Dolichospermum evolution during during cyanobacterial bloom: insights from metatranscriptomics.

Sébastien Renaut[1,2], Nathalie Fortin[3], ,Charles W. Greer[3], B. Jesse Shapiro[2]

[1]Institut de Recherche en Biologie Végétale, Université de Montréal, Montréal, Québec, Canada. [2]Department of Biological Sciences, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada. [3]Energy Mining and Environment, National Research Council Canada, Montreal, QC, Canada

sebastien.renaut@umontreal.ca @seb_renaut seb951.github.io/atrapp/RENAUT_evol_RI2019.pdf

ATRAPP - Algal Blooms, Treatment, Risk Assessment, Prediction and Prevention Through Genomics

Stuff you need to know

- Specific environmental conditions, such as elevated levels of nutrients from human activities, warmer temperatures, still water, and sunlight promote the growth of cyanobacteria. At high density, they form **harmfull 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.

Big sciency questions

- Use a **metatranscriptomics** approach that targets genes expressed in the water to identify changes in the bacterial community during the summer.

- **Identify cyanobacteria genes** that change in allele (SNPs) frequency during the bloom.

Fancy analyses (i.e boring stuff)

1. Water sampled at three location in Lake Champlain (Baie Missisquoi) during the cyanobacterial bloom (June 1st – October 6th 2016).

2. RNA extracted, followed by High Throughput sequencing (100bp paired end Illumina HiSeq).

3. Metatranscriptomes processed using SAMSA2 pipeline (trimming, merging, removal of rRNA, annotation)[1].

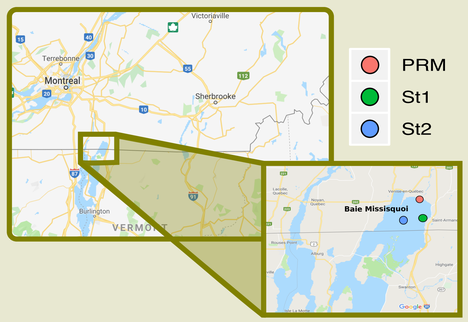
4. Sequences specifically annotated to Dolichospermum (the main cyanobacteria responsible for the bloom) used to produce a de novo reference transcriptome assembly using Trinity[2].

5. Dolichospermum sequences aligned to transcriptome using bowtie2[3]. SNP called using SAMtools[4].

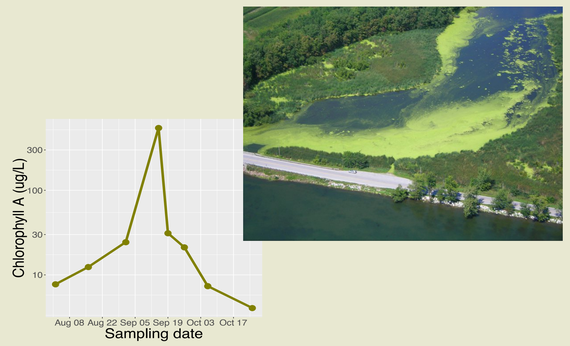
6. Changes in allele frequency quantified using regressions (logistic for St1 & St2, linear for PRM).

7. Candidate genes showing significant changes (q-value < 0.05) in allele frequency BLASTxed to reference database.

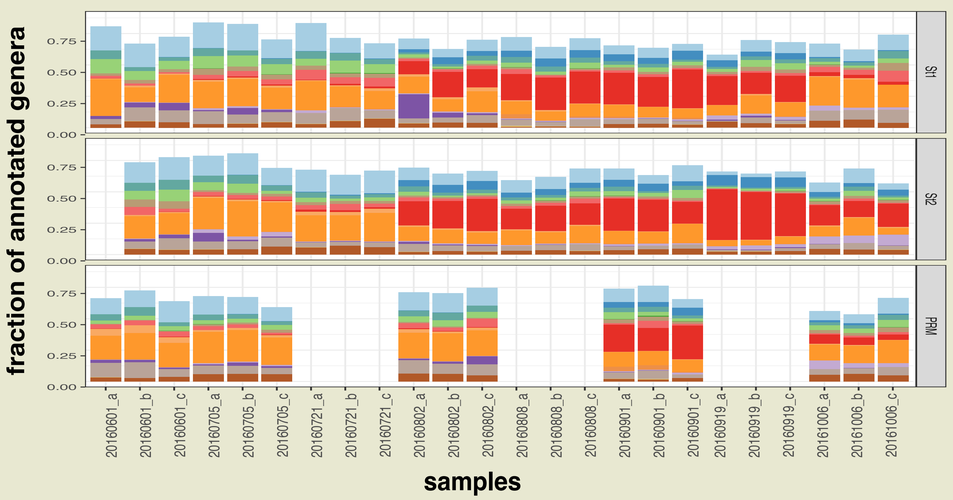
Things we find in the lake (i.e. cool stuff)



Sampling sites



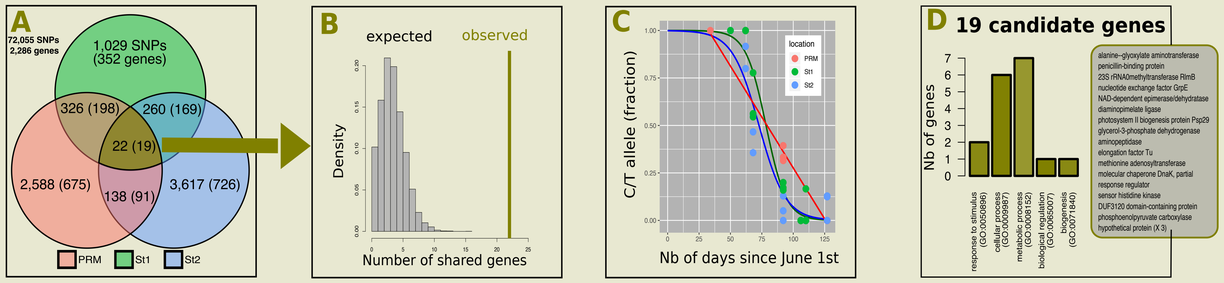
[Chlorophyll] peaked in mid-September resulting in a cyanobacterial bloom



Taxonomy

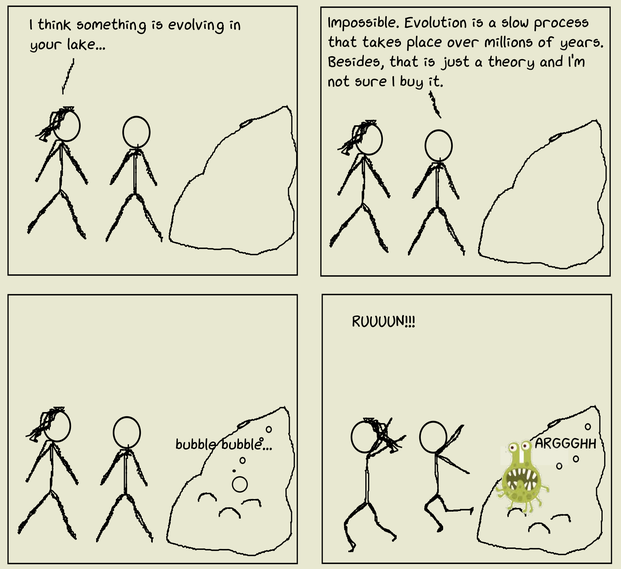
- ***Acinetobacter***
- *Anabaena*
- *Bacillus*
- *Bacteria*
- *Blautia*
- *Candidatus*
- *Chlamydia*
- ***Dolichospermum***
- *Elizabethkingia*
- ***Fusobacterium***
- *Inediibacterium*
- ***Microcystis***
- *Salmonella*
- *Staphylococcus*
- *Streptococcus*
- *Synechococcus*

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.



A. 72k SNPs identified in 2,286 Dolichospermum genes. Venn: Number of SNPs (genes) showing significant changes in allele frequency from beginning (June 1st) to end of bloom (October 6th). **B.** Example of a SNP showing a change from C to T allele in both St1, St2 (logistic regressions) and PRM (linear regression). **C. 22 SNPs showed significant changes in allele frequency in all three sites compared to three expected (p-value < 0.0001). D.** Gene ontology and list of candidate genes.

Possible outcomes



- **Metatranscriptomics is highly effective** in identifying bloom causing cyanobacteria.

- **Dynamic evolution of the system**: ~7,000 SNPs (10% of all SNPs) changed in allele frequencies during the bloom in at least one of the sampling site.

- Nineteen genes showed repeatable changes in allele frequencies in all sampling sites. These represent **candidate genes which may help to predict blooms.**

- **Further sampling in other lakes** will confirm patterns observed here.

Papers I read

[1] Samuel T Westreich et al. SAMSA2: a standalone meta transcriptome analysis pipeline. BMC bioinformatics 19 (2018)

[2] Manfred G Grabherr et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology 29 (2011)

[3] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie2 Nature methods 9 (2012)

[4] Heng Li et al. The sequence alignment / map format and SAM-tools. Bionformatics 25 (2009)

[∞] A bunch more paper and fancy methods I read & used, but not cited here

People paying for the science

GenomeCanada GenomeQuébec Fonds de recherche sur la nature et les technologies Québec

CENTRE DE LA SCIENCE DE LA BIODIVERSITÉ DU QUÉBEC QUEBEC CENTRE FOR BIODIVERSITY SCIENCE

Université de Montréal