

Poster for Evolution meeting (Providence, R. I. June 21-25, 2019)

May 06, 2019

Title

Dolichospermum evolution during during cyanobacterial bloom: insights from metatranscriptomics.

Authors & Affiliations

Sébastien Renaut^{1,2} Nathalie Fortin³ Charles W. Greer³ B. Jesse Shapiro²

¹Institut de Recherche en Biologie Végétale, Université de Montréal, Montréal, Québec, Canada

²Department of Biological Sciences, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada

³Energy Mining and Environment, National Research Council Canada, Montreal, QC, Canada

Abstract

Specific environmental conditions, such as elevated levels of nutrients from human activities, warmer temperatures, still water, and sunlight promote the growth of cyanobacteria, which at high density, form harmful algae blooms. Each summer, cyanobacteria make an appearance in Lake Champlain (Baie Missisquoi, Qc) and pose health risks to humans and animals due to the production of cyanotoxins. Here, we used a metatranscriptomics approach that targets genes expressed in the water to identify changes in the bacterial community during the summer. At the height of the bloom (mid-September), chlorophyll A levels increased by nearly 100 fold and communities were dominated by the cyanobacteria *Dolichospermum circinale* and to a lesser extent *Microcystis aeruginosa*. We identified 7,000 *Dolichospermum* SNPs (10% of all SNPs) that showed changes in allele frequencies from the beginning of the bloom to the end of the season, showing a highly dynamic evolution of the system. Nineteen genes showed the same repeatable patterns of evolution in all three sampling sites (expected: three) and these represent good candidates to predict blooms in the future.

Introduction

- Specific environmental conditions, such as elevated levels of nutrients from human activities, warmer temperatures, still water, and sunlight promote the growth of cyanobacteria, which at high density, form harmful algae blooms.
- Each summer, cyanobacteria make an appearance in Lake Champlain and pose health risks to humans and animals due to the production of cyanotoxins.

Objectives

- Use a metatranscriptomics approach that targets genes expressed in the water to identify changes in the bacterial community during the summer.
- Identify *Dolichospermum* (main bloom causing cyanobacteria) genes that change in allele (SNPs) frequency from the beginning to the end of the bloom. These represent candidate genes whose changes in allele frequencies may help to predict blooms in the future.

Method

1. Water sampled at three locations in Lake Champlain (Baie Missisquoi: St1, St2, PRM, **see Figure 1**) during the cyanobacterial bloom in the summer of 2016 (June 1st - October 10th).
2. RNA extracted, followed by High Throughput sequencing (100bp paired end Illumina HiSeq).

3. Metatranscriptomes processed (trimming, merging, removal of rRNA, annotation) using SAMSA2 pipeline¹.
4. Sequences specifically annotated to *Dolichospermum* (the main cyanobacteria responsible for the bloom) used to produce a *de novo* reference transcriptome assembly using Trinity²
5. *Dolichospermum* sequences aligned to transcriptome using bowtie2³. SNP called using SAMtools⁴.
6. Changes in allele frequency quantified using regressions (logistic for St1 & St, linear for PRM) in R.
7. Candidate genes showing significant changes (FDR corrected p -value < 0.05) in allele frequency matched to reference database and compared against reference transcriptome.

Results

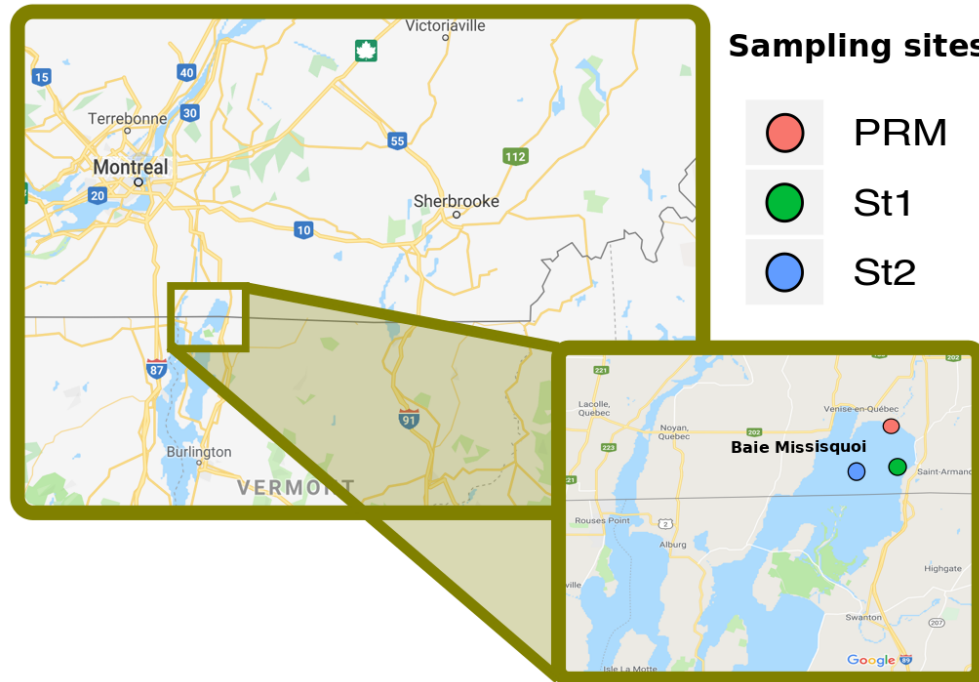


Figure 1: A map of sampling sites

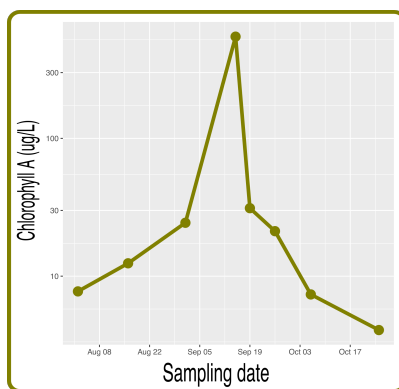


Figure 2: Chlorophyll concentration peaked in mid-September (August 1st: 7.8 ug/L, Sept 15th: 547 ug/L) resulting in a cyanobacterial bloom

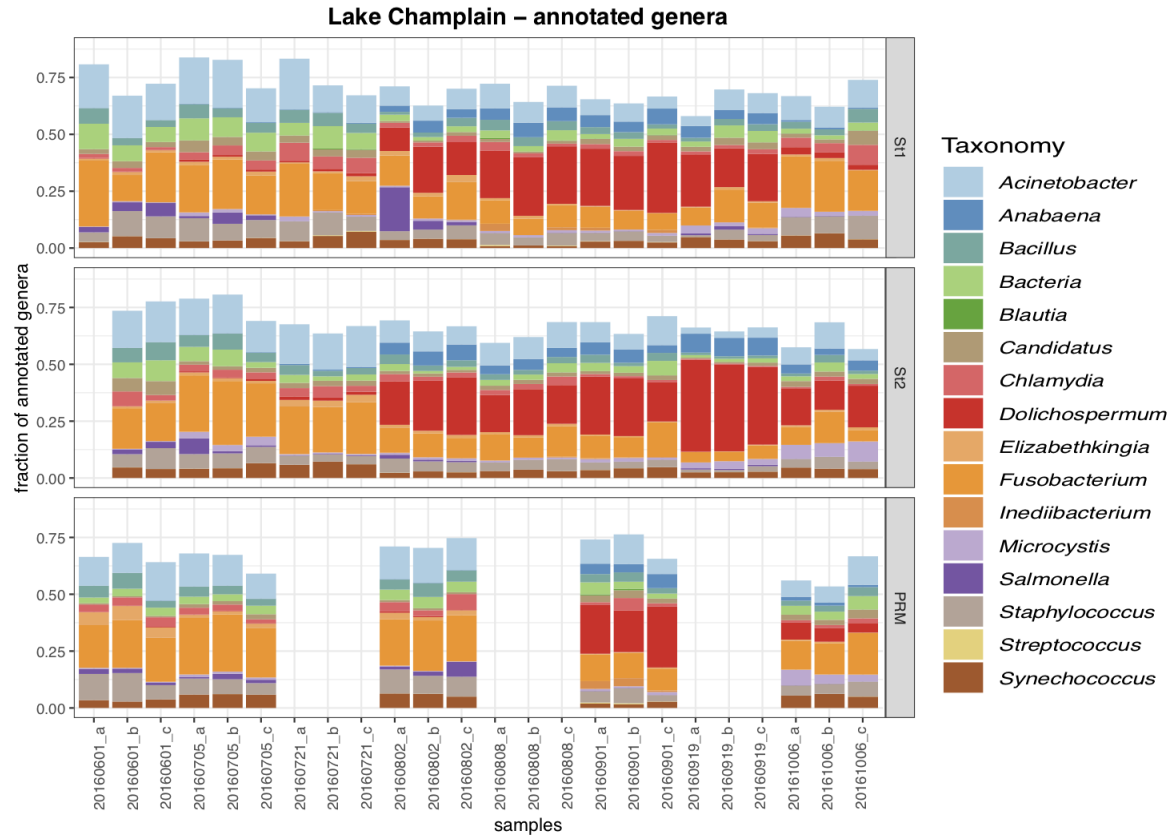


Figure 3: Communities dominated by Gram negative *Fusobacterium* & *Acinetobacter* prior to bloom. During Bloom, *Dolichospermum circinale* rose in frequency, followed to a lesser extent by *Microcystis aeruginosa* in September.

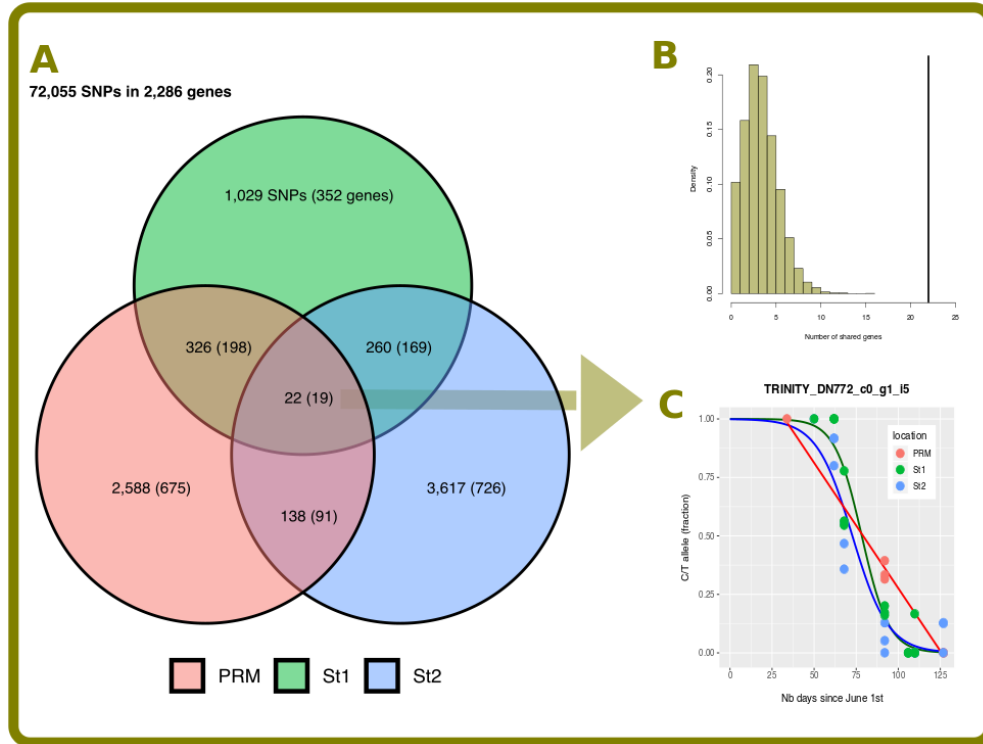


Figure 4: 72,000 SNPs identified in 2,286 *Dolichospermum* genes. A. Venn diagram: Number of SNPs (genes) showing significant changes in allele frequency from beginning (June 1st) to end of bloom (October 6th). C. Null distributions of the number of SNPs expected to show significant allele frequency change in all three sampling sites compared to observed (black vertical line, 22 SNPs observed, p -value < 0.0001). B. Example of a significant SNP in both St1, St2 (logistic regressions) and PRM (linear regression) sites.

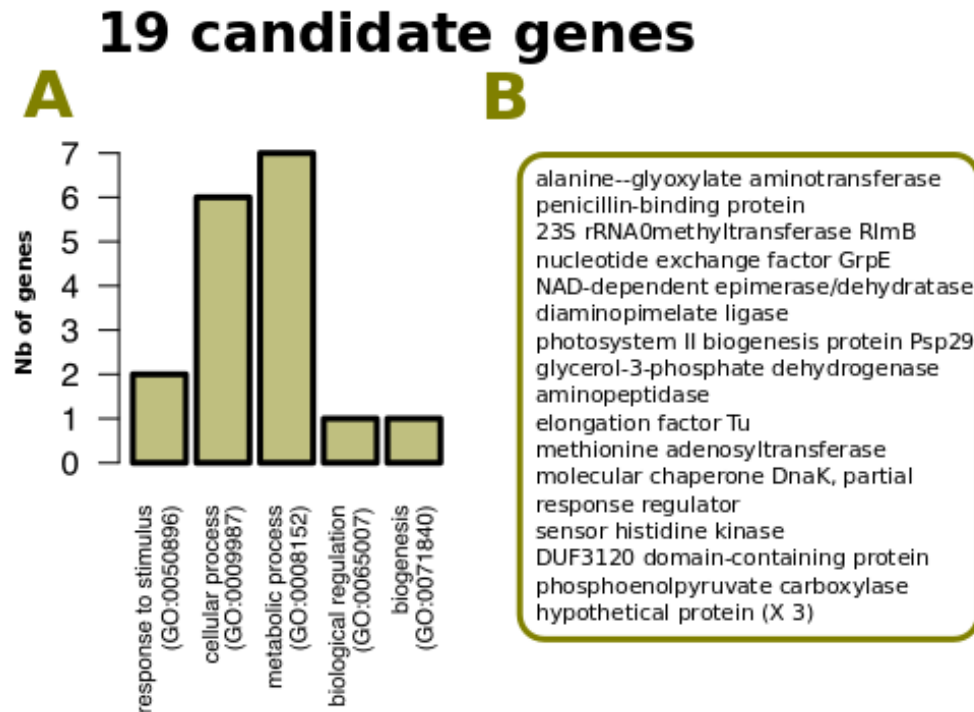


Figure 5: A. Gene ontology groups for candidate genes (no significant over-represented groups compared to reference transcriptome). B. List of candidate genes.

Conclusions

- Metatranscriptomics can identify cyanobacteria known to cause blooms (*Dolichospermum circinale*, *Microcystis aeruginosa*).
- ~7,000 SNPs (10% of all SNPs) show changes in allele frequencies during the bloom in at least one of the sampling site, showing a highly dynamic evolution of the system. Nineteen genes show repeatable changes in allele frequencies in all three sampling sites. These represent good candidates to predict bloom.
- Further sampling and sequencing in progress in other lakes in order to confirm patterns observed here.

References

1. Westreich, S. T., Treiber, M. L., Mills, D. A., Korf, I. & Lemay, D. G. SAMSA2: A standalone metatranscriptome analysis pipeline. *BMC bioinformatics* **19**, 175 (2018).
2. Grabherr, M. G. *et al.* Trinity: Reconstructing a full-length transcriptome without a genome from rna-seq data. *Nature biotechnology* **29**, 644 (2011).
3. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with bowtie 2. *Nature methods* **9**, 357 (2012).
4. Li, H. *et al.* The sequence alignment/map format and samtools. *Bioinformatics* **25**, 2078–2079 (2009).