

# Genome mining of biosynthetic gene clusters in *Nostoc* sp. CACIAM 19, a cyanobacterium from an Amazonian environment

David Batista Maués<sup>1</sup>, Alex Ranieri Jerônimo Lima<sup>1</sup>, Pablo Henrique Gonçalves Moraes<sup>2</sup>, Andrei Santos Siqueira<sup>1</sup>, Leonardo Teixeira Dall'Agnol<sup>3</sup>, Evonnildo Costa Gonçalves<sup>1</sup>

<sup>1</sup>Laboratório de Tecnologia Biomolecular, Instituto de Ciências Biológicas, Universidade Federal do Pará

<sup>2</sup>Laboratório de Farmacognosia 2, Campus São Luís, Universidade Federal do Maranhão

<sup>3</sup>Campus Bacabal, Universidade Federal do Maranhão

Cyanobacteria comprise the largest, most diverse, and the most widely distributed group of photosynthetic prokaryotes, being found in all types of terrestrial and aquatic ecosystems, and constitutes a rich source of compounds of great significance in biotechnology. The *Nostoc* genus, constituted by filamentous heterocystic cyanobacteria, nitrogen fixing, has been utilized as a source of vitamins, proteins and fatty acids, also of secondary metabolites with anticancer, antimicrobial and antiviral activities. The number of these organisms' genomes publicly available has rapidly grown in the last few years. In this sense, genome mining of their biosynthetic gene clusters has become a key approach for novel compound discovery. In order to evaluate its biotechnological potential, we obtained the draft genome of *Nostoc* sp. CACIAM 19. This strain was isolated from a water sample from Bolonha lake, localized in Belém, Pará. After DNA extraction of the cyanobacterial non-axenic culture, two sequencing runs were performed on the GS FLX 454 (Roche Life Sciences) platform using non-paired libraries, and one sequencing run was carried out on the Illumina MiSeq platform using a paired-end library with 150 bp read length. A co-assembly of all reads was performed by Newbler 2.9 and the resulting scaffolds were binned with MaxBin 2.2.1 and then submitted to the antiSMASH 3 tool. The predicted clusters were manually evaluated using the data banks Pfam, Uniprot, NCBI and Interproscan. Thirty-two clusters of biosynthetic genes for secondary metabolites were identified, among them: 12 exopolysaccharides biosynthesis clusters, 3 bacteriocins biosynthetic clusters, 1 cyanobactin biosynthetic cluster, 4 fatty acid production clusters, 4 hydrocarbon production clusters, 1 microviridin producing cluster, 1 phosphonate biosynthetic cluster and 6 NRPKS/KS clusters. The terpenoid and hydrocarbon clusters, whose products can be utilized in the chemical industry, showed up to 35% of similarity with the clusters present in *Nostoc punctiforme* PCC 73102. Such results show the potentiality and diversity of secondary metabolites produced by the amazon cyanobacteria *Nostoc* sp. CACIAM19, which can be explored in diverse sectors of the biotechnology industry.

Support: CNPq, CAPES, CIT-IEC, FAPESPA.