Cyanobacteria phylogenetic pipeline
Agnieszka Michalak

I. Introduction

Cyanobacteria are widely known microorganisms responsible for water blooms, producing neurotoxic, hepatotoxic, and cytotoxic substances (Cheung et al., 2013). Along with the toxins, they produce a wide variety of compounds such as phenolic acids, chlorophyll a, carotenoids, and phycocyanin, enabling the photosynthetic process and causing an antioxidant effect (Miranda et al., 1998).

Genus *Arthrospira* is an interesting member of cyanobacteria, unlike the majority of that group, they do not produce toxic substances. The main biotechnological application of *Arthrospira* is the production of biomass commercially known as Spirulina. It is a common confusion arising from the merging of the genus *Arthrospira* and *Spirulina* in 1932 (Geitler, L. 1932). Due to the development of molecular methods, it became clear that *Arthrospira* and *Spirulina* are very distinct organisms (Herdman et al., 1979; Guglielmi et al., 1993; Nelissen et al., 1994).

Although cyanobacteria prefer harsh, highly alkalic, halophilic environments, it is troublesome work to obtain axenic culture. Interactions of cyanobacteria and heterotrophic bacteria have been seen in a number of studies [Alex et al., 2012); Brunberg, 2006; (Worm 1998; Erwin et al., 2011; Grossart et al., 2006; SHI et al., 2009].

In the work of Leboulanger et al., 2017 the environment of Dziani Dzaha lake was investigated. Dziani Dzaha is a highly saline, alkalic lake with permanent green water and dominance of *Arthrospira*. The authors compared 16 S rRNA of the *Arthrospira* isolates with the other cyanobacteria which 16 S rRNA sequences were acquired from public databases and constructed a consensus phylogenetic tree using maximum likelihood (ML) method (Fig.1).
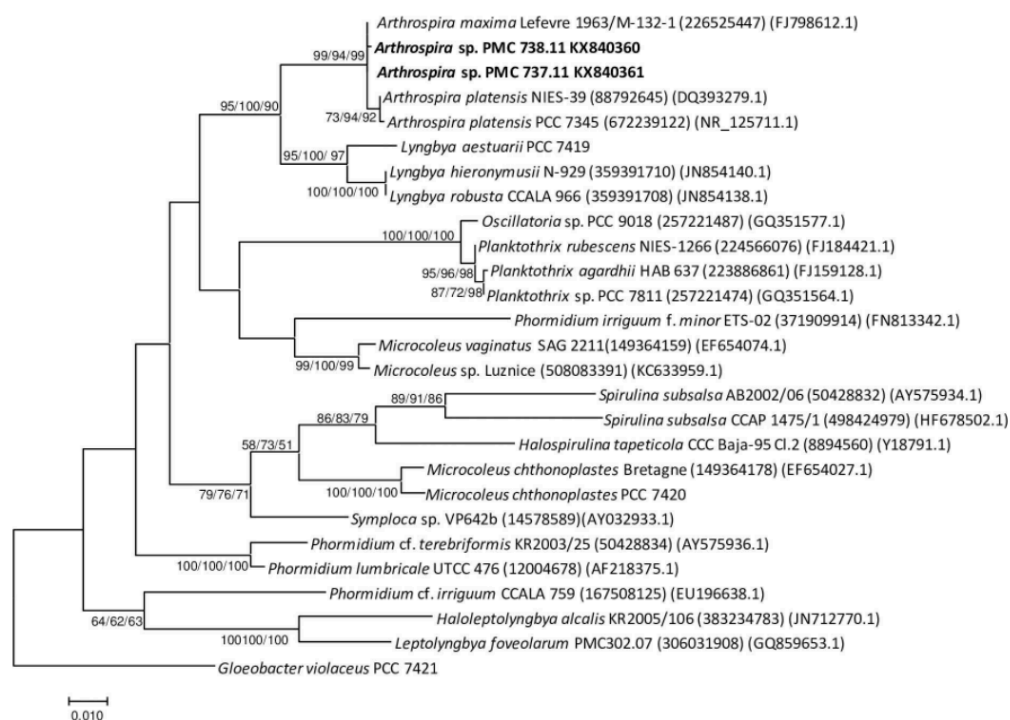

Fig. 1. Consensus phylogenetic tree (ML) constructed from 16 S rRNA, (Leboulanger et al., 2017)

II.   Methods

Acquisition of proteomes

22 protein sequences of cyanobacteria were acquired from the NCBI Genome database. The main criterion in choosing the sequences was the quality, therefore only reference sequences were selected.

Identification of protein families

Sequences were clustered with the MMseqs2 program to obtain protein families. The program was run with the default parameters. For further analyses, the clusters were filtered based on 1) the minimum number of sequences in the cluster and 2) the possibility of analyzing orthologs and paralogs.

Multialignment of protein families

Each protein family was aligned with MAFFT and the default parameters.

Inference of the protein family trees

ML trees were inferred using the iqtree program run with 1) the default parameters and 2) the bootstrapping.

Consensus tree

The consensus tree was constructed using the iqtree majority rule with a minimum support threshold of 0.5.

Supertree

Supertree was constructed using the Fasturec program with standard parameters.

Computer resources - student server:
- Cores: 96
- Memory: 255 GB

III. Results

434 ortholog clusters were obtained. Shih and colleagues (2012) did an extensive cyanobacteria phylogenetics modeled on the phylogenetically driven Genomic Encyclopedia of Bacteria and Archaea. Their Maximal-likelihood phylogeny on whole genomes will be used to compare the samples (Fig. 2).
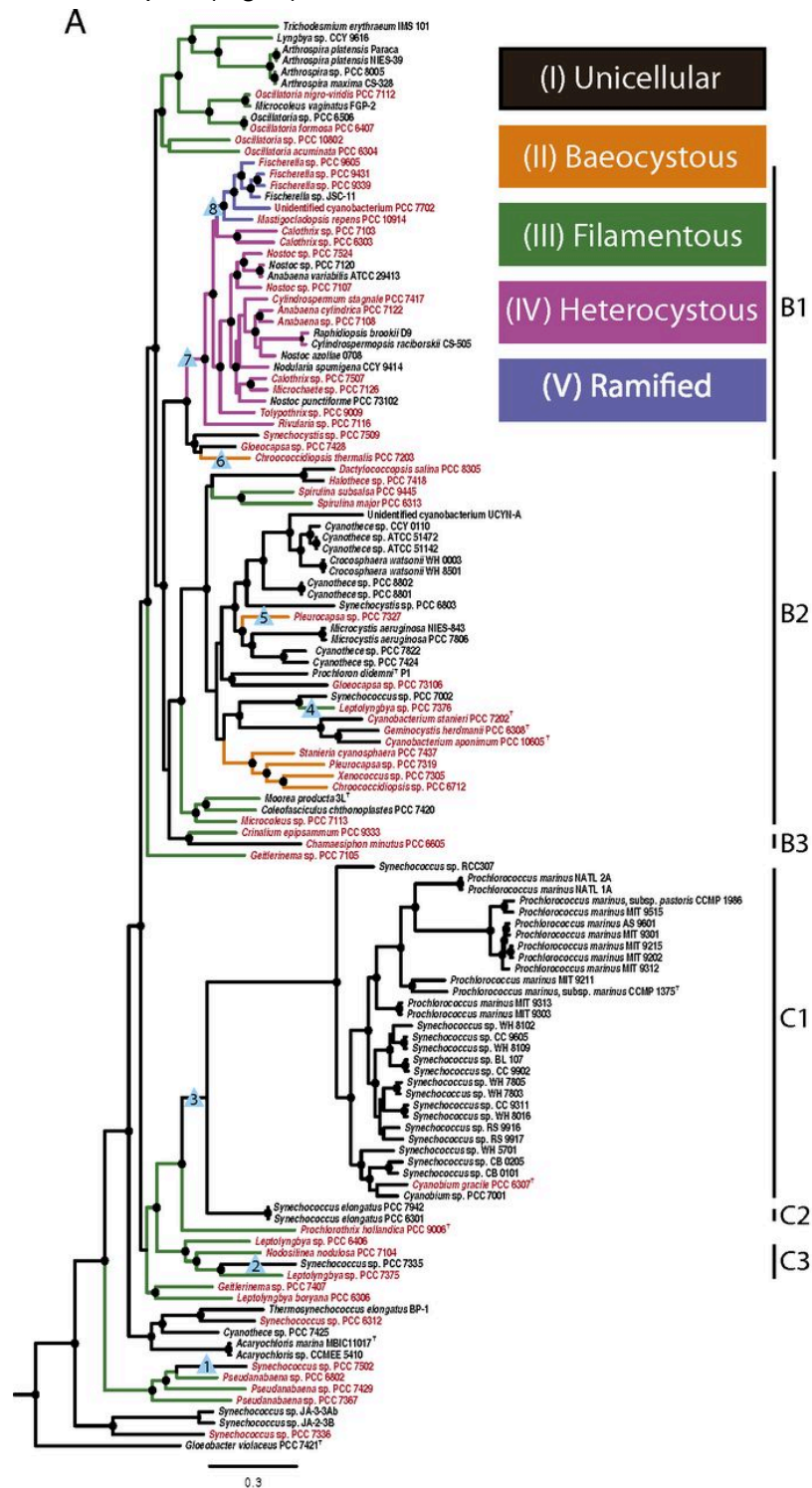

Fig. 2. ML phylogeny. Branches are color-coded according to to morphological subsection, Shih, et al., (2012).
Figure 3 compares the ortholog consensus tree with the bootstrapped ortholog consensus tree.
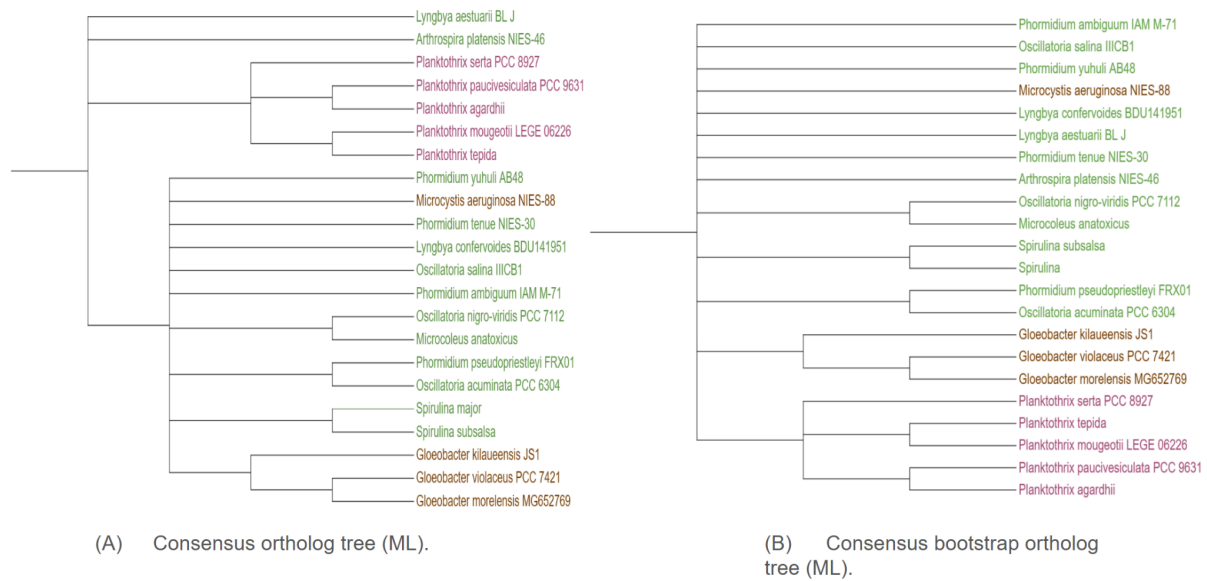
Fig. 3. (A) Consensus ortholog tree (ML), (B) Consensus bootstrapped ortholog tree. The colored species correspond to colored morphological subsection: green - filamentous, brown - unicellular, pink - heterocystous.

Robinson-Foulds distance for the consensus ortholog tree and the consensus bootstrapped ortholog tree is 0.111.

The heterocystous morphology groups together according to their genus, *Planktothrix*. Similarly *Gloeobacter* which represents unicellular morphology, however, *Microcystis aeruginosa* remains separate. According to the known phylogeny (Shih, et al., (2012), Leboulanger et al., 2017) *Lyngbya* and *Arthrospira* should group, however on both trees are separate or are within different clades, which is incorrect. Similarly, *Phormidium* does not group within its genus.

Figure 4 compares the obtained ortholog supertree with the bootstrapped ortholog supertree.
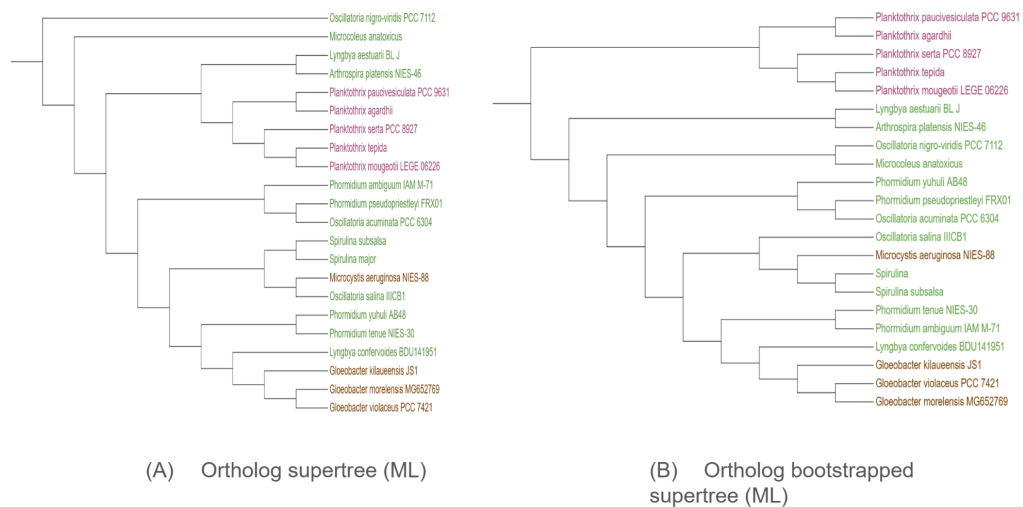

Fig. 4. (A) Ortholog supertree (ML), (B) bootstrapped ortholog supertree. The colored species correspond to colored morphological subsection: green - filamentous, brown - unicellular, pink - heterocystous.

Robinson-Foulds distance for the ortholog supertree and bootstrapped ortholog supertree is 0.179.

All morphological subsections group together, with the exception of *Microcystis aeruginosa.* According to the publication Shih, et al., (2012) *Microcystis aeruginosa* and *Spirulina* are within the same clade, so the *Microcystis aeruginosa* placement on the supertrees seems correct. *Lyngbya aestuarii* groups correctly with *Arthrospira platensis*, however, other representatives of the *Lyngbya* genus are scattered on the tree, similarly *Phormidium* genus. *Gloeobacter* correctly groups with its genus, however, the clade placement on the tree is incorrect. In the Leboulanger et al., (2017), the *Gloeobacter* genus represents the outgroup. Similarly, in the Shih, et al., (2012) phylogeny, *Gloeobacter* is separated.

IV.   Discussion

The TimeTree of Life integrates molecular data from publications to infer divergence about the species. In Figure 5 there is Timetree inferred for the species of this study.
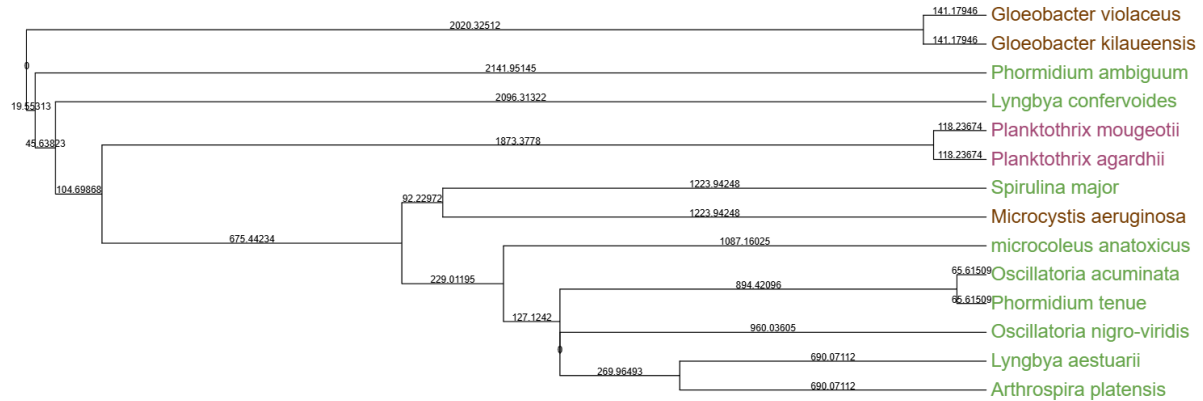

Figure 5. TimeTree inferred for the cyanobacteria species.

Since TimeTree integrates many sources of data, we can treat its result as the reference. Interestingly, similarly to my results, *Phormidium* is not clustered within its genus. According to the NCBI Taxonomy Browser, *Phormidium ambiguum* and *Phormidium tenue* belong to different orders, *Aerosakkonematales* and *Oscillatoriales* respectively, and that might be the reason for the observed data. In the case of the *Lyngbya* genus that is scattered on the plot, the reason behind this observation remains unclear, since for both species, the lineage is the same. However, *Lyngbya aestuarii* clusters with *Arthrospira platensis*, as it was mentioned before, it is phylogenetically consistent with literature.

Gloeobacter genus on the TimeTree is correctly separated from the other clusters.

Similarly to my results, *Spirulina* and *Microcystis* cluster together, even though they differ morphologically.

V. Literature

Cheung, M. Y., Liang, S., & Lee, J. (2013). Toxin-producing cyanobacteria in freshwater: A review of the problems, impact on drinking water safety, and efforts for protecting public health. In *Journal of Microbiology* (Vol. 51, Issue 1, pp. 1–10). https://doi.org/10.1007/s12275-013-2549-3

Miranda, M. S., Cintra, R. G., Barros, S. B. M., & Mancini-Filho, J. (1998). Antioxidant activity of the microalga Spirulina maxima. In Braz J Med Biol Res (Vol. 31, Issue 8).

Geitler, L. (1932). "Cyanophyceae" in Kryptogamen-Flora von Deutschland, Vol. 14, ed L. Rabenhorst (Leipzig: Akademische Verlagsgesellschaft), 673–1196.

Alex, A., Vasconcelos, V., Tamagnini, P., Santos, A., & Antunes, A. (2012). Unusual Symbiotic Cyanobacteria Association in the Genetically Diverse Intertidal Marine Sponge Hymeniacidon perlevis (Demospongiae, Halichondrida). PLoS ONE, 7(12). https://doi.org/10.1371/journal.pone.0051834

Brunberg, A.-K. (2006). Contribution of bacteria in the mucilage of Microcystis spp. (Cyanobacteria) to benthic and pelagic bacterial production in a hypereutrophic lake. FEMS Microbiology Ecology, 29(1), 13–22. https://doi.org/10.1111/j.1574-6941.1999.tb00594.x

Worm J, Sondergaard M. Dynamics of heterotrophic bacteria attached to Microcystis spp. (cyanobacteria). Aquat Microb Ecol. 1998;14:19–28.

Erwin, P. M., Olson, J. B., & Thacker, R. W. (2011). Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the northern Gulf of Mexico. PLoS ONE, 6(11). https://doi.org/10.1371/journal.pone.0026806

Grossart, H. P., Czub, G., & Simon, M. (2006). Algae-bacteria interactions and their effects on aggregation and organic matter flux in the sea. Environmental Microbiology, 8(6), 1074–1084. https://doi.org/10.1111/j.1462-2920.2006.00999.x

SHI, L., CAI, Y., YANG, H., XING, P., LI, P., KONG, L., & KONG, F. (2009). Phylogenetic diversity and specificity of bacteria associated with Microcystis aeruginosa and other cyanobacteria. Journal of Environmental Sciences, 21(11), 1581–1590. https://doi.org/10.1016/S1001-0742(08)62459-6

Leboulanger, C., Agogué, H., Bernard, C., Bouvy, M., Carré, C., Cellamare, M., Duval, C., Fouilland, E., Got, P., Intertaglia, L., Lavergne, C., Le Floc'h, E., Roques, C., & Sarazin, G. (2017). Microbial Diversity and Cyanobacterial Production in Dziani Dzaha Crater Lake, a Unique Tropical Thalassohaline Environment. PloS one, 12(1), e0168879. https://doi.org/10.1371/journal.pone.0168879

Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Natl Acad Sci U S A. 2013 Jan 15;110(3):1053-8. doi: 10.1073/pnas.1217107110. Epub 2012 Dec 31. PMID: 23277585; PMCID: PMC3549136.