BIOINFORMATICS APPROACHES ON THE STUDY OF CARBON SEQUESTERING AND BIOFUEL PRODUCTION USING CYANOBACTERIA

Tulip S Nandu, Meeta Pradhan (Mathew J Palakal)
School of Informatics, Indiana University-Purdue University, Indianapolis, Indiana, 46204.
This project is funded by International Development Fund (IDF) from the office of Vice Chancellor of Research.

Abstract:
Background: Bioenergy production has recently become a topic of intense interest due to increased concern regarding limited petroleum-based fuel supplies and the contribution of the use of these fuels to atmospheric CO$_2$ levels. Cyanobacteria in general have a diversity of pathways that allows these organisms to succeed in a wide variety of environments and which provides a wealth of targets for metabolic engineering of energy-rich biomolecules. This is an International Collaborative Project with the Indian Institute of Technology (IITB), Bombay, India.

Objective: The objective is to carry out an initial study of the 35 cyanobacterial species and identify the most likely species that would be best suited for manipulating the regulatory networks to enable biofuel production. Methodology: The sequence data from NCBI, Cyanobase, and protein information was used to predict operons across 35 species of cyanobacteria. This data is then used to build an initial network of gene/protein interaction of cyanobacteria and homologous species having maximum hits across the species using biomedical text mining techniques. The network is further expanded by selecting an organism that is less studied and mapping the results of the interactions to this organism using BLAST results. Finally, using Machine Learning methods the significant genes and significant motifs in similar and dissimilar network of cyanobacteria species are identified.

Results: We have identified about 100 homologous species for each gene within 35 cyanobacteria species. The organism
Paulinella chromatophora chromatophore had approximately 400 genes similar across 23 species and Clostridium botulinum and Halanaerobium praevalen were similar to 12 species of cyanobacteria. We can now use these species to build gene/protein interaction network to discover the functions of the genes within these species.