

Metagenomics of Cyanobacterial Blooms

Phillip Pope – PhD Thesis Abstract

School of Biomolecular and Biomedical Science
Griffith University

2007

Cyanobacteria are a diverse and widely distributed group of organisms common in soil and in both marine and freshwater. Under favorable conditions they can reproduce explosively, forming dense concentrations called blooms. Fresh water cyanobacterial blooms in particular are commonly associated with toxin production in drinking water supplies and are increasingly becoming a risk to human health. Beyond toxin production these extremely complex, constantly interacting and changing microbial communities have vast impacts on their surrounding ecosystem. The triggers that initiate bloom formation and/or toxin production remain poorly understood. This stems from the fact that there is still very little known of cyanobacterial bloom population structure and their function in the real environment.

A greater understanding of the interactions of different microbial populations and their functions in the blooming process leading to toxin production could come from using metagenomics to investigate the genetic and metabolic diversity of the mixed populations rather than the difficult to culture cyanobacteria. Two distinct cyanobacterial bloom communities existing in contrasting Australian freshwater lakes were selected and high molecular weight DNA extracted. PCR-amplified 16S rRNA genes were subsequently cloned and a total of 75 clones from Lake Samsonvale and 50 clones from Lake Ainsworth were examined. Sequences identified belonged to species from 6 different phyla from the Bacterial domain, including *Cyanobacteria*, *Actinobacteria*, *Firmicutes*, *Verrucomicrobium*, *Bacteroidetes*, and α -, β - and γ -*Proteobacteria*. The majority of the bacterial sequences were most closely related to sequences recovered from other freshwater clones or isolates (<80% homology), whilst few were closely related to sequences recovered from soil or marine habitats. In particular 9 % of the total sequences were most closely related to sequences recovered from freshwater lakes that are susceptible to cyanobacterial blooms. A total of 12 novel clusters consisting of 22 sequences were noted spanning all divisions represented in the analysis. Of this, 7 were found to lack any close relatives suggesting that sequences in these clusters may be characteristic for bloom events. Preliminary results also indicate that physio-chemical differences in lake character appear to influence bacterial community composition associated with cyanobacterial blooms.

Bloom communities from Lake Samsonvale demonstrated high levels of toxin-producing *Cyanobacteria* and uncultured *Actinobacteria*. These findings were used to justify its selection for further metagenomic analysis to gain insights into the genomes of these and other organisms. DNA was fractionated and used to construct a bacterial artificial chromosome library (CBNPD1) of 2,850 clones which had an average insert size of 27 kb. A PCR-based single-gene polyketide synthase library was constructed in tandem and used as an additional assurance that high quality DNA was being extracted and cloned. Phylogenetic analysis of gene sequences

recovered from this library demonstrated an abundance of novel bacterial polyketide synthase genes.

Sequence-based screening of library CBNPD1 was performed to identify clones of interest and provide a physiological insight within cyanobacterial blooms. A random BAC-end sequence survey generated 67 sequences (40 kb in total) from 36 randomly selected clones. G+C composition ranged from 33.33 to 72.91%. Fifteen sequence tags (22%) were found most similar to sequences affiliated to genera with no available genome. Another 17 sequence tags (25%) were most similar to sequences affiliated to genera with available genomes, however similarities were less than 80%. Sequence tags were also found with affiliation to proteins involved in a wide array of cell metabolism processes including amino acid metabolism (e.g. methionine synthase), carbohydrate metabolism (cellulose), inorganic ion metabolism (nitrite/sulfite reductase), and lipid metabolism (fatty acid hydroxylase). A number of genes involved in cell structures (e.g. flagella), DNA processes, energy production (photosynthetic reaction center L subunit) and defense mechanisms (nucleases) were also affiliated to sequence tags. PCR screening of CBNPD1 was used to detect clones containing 16S rDNA to establish a link between physiological and phylogenetic information of uncharacterized microorganisms in cyanobacterial blooms. Screens from 480 clones identified 2 clones containing a 16S rRNA gene. Clone 545 and 578 contained 16S rDNA affiliated to 2 different phylogenetic genera within the *Proteobacteria* division, *Pseudomonas* and *Roseateles* respectively.

From library screens 7 BAC inserts were selected and sequenced to completion comprising 144 kb of a cyanobacterial bloom metagenome and spanning 3 phyla including *Proteobacteria*, *Actinobacteria* and *Bacteroidetes*. 130 genes have been identified and assigned to COG (clusters of orthologous groups of proteins) functional categories. Also identified, were many housekeeping proteins spanning the majority of the COG functional groups as well as physiologically and ecologically important proteins some of which were looked at more in depth. These include a putative phenylacetyl catabolon, a putative RTX toxin, several putative oxidoreductases and several putative bacterial transcriptional regulators that are inferred in controlling a wide variety of activities in various biological processes, the most notable being quorum sensing.

This culture-independent experimental approach has provided a phylogenetic community snap shot of the cyanobacterial bloom community structure and their physiological functions within the bloom. Moreover it represents an important biodiversity resource which has already been shown to contain novel biomolecular biodiversity.