



Title: Amplicon Sequencing of ATP³ Pond Samples for Pond Crash Forensics

**Kunal Poorey & Todd W. Lane,
Sandia National Labs**

Sandia National Laboratories is a multi-program laboratory 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.

Introduction

Algaculture usage

- Food
 - Spirulina, Chlorella, Irish moss, Sea lettuce etc.
 - Algae oil (Omega-3 and Omega-6)
- Fertilizer and agar
- Pollution control
 - Waste water treatment, CO₂ reduction
- Energy production
 - biodiesel, biobutanol, biogasoline, methane, ethanol
- Other usage
 - Cosmetics animal feed, bioplastic, pharmaceutical ingredients
- Job Creation



Technical Challenges

For sustainable reliable production / cultivation of algae.

- Track the development and progression of microbial communities for targeted algal strains cultivation and using culture-based methods and deep sequencing of small-subunit ribosomal RNA genes.
- Find probiotic communities to favor yield, harvesting and also finding deleterious species causing pond crashes
- Develop methods to detect deleterious species causing pond crashes

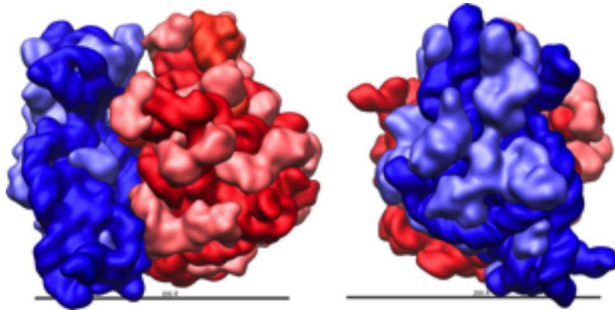
Microscopy methods

- Current methods: Microscopy and qPCR sampling
 - Disadvantages
 - Prior expertise needed to identify organisms (Micro)
 - Sampling bias in identifications (some predators are easy to spot 😊)
- Exploiting the diversity in ssu rRNA gene we sequence DNA targeting rRNA to classify the species present in algae ponds.
 - Advantages:
 - No prior knowledge needed to classify.
 - No sample based bias
 - time saver
 - Able to classify a larger variety of organisms



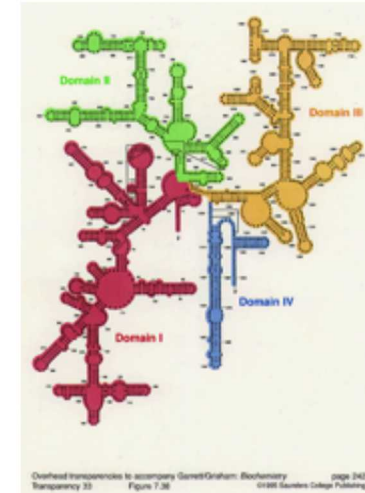
Small subunit ribosomal RNA (SSU rRNA) gene is used for phylogenetic analysis

Ribosome structure

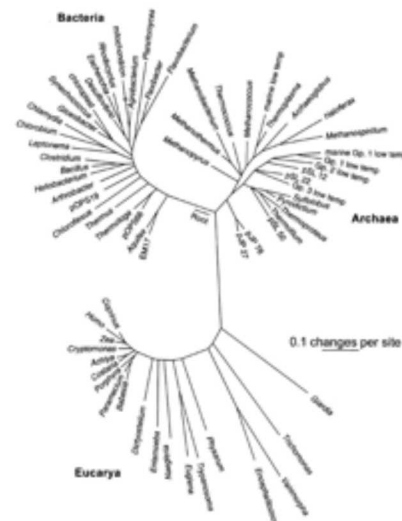


Large Subunit, Small Subunit

SSU rRNA



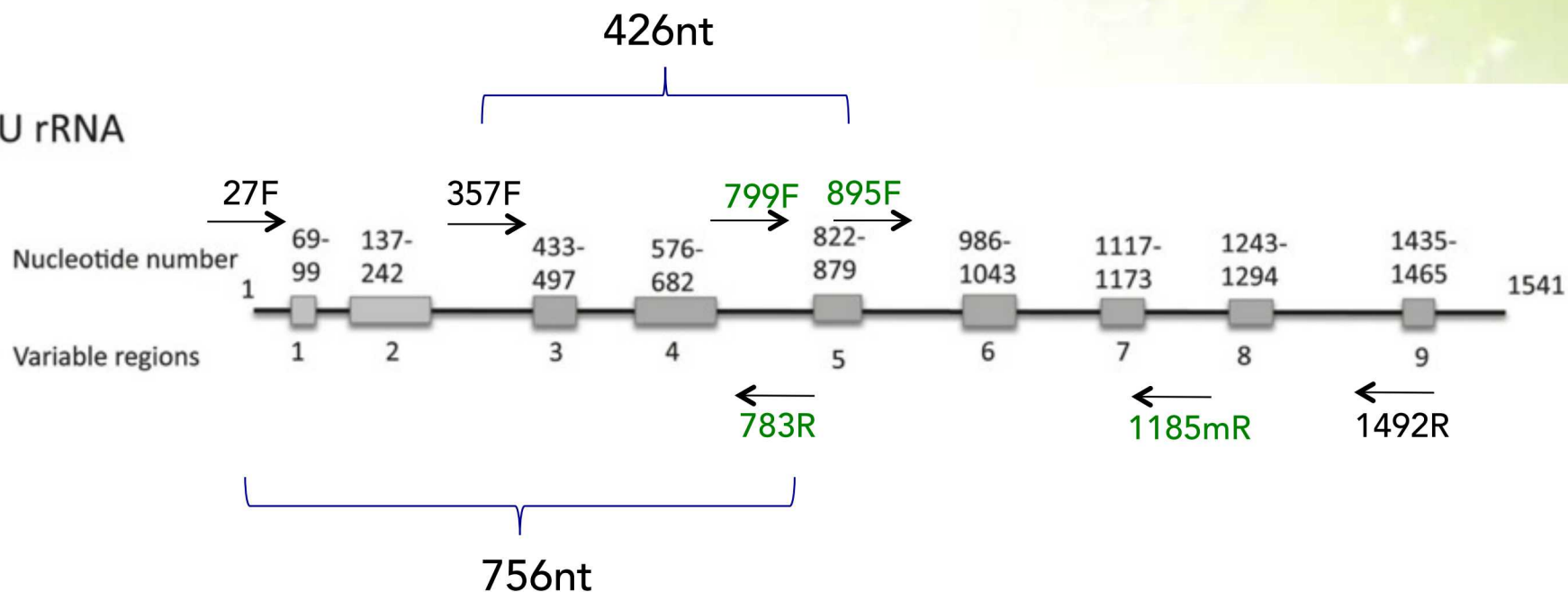
Phylogenetic “Tree of Life”
based on SSU rRNA
sequences



SSU rRNA gene sequencing strategy

600nt kit (current)

SSU rRNA



800nt kit (future)

Microbial community analysis



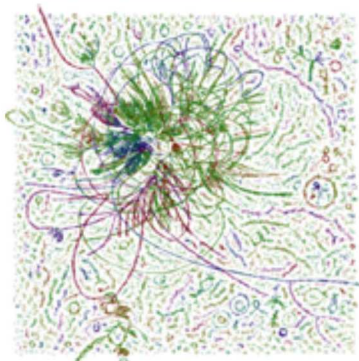
Pond



Sequencer



Microbiome

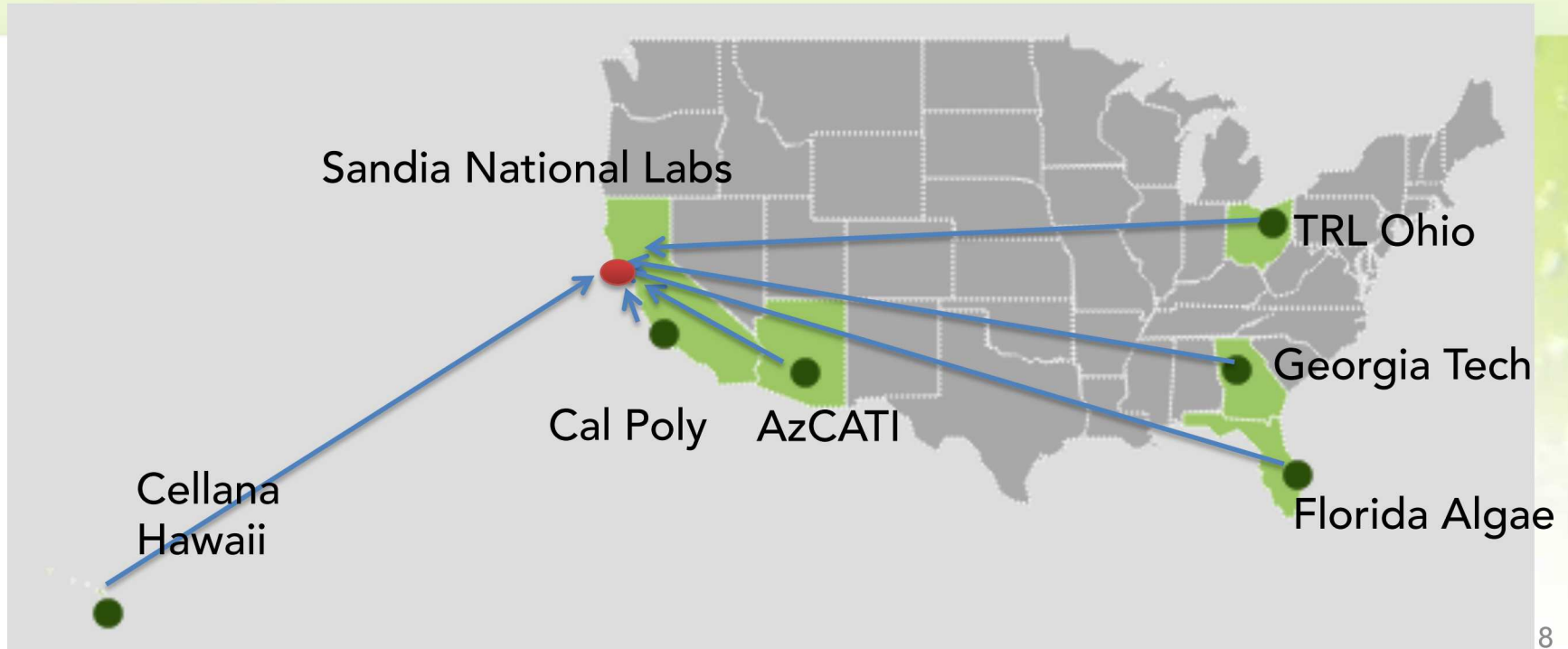


V Iverson/U of Washington

From “Pond Crash Forensics” and ATP³, we have developed methods for all the steps in the process

- Sample acquisition
 - Stabilization for shipment
- Nucleic acid extraction
 - Extraction bias
- Sequencing library prep
 - Plastid exclusion
 - Amplification bias
- Data analysis
- qPCR primer design

ATP³ sites



8



Sample acquisition

- Harvesting by centrifugation: 50 mL samples
- Resuspension in RNALater
- Storage at 4° overnight followed by -20°
- Shipment on wet ice.

16S/18S Amplicon sequencing and analysis workflow

Microbial
community
sample

Extract
DNA

PCR Amplification
(Variable region specific
16S or 18S) → indexing &
final library preparation

Illumina
Sequencing

MAGPie Analysis Pipeline
(Metagenomics and Amplicon
Sequencing Pipeline)

Quality
Filtering

De replication

OTU
clustering

OTU classification
using
GreenGenes /
Silva databases

Machine learning tools & visualization;
Krona, clustering/heatmaps, PCA

MAGPie

100%

→ Genus

70%

→ Species

Illumina

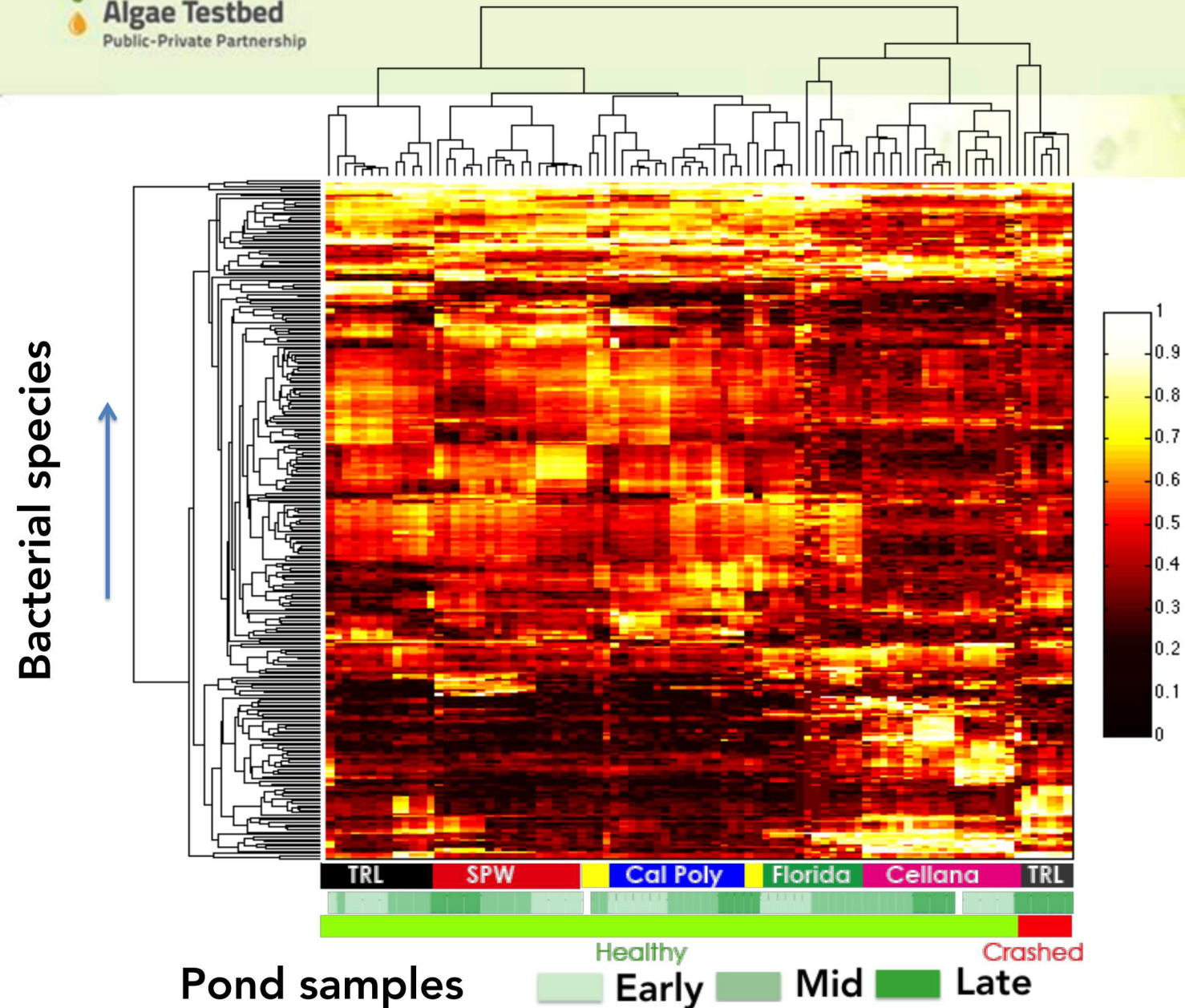
65%

→ Genus

30%

→ Species

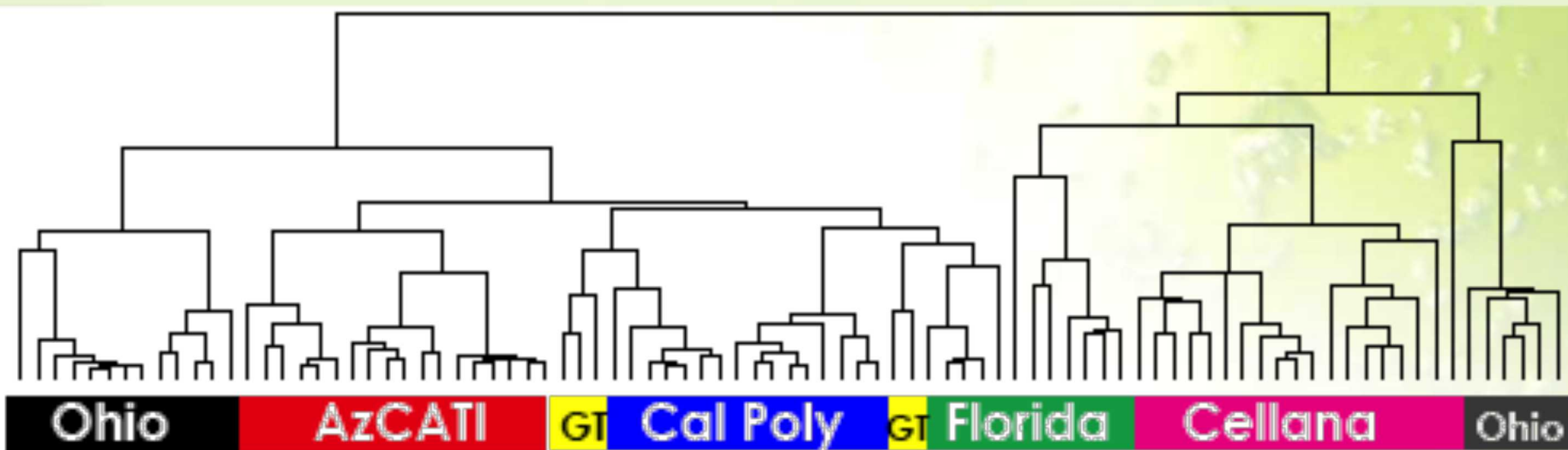
16S Analysis of spring samples



- Meta Summary for spring data from all sites
- Heatmap was made from clustering the classified OTU abundances for both species and samples.
- The Ponds from the same site cluster together

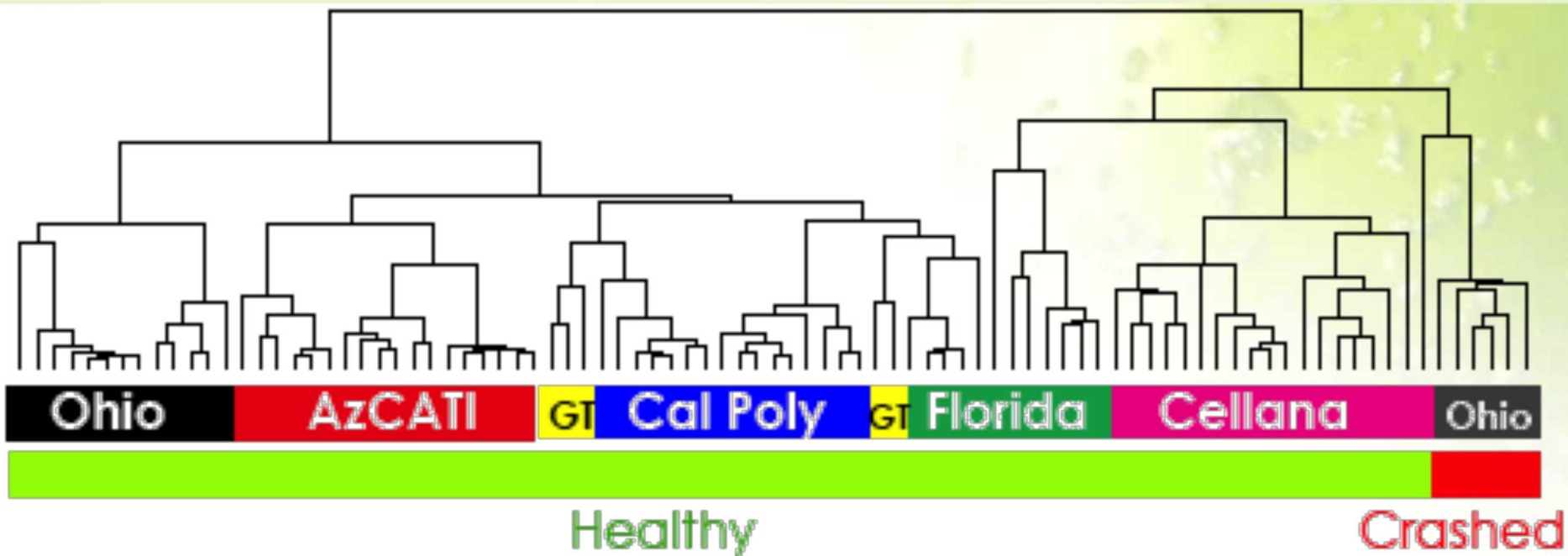


16S Analysis of spring samples



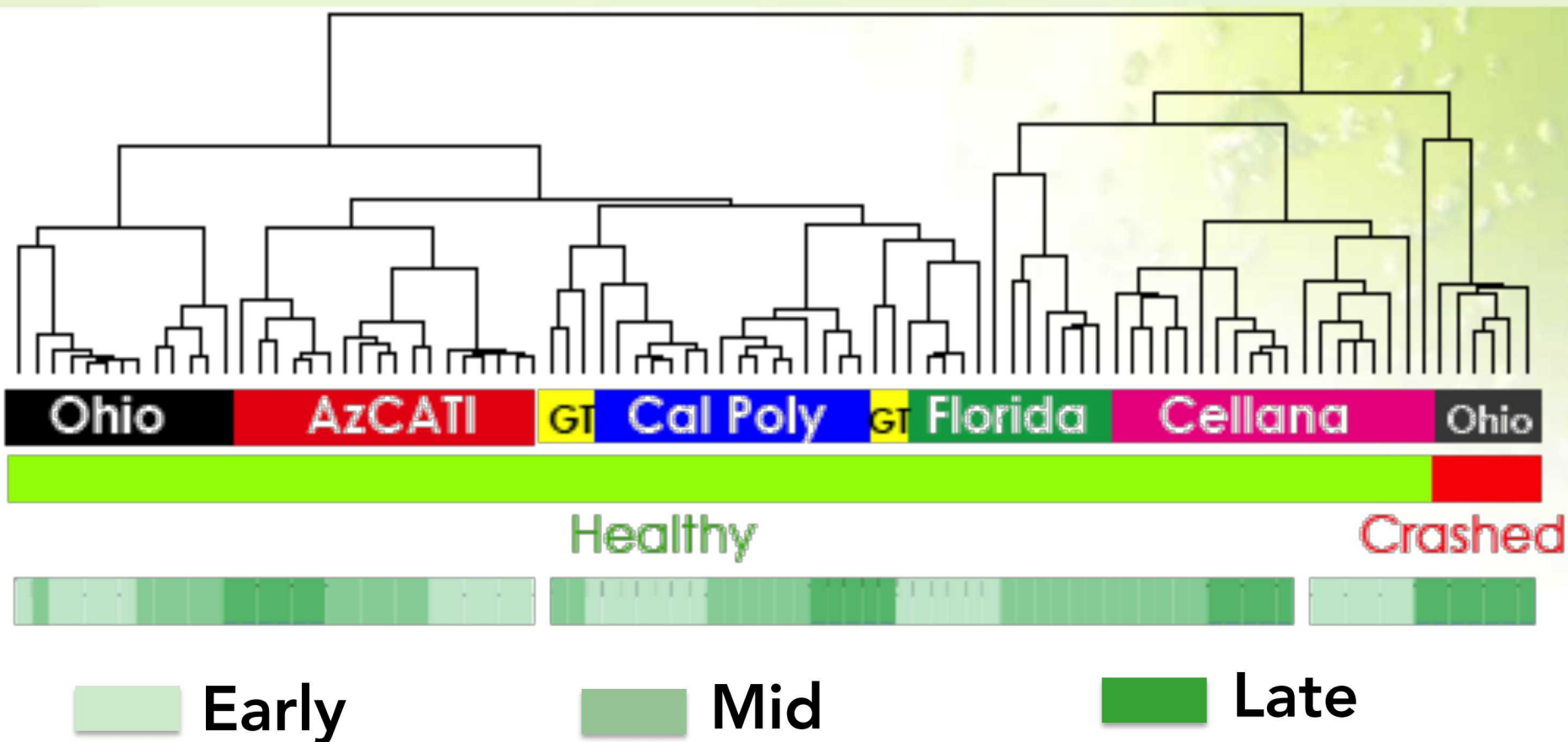
Bacterial communities can identify samples from the same site

16S Analysis of spring samples



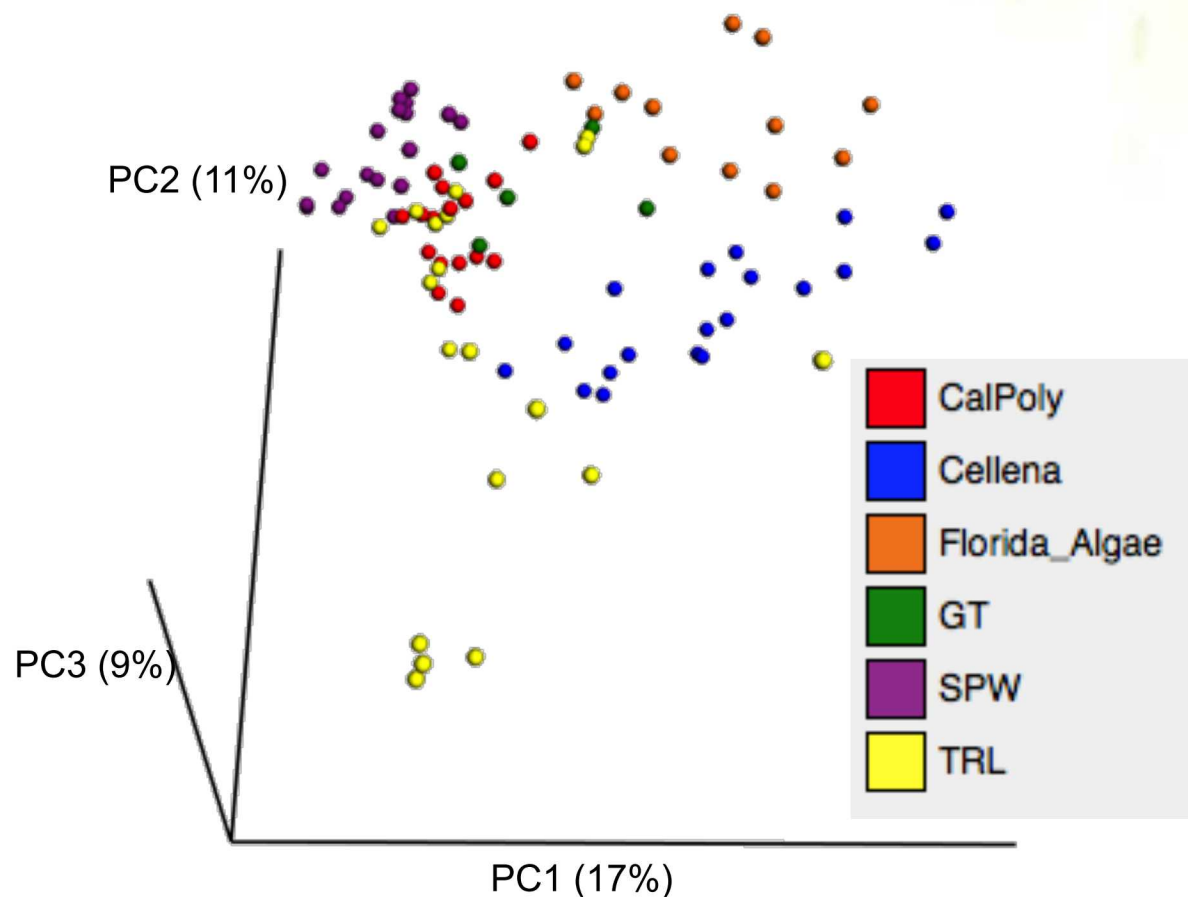
Crashed ponds had a different community structure than healthy ponds

16S Analysis of spring samples



And within same sites the community structure gradually changed with time

Principal Component Analysis (PCA)

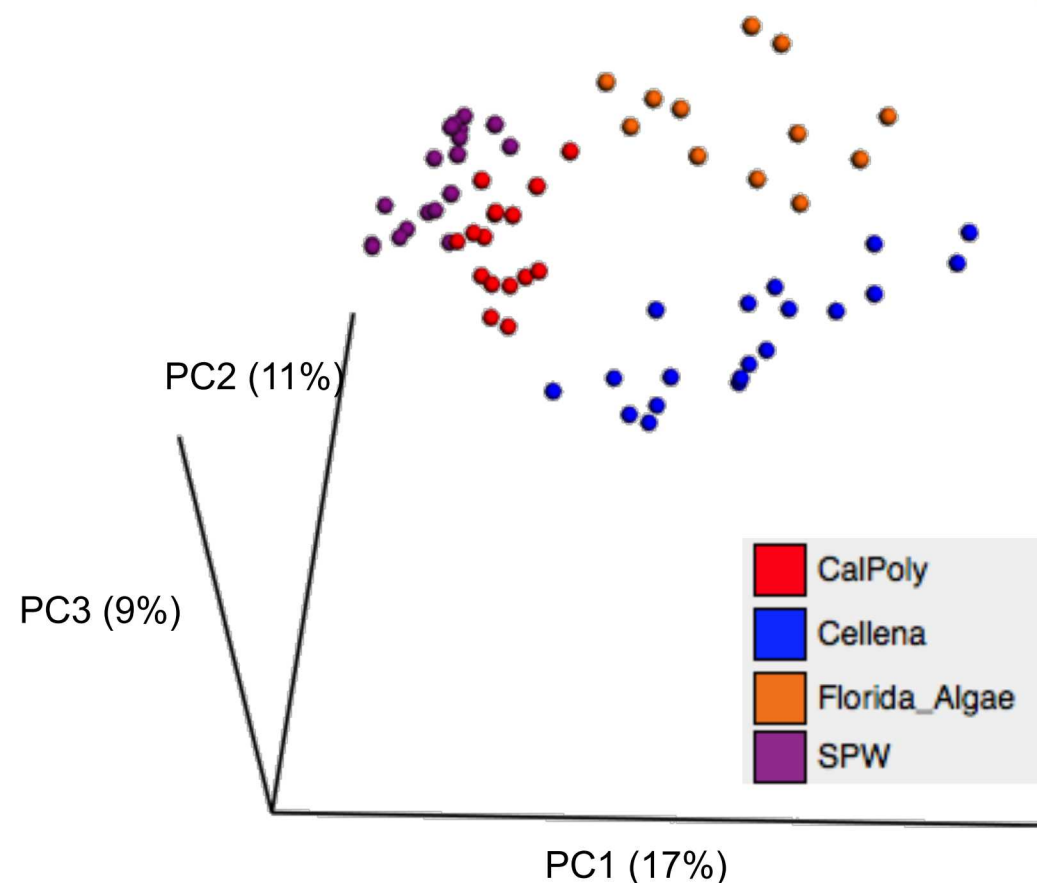


- PCA was carried out using the taxonomy structure of each pond.
- β -diversity index was used as a distance between ponds.
- Here we show 3-D scatter plot of first 3 principal components

Each site has its own signature bacterial communities and this concurs with our initial clustering analysis



Principal Component Analysis (PCA)

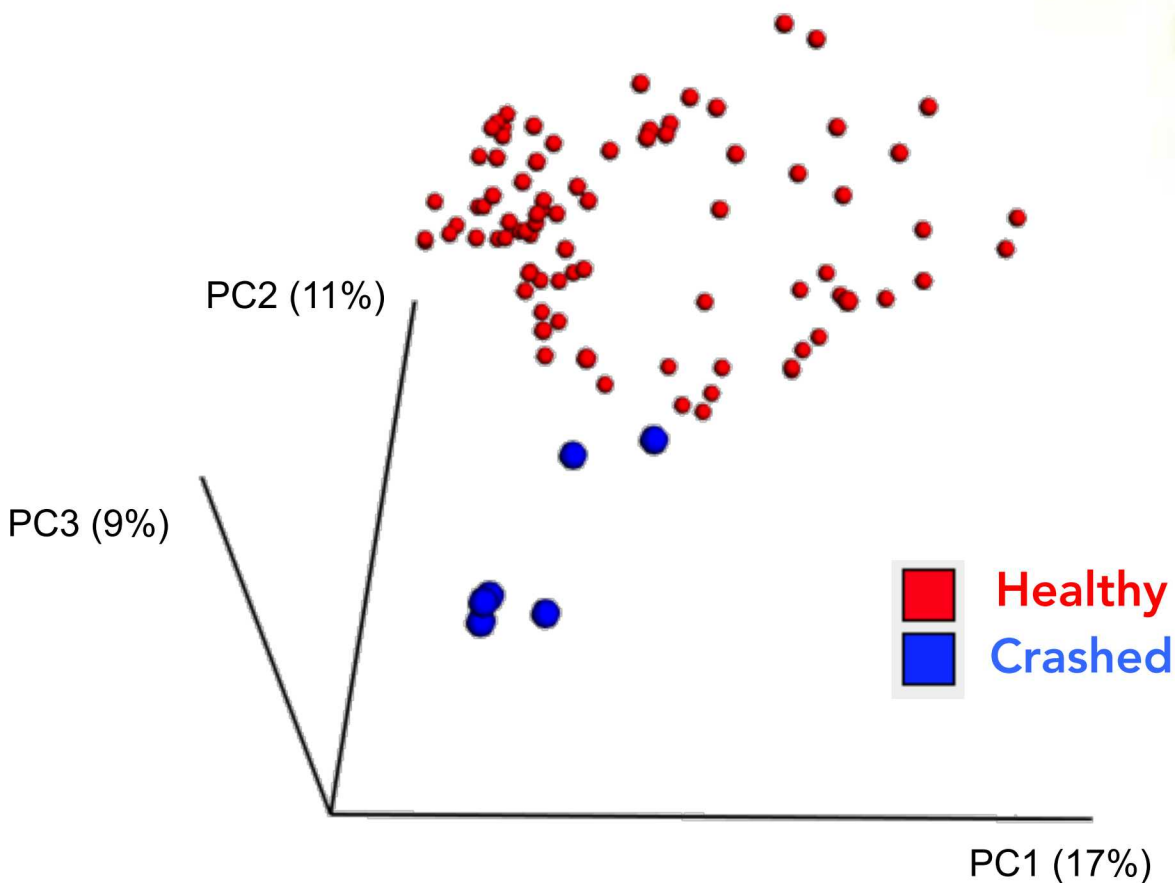


- PCA was carried out using the taxonomy structure of each pond.
- β -diversity index was used as a distance between ponds.
- Here we show 3-D scatter plot of first 3 principal components

Each site has its own signature bacterial communities and this concurs with our initial clustering analysis

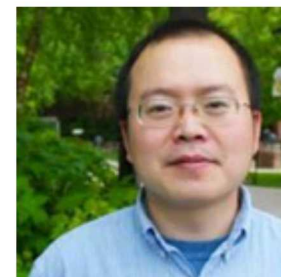


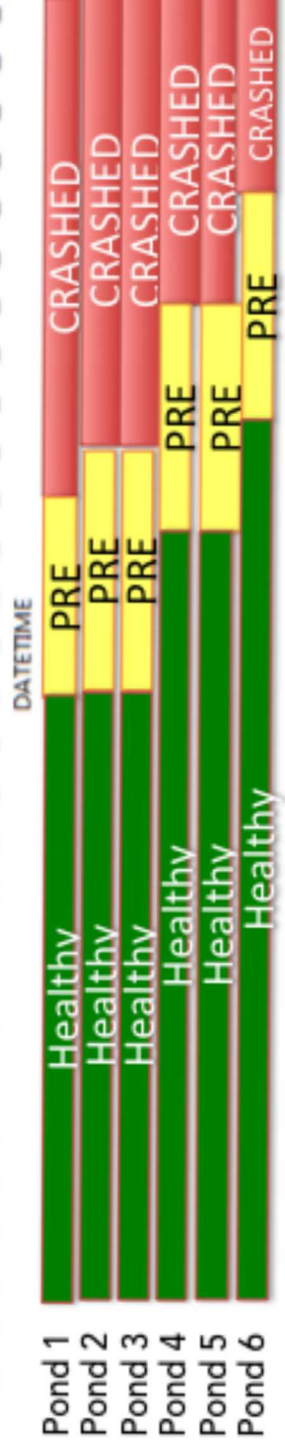
Principal Component Analysis (PCA)



- PCA was carried out using the taxonomy structure of each pond.
- β -diversity index was used as a distance between ponds.
- Here we show 3-D scatter plot of first 3 principal components

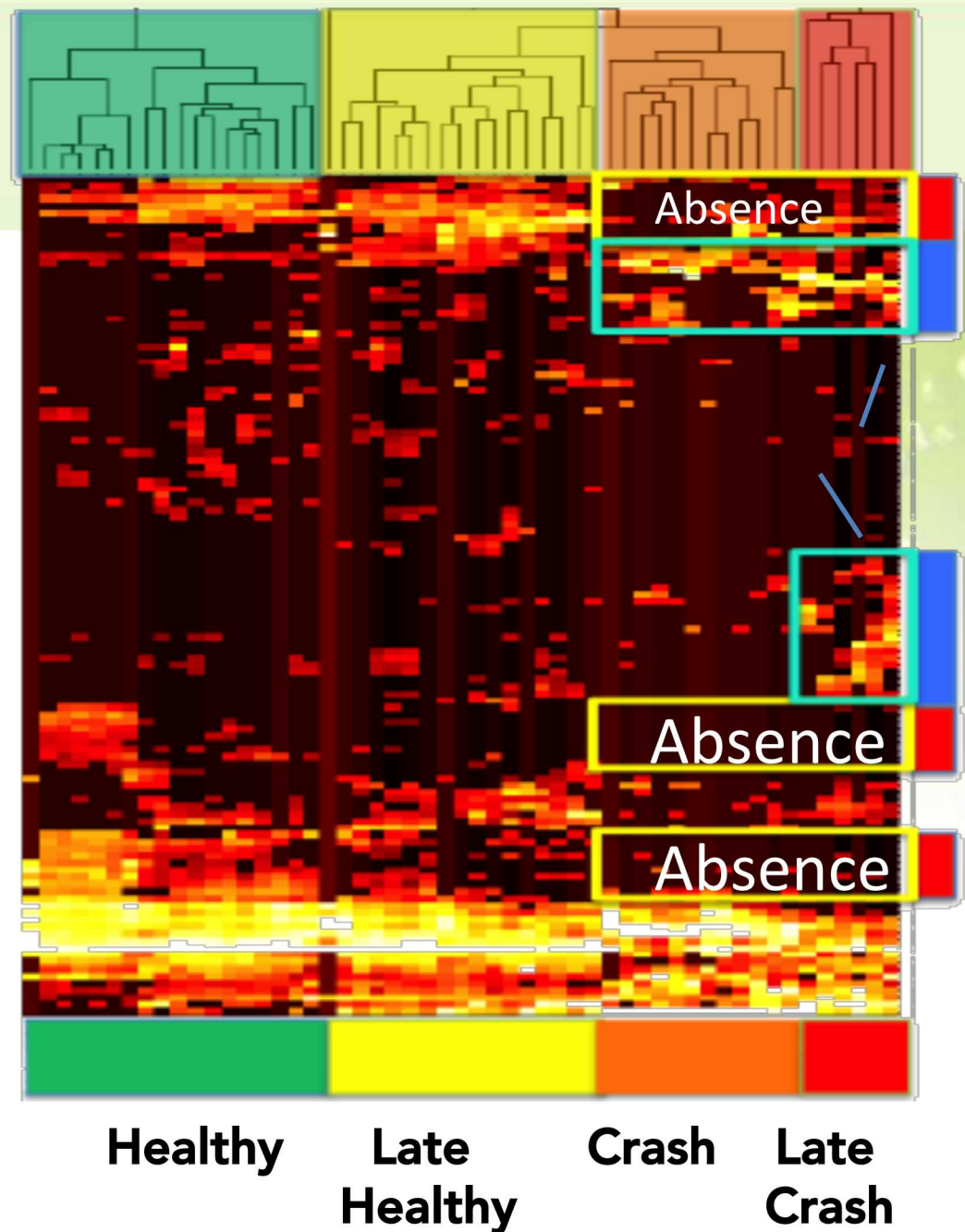
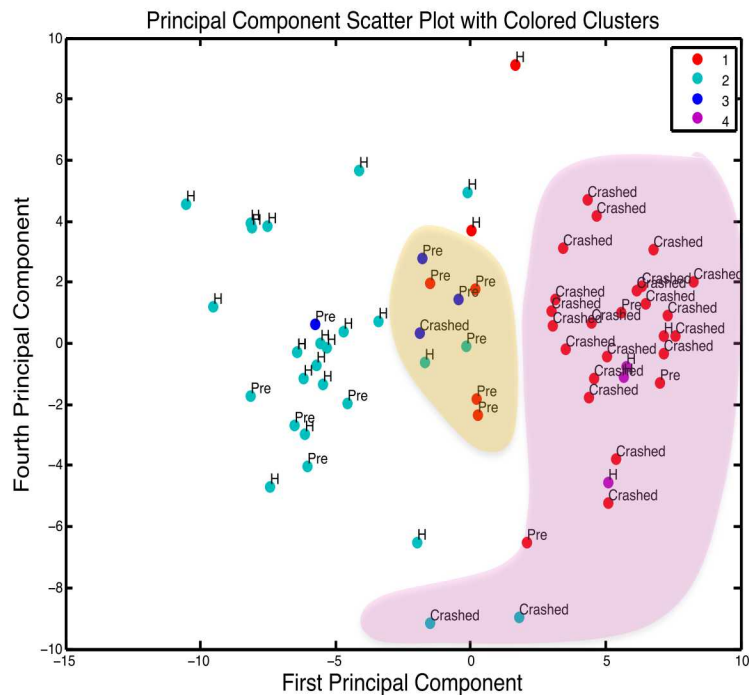
Crashed ponds have its own signature of microbiome





Clustering analysis on the ponds could differentiate the healthy and crashed ponds.

Principle component analysis also confirmed that crashed ponds have a different microbiome signature



all

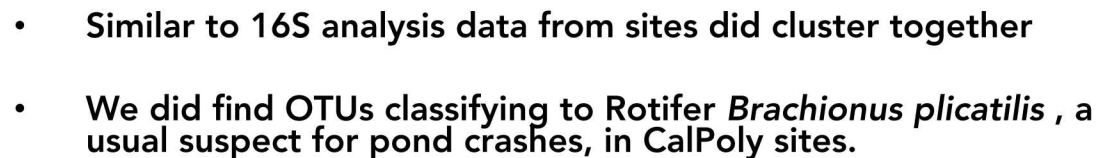
Alteromonas
Actinobacteria C111
Rhodobacteriace
Marvita
Rhylobacteriaceae
Devosia
Methylophaga
Stramenopiles
Rhizobiales
Alphaprotobacterias

absent in crash

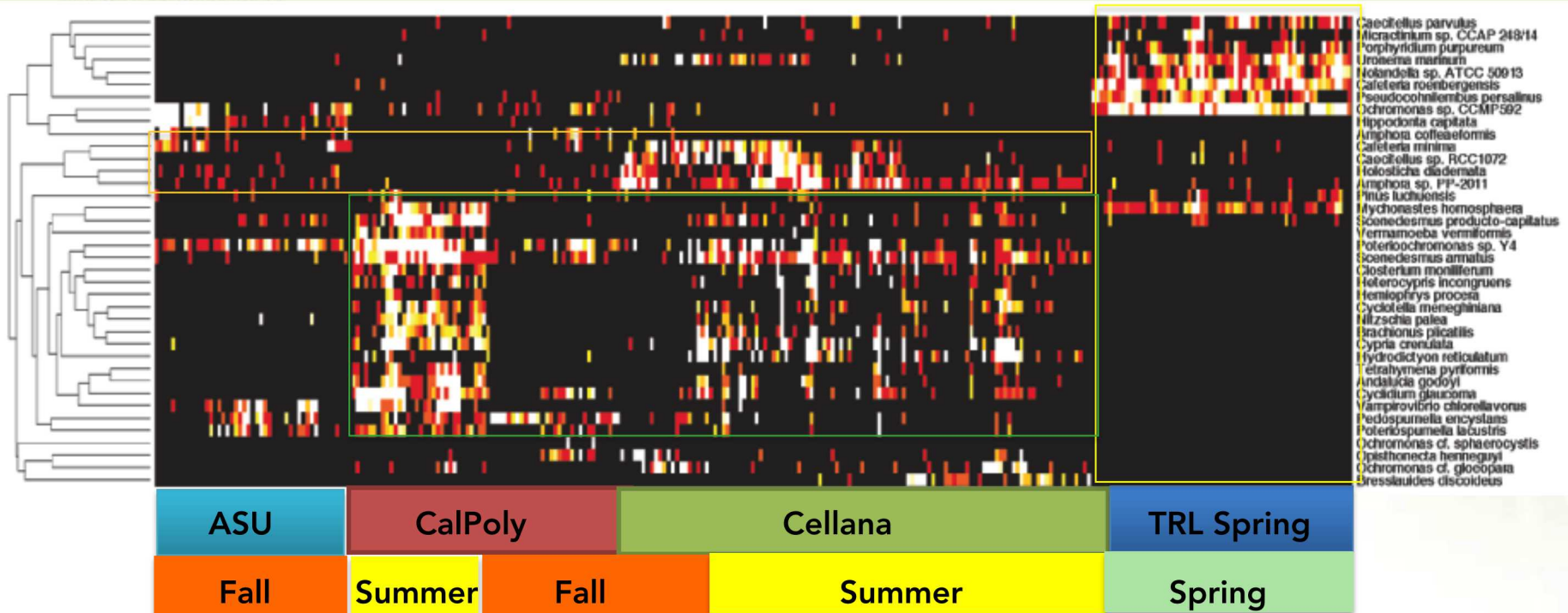
Bejerinckeaceae
Cryomorphaceae
Flammeovirgaceae
HTCC
Rhodobacteraceae
Pseudomonadaceae
Balneolaceae
Halomonas
Loktanella
Alcanivorax
Pseudomonas

present in crash

OM60
Pseudoalteromonas
Hyphomicrobacterium
Physisphaerales
Baleola
Maricaulis
Cynomorphaceae
Marinobacter



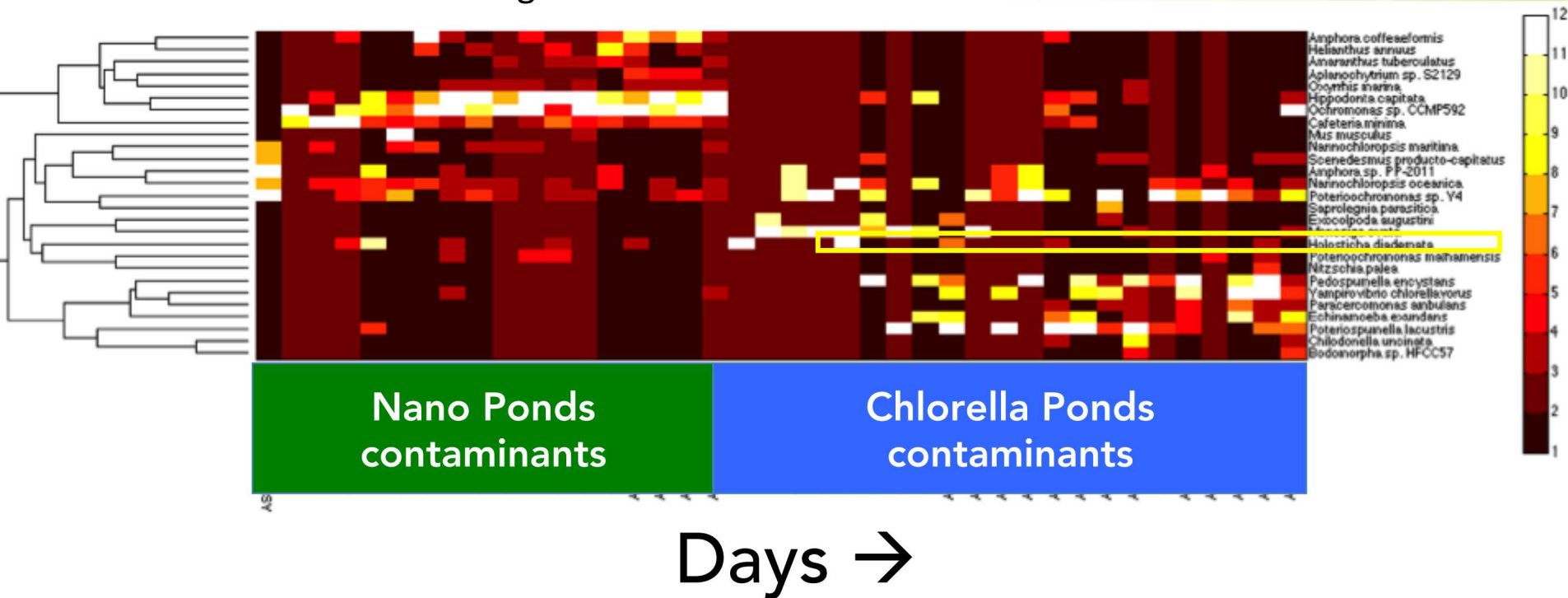
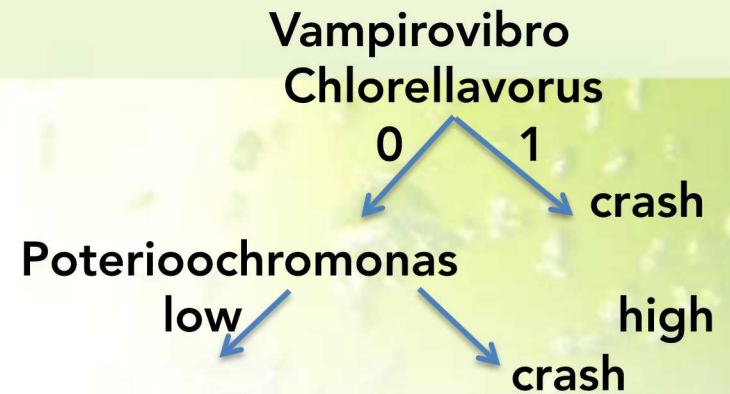
18S Analysis Meta analysis



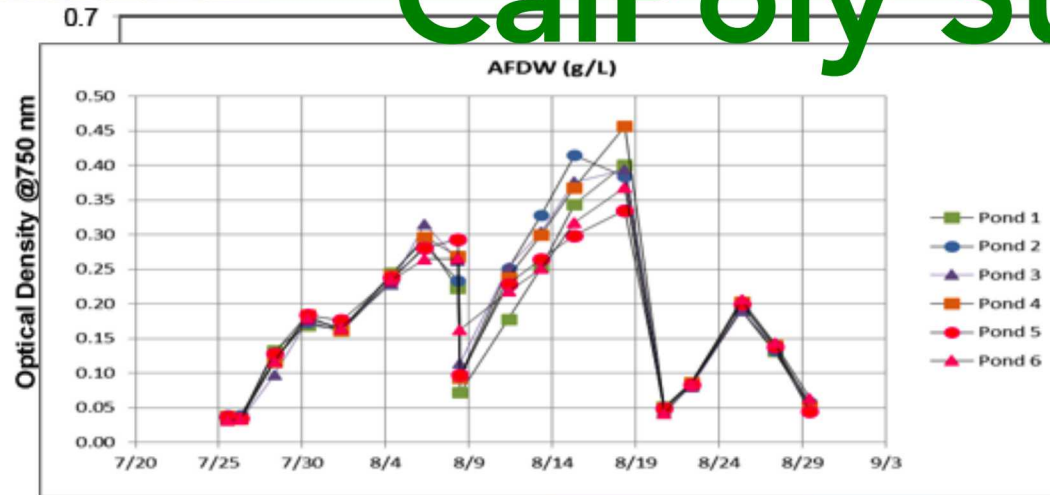
- Similar to 16S analysis data from sites did cluster together
- Some of the contaminants were common for sites (green box) and for season (orange box)
- We did find OTUs classifying to Rotifer *Brachionus plicatilis* , a usual suspect for pond crashes, in CalPoly sites.

ASU Fall

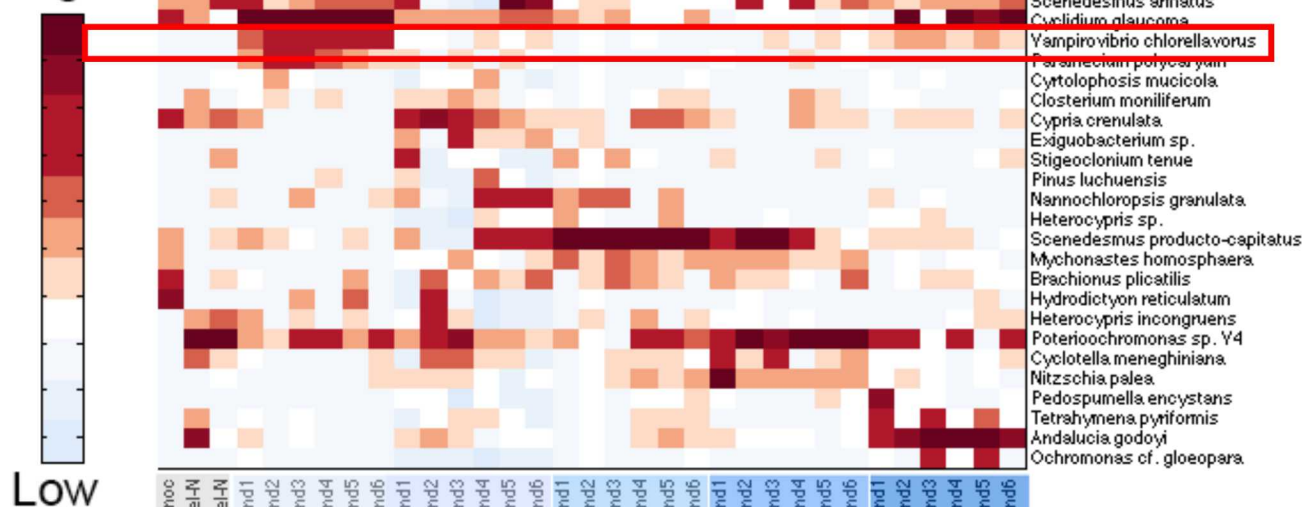
Classification tree analysis to determine the signature of the species for crashes show the most significant variables for machine learning



CalPoly Summer



High



Early

Late

Cyclidium
glaucoma

Poterioochromonas

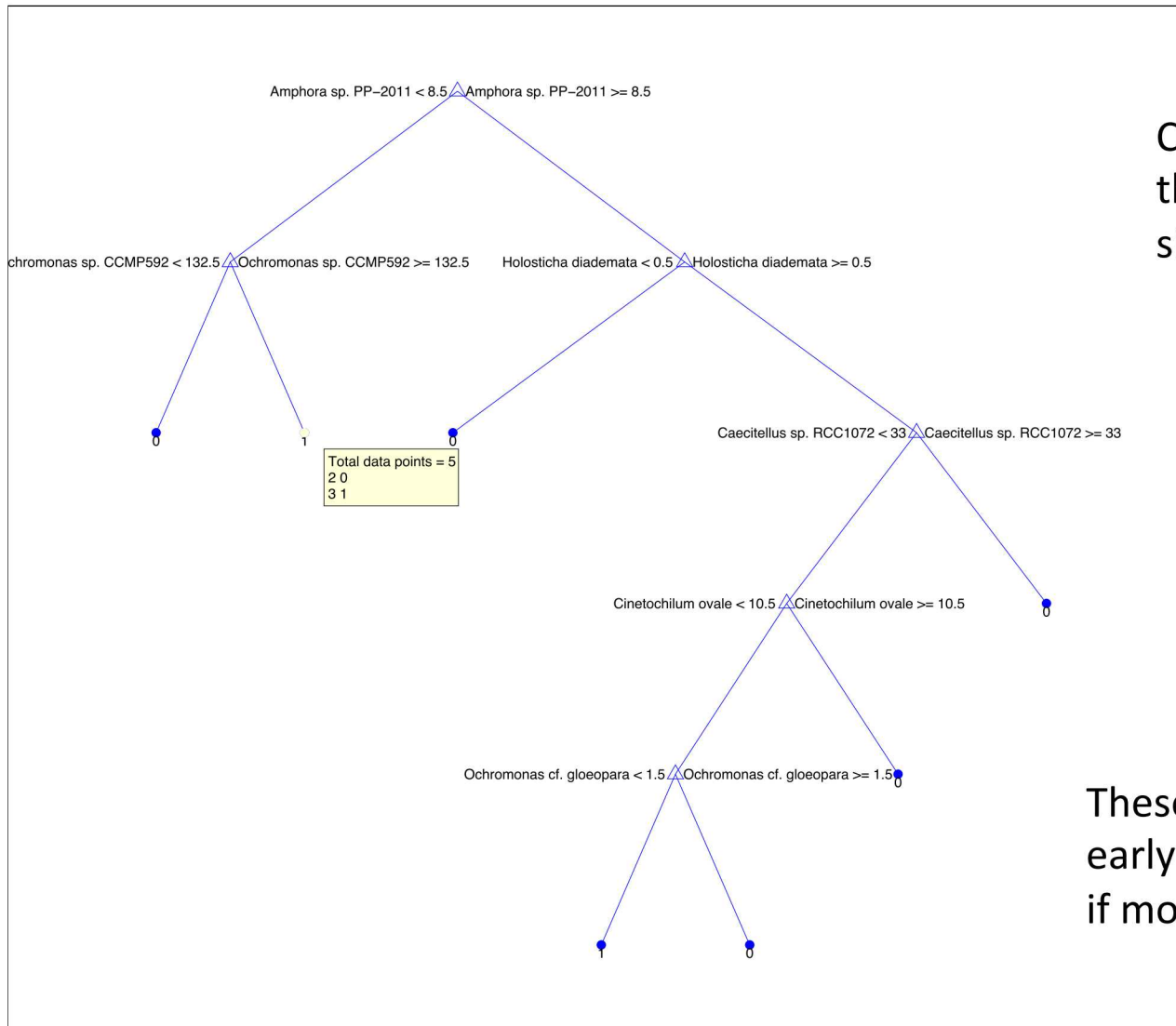
Validated tree showing species determining the presence of crash ponds

```
graph TD; A["Vampirovibrio chlorellavorus < 1"] --> B["Vampirovibrio chlorellavorus >= 1"]; A --> C["Porphyridium purpureum < 0.5"]; B --> D["Crash Pond = 1"]; C --> E["Porphyridium purpureum >= 0.5"]; E --> F["Crash Pond = 1"]; C --> G["Ochromonas cf. gloeopara < 12"]; G --> H["Ochromonas cf. gloeopara >= 12"]; H --> I["Uronema marinum < 39"]; H --> J["Crash Pond = 0"]; I --> K["Uronema marinum >= 39"]; K --> L["Mychonastes homosphaera < 2.5"]; K --> M["Crash Pond = 0"]; L --> N["Mychonastes homosphaera >= 2.5"]; N --> O["Caecitellus parvulus < 5.5"]; N --> P["Crash Pond = 0"]; O --> Q["Caecitellus parvulus >= 5.5"]; Q --> R["Poteriospumella lacustris < 53.5"]; Q --> S["Crash Pond = 0"]; R --> T["Poteriospumella lacustris >= 53.5"]; T --> U["Crash Pond = 1"]; R --> V["Crash Pond = 0"]; S --> V; V --> W["Crash Pond = 0"];
```

These species will serve as a early warning alarm for crash if monitored

These species will serve as a
early warning alarm for crash
if monitored

Decision Tree Nano



Cross-validated tree showing the species determining the signature of crash ponds

These species will serve as a early warning alarm for crash if monitored



Summary

- We have nice workflow for sequencing high throughput 16S and 18S metagenomic samples
- New bioinformatics software pipeline
 - which is fast and efficient to carry out the analysis and producing user friendly outputs.
 - pipelines used and developed:
 - MAGPie (Metagenomics and Amplicon Sequencing Pipeline)
 - & RapTOR (Rapid Threat Organism Recognition)
- And we are in process of developing and improving our data-mining tools for in-depth analysis of these rich datasets.

Questions ?

Sandia National Labs

- **Todd Lane**
- **Kunal Poorey**
- **Deanna Curtis**
- **Haifeng Gang**
- **Pamela Lane**
- **Kelly Williams**

Todd Lane



Deanna Curtis



Pamela Lane



Haifeng Gang



Thank you for your attention!

Acknowledgments

ASU

Gary Dirks
John McGowen
Thomas Dempster
Milt Sommerfeld
William Brandt
Pete Lammers
Jessica Cheng
Jordan McAllister
Sarah Arrowsmith
David Cardello
Theresa Rosov
Mary Cuevas
Jeffrey Prairie
Richard Malloy
Xuezhi Zhang
Henri Gerken
Pierre Wensel
Linda Boedeker
Sarah Mason
Travis Johnson
Sydney Lines
Wyatt Western
Mariah Patton
Maria Bautista

Cal Poly

Tryg Lundquist
Braden Crowe
Garrett Murawsky
Eric Nicolai
Aydee Melgar
Gulce Ozturk
Kaitlyn Jones
Michael Antoine
Trung K Tran
Jake Bender
Heather Freed
Daniel McBroom
Michele
Hendrickson
Gerard Nguyen
Deven Diliberto
Jack Sunderland
Dan Averbuj
Ann Marie Sequeira
Lauren Miller
Michele
Hendrickson
Emily Wang
Jack Sunderland
Ann Marie Sequeira
Soroush Aboutaleb
Lauren Miller

Cellana

Marcella Saracco
Egan Rowe
Emily Knurek
Peter Prentiss
Reyna Javar
Kari Wolff
Keao Bishop-Yuan
Christina Boyko

Florida Algae

Steven Schlosser
Chris Withstandley
Mary Riddle
Nancy Pham Ho (FIT)

Georgia Tech

Thomas Igou
Yongsheng Chen
Steven Van Ginkel
Thomas Igou
Zixuan Hu
Fariha Hassan
Jerry Duncan
Frazier Woodruff
Shusuke Doi
Hao Fu
Patricia Penalver-Argueso
Allison Dunbar
Allison Carr

Sandia National Labs

Ron Pate
Todd Lane
Patricia Gharagozloo
Thomas Reichardt
Pamela Lane
Deanna Curtis
Kunal Poorey

NREL

Phil Pienkos
Lieve Laurens
Ed Wolfrum
David Crocker
Stefanie Van
Wyche

UTEX

Schonna Manning
Jerry Brand

Commercial Algae Management

Albert Vitale
Robert Vitale

Harmon Consulting

Valerie Harmon



U.S. DEPARTMENT OF
ENERGY