

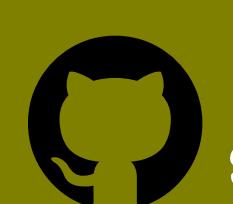
# Dolichospermum evolution during cyanobacterial bloom: insights from metatranscriptomics

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, QC, Canada <sup>2</sup>Département des Sciences Biologiques, Université de Montréal. <sup>3</sup>Energy Mining and Environment, National Research Council Canada, Montreal.

sebastien.renaud@umontreal.ca

@seb\_renaud



seb951.github.io/atrapp/RENAUT\_evol2019.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 algal 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

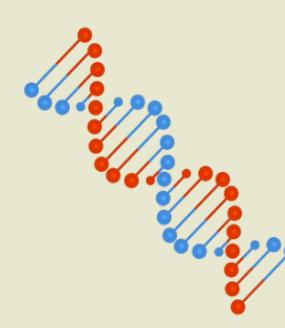
- Use a metatranscriptomics approach that targets genes expressed in the water to identify changes in the bacterial community during the summer.
- Identify cyanobacteria genes that change in allele (SNPs) frequency during the summer to help predict future blooms.

## Fancy tools and analyses

1. Water sampled in Lake Champlain (3 locations X 8 time points X triplicates).



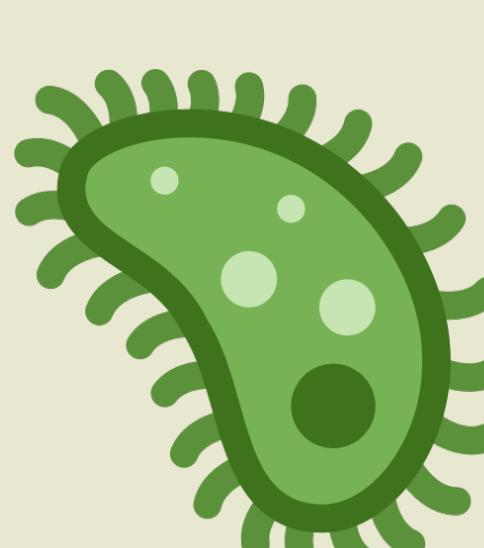
2. RNA extracted, High throughput sequencing (100bp paired end Illumina HiSeq).



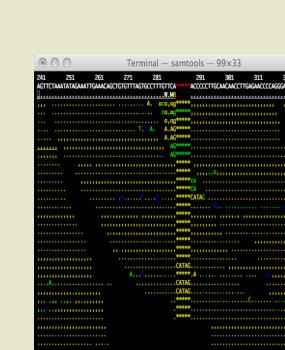
3. Metatranscriptomes processed using SAMSA2 pipeline (trimming, merging, removal of rRNA, annotation).



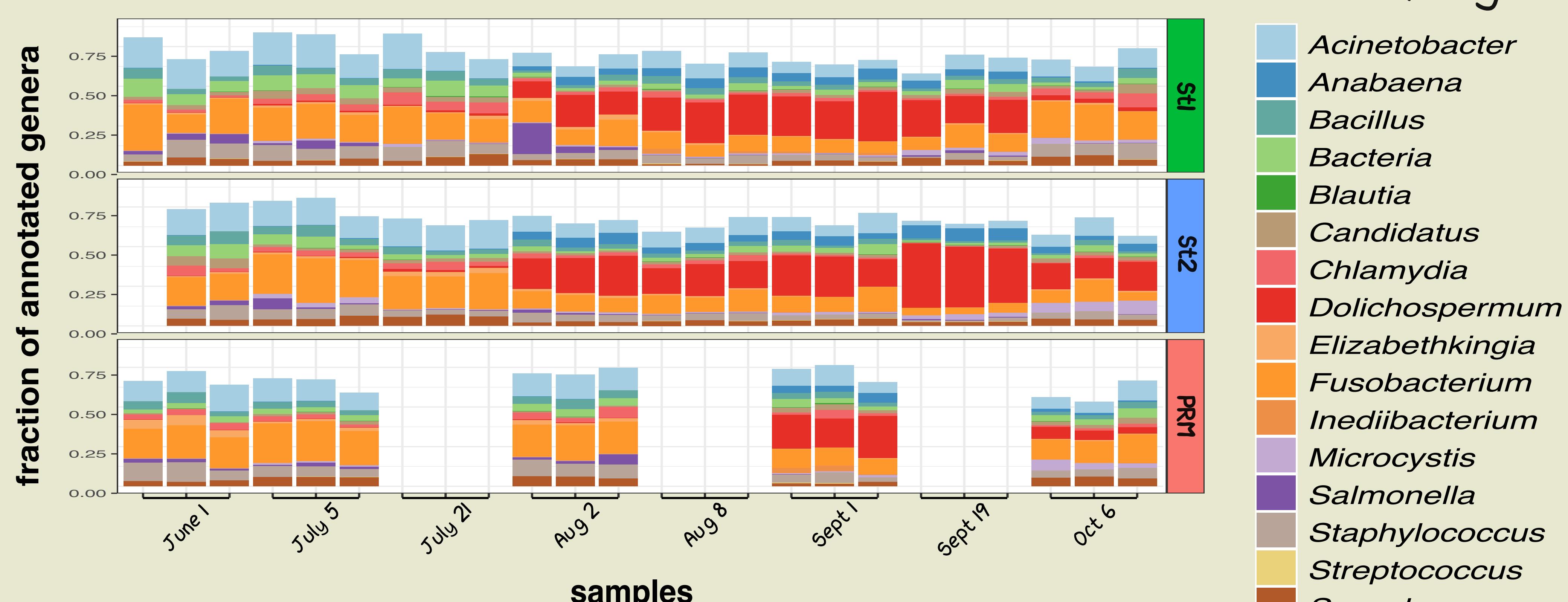
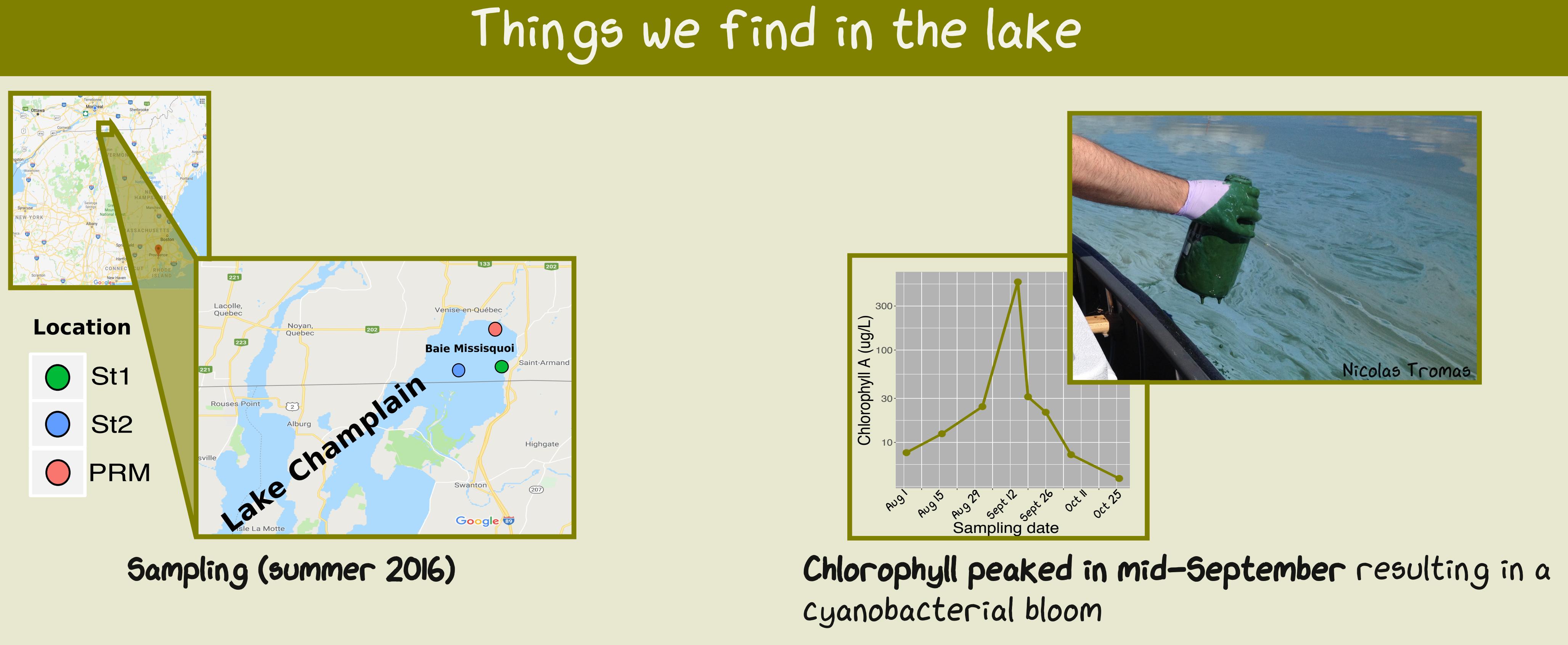
4. Sequences annotated to Dolichospermum (the main cyanobacteria responsible for the bloom) used to produce a de novo reference transcriptome using Trinity.<sup>2</sup>



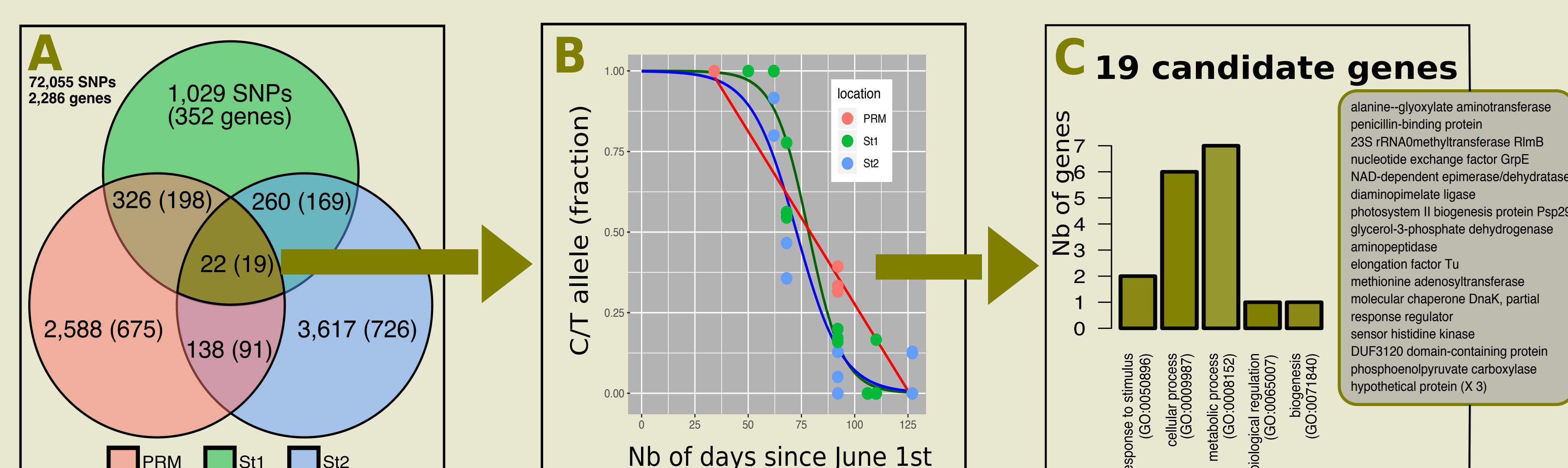
5. Dolichospermum sequences aligned to transcriptome using bowtie2<sup>3</sup>. SNP called using SAMtools<sup>4</sup>.



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

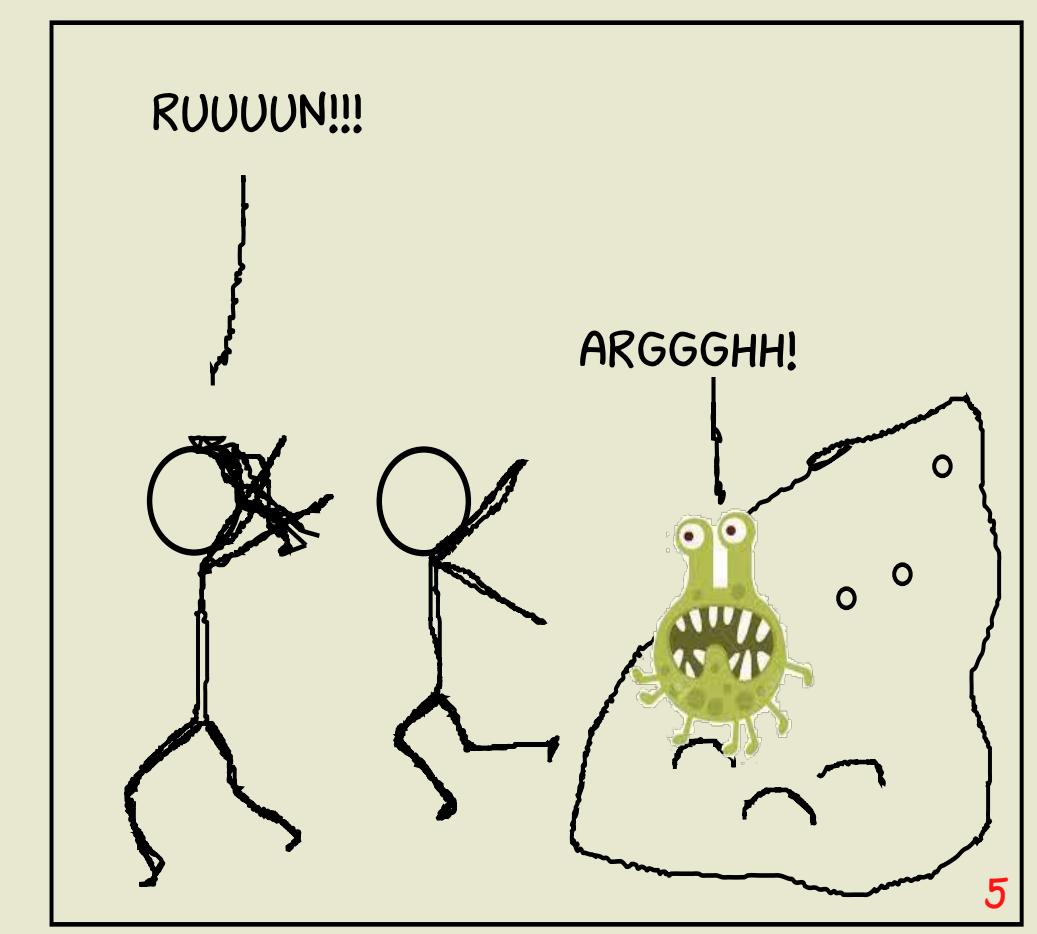
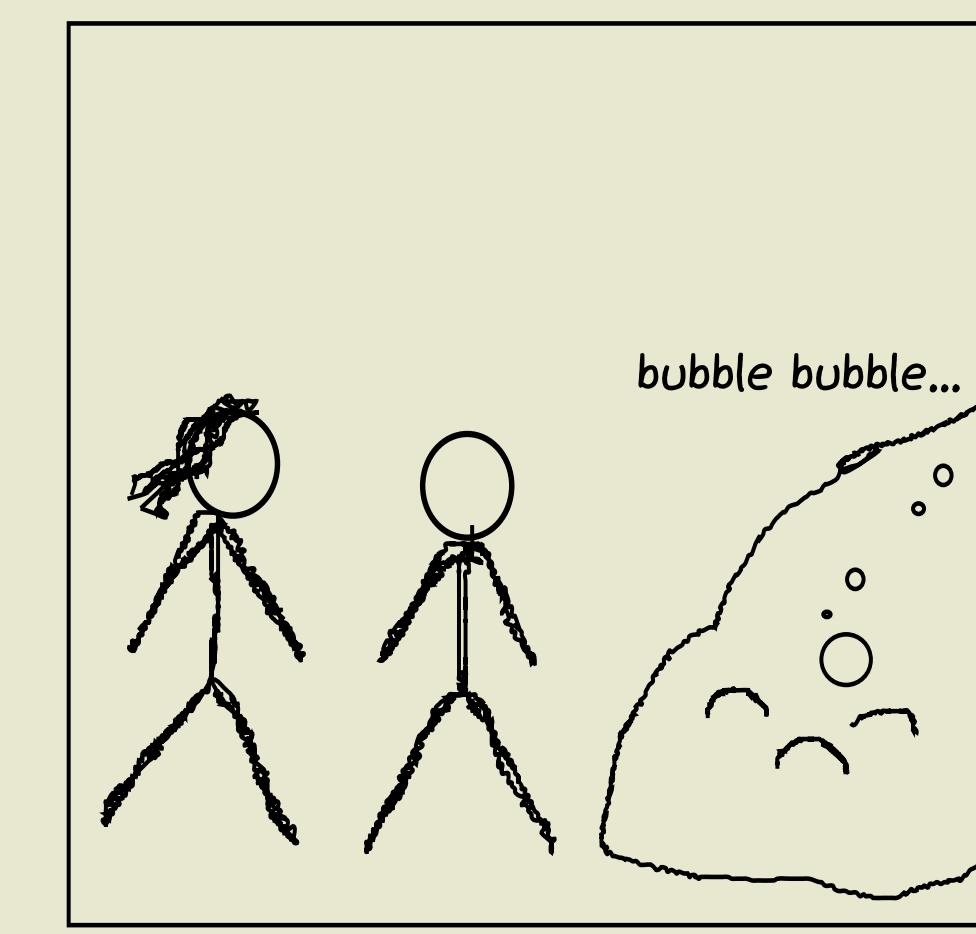
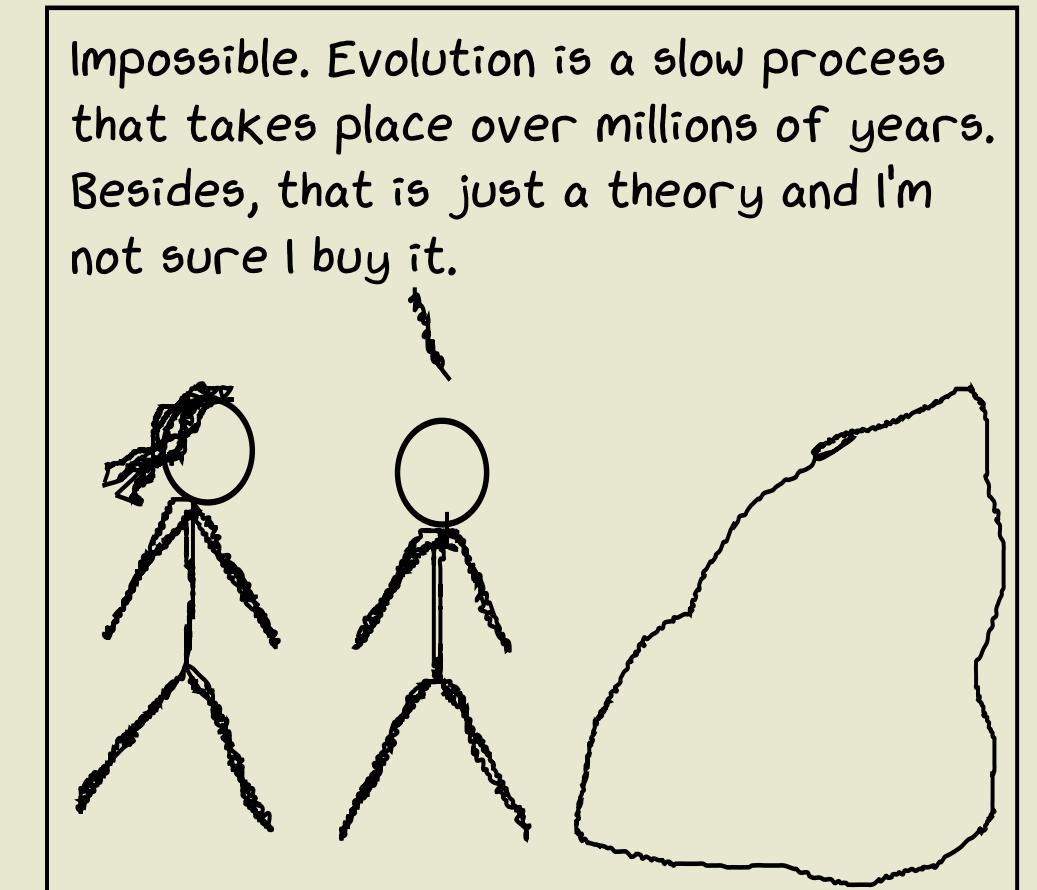
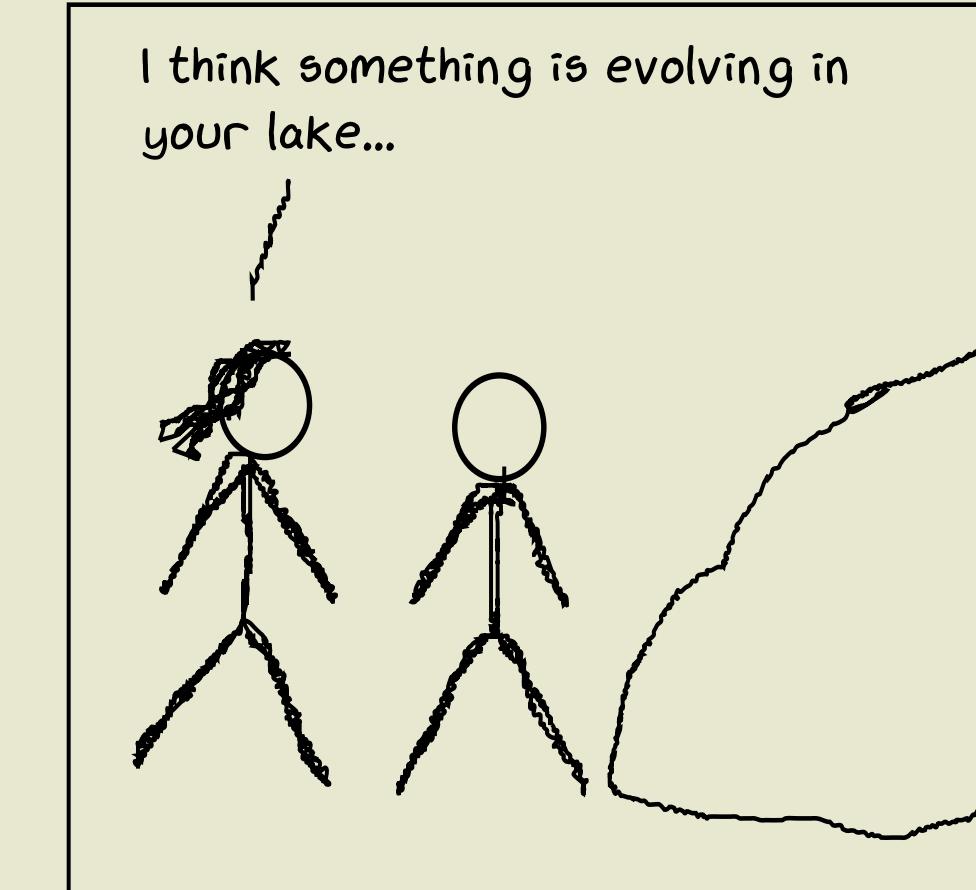


Communities dominated by Gram negative *Fusobacterium* & *Acinetobacter* prior to bloom. During bloom, the cyanobacteria *Dolichospermum* rose in frequency, followed to a lesser extent by *Microcystis* in September.



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). 22 SNPs changed in allele frequency in all three sites (compared to three expected, p-value < 0.0001). B. Example of a SNP showing a change from C to T in all 3 sites during bloom. C. 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 sites.
- Nineteen genes showed repeatable changes in allele frequencies in all sampling sites. These represent candidate genes which may help to predict future blooms.
- To do: Position SNPs on genome. More sampling to confirm candidates.

## Papers we 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)

[5] Poster font and cartoon were derived from the xkcd.com webcomic

## People paying for the science

