TITLE: METAGENOMICS FOR ASSESSING THE ECOLOGY OF FRESHWATER CYANOBACTERIA

AUTHORS: MOREIRA-SILVA[1], E.; GARCIA S.L.[2]; NUY, J.K.[2]; PEURA, S.[3]; KLUGE M.[3]; MEDINA-SILVA, R[1].

INSTITUTIONS:

[1]PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL, PORTO ALEGRE, RS (AVENIDA IPIRANGA, 6681, CEP 90619-900, PORTO ALEGRE – RS, BRAZIL)

[2]DEPARTAMENT OF ECOLOGY, ENVIRONMENT AND PLANT SCIENCES, SCIENCE FOR LIFE LABORATORY, STOCKHOLM UNIVERSIY, STOCKHOLM (FRESCATIVAGEN, 10691, STOCKHOLM – SWEDEN)

[3]DEPARTMENT OF FOREST MYCOLOGY AND PLANT PATHOLOGY, SCIENCE FOR LIFE LABORATORY, SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES, UPPSALA (ALMAS ALLÉ 5, 750 07, UPPSALA – SWEDEN)

ABSTRACT:

Cyanobacteria is a phylum of gram-negative free-living photosynthetic bacteria that are commonly known as blue-green algae. These bacteria can be encountered on oceans, freshwater, soil, and even extreme environments like bare rocks, hot springs, deserts, and Antarctic ice. Besides being major contributors to the primary production of numerous environments (e.g. oceans, hot springs, and hypersaline environments), they also play an important role in the global biogeochemical cycles, as *Cyanobacteria* fixate a significant amount of carbon and nitrogen on marine waters. However, there is a lack of studies regarding *Cyanobacteria* ecogenomics beyond marine environments in the scientific literature. Freshwater environments are exceptionally relevant to global biogeochemical cycles, especially to the carbon cycle (e.g. Lakes store about 30-60% as much organic C per year compared to the ocean, possessing less than 2% of the area of the seas). Therefore, studies exploring the diversity and functions of freshwater cyanobacteria are essential to further elucidate their role in freshwater environments, such as lakes, rivers, and ponds. This project aims to use metagenomic tools to describe the ecologic features *of Cyanobacteria*, such as diversity, abundance, and genomic content in multiple freshwater environments across the globe. To do so, we have gathered metagenomic data (both, publicly available and generated for this study) of water and sediment samples from 74 freshwater sites, including lakes, rivers, and ponds. *Cyanobacteria* genomes derived from our Brazilian samples were assembled using *Megahit* software with a minimum contig size of 1000 kb. Additionally, we have downloaded almost four thousand cyanobacterial genomes from *Integrated Microbial Genomes and Microbiomes* (img.jgi.doe.gov) and *Genome Taxonomy Database* (gtdb.ecogenomic.org). Currently, we are estimating genomes' completeness and contamination values with *CheckM*. Only genomes with at least 90% completeness and less than 5% contamination will be kept. The remaining genomes will have their taxonomy reassigned using GTDBtk. Then, genomes will be dereplicated and merged into mOTUs by *mOTUlizer*. Posteriorly, all mOTUs will be mapped across all freshwater samples'

reads using *BBmap* with 100% identity. Finally, mOTUs' distribution, abundances, diversity, and metabolic potential will be observed using R packages, such as *phyloseq* and *vegan* packages.

Keywords: ecogenomics, algae, diversity, lake, river, pond

Development Agency: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)