AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

MICROBIAL ECOLOGY

Genotyping and Multivariate Regression Trees Reveal Ecological Diversification within the *Microcystis aeruginosa* Complex along a Wide Environmental Gradient

Gabriela Martínez de la Escalera,[a] Angel M. Segura,[b] Carla Kruk,[b,c] Badih Ghattas,[d] Frederick M. Cohan,[e] Andrés Iriarte,[f] Claudia Piccini[a]

[a]Departamento de Microbiología, Instituto de Investigaciones Biológicas Clemente Estable (IIBCE), Montevideo, Uruguay
[b]Modelización Estadística de Datos e Inteligencia Artificial, CURE-Rocha, Universidad de la República, Rocha, Uruguay
[c]Limnología, IECA, Facultad de Ciencias, Universidad de la República, Rocha, Uruguay
[d]Institut de Mathématiques de Marseille, Aix-Marseille University, Marseille, France
[e]Department of Biology, Wesleyan University, Middletown, Connecticut, USA
[f]Laboratorio de Biología Computacional, Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina, Montevideo, Uruguay

ABSTRACT Addressing the ecological and evolutionary processes underlying biodiversity patterns is essential to identify the mechanisms shaping community structure and function. In bacteria, the formation of new ecologically distinct populations (ecotypes) is proposed as one of the main drivers of diversification. New ecotypes arise when mutations in key functional genes or acquisition of new metabolic pathways by horizontal gene transfer allow the population to exploit new resources, permitting their coexistence with the parental population. We previously reported the presence of microcystin-producing organisms of the *Microcystis aeruginosa* complex (toxic MAC) through an 800-km environmental gradient ranging from freshwater to estuarine-marine waters in South America. We hypothesize that the success of toxic MAC in such a gradient is due to the existence of very closely related populations that are ecologically distinct (ecotypes), each specialized to a specific arrangement of environmental variables. Here, we analyzed toxic MAC genetic diversity through quantitative PCR (qPCR) and high-resolution melting analysis (HRMA) of a functional gene (*mcyJ*, microcystin synthetase cluster). We explored the variability of the *mcyJ* gene along the environmental gradient by multivariate classification and regression trees (*m*CART). Six groups of *mcyJ* genotypes were distinguished and associated with different combinations of water temperature, conductivity, and turbidity. We propose that each *mcyJ* variant associated with a defined environmental condition is an ecotype (or species) whose relative abundances vary according to their fitness in the local environment. This mechanism would explain the success of toxic MAC in such a wide array of environmental conditions.

IMPORTANCE Organisms of the *Microcystis aeruginosa* complex form harmful algal blooms (HABs) in nutrient-rich water bodies worldwide. MAC HABs are difficult to manage owing to the production of potent toxins (microcystins) that resist water treatment. In addition, the role of microcystins in the ecology of MAC organisms is still elusive, meaning that the environmental conditions driving the toxicity of the bloom are not clear. Furthermore, the lack of coherence between morphology-based and genomic-based species classification makes it difficult to draw sound conclusions about when and where each member species of the MAC will dominate the bloom. Here, we propose that the diversification process and success of toxic MAC in a wide range of water bodies involves the generation of ecotypes, each specialized in a particular niche, whose relative abundance varies according to its fitness in the local environment. This knowledge can improve the generation of accurate

Editor Hideaki Nojiri, University of Tokyo

Address correspondence to Claudia Piccini, cpiccini@iibce.edu.uy.
The authors declare no conflict of interest.

Received 26 July 2021
Accepted 17 November 2021
Accepted manuscript posted online 24 November 2021
Published 8 February 2022

prediction models of MAC growth and toxicity, helping to prevent human and animal intoxication.

KEYWORDS ecotypes, multivariate CART, *mcyJ*, *Microcystis aeruginosa* complex, HRMA

When a group of closely related bacteria dominates a broad diversity of habitats, two possibilities emerge to explain the group's success. First, each individual organism may be a generalist, able to succeed in a broad range of environments (1), possibly with some help by environmentally induced plasticity (2). The other possibility is that the group of organisms is an amalgam of populations that are ecologically distinct yet very closely related (3). Some studies have identified closely related groups of bacterial generalists, based on the abundance of the group in disparate habitats (1). However, these cases can be misleading when the group claimed to be an ecological generalist is broadly defined by membership in a species taxon. This is because the species taxa recognized by systematics frequently contain a diversity of ecologically distinct populations known as ecotypes (3). To determine whether a bacterial clade dominating diverse habitats is composed of interchangeable generalists or is an amalgam of specialists, we require a classification that delves below the species taxon level and focuses on the ecological preferences of individual ecotypes.

In the cyanobacteria, studies of ecological diversification have favored the specialist hypothesis. The picoplankton marine species *Prochlorococcus marinus* is found at a diversity of latitudes, temperatures, and light and nutrient levels, but these organisms are not individually generalists; the species taxon consists of many ecotypes that are specialized to different combinations of these environmental dimensions (4), such as temperature (5) and light intensities (6). This strategy allows this group of specialists to occupy the entire euphotic zone, together accounting for 50% of the total chlorophyll and 4 gigatons of carbon fixed per year in vast areas of the surface ocean (7). Also, extremely closely related hot spring bacteria of *Synechococcus* thrive in a diversity of temperatures and light levels in the spring mats, but they are not generalists either; colonization and success in different habitats is due to specialization among ecotypes adapted to different conditions and resources (8).

Slowly evolving phylogenetic markers, such as ribosomal genes, do not allow us to elucidate changes occurring at the rapid interface between ecological and evolutionary processes. The ecotype theory of bacterial species (3, 9) defines the ecotype as a clade of phylogenetically related microorganisms that differ in their ecological characteristics. An ecotype generates after a single individual experiences a mutation or recombination event that changes its autoecology, allowing the utilization of a new set of resources or the ability to thrive under a particular environmental condition (10), and has been proposed as a main driver of bacterial speciation (11, 12). Under this framework, ecotypes are defined to be ecologically distinct from one another, while each is ecologically homogeneous, and the species taxa recognized by bacterial systematics often contain multiple ecotypes.

The *Microcystis aeruginosa* complex (MAC) comprises the monophyletic clade of all the species of the *Microcystis* genus and related genera (e.g., *Radiocystis* and *Sphaerocavum*) that share phenotypic and phylogenetic characteristics with the species *Microcystis aeruginosa* (13–15). They are the most common bloom-forming cyanobacteria worldwide, able to develop large blooms in freshwater and brackish ecosystems (16–20) and produce microcystin, a toxin known to cause serious liver and neural damage in humans and other vertebrate and invertebrate animals (21–26). Ten genes spanning 55 kb have been described as involved in microcystin production, *mcyA* to *mcyJ* (27). The use of *mcy* genes to analyze toxic MAC diversity has been previously reported by several authors (28–30). For example, Kim et al. (29) used a *mcyJ* gene fragment as a surrogate to detect toxic *Microcystis* and to characterize and address the dynamics of different *mcyJ* genotypes using denaturing gradient gel electrophoresis (DGGE) (29). Similarly, Hu et al. (30) performed DGGE, but using *mcyA* amplicons, and found that the dynamics of toxin

production was related to *Microcystis* community structure (30). However, multiple recombination events, gene loss, or horizontal gene transfer have been detected within the *N*-methyltransferase-domain of *mcyA* and the adenylation-domain of *mcyB* and *mcyC* sequences (31, 32). In the case of the *mcyJ* gene, it has been found as a single copy in the genome of toxic strains, and no recombination has been detected, meaning that it could be suitable for studying a toxic population's variability (33).

There are several morphologically distinct groups within the MAC sharing nearly identical 16S rRNA gene sequences but exhibiting ecological distinctness and different toxicity potentials (27, 34, 35), which have been classified as different species of the *Microcystis* genus. These morphology-based species assignments are not always genomically or phylogenetically coherent (36), making it difficult to better understand their ecological features in order to predict their presence and toxicity based only on the 16S rRNA gene. So, in order to understand MAC success across aquatic ecosystems worldwide, an adequate and sensitive approach to define and detect phylogenetic and ecologically coherent taxa (or ecotypes) is needed. Recently, we published evidence that the *Microcystis aeruginosa* complex thrives along an extended environmental gradient in the Uruguay River-Río de la Plata estuary, from its headwaters to its brackish estuary (37). As a result, *mcyB*, *mcyD*, *mcyE*, and *mcyJ* genes were detected by quantitative PCR (qPCR) throughout the ecosystem. The highest number of *mcy* gene copies was detected in summer (2×10^4 to 22×10^4 copies/mL) and spatially decreased from the reservoir to the marine sites (7 to 250 copies/mL). Moreover, the abundance of *mcy* genes was correlated positively with traditional phytoplankton indicators such as chlorophyll *a*, total phytoplankton biovolume, and cyanobacterial biovolume. Therefore, as there is no evidence of recombination in the *mcyJ* gene, it is present in all the assessed genomes of toxic MAC, and there are primers available to study its variability (29) that were already applied to our study system (37), we selected the *mcyJ* gene as a genetic marker to address toxic *Microcystis* diversification.

Here, we investigate whether the success of the MAC group across disparate habitats along the Uruguay River is due to ecological diversification into multiple ecotypes specialized on different conditions. The methods and algorithms used to identify bacterial ecotypes are related to PCR amplification of phylogenetic marker genes and posterior analysis of the amplicons, either by fingerprinting methods (e.g., DGGE, temperature gradient gel electrophoresis [TGGE]) (38) or by sequencing (4, 5). Fingerprinting-based methods allow distinction of sequences differing by as little as a single nucleotide, and in the case of sequencing (e.g., amplicon sequencing of 16S genes or functional genes such as phycocyanin operon), a similarity cutoff is usually applied to define taxa at different phylogenetic levels. This cutoff can produce different results, depending on how broadly the taxa are defined. Thus, identifying ecotypes by this approach depends on the stringency of taxon assignment (4). High resolution melting analysis (HRMA) is a fingerprinting-based method that has been used to study diversity at the genotype level (39, 40). The HRMA denaturation curves give information about the whole amplicon sequence, detecting single nucleotide polymorphisms (SNPs). Due to great precision at high throughput and relatively low cost, it has been applied to genotyping, mutation scanning, and SNP detection in human diseases (41) and bacterial populations (42–45) and to address the diversity of microbial communities (39, 40, 46). The output of HRMA denaturation curves presents infinite dimensionality and autocorrelation, which require the use of functional data approaches. HRMA curves reflect the nonlinear adaptive response of microbial communities to environmental conditions amenable to analysis by robust machine learning techniques (47).

We developed, tested, and applied an approach to detect microbial ecotypes by combining molecular and machine learning tools to evaluate the hypothesis that toxic MAC is composed of multiple ecotypes adapted to thrive in environmental conditions ranging from freshwater to a brackish estuary and from warm to cold water. We used water samples taken from a large environmental and spatial gradient (~800 km), ranging from a freshwater reservoir (Salto Grande reservoir in the Uruguay River) to the

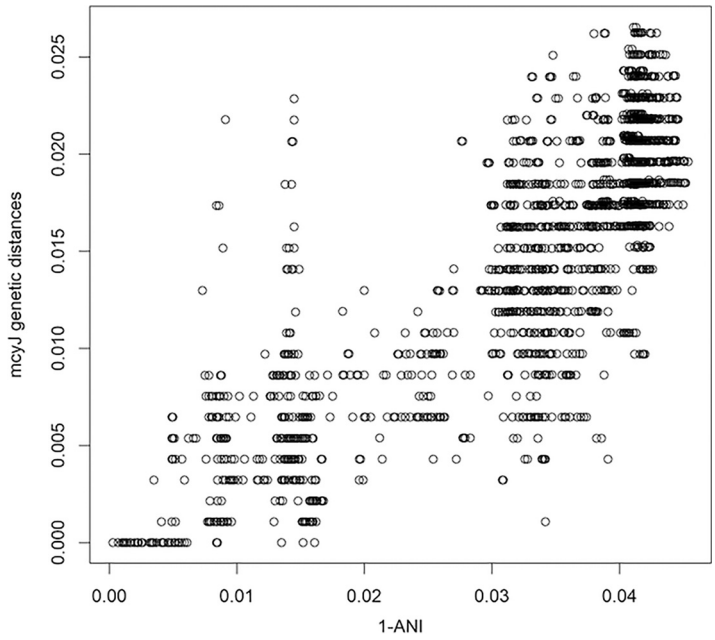

FIG 1 Spearman correlation between the nucleotide distances of the *mcyJ* gene and the whole genomic distance, measured as 1 − ANI, for the toxic MAC species ($r_s = 0.78$, $P < 0.05$).

mouth of the largest South American estuary (Río de la Plata). Under the working hypothesis that the success of toxic MAC in such a gradient is due to the existence of very closely related populations that are ecologically distinct (ecotypes), we used large-scale microbial data obtained from the samples to detect different genotypes of toxic MAC using the HRMA curves generated by amplicons of a gene involved in microcystin biosynthesis (*mcyJ*). First, we validated the rationale of the method using *in silico* simulated data (see below in Materials and Methods). Then the method was applied to natural samples from the gradient, and the genotypes were classified based on sequence differences and environmental variables by a multivariate classification and regression trees approach (*m*CART). We discuss the results in the framework of the ecotype theory and its impact to detect processes at the interface between ecology and evolution.

RESULTS

***mcyJ* as a tool for ecotype identification.** In order to confirm the suitability of the gene *mcyJ* as a phylogenetic marker of toxic MAC, we first analyzed the average nucleotide identity (ANI) between all of the published genomes of MAC and found that it ranged from 0.95 (±0.02) to 1.00 (±0.53), indicating that according to standard criteria, all the available *Microcystis* genomes belong to the same species (48, 49). The estimated nucleotide distance of *mcyJ* sequences taken from the available genomes of toxic MAC species was positively and significantly correlated with the average nucleotide distance for the whole genomes, measured as 1 − ANI (used as a dissimilarity measure) ($r_s = 0.78$, $P < 0.05$) (Fig. 1). A maximum likelihood phylogenetic tree was built for *mcyJ* and then compared with a reference tree built based on the concatenated alignment of 28 highly conserved ribosomal proteins from the same genomes. The estimated global distance between both phylogenetic trees was 0.53, and when the similarity between both trees was calculated for each node we found that the most recent ones tend to have a higher similarity than most basal ones (see Fig. S1 and S2 in the supplemental material).

Rationale for ecotype identification. To explain the basic ideas on how the ecotypes of toxic MAC were identified, we give a synthetic example with the operative steps of the method. First, we assumed that the *mcyJ* melting profile obtained from a

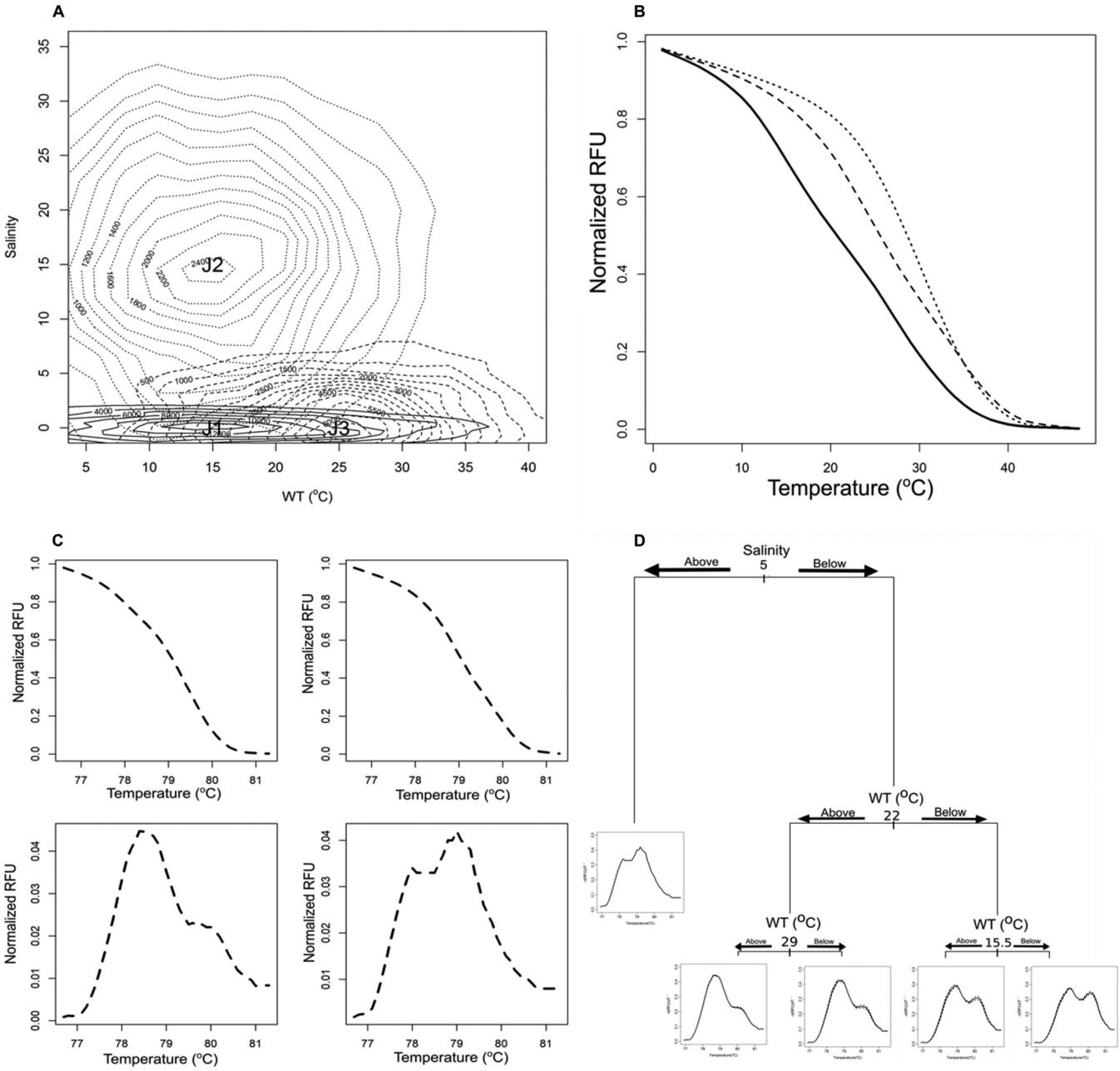

FIG 2 Synthetic example with the rationale and operative steps followed in the proposed methodology. (A) Abundance and distribution of three simulated *mcyJ* genotypes (J1, J2, and J3) according to environmental characteristics (e.g., temperature and salinity). (B) Normalized melting curves from the three HRMA curves obtained from three cloned *mcyJ* genes individually analyzed—J1 (solid line), J2 (dashed line), and J3 (dotted line). (C) Normalized HRMA curves obtained from a community composed of J1, J2, and J3 genotypes. Upper left, curve obtained from freshwater at 33°C and its derivative below; upper right, curve obtained from marine water (salinity 33) at 13°C with its derivative below. (D) Optimal multivariate regression tree defining ecotypes sampled from specific environmental conditions. In each node, the environmental variable and its threshold value are shown. Water temperature (WT) and salinity are shown. At the end of each branch the average melting peak (solid line) and its standard deviation (dashed line) representing the toxic genotype community are shown.

natural water sample containing MAC organisms (e.g., a sample with three different *mcyJ* genotypes) represents the abundance-weighted average profile of the genotypes present in that sample. Here, we used three *mcyJ* genotypes (J1, J2, and J3) belonging to previously obtained *mcyJ* clones showing different melting curves (Fig. 2B) (34) to generate a synthetic MAC community. After running simulations and in agreement with our assumption, the combination of the hypothetical environmental drivers I and II (e.g., temperature and salinity) (Fig. 2a) deterministically defined the relative

abundance of the three *mcyJ* genotypes, which is revealed in their individual melting profiles (Fig. 2B). So, under a particular environmental condition (I_1 and II_1), we expect to find a particular combination of genotype relative abundances in the sample (Fig. 2C). The relative abundance of each genotype under each environmental array of conditions determines a distinctive melting profile of that sample, which accounts for the dominant ecotype (Fig. 2C; dashed line). To put it another way, the melting profile of a given sample results from an assemblage of distinct melting profiles whose relative abundances will depend on the values of the environmental variables. For our simulation, the relationship between abundance and environmental conditions was constructed from bivariate normal distributions with average $\mu_i = \{\mu_I, \mu_{II}\}$ and covariance matrix σ_i specified for each genotype, where $i = 1{:}3$. After obtaining the melting curves from each genotype, we randomly sampled environmental conditions 50 times, and for each condition (defined by the pair {I,II}), the relative abundance of each genotype was estimated and the melting profile was constructed. Under our hypothesis, each region in the environmental space should be characterized by an identical melting profile corresponding to the dominant toxic ecotype (Fig. 2D). It is important to recall that environmental drivers can be continuous or categorical and explicitly include biotic interactions among the defining variables. Therefore, the ecotypes can be either completely specialized into different environments or just quantitatively specialized, where ecotypes differ in their preferences but overlap in the environmental distributions. In the extreme case where the abundances of the three *mcyJ* amplicons are not determined by the environmental variables selected ($\mu_i = \mu_j$ and $\sigma_i = \sigma_j$ for all i, j), a single ecotype is expected. Under this situation, the multivariate CART would not partition the data into separated ecotypes, and the tree will remain as a root tree.

Defining ecotypes of MAC. The analysis of HRMA profiles using the *m*CART technique yielded the responses of the toxic MAC community to a wide environmental gradient, splitting the data into 6 groups of *mcyJ* genotypes with distinct and specific environmental preferences. We propose these groups as toxic MAC ecotypes. These ecotypes (named A to F) were detected using 72 normalized melting profiles. Three environmental variables were selected by the model as the most relevant explaining the differences between the ecotypes: temperature, salinity, and turbidity (Fig. 3; Table 1). The variables wind intensity, total nitrogen (TN), and total phosphorus (TP) were not selected by the model. Figure 4 shows the average prediction error obtained from the optimal tree (Fig. 3) and the prediction errors from each permuted tree (see "Multivariate CART for functional analysis," below). The average prediction error of the optimal tree is represented as a line, and the prediction errors from permuted trees are represented as a density plot (Fig. 4). The optimal tree had an average prediction error significantly lower than the average error obtained from permuted trees (log-likelihood ratio test [LRT], $P < 0.05$; Fig. 4).

Water temperature was the first selected variable, splitting groups of toxic MAC into those thriving at temperatures higher (above) or lower (below) than 14.5°C (Fig. 3). The next two selected variables were water turbidity and conductivity, with threshold values of 14.2 nephelometric turbidity units (NTU) and 0.104 millisiemens (mS) cm^{-1}, respectively. Then, intermediate conductivity values (9.16 mS cm^{-1}) and high water temperature (23.8°C) were selected as splitting variables. Total nutrients (nitrogen and phosphorus mean values of 0.74 mg L^{-1}, 52 ug L^{-1}, respectively) and wind intensity were not relevant to define toxic ecotypes. Ecotype A was present in brackish waters at temperatures higher than 14.5°C and more than 14.2 turbidity units, while ecotype B was found under the same turbidity conditions as A but in freshwater and at high water temperature (>23.8°C; Table 1). Ecotype C was the most frequent; it preferred warm freshwaters and had a large amount of potential toxicity (assessed by qPCR as the number of *mcyJ* copies per mL), conditions found at the reservoir. On the other hand, ecotype D inhabits water with a temperature of >14.5°C but has a wide range of conductivity preferences. Finally, ecotypes E and F occurred in cold water (water temperature, <14.5°C) but slightly differed in their conductivity preferences (0.104 mS cm^{-1} conductivity threshold) (Fig. 3). Ecotypes B and C, which belong to the reservoir, were those with a higher

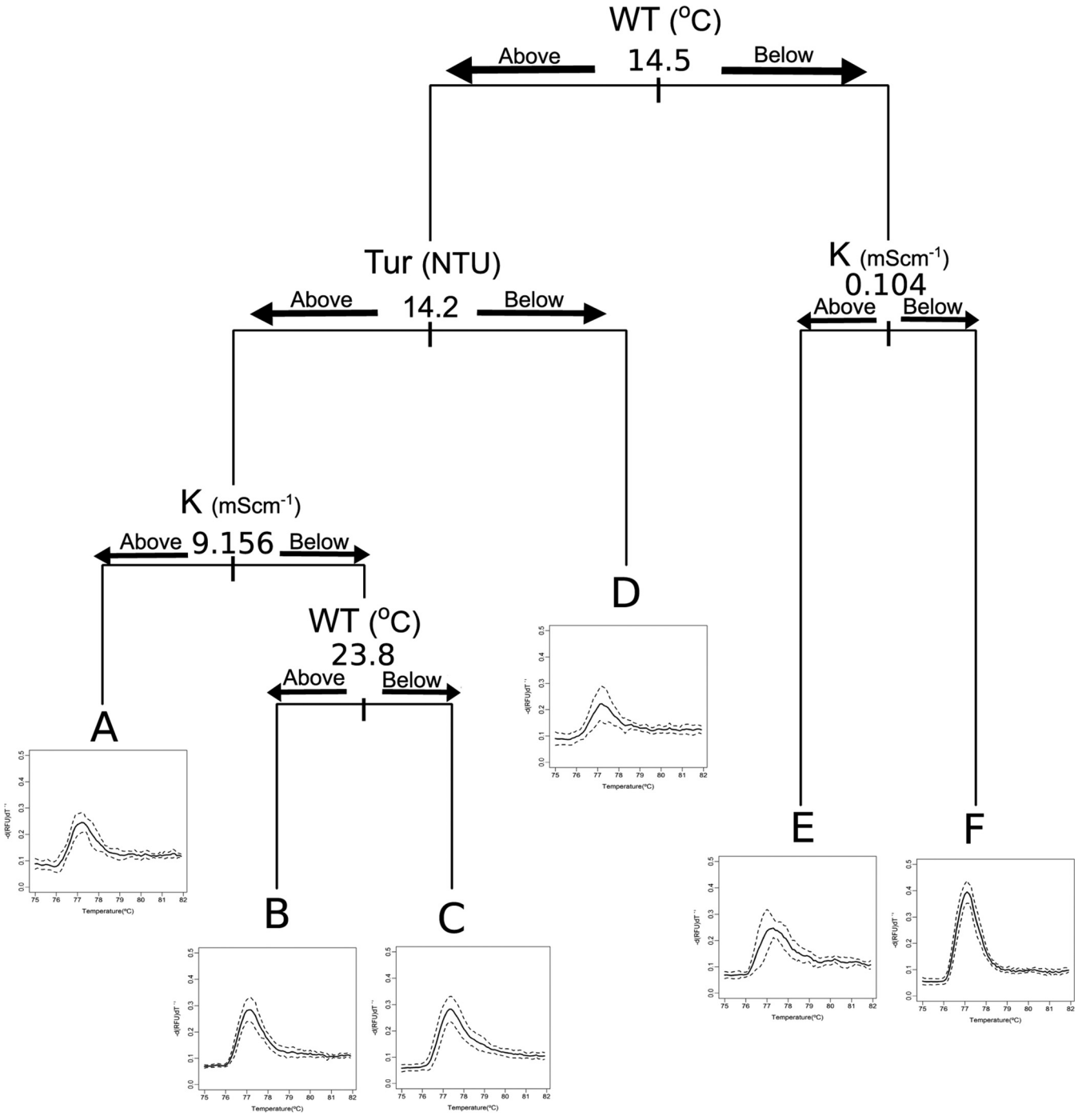

FIG 3 Ecotypes defined by the multivariate regression tree (*m*CART). Multivariate regression tree showing the main environmental variables explaining the profile diversity of toxic genotypes. The selected environmental variable and its threshold value are shown at each node—water temperature (WT), turbidity (Tur), and conductivity (K). At the end of each branch, the abundance-weighted average melting curves for each toxic MAC sample (solid lines) and its standard deviation (dashed lines) are shown.

abundance of toxic cells per mL, as evidenced by qPCR. In contrast, ecotype E, belonging to the middle and outer estuary, exhibited the lowest number of toxic cells (Table 1).

DISCUSSION

Microbial species can be seen as genetically, phenotypically, and ecologically similar units that are selectively optimized to either coexist occupying different niches or to overlap, not only genetically but also ecologically (50). Here, we used the communities

TABLE 1 Description of niche and main traits of each ecotype of toxic MAC[a]

Ecotype	Niche	Zone	Temp (°C)	Cond (mS cm^{-1})	Tur (NTU)	*mcyJ* gene (copies ml^{-1})	Frequency of occurrence
A	Warm brackish-marine water	Middle and outer estuary	19.1 (14.8–22.9)	36.4 (18.2–52.0)	40.9 (15.4–89.6)	34.9 (BDL–96.0)	0.14
B	Hot freshwater	Reservoir and riverine freshwater	27.2 (24.3–33.6)	0.041 (0.023–0.054)	28.5 (14.3–47.9)	2.30E4 (BDL–1.14E5)	0.14
C	Warm freshwater	Reservoir and riverine freshwater	20.5 (15.6–23.4)	0.074 (0.033–0.113)	56.0 (19.9–127.0)	1.21E4 (BDL–4.54E4)	0.31
D	Warm water and low turbidity	Inner and middle estuary	20.9 (16.1–25.9)	14.02 (0.048–52.7)	7.6 (BDL–14.2)	46.2 (BDL–148.7)	0.20
E	Cold brackish-marine water	Middle and outer estuary	11.9 (11.2–12.3)	29.13 (0.113–55.8)	20.8 (BDL–49.0)	11.7 (BDL–44.7)	0.11
F	Cold freshwater	Reservoir and riverine freshwater	12.7 (11.0–14.2)	0.061 (0.036–0.095)	17.5 (0.9–42.2)	192.7 (0.3–437.2)	0.09

[a]Mean values and ranges of environmental variables associated with each ecotype: temperature (Temp, °C), conductivity (Cond, mS cm^{-1}), turbidity (Tur, NTU), *mcyJ* gene abundance (copies ml^{-1}), and the relative frequency of occurrence of each ecotype. BDL, below detection limit.

of toxic *Microcystis* spp. as model organisms to determine whether this clade, able to succeed in water bodies around the world (from freshwater lakes to estuaries), is composed of interchangeable generalists or is a combination of ecologically distinct specialists that proliferate when the environmental conditions are ideal for each specialized lineage.

Because the identification of toxic MAC species using classical molecular marker genes such as the 16S rRNA gene is not always possible (51, 52), the present approach involved the analysis of the variability of a functional gene related to microcystin synthesis (*mcyJ*) as a molecular marker to differentiate between toxic taxa. Here, we combined a molecular method for genotyping (HRMA of *mcyJ* gene amplicons) with machine learning techniques (*m*CART) to achieve a high discriminatory power and to discover the relationships between MAC community structure and the environment.

We used two criteria to determine if *mcyJ* could yield an accurate measure of the evolutionary history of this group. First, we found a significant correlation between the pairwise distances of *mcyJ* genes with the average genomic distances (1-ANI), and second, we performed a comparison between the phylogeny constructed from orthologous genes (highly conserved ribosomal-protein sequences) and the *mcyJ*-based phylogeny. We found that the evolution of the *mcyJ* gene mirrors the evolution of the different analyzed lineages (Fig. S2). Since the global distance between the trees was 53%, we analyzed the similarity node by node and found that is generally higher at the tips of the trees, implying that

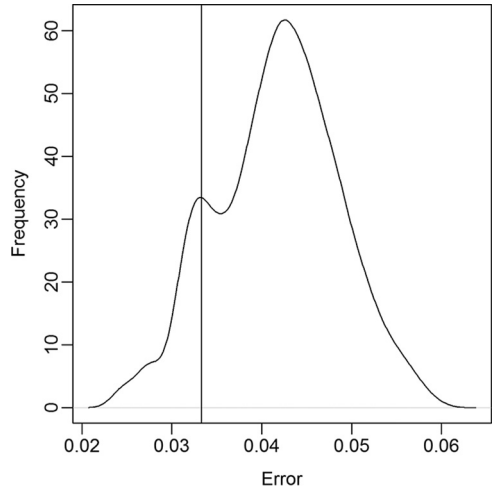

FIG 4 Errors obtained for multivariate regression trees. The density of the tree-error performed by randomly switching the values of environmental variables and vertical line denotes the mean value of the error obtained using the original data.

mcyJ is a suitable phylogenetic marker for this group, especially for recently divergent lineages of *Microcystis*. Other well-known markers, such as 16S rRNA or ribosomal protein-coding genes (data not shown) were also tested, but these genes were found to be too conserved for this clade. As a result, the resolution power of these markers for this genus when used independently is marginal. This is why we explored other phylogenetically informative genes in order to detect ecologically coherent genotypes. Rinta-Kanto and Wilhelm used the *mcyA* gene to address the genetic diversity of potentially toxic *Microcystis* (28). Similarly, Hu et al. applied DGGE to *mcyA* amplicons and found that microcystin variants were related to the obtained band pattern, implying that the composition of the *Microcystis* community determined the type of microcystin produced (30). However, it has been shown that *mcyA* has several recombination regions (32), and multiple recombination events were detected within the *N*-methyltransferase domain of *mcyA* and the adenylation-domain of *mcyB* and *mcyC* sequences, suggesting genetic exchange within and between *mcy* genes. This would hinder the use of *mcyA to mcyC* genes as phylogenetic markers for toxic populations. On the other hand, no recombination was detected in the *mcyJ* gene, which is present at a single copy in toxic strains and is conserved enough to address the variability of toxic populations (32).

Different roles related to the abiotic and biotic environment have been proposed for microcystin toxins, among them to increase organisms' fitness by adaptation to low CO_2 levels (53, 54), a protein-modulating role (55), for colony formation (56), and as a protection against oxidative stress (57). Thus, genetic variation in the microcystin synthesis genes might reflect different adaptations to the local environment. In this context, the present approach is a novel way to gain insight into MAC diversity and to objectively define groups with a genetic signature underlying the worldwide success of this bloom-forming cyanobacterium. An advantage of using an HRMA-based method to explore sequence diversity instead of amplicon sequencing and further analysis of the reads relies on the fact that, although both are based on PCR of a target gene, the former is faster (real-time data acquisition) and does not depend on identity cutoffs to define operational taxonomic or phylogenetically coherent units (58). This study complements and expands on previous work using HRMA to analyze the diversity of natural microbial communities (based on the 16S rRNA gene) (39, 40, 46) by combining new genetic regions as diversity markers and machine learning tools. In addition, *m*CART provides a nonarbitrary cutoff that is related to the ecological distinctness of the genotypes.

The present research shows that natural communities of MAC along the studied gradient are composed of genotypes that are ecologically coherent specialists or ecotypes. This suggests that the mechanism of the MAC's success over diverse habitats has been the generation of a number of ecotypes with distinct and specific niches and agrees with previous work showing that communities of toxic MAC cyanobacteria are able to inhabit a wide range of environmental conditions (37). It is also consistent with the worldwide distribution and current proliferation of the MAC and has been found for other cyanobacteria such as *Prochlorococcus* (59), for which at least six ecotypes differing in physiology and occupying distinct niches in the ocean have been found (4, 59, 60). Different ecotypes of *Cylindrospermopsis raciborskii* have also been described based on phylogenetic markers, morphology, tolerance to different light intensities, affinities to low or high phosphate concentrations, and toxicity (61–63). Altogether, the findings suggest that the generation of ecotypes might be a common cyanobacterial strategy to proliferate and succeed in aquatic ecosystems all over the world.

We found that variability in physical and hydrological variables (temperature, salinity, and turbidity) defined MAC ecotypes, while nutrient concentrations did not, suggesting a prominent role of local conditions over trophic state, probably due to the high availability of nutrients at all sites and sampling times (TP, $\sim$60 μg L^{-1}; TN, $\sim$0.9 mg L^{-1}) (Fig. 3, Table 1). These variables are directly related to the physical aspects that characterize ecosystem dynamics in space and time. For example, turbidity and conductivity define the estuarine portion of the system, while temperature

defines seasonal changes. The ecotypes of toxic MAC that were associated with brackish water (A and E) had a small number of toxic cells, meaning that toxic MAC are rare in the estuary. These results agreed with the general decrease of MAC biomass and abundance of *mcy*-harboring cells found from fresh to marine water (37, 64). Salinity concentration above the organisms' optimum causes osmotic stress, decreases photosynthesis rates, and might induce cell lysis (65–68), leading to a decrease in abundance and biomass and precluding their detection by classical microscopy counts (69). Recent studies demonstrated that some *M. aeruginosa* strains acquired the ability to produce an osmoprotectant (such as sucrose) by horizontal gene transfer, generating salt-tolerant genotypes (70). Thus, toxic MAC ecotypes found in the estuary may be composed by this kind of salt-tolerant organism.

Several studies identified different toxic MAC genotypes along temporal and spatial gradients in freshwater ecosystems, related mainly to nutrient availability (71–74). However, under eutrophic or hypereutrophic conditions, such as in the ecosystems studied here, nutrients are no longer shaping genotype richness. Kim et al. found that *mcyJ* gene diversity assessed by DGGE was reduced during summer compared with spring and autumn, pointing to an effect of water temperature in the selection of different toxic populations, as found here (29).

The six ecotypes of toxic MAC associated with different environmental settings through the assessed environmental and spatial (~800 km) gradient led us to hypothesize that the reservoir, which displays high MAC biomass and diversity through the whole year, might act as a source or seed bank of toxic ecotypes. As toxic MAC organisms are transported downstream through the Uruguay River and into the Río de la Plata estuary, populations would be locally selected by environmental conditions, allowing the dominance of different ecotypes. Sabart et al. studied spatial-temporal changes of *Microcystis* diversity in interconnected freshwater ecosystems (reservoirs, ponds, and a river) based on the internal transcribed spacer (ribosomal ITS) (66). They found that *Microcystis* populations were genetically different over short distances (~20 km) and that populations observed in the main reservoir were different from those found downstream. However, at a larger, global geographical scale, the connections between phylogenetic relationships of *Microcystis* communities and the environmental conditions were not detected (75). A possible explanation is that fast-evolving molecules, such as ribosomal ITS, can exhibit high levels of homoplasy, which increases the noise in the phylogenetic signal and avoids pattern detection. Here, the same ecotypes were detected in sites as far as ~500 km away, revealing that local environmental conditions are more relevant than distance for selection. This rules out alternative hypotheses of ecotype generation in subsidiary ecosystems. Nevertheless, more work will be needed to describe these ecotypes genomically and phylogenetically to elucidate the relative ages of the lineages' splitting.

In sum, by combining a high-resolution molecular method with a machine learning technique, we identified six ecotypes of cyanobacteria from the *Microcystis* genus, which are increasingly found in freshwater to brackish water blooms worldwide. This finding provides new insight into the ecological and evolutionary strategy that makes this taxon so successful across a range of environmental conditions.

MATERIALS AND METHODS

Study site. The study area is located in the subtropical region of South America and covers an extension of ca. 800 km, from the Salto Grande reservoir in the Uruguay River (31° 11′ latitude, 57° 52′ longitude) to Punta del Este (34° 57′ latitude, 55° 02′ longitude) at the marine end of the Río de la Plata estuary. Six sites were sampled every 2 months for 1 year (from January 2013 to March 2014), and subsurface (~0.5 m) samples were taken at coastal stations (0.01 to 0.5 km) (for more details see references 37 and 64). In total, 36 water samples were analyzed. The system presents strong temporal and spatial gradients in terms of temperature, conductivity (a proxy of salinity), and turbidity (37, 76–78). The highest surface water temperatures in the Salto Grande reservoir are usually recorded during summer (January to March, 33°C), while the lowest temperatures belong to the outer marine zone of the Río de la Plata during winter to early spring (June to October, 11°C). Conductivity is at a minimum at the freshwater sites (0.023 mS cm^{-1}; Salto, Fray Bentos, Carmelo, and Colonia) and maximum at the marine end of the estuary (55 mS cm^{-1}; Montevideo and Punta del Este). Turbidity ranged from 0 to 187 NTU, with

higher values at the middle of the gradient (Carmelo and Colonia). Higher concentrations of nutrients were measured in the Salto Grande reservoir and Montevideo (total phosphorus, 60 mg L^{-1}; total nitrogen, 0.9 mg L^{-1}). MAC biovolume, calculated by multiplying mean individual volume by its abundance in each sample measured by microscopy (see reference 37 for further information) was highest in the Salto Grande reservoir during summer and decreased toward the marine end (Punta del Este).

Strategy. High-resolution melting analysis (HRMA) is a post-real-time PCR method used to identify genotypes based on the detection of single nucleotide polymorphisms (SNPs). After real-time PCR, a melting analysis is performed by gradually heating the amplicons at 0.1°C steps. During this process, as the temperature increases, the melting point of the amplicon is reached and amplicon DNA denatures, melting apart the double strand and causing the fluorescence of the attached dye, used to visualize the amplification, to fade away. This melting behavior and concomitant fluorescence decay are represented as melting curves (fluorescence decay during melting temperature increase) and are distinctive of each sequence within a mixed sample. Thus, they can discriminate samples according to their sequence length, GC content, and strand complementarity. This is the case when analyzing sequences from single isolates; however, in this work our goal was to detect variations between melting curves of the *mcyJ* gene from different MAC lineages that could be present in an environmental, complex sample. Therefore, the resulting melting curve corresponds to the abundance-weighted average melting profile of all the *mcyJ* gene sequences present in the sample. The relationships between these melting profiles and the environmental variables were evaluated with multivariate classification and regression trees (*m*CART) (79, 80).

Classification and regression trees (CART) is a machine learning technique based on computational statistical methods (81, 82). A classification or regression tree is constructed by recursive binary partitioning of the response variable into regions that are increasingly homogeneous (i.e., nodes) until no improvement is possible, and the final nodes are called leaves. In regression trees, each node contains the predictor variable (e.g., an environmental variable) that results in the most homogeneous partition of the response variable (e.g., a biological variable) measured by the sum of squared errors (SSE), whose selection is based on an optimization process (83). This keeps on going until no more reduction of SSE is achieved. These methods are easily interpretable and provide simple above (>) or below (<) decision trees (78). Multivariate CART (*m*CART) is an extension of classical CART used in ecology (79, 80, 84). In *m*CART, the response variable is no longer an individual value but a variety of independent values (80). In the present context, the response variables are the coefficients of the melting curve functions. This method was used to evaluate the relation of environmental variables and toxic genotypes and identify groups of closely related toxic genotypes of MAC (ecotypes) exhibiting the same environmental preferences.

***mcyJ* gene variation analysis.** We performed further analyses to support the use of *mcyJ* as a proxy for the whole-genome divergence. First, we compared *mcyJ* variability with that of the *Microcystis* genome using the genomes available at the NCBI assembly database (Table S1). We found *mcyJ* genes in 91 of 165 available genomes. The pairwise two-way ANI score was computed among *Microcystis* genomes using the ani.rb script (71) available at enveomics.blogspot.com. The ANI score is the average nucleotide identity across the part of the genome that is orthologous between two organisms. Then, the *mcyJ* sequences were retrieved from each genome using the tblastn tool (85), and the pairwise genetic distances were calculated using the Kimura two-parameter substitution model with gamma distribution rate variation among sites in MEGA7 software (86). The two measures of genetic distance were compared by means of the Spearman correlation coefficient (r_s). We also compared the phylogenetic tree generated with *mcyJ* with a reference tree generated with a concatenated alignment of 28 highly conserved ribosomal-protein sequences (Fig. S1). In brief, alignments were performed with Clustal Omega (87). Model selection and phylogenetic analyses were performed using IQ-Tree (88) under ultrafast bootstrap modality (89). The model selected by IQ-Tree was JTTDCMut+F+I+G4, that is, the revised JTT matrix published by Kosiol and Goldman (90), plus empirical amino acid frequency, invariable sites, and discrete Gamma model (91) with default 4 rate categories.

Comparisons between the two generated trees were performed with the treeDistance function of the TreeDist package in R. This function calculates the global distance based on the amount of phylogenetic information that both trees hold in common as proposed by Smith (92). The similarity between different lineages was calculated and visualized using the phylo.io server (93).

DNA extraction. For the DNA extraction, 250- to 300-mL amounts of the subsurface water samples were collected with clean plastic 20-L carboys and then filtered through a 0.22-μm sterile polycarbonate membrane (Millipore, Darmstadt, Germany), which was immediately frozen at −20°C until processing. Procedures for nucleic acid extraction were performed as described in reference 94.

Real-time PCR. (i) Quantification of the *mcyJ* gene in the environmental samples. The density of toxic cells in each ecotype was quantified based on the amount of *mcyJ* amplified in qPCR. First, 2 μL of DNA extracts from each sample (ca. 50 ng DNA) were applied to the Power SYBR green PCR kit (Invitrogen) with a final reaction mixture volume of 20 μL. Primers for the *mcyJ* gene were those from Kim et al. (29). Cycling conditions were 2 min at 50°C, 15 min at 95°C, and 40 cycles of 15 s at 94°C, 30 s at 60°C, and 30 s at 72°C, including a final melting step from 65 to 95°C, at increases of 1°C each 4 s (34). A 96 FLX Touch TM thermal cycler (Bio-Rad) was used. To quantify the abundance of the *mcyJ* gene, cloned amplicons (34) were used to perform the calibration curves. Curves were achieved using five serial dilutions from 1/10 to 1/100,000 of the cloned genes (in quintuplicate) and applied to qPCR in the same PCR plate where the samples were assayed. Samples were run in triplicate.

(ii) High-resolution melting analysis (HRMA) of *mcyJ* amplicons. Amplification of the *mcyJ* gene was performed using the HRMA primers described in the literature (83). PCR amplification was

conducted using a 96 FLX Touch TM thermal cycler (Bio-Rad, California, USA). First, 2 μL of DNA extracts from each sample (ca. 50 ng DNA) were applied to the MeltDoctor HRM master mix (Applied Biosystems, California, USA) with a final reaction mixture volume of 20 μl. Cycling conditions were 2 min at 50°C, 15 min at 95°C, and 40 cycles of 15 s at 94°C, 30 s at 60°C, and 30 s at 72°C. To generate the HRMA melting profiles, the fluorescence obtained from each sample (in relative fluorescence units [RFU]) was recorded at 0.02°C/s increases within a melting region of 65°C to 95°C. HRM data were acquired using Bio-Rad precision melt analysis, and each sample was run in duplicate. All the samples were run in the same PCR plate. Melting curves were normalized to the same fluorescence level (RFU) using the pre- and postmelt regions (before and after the melting region, respectively), which were selected based on the specific melt region of the *mcyJ* amplicon (75°C to 82°C; melting temperature [T_m], 79.5°C). These RFU values were used for the statistical analysis.

Multivariate CART for functional analysis. The melting curves (RFU) were first represented in a functional basis using a nonperiodic β-splin basis of order 4 (this was the optimal choice after several experiments). The multivariate CART (77, 81) was used to model these coefficients (output variables) using the following explanatory variables as input: total nutrients (total nitrogen [TN] and total phosphorus [TP]), wind intensity (WI), water temperature (WT), turbidity (Turb), and conductivity (K). Cross-validation was used to optimize the size (number of leaves) of the final tree. To assess the performance (prediction error) and the reliability of such an optimal tree, we repeated the following steps 100 times: (i) random split of the data in two parts, learning and test sample (in proportion 2/3 and 1/3, respectively), (ii) building of an optimal tree over the learning sample, (iii) random permutation of the observed input variable and building of another optimal tree over the permuted data set, and (iv) computing prediction error of optimal and permuted trees separately. If both trees (optimal and permuted) are similar, it indicates that the optimal tree was generated randomly and no classification was achieved. In order to test that, the error distributions obtained from both trees (100 errors calculated per tree) were compared using a log-likelihood ratio test (LRT). Once an optimal tree was obtained, a test was applied to detect if the counts of *mcyJ* gene differed between the groups of samples given by the tree using LRT and Tukey's pairwise *post hoc* comparisons. All statistical analyses were performed with the free software R, version 3.6.1 using the fda, rpart, nlme, and PMCMR packages (95–99).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.02 MB.

SUPPLEMENTAL FILE 2, PDF file, 0.8 MB.

ACKNOWLEDGMENTS

We thank Asociación Honoraria de Salvamentos Marítimos y Fluviales (ADES) and Comisión Técnico-Mixta de Salto Grande (CTM-Salto Grande) for their valuable help with performing samplings. This study was supported by grant ANII-ALGAS (Laboratorio Tecnológico del Uruguay), by PEDECIBA-Biología, and by grant FCE_1_2019_1_156308 of the Agencia Nacional de Investigación e Innovación of Uruguay (ANII) and Proyecto ECOS (Aprendizaje Estadístico para la Modelización y Análisis de Recursos Naturales).

(This work was carried out by G. Martínez de la Escalera in partial fulfillment of the requirements for a Ph.D. from PEDECIBA [University of Uruguay, Montevideo, Uruguay, 2021].)

REFERENCES

1. Sriswasdi S, Yang C, Iwasaki W. 2017. Generalist species drive microbial dispersion and evolution. Nat Commun 8:1162. https://doi.org/10.1038/s41467-017-01265-1.
2. Sultan SE. 2015. Organism and environment: ecological development, niche construction, and adaption. Oxford University Press, New York, NY.
3. Cohan FM. 2017. Transmission in the origins of bacterial diversity, from ecotypes to phyla. Microbiol Spectr 5:5.5.13. https://doi.org/10.1128/microbiolspec.MTBP-0014-2016.
4. Martiny AC, Tai AP, Veneziano D, Primeau F, Chisholm SW. 2009. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. Environ Microbiol 11:823–832. https://doi.org/10.1111/j.1462-2920.2008.01803.x.
5. Chandler JW, Lin Y, Gainer PJ, Post AF, Johnson ZI, Zinser ER. 2016. Variable but persistent coexistence of *Prochlorococcus* ecotypes along temperature gradients in the ocean's surface mixed layer. Environ Microbiol Rep 8:272–284. https://doi.org/10.1111/1758-2229.12378.
6. Moore LR, Goericke R, Chisholm SW. 1995. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. Mar Ecol Prog Ser 116:259–275. https://doi.org/10.3354/meps116259.
7. Biller SJ, Berube PM, Lindell D, Chisholm SW. 2015. *Prochlorococcus*: the structure and function of collective diversity. Nat Rev Microbiol 13:13–27. https://doi.org/10.1038/nrmicro3378.
8. Melendrez MC, Becraft ED, Wood JM, Olsen MT, Bryant DA, Heidelberg JF, Rusch DB, Cohan FM, Ward DM. 2015. Recombination does not hinder formation or detection of ecological species of *Synechococcus* inhabiting a hot spring cyanobacterial mat. Front Microbiol 6:1540. https://doi.org/10.3389/fmicb.2015.01540.
9. Ward DM, Ferris MJ, Nold SC, Bateson MM, Kopczynski ED, Ruff-Roberts AL. 1994. Species diversity in hot spring microbial mats as revealed by both molecular and enrichment culture approaches: relationship between biodiversity and community structure, p 33–44. *In* Stal LJ, Caumette P (ed), Microbial mats. Springer, Berlin, Germany.
10. Cohan FM. 2002. What are bacterial species? Annu Rev Microbiol 56:457–487. https://doi.org/10.1146/annurev.micro.56.012302.160634.
11. Cohan FM. 2006. Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. Philos Trans R Soc Lond B Biol Sci 361:1985–1996. https://doi.org/10.1098/rstb.2006.1918.

12. Cohan FM, Perry EB. 2007. A systematics for discovering the fundamental units of bacterial diversity. Curr Biol 17:R373–R386. https://doi.org/10.1016/j.cub.2007.03.032.
13. Otsuka S, Suda S, Li R, Matsumoto S, Watanabe MM. 2000. Morphological variability of colonies of *Microcystis* morphospecies in culture. J Gen Appl Microbiol 46:39–50. https://doi.org/10.2323/jgam.46.39.
14. Komárek J, Komárková J. 2002. Review of the European *Microcystis* morphospecies (Cyanoprokaryotes) from nature. Fottea 2:1–24.
15. Rigonato J, Sant'Anna CL, Giani A, Azevedo MTP, Gama WA, Viana VF, Fiore MF, Werner VR. 2018. *Sphaerocavum*: a coccoid morphogenus identical to *Microcystis* in terms of 16S rDNA and ITS sequence phylogenies. Hydrobiologia 811:35–48. https://doi.org/10.1007/s10750-017-3312-2.
16. De Leon L, Yunes JS. 2001. First report of a microcystin-containing bloom of the cyanobacterium *Microcystis aeruginosa* in the La Plata River, South America. Environ Toxicol 16:110–112. https://doi.org/10.1002/1522-7278(2001)16:1<110::AID-TOX1012>3.0.CO;2-Z.
17. González-Piana M, Fabián D, Delbene L, Chalar G. 2011. Toxics blooms of *Microcystis aeruginosa* in three Rio Negro reservoirs, Uruguay. Harmful Algae News 43:16–17.
18. O'Neil J, Davis T, Burford M, Gobler C. 2012. The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. Harmful Algae 14:313–334. https://doi.org/10.1016/j.hal.2011.10.027.
19. Srivastava A, Singh S, Ahn C-Y, Oh H-M, Asthana RK. 2013. Monitoring approaches for a toxic cyanobacterial bloom. Environ Sci Technol 47:8999–9013. https://doi.org/10.1021/es401245k.
20. Kruk C, Martínez A, Martínez de la Escalera G, Trinchin R, Manta G, Segura AM, Piccini C, Brena B, Yannicelli B, Fabiano G, Calliari D. 2021. Rapid freshwater discharge on the coastal ocean as a mean of long distance spreading of an unprecedented toxic cyanobacteria bloom. Sci Total Environ 754:142362. https://doi.org/10.1016/j.scitotenv.2020.142362.
21. Azevedo SM, Carmichael WW, Jochimsen EM, Rinehart KL, Lau S, Shaw GR, Eaglesham GK. 2002. Human intoxication by microcystins during renal dialysis treatment in Caruaru: Brazil. Toxicology 181–182:441–446. https://doi.org/10.1016/S0300-483X(02)00491-2.
22. Dittmann E, Wiegand C. 2006. Cyanobacterial toxins: occurrence, biosynthesis and impact on human affairs. Mol Nutr Food Res 50:7–17. https://doi.org/10.1002/mnfr.200500162.
23. Milutinović A, Živin M, Zorc-Pleskovič R, Sedmak B, Šuput D. 2003. Nephrotoxic effects of chronic administration of microcystins -LR and -YR. Toxicon 42:281–288. https://doi.org/10.1016/s0041-0101(03)00143-0.
24. Vidal F, Sedan D, D'Agostino D, Cavalieri ML, Mullen E, Parot Varela MM, Flores C, Caixach J, Andrinolo D. 2017. Recreational exposure during algal bloom in Carrasco Beach, Uruguay: a liver failure case report. Toxins 9:267. https://doi.org/10.3390/toxins9090267.
25. Massey IY, Yang F, Ding Z, Yang S, Guo J, Tezi C, Al-Osman M, Kamegni RB, Zeng W. 2018. Exposure routes and health effects of microcystins on animals and humans: a mini-review. Toxicon 151:156–162. https://doi.org/10.1016/j.toxicon.2018.07.010.
26. Onyango DM, Orina PS, Ramkat RC, Kowenje C, Githukia CM, Lusweti D, Lung'ayia HBO. 2020. Review of current state of knowledge of microcystin and its impacts on fish in Lake Victoria. Lakes Reserv 25:350–361. https://doi.org/10.1111/lre.12328.
27. Tillett D, Dittmann E, Erhard M, von Döhren H, Börner T, Neilan BA. 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. Chem Biol 7:753–764. https://doi.org/10.1016/s1074-5521(00)00021-1.
28. Rinta-Kanto JM, Wilhelm SW. 2006. Diversity of microcystin-producing cyanobacteria in spatially isolated regions of Lake Erie. Appl Environ Microbiol 72:5083–5085. https://doi.org/10.1128/AEM.00312-06.
29. Kim S-G, Joung S-H, Ahn C-Y, Ko S-R, Boo SM, Oh H-M. 2010. Annual variation of *Microcystis* genotypes and their potential toxicity in water and sediment from a eutrophic reservoir. FEMS Microbiol Ecol 74:93–102. https://doi.org/10.1111/j.1574-6941.2010.00947.x.
30. Hu C, Rea C, Yu Z, Lee J. 2016. Relative importance of *Microcystis* abundance and diversity in determining microcystin dynamics in Lake Erie coastal wetland and downstream beach water. J Appl Microbiol 120:138–151. https://doi.org/10.1111/jam.12983.
31. Mikalsen B, Boison G, Skulberg OM, Fastner J, Davies W, Gabrielsen TM, Rudi K, Jakobsen KS. 2003. Natural variation in the microcystin synthetase operon *mcyABC* and impact on microcystin production in *Microcystis* strains. J Bacteriol 185:2774–2785. https://doi.org/10.1128/JB.185.9.2774-2785.2003.
32. Tanabe Y, Kaya K, Watanabe MM. 2004. Evidence for recombination in the microcystin synthetase (*mcy*) genes of toxic cyanobacteria *Microcystis* spp. J Mol Evol 58:633–641. https://doi.org/10.1007/s00239-004-2583-1.
33. Tanabe Y, Sano T, Kasai F, Watanabe MM. 2009. Recombination, cryptic clades and neutral molecular divergence of the microcystin synthetase (*mcy*) genes of toxic cyanobacterium *Microcystis aeruginosa*. BMC Evol Biol 9:115. https://doi.org/10.1186/1471-2148-9-115.
34. Otsuka S, Suda S, Li R, Watanabe M, Oyaizu H, Matsumoto S, Watanabe MM. 1998. 16S rDNA sequences and phylogenetic analyses of *Microcystis* strains with and without phycoerythrin. FEMS Microbiol Lett 164:119–124. https://doi.org/10.1111/j.1574-6968.1998.tb13076.x.
35. Dittmann E, Neilan BA, Erhard M, Von Döhren H, Börner T. 1997. Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. Mol Microbiol 26:779–787. https://doi.org/10.1046/j.1365-2958.1997.6131982.x.
36. Pérez-Carrascal OM, Terrat Y, Giani A, Fortin N, Greer CW, Tromas N, Shapiro BJ. 2019. Coherence of *Microcystis* species revealed through population genomics. ISME J 13:2887–2900. https://doi.org/10.1038/s41396-019-0481-1.
37. Martínez de la Escalera G, Kruk C, Segura AM, Nogueira L, Alcántara I, Piccini C. 2017. Dynamics of toxic genotypes of *Microcystis aeruginosa* complex (MAC) through a wide freshwater to marine environmental gradient. Harmful Algae 62:73–83. https://doi.org/10.1016/j.hal.2016.11.012.
38. Ferris MJ, Kühl M, Wieland A, Ward DM. 2003. Cyanobacterial ecotypes in different optical microenvironments of a 68 degrees C hot spring mat community revealed by 16S-23S rRNA internal transcribed spacer region variation. Appl Environ Microbiol 69:2893–2898. https://doi.org/10.1128/AEM.69.5.2893-2898.2003.
39. Hjelmsø MH, Hansen LH, Baelum J, Feld L, Holben WE, Jacobsen CS. 2014. High-resolution melt analysis for rapid comparison of bacterial community compositions. Appl Environ Microbiol 80:3568–3575. https://doi.org/10.1128/AEM.03923-13.
40. Kim J, Lee C. 2014. Rapid fingerprinting of methanogenic communities by high-resolution melting analysis. Bioresour Technol 174:321–327. https://doi.org/10.1016/j.biortech.2014.10.037.
41. Li B-S, Wang X-Y, Ma F-L, Jiang B, Song X-X, Xu A-G. 2011. Is high resolution melting analysis (HRMA) accurate for detection of human disease-associated mutations? A meta analysis. PLoS One 6:e28078. https://doi.org/10.1371/journal.pone.0028078.
42. Hofinger BJ, Jing H-C, Hammond-Kosack KE, Kanyuka K. 2009. High-resolution melting analysis of cDNA-derived PCR amplicons for rapid and cost-effective identification of novel alleles in barley. Theor Appl Genet 119:851–865. https://doi.org/10.1007/s00122-009-1094-2.
43. Thomsen N, Ali RG, Ahmed JN, Arkell RM. 2012. High resolution melt analysis (HRMA): a viable alternative to agarose gel electrophoresis for mouse genotyping. PLoS One 7:e45252. https://doi.org/10.1371/journal.pone.0045252.
44. Tong SYC, Giffard PM. 2012. Microbiological applications of high-resolution melting analysis. J Clin Microbiol 50:3418–3421. https://doi.org/10.1128/JCM.01709-12.
45. Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ. 2003. High-resolution genotyping by amplicon melting analysis using LCGreen. Clin Chem 49:853–860. https://doi.org/10.1373/49.6.853.
46. Zeyoudi M, Altenaiji E, Ozer LY, Ahmed I, Yousef AF, Hasan SW. 2015. Impact of continuous and intermittent supply of electric field on the function and microbial community of wastewater treatment electro-bioreactors. Electrochimica Acta 181:271–279. https://doi.org/10.1016/j.electacta.2015.04.095.
47. Libbrecht MW, Noble WS. 2015. Machine learning applications in genetics and genomics. Nat Rev Genet 16:321–332. https://doi.org/10.1038/nrg3920.
48. Konstantinidis KT, Tiedje JM. 2005. Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258–6264. https://doi.org/10.1128/JB.187.18.6258-6264.2005.
49. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106:19126–19131. https://doi.org/10.1073/pnas.0906412106.
50. Shapiro BJ, Polz MF. 2014. Ordering microbial diversity into ecologically and genetically cohesive units. Trends Microbiol 22:235–247. https://doi.org/10.1016/j.tim.2014.02.006.
51. Kato T, Watanabe MF, Watanabe M. 1991. Allozyme divergence in *Microcystis* (*Cyanophyceae*) and its taxonomic inference. Arch Hydrobiol Suppl:129–140.

52. Otsuka S, Suda S, Shibata S, Oyaizu H, Matsumoto S, Watanabe MM. 2001. A proposal for the unification of five species of the cyanobacterial genus Microcystis Kützing ex Lemmermann 1907 under the rules of the Bacteriological Code. Int J Syst Evol Microbiol 51:873–879. https://doi.org/10.1099/00207713-51-3-873.
53. Van de Waal DB, Verspagen JMH, Finke JF, Vournazou V, Immers AK, Kardinaal WEA, Tonk L, Becker S, Van Donk E, Visser PM, Huisman J. 2011. Reversal in competitive dominance of a toxic versus non-toxic cyanobacterium in response to rising CO_2. ISME J 5:1438–1450. https://doi.org/10.1038/ismej.2011.28.
54. Liu J, Van Oosterhout E, Faassen EJ, Lürling M, Helmsing NR, Van de Waal DB. 2016. Elevated pCO2 causes a shift towards more toxic microcystin variants in nitrogen-limited *Microcystis aeruginosa*. FEMS Microbiol Ecol 92:fiv159. https://doi.org/10.1093/femsec/fiv159.
55. Zilliges Y, Kehr J-C, Meissner S, Ishida K, Mikkat S, Hagemann M, Kaplan A, Börner T, Dittmann E. 2011. The cyanobacterial hepatotoxin microcystin binds to proteins and increases the fitness of *Microcystis* under oxidative stress conditions. PLoS One 6:e17615. https://doi.org/10.1371/journal.pone.0017615.
56. Gan N, Xiao Y, Zhu L, Wu Z, Liu J, Hu C, Song L. 2012. The role of microcystins in maintaining colonies of bloom-forming *Microcystis* spp. Environ Microbiol 14:730–742. https://doi.org/10.1111/j.1462-2920.2011.02624.x.
57. Schuurmans JM, Brinkmann BW, Makower AK, Dittmann E, Huisman J, Matthijs HC. 2018. Microcystin interferes with defense against high oxidative stress in harmful cyanobacteria. Harmful Algae 78:47–55. https://doi.org/10.1016/j.hal.2018.07.008.
58. Vossen RH, Aten E, Roos A, den Dunnen JT. 2009. High-resolution melting analysis (HRMA): more than just sequence variant screening. Hum Mutat 30:860–866. https://doi.org/10.1002/humu.21019.
59. Moore LR, Rocap G, Chisholm SW. 1998. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. Nature 393:464–467. https://doi.org/10.1038/30965.
60. Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, Roggensack S, Berube PM, Henn MR, Chisholm SW. 2013. Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. ISME J 7:184–198. https://doi.org/10.1038/ismej.2012.89.
61. Dokulil MT, Teubner K. 2000. Cyanobacterial dominance in lakes. Hydrobiologia 438:1–12. https://doi.org/10.1023/A:1004155810302.
62. Piccini C, Aubriot L, Fabre A, Amaral V, González-Piana M, Giani A, Figueredo CC, Vidal L, Kruk C, Bonilla S. 2011. Genetic and eco-physiological differences of South American *Cylindrospermopsis raciborskii* isolates support the hypothesis of multiple ecotypes. Harmful Algae 10:644–653. https://doi.org/10.1016/j.hal.2011.04.016.
63. Becraft ED, Wood JM, Rusch DB, Kühl M, Jensen SI, Bryant DA, Roberts DW, Cohan FM, Ward DM. 2015. The molecular dimension of microbial species. 1. Ecological distinctions among, and homogeneity within, putative ecotypes of *Synechococcus* inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. Front Microbiol 6:590. https://doi.org/10.3389/fmicb.2015.00590.
64. Kruk C, Segura AM, Nogueira L, Alcántara I, Calliari D, Martínez de la Escalera G, Carballo C, Cabrera C, Sarthou F, Scavone P, Piccini C. 2017. A multilevel trait-based approach to the ecological performance of *Microcystis aeruginosa* complex from headwaters to the ocean. Harmful Algae 70:23–36. https://doi.org/10.1016/j.hal.2017.10.004.
65. Orr PT, Jones GJ, Douglas GB. 2004. Response of cultured *Microcystis aeruginosa* from the Swan River, Australia, to elevated salt concentration and consequences for bloom and toxin management in estuaries. Mar Freshw Res 55:277–283. https://doi.org/10.1071/MF03164.
66. Sabart M, Pobel D, Latour D, Robin J, Salençon M, Humbert J. 2009. Spatiotemporal changes in the genetic diversity in French bloom-forming populations of the toxic cyanobacterium, *Microcystis aeruginosa*. Environ Microbiol Rep 1:263–272. https://doi.org/10.1111/j.1758-2229.2009.00042.x.
67. Zhang T, Gong H, Wen X, Lu C. 2010. Salt stress induces a decrease in excitation energy transfer from phycobilisomes to photosystem II but an increase to photosystem I in the cyanobacterium *Spirulina platensis*. J Plant Physiol 167:951–958. https://doi.org/10.1016/j.jplph.2009.12.020.
68. Chen L, Mao F, Kirumba GC, Jiang C, Manefield M, He Y. 2015. Changes in metabolites, antioxidant system, and gene expression in *Microcystis aeruginosa* under sodium chloride stress. Ecotoxicol Environ Saf 122:126–135. https://doi.org/10.1016/j.ecoenv.2015.07.011.
69. Segura AM, Piccini C, Nogueira L, Alcántara I, Calliari D, Kruk C. 2017. Increased sampled volume improves *Microcystis aeruginosa* complex (MAC) colonies detection and prediction using random Forests. Ecological Indicators 79:347–354. https://doi.org/10.1016/j.ecolind.2017.04.047.
70. Tanabe Y, Hodoki Y, Sano T, Tada K, Watanabe MM. 2018. Adaptation of the freshwater bloom-forming cyanobacterium *Microcystis aeruginosa* to brackish water is driven by recent horizontal transfer of sucrose genes. Front Microbiol 9:1150. https://doi.org/10.3389/fmicb.2018.01150.
71. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints 4:e1900v1. https://doi.org/10.7287/peerj.preprints.1900v1.
72. Berry MA, White JD, Davis TW, Jain S, Johengen TH, Dick GJ, Sarnelle O, Denef VJ. 2017. Are oligotypes meaningful ecological and phylogenetic units? A case study of *Microcystis* in freshwater lakes. Front Microbiol 8:365. https://doi.org/10.3389/fmicb.2017.00365.
73. Wang X, Sun M, Wang J, Yang L, Luo L, Li P, Kong F. 2012. *Microcystis* genotype succession and related environmental factors in Lake Taihu during cyanobacterial blooms. Microb Ecol 64:986–999. https://doi.org/10.1007/s00248-012-0083-1.
74. Tromas N, Taranu ZE, Martin BD, Willis A, Fortin N, Greer CW, Shapiro BJ. 2018. Niche separation increases with genetic distance among bloom-forming cyanobacteria. Front Microbiol 9:438. https://doi.org/10.3389/fmicb.2018.00438.
75. Van Gremberghe I, Leliaert F, Mergeay J, Vanormelingen P, Van der Gucht K, Debeer A-E, Lacerot G, De Meester L, Vyverman W. 2011. Lack of phylogeographic structure in the freshwater cyanobacterium *Microcystis aeruginosa* suggests global dispersal. PLoS One 6:e19561. https://doi.org/10.1371/journal.pone.0019561.
76. Acha EM, Mianzan H, Guerrero R, Carreto J, Giberto D, Montoya N, Carignan M. 2008. An overview of physical and ecological processes in the Rio de la Plata Estuary. Continental Shelf Res 28:1579–1588. https://doi.org/10.1016/j.csr.2007.01.031.
77. Ferrari G, Pérez M de C, Dabezies M, Míguez D, Saizar C. 2011. Planktic cyanobacteria in the lower Uruguay river, South America. Fottea 11:225–234. https://doi.org/10.5507/fot.2011.021.
78. García-Rodríguez F, Brugnoli E, Muniz P, Venturini N, Burone L, Hutton M, Rodríguez M, Pita A, Kandratavicius N, Pérez L, Verocai J. 2014. Warm-phase ENSO events modulate the continental freshwater input and the trophic state of sediments in a large South American estuary. Mar Freshw Res 65:1–11. https://doi.org/10.1071/MF13077.
79. Breiman L, Friedman J, Stone CJ, Olshen RA. 1984. Classification and regression trees. CRC Press, Boca Raton, FL.
80. Nerini D, Ghattas B. 2007. Classifying densities using functional regression trees: applications in oceanology. Comput Stat Data Anal 51:4984–4993. https://doi.org/10.1016/j.csda.2006.09.028.
81. De'ath G, Fabricius KE. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192. https://doi.org/10.1890/0012–9658(2000)081[3178:CARTAP]2.0.CO;2.
82. Cutler DR, Edwards TC, Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007. Random forests for classification in ecology. Ecology 88:2783–2792. https://doi.org/10.1890/07-0539.1.
83. Breiman L. 2001. Random Forests. Machine Learning 45:5–32. https://doi.org/10.1023/A:1010933404324.
84. De'ath G. 2002. Multivariate regression trees: a new technique for modelling species-environment relationship. Ecology 83:1105–1117. https://doi.org/10.1890/0012-9658(2002)083[1105:MRTANT]2.0.CO;2.
85. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
86. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.
87. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75.
88. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.

89. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281.
90. Kosiol C, Goldman N. 2005. Different versions of the Dayhoff rate matrix. Mol Biol Evol 22:193–199. https://doi.org/10.1093/molbev/msi005.
91. Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39:306–314. https://doi.org/10.1007/BF00160154.
92. Smith MR. 2020. Information theoretic generalized Robinson-Foulds metrics for comparing phylogenetic trees. Bioinformatics 36:5007–5013. https://doi.org/10.1093/bioinformatics/btaa614.
93. Robinson O, Dylus D, Dessimoz C. 2016. Phylo.io: interactive viewing and comparison of large phylogenetic trees on the web. Mol Biol Evol 33:2163–2166. https://doi.org/10.1093/molbev/msw080.
94. Martínez de la Escalera G, Antoniades D, Bonilla S, Piccini C. 2014. Application of ancient DNA to the reconstruction of past microbial assemblages and for the detection of toxic cyanobacteria in subtropical freshwater ecosystems. Mol Ecol 23:5791–5802. https://doi.org/10.1111/mec.12979.
95. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2016. nlme: linear and nonlinear mixed effects models. R package version 3:111. http://CRAN.R-project.org/package=nlme.
96. Pohlert T. 2014. The pairwise multiple comparison of mean ranks package (PMCMR). R package 27:9. http://CRAN.R-project.org/package=PMCMR.
97. R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
98. Ramsay JO, Wickham H, Graves S, Hooker G. 2014. fda: functional data analysis. R package version 2:142.
99. Therneau T, Atkinson B, Ripley B, Ripley MB. 2015. Package 'rpart'. cranmaicacuk/web/packages/rpart/rpart.pdf. Accessed 20 April 2016.