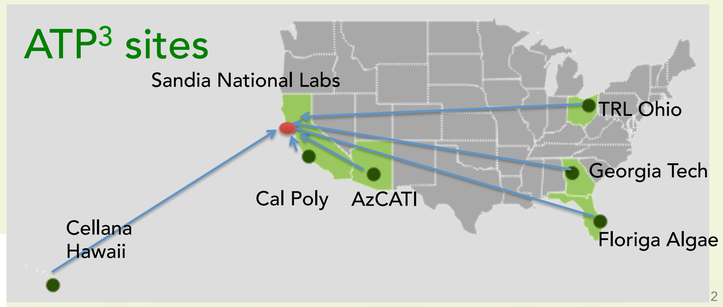
SAND2015-4942C

Genetic Evaluation of Pond Crashes during ATP³ Unified Field Study

Kunal Poorey, Deanna J. Curtis, Haifeng Geng

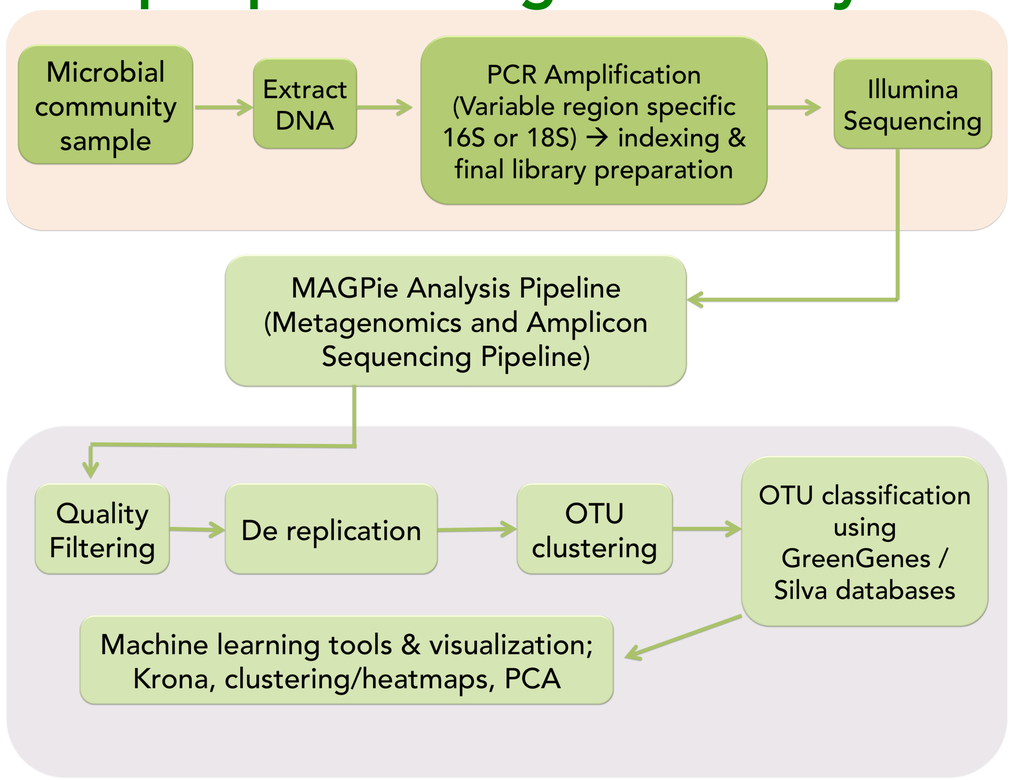
Todd W. Lane, Sandia National Labs



Abstract

The Algae Testbed Public Private Partnership, ATP³, is a Department of Energy supported consortium funded to generate high impact data in order to quantify the seasonal and geographic variations on the production of microalgae biomass for use in production of biofuels and bio-products. We use metagenomic amplicon sequencing to identify deleterious species that contribute to pond instability and consequent reduction in annualized production. Such diagnosis of the root causes of pond crashes is critical to for development of inexpensive screening and monitoring tools for early crash detection, as well as engineering countermeasures. For this Unified Field Studies, cultivation trials, two strains, *Nannochloropsis oceanica* and *Chlorella vulgaris*, were grown over four seasons: Spring, Summer, Fall and Winter at geographically distinct consortium member sites. Routine samples were preserved and archived onsite for genetic analysis in event of pond crash. At Sandia National Labs microbiome analysis by next generation DNA sequencing was carried out to provide a presumptive identification of the biological agent(s) responsible for the pond crash. Fragments of the small subunit ribosomal RNA gene were amplified by PCR and sequenced. Though analysis of the sequencing data with novel Metagenomics Pipeline (MAGPie) combined with data-mining tools we identified bacterial and eukaryotic communities responsible for pond crashes.

Sample processing and Analysis Pipeline

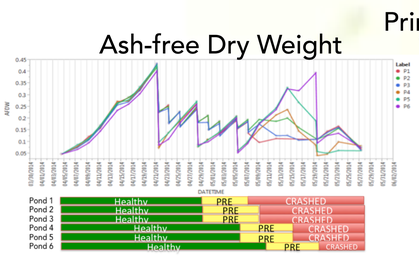


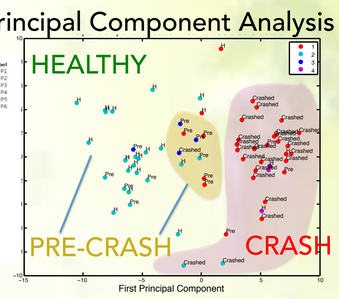
The biomass samples from the sites were sent to Sandia from the six ATP3 sites is processed to extract DNA followed by 2 steps of PCR amplification for a specific variable region in rRNA to amplify and then sequence. For these samples we amplify V3-V4 region of 16s rRNA to target prokaryotes and V4 region of 18s rRNA to target Eukaryotes.

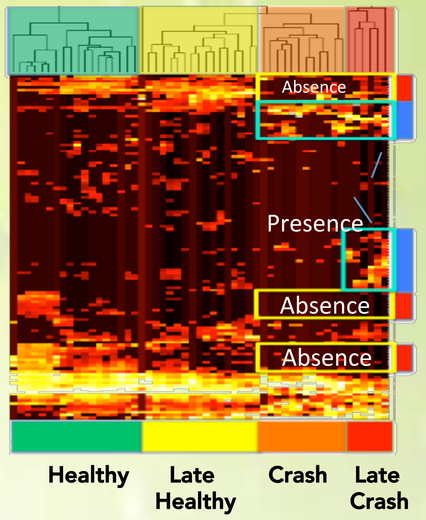
The sequence reads were analyzed by new software pipeline **MAGPie (MetAGenomics and Amplicon Sequencing PIpEline)** for OTU clustering and classification. We have integrated Usearch program for analysis and has significantly faster than other traditional softwares (Qiime). Further we use machine learning tools to study and identify contaminants or deleterious species on our datasets

TRL Spring 16s analysis

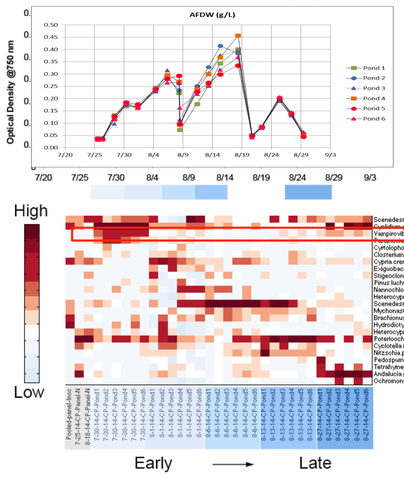
For site at TRL we did an in-depth analysis for its spring run to study a unique crash event. We focused on 16s prokaryotic community structures to see the smoking gun of the crash by clustering MAGPie output both w.r.t the samples and bacterial taxa. 4 major clusters were observed for healthy, pre-crash, crash, late crash samples. And similar results were obtained by principal component analysis







CalPoly Summer



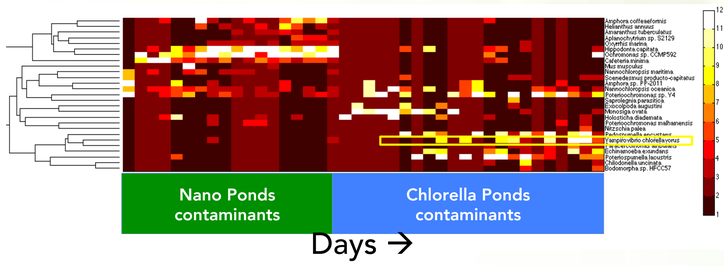
The figure on the left shows OD measures in the samples. The bands below going from white to blue in gradient shows the time course of the sample collection. The heat map shows eukaryotic species clustering on the y axis to show the emergence of deleterious species in the ponds.

Microscopy observations at the sites confirmed/ identified the contamination as: ciliates (*Cyclidium glaucoma*), chrysophyte (*Poterioochromonas sp*) , flagellate (*Andalucia godoyi*). *Scenedesmus*, a few diatoms and **Vampirovibrio chlorellavorus**

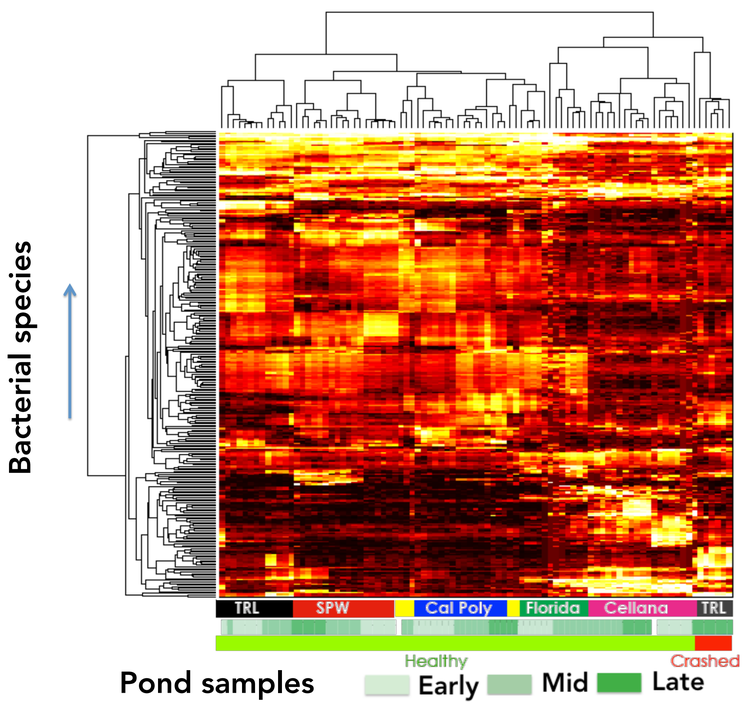
AzCATI Fall

Similar to previous examples we show a heat map of contaminants for the Fall season in which both Nano and Chlorella were grown

Various contaminants found in these ponds were: :Flagellate: *Cafeteria minima, Monosiga ovata, Bodomorpha*, Ciliates: *Chilondella ucinata, Holosticha diademata, Exocolpoda augustini*, dinoflagellate: *Oxyrrhis marina*, diatoms: *Hippodonta capitata, Amphora coffeaformis, Chrysophysceae: Pedospumella encystan*, fungus:*Saprolegnia parasitica*, along with **Vampirovibrio chlorellavorus**



Results: 16s analysis for spring season



The figure on the left shows a heat map of bacterial community structure for spring season for all the 6 sites. The data was clustered both with respect to column and row using a correlation coefficient (below) and also by rows

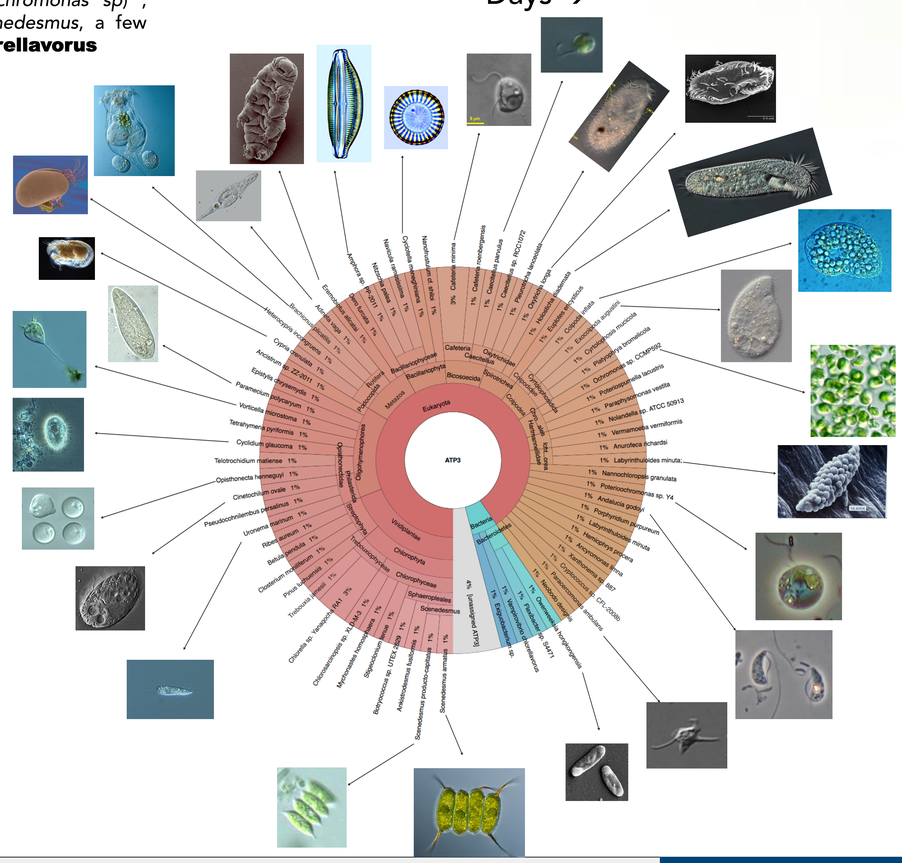
$$d_{st} = 1 - \frac{(x_s - \bar{x}_s)(x_t - \bar{x}_t)'}{\sqrt{(x_s - \bar{x}_s)(x_s - \bar{x}_s)'}\sqrt{(x_t - \bar{x}_t)(x_t - \bar{x}_t)'}}$$

The first set of rectangular bands at the bottom of the heat map shows pond location. It is clear that each APT³ site has its own signature microbial community. One other key observation can be drawn from the heat map and the column dendogram that the microbial community structure of site in Cellana is quite distant to all the other sites.
The second set of bands are colored with respect to the date the samples was taken. It was observed that the microbial community structure can be clustered according to early, mid and late cultivation days. This was replicated for all 6 ponds in the site. And last the pond sample corresponding to a crash had a unique and significantly different community structure than normals and we were able to detect the change

Summary

On the right we show all the eukaryotic contaminants identified in the pond crashes by sequencing and computational pipelines. They range from diatomes. Cilliates, bacterias, rotifers, amoebas, and competing algae species. These results from sequencing analysis did match the microscopy observations at sites.

The advantages of using NGS methods is that we get a rough estimate of species abundance and making observation is not dependent on any *a priori* knowledge. Although several potential sources of bias can distort abundances and presence does not imply causality.



Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.