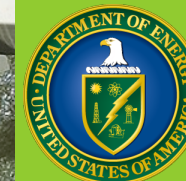




SAND2016-6141C



U.S. DEPARTMENT OF
ENERGY



**Sandia
National
Laboratories**



Capturing Pond Crash Signature in ATP3 sites using Machine
Learning

Kunal Poorey, Todd W. Lane

Deanna J. Curtis, Haifeng Gang, Valerie Harmon, John McGowen

**Sandia National Labs
and ATP³ team**

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.

Sandia Team

Sandia National Labs

Todd Lane

Kunal Poorey

Deanna Curtis

Haifeng Gang

Pamela Lane

Todd Lane



Kunal Poorey



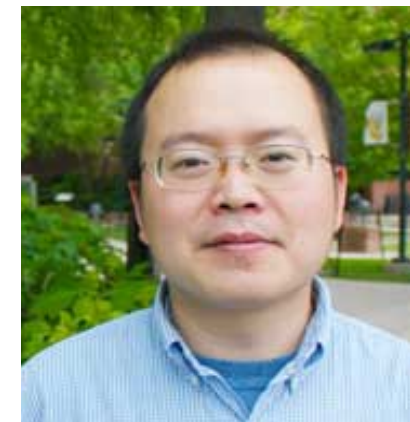
Pamela Lane



Deanna Curtis



Haifeng Gang



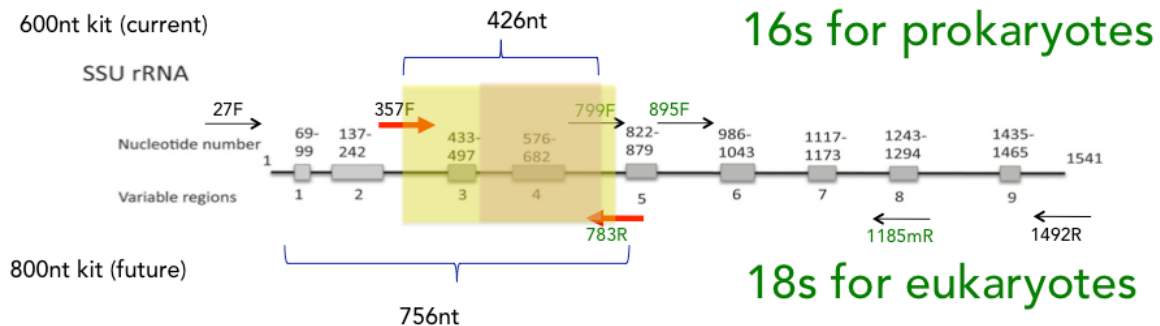
Technical Challenges & Sequencing Methods

Goal : For sustainable reliable production / cultivation of algae.

- Find probiotic communities to favor yield, harvesting and also finding deleterious species causing pond crashes
- Develop methods to detect deleterious species causing pond crashes

Current detection methods: Microscopy and qPCR sampling

- Disadvantages
 - Prior expertise needed to identify organisms (Micro)
 - Sampling bias in identifications (some predators are easy to spot 😊)



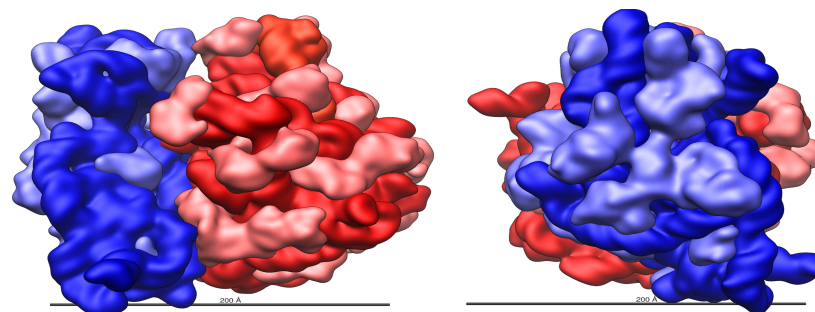
Exploiting the diversity in ssu rRNA gene we sequence DNA targeting rRNA to classify the species present in algae ponds.

- 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

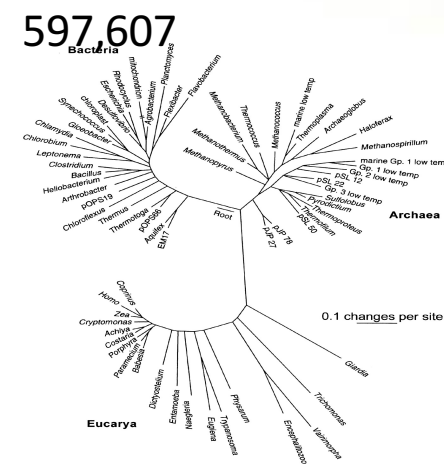
Phylogenetic “Tree of Life” based on SSU rRNA sequences

Ribosome structure



Large Subunit, Small Subunit

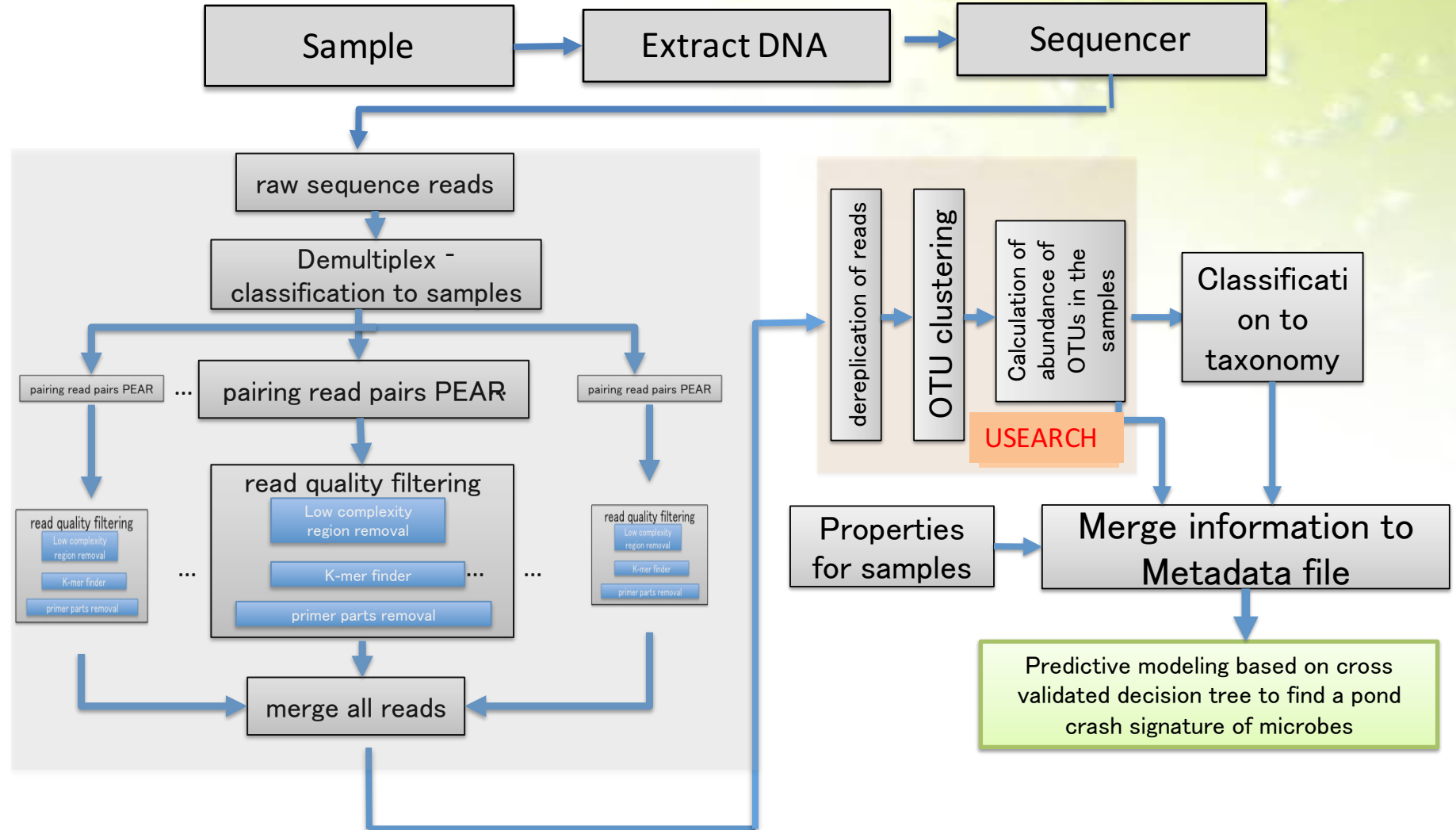
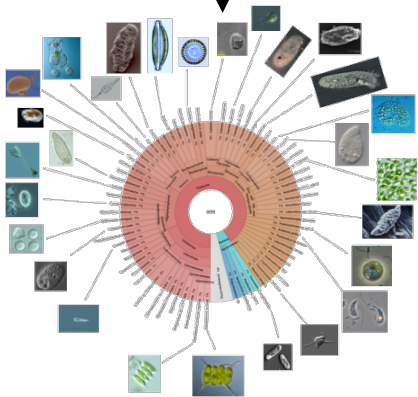
Entries in Silva SSU database



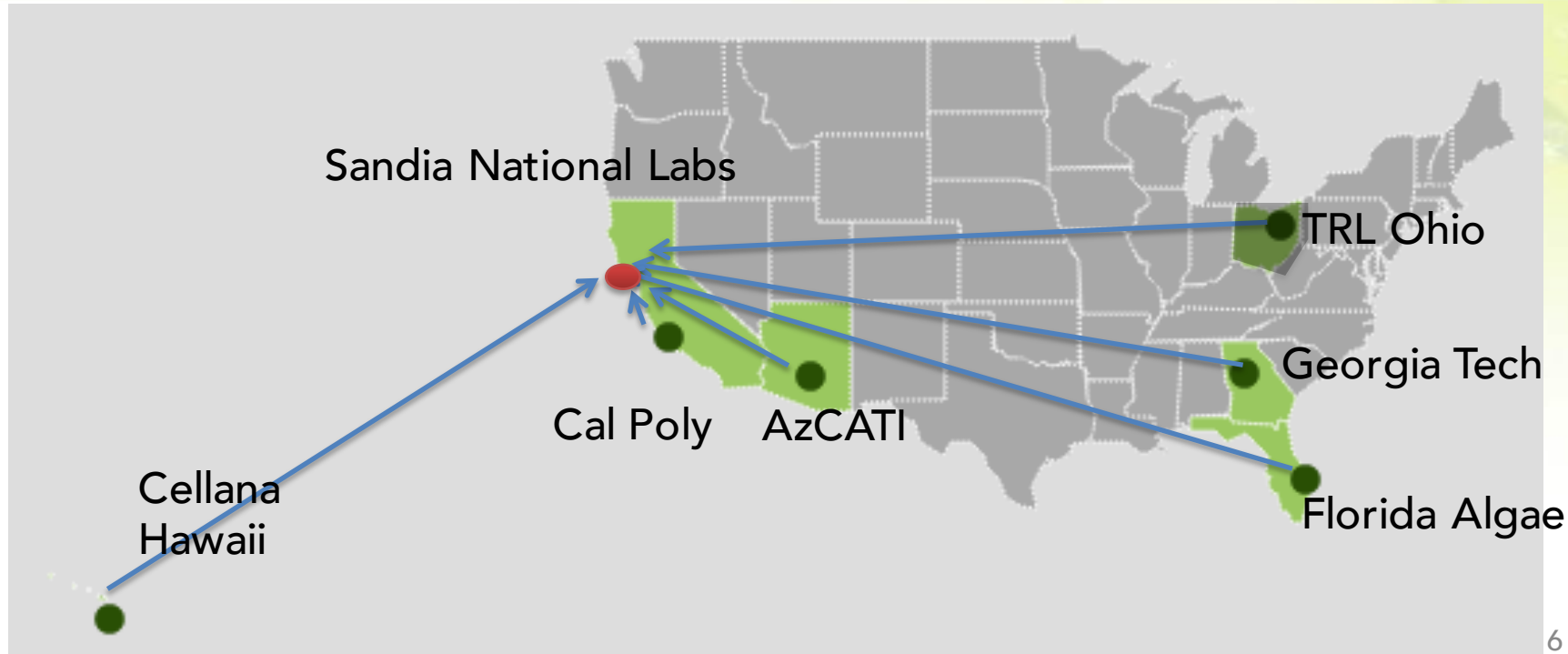
Microbial Community Analysis



Microbiome



ATP³ sites



Sample acquisition

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

Map of sequence reads to the 16s DNA locus

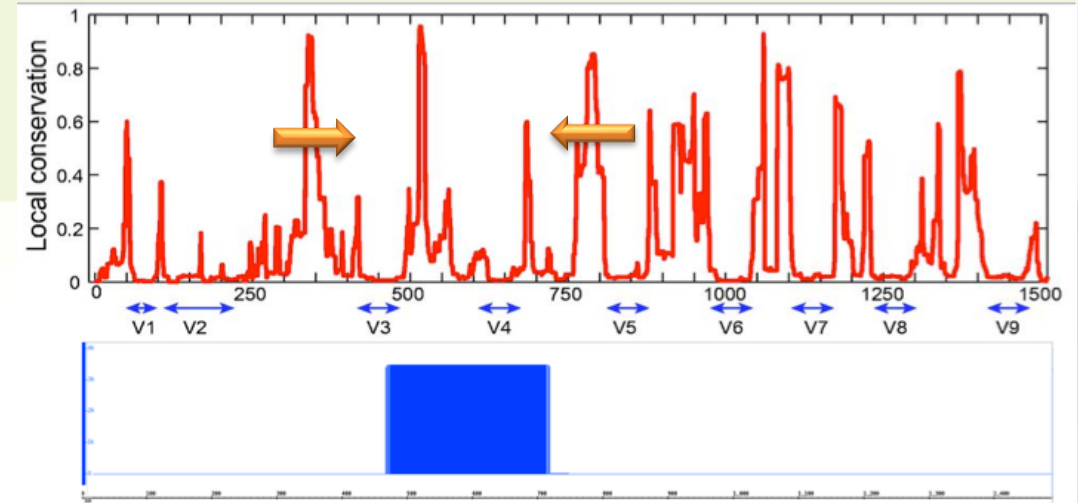
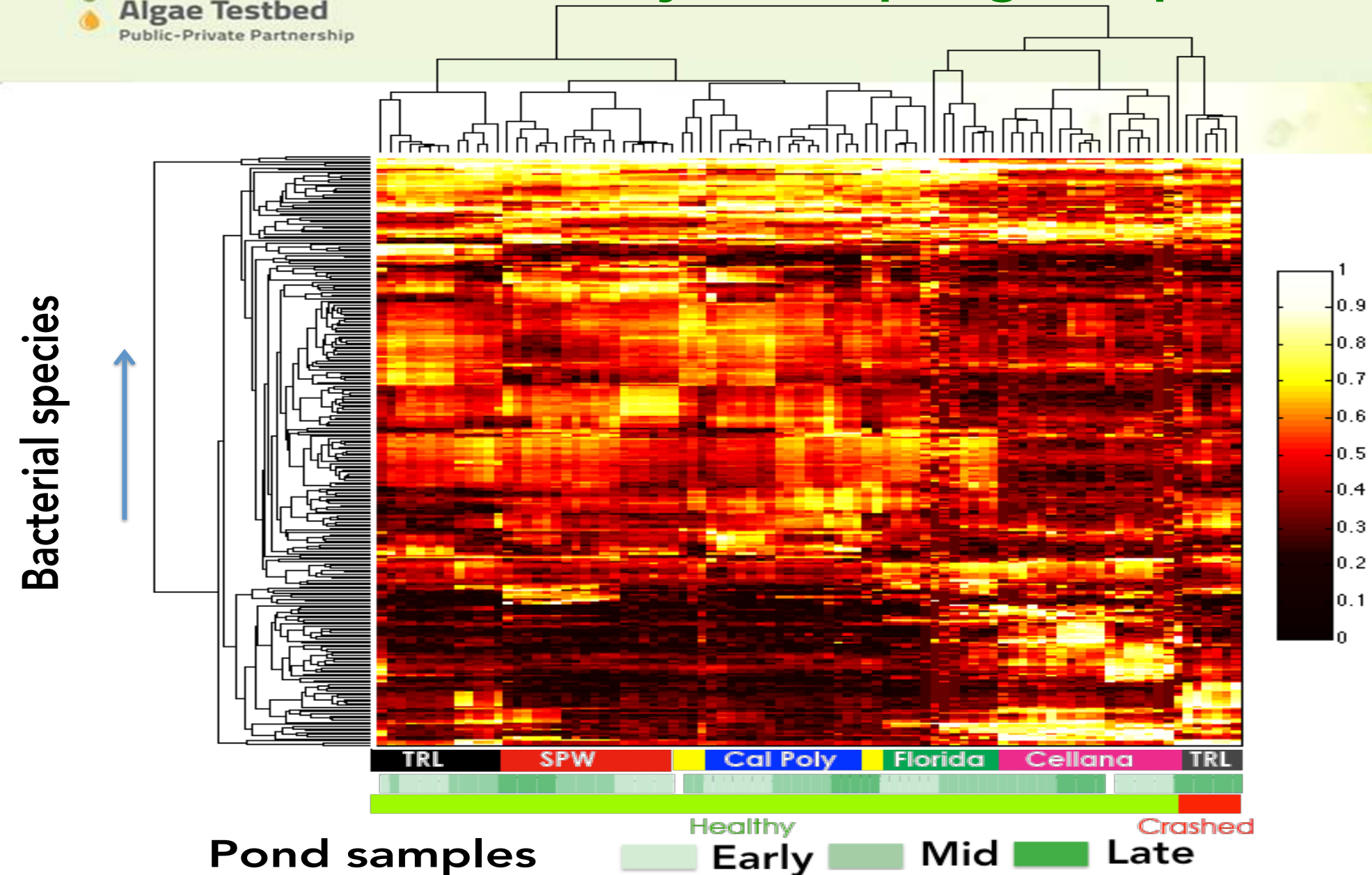


Table of classified OTUs and their abundances can be displayed on websites as well

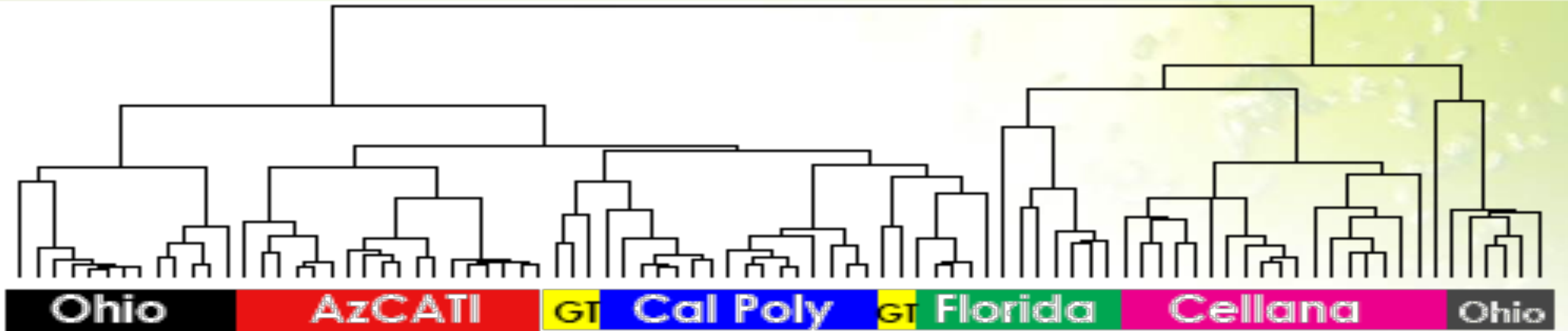
taxid in DB	Taxonomy	taxa	Pond 1	Pond 2	Pond 3	Pond 4	Pond 5	Pond 6
U41092.1.1792	Eukaryota;Eustigmatophyceae;Eustigmatales;Monodopsidacea	Nannochloropsis granulata;	13379	1	1	1	19644	21286
Y14950.1.1794	Eukaryota;Viridiplantae;Chlorophyta;Trebouxiophyceae;Chlore	Chlorella sp. Yanaqocha RA1;	3375		216	113	677	202
JF834543.1.1228	Eukaryota;Bacillariophyta;Bacillariophyceae;Thalassiosphaerales;	Amphora sp. PP-2011;		1		1	1	1
DQ059583.1.1784	Eukaryota;Spirotrichea;Urostylida;Holostichidae;Holosticha;Ho	Holosticha diademata;	1					
EF165112.1.1687	Eukaryota;Chrysophyceae;Chromulinales;Chromulinaceae;Och	Ochromonas cf. gloeopara;	1		1		8	
FR865727.2.1770	Eukaryota;Viridiplantae;Chlorophyta;Chlorophyceae;Sphaeropl	Scenedesmus armatus;	12		27	10	1	4
AY520448.1.1735	Eukaryota;Bicosoecida;Cafeteriaeae;Cafeteria;Cafeteria minim	Cafeteria minima;		6	13	3	1	17
AY622191.1.1753	Eukaryota;Metazoa;Arthropoda;Ostracoda;Podocopida;Cyclopo	Cypria crenulata;			36	12	30	3
HM161745.1.1787	Eukaryota;Synurophyceae;Ochromonadales;Ochromonadaceae	Poterioochromonas sp. Y4;	11		7			3
L27634.3.1794	Eukaryota;Labyrinthulomycetes;Thraustochytriaceae;Labyrinth	Labyrinthuloides minuta;						
EU106848.1.1693	Eukaryota;Bicosoecida;Caecitellus;Caecitellus sp. RCC1072;	Caecitellus sp. RCC1072;		2	14		21	63
M74497.1.1789	Eukaryota;Viridiplantae;Chlorophyta;Chlorophyceae;Sphaeropl	Hydrodictyon reticulatum;	7		12	11	20	
JN120201.1.1541	Eukaryota;Oligohymenophorea;Opisthonectidae;Opisthonecta	Opisthonecta henneguyi;						
DQ514856.1.1722	Eukaryota;Bacillariophyta;Coscinodiscophyceae;Stephanodisca	Cyclotella meneghiniana;			5		1	
EU032356.1.1751	Eukaryota;Oligohymenophorea;Pleuronematida;Cyclidiidae;Cy	Cyclidium glaucoma;	1		9	1	8	1
JF489982.1.1752	Eukaryota;Eustigmatophyceae;Eustigmatales;Monodopsidacea	Nannochloropsis oceanica;						1
EU039885.1.1700	Eukaryota;Colpodea;Colpodida;Hausmanniellidae;Bresslauides	Bresslauides discoideus;						
U49911.1.1808	Eukaryota;Metazoa;Rotifera;Monogononta;Ploimida;Brachioni	Brachionus plicatilis;			11	2	9	
AF352222.1.2260	Eukaryota;Viridiplantae;Streptophyta;Zygnemophyceae;Desmi	Closterium moniliferum;						
HQ912576.1.1744	Eukaryota;Bacillariophyta;Coscinodiscophyceae;Stephanodisca	Cyclotella meneghiniana;						

16S Analysis of spring samples



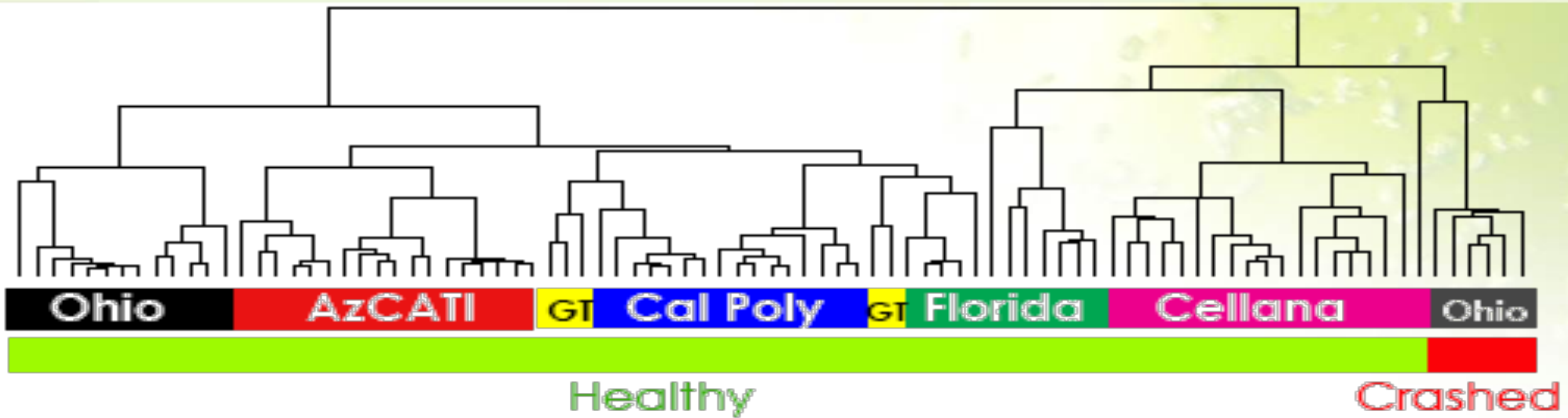
- Meta Summary for spring data from all sites
- Heatmap was made from clustering the classified OTU abundances for both species and samples.
- The Ponds from the same site cluster together

16S Analysis of spring samples



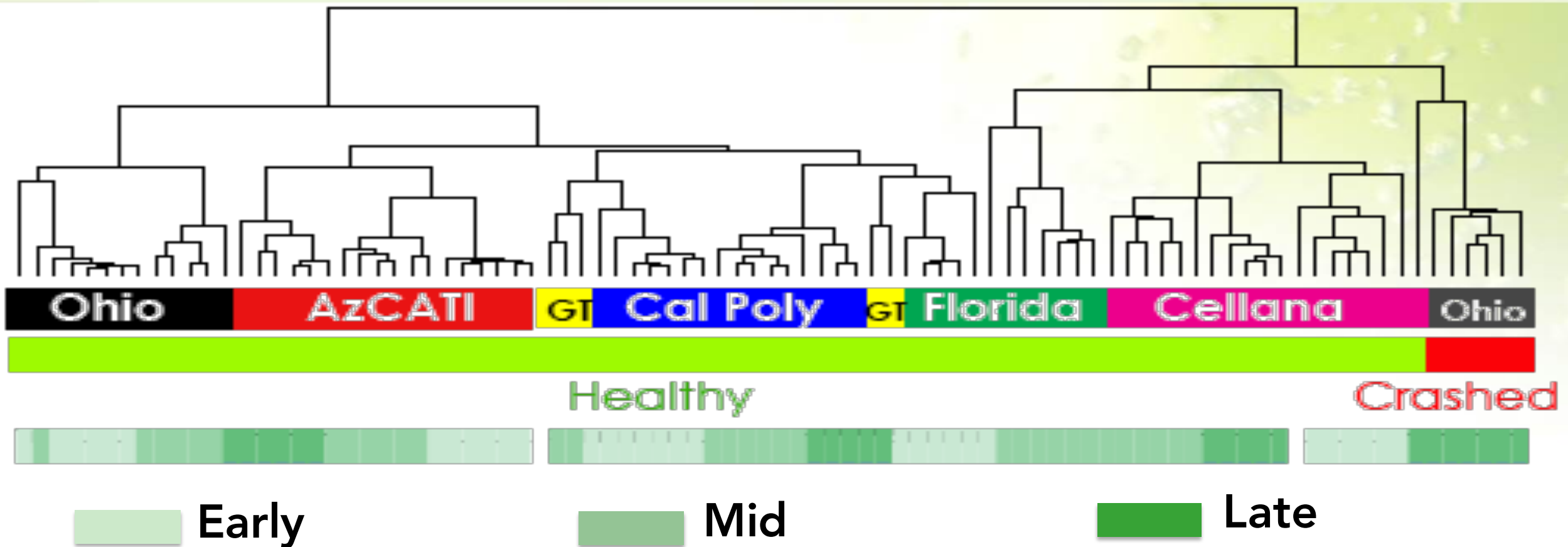
Bacterial communities can identify samples from the same site

16S Analysis of spring samples



Crashed ponds had a different community structure than healthy ponds

16S Analysis of spring samples

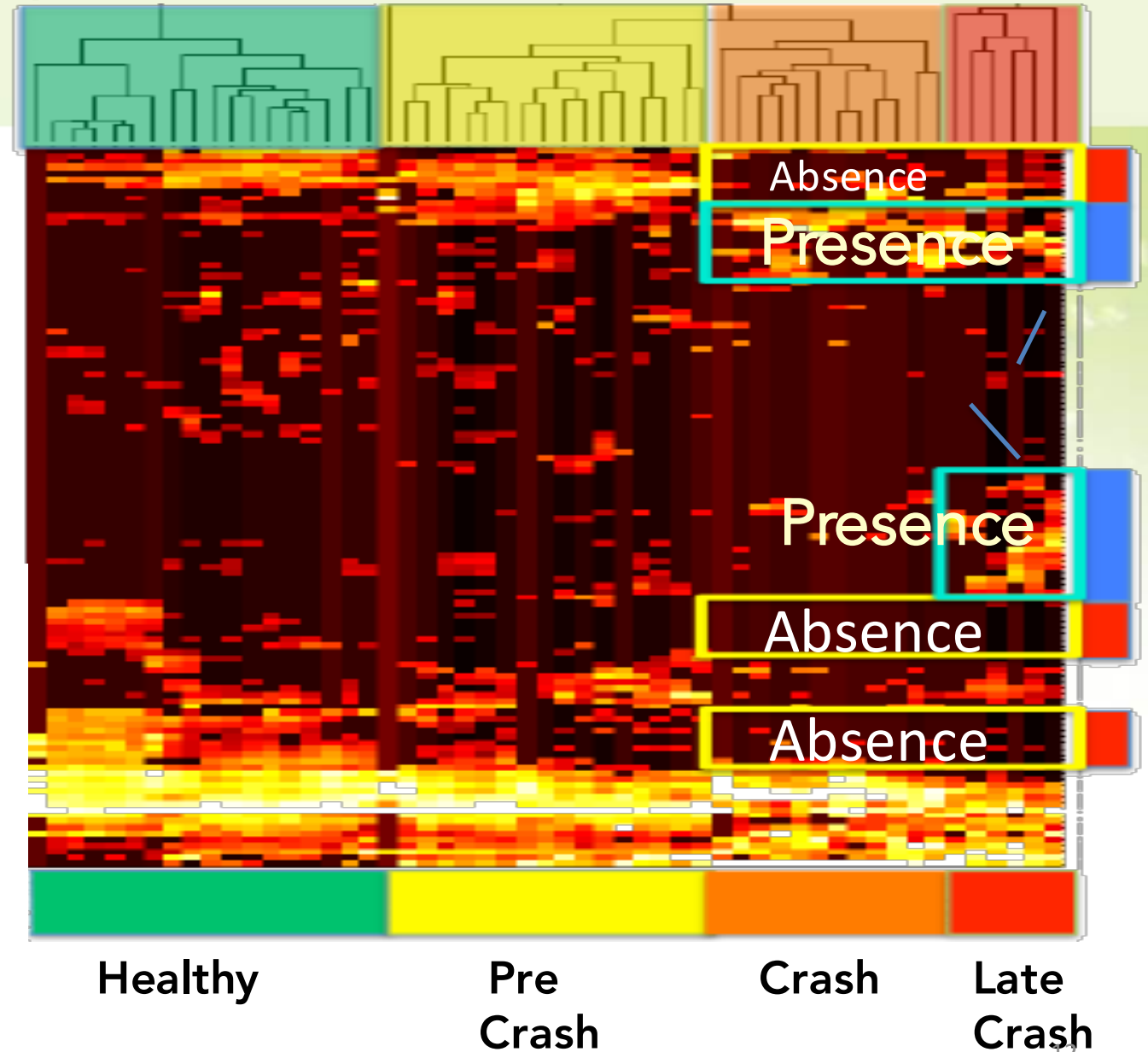
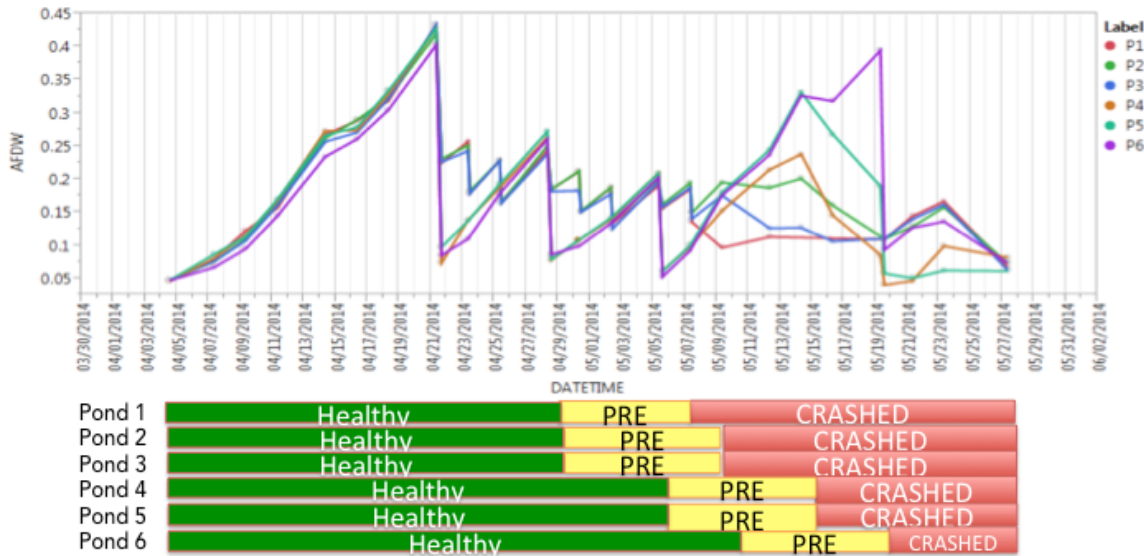


And within same sites the community structure gradually changed with time

TRL – Ohio site Spring 2014

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

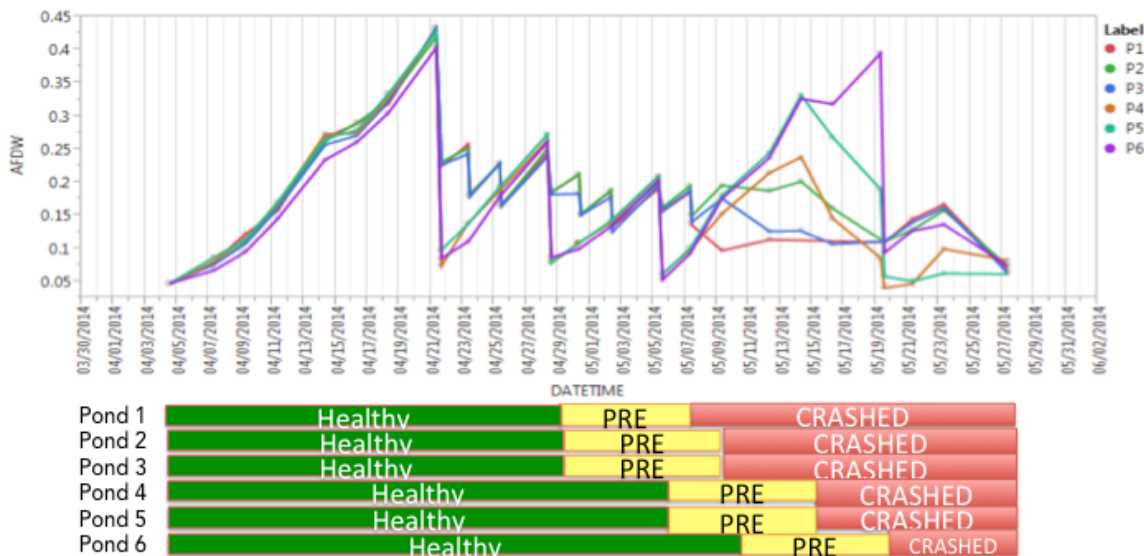
Ash-Free Dry Weight



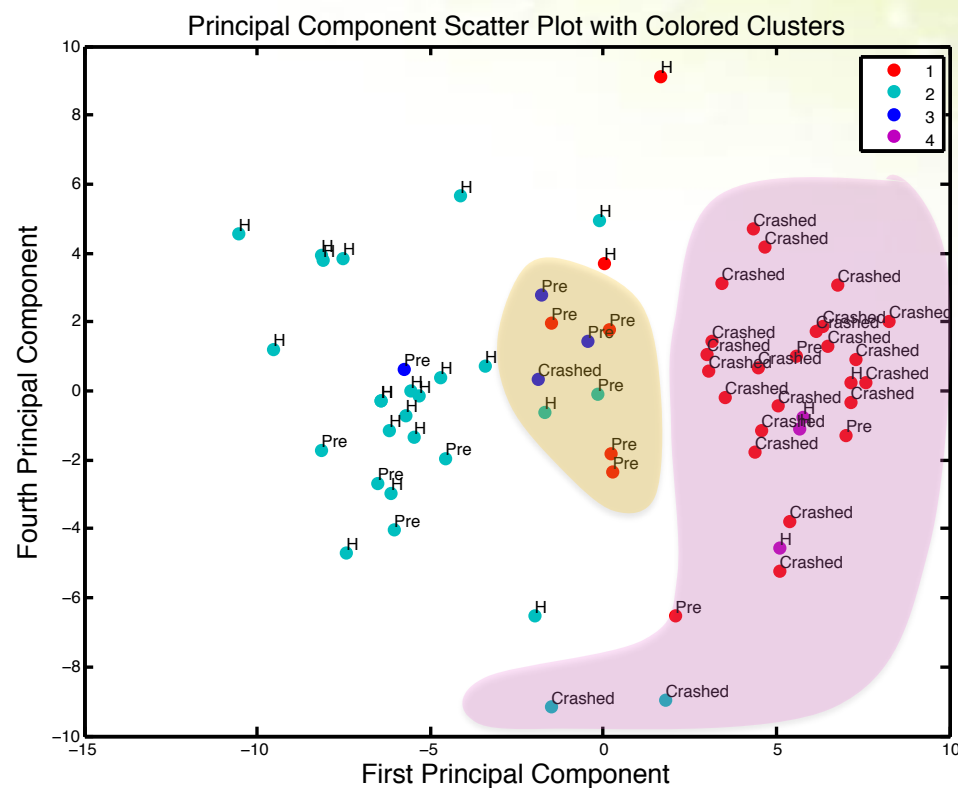
TRL – Ohio site Spring 2014

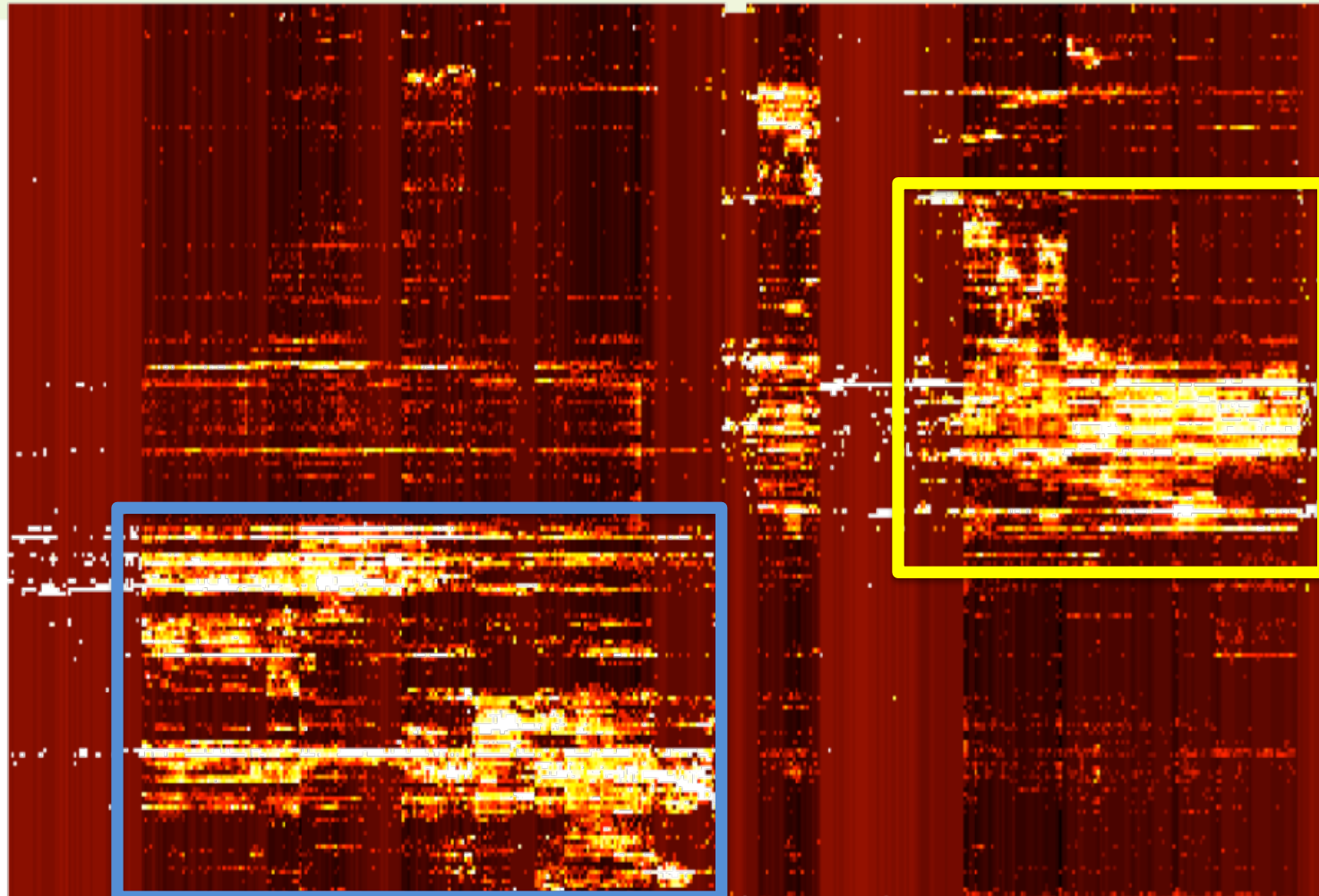
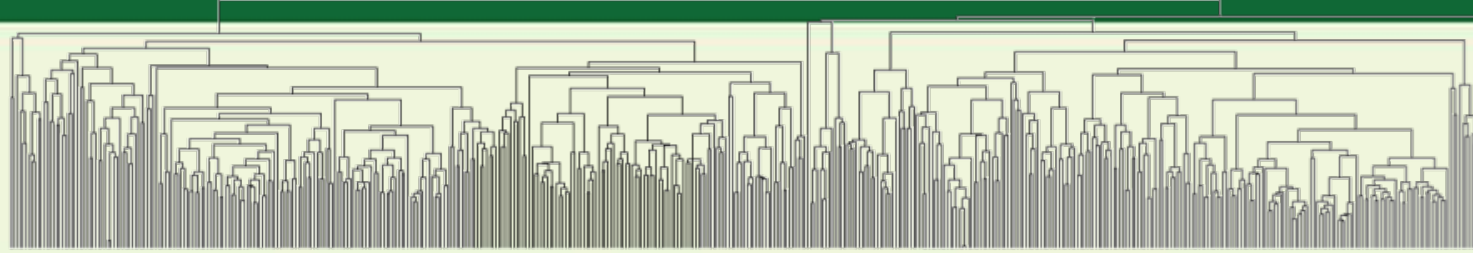
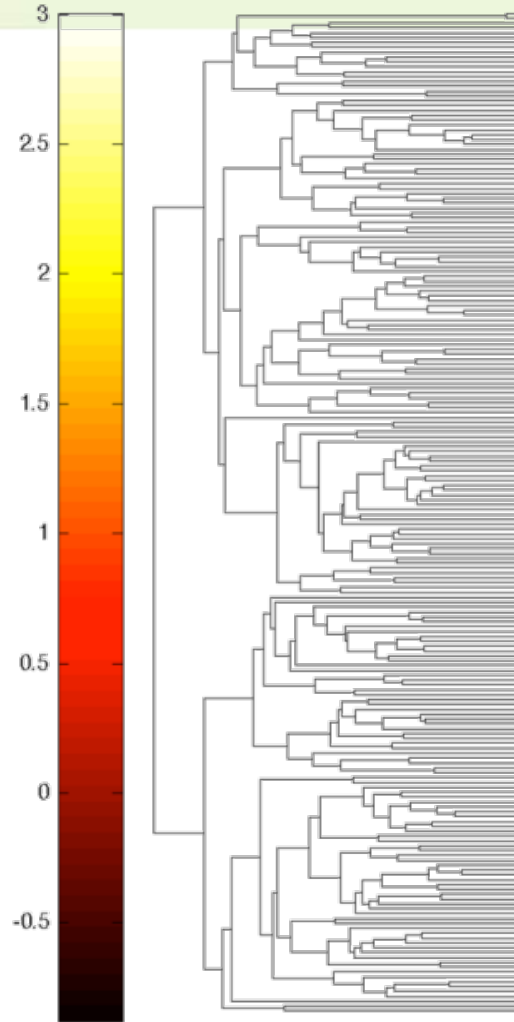
Clustering analysis on the ponds could differentiate the healthy and crashed ponds.

Ash-Free Dry Weight



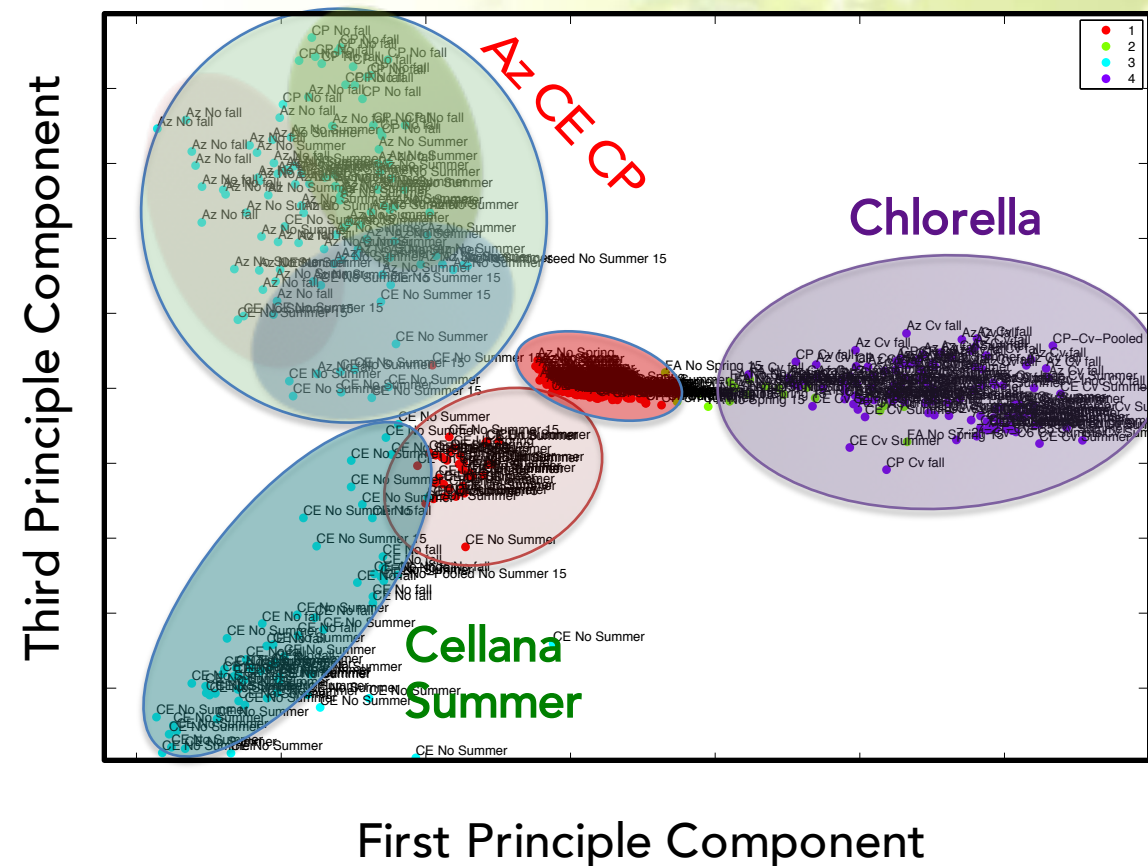
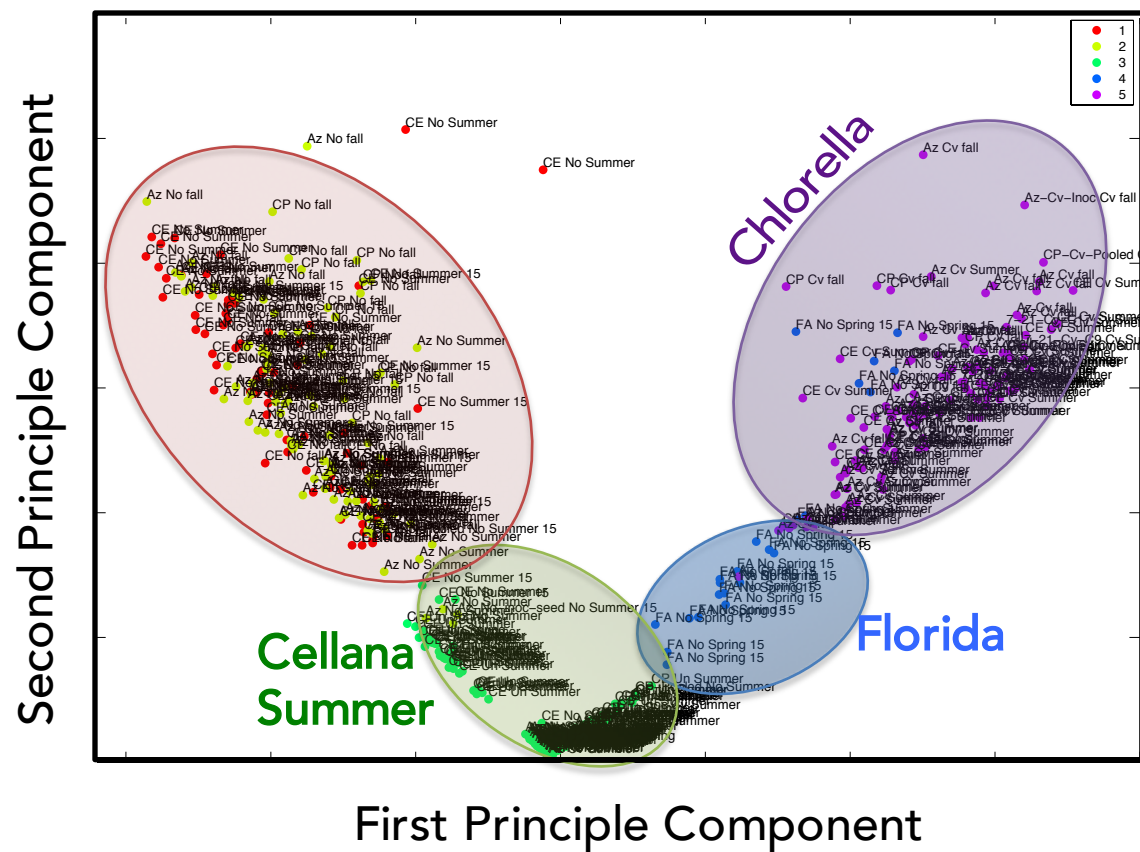
Principle component analysis also confirmed that crashed ponds have a different microbiome signature





AZ CP Cell FA AZ CP AZ Cell AZ
 Nannochloropsis Chlorella

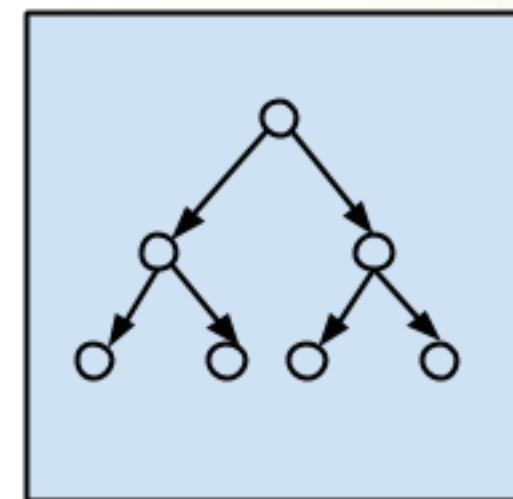
Spring Fall Summer Fall



Identifying Trends in Pond Crashes

	Species 1	Species 2	Species 3		Species M	Result
Pond1_Site1	34	456	2		78	Crash
Pond2_Site2	0	2	45		900	Healthy
Pond3_Site3	765	0	4		22	Healthy
Pond4_Site4	44	334	11		12	Healthy
Pond5_Site5	73	543	7		5	Crash
....	...					
Pond6_SiteN	456	100	233		33	Healthy

CART Classification Regression Tree

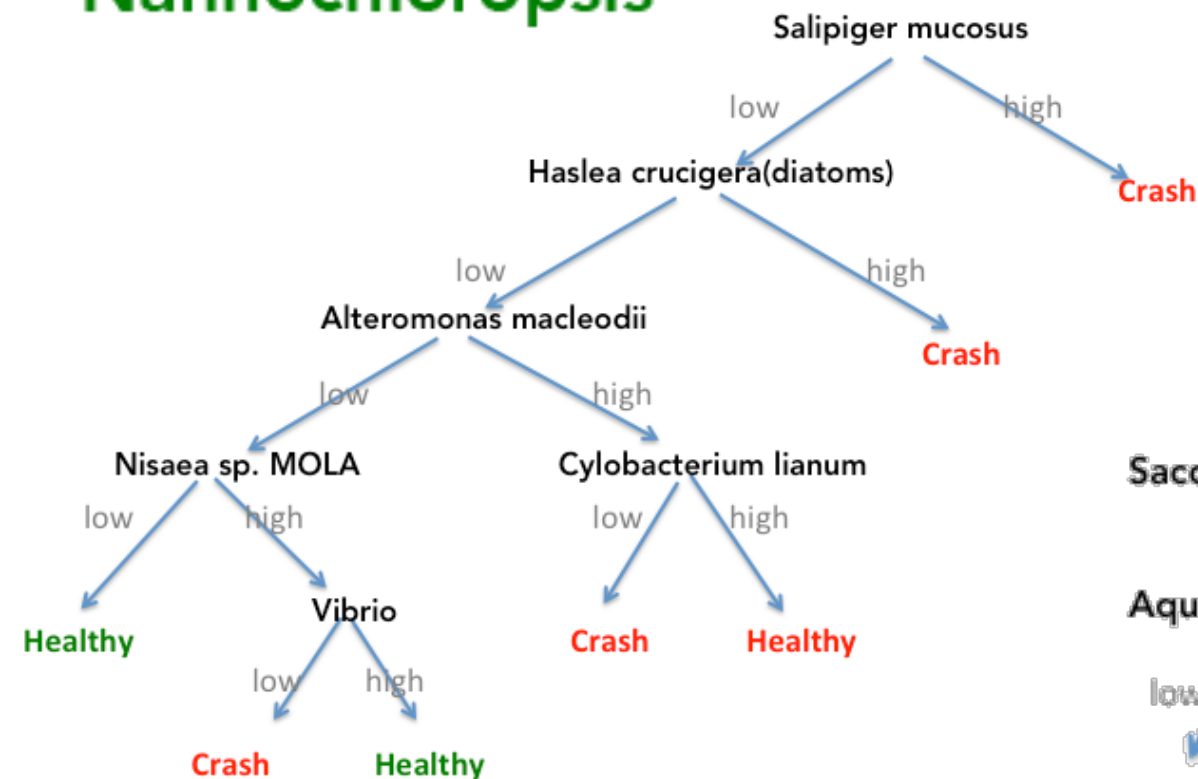


Decision Tree
Algorithms

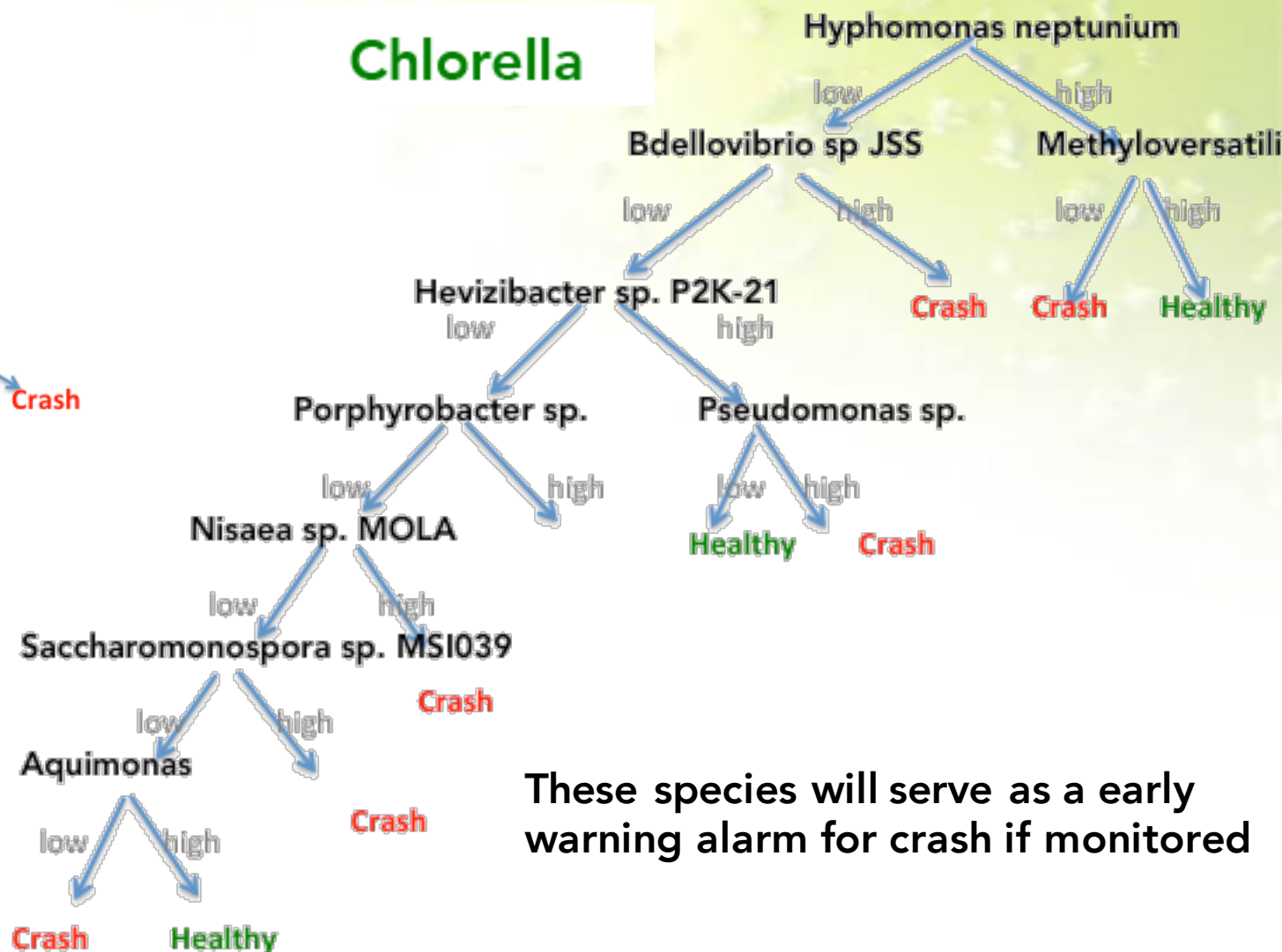
Applying data driven approaches Decision Tree

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

Nannochloropsis



Chlorella

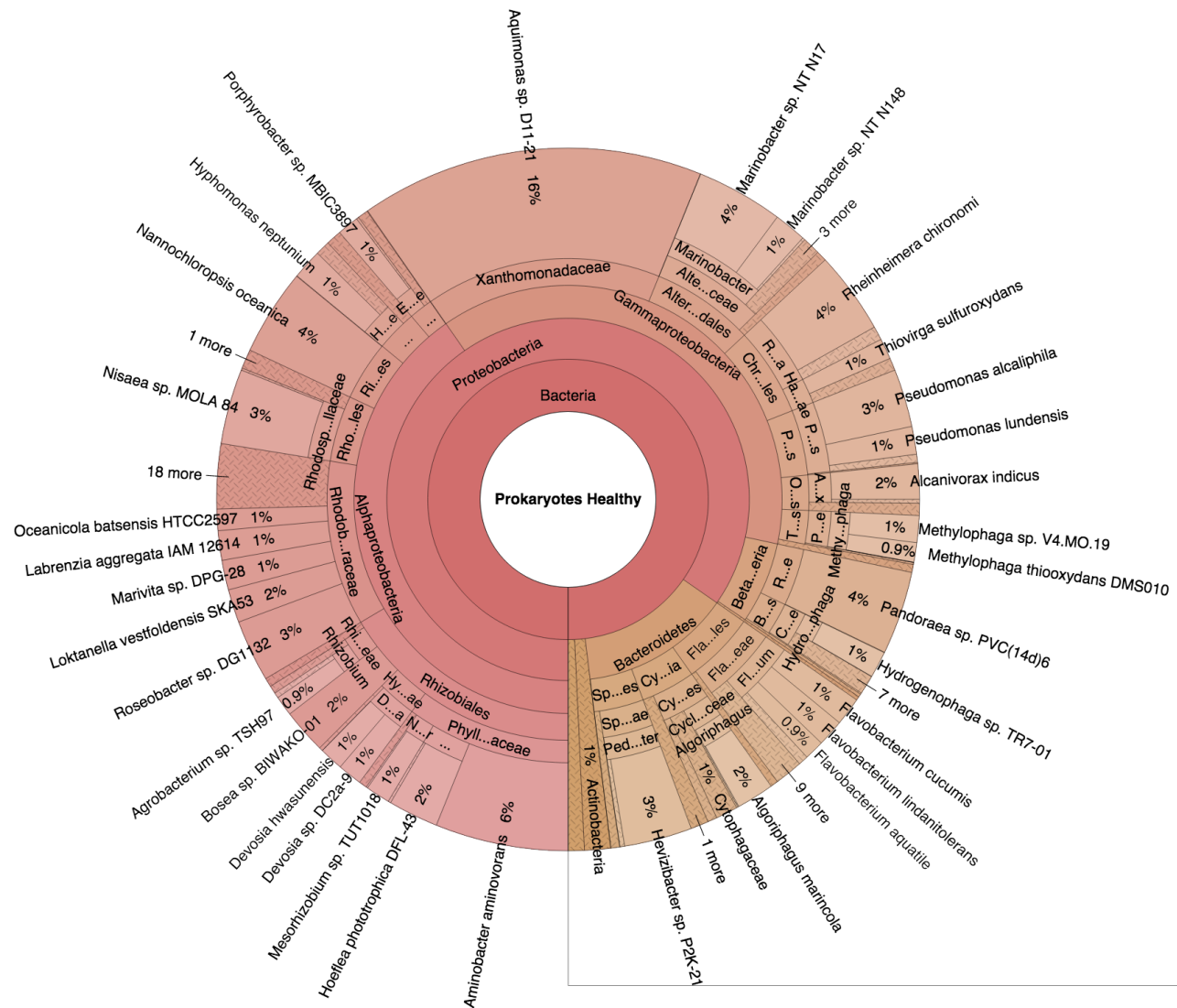


These species will serve as a early warning alarm for crash if monitored

Crash

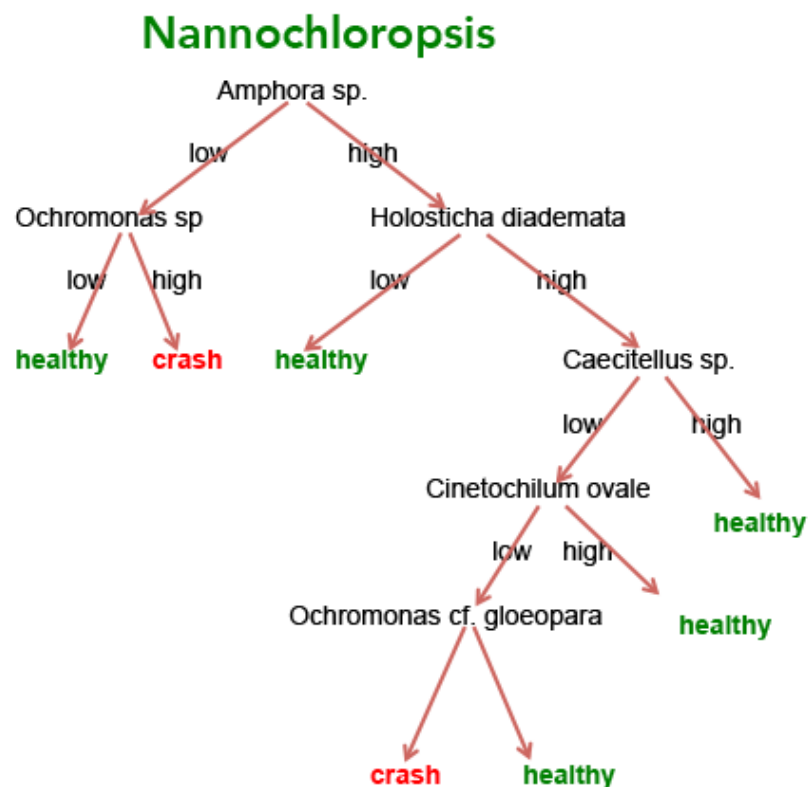


Healthy

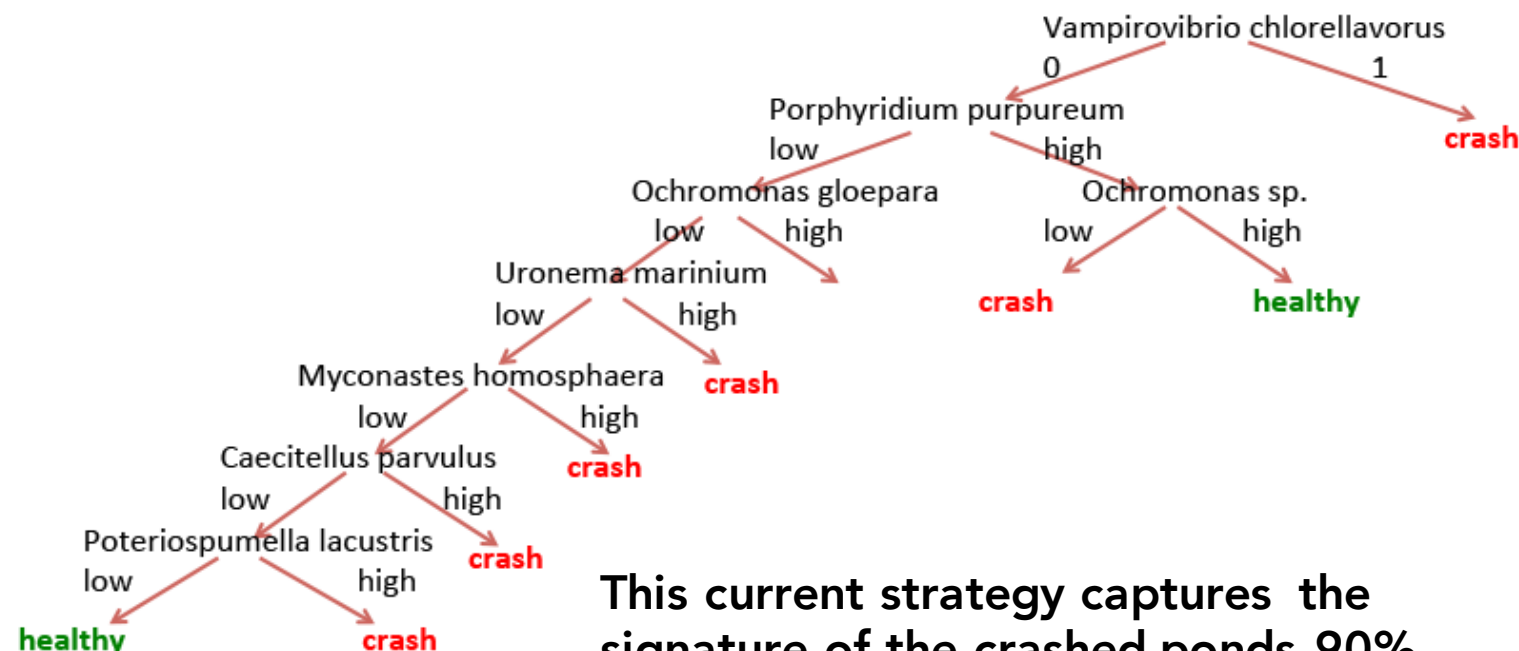


Applying data driven approaches Decision Tree

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



Chlorella

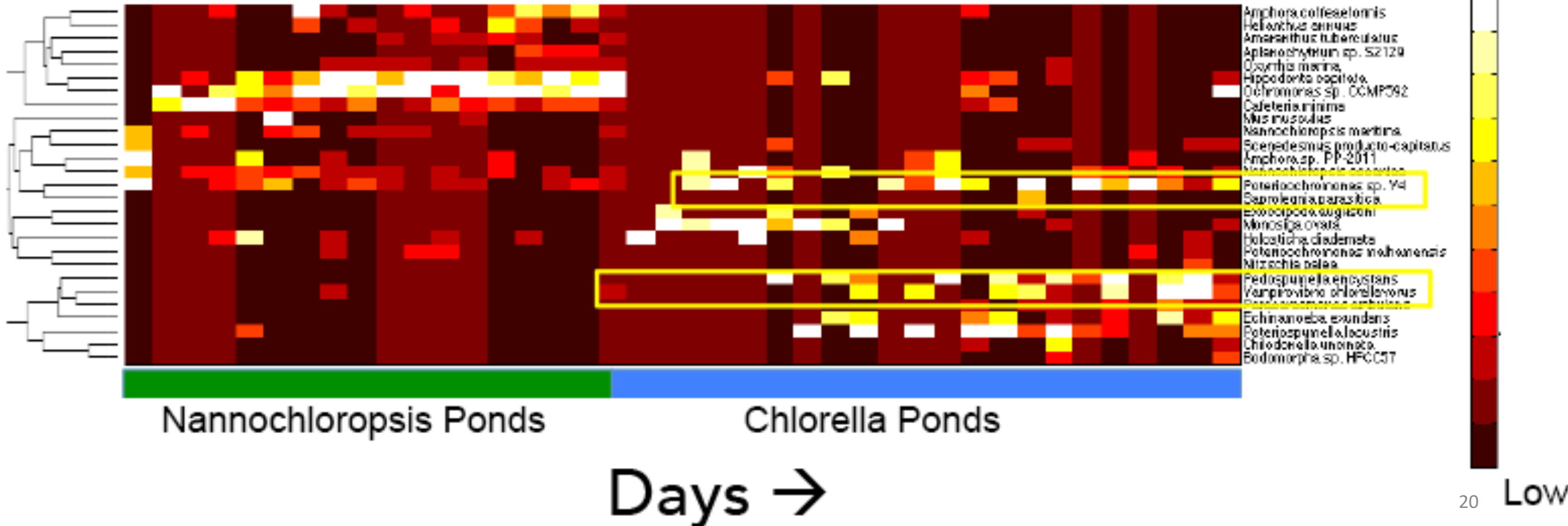
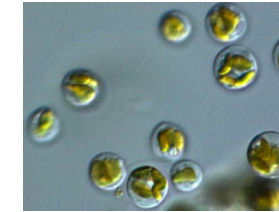
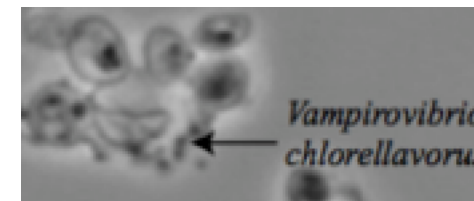
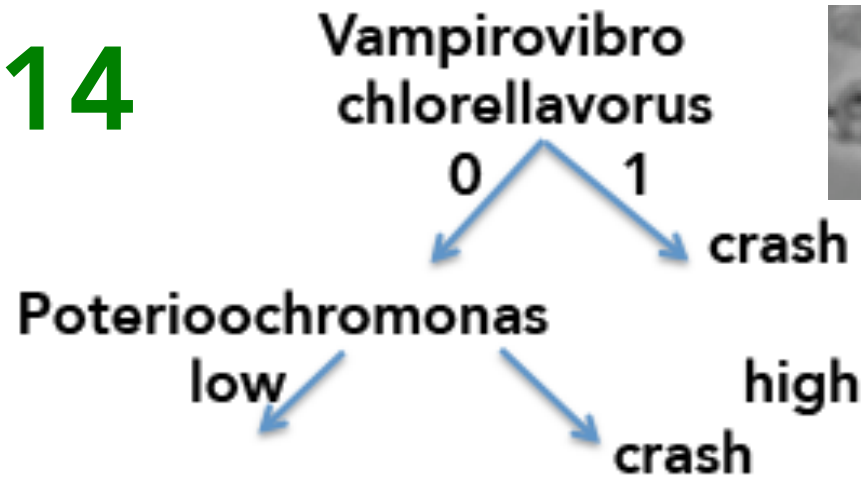


This current strategy captures the signature of the crashed ponds 90% accurately

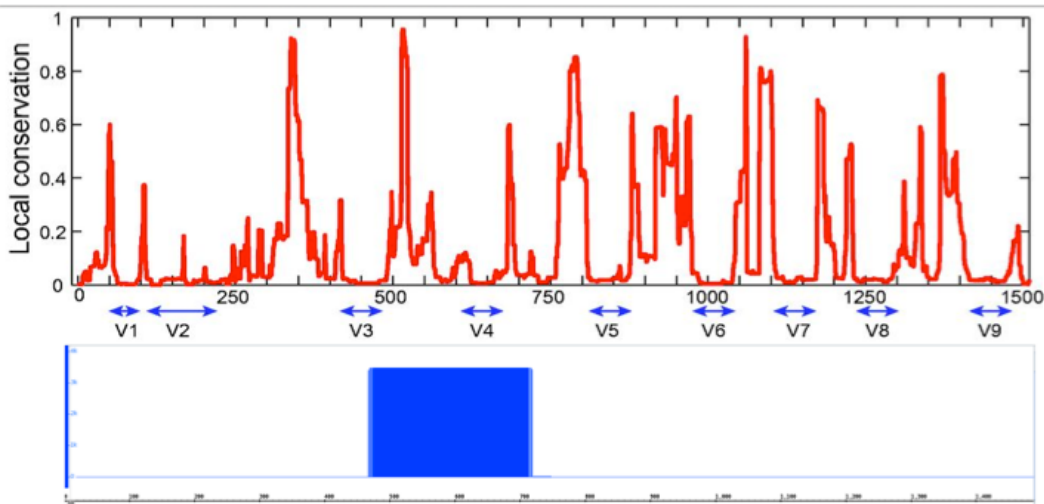
These species will serve as a early warning alarm for crash if monitored

ASU Fall 2014

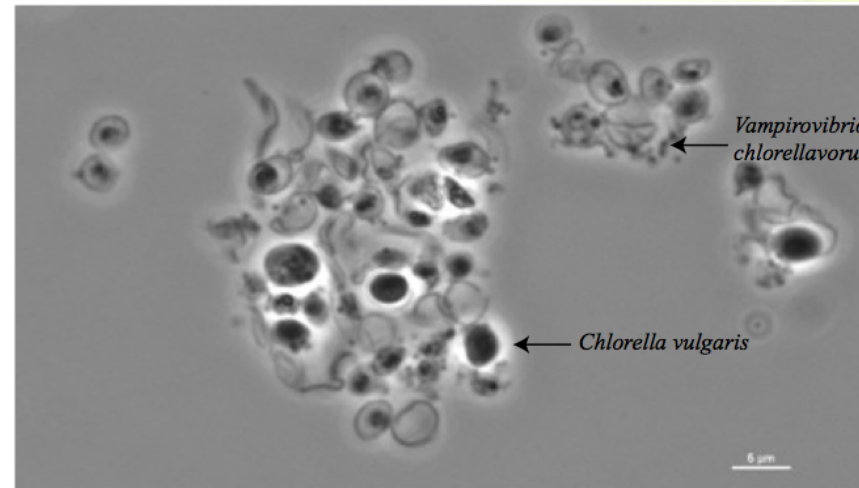
Classification tree analysis to determine the signature of the species for crashes show the most significant variables for machine learning



Vampirovibrio chlorellavorus



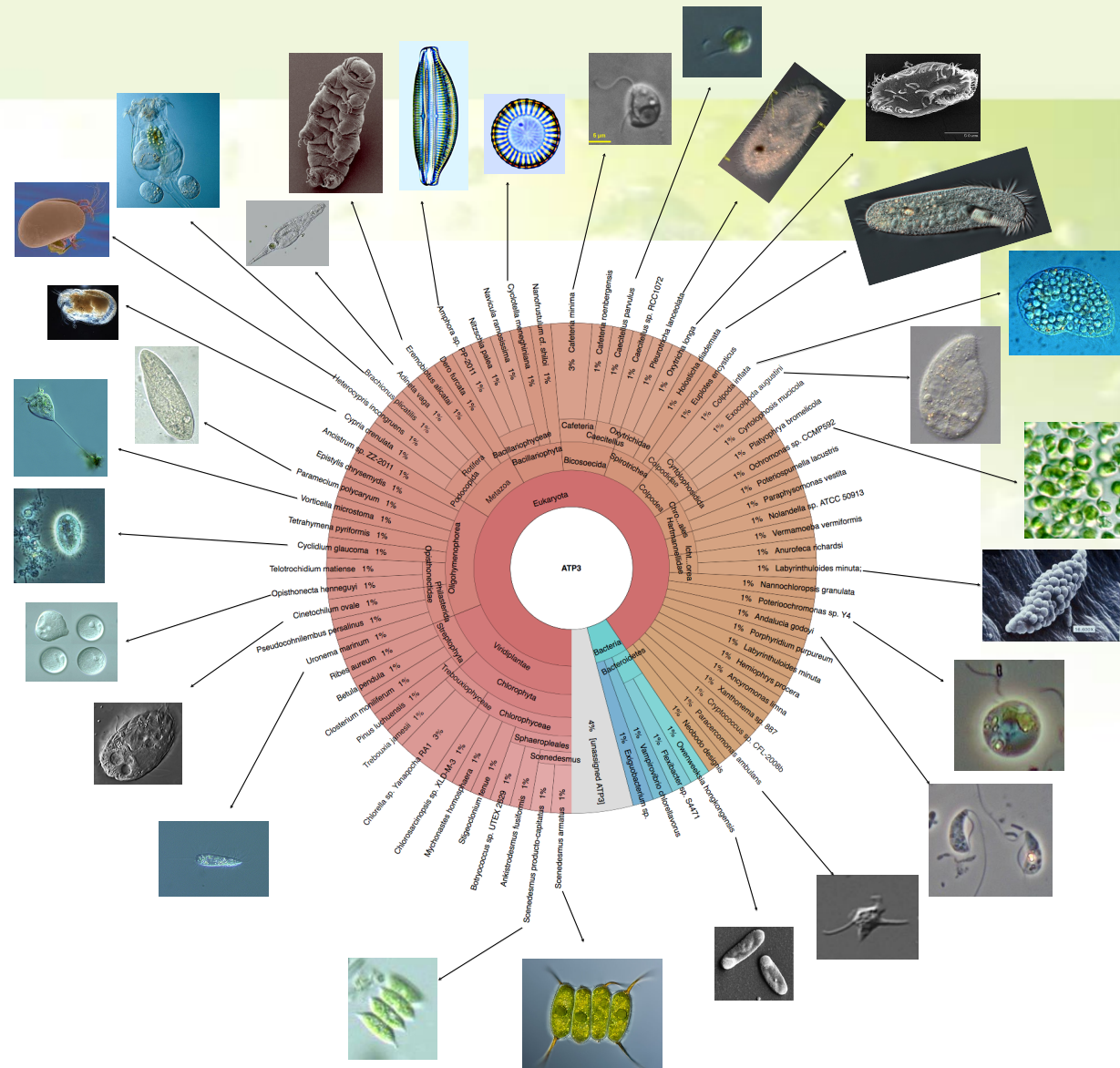
The OTU sequence called is a genuine bacterial sequence and had 99% identity to the 16s
Vampirovibrio chlorellavorus



(Rochelle Melissa Soo 2015) PhD dissertation

Micrograph of Vampirovibrio chlorellavorus

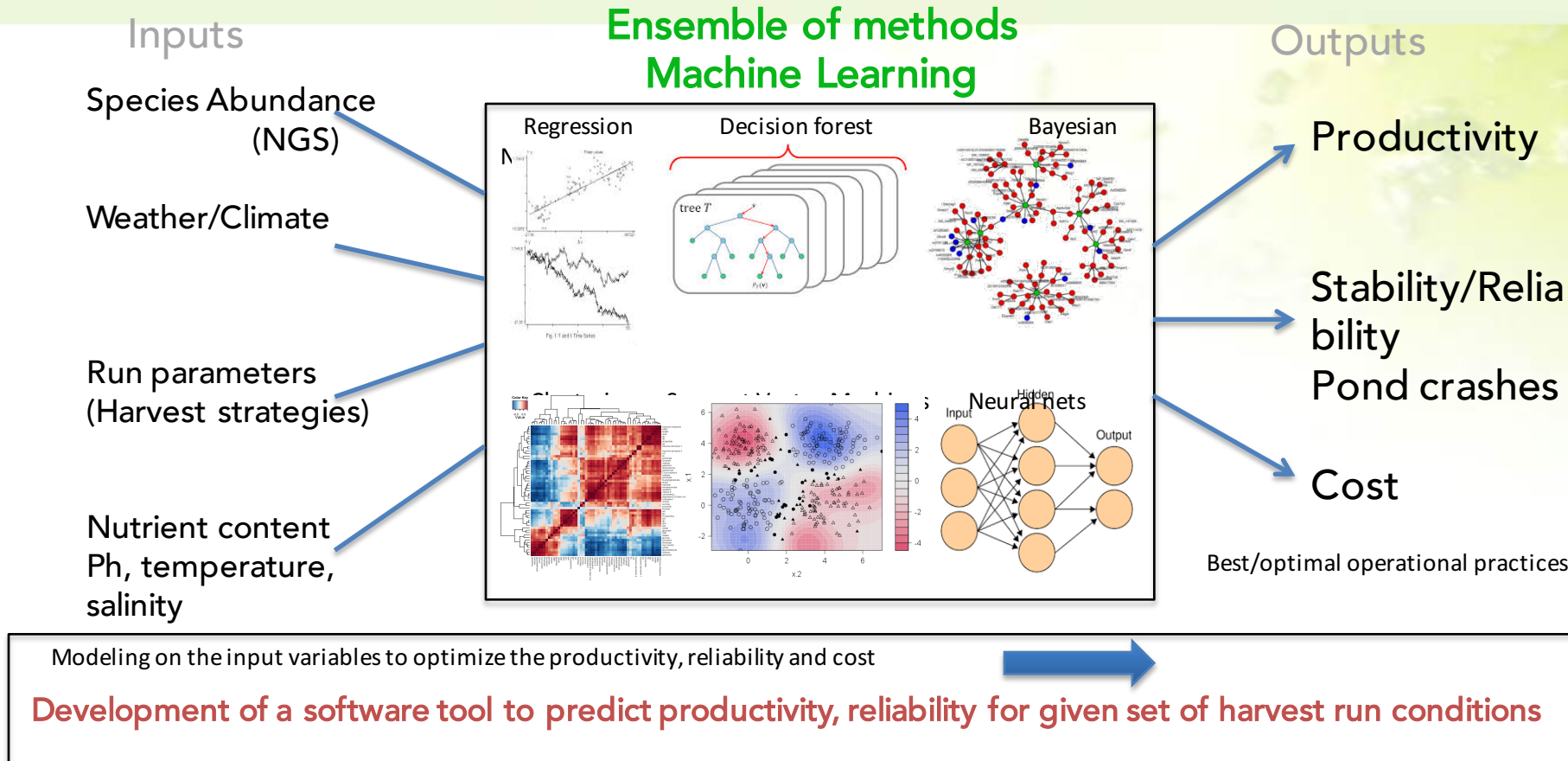




Summary

- We have smooth workflow for sequencing high throughput 16S and 18S amplicon sequencing samples with a quick turnover.
- New bioinformatics software pipeline
 - fast and efficient to carry out the analysis and producing user friendly outputs.
 - pipelines used and developed:
 - **MAGPie** (Metagenomics and Amplicon Sequencing Pipeline)
 - **ASAP** (Amplicon Sequencing Analysis Pipeline)
 - & **RapTOR** (Rapid Threat Organism Recognition)
- We have identified Eukaryotic and Prokaryotic community structure associated with pond crashes

The Future



Species Abundance
(NGS, microscopy counts)

Run parameters
(Harvest strategies)

Modeling of contaminant growth within the pond

$$\dot{a} = \alpha a \left[1 - a \left(\frac{K_a \varepsilon b}{b_c + \varepsilon b} + K_v \right)^{-1} \right],$$

Contaminant growth model

Optimal harvest conditions

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 Aboutalebi
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 Wychen
Eric Knoshaug
Ryan Davis

UTEX

Schonna Manning
Jerry Brand

Commercial Algae Management

Albert Vitale
Robert Vitale

Harmon Consulting

Valerie Harmon



U.S. DEPARTMENT OF
ENERGY