Unraveling Evolutionary Paths: Genomic Divergence and Geographic Secrets of *Cylindrospermopsis* and *Sphaerospermopsis*

Reno S. Nooblath[1], Daniel H. Gomes[1], Sintia S. Almeida[2], Vinícius A. Abreu [1]

[1]Institute of Exact and Natural Sciences – Federal University of Pará (UFPA)
CEP 66075–110 – Belém – PA – Brazil

[2]Evandro Chagas Institute (IEC) – Federal Government
Ananindeua, PA.

{reno.bioinfo,daniel.h.gomes,sintiaalmeida}@gmail.com,vabreu@ufpa.br

Abstract. *This study aims to investigate the biological diversity of 21 genomes of Cylindrospermopsis and 6 genomes of Sphaerospermopsis belonging to the family Aphanizomenonaceae using bioinformatics methods in comparative genomics. The comparative analysis of the lineages of the groups From the Americas, Non-Americas and Sphaerospermospsis revealed conserved central genome but with different accessory genomes by geographic region. The variations observed in the organization of saxitoxin and cylindrospermopsin genes in the accessory genome of Cylindrospermopsis raciborskii from the Americas and Non-Americas, suggests in the future the formation of a new taxonomic group, due to independent evolutionary trajectories.*

1. Introduction

Environmental factors, such as eutrophication, the increase in greenhouse gases and global warming [Huisman et al. 2018, Litchman 2023], have driven the geographical expansion of *Cylindrospermopsis raciborskii* and *Sphaerospermopsis*, both belonging to the Aphanizomenonaceae family. This expansion has occurred in more northerly regions and at higher latitudes in Europe, America and Asia [Kokociński et al. 2017, Kim et al. 2020]. Some research proposes the creation of a new genus based on the morphological and molecular characteristics shared between these genera [Sant'Anna et al. 2019]. In addition, there are proposals to form a taxonomic group that includes the genera *Anabaena, Dolichospermum* and *Aphanizomenon* (*ADA*), which belong to the same family[Driscoll et al. 2018, Teikari et al. 2019, Dreher et al. 2021].

Cylindrospermopsis raciborskii, initially documented in Java in 1912, is known to produce toxins such as cylindrospermopsin and saxitoxin, which have been linked to adverse effects on human and environmental health [Hoff-Risseti et al. 2013, Fuentes-Valdés et al. 2018]. *Sphaerospermopsis aphanizomenoides*, as identified by Horecka and Komarek [Zapomělová et al. 2009], has been linked to microcystin and anatoxin [Li et al. 2017], both of which have been demonstrated to have significant toxic effects.

In light of the recent identification of new cyanobacterial lineages emerging from various locations worldwide [Antunes et al. 2015, Sidelev et al. 2020, Laux et al. 2023],

it is imperative to conduct comprehensive studies and taxonomic revisions. This is essential for a more accurate understanding of the phylogenetic relationships among these cyanobacteria and for monitoring changes in genetic diversity as new lineages and species are identified.

To understand the detrimental effects of the proliferation of the cyanobacteria *Cylindrospermopsis* and *Sphaerospermopsis*, it is important to advance knowledge of their diversity, genetic variation, and evolutionary processes within populations. An effective approach to explore this diversity and the evolution of these species is through bioinformatics methods in comparative genomics. We compare the available genomes of these species based on up to 200 scaffolds to achieve greater reliability in the results of genomic analyses [Klassen and Currie 2012, Setubal et al. 2018].

The specific objective of this study was to analyze genetic and ecological variations among *Cylindrospermopsis raciborskii* and *Sphaerospermopsis*, and to investigate their evolutionary relationships.

2. Materials and Methods

2.1. Cyanobacteria strains

In this study, 27 cyanobacterial genomes from the GenBank database of the National Center for Biotechnology Information (NCBI) were used, with up to 200 scaffolds each (Table S1). This included 21 genomes from *Cylindrospermopsis* lineages and 6 genomes from *Sphaerospermopsis* lineages within the family Aphanizomenonaceae. The KLL 07 lineage of *Cylindrospermopsis raciborskii* was selected as the reference lineage for both the *Cylindrospermopsis* and *Sphaerospermopsis* genera due to its closed genome. Our results showed that, when compared with the 16S rRNA gene of the same genus in the SILVA rRNA database project[Quast et al. 2013], this reference genome provided a better percentage of identity for the sequence.

2.2. Phylogenetic analysis and Phylogenomic

The phylogenetic tree of the 16S rRNA gene was inferred using a Bayesian algorithm implemented with MrBayes 3.2.2 [Nielsen 2005], employing the GTR evolutionary model with 42 cyanobacterial nucleotide sequences over 5.100.000 generations [Darriba et al. 2012]. The phylogenomic tree was constructed using 31 conserved protein sequences [Wu and Eisen 2008], and was compared using the BlastP tool, which was employed to identify homologous proteins across 33 cyanobacterial genomes.

The alignment of the 31 conserved protein sequences was performed using Mega 11 [Tamura et al. 2021] and aligned with the Muscle algorithm [Edgar 2004]. Phylogenetic inference was conducted using Maximum Likelihood (ML) and Neighbor-Joining (NJ) methods [Saitou and Nei 1987], applying the PROTGAM-MAGTR amino acid substitution model and a bootstrap value of 1000 replicates. The 16S rRNA sequences were standardized to up to 1384 base pairs when compared with the reference genome sequence. Sequences of 16S rRNA that did not meet this standard were excluded.

2.3. Annotation of genes and genomes

Genome annotation was performed using the Rapid Annotation using Subsystems Technology (RAST) server [Aziz et al. 2008], employing the RASTtk annotation pipeline.

Gene clusters associated with secondary metabolites were predicted using antiSMASH [Medema et al. 2011], with all available additional features provided by the tool selected.

2.4. Synteny analysis genes and genomes

The synteny analysis of gene and metabolite clusters was conducted by geographic region using the pyGenomeViz tool (https://github.com/moshi4/pyGenomeViz) with the MMseqs algorithm and an identity threshold of 50%. To investigate the conservation, genomic location, and organization of gene clusters, the Blast Atlas method available in the GView Server software [Petkau et al. 2010] was employed. The KLL 07 reference lineage was compared with lineages from Asia, Europe, and Oceania. The genome of *Raphidiopsis brookii D9*, located in the Americas and with a smaller scaffold, was selected as the reference for comparison with lineages from the Americas. Additionally, the *Sphaerospermopsis torques-reginae* ITEP-024 lineage, from Brazil and with a closed genome, was chosen as the reference for the genus *Sphaerospermopsis* and compared with the five lineages of the same genus.

2.5. Pan-genomics analysis

The core genome and pangenome of *Cylindrospermopsis* and *Sphaerospermopsis* lineages were estimated using the Best Database Hits (BDBH) [Contreras-Moreira and Vinuesa 2013], Markov Cluster (OMCL) [Li et al. 2003], and lowpolynomial algorithm (COGtriangles) [Kristensen et al. 2010] implemented in the Get Homologues tool [Contreras-Moreira and Vinuesa 2013], with an e-value threshold of 1e-05 and a coverage of 75% as default settings. This analysis was performed based on sequence similarity among all genomes obtained in GenBank format through the RAST Server [Aziz et al. 2008].

3. Results

3.1. Geographical distribution of selected cyanobacteria of the genera *Cylindrospermopsis* and *Sphaerospermopsis*.

Based on data extracted from NCBI, we identified the species of *Cylindrospermopsis*, represented graphically as spheres, and the species of *Sphaerospermopsis*, represented as squares. Both species were grouped by geographic region (Figure 1) .

3.2. Phylogenetic Analyses Based on the 16S rRNA Gene and Genomes

The phylogenomic tree (Figure 2A) and the phylogenetic tree based on the 16S rRNA gene(Figure 2B) revealed a speciation process resulting in the formation of two distinct clades. The first clade comprises the *Cylindrospermopsis* group, while the second clade consists of the *Sphaerospermopsis* group. The *Cylindrospermopsis* clade is further divided into two subclades: The Americas clade, which includes samples from South America, North America, and Central America, and the Non-Americas clade, encompassing Asia, Europe, and Oceania.

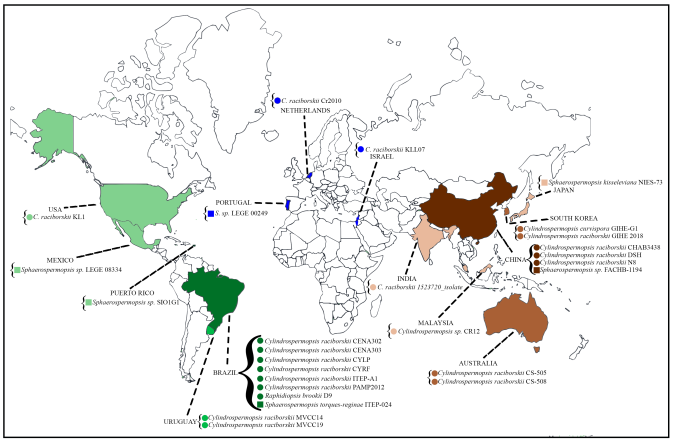

Figure 1. Geographic Distribution of Collected Lineages.

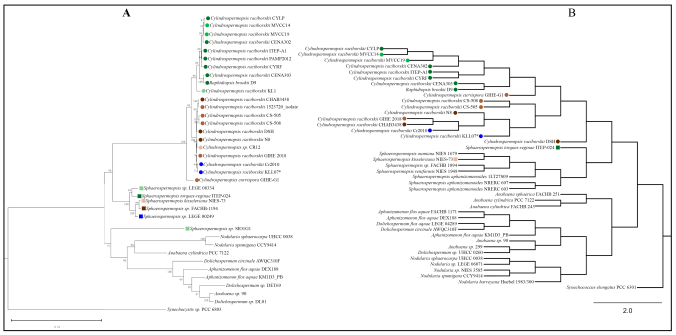

Figure 2. (A) Maximum Likelihood phylogenetic tree constructed from 31 conserved protein sequences of 37 cyanobacterial lineages.(B) Bayesian Inference phylogenetic tree derived from 16S rRNA gene nucleotide sequences of 42 cyanobacterial lineages.

3.3. Genome Annotation Analysis

The annotation results for the lineages analyzed in this study revealed significant genomic similarities even among geographically distant lineages within the same genus. Annotation of subsystems, conducted using the RAST server and SEED viewer, indicated a predominance of functions in the following categories: (I) Protein Metabolism; (II) Cofactors, Vitamins, Prosthetic Groups, and Pigments; (III) Amino Acids and Derivatives; and (IV) Carbohydrates (Figure S1 and Table S13).

3.4. Gene Clusters by Region in the Genome

Gene clusters were observed by region in the genome (Figure 3), responsible for the synthesis of non-ribosomal peptides identified as cylindrospermopsin (*cyr*), saxitoxin (*sxt*), hassalidin (*has*), anabaenopeptin (*apt*), esfaerocyclamide (*sph*), and nocuolin (*noc*). Among these clusters, the *hgl* gene cluster related to heterocyst glycolipids and the *Nih* gene cluster involved in nitrogen fixation (Figure S2) are the most prevalent.

The *sxt* gene cluster, which regulates saxitoxin production, is present in both genera found in the Americas, identified in the lineages *R. brookii D9, C. raciborskii* CENA 302, MVCC14, MVCC19, CYRF, and ITEP-A1 (Figure S3), as well as in the *Sphaerospermopsis* group in the lineages *S. torques-reginae* ITEP-024 and *S. sp. FACHB-1194*

(Figure S4). The presence of the *sxtK* gene (unknown protein) and *sxtJ* gene (regulatory protein) was observed in *S. torques-reginae* ITEP-024 and *S. sp.* LEGE 08334 (Table S2). The *sxtK* gene was also present in *C. raciborskii* CENA 302, while the *sxtJ* gene was found in *R. brookii* D9 and *C. raciborskii* MVCC14 (Table S3).

The *cyr* gene cluster, which regulates cylindrospermopsin synthesis, was identified in the lineages *C. raciborskii* CS-505, CR12, CHAB-3438, and DSH located in Oceania, Malaysia, and Asia (Figure S5 and Table S4).

The *apt* gene cluster, which regulates anabaenopeptin synthesis, was identified in the lineages *S. torques-reginae* ITEP-024 (Brazil), *S. sp.* LEGE 08334 (Mexico), *S. sp.* SIO1G1 (Puerto Rico), *S. kisseleviana* NIES-73 (Japan), and *S. sp. FACHB-1194* (China). The *apt* cluster is represented by the genes *aptA1, aptA2, aptB, aptC, aptD, aptE,* and *aptF* (Figure S6 and Table S5).

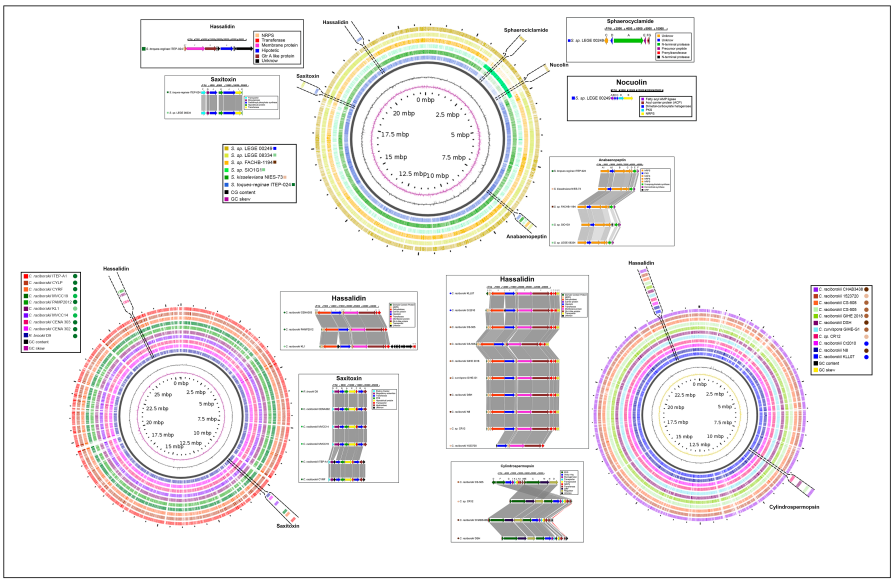

Figure 3. Blast Atlas analysis for 10 *Cylindrospermopsis raciborskii* strains from the Americas, 11 *Cylindrospermopsis raciborskii* strains Non Americas, and 6 *Sphaerospermopsis* strains. Identification of the saxitoxin (*sxt*), cylindrospermopsin (*cyr*), hassalidin (*has*), anabaenopeptin (*apt*), nocuolin (*noc*) and sphaerocyclamide (*sph*) gene clusters. These clusters are highlighted inside the boxes, and their locations in the genomes are indicated in the figure.

3.5. Comparative genomic analyses

The calculation of average nucleotide identity (ANI) (Figure S7) showed that *Cylindrospermopsis raciborskii* lineages exhibited high identity among themselves, ranging from 92.71% to 99.75%. In contrast, genetic identity among lineages within the *Sphaerospermopsis* group ranged from 84.3% to 97.16%. The pangenome analysis for *Sphaerospermopsis* estimated a total of 9.000 distinct genes, with the core genome comprising 2.375 genes (Figures S8A and S8B). For *C. raciborskii* from the Americas, the pangenome was estimated to include 6.000 distinct genes, with the core genome consisting of 1.410 genes (Figures S9A and S9B). Similarly, the pangenome of *C. raciborskii* from non-American regions was estimated to contain 5.000 distinct genes, with a core genome of 2.249 genes (Figures S10A and S10B).

4. Discussion

The analysis of biological diversity between the genera *Cylindrospermopsis* and *Sphaerospermopsis* reveals an intricate evolutionary history, shaped by multiple factors, including historical biogeographic events, environmental pressures, and genetic diversification. The phylogenetic separation of these lineages based on geographic regions suggests that their current distribution is rooted in ancient tectonic events, such as the breakup of Pangea [Stampfli et al. 2013]. Geographic isolation, combined with other evolutionary forces, contributed to the divergence of these genera into distinct clades, each adapted to specific environmental conditions.

The evolutionary history of cyanobacterial lineages, such as *C. curvispora* from South Korea and *S. torques-reginae* from Brazil, sheds light on the processes of long-distance dispersal and vicariance events. The phylogenetic relationship between *C. curvispora* and lineages from the Americas exemplifies how ancient populations may have dispersed over vast distances. This dispersal could have occurred through natural mechanisms, such as oceanic currents or atmospheric circulation, or through historical biogeographic events, such as continental drift. The presence of a sister clade relationship suggests that these lineages share a common evolutionary origin, which has diverged over time due to geographic isolation and environmental pressures.

The observed high genetic similarity (99.75% ANI) between the *S. torques-reginae* lineage from Brazil and *S. sp.LEGE 08334* from Mexico despite their geographical separation, underscores a notable evolutionary connection (Figure S7). Such genetic similarities suggest that ancient migrations and gene flow, possibly facilitated by environmental changes or human activity, have influenced their shared evolutionary history.The consistency of genetic markers across distant populations, such as those seen between *S. kisseleviana* from Japan and *S. sp. FACHB-1194* from China (ANI of 98.93%), supports the notion that historical migrations and genetic drift have played significant roles in the diversification of these cyanobacteria.

The *S. sp. LEGE 00249* lineage from Portugal acts as a common ancestor among the analyzed lineages, while *S. sp. SIO1G1* from Puerto Rico exhibits a lower genetic identity compared to other lineages within the same group. This lower similarity suggests the potential need for taxonomic reclassification, (Figure S7), indicating that current classification systems may not fully capture the evolutionary complexity of these organisms. In the phylogenomic tree (Figure 2A) , the lineage *S. sp. SIO1G1* (Puerto Rico) forms a clade with genera such as *Anabaena, Dolichospermum*, and *Aphanizomenon (ADA)*, indicating shared genetic traits among these groups. Notably, *S. sp. SIO1G1* (Puerto Rico) exhibits a higher average nucleotide identity (ANI) with *S. sp. FACHB-1194* (China) and *S. kisseleviana NIES-73* (Japan) than with the geographically closer *S. sp. LEGE 08334* (Mexico) (Figure S7). This counterintuitive pattern of genetic similarity suggests complex evolutionary dynamics, including historical gene flow and genetic drift, that may not align neatly with present-day geographic proximity.

These findings point to a potential need for re-evaluation of the current results and suggest a potential need for adjustment in taxonomic classification. To enhance future analyses of the taxonomic placement of *S. sp. SIO1G1* (Puerto Rico) within *Sphaerospermopsis*, it will be essential to incorporate additional genomic sequences from *Sphaerospermopsis* species and from the *ADA* group, particularly *Anabaena, Dolichospermum*,

and *Aphanizomenon*.

The presence of an open pangenome curve and high diversity in the accessory genome within the *Sphaerospermopsis* groups (Figures S8A and S8B), *C. raciborskii* from the Americas (Figures S9A and S9B), and non-Americas (Figures S10A and S10B) indicates that genetic diversity within these clades is substantial and not yet saturated. This extensive genetic diversity suggests a high potential for adaptation and ecological plasticity, enabling these lineages to thrive in a variety of environmental conditions. The continuous acquisition of new genes points to an ongoing evolutionary process characterized by horizontal gene transfer, recombination, and selective pressures that favor novel genetic traits.

Our synteny analysis further reveals that the gene clusters responsible for the production of secondary metabolites, such as cylindrospermopsin and saxitoxin, exhibit differences in gene organization (Figure 3). Despite their roles as important phylogenetic markers, the distinct arrangements of these gene clusters indicate either independent evolutionary origins or divergent genetic rearrangements over time. These differences may reflect adaptive responses to specific environmental pressures or historical events, such as the migration of lineages across different continents and habitats.

Comparative analyses of gene clusters among cyanobacterial lineages have revealed that certain groups, despite being geographically distant, possess conserved gene clusters responsible for the production of secondary metabolites. This finding suggests that these clusters play essential roles in the adaptive strategies of these organisms, transcending geographic and ecological boundaries. For instance, the hassalidin biosynthetic gene cluster *(has)* is responsible for producing antifungal compounds and is found in lineages from diverse regions: CENA 303, KL1, and PAMP2012 from the Americas, KLL07 from Europe, CS505 from Oceania, and N8, CR2010, GIHEG1, and GIHE2018 from Asia (Figure 3 and Table S8). Notably, within the *Sphaerospermopsis* group, the Brazilian lineage *S. torques-reginae* ITEP024 is the sole representative with the *has* gene cluster. This unique occurrence within *Sphaerospermopsis* suggests a distinctive evolutionary event, possibly facilitated by horizontal gene transfer from other cyanobacterial groups or through retention from a common ancestor that has since lost this capability in other lineages. The genes within this cluster *(hasL, hasM, hasN, hasO, hasQ, hasT, hasU, hasV, hasX, and hasY)* are not only essential for hassalidin synthesis but also highlight the cluster's versatility. Among these, *hasL* stands out due to its potential role in antibiotic production, underscoring the biotechnological value of these secondary metabolites (Figure 3 and Table S8).

Some studies indicate that saxitoxin, a potent neurotoxin, is predominantly in *Cylindrospermopsis* lineages located in South America, highlighting a potential biogeographic specificity in toxin production [Antunes et al. 2015]. However, the exclusivity of saxitoxin production by *Sphaerospermopsis* cyanobacteria in this region remains uncertain. To validate this hypothesis, further studies involving additional members of the *Sphaerospermopsis* genus from South America are recommended, as broader sampling could provide a more comprehensive view of the distribution and genetic diversity of the *sxt* gene cluster.

The comparative analysis of the *sxt* gene cluster between *Sphaerospermopsis* and

the *C. raciborskii* group from the Americas reveals significant differences in both sequence size, which can reach up to 5000 base pairs, and gene organization, particularly in genes such as *sxtJ* and *sxtK* (Figures S3, S4) and (Tables S2, S3). These variations suggest that the *sxt* cluster has undergone considerable evolutionary divergence, reflecting localized adaptations and genetic rearrangements that may influence saxitoxin production. In contrast, the lineages from both organism groups in Asia, Europe, and Oceania do not show the presence of *sxt* gene clusters, suggesting that saxitoxin production is not a universal trait among these cyanobacteria. However, some lineages, such as *C. raciborskii CHAB3438* (China), *C. raciborskii DSH* (China), *C. raciborskii CR12* (Malaysia), and *C. raciborskii CS505* (Australia), have been identified as producers of cylindrospermopsin [Scarlett et al. 2020]. The analysis of the *cyr* gene cluster in the CR12 lineage reveals gene rearrangements and the absence of key genes *(cyrL, cyrM, cyrN, and cyrO)* compared to the CS505 lineage, suggesting independent evolutionary trajectories and functional diversification in toxin production [Abreu et al. 2018].

One evolutionary hypothesis posits that the ability to produce cylindrospermopsin originated in ancestral populations that migrated to Sub-Saharan Africa, subsequently expanding to Oceania and South America [Vico et al. 2020]. During these migrations, South American populations may have lost the *cyr* cluster and acquired the *sxt* cluster via horizontal gene transfer, highlighting the fluid and dynamic nature of microbial genomes. This genetic exchange likely conferred an adaptive advantage, facilitating the survival and spread of these cyanobacteria in new environments. These *sxt* producing lineages are thought to have later migrated to North America, further dispersing this evolutionary innovation [Padisák 1997].

Additionally, the *S. sp. LEGE 00249* lineage from Portugal possesses unique biosynthetic gene clusters, such as the *sph* cluster, which exhibits antimicrobial activity against the bacterium Halomonas aquamarina CECT 5000 [Martins et al. 2018], and the *noc* cluster, responsible for the synthesis of nocuolin, a compound with potential anticancer properties, [Voráčová et al. 2017]. These examples underscore the diverse functional capabilities of cyanobacteria and their potential applications in biotechnology and medicine.

The gene clusters *(nih)*, responsible for nitrogen fixation, and *(hgl)*, involved in heterocyst glycolipid production, are the most conserved among the groups (Figure S2). The Brazilian lineages *C. raciborskii* CENA 303 and *R. brookii D9* did not exhibit these gene clusters [Abreu et al. 2018]. When compared to the lineages *C. raciborskii* CENA 302 (Brazil), MVCC 14 (Uruguay), ITEPA-A1 (Brazil), CR12 (Malaysia), CS505 (Australia), and CS508 (Australia), *C. raciborskii* CENA 303 (Brazil) demonstrated lower similarity among them [Hoffmann et al. 2017]. This difference is attributed to the absence of these genes compared to the other lineages within the *Sphaerospermopsis* group (Tables S9, S10, S11 and S12).

Despite geographic distances, the *Sphaerospermopsis* group frequently exhibits the *apt* gene cluster, responsible for anabaenopeptin production, with the exception of the *S. sp. LEGE 00249* lineage from Portugal (Figure S6 and Table S5). In contrast, the absence of these gene clusters in *Cylindrospermopsis* suggests underlying genetic differences between the two genera. Anabaenopeptins, recognized for their inhibitory effects on phosphatases and proteases, may play a crucial role in the ecological adaptability and

toxicity of these cyanobacteria, enhancing their defense mechanisms against zooplankton and crustaceans, thus contributing to their survival and competitive advantage in various environments [Monteiro et al. 2021].

Although some studies have proposed the formation of a new taxonomic group based on *Cylindrospermopsis* and *Sphaerospermopsis*, our genomic analysis does not support this classification. Comparative analysis reveals that, despite phenotypic similarities, *Cylindrospermopsis* and *Sphaerospermopsis* exhibit significant evolutionary differences (Figure S13). Therefore, the data do not support the creation of a new genus, as the division into three distinct groups does not allow these lineages to be classified as a cohesive taxonomic group.

5. Conclusion

The study demonstrates that *Cylindrospermopsis* and *Sphaerospermopsis* are taxonomically distinct at the genus level, with a conserved central genome but variations in the accessory genome according to geographic region, indicating adaptation to different environmental conditions. The absence of the cylindrospermopsin gene cluster in South American strains of *Cylindrospermopsis raciborskii* may lead to future evolutionary divergence from strains on other continents.

For a more comprehensive understanding of cyanobacterial genetic diversity, it is necessary to develop a comprehensive database using mechanistic models based on traits such as toxin production, nitrogen fixation and environmental adaptations. Integrating these data into statistical models will allow detailed analysis and prediction of harmful algal blooms. Future research should include lineages from under-represented regions, such as Africa, and the use of more *Sphaerospermopsis* genomes, as well as exploring genetic and functional variation and comparison with related genera.

6. Acknowledgment

Research developed with the support of the High-Performance Computing Center of the Federal University of Pará.

References

[Abreu et al. 2018] Abreu, V. A. C., Popin, R. V., Alvarenga, D. O., Schaker, P. D. C., Hoff-Risseti, C., Varani, A. M., and Fiore, M. F. (2018). Genomic and Genotypic Characterization of Cylindrospermopsis raciborskii: Toward an Intraspecific Phylogenetic Evaluation by Comparative Genomics. *Frontiers in Microbiology*, 9. Publisher: Frontiers.

[Antunes et al. 2015] Antunes, J. T., Leão, P. N., and Vasconcelos, V. M. (2015). Cylindrospermopsis raciborskii: review of the distribution, phylogeography, and ecophysiology of a global invasive species. *Frontiers in Microbiology*, 6. Publisher: Frontiers.

[Aziz et al. 2008] Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9(1):75.

[Contreras-Moreira and Vinuesa 2013] Contreras-Moreira, B. and Vinuesa, P. (2013). GET_homologues, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. *Applied and Environmental Microbiology*, 79(24):7696–7701. Publisher: American Society for Microbiology.

[Darriba et al. 2012] Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and high-performance computing. *Nature methods*, 9(8):772.

[Dreher et al. 2021] Dreher, T. W., Davis, E. W., Mueller, R. S., and Otten, T. G. (2021). Comparative genomics of the ADA clade within the Nostocales. *Harmful Algae*, 104:102037.

[Driscoll et al. 2018] Driscoll, C. B., Meyer, K. A., Šulčius, S., Brown, N. M., Dick, G. J., Cao, H., Gasiūnas, G., Timinskas, A., Yin, Y., Landry, Z. C., Otten, T. G., Davis, T. W., Watson, S. B., and Dreher, T. W. (2018). A closely-related clade of globally distributed bloom-forming cyanobacteria within the Nostocales. *Harmful Algae*, 77:93–107.

[Edgar 2004] Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797.

[Fuentes-Valdés et al. 2018] Fuentes-Valdés, J. J., Soto-Liebe, K., Pérez-Pantoja, D., Tamames, J., Belmar, L., Pedrós-Alió, C., Garrido, D., and Vásquez, M. (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.

[Hoff-Risseti et al. 2013] Hoff-Risseti, C., Dörr, F. A., Schaker, P. D. C., Pinto, E., Werner, V. R., and Fiore, M. F. (2013). Cylindrospermopsin and saxitoxin synthetase genes in Cylindrospermopsis raciborskii strains from Brazilian freshwater. *PloS One*, 8(8):e74238.

[Hoffmann et al. 2017] Hoffmann, L., Ramos, R. J. T., Guedes, I. A., Costa, P. F., Miguel, C. R. D., Azevedo, S. M. F. d. O. e., and Silva, R. (2017). Draft Genome Sequences of Two Brazilian Cyanobacterial Strains of Cylindrospermopsis raciborskii: Differences in Membrane Transporters, Saxitoxin Production, and Antioxidant Activities. *Genome Announcements*, 5(43):10.1128/genomea.00879–17. Publisher: American Society for Microbiology.

[Huisman et al. 2018] Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H., and Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, 16(8):471–483. Publisher: Nature Publishing Group.

[Kim et al. 2020] Kim, Y.-J., Park, H.-K., and Kim, I.-S. (2020). Invasion and toxin production by exotic nostocalean cyanobacteria (*Cuspidothrix, Cylindrospermopsis*, and *Sphaerospermopsis*) in the Nakdong River, Korea. *Harmful Algae*, 100:101954.

[Klassen and Currie 2012] Klassen, J. L. and Currie, C. R. (2012). Gene fragmentation in bacterial draft genomes: extent, consequences and mitigation. *BMC Genomics*, 13(1):14.

[Kokociński et al. 2017] Kokociński, M., Gągała, I., Jasser, I., Karosienė, J., Kasperovičienė, J., Kobos, J., Koreivienė, J., Soininen, J., Szczurowska, A., Woszczyk, M., and Mankiewicz-Boczek, J. (2017). Distribution of invasive Cylindrospermopsis raciborskii in the East-Central Europe is driven by climatic and local environmental variables. *FEMS Microbiology Ecology*, 93(4):fix035.

[Kristensen et al. 2010] Kristensen, D. M., Kannan, L., Coleman, M. K., Wolf, Y. I., Sorokin, A., Koonin, E. V., and Mushegian, A. (2010). A low-polynomial algorithm for assembling

clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics*, 26(12):1481–1487.

[Laux et al. 2023] Laux, M., Piroupo, C. M., Setubal, J. C., and Giani, A. (2023). The *Raphidiopsis* (= *Cylindrospermopsis*) *raciborskii* pangenome updated: Two new metagenome-assembled genomes from the South American clade. *Harmful Algae*, 129:102518.

[Li et al. 2003] Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research*, 13(9):2178–2189. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[Li et al. 2017] Li, X., Yang, Y., Cai, F., Yang, P., Song, L., and Li, R. (2017). New insights into the taxonomy of the genus Sphaerospermopsis (Nostocales, Cyanobacteria) with the description of Sphaerospermopsis crassa sp. nov. *Phycologia*, 56(2):147–158. Publisher: Taylor & Francis.

[Litchman 2023] Litchman, E. (2023). Understanding and predicting harmful algal blooms in a changing climate: A trait-based framework. *Limnology and Oceanography Letters*, 8(2):229–246. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/lol2.10294.

[Martins et al. 2018] Martins, J., Leikoski, N., Wahlsten, M., Azevedo, J., Antunes, J., Jokela, J., Sivonen, K., Vasconcelos, V., Fewer, D. P., and Leão, P. N. (2018). Sphaerocyclamide, a prenylated cyanobactin from the cyanobacterium Sphaerospermopsis sp. LEGE 00249. *Scientific Reports*, 8(1):14537. Publisher: Nature Publishing Group.

[Medema et al. 2011] Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, 39(suppl_2):W339–W346.

[Monteiro et al. 2021] Monteiro, P. R., do Amaral, S. C., Siqueira, A. S., Xavier, L. P., and Santos, A. V. (2021). Anabaenopeptins: What We Know So Far. *Toxins*, 13(8):522. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

[Nielsen 2005] Nielsen, R. (2005). *Statistical Methods in Molecular Evolution*. Statistics for Biology and Health. Springer, New York, NY.

[Padisák 1997] Padisák, J. (1997). Cylindrospermopsis raciborskii (Woloszynska) Seenayya et Subba Raju, an expanding, highly adaptive cyanobacterium: worldwide distribution and review of its ecology. *ARCHIV FÜR HYDROBIOLOGIE SUPPLEMENTBAND MONOGRAPHISCHE BEITRAGE*, 107(4):563–593. Number: 4 Publisher: Schweizerbart.

[Petkau et al. 2010] Petkau, A., Stuart-Edwards, M., Stothard, P., and Van Domselaar, G. (2010). Interactive microbial genome visualization with GView. *Bioinformatics*, 26(24):3125–3126.

[Quast et al. 2013] Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1):D590–D596. Accessed on:2022-08-19.

[Saitou and Nei 1987] Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4):406–425.

[Sant'Anna et al. 2019] Sant'Anna, C. L., Gama, W. A., Rigonato, J., Correa, G., Mesquita, M. C. B., and Marinho, M. M. (2019). Phylogenetic connection among close genera of Aphanizomenonaceae (Cyanobacteria): *Amphiheterocytum* gen. nov., *Cylindrospermopsis* and *Sphaerospermopsis*. *Algal Research*, 37:205–214.

[Scarlett et al. 2020] Scarlett, K. R., Kim, S., Lovin, L. M., Chatterjee, S., Scott, J. T., and Brooks, B. W. (2020). Global scanning of cylindrospermopsin: Critical review and analysis of aquatic occurrence, bioaccumulation, toxicity and health hazards. *Science of The Total Environment*, 738:139807.

[Setubal et al. 2018] Setubal, J. C., Almeida, N. F., and Wattam, A. R. (2018). Comparative Genomics for Prokaryotes. In Setubal, J. C., Stoye, J., and Stadler, P. F., editors, *Comparative Genomics: Methods and Protocols*, pages 55–78. Springer, New York, NY.

[Sidelev et al. 2020] Sidelev, S., Koksharova, O., Babanazarova, O., Fastner, J., Chernova, E., and Gusev, E. (2020). Phylogeographic, toxicological and ecological evidence for the global distribution of *Raphidiopsis raciborskii* and its northernmost presence in Lake Nero, Central Western Russia. *Harmful Algae*, 98:101889.

[Stampfli et al. 2013] Stampfli, G. M., Hochard, C., Vérard, C., Wilhem, C., and vonRaumer, J. (2013). The formation of Pangea. *Tectonophysics*, 593:1–19.

[Tamura et al. 2021] Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, 38(7):3022–3027.

[Teikari et al. 2019] Teikari, J. E., Popin, R. V., Hou, S., Wahlsten, M., Hess, W. R., and Sivonen, K. (2019). Insight into the genome and brackish water adaptation strategies of toxic and bloom-forming Baltic Sea Dolichospermum sp. UHCC 0315. *Scientific Reports*, 9(1):4888. Publisher: Nature Publishing Group.

[Vico et al. 2020] Vico, P., Bonilla, S., Cremella, B., Aubriot, L., Iriarte, A., and Piccini, C. (2020). Biogeography of the cyanobacterium *Raphidiopsis* (*Cylindrospermopsis) raciborskii*: Integrating genomics, phylogenetic and toxicity data. *Molecular Phylogenetics and Evolution*, 148:106824.

[Voráčová et al. 2017] Voráčová, K., Hájek, J., Mareš, J., Urajová, P., Kuzma, M., Cheel, J., Villunger, A., Kapuscik, A., Bally, M., Novák, P., Kabeláč, M., Krumschnabel, G., Lukeš, M., Voloshko, L., Kopecký, J., and Hrouzek, P. (2017). The cyanobacterial metabolite nocuolin a is a natural oxadiazine that triggers apoptosis in human cancer cells. *PLOS ONE*, 12(3):e0172850. Publisher: Public Library of Science.

[Wu and Eisen 2008] Wu, M. and Eisen, J. A. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biology*, 9(10):R151.

[Zapomělová et al. 2009] Zapomělová, E., Jezberová, J., Hrouzek, P., Hisem, D., Řeháková, K., and Komárková, J. (2009). Polyphasic Characterization of Three Strains of Anabaena Reniformis and Aphanizomenon Aphanizomenoides (cyanobacteria) and Their Reclassification to Sphaerospermum Gen. Nov. (incl. Anabaena Kisseleviana). *Journal of Phycology*, 45(6):1363–1373. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1529-8817.2009.00758.x.