# Uncovering novel biosynthetic gene clusters and chemical diversity in marine cyanobacterium Capilliphycus salinus ALCB114379 through metabologenomics

Gabriel Schimmelpfeng Passos[1], Fernanda Rios Jacinavicius[1], Tiago Ferreira Leão[2], Ian Castro-Gamboa[2], Marli de Fátima Fiore[1]

[1.] CENA/USP, Center for Nuclear Energy in Agriculture at the University of São Paulo, Av. Centenário, 303 - São Dimas, Piracicaba - SP, 13400-970, Brazil.;
[2.] Unesp, São Paulo State University, Av. Professor Francisco Degni, 55 - Jardim Quitandinha, Araraquara - SP, 14800-900, Brazil;

Marine environments are among the most dynamic on Earth, inhabited by a diverse array of microorganisms that play crucial roles in biogeochemical cycles. Cyanobacteria, capable of oxygenic photosynthesis, are found in various marine landscapes. Their ability to thrive in different environments is linked to a complex genetic arsenal for producing secondary metabolites, such as cyanopeptides, which often have bactericidal and antifungal properties. Advances in genomic sequencing are revealing the taxonomic and functional diversity of the marine microbiome. When combined with metabologenomics, it becomes possible to detect unique natural compounds and link them to their biosynthetic genes. However, only 19% of natural cyanobacterial products are associated with sequenced organisms due to the limited availability of genomes in databases, particularly from the order Oscillatoriales. This study aims to identify biosynthetic gene clusters (BGCs) for cyanopeptide production in the marine cyanobacterium *Capilliphycus salinus* ALCB114379, a recently classified type strain in a new genus, and link them to their metabolites. We selected 62 cyanobacterial genomes, including *C. salinus* ALCB114379, from the National Center for Biotechnology Information (NCBI) and subjected them to BGC annotation using the AntiSMASH tool with default parameters in relaxed settings. Kinship analysis among the 387 potential BGCs was performed by generating gene cluster families (GCFs) with BigSCAPE using default parameters. In parallel, the untargeted metabolomic analysis involved cultivating *C. salinus* ALCB114379 in Z8 medium for 30 days in a BOD chamber. 20 mg of freeze-dried biomass were extracted with 100% acetonitrile. The supernatants were filtered through 0.22 ?m PVDF filters and transferred into HPLC vials before LC/MS analysis using an isocratic method on an LC/MS QTOF Xevo G2-XS. The raw mass spectrometry data were converted into mzXML file using MSConvert, and the resulting files were input into Molecular networks obtained using the Global Natural Product Social Molecular Networking (GNPS) platform for spectral molecular network construction and annotation by comparing MS/MS spectra against in-house databases and the literature. BigSCAPE analysis identified 5 GCFs: 2 RiPPs (one shared with *Microcystis* and the other with *Limnoraphis*), 1 NRPS (shared with *Lyngbya* and *Limnoraphis*), 1 PKSI (shared with *Limnoraphis*), and a terpene (shared with *Okeania* and *Limnoraphis*). Except for terpenes, these metabolite families generally include cyanopeptides. GNPS results recorded an average of 982.5 MS1 and 4,269.5 MS2 counts, with only 17 hits in the GNPS database, indicating a vast diversity of unknown metabolites. This diversity underscores the potential of marine cyanobacterium in producing natural products. Combining data from AntiSMASH, BigSCAPE, and GNPS, the NPLinker tool can link identified metabolites to their respective BGCs, paving the way for discovering new natural products of interest.

***Agradecimentos:*** Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Cyanos research group.