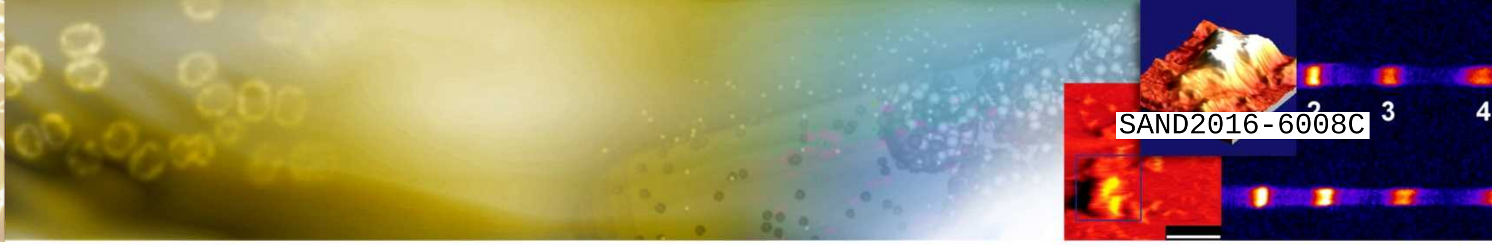
SANDIA
BIOSCIENCE



SAND2016-6008C

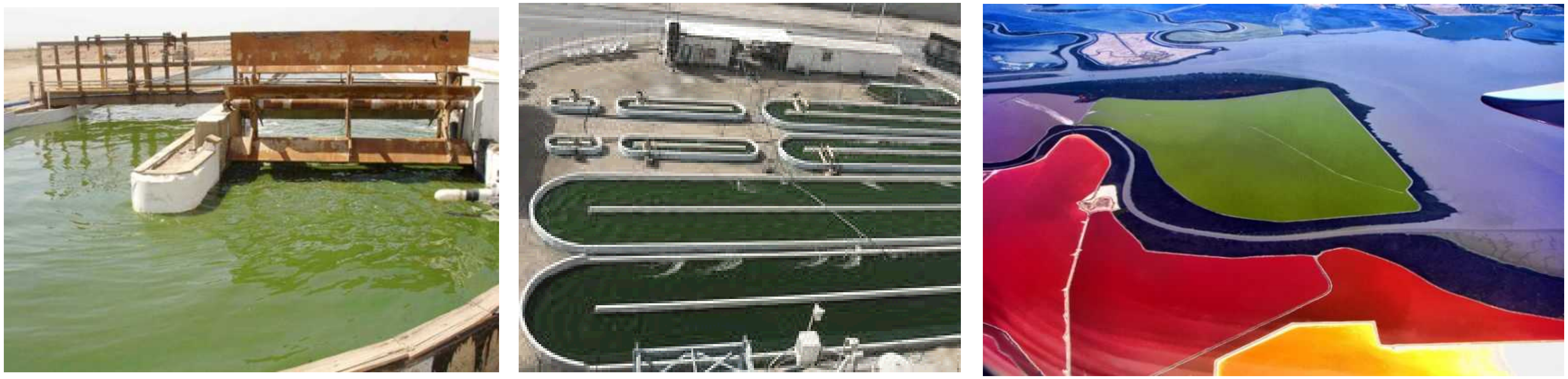
Longitudinal analysis of microbiota in microalga *Nannochloropsis salina* cultures

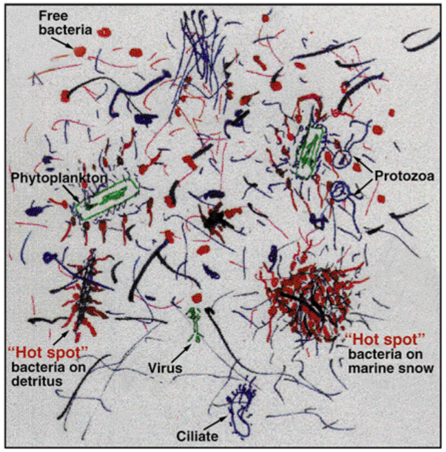
Haifeng Geng

Department of Systems Biology
Sandia National Laboratories

Sandia National Laboratories

The algae and microbiota form complex consortia that drive system level functions

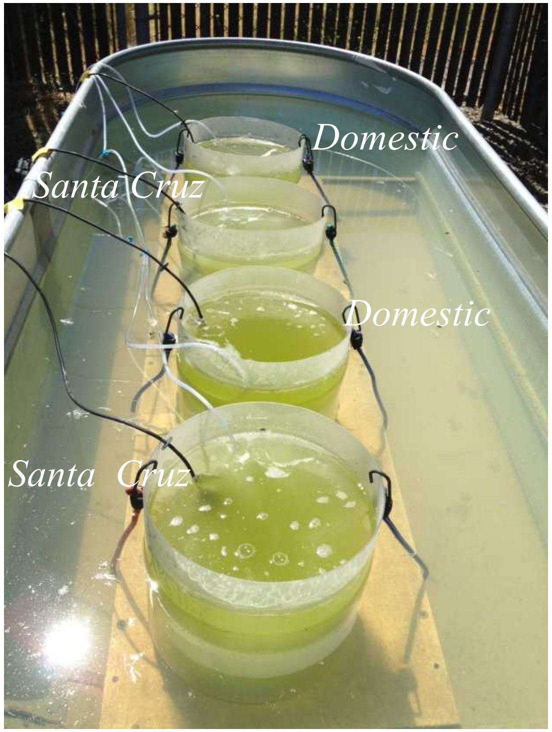




Stable algal communities are complex systems whose stability emerges from the collective behaviors of interacting species

Understand key principles that maintain stable algal cultures in pond systems and apply this knowledge to create stable microbial consortia

System-level analysis of algae-bacteria interactions

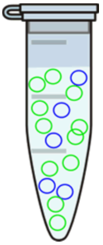


Environment data

Temp, pH, NH_3, O_2, LUX, PAR

DNA isolation

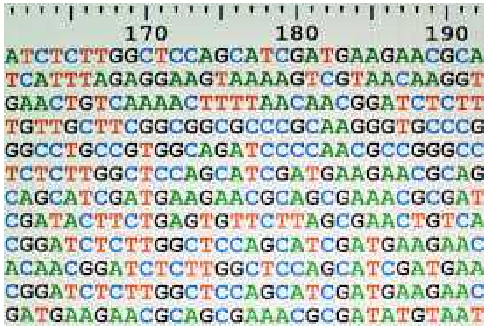
↓ 16S rDNA amplification



Microbial 16S rDNA

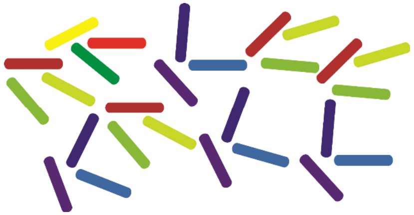
↓ Next generation sequencing

Microbial sequences

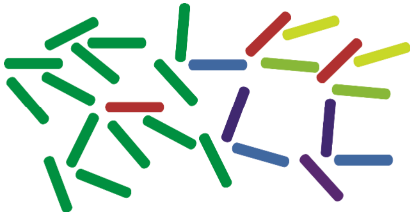


```
      170       180       190
ATCTCTTGGCTCCAGCATCGATGAAGAACGCA
TCATTTAGAGGAAGTAAAAGTCGTAACAAGGT
GAACTGTCAAAACTTTTAACAACGGATCTCTT
TGTTGCTTCGGCGGCGCCCGCAAGGGTGCCCG
GGCCTGCCGTGGCAGATCCCCAACGCCGGGCC
TCTCTTGGCTCCAGCATCGATGAAGAACGCAG
CAGCATCGATGAAGAACGCAGCGAAACGCGAT
CGATACTTCTGAGTGTTCTTAGCGAACTGTCA
CGGATCTCTTGGCTCCAGCATCGATGAAGAAC
ACAACGGATCTCTTGGCTCCAGCATCGATGAA
CGGATCTCTTGGCTCCAGCATCGATGAAGAAC
GATGAAGAACGCAGCGAAACGCGATATGTAAT
```

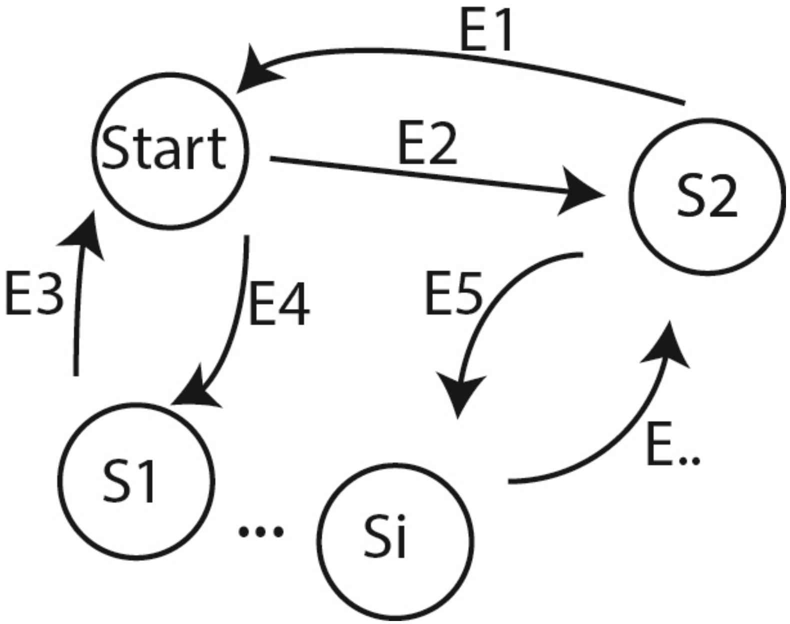
Microbial community modeling of biological state and transition



State i



State j



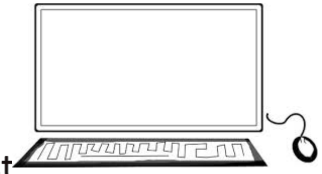
Markov chain

Random process

"memorylessness"

16S amplicon bioinformatics framework

Pick OTUs



cluster similar sequences:
cd-hit, blast, Mothur, uclust

Representative sequence/OTU

BLAST; RDP; RTAX

Assign taxonomy

de novo: MUSCLE;
reference based: PyNAST

Align sequence

Filter alignment

Build tree

Build an OTU table

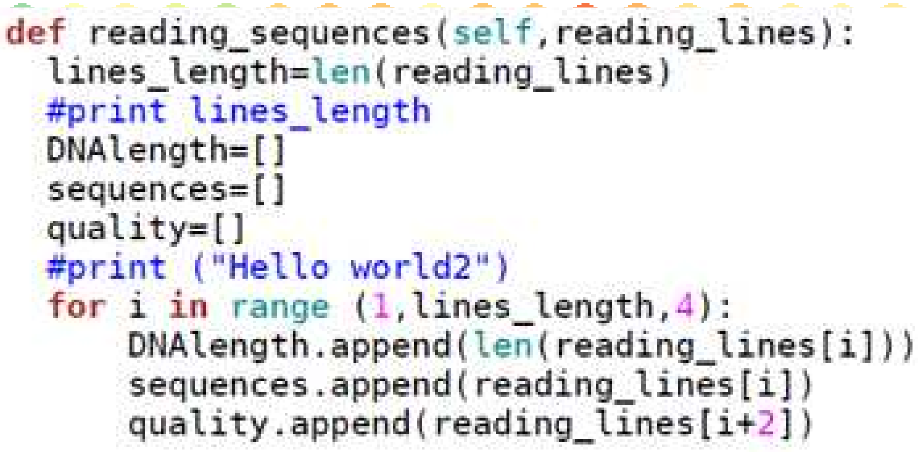
OTU table: 47 X 130

	SC3day2	D2day4	SC3day4	...
f_*Alcanivoracaceae*	167	280	102	...
f_*Flavobacteriaceae*	10	366	31	...
	...	...	...	

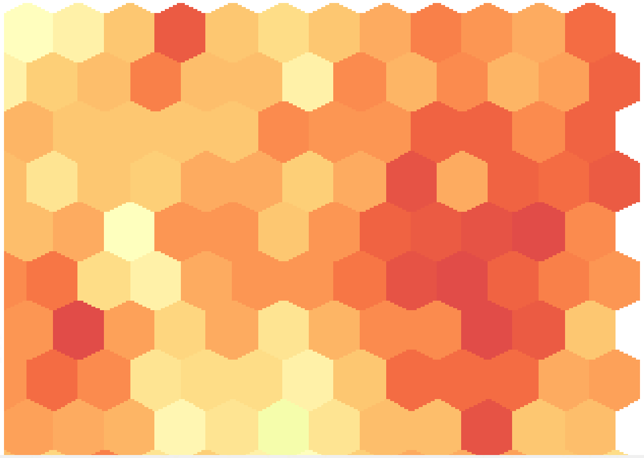
16S Operational Taxonomic Units (OTUs)

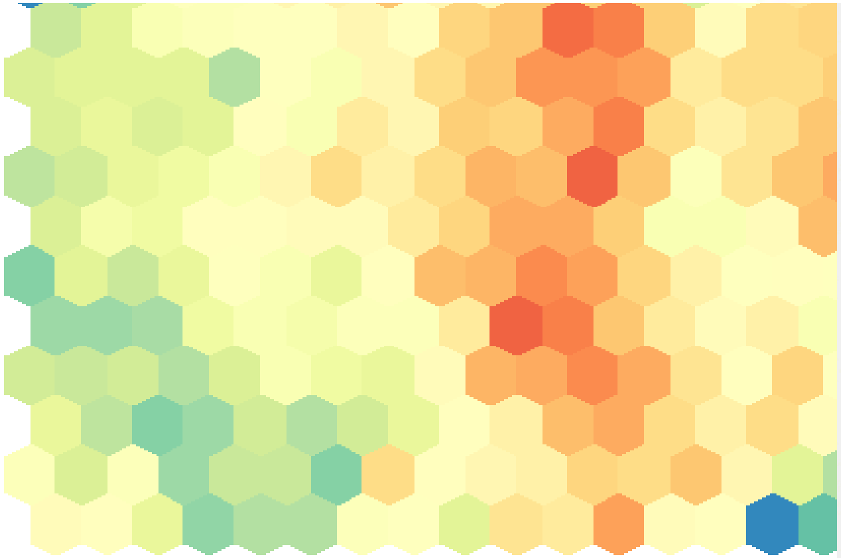
QIIME, Nature Methods 7, 335-336 (2010)

Metadata informatics integration from biology to data science



```
def reading_sequences(self,reading_lines):
    lines_length=len(reading_lines)
    #print lines_length
    DNAlength=[]
    sequences=[]
    quality=[]
    #print ("Hello world2")
    for i in range (1,lines_length,4):
        DNAlength.append(len(reading_lines[i]))
        sequences.append(reading_lines[i])
        quality.append(reading_lines[i+2])
```



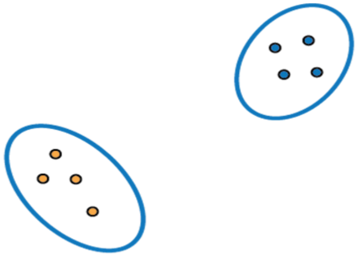


Metadata sets in microalgae microbtiota

High-dimensional

Multiscale

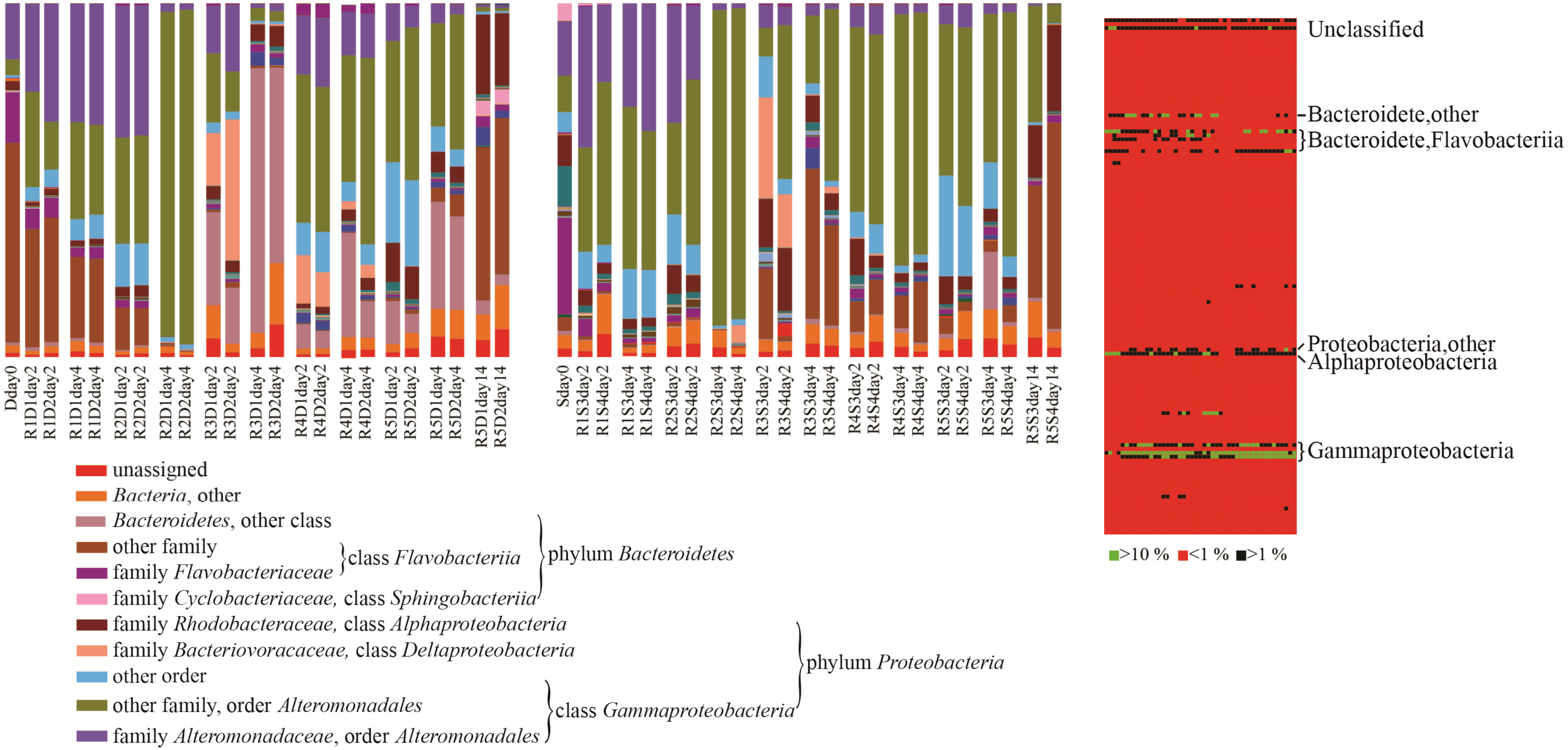
Informative Integrative Interpretative



Effective Insight

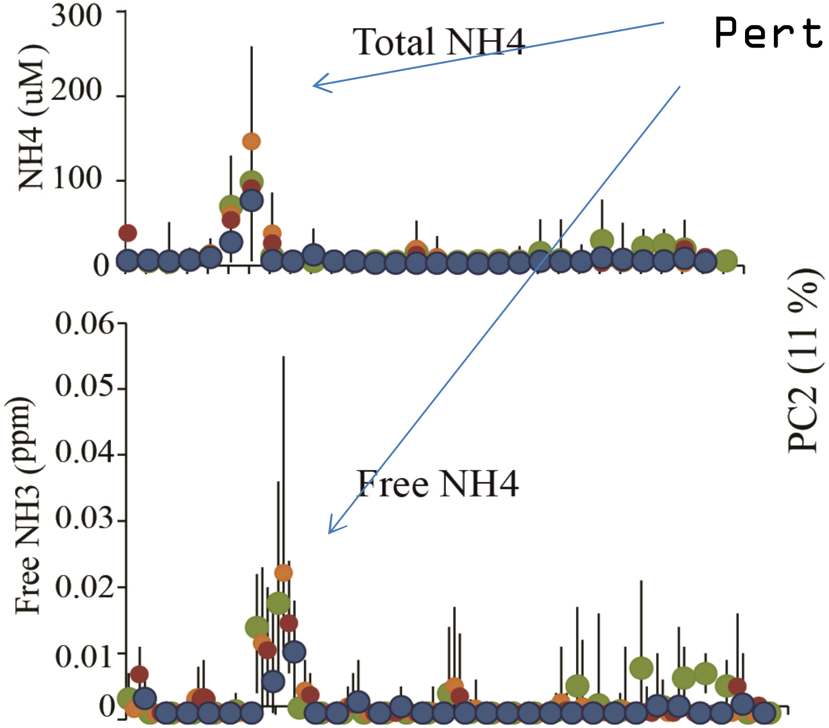
Valuable Innovation

Three major phylotypes in *N. salina* microbiota

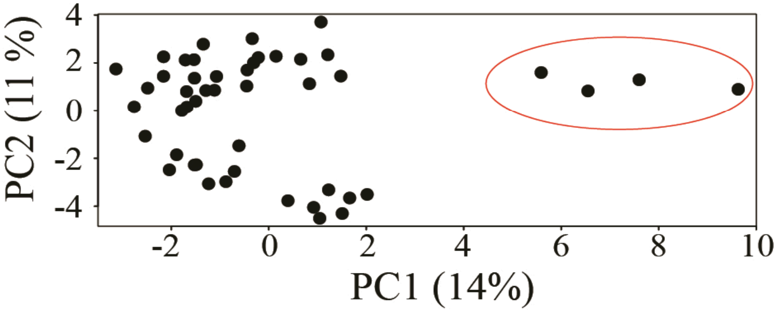


Geng et al. ***Microb Ecol***. 2016 Jul;72(1):14-24

Changes in microbiota diversity corresponds to environmental disturbance



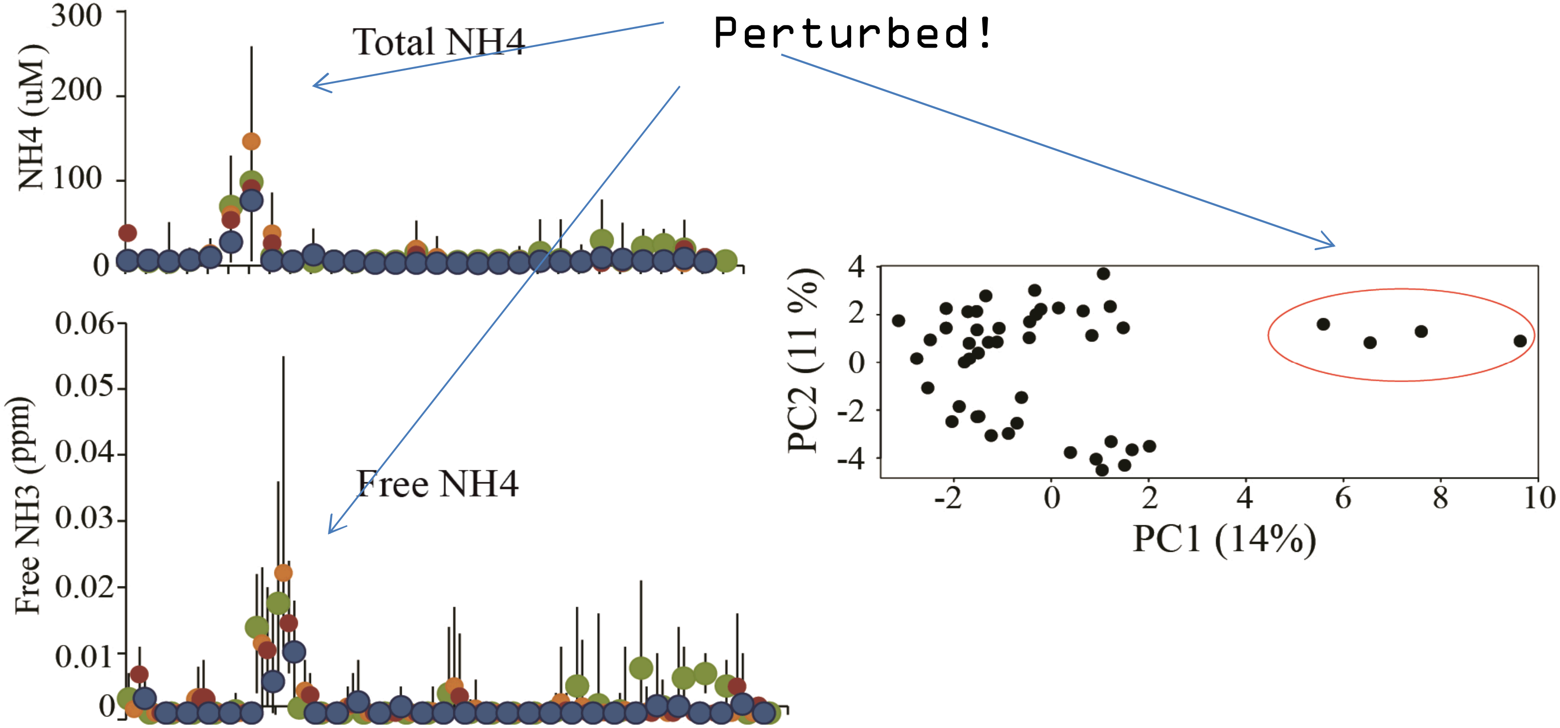
Perturbed!



- ❑ Profiles of total and free ammonia
- ❑ PCA ordination of the unweighted UniFrac matrix
- ❑ A **disturbance** in community **occurs** in samples, where **total and free ammonia spiked:** following disturbance, the community **rebound back** to sample cluster

Geng et al. ***Microb Ecol***. 2016 Jul;72(1):14-24

Changes in microbiota diversity corresponds to environmental disturbance

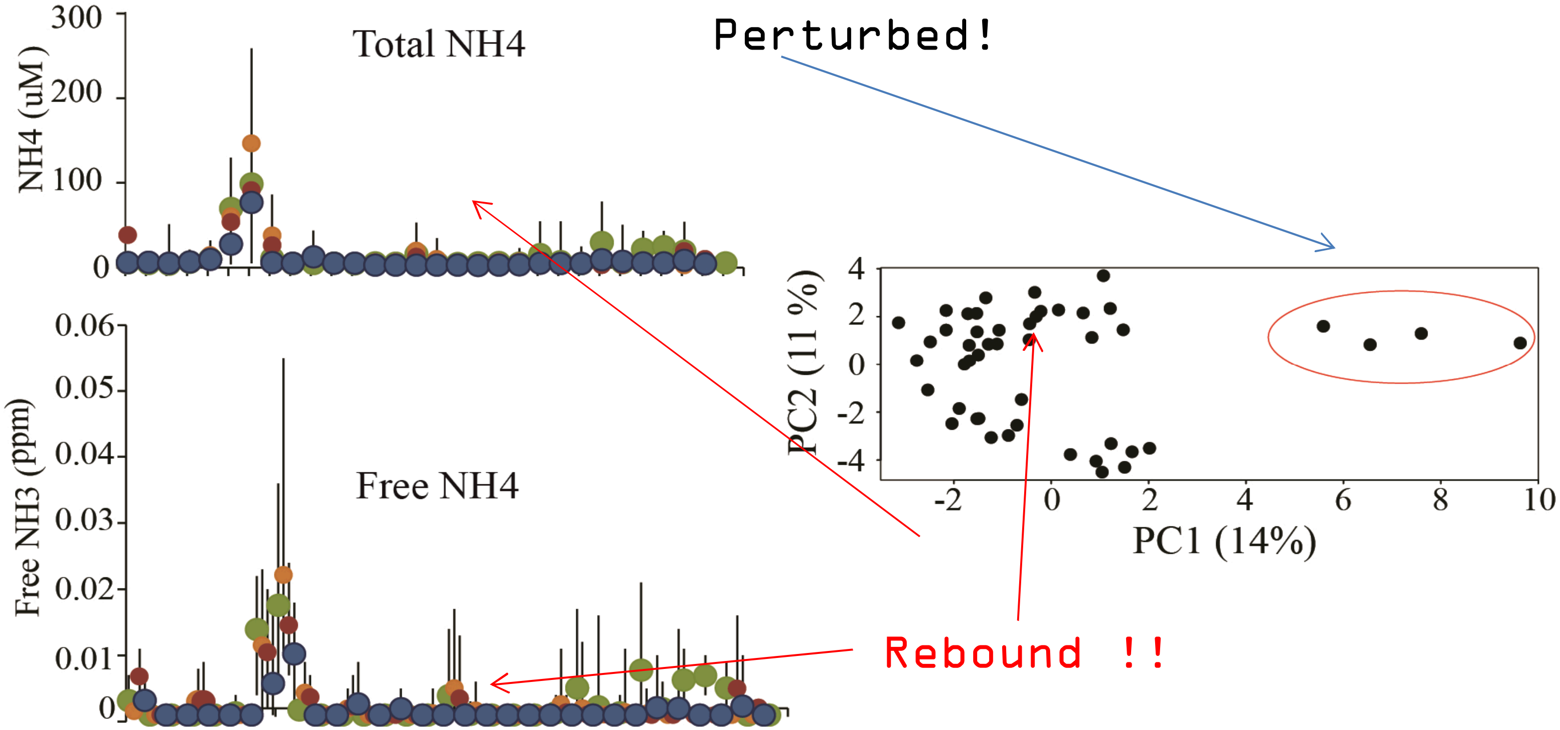


Perturbed!

- Profiles of total and free ammonia
- PCA ordination of the unweighted UniFrac matrix
- A **disturbance** in community **occurs** in samples, where **total and free ammonia spiked**; following disturbance, the community **rebound back** to sample cluster

Geng et al. ***Microb Ecol***. 2016 Jul;72(1):14-24

Changes in microbiota diversity corresponds to environmental disturbance



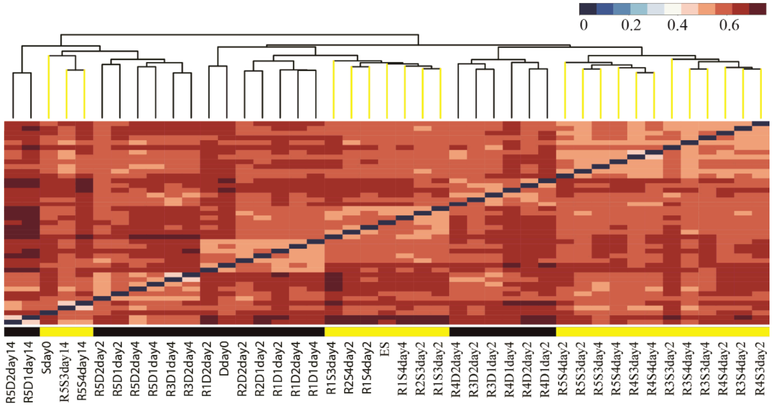
Perturbed!

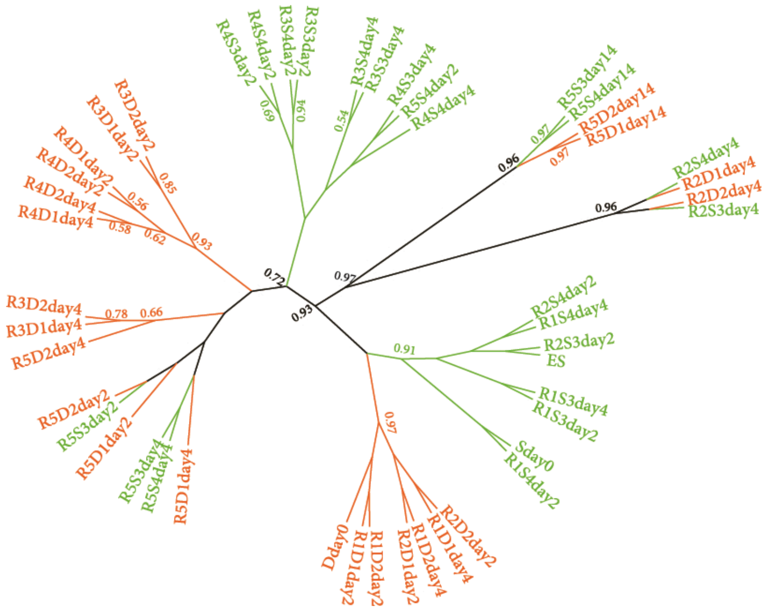
Rebound !!

- Profiles of total and free ammonia
- PCA ordination of the unweighted UniFrac matrix
- A **disturbance** in community **occurs** in samples, where **total and free ammonia spiked**; following disturbance, the community **rebound back** to sample cluster

Geng et al. ***Microb Ecol***. 2016 Jul;72(1):14-24

Microbial community founder determines microbiota succession

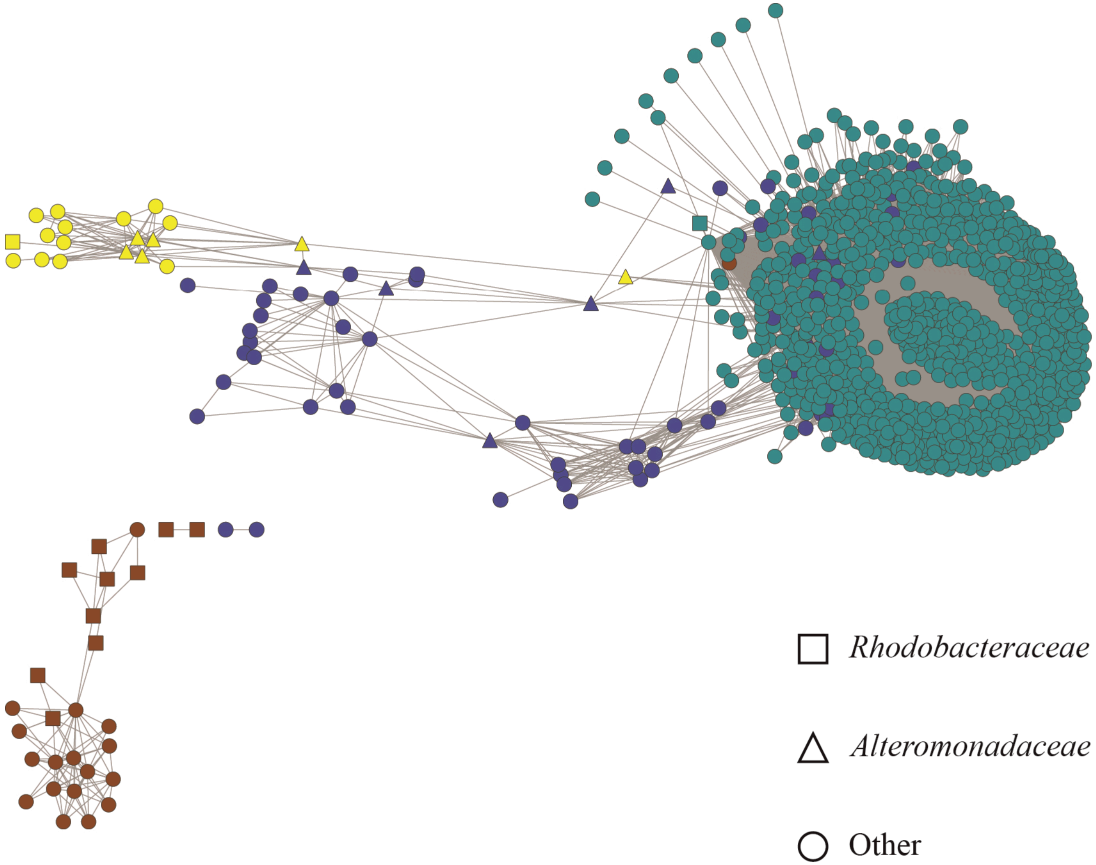




- ❑ Heat map showing pair-wise comparison based on unweighted Unifrac distance matrix
- ❑ Unsupervised hierarchical clustering; Jackknifing of tree
- ❑ The outgroup samples in the tree are the ammonia-perturbed samples
- ❑ Clustering of microbial diversity (β-diversity) showing distinction between samples from different founder cohorts
- ❑ Demonstrated microbiota stability

Geng et al. ***Microb Ecol***. 2016 Jul;72(1):14-24

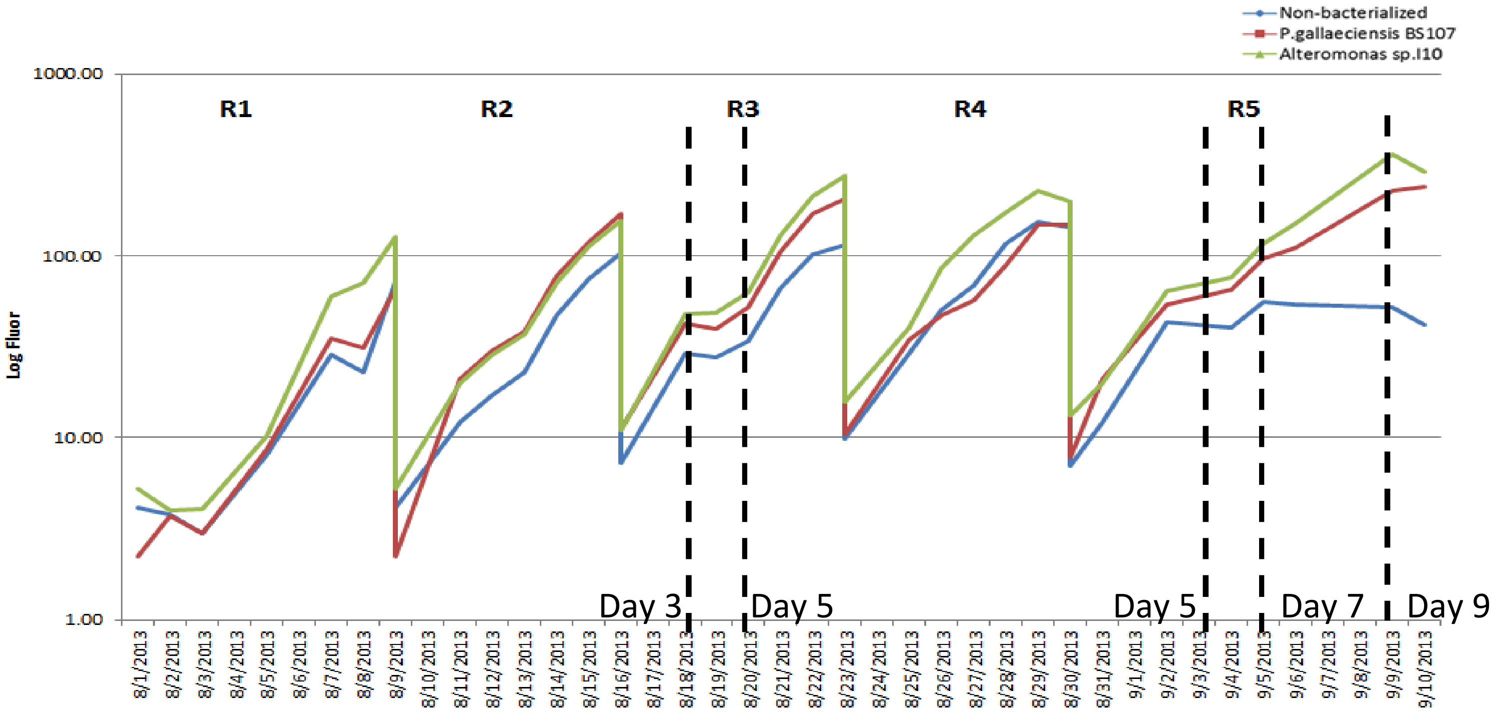
Microbiota stability property stems from formation of error-tolerance microbial community network

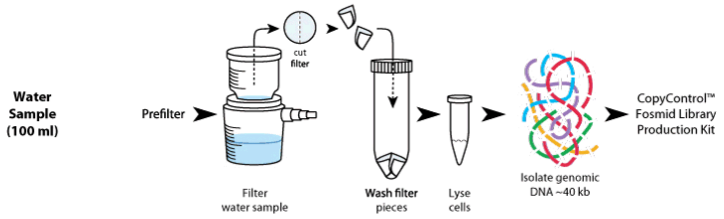


❑ Proposed "keystone" species for well-beings of algal cultures are from family *Rhodobacteraceae* and *Alteromonadaceae !*

Geng et al. ***Front Microbiol***. accepted

Whole-Genome Sequencing (WGS) high resolution quantitative characterization of microbiota




Water Sample (100 ml)

Prefilter

cut filter

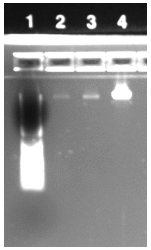
Filter water sample

Wash filter pieces

Lyse cells

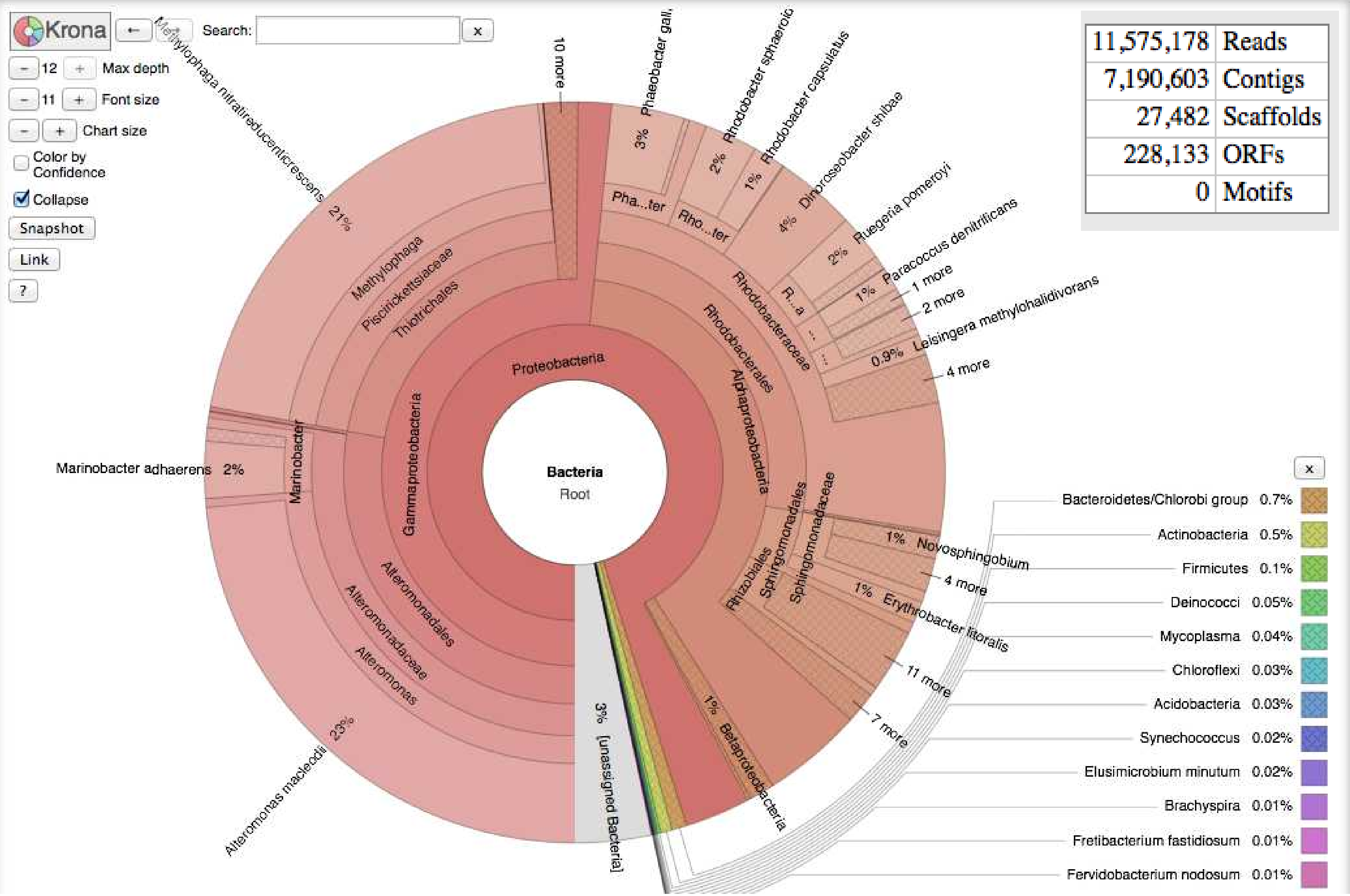
Isolate genomic DNA ~40 kb

CopyControl™ Fosmid Library Production Kit



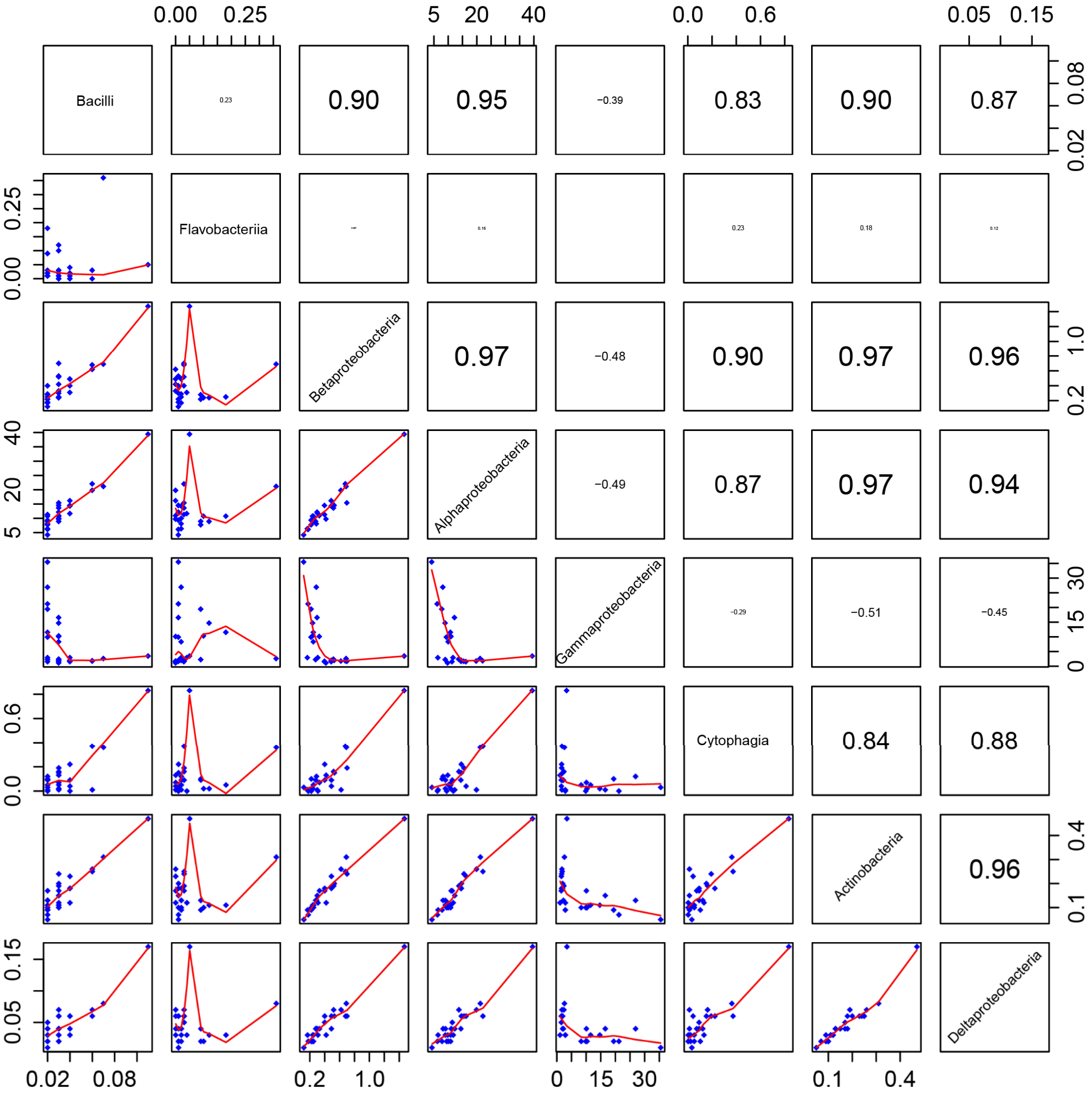
8-week outdoor cultivation period
(20L mesocosms)

Species and genes finding with whole genome shotgun



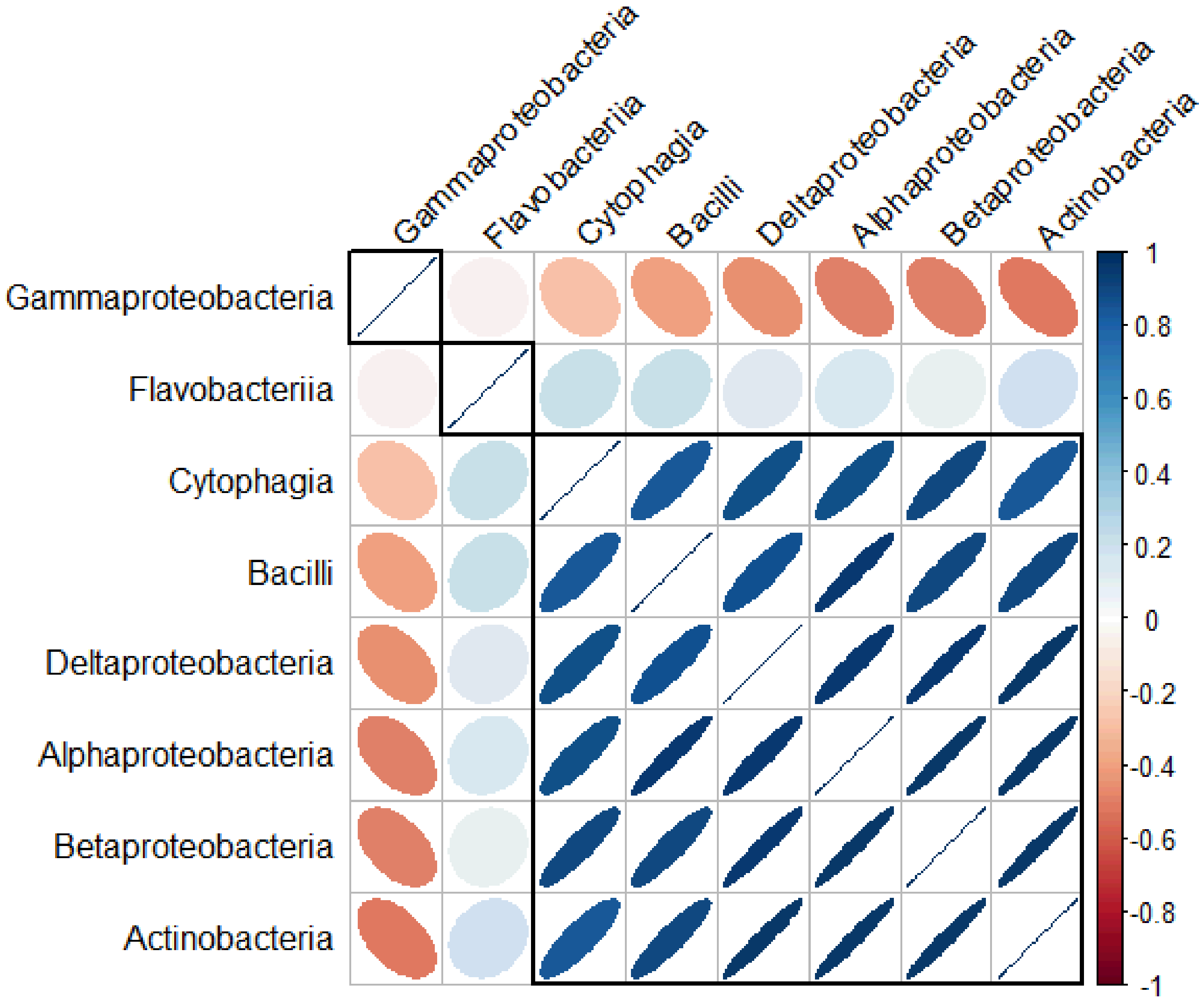
11,575,178	Reads
7,190,603	Contigs
27,482	Scaffolds
228,133	ORFs
0	Motifs

Microbial classes profiling cross samples



- 26 samples x 39 classes
- Cutoff for eight bacterial classes:
- Minimal abundance 1 % at least 50 % of samples
  - Bacilli,
  - Flavobacteriia,
  - Betaproteobacteria
  - Alphaproteobacteria
  - Gammaproteobacteria
  - Cytophagia
  - Actinobacteria
  - Deltaproteobacteria

Correlation between eight major bacterial classes in *N. salina* microbiota



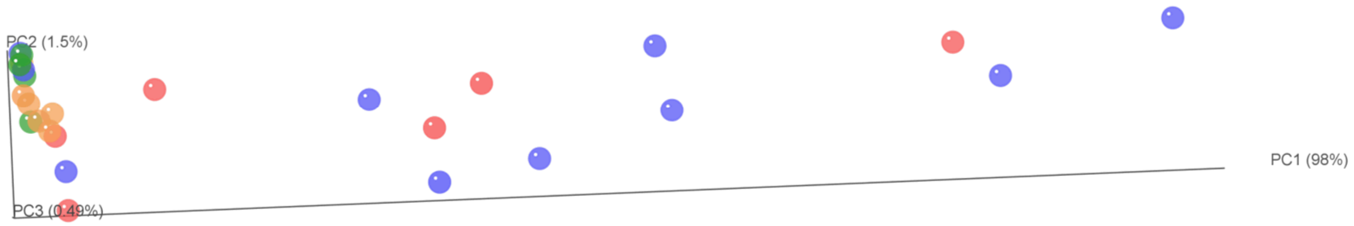
Microbial community patterns transitions between early and sessile phases

sessile

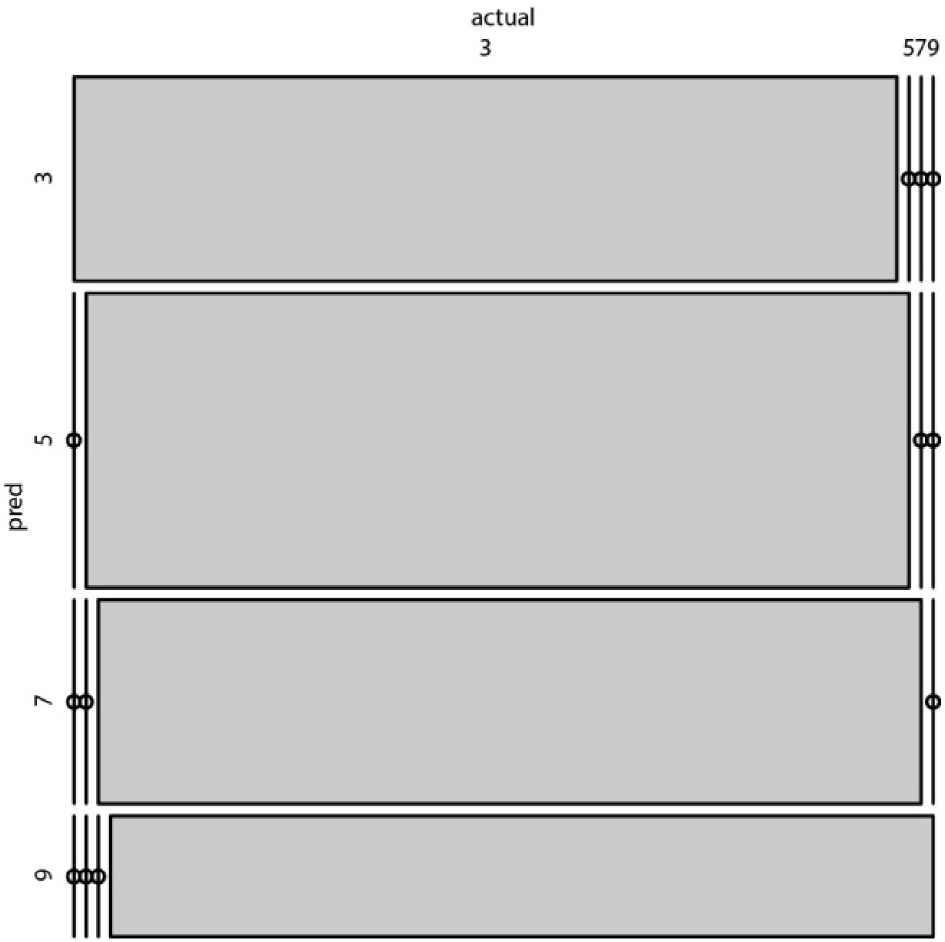
Microbial community patterns transitions between early and sessile phases

sessile

early



Biphasic lifestyle classified by classification tree in machine learning

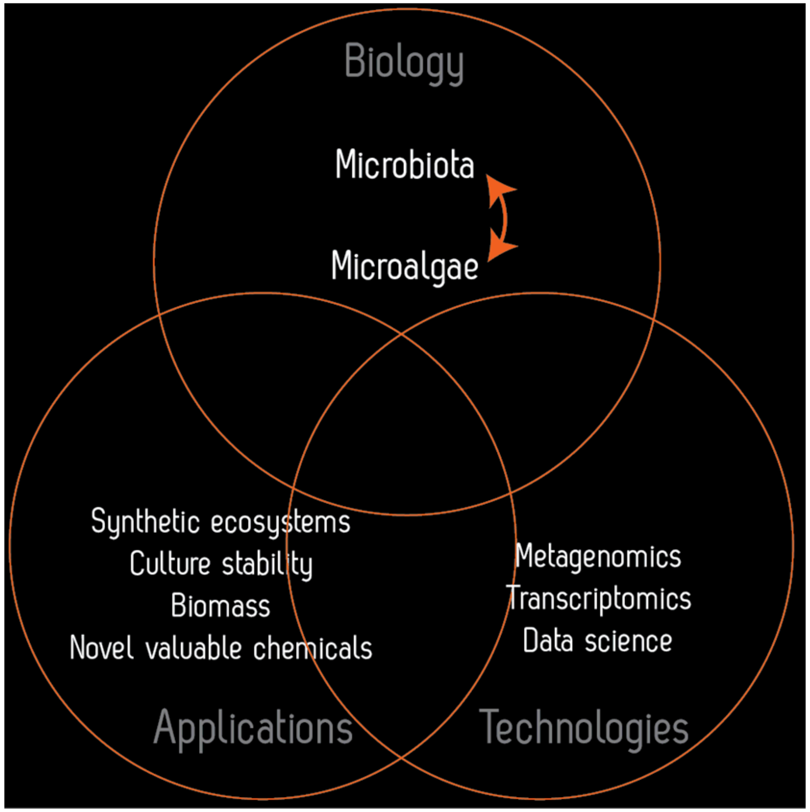


Training

```
> confusion_table
  Biom_pred
   3 5 7 9
 3 0 0 0 0
 5 1 1 0 0
 7 0 0 1 0
 9 0 0 0 1
```

Testing: 75 % accuracy

Summary



- Informative, integrative, and interpretative to quantitative analysis of metagenomics metadata sets
- Developed and and implemented algae systems biology tools
- Community stability and resilience.
- Microbial complexity network prone to small-world network: sparse matrix error tolerance
- Machine Learning Learnability

Acknowledgments

- Sandia National Laboratories
  - Todd W. Lane
  - Kenneth L. Sale
  - Mary Bao Tran-Gyamfi
  - Masood Z. Hadi
  - Eizadora T. Yu
  - Ryan Davis
- Sequencing Centers
  - Sandia National Laboratories
  - QB3-Berkeley
- Institutional Support
  - Sandia LDRD
  - Department of Energy (DOE) Genomic Science Program

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.