

# Dolichospermum evolution during cyanobacterial bloom: insights from metatranscriptomics.

Sébastien Renaud<sup>1,2</sup>, Nathalie Fortin<sup>3</sup>, Charles W. Greer<sup>3</sup>, B. Jesse Shapiro<sup>2</sup>

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



sebastien.renaud@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 **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.

## 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 tools and analyses

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

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

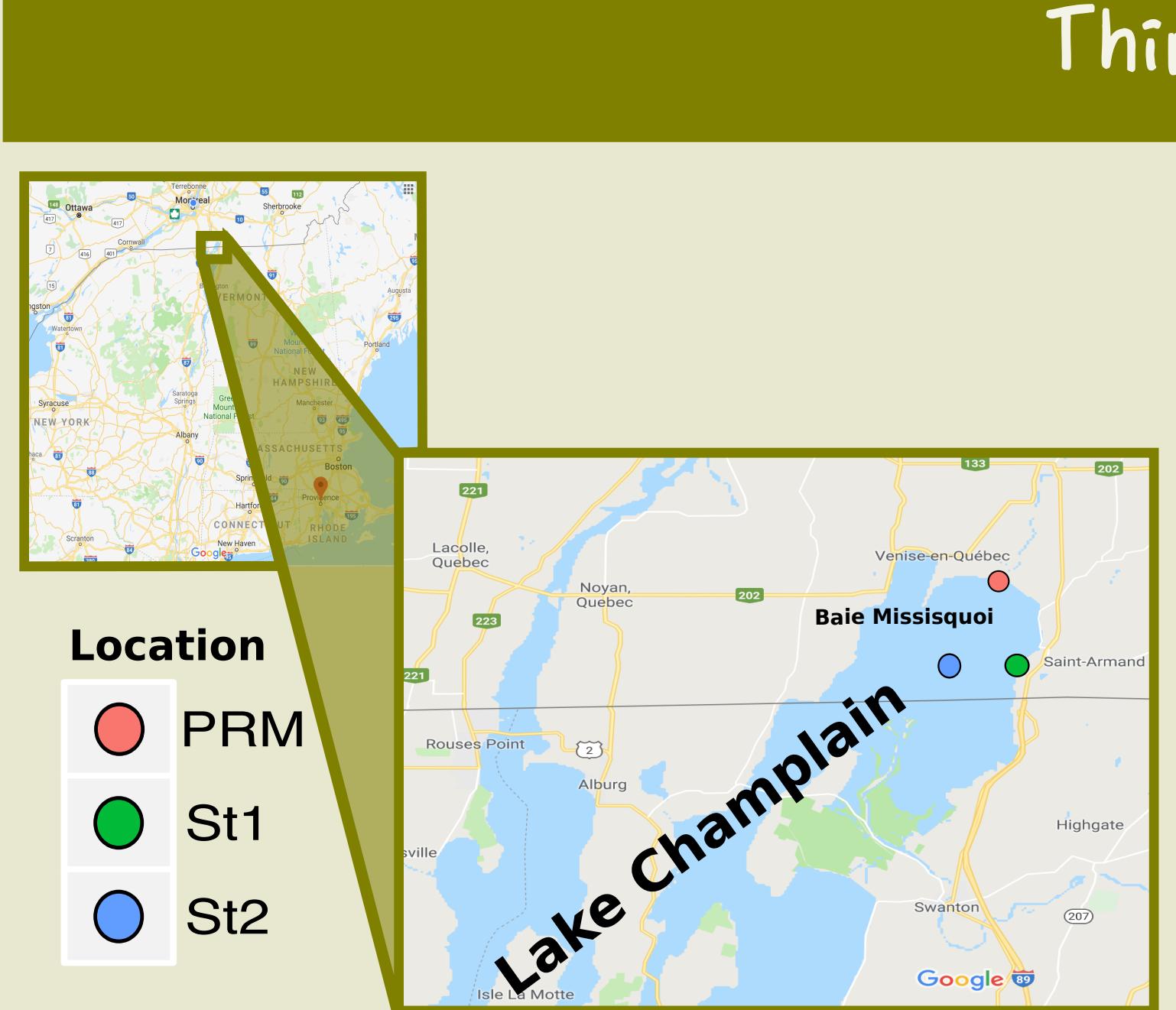
3. Metatranscriptomes processed using SAMSA2 pipeline (trimming, merging, removal of rRNA, annotation).

4. Sequences annotated to Dolichospermum (the main cyanobacteria responsible for the bloom) used to produce a *de novo* reference transcriptome using Trinity<sup>2</sup>.

5. Dolichospermum sequences aligned to transcriptome using bowtie2<sup>3</sup>. SNP called using SAMtools<sup>4</sup>.

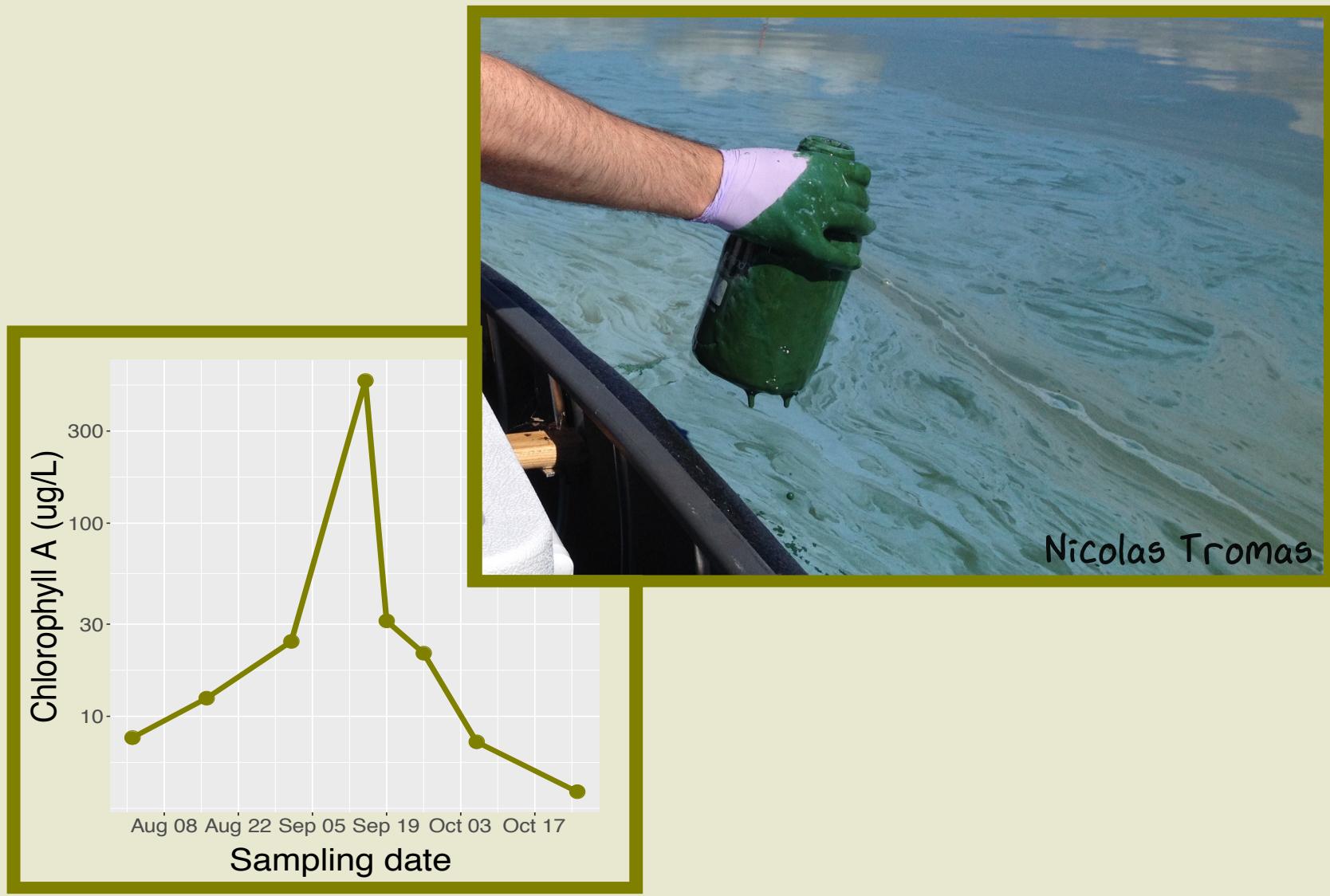
6. Changes in allele frequency quantified using regressions (logistic for St1 & St2, linear for PRM).

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

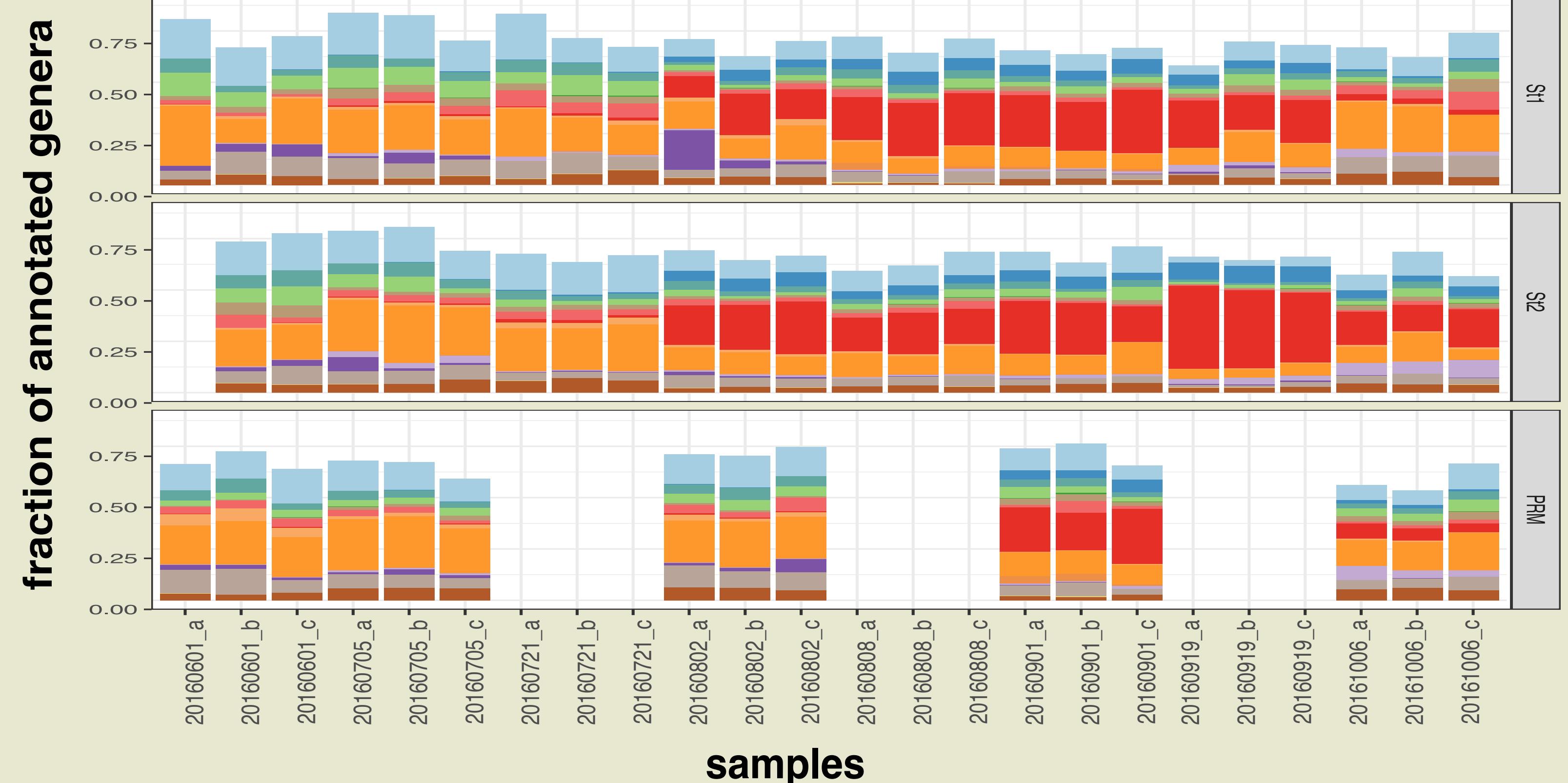


Sampling sites

## Things we find in the lake



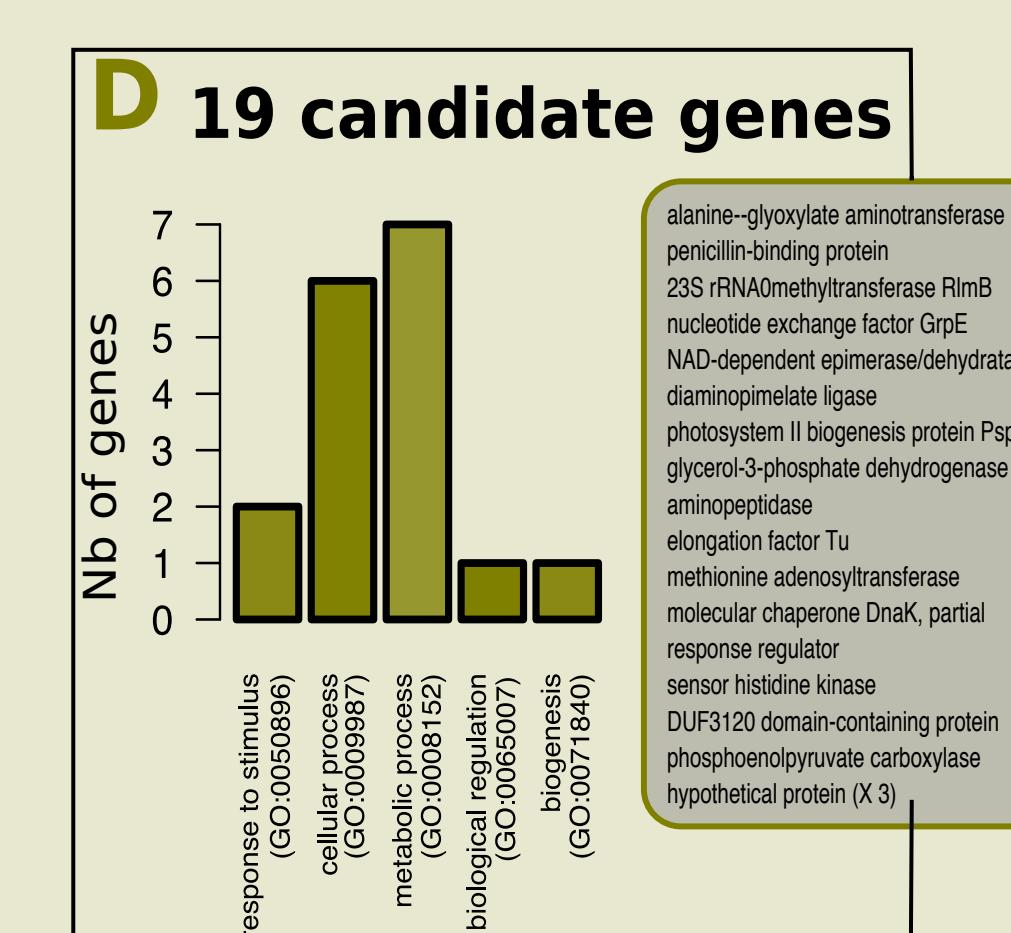
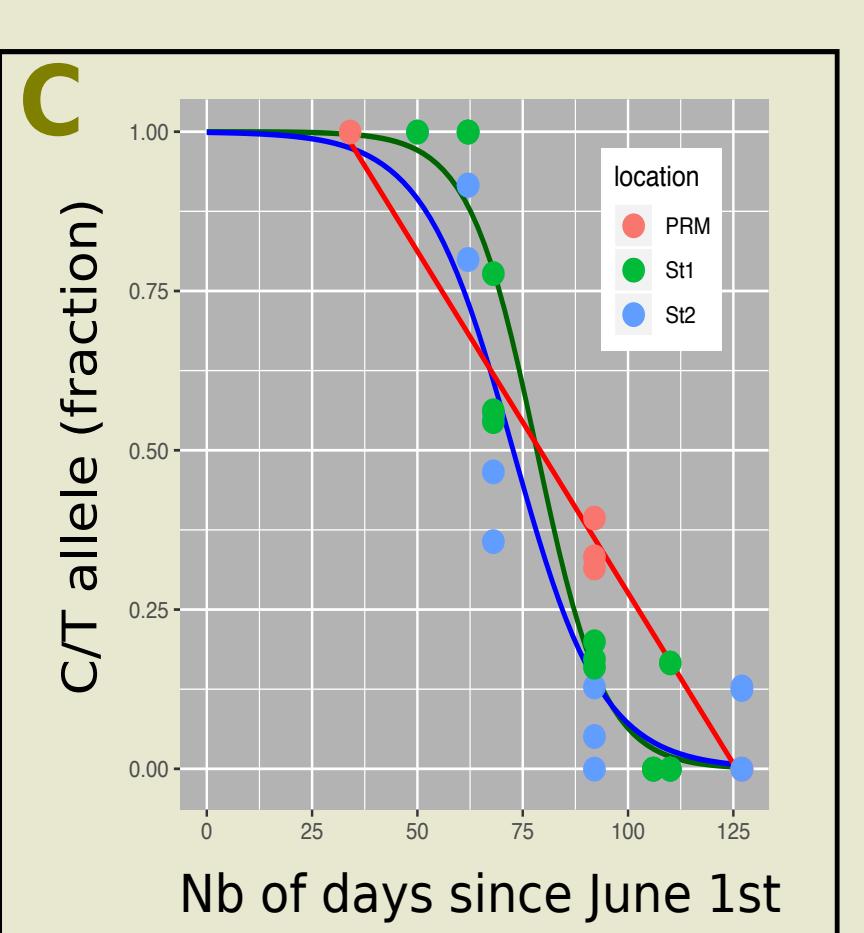
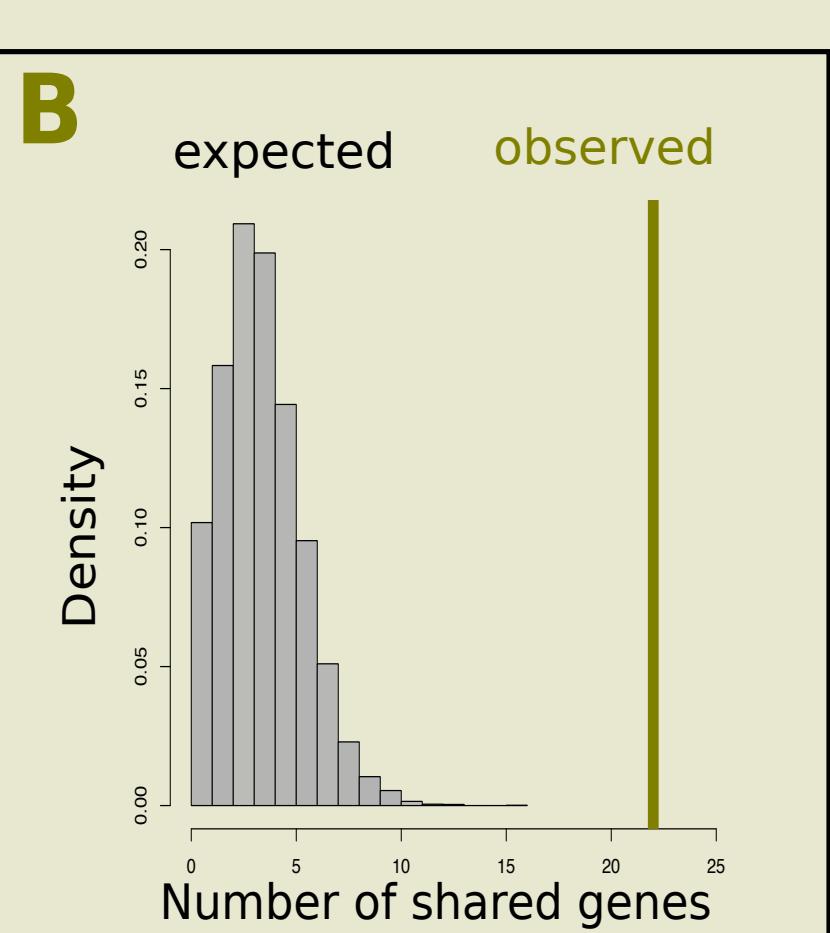
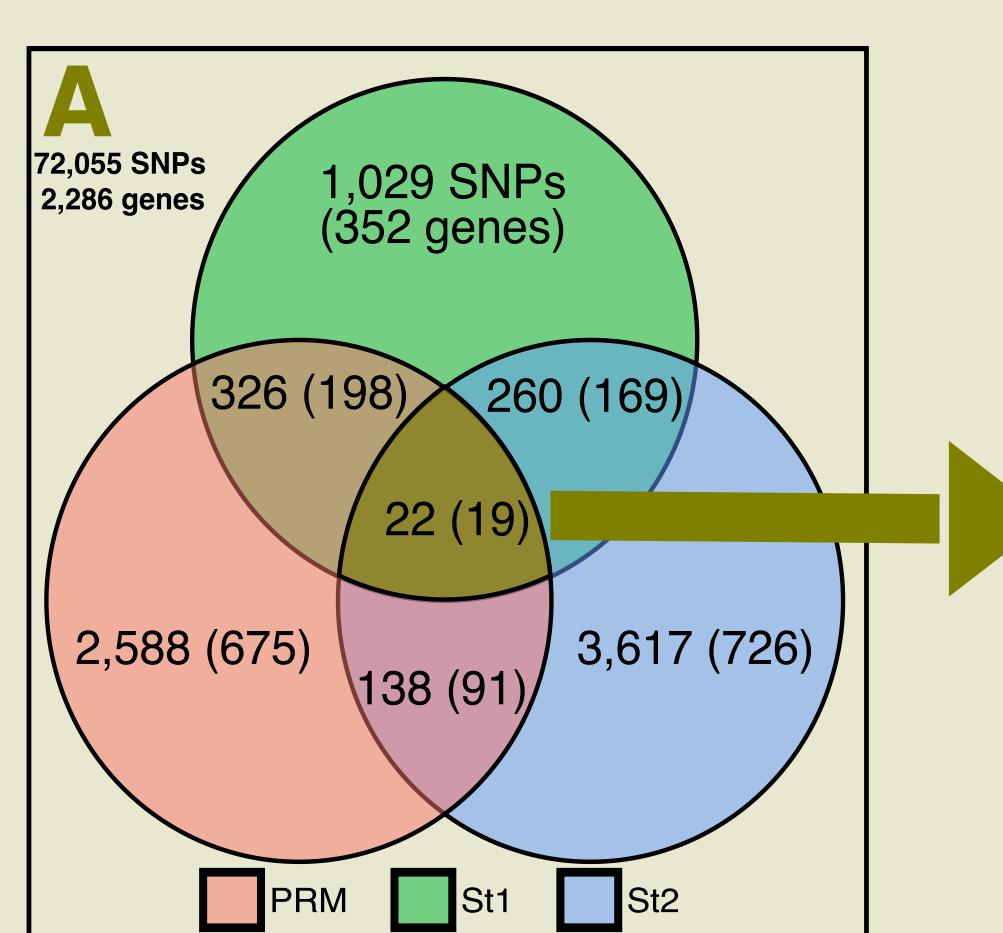
[Chlorophyll] peaked in mid-September resulting in a cyanobacterial bloom



## Taxonomy

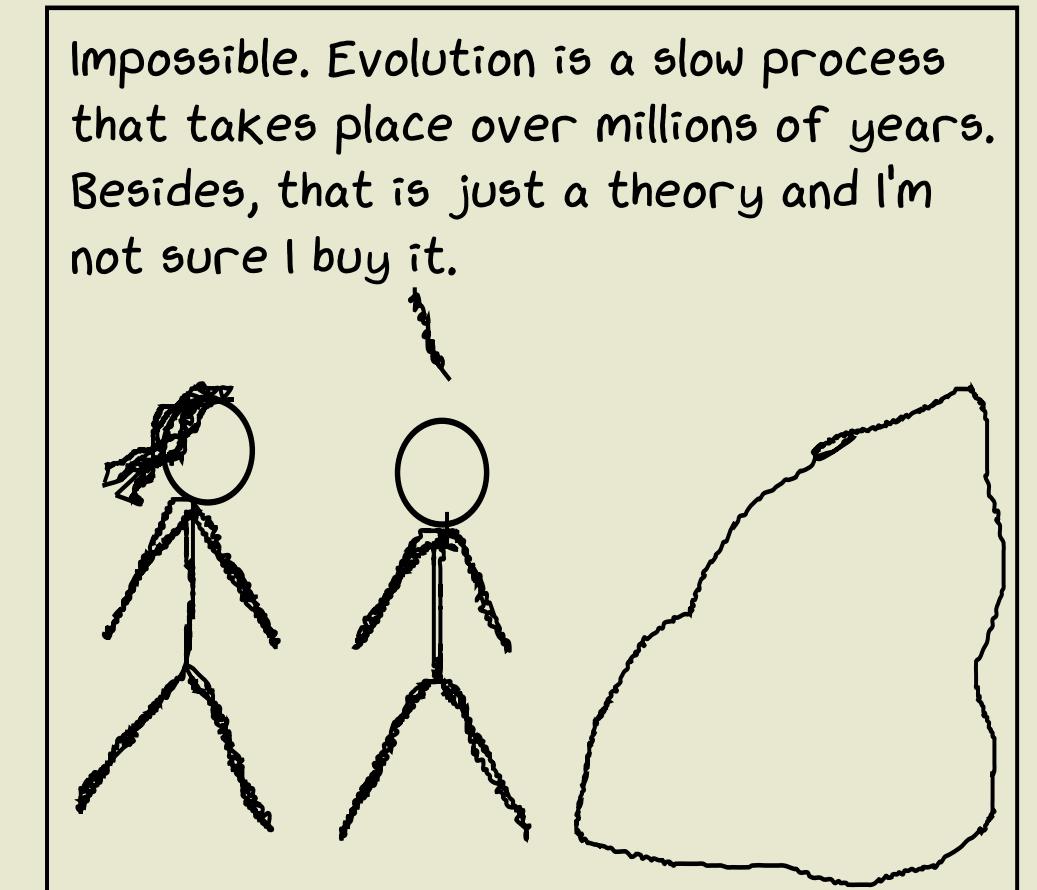
<i>Acinetobacter</i>
<i>Anabaena</i>
<i>Bacillus</i>
<i>Bacteria</i>
<i>Blautia</i>
<i>Candidatus</i>
<i>Chlamydia</i>
<b>Dolichospermum</b>
<i>Elizabethkingia</i>
<i>Fusobacterium</i>
<i>Inediibacterium</i>
<b>Microcystis</b>
<i>Salmonella</i>
<i>Staphylococcus</i>
<i>Streptococcus</i>
<i>Synechococcus</i>

Communities dominated by Gram negative *Fusobacterium* & *Acinetobacter* prior to bloom. During bloom, the cyanobacteria *Dolichospermum* rose in frequency, followed to a lesser extent by *Microcystis* 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. 22 SNPs showed significant changes in allele frequency in all three sites compared to three expected ( $p\text{-value} < 0.0001$ ). C. Example of a SNP showing a change from C to T allele in both St1, St2 (logistic regression) and PRM (linear regression). 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.

• To do: Position SNPs on genome. More sampling.

## 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. Bioinformatics 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