table S4. Among others, development in different geographic regions can be reflected by the presence of distinctive defence mechanisms, exclusive enrichment of the mobilome associated with prophages and transposons, carbohydrate transport (linked to utilisation of external organic compounds), exclusive presence of metal efflux pumps linked to specific resistance to toxic metal ions, differential regulation of gene expression and maintenance of cell wall functionality.

As expected, differences may also manifest in genomes isolated from the same region. For example, the differences apparent in Figure 2 between the 'straight' S01–S14 and 'coiled' C03–C07 strains isolated from Lake Wivenhoe (Australia) were analysed in detail by Willis et al. (2018), who concluded that genomic variability exists among co-occurring strains and may be the basis for strain phenotypic differences and population plasticity.

4.4 | Geographical Distribution of *Raphidiopsis*

Among speciation processes, differentiation between bacterial populations may occur at the subspecies level within ecological niches (ecovars, ecological variants) and due to geographical separation (geovars, geographical variants) (Staley 2006). A well-known example of ecovars is the different strains of *Prochlorococcus* adapted to different levels of light intensity

(Delmont and Eren 2018; Ulloa et al. 2021), while several pieces of evidence have been accumulated on the differentiation between species (geovars) from different regions on Earth (Staley 2006; Choudoir et al. 2016). In all cases, HGT is pivotal in contributing to cyanobacterial differentiation, including transduction (Rai et al. 2021).

In this regard, the presence of genomic islands characterised by different physiological and toxigenic traits is consistent with the gain and loss of genetic traits after the geographical separation of populations. Within the shell and cloud genome, genomic islands are the results of the different histories of adaptation and changes mediated by HGT, contributing to expanding the accessory genome and differentiation of *R. raciborskii* in different geovars. The ANI differences between separated geographical groups in the limbo of values between 0.92 and 0.94 would suggest the existence of an active plastic pangenome and, possibly, a (continuing) separation into different geospecies (geographically distinct species).

The close evolutionary and functional connection of the Lake Trasimeno *R. raciborskii* strain with the North American assemblies is consistent with the results obtained in previous phylogenetic studies based on the analysis of 16S rRNA and housekeeping genes, which found a tight coupling between strains isolated from North America and strains isolated from Southern Europe, that is Spain (Cirés et al. 2014) and Greece (Panou et al. 2018). Considering also the close coupling found between Tunisian, Greek and USA isolates (Moreira et al. 2015; Panou et al. 2018), this work adds new geographic and genomic evidence for a close connection between North America and (part) of southern European populations, calling for new assumptions on the phylogeography and dispersal routes of *R. raciborskii*, and adding new evidence to a connection of the species between the north American continent and the Mediterranean area (Haande et al. 2008; Cirés et al. 2014; Panou et al. 2018), even though the transport mechanisms remain to be clarified (Sha et al. 2025).

4.5 | Study Methodological Limitations

The results are strongly influenced by the limited number of genomes characterised to date and available in public repositories. The lack of *R. raciborskii* genome coverage in Africa and the few strains available in Europe, North America, and North Asia make it difficult to reach firm conclusions. The open pangenome is expected to grow as more genomes are included, in particular increasing the cloud and (to a lesser extent) reducing the core component (Cai et al. 2023). Greater geographic coverage of genomes will allow for a more robust understanding of taxonomy and fine-scale changes and adaptations captured in the shell and cloud pangenomes, although reliable interpretations may still depend on the completeness of metabolic databases. Furthermore, potential limitations of full shotgun sequencing technologies (which result in broken assemblies and variable number of contigs, thus also affecting functional mapping; Eisenhofer et al. 2023) can be solved or mitigated by the adoption of new state-of-the-art long-read technologies that allow annotation of circular genomes (Driscoll et al. 2017). Finally, while functional genomic analyses indicate the presence and potential

expression of genes, transcriptomic approaches are required to assess their expression (Choi et al. 2016).

5 | Conclusion

The genomic analysis of the LT_0923 strain of *R. raciborskii* isolated from Lake Trasimeno allowed us to understand the taxonomic position and functional characteristics of a strain isolated in the Mediterranean region compared with other assemblies distributed worldwide. Phylogenomic analyses based on taxonomic markers indicated a strong connection between the Mediterranean strain and the corresponding strains from North America and South America, as well as a separation from the Australian, Asian, and Central European strains. The results were confirmed by the analysis of gene clusters in the core genome of the whole set of *Raphidiopsis* genomes, indicating conservation of the phylogenetic signal through vertical inheritance, but with a concomitant high gene flow confirmed by the high proportion of cloud genes. The new results confirmed the ancestral character of nitrogen fixation, which was lost only in a few strains. Conversely, the presence of STX and CYN did show a circumscribed distribution in the South American, and the Australian and Asian strains, respectively. Causes of the absence of toxicity in the North American/Mediterranean genomes (STX), and in the European and the majority of Asian strains can be due to several factors, including operon loss/changes (evidenced by cyanotoxins gene vestiges), differential HGT and genetic drift. The open pangenome, and the large fraction and geographically distinct gene families identified in the cloud genome, which is enriched with genes specialised in environmental-specific functions and defence mechanisms, are consistent with the development of *Raphidiopsis* in geographically distinct regions. Though this work contributed to filling a knowledge gap concerning the taxonomy and functions of a *Raphidiopsis* strain isolated in the Mediterranean region, the overall knowledge of the taxonomy of this genus is far from being considered complete and strongly biased by the small number of genomes so far characterised.

Author Contributions

Nico Salmaso: conceptualization, software, formal analysis, writing – original draft, writing – review and editing, investigation. **Leonardo Cerasino:** methodology, investigation, writing – review and editing. **Margherita Di Brizio:** methodology, investigation, writing – review and editing. **Massimo Pindo:** methodology, investigation, writing – review and editing. **Adriano Boscaini:** methodology, investigation, data curation, writing – review and editing.

Acknowledgements

We thank the personnel of ARPA Umbria involved in the environmental monitoring of Lake Trasimeno. We thank A. Zampedri, A. Paoli, S. Larger and E. Stefani for technical assistance in the laboratory analyses. This study was carried out within the National Biodiversity Future Centre and received funding from the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR)—Missione 4 Componente 2, Investimento 1.4—D.D. 1034 17/06/2022, CN00000033, CUPD43C22001280006). This manuscript reflects only the authors' views and opinions; neither the European Union nor the European Commission can be considered responsible for them. We are grateful to two anonymous reviewers for valuable comments and suggestions on an earlier version of the manuscript.

Ethics Statement

All prevailing local, national and international regulations and conventions, and normal scientific ethical practices, have been respected.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JBMNDW000000000. The version described in this paper is version JBMNDW010000000. The sequences are publicly available under BioProject accession number PRJNA1215010, BioSample accession number SAMN46392051, and SRA accession number SRR32208149. The accession number for the 16S rRNA gene obtained from Sanger sequencing is PV054108.

References

- Abreu, V. A. C., R. V. Popin, D. O. Alvarenga, et al. 2018. "Genomic and Genotypic Characterization of *Cylindrospermopsis raciborskii*: Toward an Intraspecific Phylogenetic Evaluation by Comparative Genomics." *Frontiers in Microbiology* 9: 306.
- Aguilera, A., E. B. Gómez, J. Kaštovský, R. O. Echenique, and G. L. Salerno. 2018. "The Polyphasic Analysis of Two Native *Raphidiopsis* Isolates Supports the Unification of the Genera *Raphidiopsis* and *Cylindrospermopsis* (Nostocales, Cyanobacteria)." *Phycologia* 57: 130–146.
- Alneberg, J., B. S. Bjarnason, I. de Bruijn, et al. 2014. "Binning Metagenomic Contigs by Coverage and Composition." *Nature Methods* 11: 1144–1146.
- Antunes, J. T., P. N. Leão, and V. M. Vasconcelos. 2015. "*Cylindrospermopsis raciborskii*: Review of the Distribution, Phylogeography, and Ecophysiology of a Global Invasive Species." *Frontiers in Microbiology* 6: 473.
- Aramaki, T., R. Blanc-Mathieu, H. Endo, et al. 2020. "KofamKOALA: KEGG Ortholog Assignment Based on Profile HMM and Adaptive Score Threshold." *Bioinformatics* 36: 2251–2252.
- Austoni, M., A. Boscaini, F. Buzzi, et al. 2025. "Recent Spread of *Raphidiopsis raciborskii* in the Lake District South of the Alps." *Advances in Oceanography and Limnology* 16: 13225. <https://doi.org/10.4081/aiol.2025.13225>.
- Austoni, M., and M. Stefanelli. 2024. "First Observation of *Raphidiopsis (Cylindrospermopsis) raciborskii* (Nostocales) During a Heatwave in Lake Comabbio (Northern Italy)." *Harmful Algae News* 77: 20–21.
- Blanco-Míguez, A., F. Beghini, F. Cumbo, et al. 2023. "Extending and Improving Metagenomic Taxonomic Profiling With Uncharacterized Species Using MetaPhlan 4." *Nature Biotechnology* 41: 1633–1644.
- Blin, K., S. Shaw, H. E. Augustijn, et al. 2023. "antiSMASH 7.0: New and Improved Predictions for Detection, Regulation, Chemical Structures and Visualisation." *Nucleic Acids Research* 51: W46–W50.
- Brockhurst, M. A., E. Harrison, J. P. J. Hall, T. Richards, A. McNally, and C. MacLean. 2019. "The Ecology and Evolution of Pangenomes." *Current Biology* 29, no. 20: R1094–R1103. <https://doi.org/10.1016/j.cub.2019.08.012>.
- Buzzi, F., C. Agostinelli, R. Alber, et al. 2025. "Cyanobacterial Blooms in the Po River Basin and the Eastern Alps." *Advances in Oceanography and Limnology* 16: 12584.
- Cai, H., C. J. McLimans, J. E. Beyer, L. R. Krumholz, and K. D. Hambright. 2023. "*Microcystis* Pangenome Reveals Cryptic Diversity Within and Across Morphospecies." *Science Advances* 9: eadd3783.
- Cao, H., Y. Shimura, M. M. Steffen, et al. 2020. "The Trait Repertoire Enabling Cyanobacteria to Bloom Assessed Through Comparative Genomic Complexity and Metatranscriptomics." *MBio* 11, no. 3: e01155-20. <https://doi.org/10.1128/mbio.01155-20>.
- Cerasino, L., and N. Salmaso. 2020. "Co-Occurrence of Anatoxin-a and Microcystins in Lake Garda and Other Deep Subalpine Lakes: Co-Occurrence of Anatoxin-a and Microcystins in Lake Garda." *Advances in Oceanography and Limnology* 11: 11–21. <https://doi.org/10.4081/aiol.2020.8677>.
- Chaumeil, P.-A., A. J. Mussig, P. Hugenholtz, and D. H. Parks. 2022. "GTDB-Tk v2: Memory Friendly Classification With the Genome Taxonomy Database." *Bioinformatics* 38: 5315–5316.
- Chen, Z.-J., Z.-X. Ruan, N. Cheng, et al. 2022. "Whole-Genome Sequencing and Phosphorus Uptake and Transport Pathway Comparative Analysis of *Cylindrospermopsis raciborskii*." *Acta Hydrobiologica Sinica* 46: 1130–1141.
- Chklovski, A., D. H. Parks, B. J. Woodcroft, and G. W. Tyson. 2023. "CheckM2: A Rapid, Scalable and Accurate Tool for Assessing Microbial Genome Quality Using Machine Learning." *Nature Methods* 20: 1203–1212.
- Choi, S. Y., B. Park, I.-G. Choi, et al. 2016. "Transcriptome Landscape of *Synechococcus elongatus* PCC 7942 for Nitrogen Starvation Responses Using RNA-Seq." *Scientific Reports* 6: 30584.
- Choudoir, M. J., J. R. Doroghazi, and D. H. Buckley. 2016. "Latitude Delineates Patterns of Biogeography in Terrestrial *Streptomyces*." *Environmental Microbiology* 18: 4931–4945.
- Cirés, S., L. Wörmer, A. Ballot, et al. 2014. "Phylogeography of *Cylindrospermopsis* and Paralytic Shellfish Toxin-Producing Nostocales Cyanobacteria From Mediterranean Europe (Spain)." *Applied and Environmental Microbiology* 80: 1359–1370.
- Coleman, M. L., M. B. Sullivan, A. C. Martiny, et al. 2006. "Genomic Islands and the Ecology and Evolution of *Prochlorococcus*." *Science* 311: 1768–1770.
- Costa, S. S., L. C. Guimarães, A. Silva, S. C. Soares, and R. A. Baraúna. 2020. "First Steps in the Analysis of Prokaryotic Pan-Genomes." *Bioinformatics and Biology Insights* 14: 38064. <https://doi.org/10.1177/1177932220938064>.
- da Costa, S. M., A. S. Ferrão-Filho, and S. M. F. O. Azevedo. 2013. "Effects of Saxitoxin- and Non-Saxitoxin-Producing Strains of the Cyanobacterium *Cylindrospermopsis raciborskii* on the Fitness of Temperate and Tropical Cladocerans." *Harmful Algae* 28: 55–63. <https://doi.org/10.1016/j.hal.2013.05.017>.
- Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10: giab008.
- Del Pasqua, M., F. Bandini, E. Morandi, R. Antonellini, M. Masi, and M. de Gioia. 2024. "Blooms of the Blue-Green Alga *Cylindrospermopsis raciborskii* in Emilia Romagna Region (Italy): Case Reports Managed by the Regional Agency for Environmental Protection and Energy of Emilia Romagna (ARPAE)." *Advances in Oceanography and Limnology* 15: 12608.
- Delmont, T. O., and A. M. Eren. 2018. "Linking Pangenomes and Metagenomes: The *Prochlorococcus* Metapangenome." *PeerJ* 6: e4320.
- Driscoll, C. B., T. G. Otten, N. M. Brown, and T. W. Dreher. 2017. "Towards Long-Read Metagenomics: Complete Assembly of Three Novel Genomes From Bacteria Dependent on a Diazotrophic Cyanobacterium in a Freshwater Lake Co-Culture." *Standards in Genomic Sciences* 12: 9.
- Eddy, S. R. 2011. "Accelerated Profile HMM Searches." *PLoS Computational Biology* 7: e1002195.
- Edgar, R. C. 2022. "Muscle5: High-Accuracy Alignment Ensembles Enable Unbiased Assessments of Sequence Homology and Phylogeny." *Nature Communications* 13: 6968.

- Ehrenfels, B., K. B. L. Baumann, R. Niederdorfer, et al. 2023. "Hydrodynamic Regimes Modulate Nitrogen Fixation and the Mode of Diazotrophy in Lake Tanganyika." *Nature Communications* 14: 6591.
- Eisenhofer, R., I. Odriozola, and A. Alberdi. 2023. "Impact of Microbial Genome Completeness on Metagenomic Functional Inference." *ISME Communications* 3: 1–5.
- Eren, A. M., E. Kiefl, A. Shaiber, et al. 2021. "Community-Led, Integrated, Reproducible Multi-Omics With Anvi'o." *Nature Microbiology* 6: 3–6.
- Fastner, J., R. Heinze, A. R. Humpage, U. Mischke, G. K. Eaglesham, and I. Chorus. 2003. "Cylindrospermopsin Occurrence in Two German Lakes and Preliminary Assessment of Toxicity and Toxin Production of *Cylindrospermopsis raciborskii* (Cyanobacteria) Isolates." *Toxicon* 42: 313–321.
- Feldgarden, M., V. Brover, N. Gonzalez-Escalona, et al. 2021. "AMRFinderPlus and the Reference Gene Catalog Facilitate Examination of the Genomic Links Among Antimicrobial Resistance, Stress Response, and Virulence." *Scientific Reports* 11: 12728.
- Fuentes-Valdés, J. J., K. Soto-Liebe, D. Pérez-Pantoja, et al. 2018. "Draft Genome Sequences of *Cylindrospermopsis raciborskii* Strains CS-508 and MVCC14, Isolated From Freshwater Bloom Events in Australia and Uruguay." *Standards in Genomic Sciences* 13: 26.
- Haande, S., T. Rohrlack, A. Ballot, et al. 2008. "Genetic Characterisation of *Cylindrospermopsis raciborskii* (Nostocales, Cyanobacteria) Isolates From Africa and Europe." *Harmful Algae* 7: 692–701.
- Hill, H. 1972. "A New *Raphidiopsis* Species (Cyanophyta, Rivulariaceae) From Minnesota Lakes." *Phycologia* 11: 213–215.
- Hoff-Risetti, C., F. A. Dörr, P. D. C. Schaker, E. Pinto, V. R. Werner, and M. F. Fiore. 2013. "Cylindrospermopsin and Saxitoxin Synthetase Genes in *Cylindrospermopsis raciborskii* Strains From Brazilian Freshwater." *PLoS One* 8: e74238.
- Huerta-Cepas, J., F. Serra, and P. Bork. 2016. "ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data." *Molecular Biology and Evolution* 33: 1635–1638.
- Kanehisa, M., S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. 2014. "Data, Information, Knowledge and Principle: Back to Metabolism in KEGG." *Nucleic Acids Research* 42: D199–D205.
- Kanehisa, M., and Y. Sato. 2020. "KEGG Mapper for Inferring Cellular Functions From Protein Sequences." *Protein Science* 29: 28–35.
- Kang, D. D., F. Li, E. Kirton, et al. 2019. "MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction From Metagenome Assemblies." *PeerJ* 7: e7359.
- Karadžić, V., G. S. Simić, D. Natić, A. Ržaničanin, M. Ćirić, and Z. Gačić. 2013. "Changes in the Phytoplankton Community and Dominance of *Cylindrospermopsis raciborskii* (Wolosz.) Subba Raju in a Temperate Lowland River (Ponjavica, Serbia)." *Hydrobiologia* 711: 43–60.
- Kokociński, M., R. Akçaalan, N. Salmaso, M. P. Stoyneva-Gärtner, and A. Sukenik. 2017. "Expansion of Alien and Invasive Cyanobacteria." In *Handbook on Cyanobacterial Monitoring and Cyanotoxin Analysis*, edited by J. Meriluoto, L. Spoof, and G. A. Codd, 28–39. Wiley.
- Komárek, J. 2013. "Cyanoprokaryota Part 3: Heterocytous Genera." In *Süßwasserflora von Mitteleuropa, Band 19/3*. Springer Spektrum.
- Komárek, J., and J. Mareš. 2012. "An Update to Modern Taxonomy (2011) of Freshwater Planktic Heterocytous Cyanobacteria." *Hydrobiologia* 698: 327–351.
- Koropatkin, N. M., D. W. Koppelaar, H. B. Pakrasi, and T. J. Smith. 2007. "The Structure of a Cyanobacterial Bicarbonate Transport Protein, CmpA." *Journal of Biological Chemistry* 282: 2606–2614.
- Kotai, J. 1972. "Instructions for Preparation of Modified Nutrient Solution Z8 for Algae." Publ B-1169 Norwegian Institute for Water Research.
- Lagos, N. 2003. "Paralytic Shellfish Poisoning Phycotoxins: Occurrence in South America." *Comments on Toxicology* 9: 175–193.
- Langmead, B., and S. L. Salzberg. 2012. "Fast Gapped-Read Alignment With Bowtie 2." *Nature Methods* 9: 357–359.
- Laux, M., C. M. Piroupo, J. C. Setubal, and A. Giani. 2023. "The *Raphidiopsis* (= *Cylindrospermopsis*) *raciborskii* Pangenome Updated: Two New Metagenome-Assembled Genomes From the South American Clade." *Harmful Algae* 129: 102518.
- Lee, M. D. 2019. "GToTree: A User-Friendly Workflow for Phylogenomics." *Bioinformatics* 35: 4162–4164.
- Li, D., C.-M. Liu, R. Luo, K. Sadakane, and T.-W. Lam. 2015. "MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph." *Bioinformatics (Oxford, England)* 31: 1674–1676.
- Li, R., W. W. Carmichael, S. Brittain, et al. 2001. "Isolation and Identification of the Cyanotoxin Cylindrospermopsin and Deoxy-Cylindrospermopsin From a Thailand Strain of *Cylindrospermopsis raciborskii* (Cyanobacteria)." *Toxicon* 39: 973–980.
- Li, W., K. R. O'Neill, D. H. Haft, et al. 2020. "RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline Reach With Protein Family Model Curation." *Nucleic Acids Research* 49: D1020–D1028.
- Ludovisi, A., D. Cappelletti, A. C. Elia, et al. 2021. "IT21-A Lago Trasimeno." In *La Rete Italiana per la Ricerca Ecologica di Lungo Termine. Lo Studio Della Biodiversità e Dei Cambiamenti*, edited by L. Capotondi, 663–673. CNR-Edizioni.
- Lund, J. W. G., C. Kipling, and E. D. Le Cren. 1958. "The Inverted Microscope Method of Estimating Algal Numbers and the Statistical Basis of Estimations by Counting." *Hydrobiologia* 11: 143–170.
- Maeda, S., G. D. Price, M. R. Badger, C. Enomoto, and T. Omata. 2000. "Bicarbonate Binding Activity of the CmpA Protein of the Cyanobacterium *Synechococcus* sp. Strain PCC 7942 Involved in Active Transport of Bicarbonate." *Journal of Biological Chemistry* 275: 20551–20555.
- Manti, G., D. Mattei, V. Messineo, et al. 2005. "First Report of *Cylindrospermopsis raciborskii* in Italy." *Harmful Algae News* 28: 8–9.
- Martin, R. M., M. Kausch, K. Yap, J. D. Wehr, G. L. Boyer, and S. W. Wilhelm. 2021. "Metagenome-Assembled Genome Sequences of *Raphidiopsis raciborskii* and *Planktothrix Agardhii* From a Cyanobacterial Bloom in Kissena Lake, New York, USA." *Microbial Resource Announcements* 10: e01380-20.
- Mesquita, M. C. B., M. Lüring, F. Dorr, E. Pinto, and M. M. Marinho. 2019. "Combined Effect of Light and Temperature on the Production of Saxitoxins in *Cylindrospermopsis raciborskii* Strains." *Toxins* 11: 38.
- Messineo, V., S. Melchiorre, A. Di Corcia, P. Gallo, and M. Bruno. 2010. "Seasonal Succession of *Cylindrospermopsis raciborskii* and *Aphanizomenon Ovalisporum* Blooms With Cylindrospermopsin Occurrence in the Volcanic Lake Albano, Central Italy." *Environmental Toxicology* 25: 18–27.
- Moreira, C., A. Fathalli, V. Vasconcelos, and A. Antunes. 2015. "Phylogeny and Biogeography of the Invasive Cyanobacterium *Cylindrospermopsis raciborskii*." *Archives of Microbiology* 197: 47–52.
- Moustaka-Gouni, M., K. A. Kormas, E. Vardaka, M. Katsiapi, and S. Gkelis. 2009. "*Raphidiopsis mediterranea* Skuja Represents Non-Heterocytous Life-Cycle Stages of *Cylindrospermopsis raciborskii* (Woloszynska) Seenayya et Subba Raju in Lake Kastoria (Greece), its Type Locality: Evidence by Morphological and Phylogenetic Analysis." *Harmful Algae* 8: 864–872.
- Mugnai, M., M. Angela Margheri, C. Cristina Sili, et al. 2008. "The Cyanobacterial Community of Lake Trasimeno." *Algological Studies* 128: 37–64.

- Natwora, K. E., and C. S. Sheik. 2021. "Assessment of Nitrogen Fixation Rates in the Laurentian Great Lakes." *Journal of Great Lakes Research* 47: 1288–1295.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32: 268–274.
- Nguyen, T. T. L., T. H. Hoang, T. K. Nguyen, and T. T. Duong. 2017. "The Occurrence of Toxic Cyanobacterium *Cylindrospermopsis raciborskii* and Its Toxin Cylindrospermopsin in the Huong River, Thua Thien Hue Province, Vietnam." *Environmental Monitoring and Assessment* 189: 490.
- Nurk, S., D. Meleshko, A. Korobeynikov, and P. A. Pevzner. 2017. "metaSPAdes: A New Versatile Metagenomic Assembler." *Genome Research* 27: 824–834.
- Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, and D. McGlinn. 2020. "Vegan: Community Ecology Package."
- Orakov, A., A. Fullam, L. P. Coelho, et al. 2021. "GUNC: Detection of Chimerism and Contamination in Prokaryotic Genomes." *Genome Biology* 22: 178.
- Padisák, J. 1997. "Cylindrospermopsis raciborskii (Woloszynska) Seenayya et Subba Raju, an Expanding, Highly Adaptive Cyanobacterium: Worldwide Distribution and Review of Its Ecology." *Arch FÜR Hydrobiol Suppl Monogr BEITRAGE* 107: 563–593.
- Pan, S., C. Zhu, X.-M. Zhao, and L. P. Coelho. 2022. "A Deep Siamese Neural Network Improves Metagenome-Assembled Genomes in Microbiome Datasets Across Different Environments." *Nature Communications* 13: 2326.
- Panou, M., S.-K. Zervou, T. Kaloudis, A. Hiskia, and S. Gkelis. 2018. "A Greek *Cylindrospermopsis raciborskii* Strain: Missing Link in Tropic Invader's Phylogeography Tale." *Harmful Algae* 80: 96–106.
- Pareeth, S., M. Bresciani, F. Buzzi, et al. 2017. "Warming Trends of Perialpine Lakes From Homogenised Time Series of Historical Satellite and In-Situ Data." *Science of the Total Environment* 578: 417–426.
- Park, S.-C., K. Lee, Y. O. Kim, S. Won, and J. Chun. 2019. "Large-Scale Genomics Reveals the Genetic Characteristics of Seven Species and Importance of Phylogenetic Distance for Estimating Pan-Genome Size." *Frontiers in Microbiology* 10: 834.
- Parks, D. H., M. Chuvochina, C. Rinke, A. J. Mussig, P.-A. Chaumeil, and P. Hugenholtz. 2022. "GTDB: An Ongoing Census of Bacterial and Archaeal Diversity Through a Phylogenetically Consistent, Rank Normalized and Complete Genome-Based Taxonomy." *Nucleic Acids Research* 50: D785–D794.
- Piccini, C., L. Aubriot, B. D'Alessandro, F. Martigani, and S. Bonilla. 2013. "Revealing Toxin Signatures in Cyanobacteria: Report of Genes Involved in Cylindrospermopsin Synthesis From Saxitoxin-Producing *Cylindrospermopsis raciborskii*." *Advances in Microbiology* 3: 289–296.
- Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. "FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments." *PLoS One* 5: e9490.
- Pritchard, L., R. H. Glover, S. Humphris, J. G. Elphinstone, and I. K. Toth. 2016. "Genomics and Taxonomy in Diagnostics for Food Security: Soft-Rotting Enterobacterial Plant Pathogens." *Analytical Methods* 8: 12–24.
- Quinlan, A. R., and I. M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics—Oxford Academic* 26: 841–842.
- Rai, R., K. K. Rai, S. Singh, A. Raj, and L. C. Rai. 2021. "Stress Proteins and Signal Transduction in Cyanobacteria." In *Ecophysiology and Biochemistry of Cyanobacteria*, edited by R. P. Rastogi, 155–180. Springer Nature.
- Rajaniemi-Wacklin, P., A. Rantala, M. A. Mugnai, et al. 2005. "Correspondence Between Phylogeny and Morphology of *Snowella* spp. and *Woronichinia Naegeliana*, Cyanobacteria Commonly Occurring in Lakes." *Journal of Phycology* 42: 226–232.
- Revell, L. J. 2024. "Phytools 2.0: An Updated R Ecosystem for Phylogenetic Comparative Methods (And Other Things)." *PeerJ* 12: e16505.
- Rodriguez-R, L. M., R. E. Conrad, T. Viver, et al. 2024. "An ANI Gap Within Bacterial Species That Advances the Definitions of Intra-Species Units." *MBio* 15: e0269623.
- Rosselló-Móra, R., and W. B. Whitman. 2019. "Dialogue on the Nomenclature and Classification of Prokaryotes." *Systematic and Applied Microbiology* 42: 5–14.
- Rott, E., N. Salmaso, and E. Hoehn. 2007. "Quality Control of Utermöhl-Based Phytoplankton Counting and Biovolume Estimates—An Easy Task or a Gordian Knot?" *Hydrobiologia* 578: 141–146.
- Rzyski, P., and B. Poniedziałek. 2014. "In Search of Environmental Role of Cylindrospermopsin: A Review on Global Distribution and Ecology of Its Producers." *Water Research* 66: 320–337.
- Saker, M. L., and D. J. Griffiths. 2000. "The Effect of Temperature on Growth and Cylindrospermopsin Content of Seven Isolates of *Cylindrospermopsis raciborskii* (Nostocales, Cyanophyceae) From Water Bodies in Northern Australia." *Phycologia* 39: 349–354.
- Salmaso, N., C. Capelli, S. Shams, and L. Cerasino. 2015. "Expansion of Bloom-Forming *Dolichospermum Lemmermannii* (Nostocales, Cyanobacteria) to the Deep Lakes South of the Alps: Colonization Patterns, Driving Forces and Implications for Water Use." *Harmful Algae* 50: 76–87.
- Salmaso, N., L. Cerasino, A. Boscaini, and C. Capelli. 2016. "Planktic *Tychonema* (Cyanobacteria) in the Large Lakes South of the Alps: Phylogenetic Assessment and Toxicogenic Potential." *FEMS Microbiology Ecology* 92: fiw155.
- Salmaso, N., L. Cerasino, M. Pindo, and A. Boscaini. 2024. "Taxonomic and Functional Metagenomic Assessment of a *Dolichospermum* Bloom in a Large and Deep Lake South of the Alps." *FEMS Microbiology Ecology* 100: fiae117.
- Schoch, C. L., S. Ciuffo, M. Domrachev, et al. 2020. "NCBI Taxonomy: A Comprehensive Update on Curation, Resources and Tools." *Database* 2020: baaa062.
- Schwengers, O., L. Jelonek, M. A. Dieckmann, S. Beyvers, J. Blom, and A. Goesmann. 2021. "Bakta: Rapid and Standardized Annotation of Bacterial Genomes via Alignment-Free Sequence Identification." *Microbial Genomics* 7: 000685.
- Sha, Z., H. Chen, L. Jin, et al. 2025. "Evidence of Global Dispersal of the Harmful Cyanobacterium, *Raphidiopsis raciborskii*, in Lentic Freshwaters Through Migratory Waterbirds." *Harmful Algae* 142: 102786.
- Shaiber, A., A. D. Willis, T. O. Delmont, et al. 2020. "Functional and Genetic Markers of Niche Partitioning Among Enigmatic Members of the Human Oral Microbiome." *Genome Biology* 21: 292.
- Shvarev, D., and I. Maldener. 2018. "ATP-Binding Cassette Transporters of the Multicellular Cyanobacterium *Anabaena* sp. PCC 7120: A Wide Variety for a Complex Lifestyle." *FEMS Microbiology Letters* 365: fny012.
- Sieber, C. M. K., A. J. Probst, A. Sharrar, et al. 2018. "Recovery of Genomes From Metagenomes via a Dereplication, Aggregation and Scoring Strategy." *Nature Microbiology* 3: 836–843.
- Sinha, R., L. A. Pearson, T. W. Davis, et al. 2014. "Comparative Genomics of *Cylindrospermopsis raciborskii* Strains With Differential Toxicities." *BMC Genomics* 15: 83.
- Skuja, H. 1937. "Süßwasseralgen aus Griechenland und Kleinasien." *Hedwigia* 77: 15–73.

- Staley, J. T. 2006. "The Bacterial Species Dilemma and the Genomic-Phylogenetic Species Concept." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 361, no. 1475: 1899–1909. <https://doi.org/10.1098/rstb.2006.1914>.
- Stucken, K., U. John, A. Cembella, et al. 2010. "The Smallest Known Genomes of Multicellular and Toxic Cyanobacteria: Comparison, Minimal Gene Sets for Linked Traits and the Evolutionary Implications." *PLoS One* 5: e9235.
- Stumm, W., and J. J. Morgan. 1996. *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters*. Environmental Science and Technology: A Wiley-Interscience Series of Texts and Monographs.
- Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. "A Genomic Perspective on Protein Families." *Science* 278: 631–637.
- Tettelin, H., D. Riley, C. Cattuto, and D. Medini. 2008. "Comparative Genomics: The Bacterial Pan-Genome." *Current Opinion in Microbiology* 11: 472–477.
- Tokodi, N., K. Kłodawska, A. Willis, et al. 2025. "Mechanisms Responsible for the Ubiquity of Cyanobacterium *Raphidiopsis raciborskii*—Is Photosynthetic Apparatus a Key Player?" *Algal Research* 85: 103870.
- Tonkin-Hill, G., N. MacAlasdair, C. Ruis, et al. 2020. "Producing Polished Prokaryotic Pangenomes With the Panaroo Pipeline." *Genome Biology* 21: 180.
- Ulloa, O., C. Henríquez-Castillo, S. Ramírez-Flandes, et al. 2021. "The Cyanobacterium *Prochlorococcus* Has Divergent Light-Harvesting Antennae and May Have Evolved in a Low-Oxygen Ocean." *Proceedings of the National Academy of Sciences* 118: e2025638118.
- van Dongen, S., and C. Abreu-Goodger. 2012. "Using MCL to Extract Clusters From Networks." *Methods in Molecular Biology (Clifton, N.J.)* 804: 281–295. https://doi.org/10.1007/978-1-61779-361-5_15.
- Vernikos, G., D. Medini, D. R. Riley, and H. Tettelin. 2015. "Ten Years of Pan-Genome Analyses." *Current Opinion in Microbiology* 23: 148–154.
- Vico, P., S. Bonilla, B. Cremella, L. Aubriot, A. Iriarte, and C. Piccini. 2020. "Biogeography of the Cyanobacterium *Raphidiopsis (Cylindrospermopsis) raciborskii*: Integrating Genomics, Phylogenetic and Toxicity Data." *Molecular Phylogenetics and Evolution* 148: 106824.
- Vico, P., A. Iriarte, S. Bonilla, and C. Piccini. 2021. "Metagenomic Analysis of *Raphidiopsis raciborskii* Microbiome: Beyond the Individual." *Biodiversity Data Journal* 9: e72514.
- Vollmers, J., S. Wiegand, F. Lenk, and A.-K. Kaster. 2022. "How Clear Is Our Current View on Microbial Dark Matter? (Re-)assessing Public MAG & SAG Datasets With MDMcleaner." *Nucleic Acids Research* 50: e76.
- Watson, A. R., J. Füssel, I. Veseli, et al. 2023. "Metabolic Independence Drives Gut Microbial Colonization and Resilience in Health and Disease." *Genome Biology* 24: 78.
- Willis, A., and J. N. Woodhouse. 2020. "Defining Cyanobacterial Species: Diversity and Description Through Genomics." *Critical Reviews in Plant Sciences* 39: 101–124.
- Willis, A., J. N. Woodhouse, S. E. Ongley, A. R. Jex, M. A. Burford, and B. A. Neilan. 2018. "Genome Variation in Nine Co-Occurring Toxic *Cylindrospermopsis raciborskii* Strains." *Harmful Algae* 73: 157–166.
- Wood, S. A., and D. J. Stirling. 2003. "First Identification of the Cylindrospermopsin-Producing Cyanobacterium *Cylindrospermopsis raciborskii* in New Zealand." *New Zealand Journal of Marine and Freshwater Research* 37: 821–828.
- Yu, G. 2020. "Using Ggtree to Visualize Data on Tree-Like Structures." *Current Protocols in Bioinformatics* 69: e96.
- Zarenezhad, S., T. Sano, M. M. Watanabe, and M. Kawachi. 2012. "Evidence of the Existence of a Toxic Form of *Cylindrospermopsis raciborskii* (Nostocales, Cyanobacteria) in Japan." *Phycological Research* 60: 98–104.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.