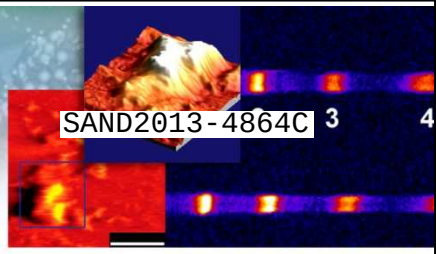
SANDIA BIOSCIENCE



Pond Crash Forensics:
Microbiome Analysis and Field Diagnostics

Todd W. Lane Ph.D.
Sandia National Laboratories

3rd International Conference on Algal Biomass, Biofuels & Bioproducts
19 June 2013

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.

U.S. DEPARTMENT OF ENERGY | NNSA National Nuclear Security Administration

U.S. DEPARTMENT OF ENERGY | Energy Efficiency & Renewable Energy
BIOENERGY TECHNOLOGIES OFFICE

Sandia National Laboratories

The difference between short term & sustained productivity is the problem

Short term areal production of 30-50 g/m^2/day
–Commonly claimed

Annualized areal production rates of 13.2 g/m^2/day: ANL, NREL, PNNL 2012

Sub optimal growth conditions lead to lower annualized production.:

- Irradiance, temp, salinity etc.
- Pond/PBR collapses caused in part by biological agents (ultimate suboptimal condition)

Real time data on predator/pathogen load enables proactive pond management: We intend to create tools to enable such management

Presence of the biological agent can be necessary but not sufficient to crash

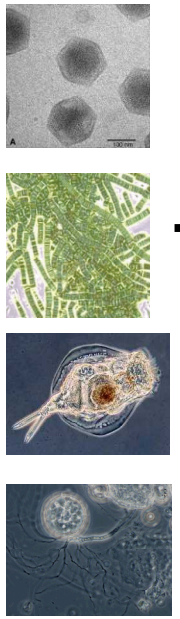
Agent *Algae* *Environment* *Collapse*

Viruses

Bacteria

Predators

Fungi



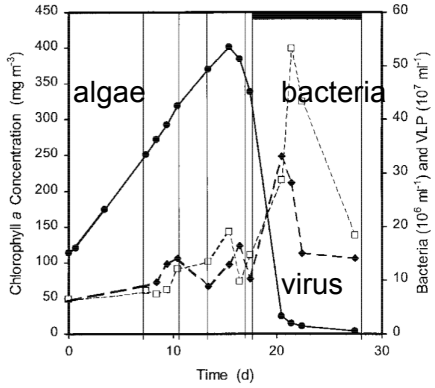
Patterson & Laderman, 2001.

+

+

Environment (Temp, salinity, pH, CO2, nutrients)

=



Herman Gons et al., Antonie van Leeuwenhoek, 81: 319-326, 2002.

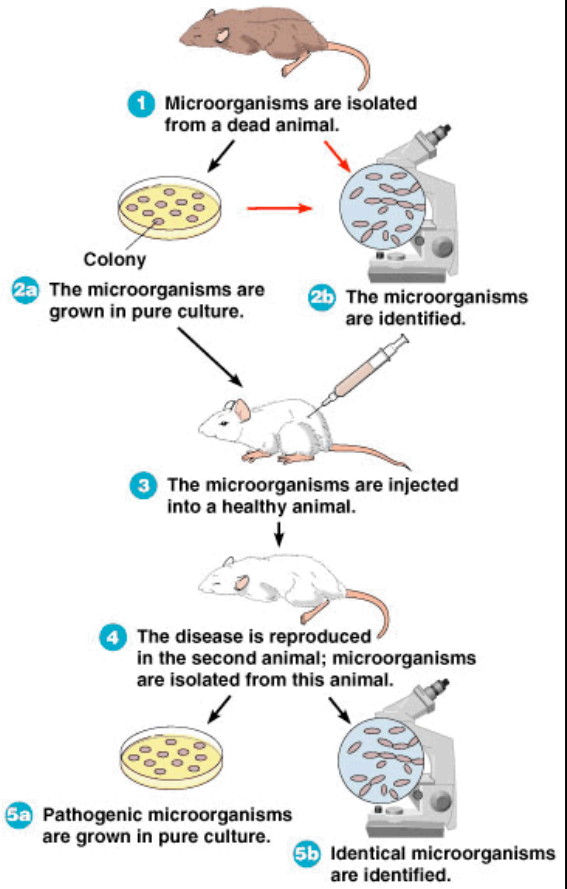
"Perhaps the most worrisome component of the large-scale algal cultivation enterprise is the fact that algal predators and pathogens are both pervasive and little understood."

- DOE Draft Algal Biofuels Technology Roadmap (2009)

SANDIA BIOSCIENCE

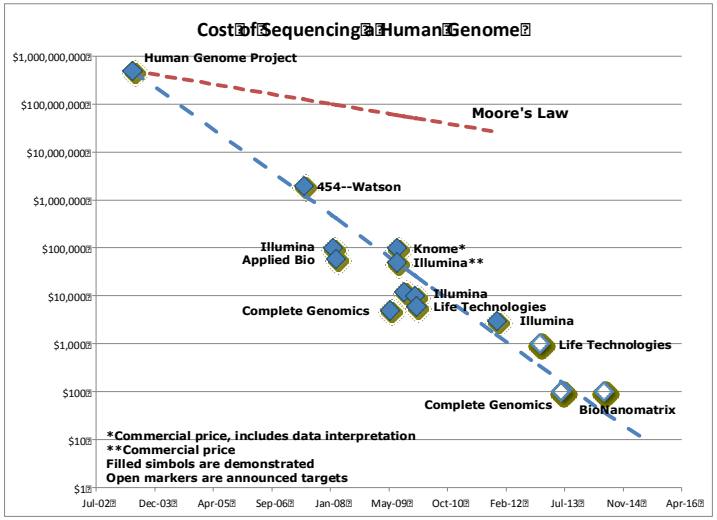
A staged approach to crash agent identification (Koch's postulates)

- **Presumptive Identification**
  - Detect the presence of the agent in crashed ponds
  - Agent absent or in lower abundance in healthy ponds
  - Complicated by environmental parameters
- **Confirmatory Identification**
  - Isolate the agent(s) and recapitulate the crash
  - Complicated by environmental parameters
- **Development of Field Assay**
  - Quantitative yet rapid, simple and cheap (dipstick assay)
- **Early detection**



Sequencing can provide presumptive identification of pond crash agent.

- The cost of next gen sequencing is falling at a rate that outstrips Moore's law
- Cost of human genome fallen by 1/2 – 2/3 since January 2011
- The amount of data per run is increasing dramatically
- Bar-coding allows full advantage of this capacity
- Key is to get more sequencing hits on target thus reducing the cost of analysis to $10s

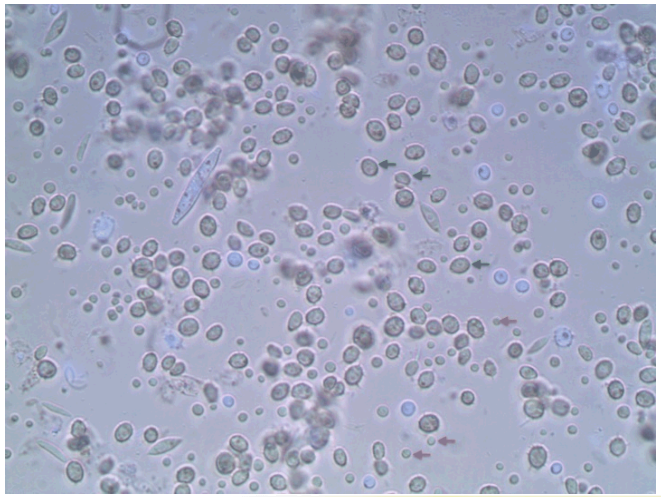


MiSeq

1X 100nt: 12hrs
340 Mb
2X150nt: 27hrs
> 1Gb

HiSeq

2X150nt: 8days
540 -600Gb



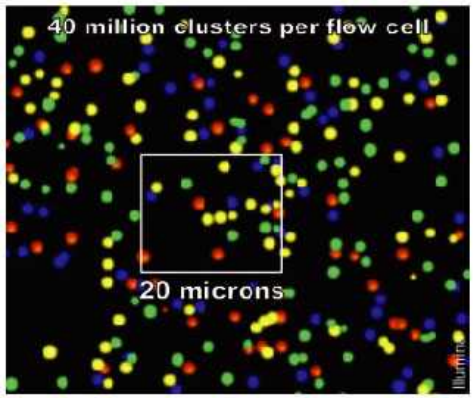
NASA OMEGA prom

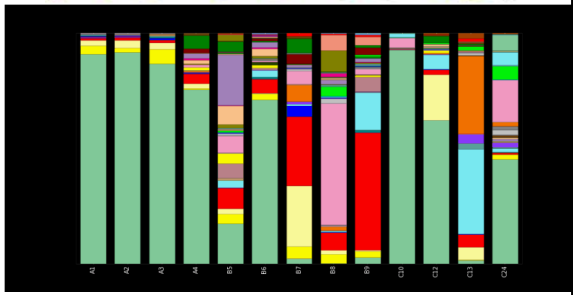
SANDIA BIOSCIENCE

Sandia National Laboratories

In the past year we have obtained samples (both healthy and infected) from a variety of collaborators

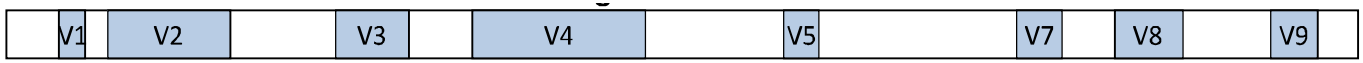
- Pilot/production scale ponds or PBRs
  - "Whole" water shipped overnight on ice
    - Culturing of agents
  - Frozen biomass/Purified nucleic acids
- Biomass harvested
- Viral fraction purified and concentrated
- Nucleic acids extracted
  - Prokaryotic analysis
  - Eukaryotic analysis
  - Viral analysis
- Library preparation and barcoding
- High throughput sequencing (Illumina)
- Bioinformatic analysis





SANDIA BIOSCIENCE

Euk SSU rRNA variable regions



- **Variable Region 4: ~ 400 bp amplicon (*S. cerevesiae*)**
  - **Significant length heterogeneity**
  - **Highest degree of variability**
  - **Generally thought to be sufficient to identify eukaryotes to the genus level**
  - **50-75 % coverage with 150bp PE (no overlap)**
  - **MiSeq 250 bp PE (Beta) 35 hours 10-14M reads**

- **Variable Region 9: ~168 bp amplicon (*S. cerevesiae*)**
  - **High variability**
  - **Used for genus level ID**
  - **MiSeq100 bp PE 14-19 hours 10-14M reads (PE)**

- Often used in combination

- Other molecular bar coding regions
  - Internal transcribed spacer (ITS): various length, allows for species level ID, applied in fungal phylogeny
  - MtCO

Bioinformatics pipeline

178 million raw paired-end reads (150PE)

In-house tools

Off-the-shelf tools

↳ Qfilter ↴

155 million quality-filtered & trimmed paired-end reads

↳ Bowtie2 ↴

145 million paired-end alignments to Silva rRNA database

silva
comprehensive ribosomal RNA databases

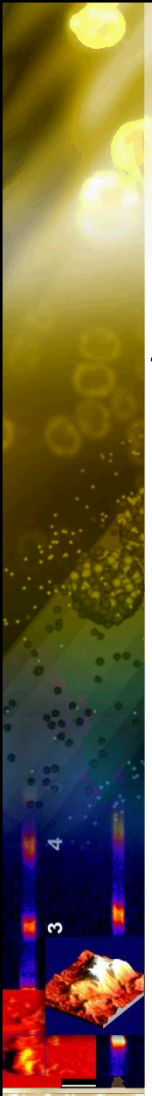
↳ sam2tax ↴

Silva results (all hits within 1% of best) mapped to NCBI taxonomy

↳ lca ↴

RapTOR

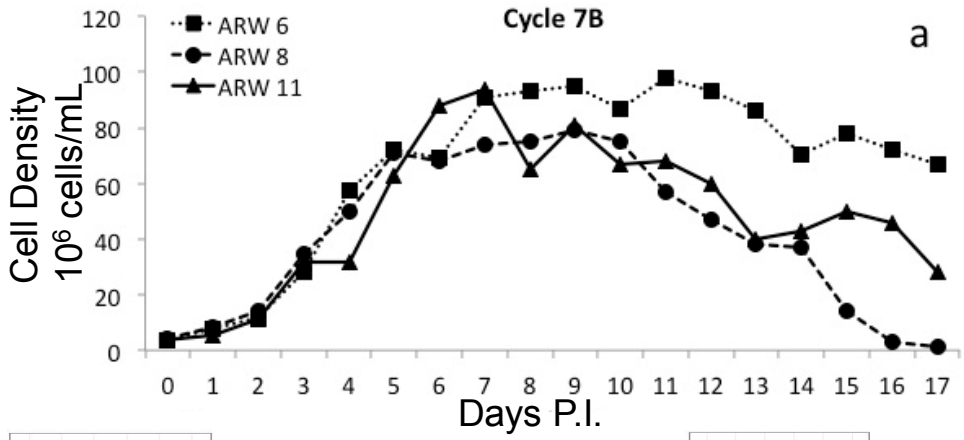
Last common ancestor based taxonomic summary for each hit:
132 million resolve at Genus level, 10 million at Species level

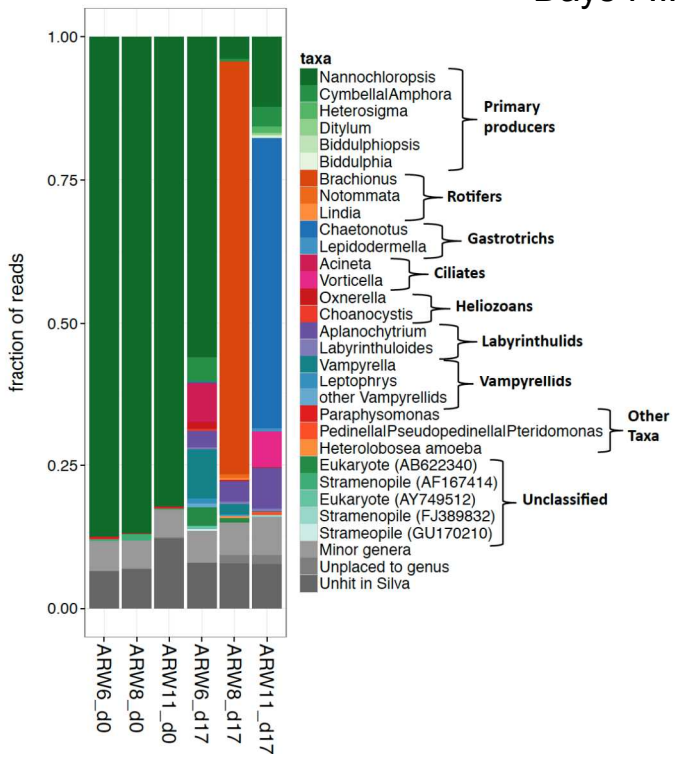


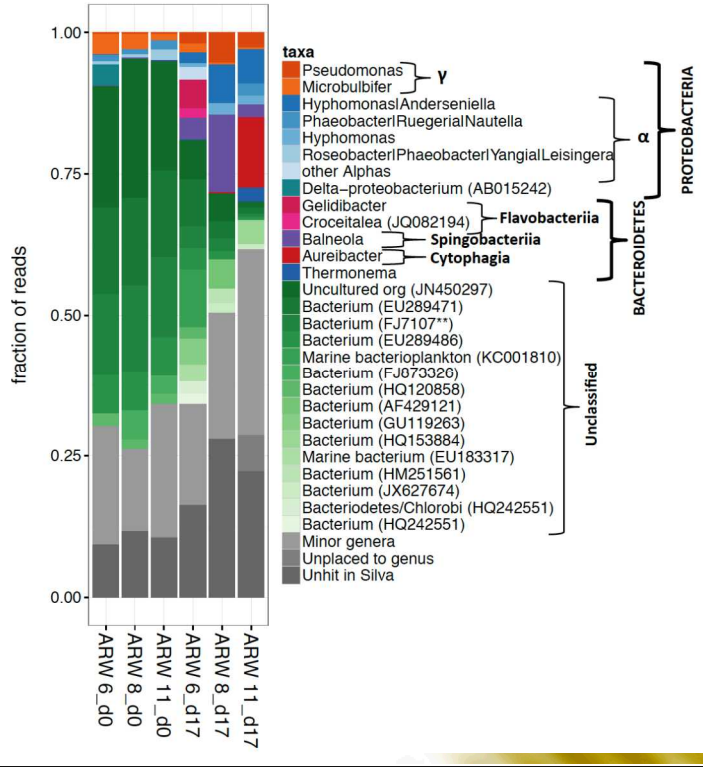
SANDIA BIOSCIENCE

Sandia National Laboratories

Forensic Analysis of Pond Crashes

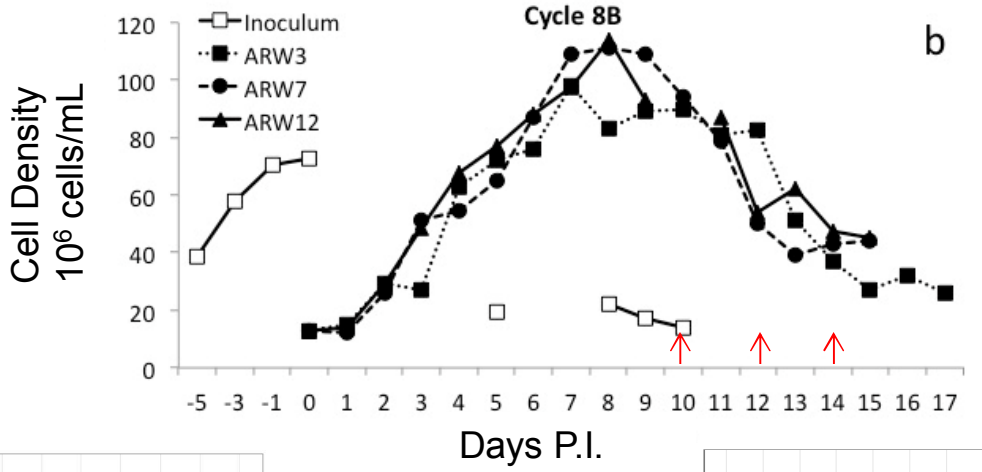


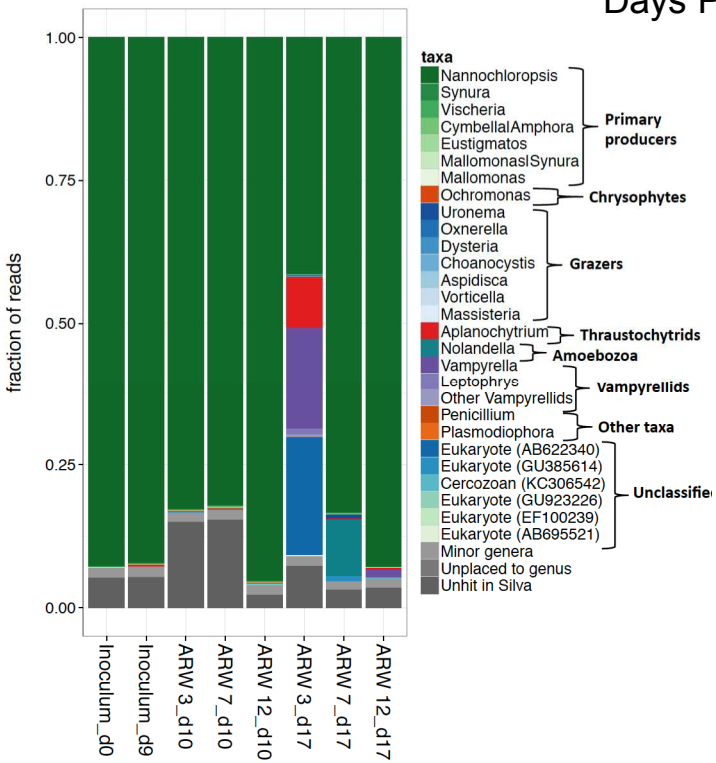


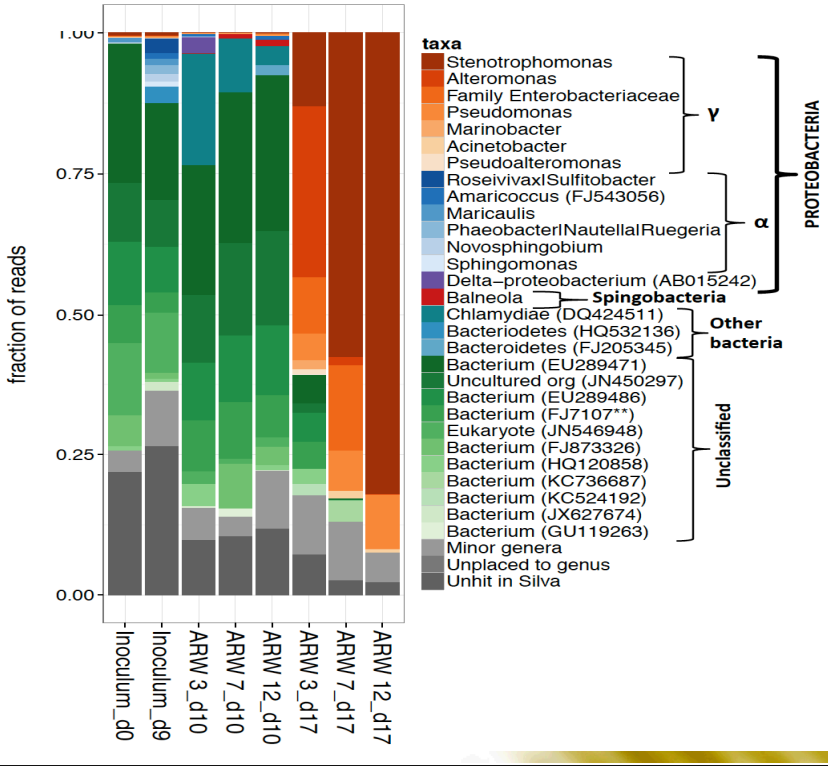


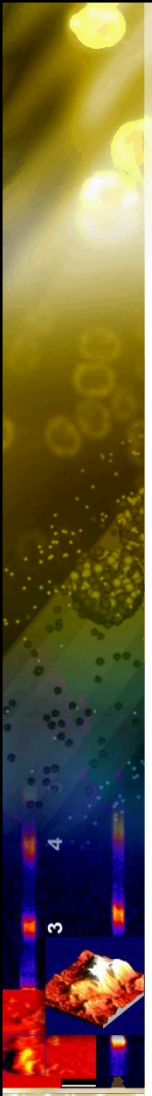
SANDIA BIOSCIENCE

Evaluation of pond intervention



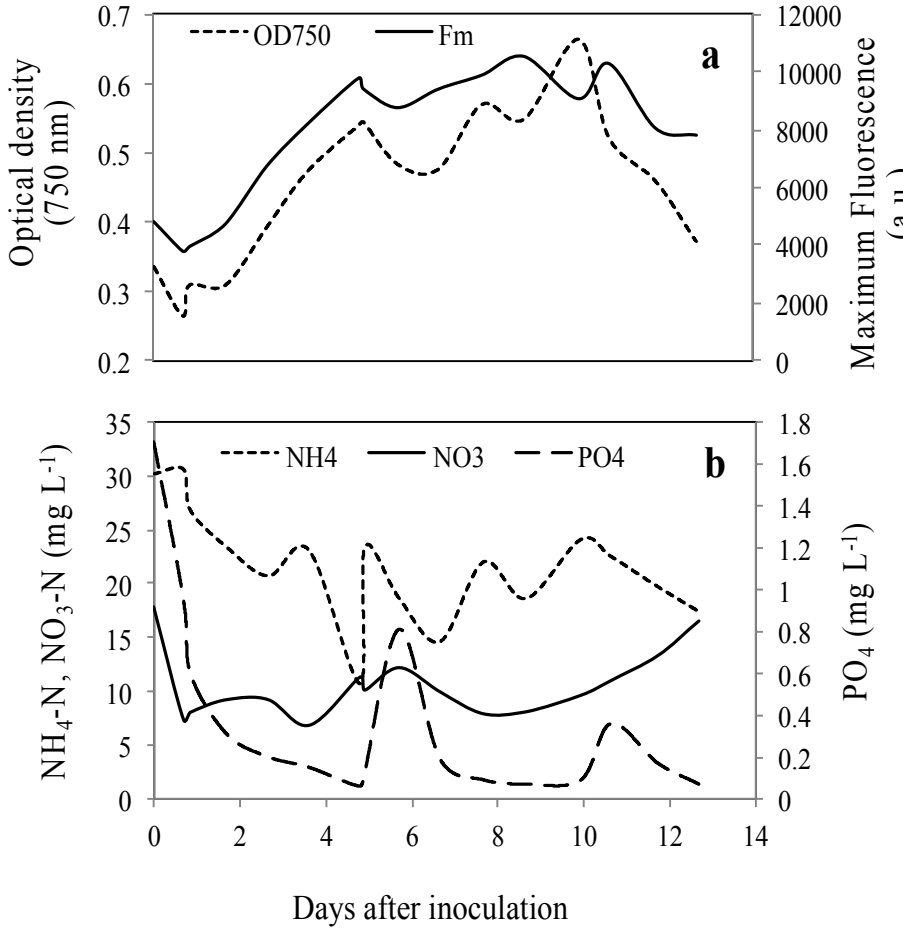






SANDIA BIOSCIENCE

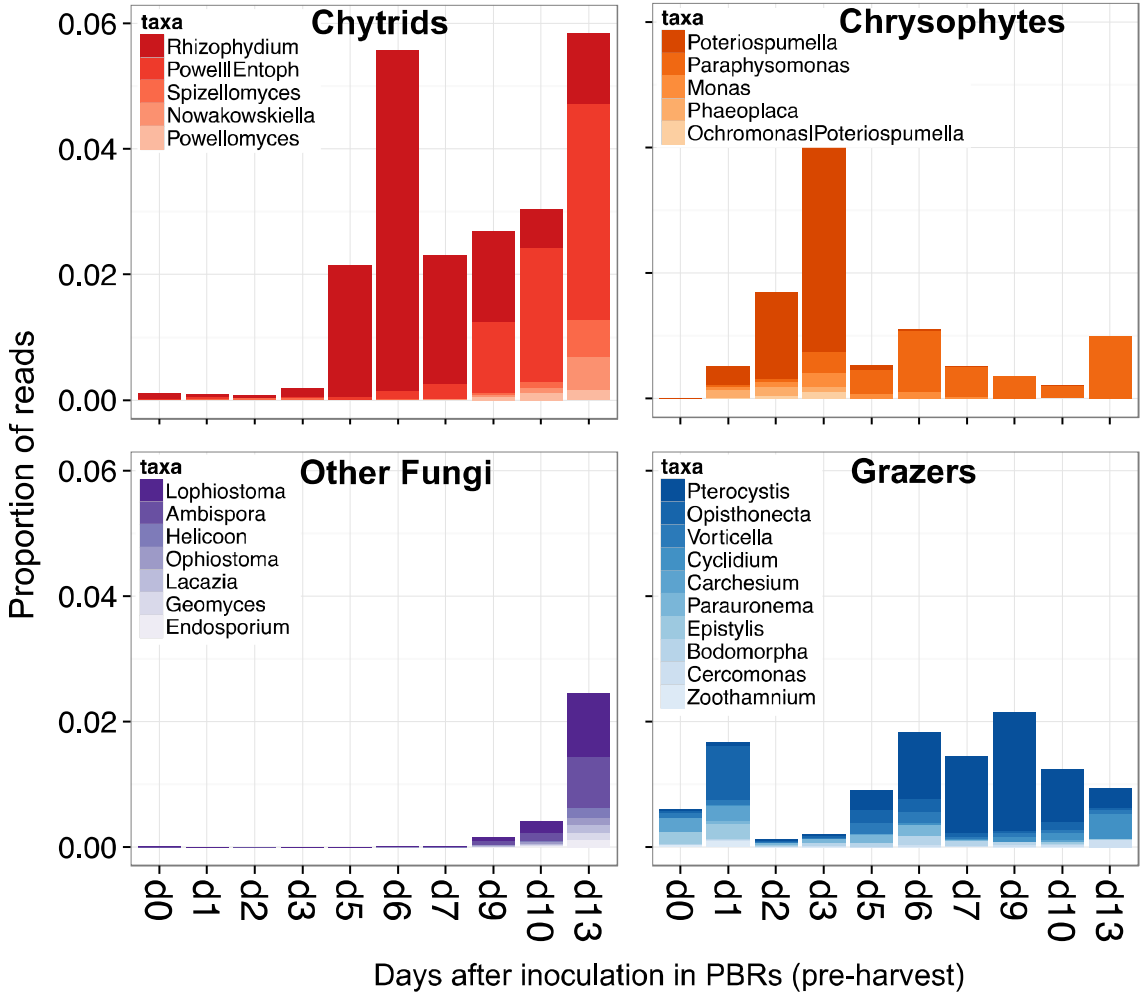
Identification of pests in OMEGA floating PBRs



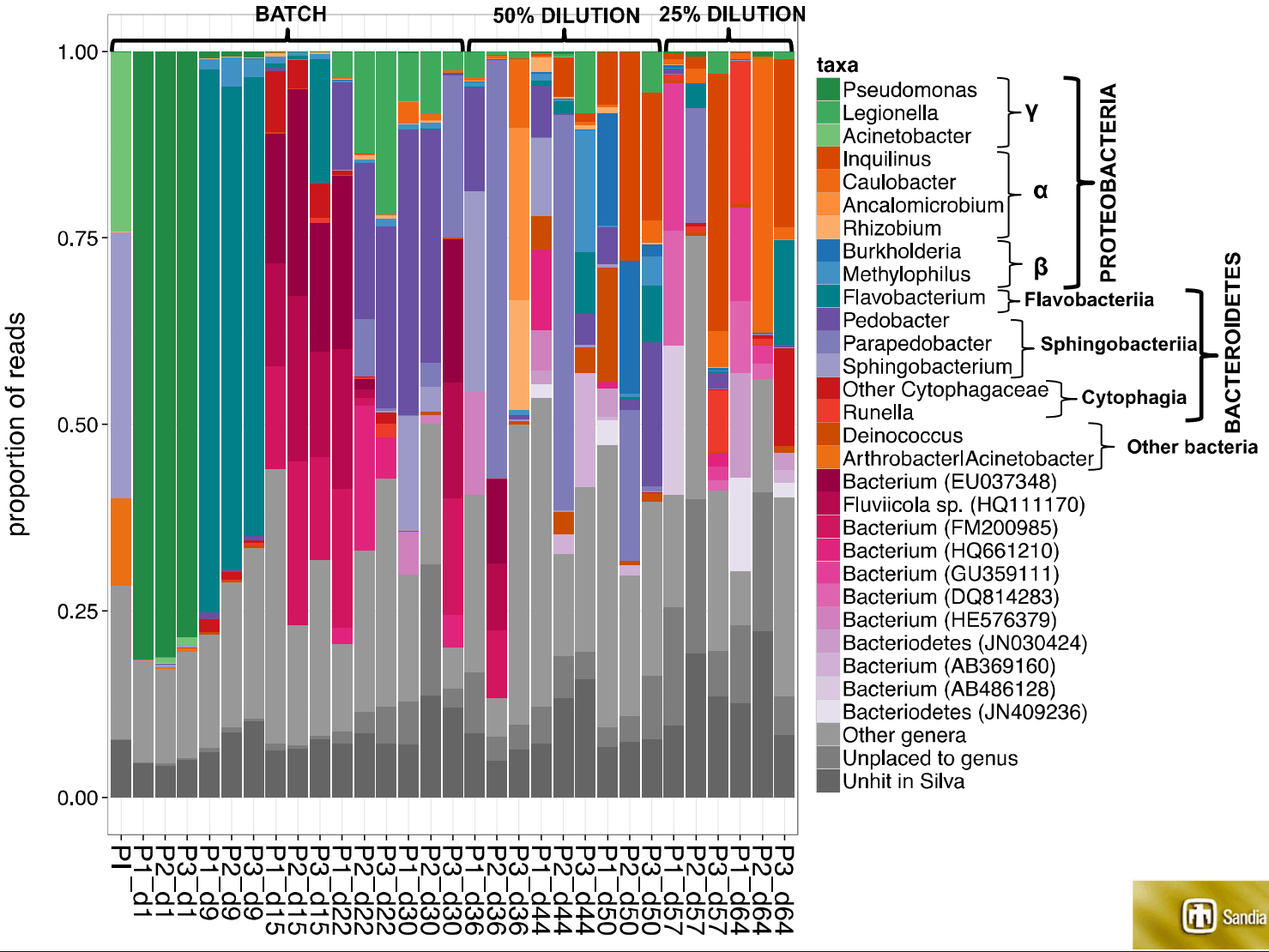
SANDIA BIOSCIENCE

Sandia National Laboratories

Most abundant Eukaryotic pest genera



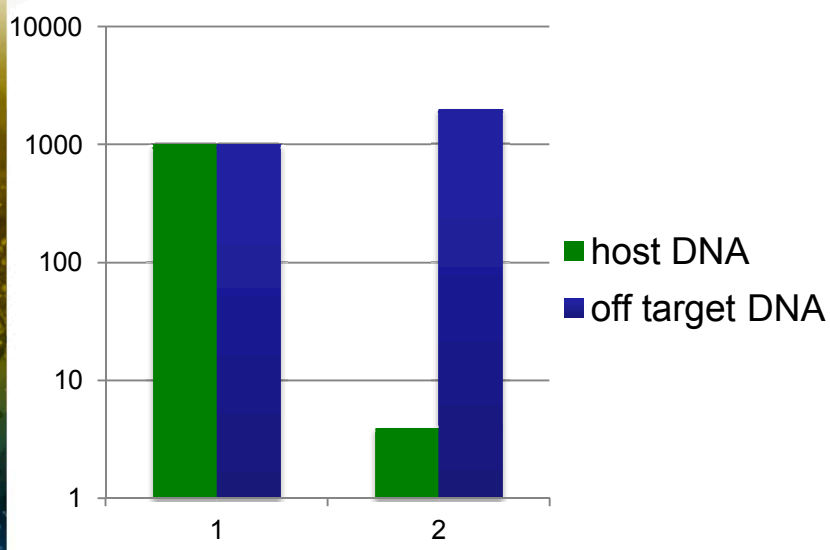
Long Term Microbiome Analysis Indicates Community Divergence

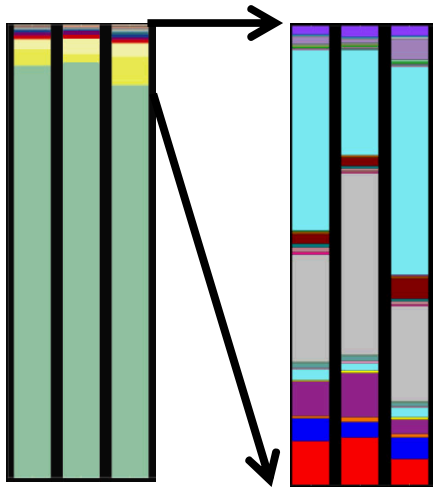


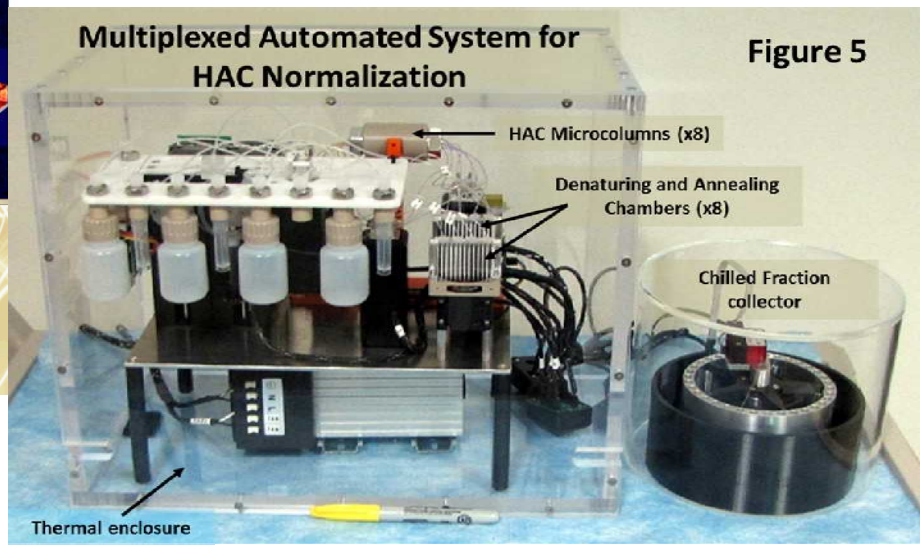
SANDIA BIOSCIENCE

Sandia National Laboratories

Automated, mutiplexed and manual systems for capture-based suppression







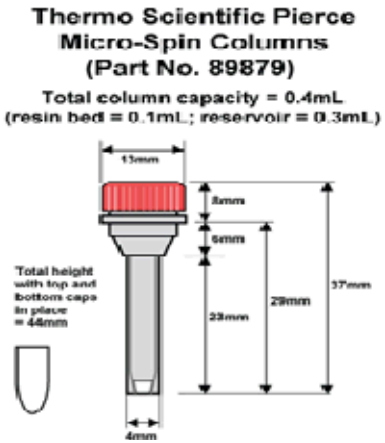
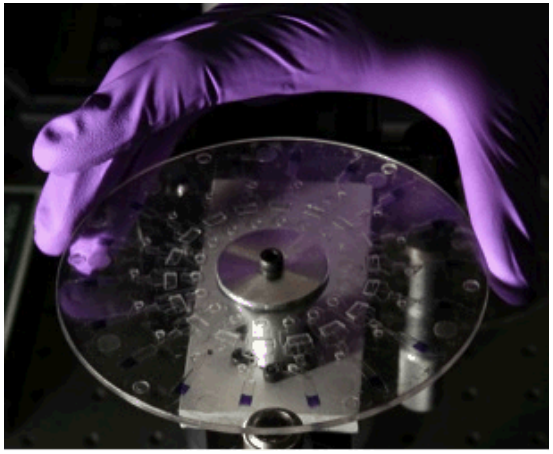
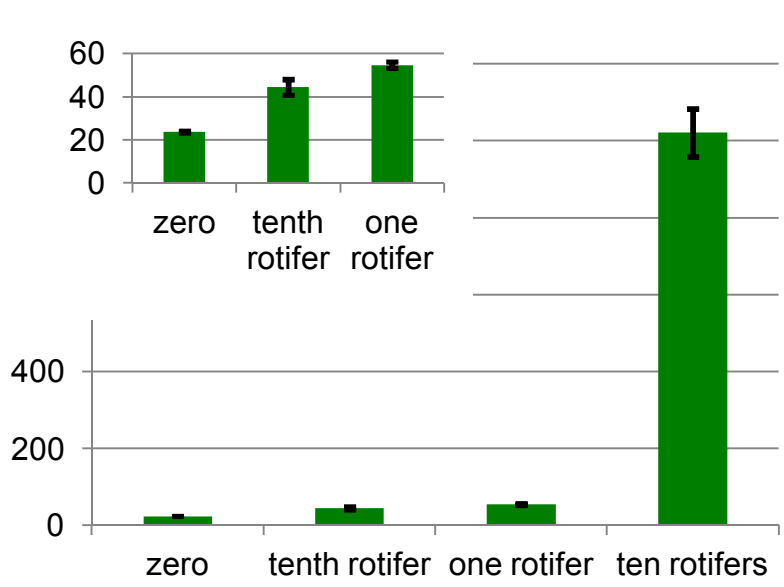


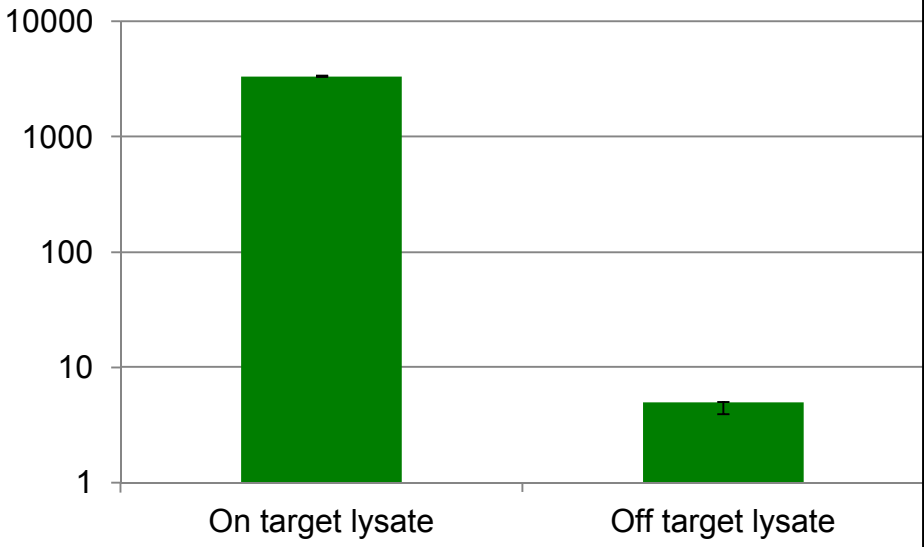
Figure 10: Dimensions of the commercially available spin columns

The goal of all this sequencing? Direct from the sequencer to assay.

- Target probes for predators pathogens and parasites in the pond
- SpinDX system of centrifugal fluidics and detection
  - Originally designed for clinical or environmental agent detection in “low resource” environments.
  - Rapid prototyping
  - Optimization
  - Validation
  - Genus level probes target agent and nearest neighbors

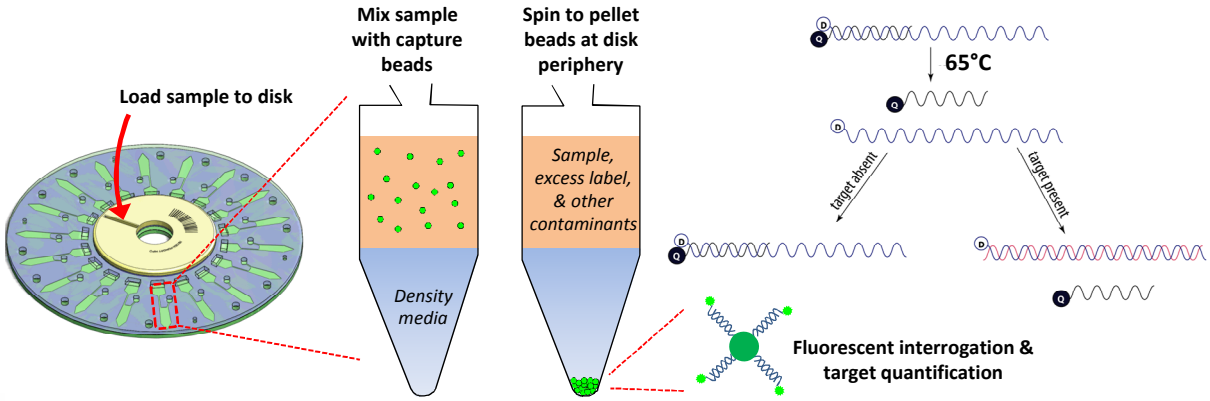


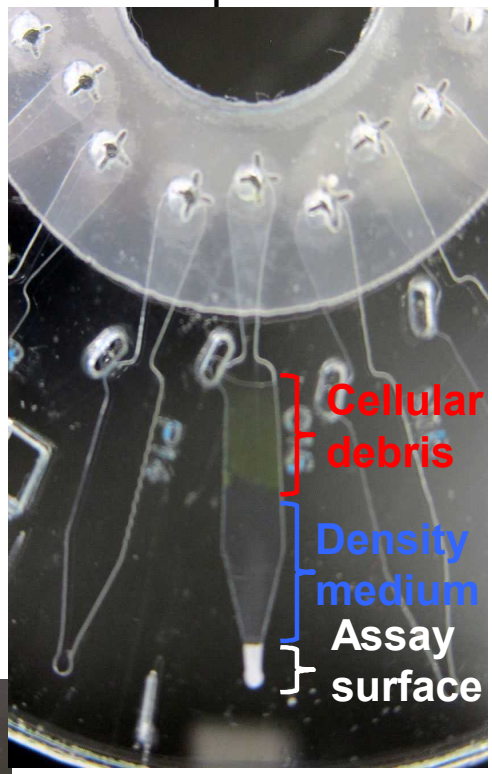




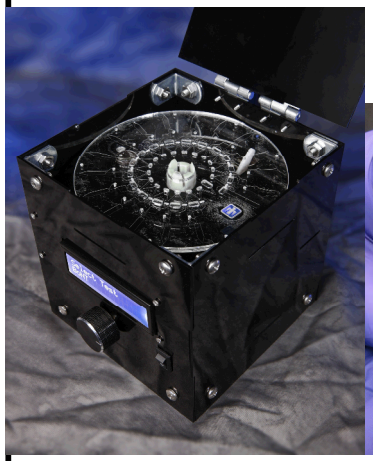
SpinDx™ has the necessary characteristics for a field assay for pond management

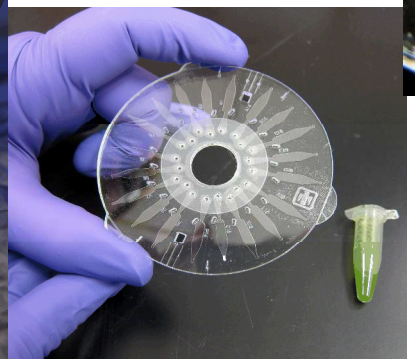
- FRET-based bead hybridization assay enabling capture and quantification of pathogen-specific RNA/DNA signatures





- Assay time: approx 30 min
- 36 channels per disc
- Potential for multiplexed assays in each channel
- Low reagent costs
- Low material costs
- Low instrument cost ($1000)
- Fieldable

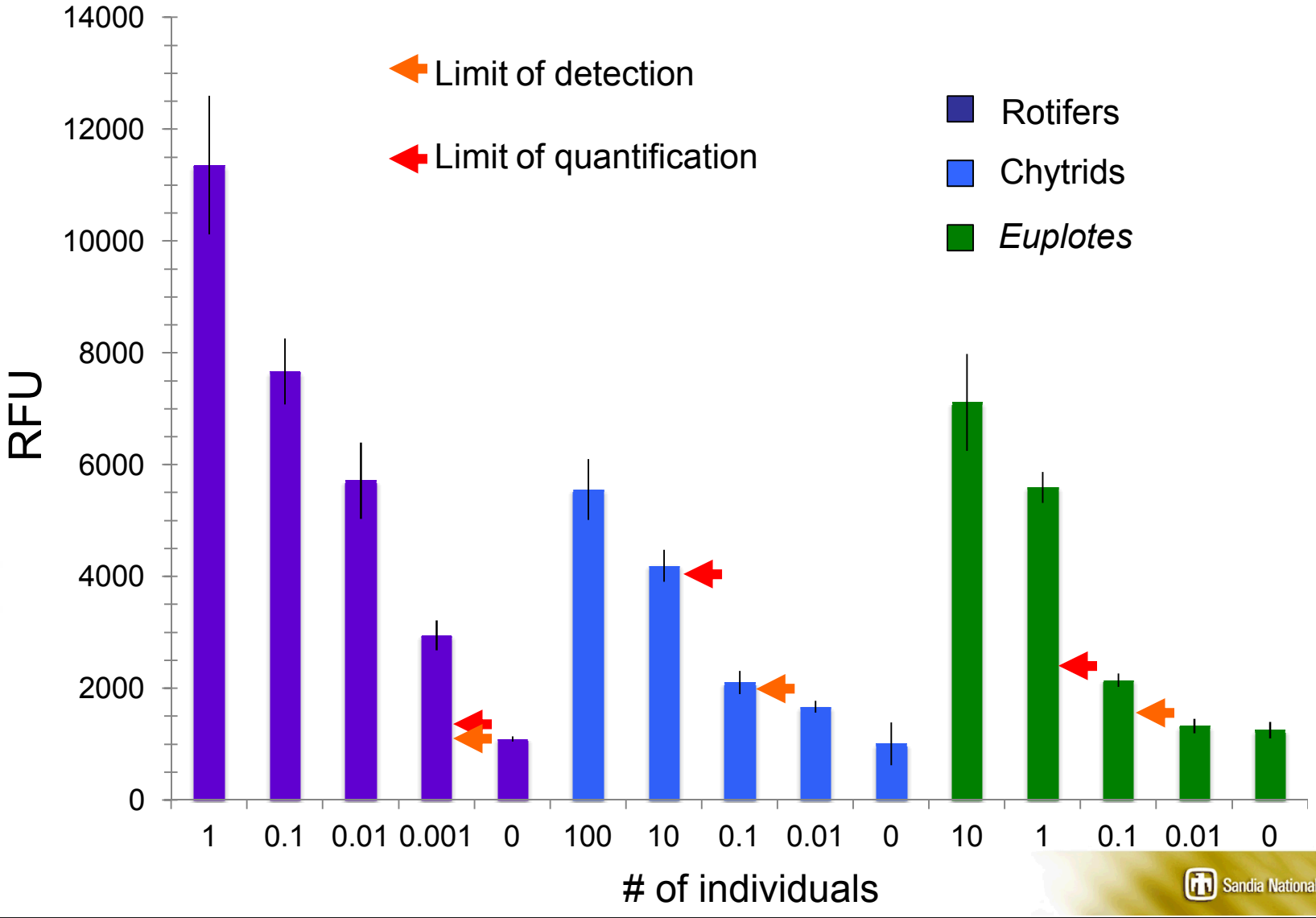




SANDIA BIOSCIENCE

Sandia National Laboratories

SpinDX detection/quantification of pest species

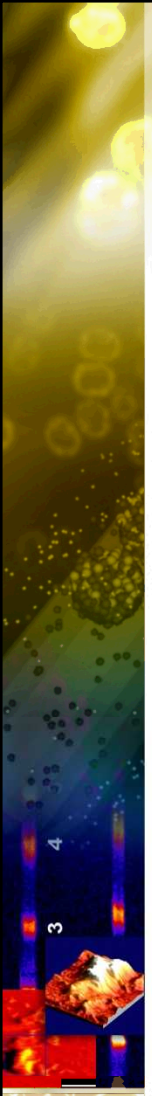


SANDIA BIOSCIENCE

Sandia National Laboratories

Summary

- Relevance of objectives
  - Development of technology to for early detection of pond crash agents, enabling the development and utilization of countermeasures, reducing costs.
- Approach
  - Second generation sequencing for identification
  - Pond side SpinDX for detection
- Technical accomplishments
  - Developed and demonstrated methods for agent identification
  - Developed and demonstrated rapid, inexpensive, pond side diagnostics
- Future work
  - Applied for funding to complete development/deployment of pond side diagnostics
  - Molecular Identification of algal predators and pathogens for ATP[3]
- Success factors and challenges
  - Increased partnership and interaction with industry
- Technology transfer
  - Companies interested in testing and evaluation of pond side diagnostic system



Acknowledgments

DOE EERE BioEnergy Technology Office

Sandia National Labs

- **Laura Carney**
- **Pamela Lane**
- **Owen Solberg**
- **Chung-Yan Koh**
- Lara Jensen
- Aaron Collins
- Deanna Curtis
- Jeri Timlin
- Kelly Williams

NAABB

- Josh Wilkenfeld
- Tzachi Samocha
- Braden Crowe
- Jonathon VanWagenen
- Michael Huesemann

NASA

- Jonathon Trent
- Sigrid Reinsch

RapTOR

Pacific Northwest NATIONAL LABORATORY

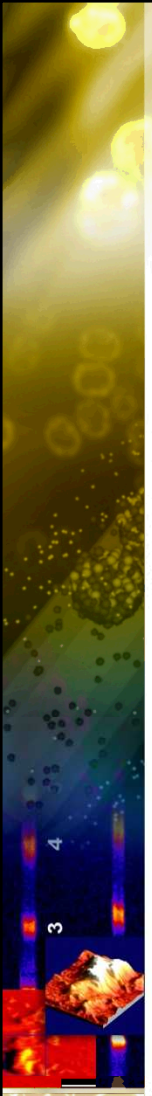
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION OMEGA PROJECT 2010 – 2012 ARC SFPUC URS

EUREKAGENOMICS

AgriLIFE RESEARCH Texas A&M System

SANDIA BIOSCIENCE

Spare slides



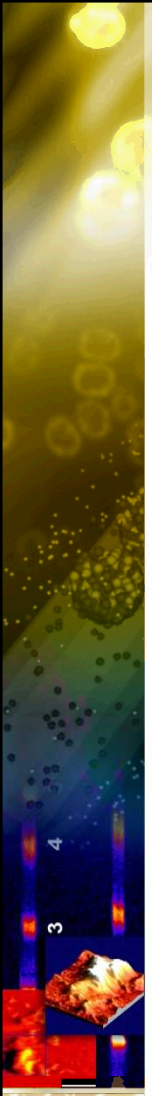
SANDIA BIOSCIENCE

Goal Statement

- The goal of this project was twofold:
  - to develop tools and methods that will be used to identify the biological agents of pond crashes through the forensic analysis crash samples.
  - To develop technology for the rapid, pond side, early detection of these pond crash agents

- The creation of tools for the diagnosis and detection of biological agents of pond crashes will be critical to informing the development of inexpensive screening and monitoring tools for early crash detection, as well as engineering and biological countermeasures that will enhance pond stability and increase long-term productivity.

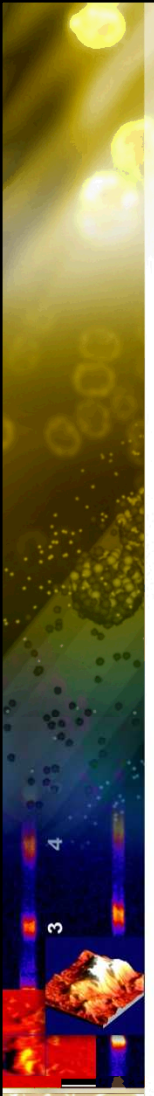
- This will decrease the loss of production time to crashes and therefore decrease the cost of the final product.



SANDIA BIOSCIENCE

Project Overview

- Annualized areal production is in part limited by pond crashes caused by biological agents. These crashes increase algal production costs and are an economic barrier to the commercialization of algal biofuels
- Goals:
  - Develop diagnostic tools and methods to identify the root causes of pond instabilities through the forensic analysis of samples taken from raceways and PBRs post-crash.
  - Identify and demonstrate potential technologies for rapid, inexpensive pond side diagnostics
  - Develop spectroscopic indicators for early stages of algal infection.
- Leverage:
  - Internal Sandia $12M investment in Biodefense technology which enables ultra high throughput sequencing to rapidly and cheaply identify an etiologigical agent without the need for isolation.
  - Internal and NIH investment in fieldable diagnostics: SpinDX
  - Sandia's hyperspectral imaging capabilities.



SANDIA BIOSCIENCE

1-Approach

Leveraged Sandia investments in biodefense and clinical diagnostics

Developed methods to enrich for nucleic acids that are likely to derive from the etiological agent of the crash.

Utilized second generation sequencing to identify agents.

Created quantitative assays that facilitate the detection of agents at low concentration.

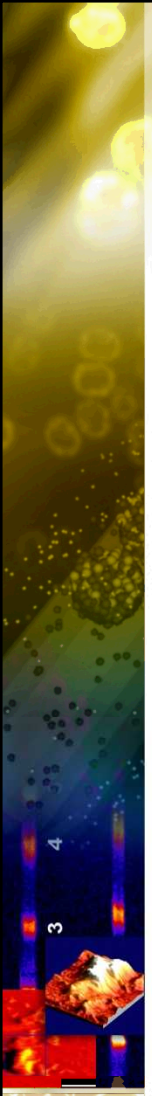
Utilize advance spectroscopic methods to detect early hallmarks of algal stress.

Technical metrics of progress

- Speed of analysis: 48 hours for identification, ~30 min for detection
- Sensitivity of detection: <1 organism
- Cost of analysis

Unique aspects:

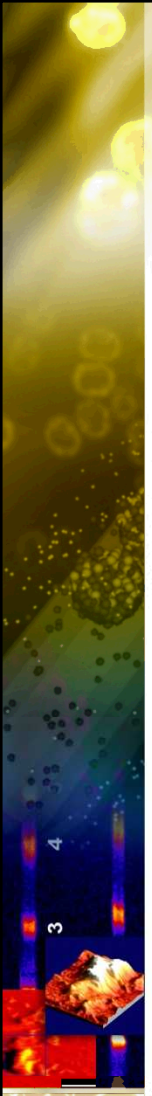
- The use of nucleic acid target enrichment followed by ultra high throughput sequencing as a strategy for agent identification
- Fieldable diagnostics for pond side detection



SANDIA BIOSCIENCE

2-Technical accomplishments.

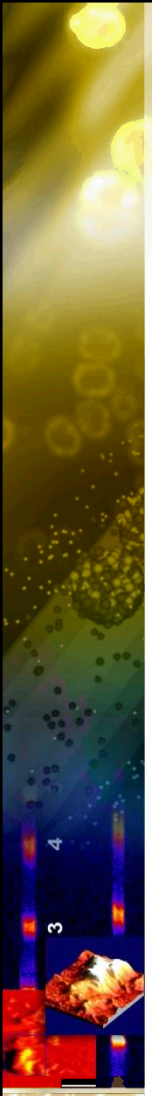
- Genetically identify unknown etiological agents of pond crashes without the need for agent isolation.
  - Developed collaborations and obtained samples from groups running pilot scale open ponds
  - Tested sample preparation and analysis trained on laboratory and field samples
  - Developed host and background subtraction reagents
  - $10-$30 per sample sequencing cost
- Developed inexpensive yet rapid and sensitive pond side diagnosis system
  - Demonstrated the bench top technique for
  - 30 minutes sample to answer
  - Single organism detection for many predators.
  - Estimated ~Dollar per sample/ ~$1000 instrument



SANDIA BIOSCIENCE

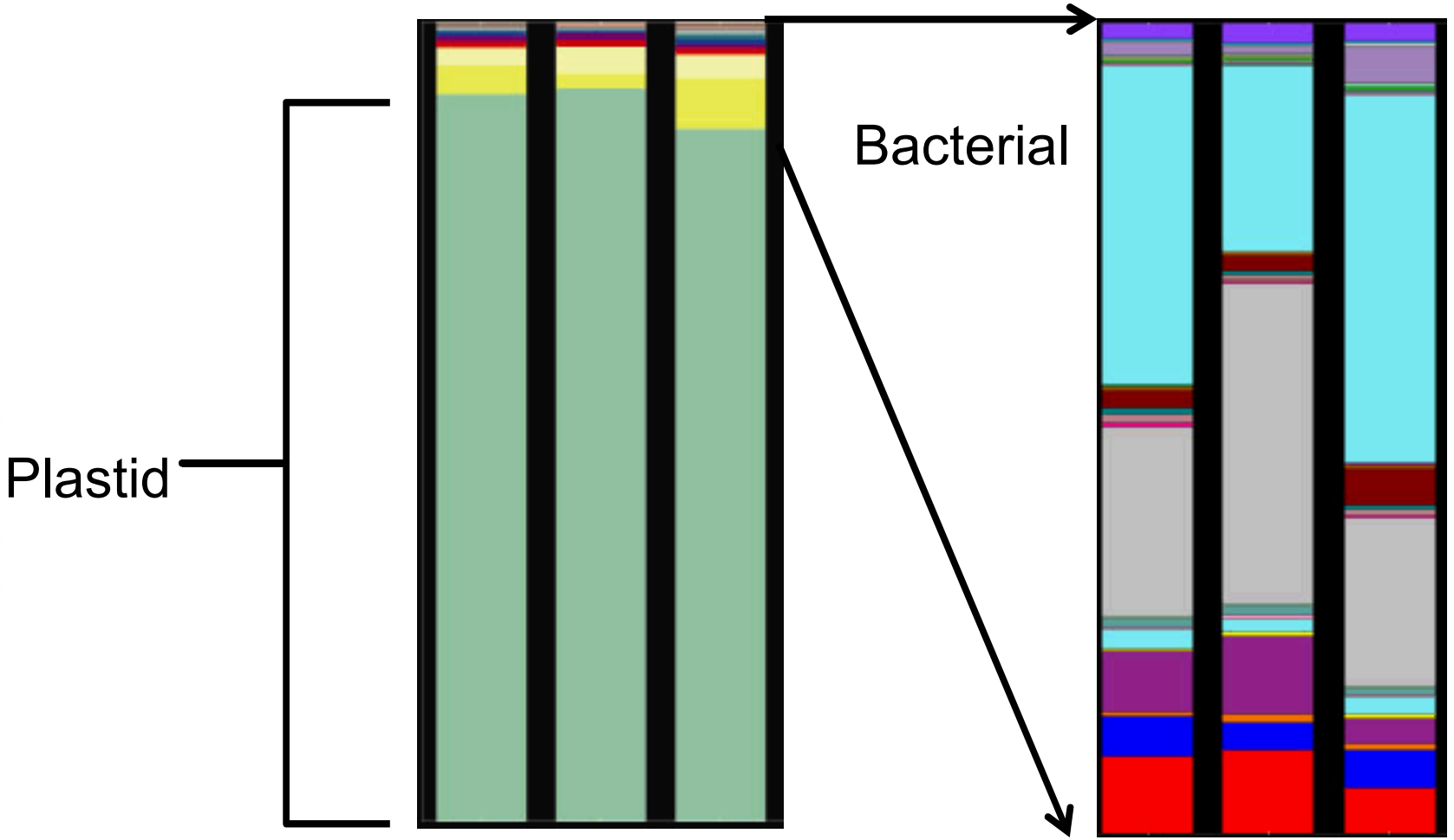
The goal of all this sequencing? From the sequencer to assay.

- Development of probe sets for persistent pest
  - Additional sequencing of molecular barcode regions
    - Full SSU, LSU, ITS regions, mtCO etc.
  - Informatic analysis
    - Define the most informative regions
  - Probe development, validation, optimization
    - qPCR
    - Specificity, cross reactivity, sensitiviithy
  - Deployment
    - Laboratories associated with ponds
  - Feedback



SANDIA BIOSCIENCE

Removal of host can increase yield of data by 20 fold



We compared high output versus rapid run Illumina systems

- Illumina sequencers are new to community profiling applications
  - Current maximal read length:150 bp PE (300bp)
  - Soon to be 250bp PE (500 bp): in beta testing.
  - Trade off of using Illumina is on of read length versus throughput (limits coverage to single VRs)
  - Accuracy is higher than longer read platforms
  - Potential for greater depth of sequencing
  - Greater multiplexing– lower per sample costs
  - JGI: 96 sample per MiSeq lane (16s analysis)\

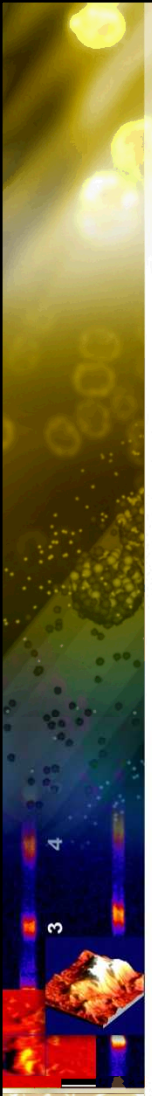
For this study:

- Miseq
  - 21 barcodes per lane (singe lane/chip)
  - 100K Reads per sample (454 run equiv)
  - 27 hours run time , 150 bp PE

- Illumina GAII
  - 18 barcodes per lane (4 lanes)
  - 2.5M Reads per sample
  - 14 day run time

Project Goals

- Rapidly identify biological agents that play a role in pond crashes
  - Next gen DNA sequencing
  - Compare healthy ponds to crashed
  - Compare time series in ponds leading to crashes
- Goal is to complete this analysis in <24 hrs
- Drive down costs
  - Removal of non-informative nucleic acids
  - Multiplexing of samples
- Bioinformatic analysis pipeline
- Create molecular assays against these agents
- Develop methods for routine isolation and culture of agents
- When possible isolate agents and reconstitute crash
  - Confirm the role of the suspected agent (Koch's postulates)
  - Determine the role of abiotic factors in modulating the crash



SANDIA BIOSCIENCE

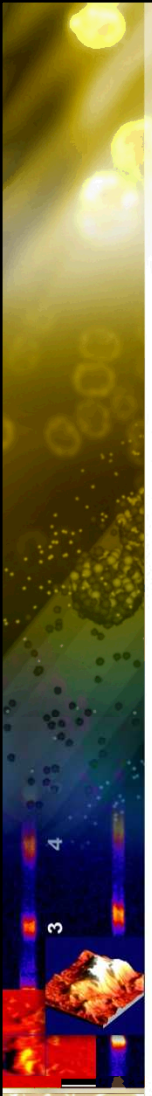
Instead of fractionating biomass extensively we create targeted libraries for sequencing

Group specific primers:
rDNA analysis
Prokaryotic amplifcation primers that exclude chloroplast

Subtraction:
rDNA analysis or metagenomic analysis
Physically removes unwanted sequences

Normalization;
rDNA analysis or metagenomic analysis
Removes or destroys high abundance sequences

Blocking primers:
rDNA analysis
3' modified primers that prevent amplification from known targets



SANDIA BIOSCIENCE

Physical separations can be time consuming and incomplete

- The final goal is to limit the extent of pre-processing
- Aside from concentrating viruses, it may not be necessary to fractionate crash samples
- Final identification through UHTS