Sandia Bioscience

SAND2013-6673C

Pond Crash Forensics:
Microbiome Analysis and Field Diagnostics

Laura Carney, Pamela Lane, Chung-Yan Koh & Todd W. Lane
Sandia National Laboratories

Heliae Development, LLC, Gilbert AZ
15 August 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

Sandia is a science-based engineering research and development laboratory

Energy, Climate & Infrastructure Security

Nuclear Weapons

International, Homeland & Nuclear Security

Defense, Systems & Assessments

Sandia is a multi-site laboratory

> 11,876 people total
~ 1180 in California
~ 1650 w/ Ph.D. (lab-wide)
~ $2.5B budget

Albuquerque

Livermore

WIPP,
Carlsbad, New Mexico

JBEI,
Emeryville, California

Pantex,
Amarillo, Texas

Tonopah Test Range,
Nevada

California Lab Capabilities

National Security Missions

Maintain a Safe, Secure, & Effective Nuclear Deterrent

Reduce the Chemical, Biological, Radiological, & Nuclear Risks

Systems Engineering & Analysis for Nuclear Weapons

CBRN Defense

CA Approach

Integrated Ops
Partnerships
Location

Defense Against Cyber Threats

Transportation, Fuels, & Engines and their Impacts

Capability Portfolios

Protect Operations & Assets in Global Cyberspace

Achieve a Secure, Sustainable Transportation Future

Sandia's bioscience program addresses the Nation's biodefense and biofuels needs

"Developing the next generation of biofuels is key to our effort to end our dependence on foreign oil and address the climate crisis – while creating millions of new jobs that can't be outsourced."

"…terrorists are more likely to be able to obtain and use a biological weapon than a nuclear weapon…The US government needs to move aggressively to limit the proliferation of biological weapons and reduce the prospect of a bioterror attack."

Our goal is to analyze, understand and control the functions of biological systems to meet national security challenges in biodefense and biofuels production.

Sandia Bioscience

Sandia has extensive experience with developing and fielding CB defense solutions

Detection Systems

- BioWatch Gen 2
- BioWatch Gen 3
- Tenix Water Monitor
- RDCDS
- μGasAnalyzer
- PROTECT
- RapTOR

ConOps Development & Support

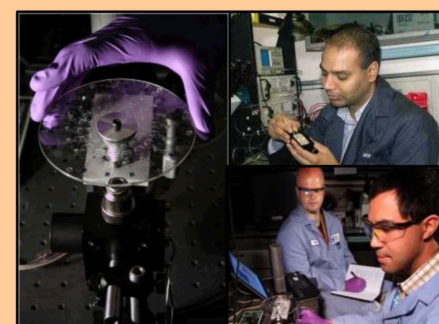
- BioNet
- BioWatch Incident Characterization (BWIC)
- BioWatch Indoor Reachback Center (BIRC)
- PROACT

Decontamination & Restoration

- Sandia Decon Foam
- DF-200
- IBRD
- WARRP
- Validated Sampling Procedures
- BROOM
- AWARE

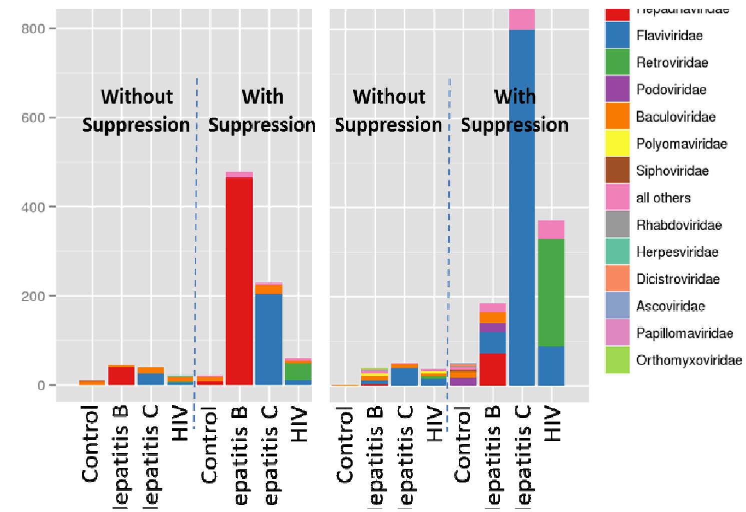
Medical Diagnostics

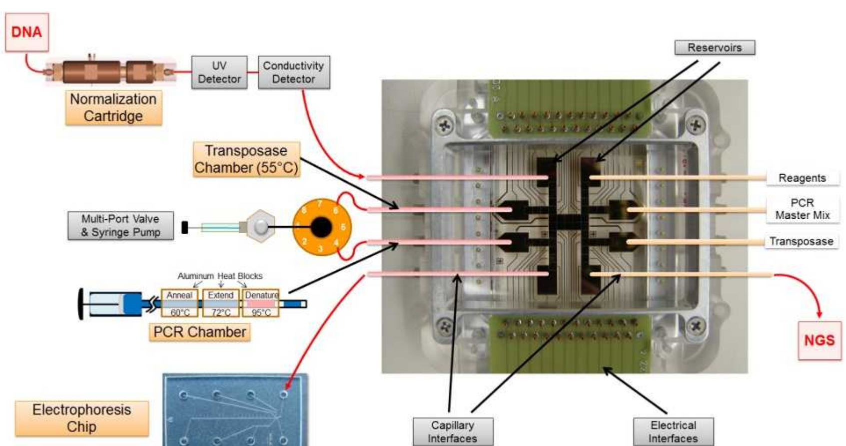
- Toxin Diagnostics
- Oral Diagnostics
- Radiation Biodosimetry
- Rickettsial Diagnostics
- NASA astronaut diagnostics

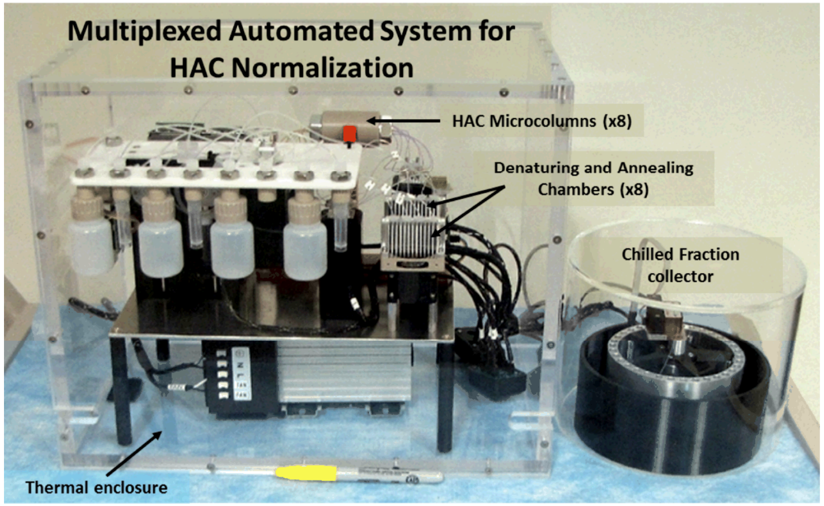


We have developed a system for ID and characterization of unknown pathogens

RapTOR rapidly and automatically applies suppressive molecular biology methods, next generation sequencing and bioinformatics to enable the identification and characterization of unknown pathogens in clinical samples, thereby enabling a <u>faster and more effective public and military health response</u>.

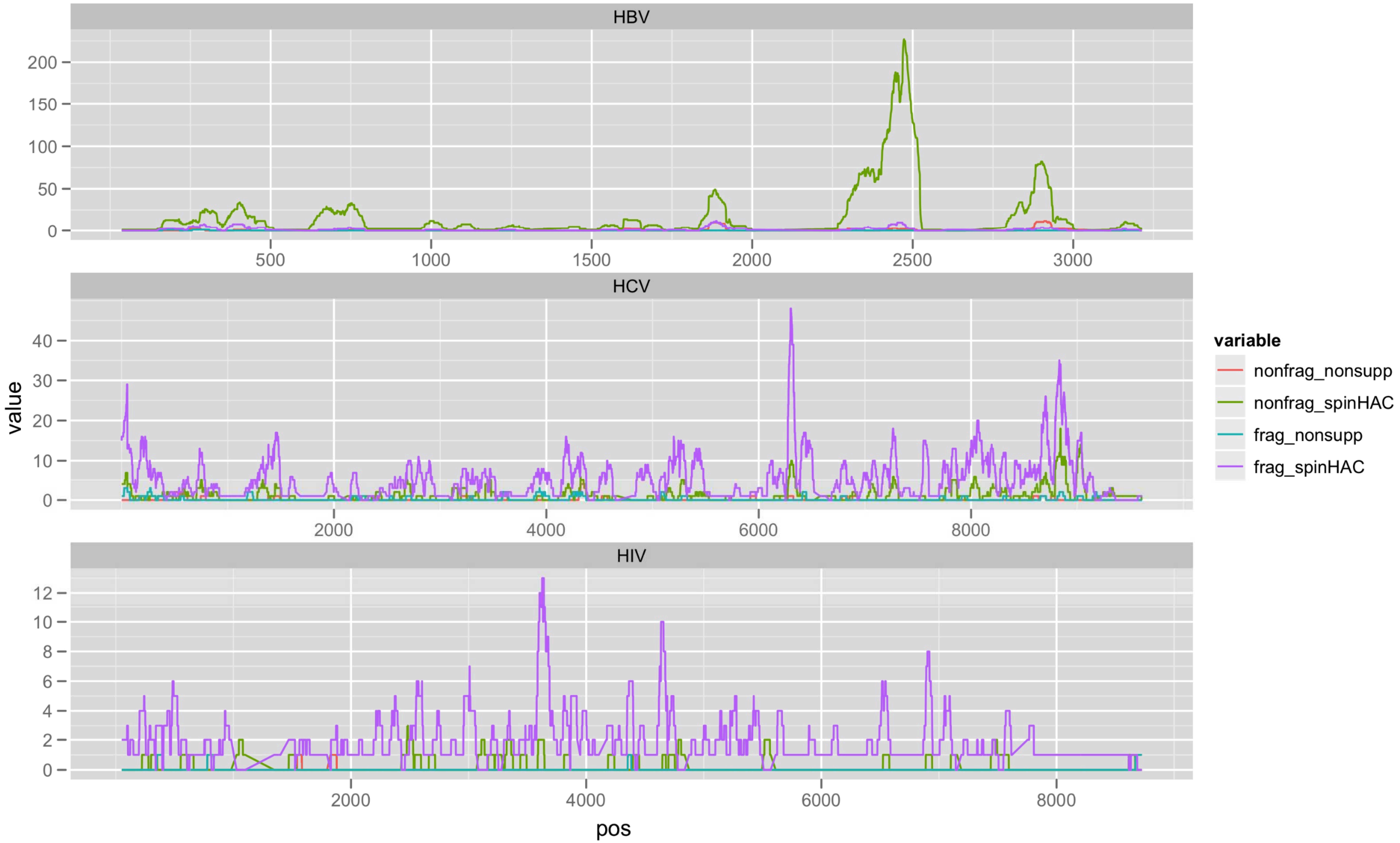






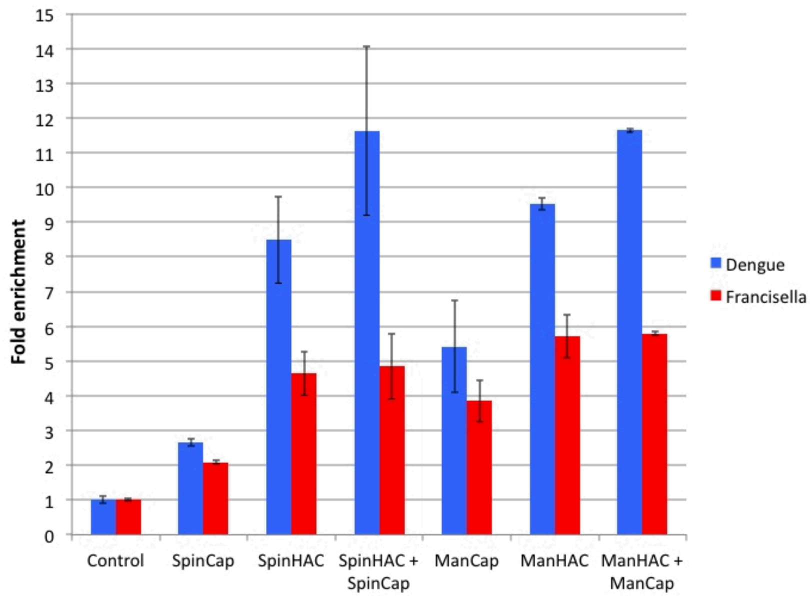
SANDIA BIOSCIENCE

Increased coverage of viral genomes through suppression

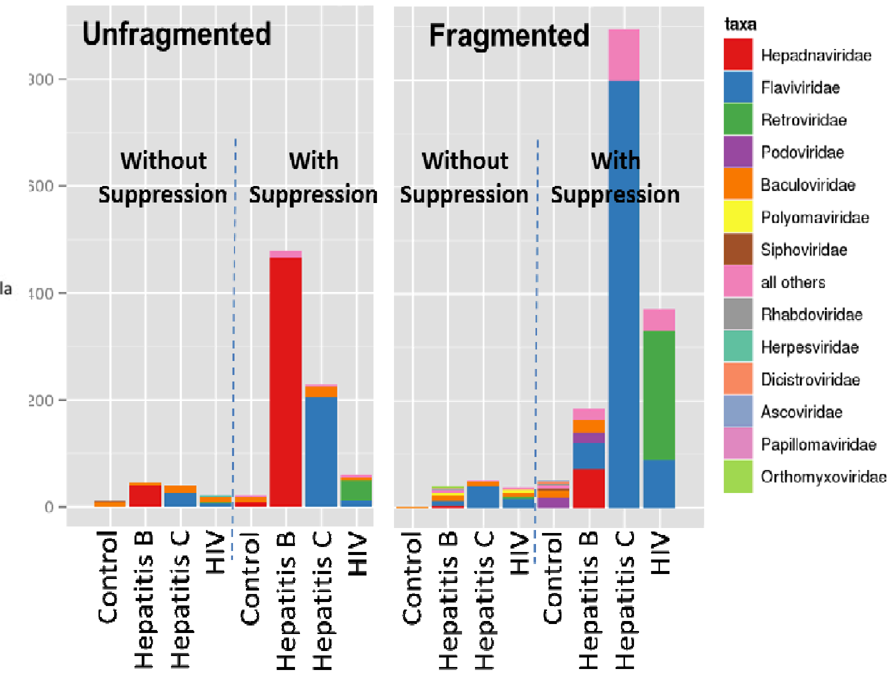


We have characterized suppression methods and applied them to clinical samples

PBMC's spiked with Dengue and Francisella



Human plasma samples



Capillary systems give most consistent results

Pathogen enrichment is dependent on sample type (5-20X)

Other sample prep parameters effect pathogen enrichment

SpinDx system for medical diagnostics

- **Rapid:** < 20 min sample-to-answer
- **Inexpensive:** < $500 per instrument, < $2.00 per consumable disk
- **Multiplexed:** Up to 64 parallel assays
- **Ultra-Sensitive:** 1 – 2 orders of magnitude more sensitive than ELISA
- **No Sample Preparation:** Direct analysis of clinical (blood, serum, etc.) or non-clinical (foods, powders, etc.) samples
- **Minimally-Invasive:** 2-µL sample per assay
- **Broad assay menu:** Proteins, nucleic acids, cells

Designed for Biodefense
Radiation Biodosimetry
Toxin & Pathogen Diagnostics
Applicable to Public Health
Spin-out company Sandstone Diagnostics

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.:

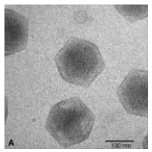
- Irradiance, temp, salinity etc.
- Pond/PBR collapses caused in part by biological agents (ultimate suboptimal condition)

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*

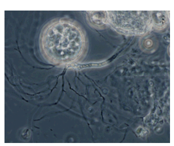
Viruses



Bacteria

Predators

Fungi



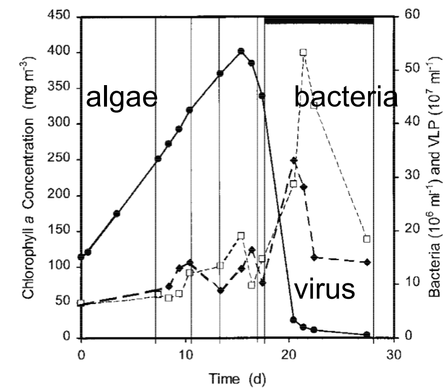
Patterson & Laderman, 2001.

+

+

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."

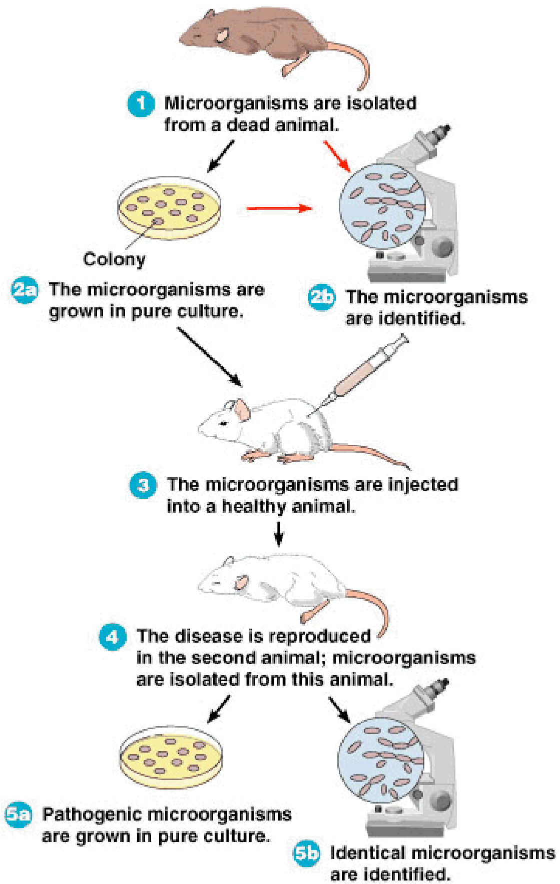
- DOE Draft Algal Biofuels Technology Roadmap (2009)

Goals of PCF for ATP3

- 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 <48 hrs
- Drive down costs
  - Removal of non-informative nucleic acids
  - Multiplexing of samples (36 X for MiSeq: >$30)
- Bioinformatic analysis pipeline
- Create molecular assays against these agents

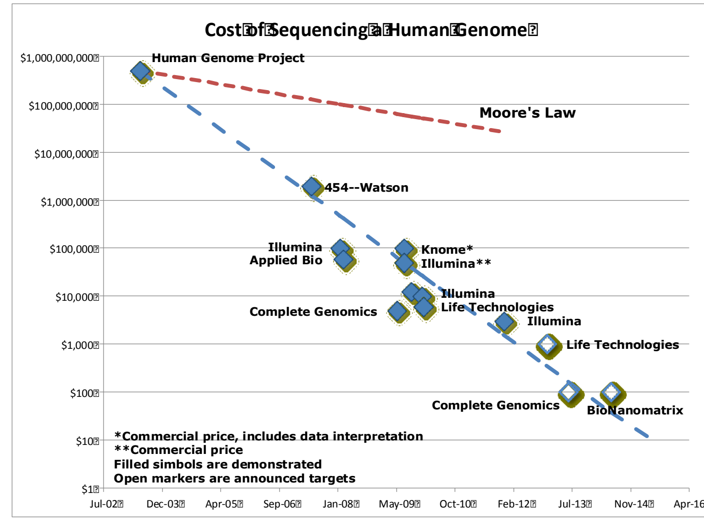
A staged approach to crash agent identification (Koch's postulates)

- **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
- <u>Key is to get more sequencing hits on target thus reducing the cost of analysis to $10s</u>

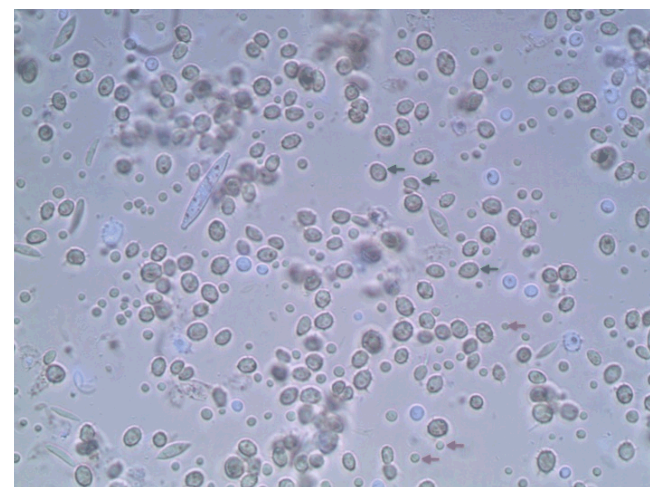


MiSeq

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

HiSeq

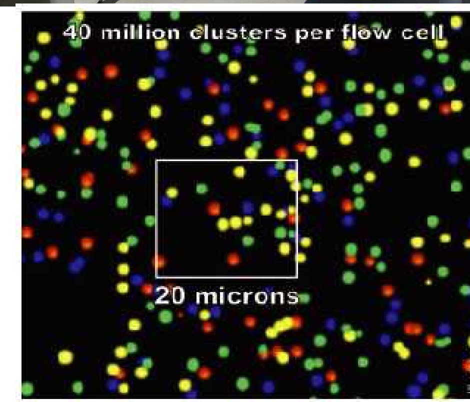
2X150nt: 8days
540 -600Gb

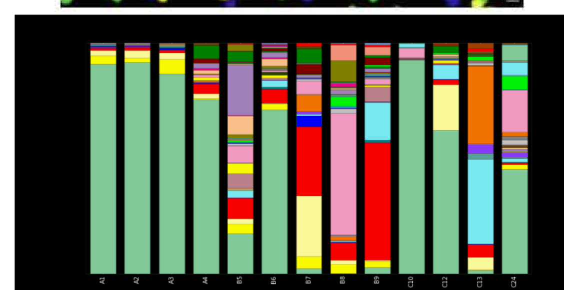


NASA OMEGA prom

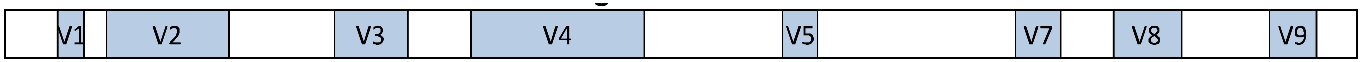
There are various forms that samples can take.

- Ponds or PBRs
  - "Whole" water shipped overnight on ice
    - Culturing of agents
  - Frozen biomass/Purified nucleic acids
    - Analysis of archived samples
    - Preserves diversity and structure
  - Viral fraction purified and concentrated
  - Nucleic acids extracted
    - Prokaryotic analysis
    - Eukaryotic analysis
    - Viral analysis
  - Library preparation and barcoding
  - High throughput sequencing (Illumina)
  - Bioinformatic analysis





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

In-house tools

Off-the-shelf tools

178 million raw paired-end reads (150PE)

Qfilter

155 million quality-filtered & trimmed paired-end reads

Bowtie2

145 million paired-end alignments to Silva rRNA database

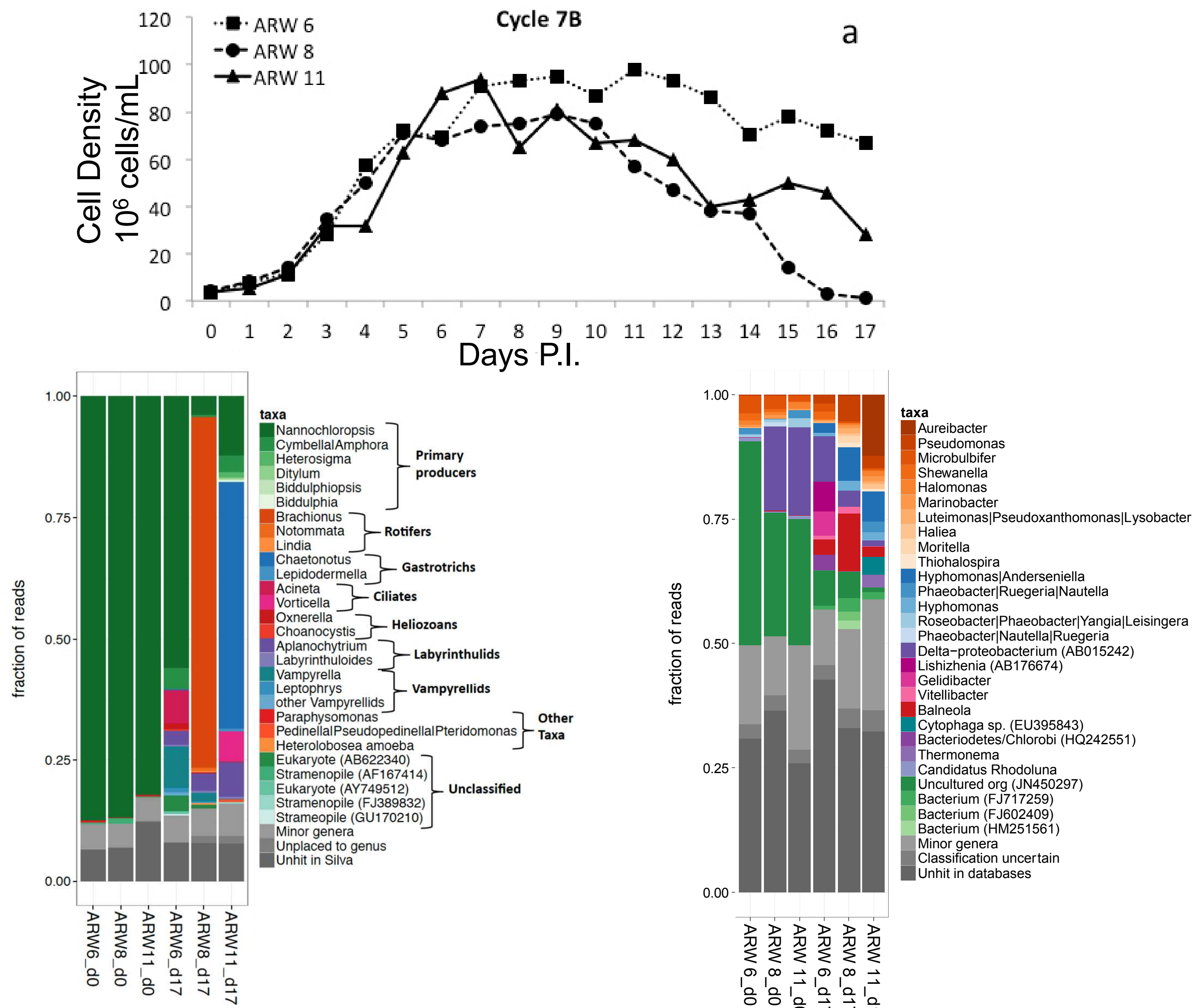
sam2tax

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

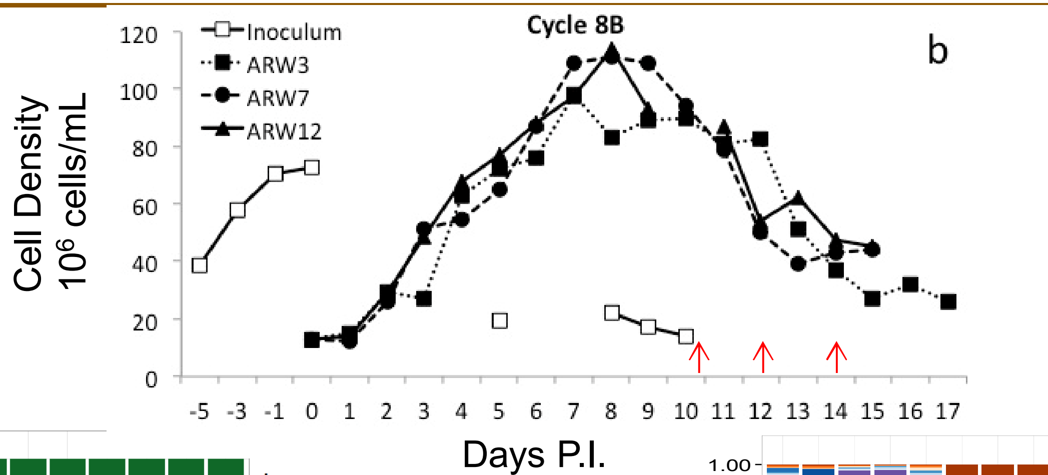
lca

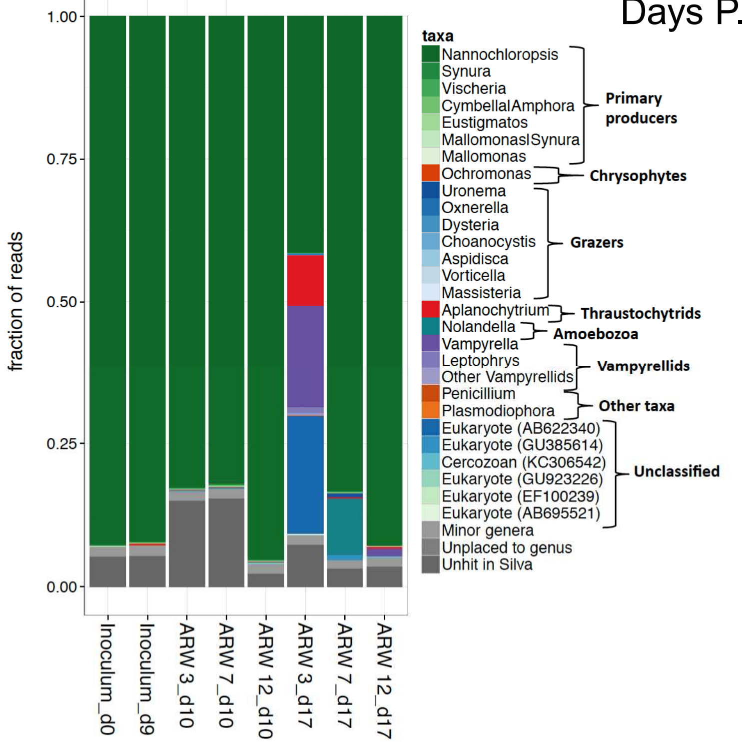
Last common ancestor based taxonomic summary for each hit:
132 million resolve at Genus level, 10 million at Species level

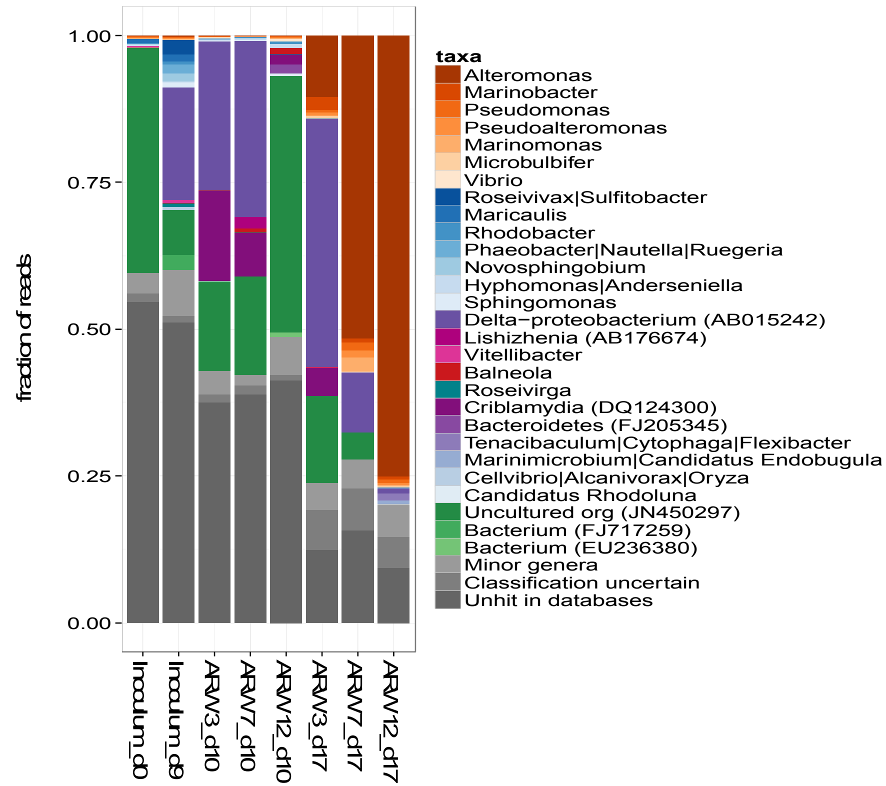
Forensic Analysis of Pond Crashes



Evaluation of pond intervention

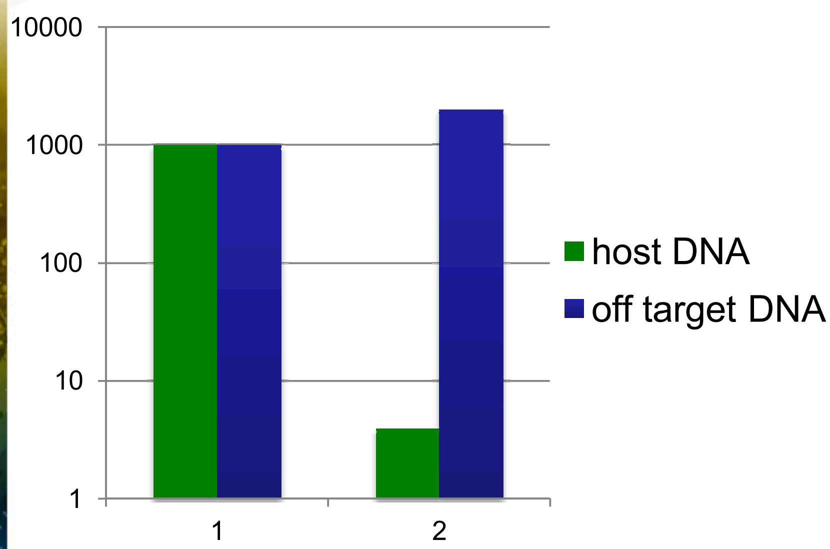


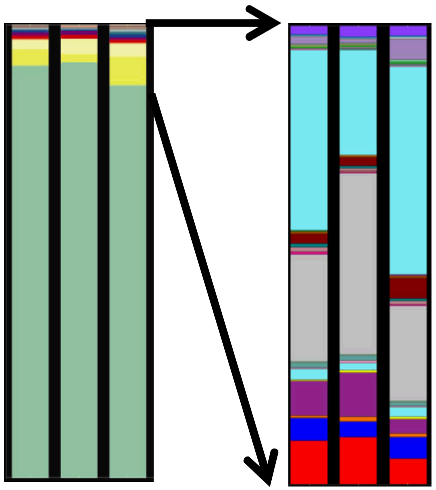


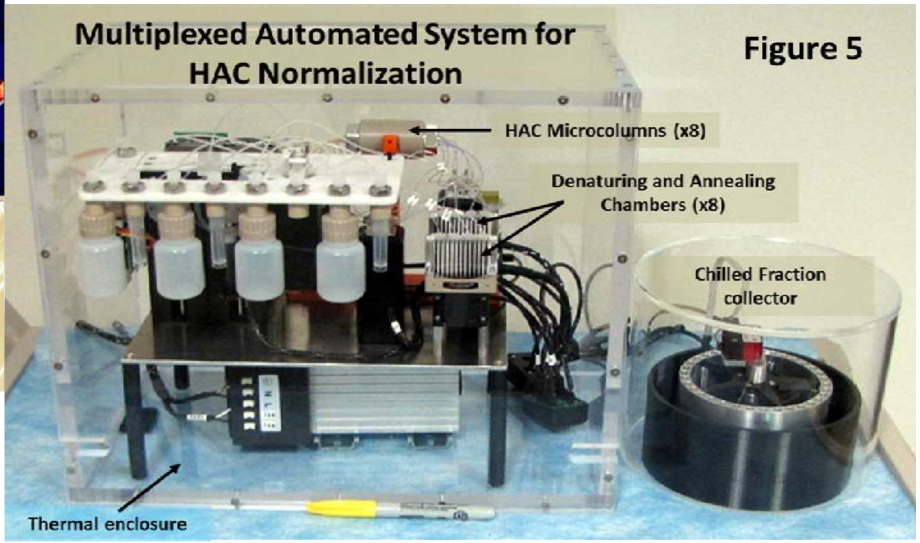


Sandia Bioscience

Automated, mutiplexed and manual systems for capture-based suppression







Thermo Scientific Pierce Micro-Spin Columns (Part No. 89879)

Total column capacity = 0.4mL (resin bed = 0.1mL; reservoir = 0.3mL)

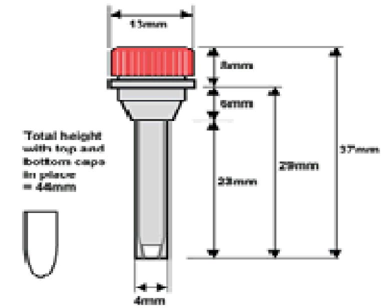
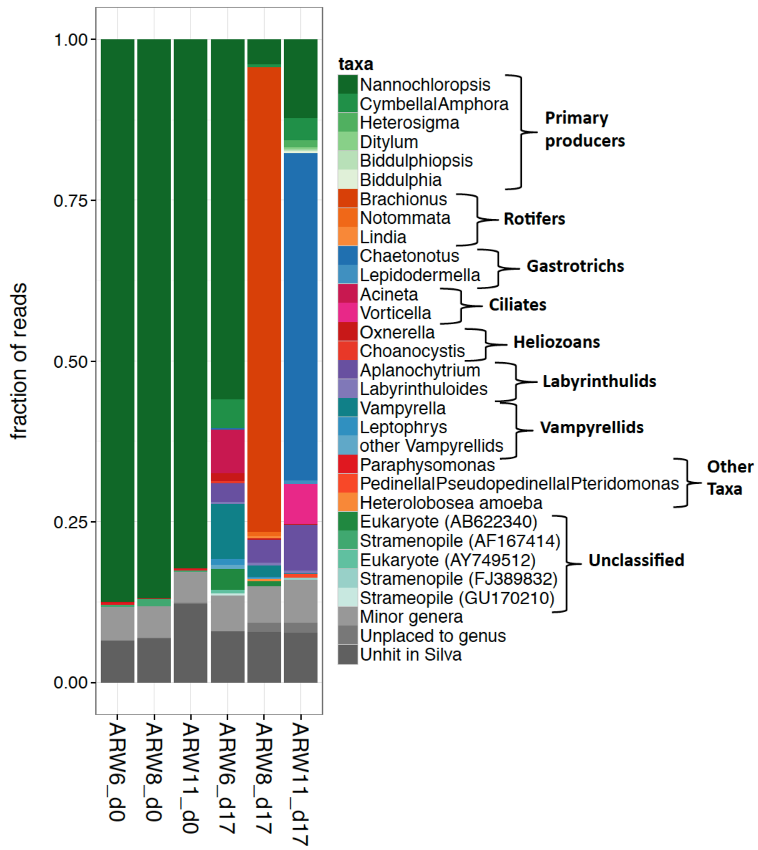
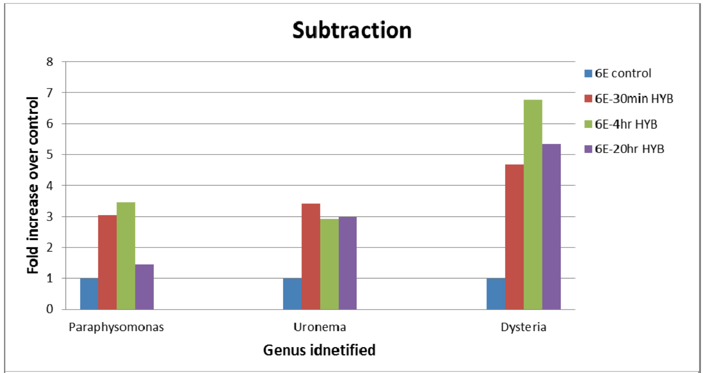


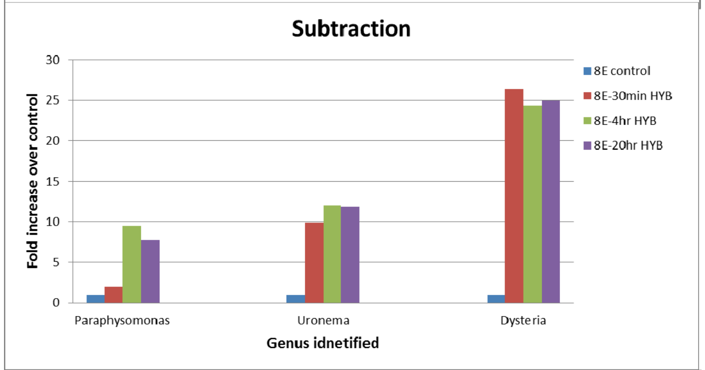
Figure 10: Dimensions of the commercially available spin columns

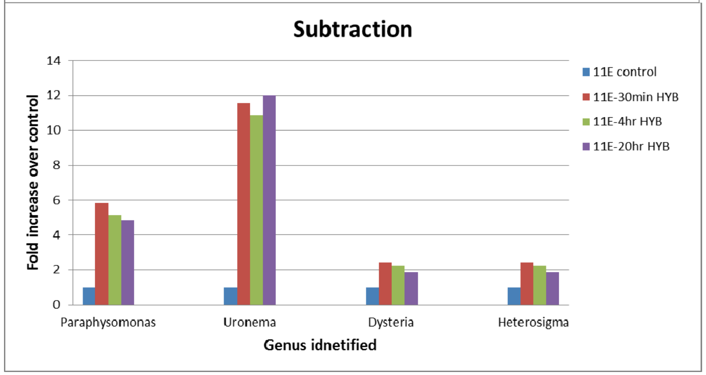
SANDIA BIOSCIENCE

Subtraction enhances early detection





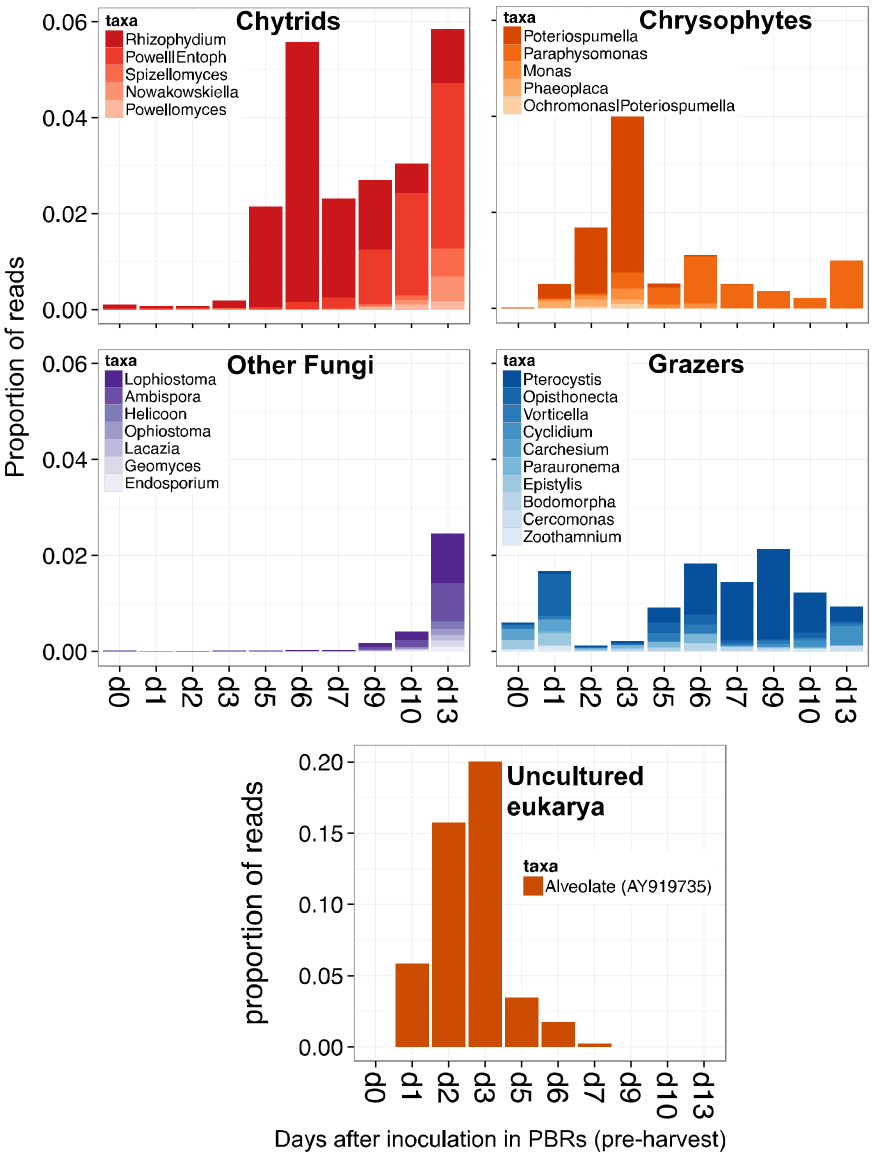




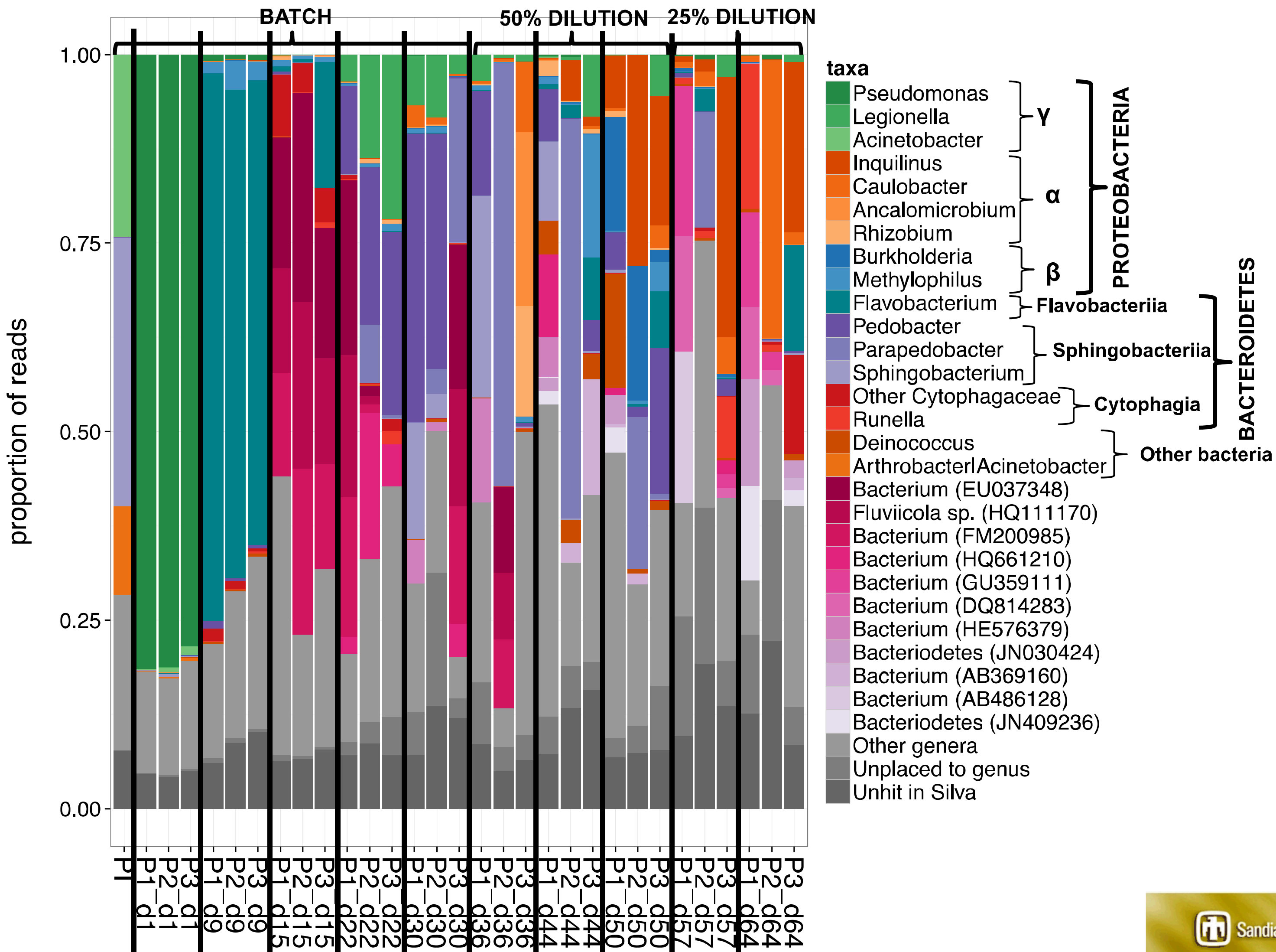
SANDIA BIOSCIENCE

Sandia National Laboratories

Deleterious species in the OMEGA system

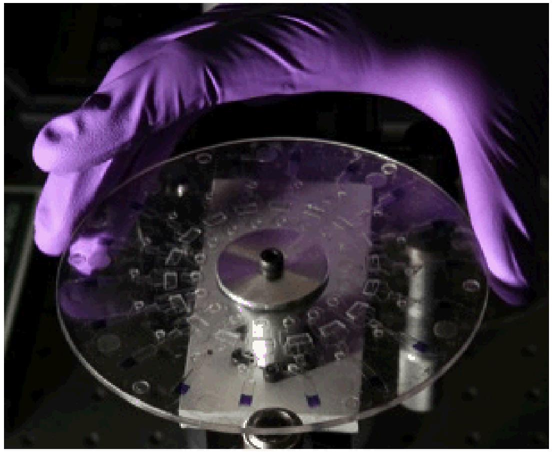


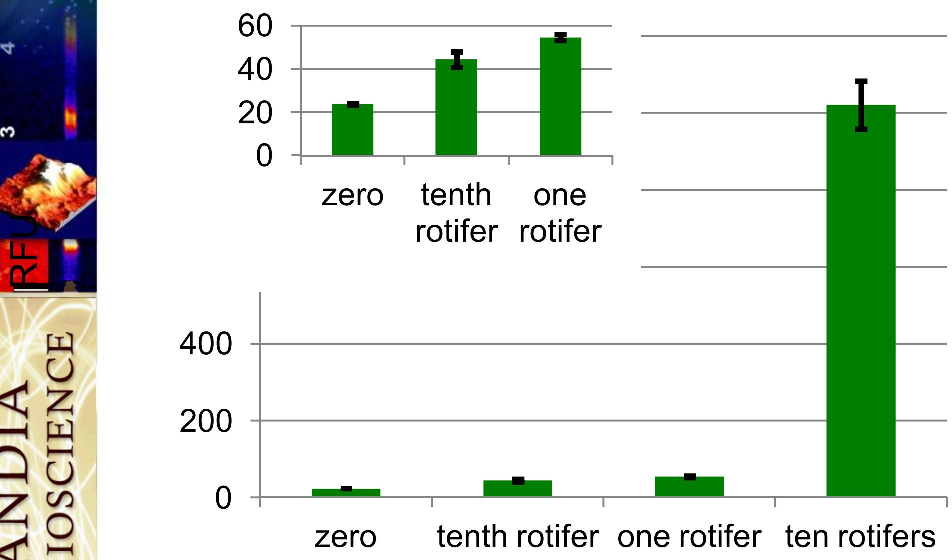
Long Term Microbiome Analysis Indicates Community Divergence

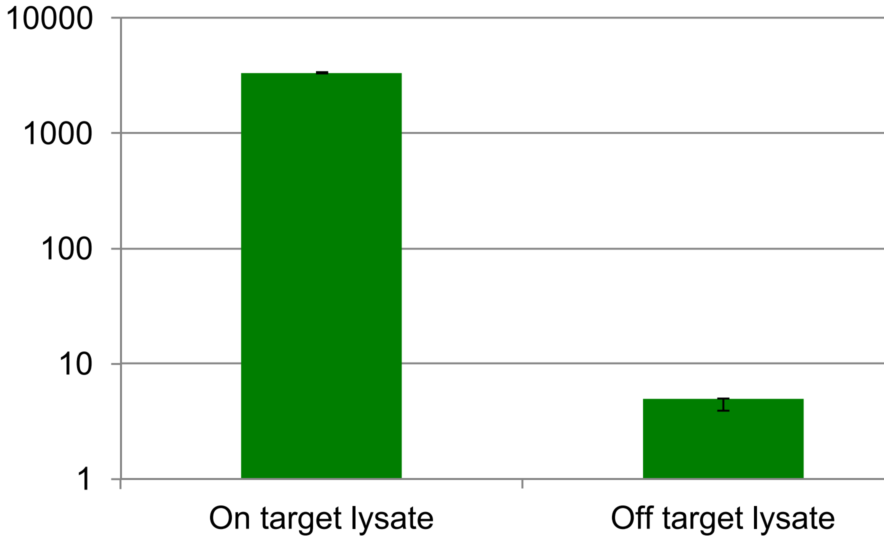


The goal of all this sequencing? Direct from the sequencer to assay.

- 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

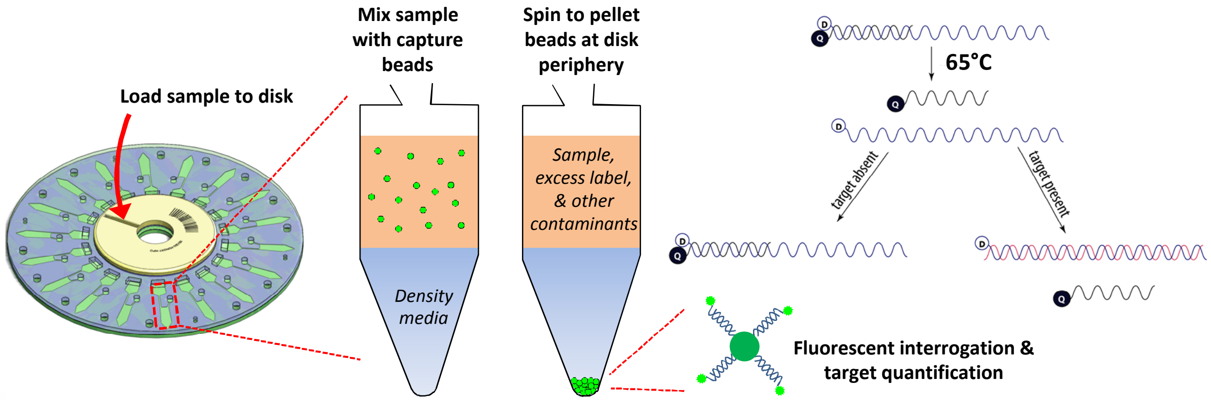


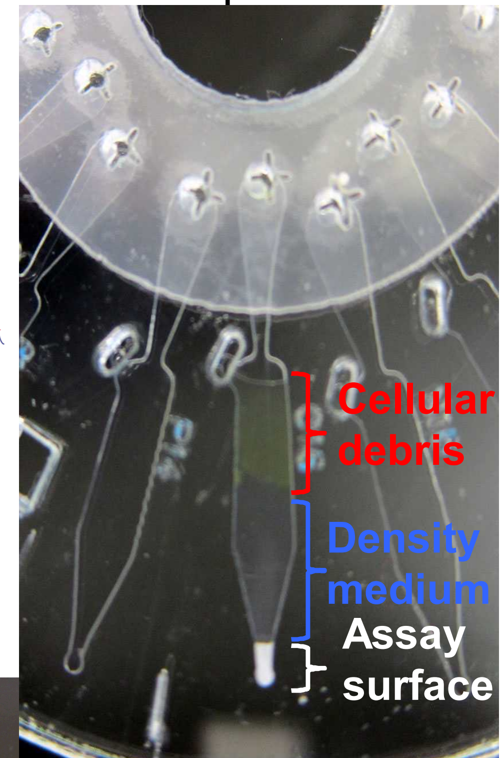




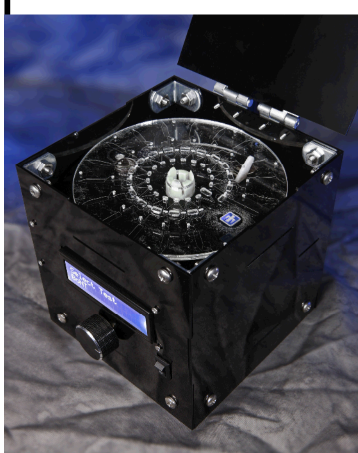
SpinDx™ has the necessary characteristics for a field assay for pond management

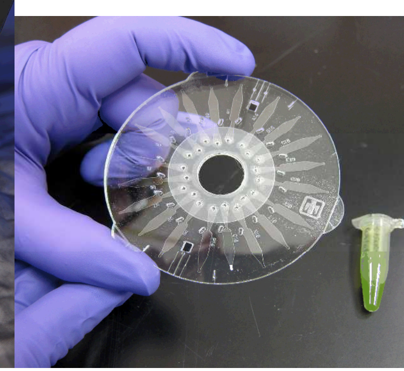
- FRET-based bead hybridization assay enabling capture and quantification of pathogen-specific RNA/DNA signatures





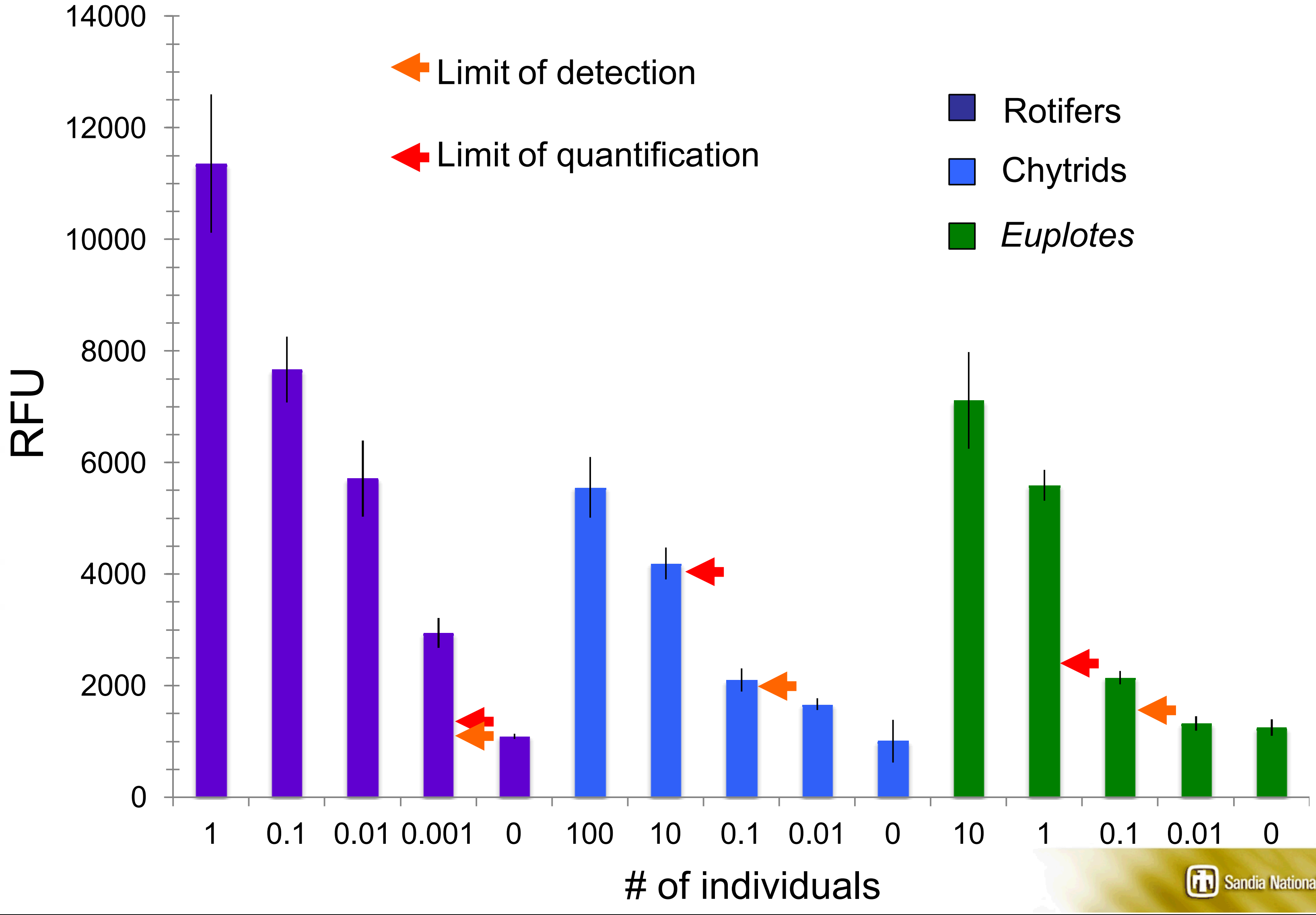
- Assay time: approx 30 min
- 36 channels per disc
- Potential for multiplexed assays in each channel
- Low reagent costs
- Low material costs
- Low instrument cost ($1000)
- Fieldable





Sandia Bioscience

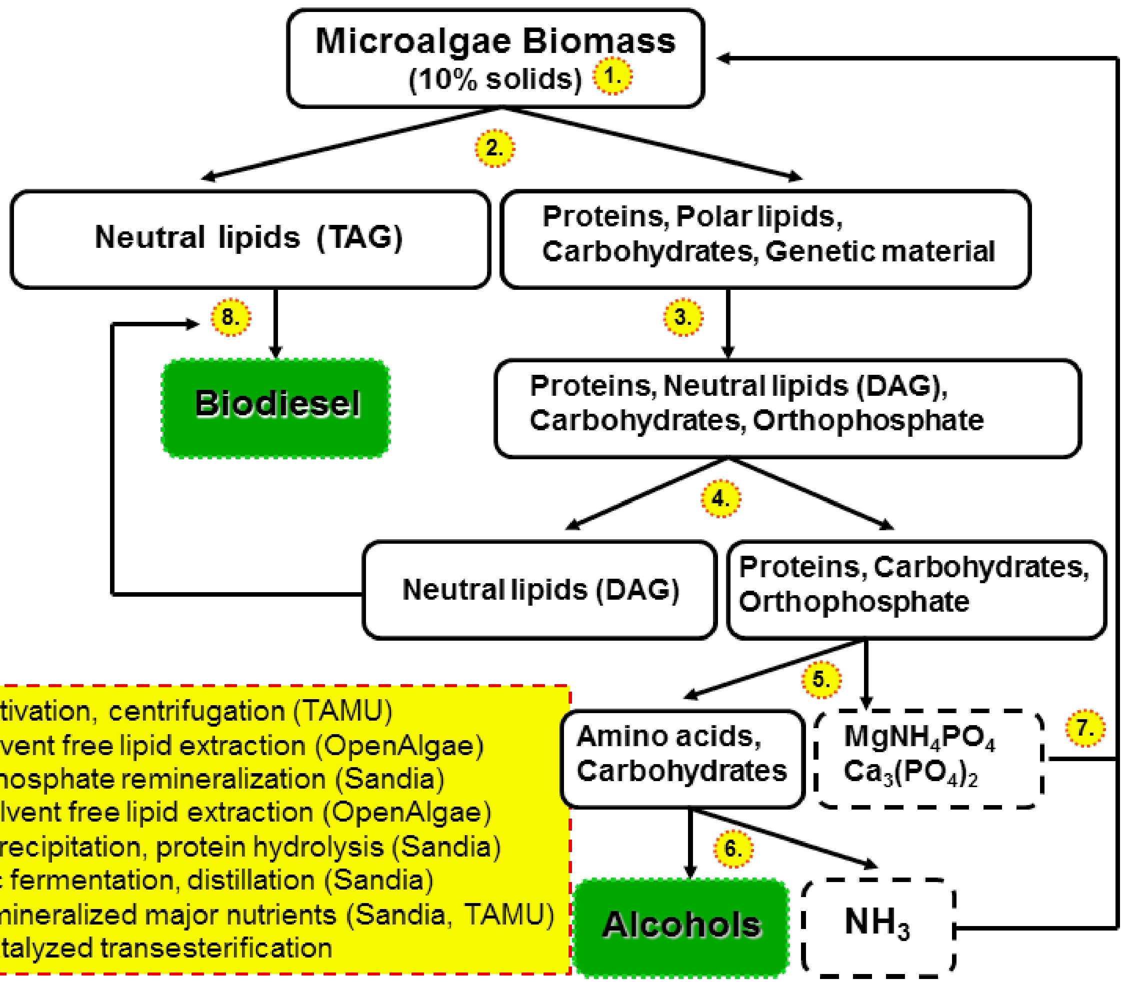
SpinDX detection/quantification of pest species



Sandia Bioscience

Sandia National Laboratories

DOE Nutrient Recycling Project



1.) Raceway cultivation, centrifugation (TAMU)
2.) 1st stage, solvent free lipid extraction (OpenAlgae)
3.) Enzymatic phosphate remineralization (Sandia)
4.) 2nd stage, solvent free lipid extraction (OpenAlgae)
5.) Phosphate precipitation, protein hydrolysis (Sandia)
6.) Microaerobic fermentation, distillation (Sandia)
7.) Return of remineralized major nutrients (Sandia, TAMU)
8.) Acid/base catalyzed transesterification

Motivation – scale of energy consumption exceeds nutrient production

- To meet 10% of liquid fuel needs (roughly 30 BGY):
  - Algae biomass: 200 - 500 M mt/yr.
  - Phosphorous: 2.4 - 6 M mt/yr
    - Compare 4.1 M mt in 2006: 61 – 146% of recent consumption.
  - Nitrogen (nitrate, ammonium, etc.) 18 - 45 M mt/yr
    - Compare 14 M mt in 2006: 130 – 320% of recent consumption.
- Food-vs-fuel concerns for nutrients.
- Nutrients (fertilizer)
  - Needed for biological productivity, not for fuel.
  - Phosphorous: mined resource, essentially nonrenewable.
  - Nitrogen: Haber-Bosch process has own energy requirements.

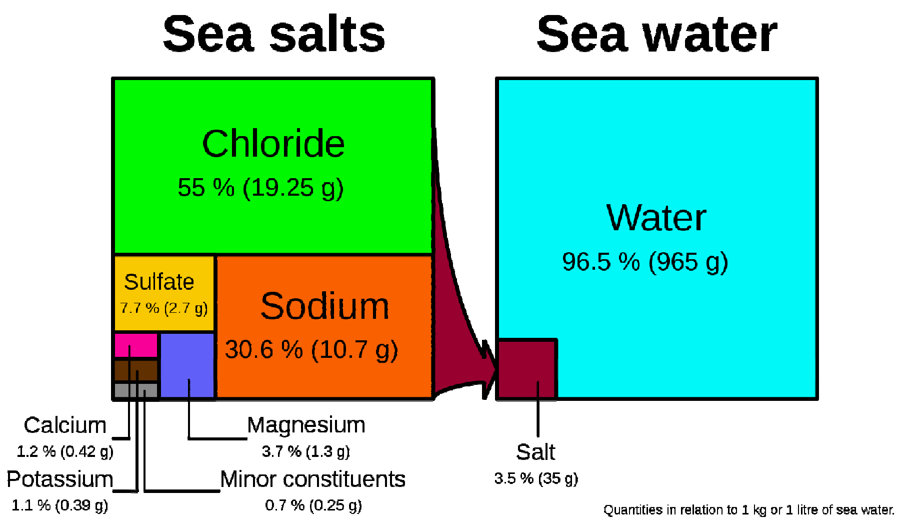
"The Achilles' Heel of Algae Biofuels: Peak Phosphate"
Forbes, Feb 2012

Need to recycle nutrients.

Cannot afford to pass through one time only.

Pate, Klise, Wu, "Resource demand implications for US algae biofuels production Scale-up," Applied Energy, 88:3377-3388 (2011).

In currrent system, Mg carried over with biomass may result in struvite fomation



$NaNO_3$ (M)	KH_2PO_4 (M)	Fe (mg/g)	Mg (mg/g)
0.006	0.0003	3.37	77.3
0.006	0.0003	4.75	67.1
0.006	0.001	3.86	78.1
0.006	0.001	3.81	98.3
0.003	0.0003	2.83	82.0
0.003	0.0003	2.91	93.5
0.003	0.001	3.46	74.8
0.003	0.001	2.41	74.4

3-4 mMol Mg/gm AFDW

Depending on chemical makeup of growth medium significant extracellular Mg may be carried over with biomass

Internal Mg stores may also be singificant: Macronutrient

The resulting P/Mg ratio may promote the formation of struvite: $MgNH_4PO_4$

We will not alter the chemistry of the pond or biomass to form or use struvite

No requirement for new Mg

SANDIA BIOSCIENCE

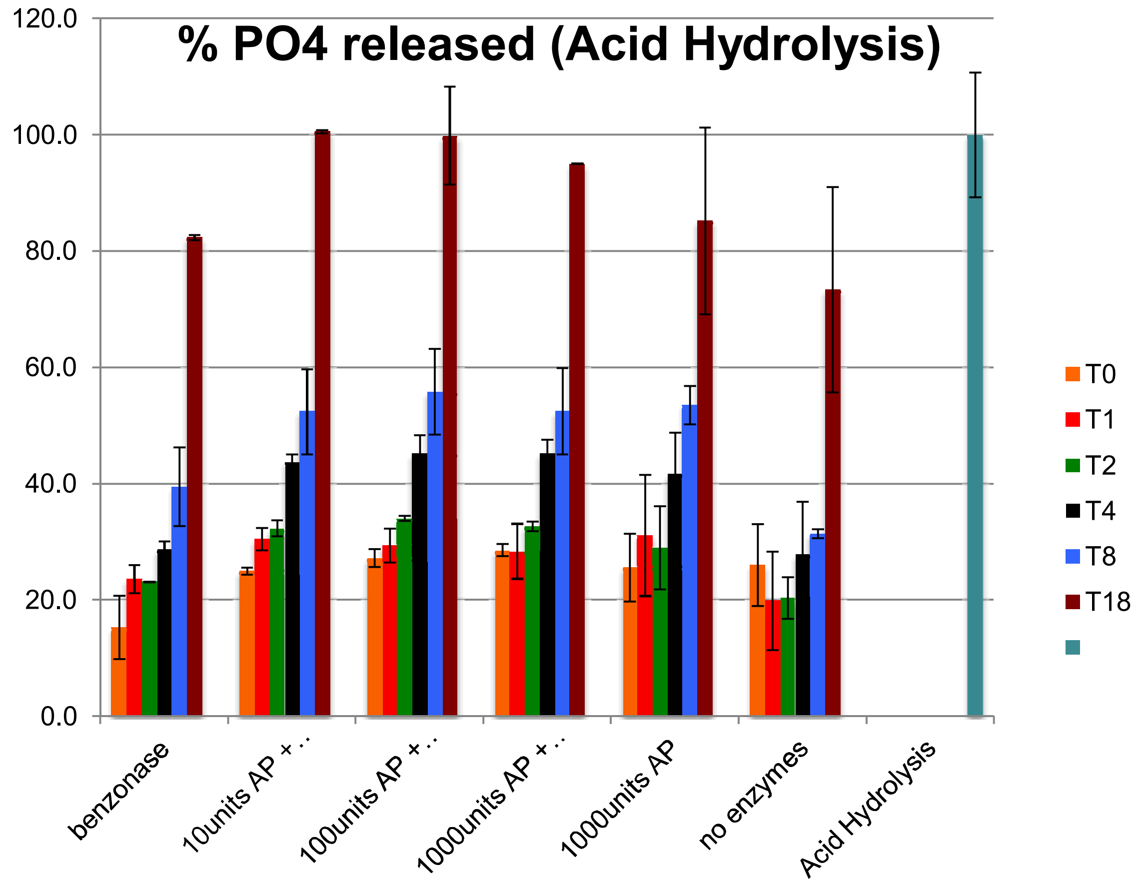
Sandia National Laboratories

Staged approach to development of a one pot reaction

- Test different cocktails for remineralization of phosphate and phospholipid conversion.
  - Determine rate of remineralization
  - Optimize reaction conditions (extraction of nutrients from solid phase)
  - Identify recalcitrant pools
  - Minimize reaction time
- Develop microbial consortia with appropriate enzymatic activities: test culture supernatants.
  - Identify candidate genes, clone, overexpress
  - Test for protein and activity level
- Grow microbial consortia on residual algal biomass—expressing enzymes *in situ* and converting amino acids to ammonium.
  - Optimize growth conditions (limit conversion to microbial biomass)
  - Optimize enzyme production on residual biomass
  - Limit uptake of inorganic phosphate by microbial consortium

Technical accomplishments.

Remineralization of "total" cellular Phosphate

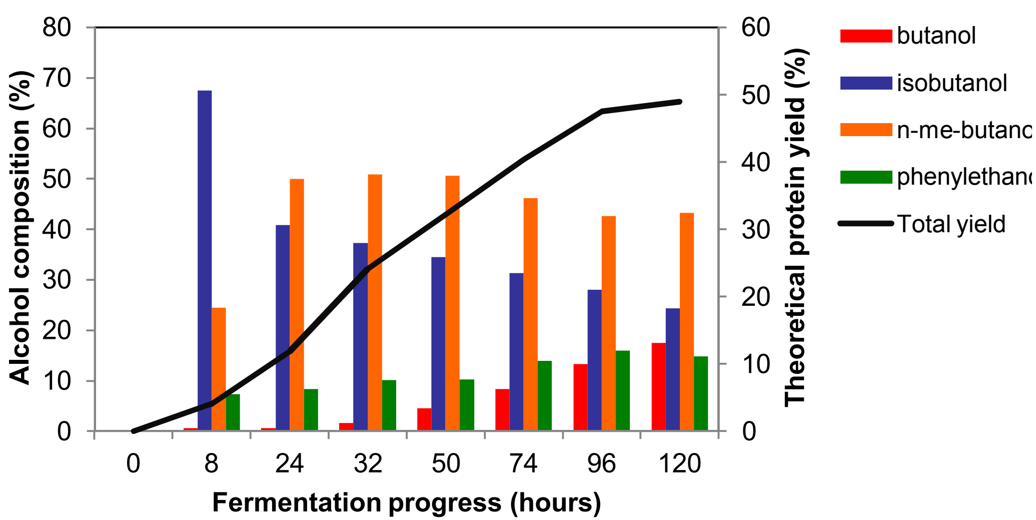


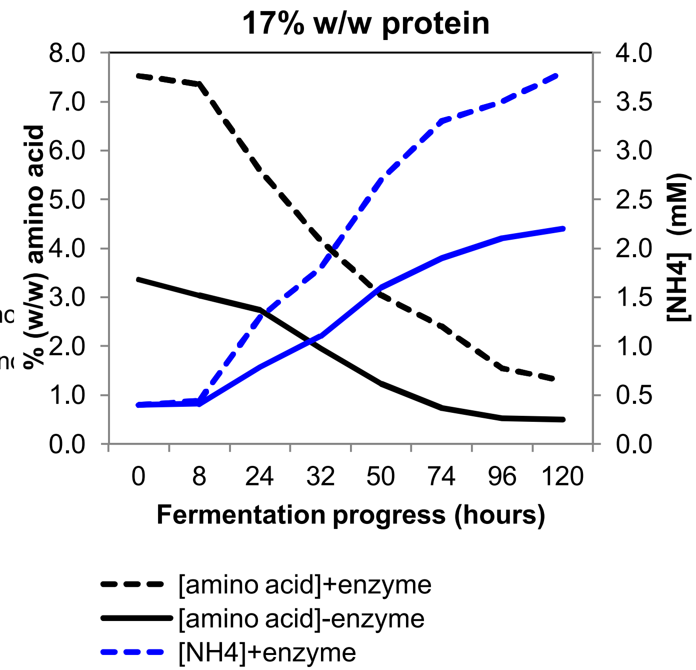
We will take advantage of NH_4 production by Previously developed fermentation process

Goal: Develop process to maximize conversion of microalgal residuals to fuels
Strategy: Subject lipid extracted biomass to microaerobic fermentation using an alcohol tolerant metabolically engineered *E. coli* strain, (Xin Huo, et al *Nature Biotech*, 2011)

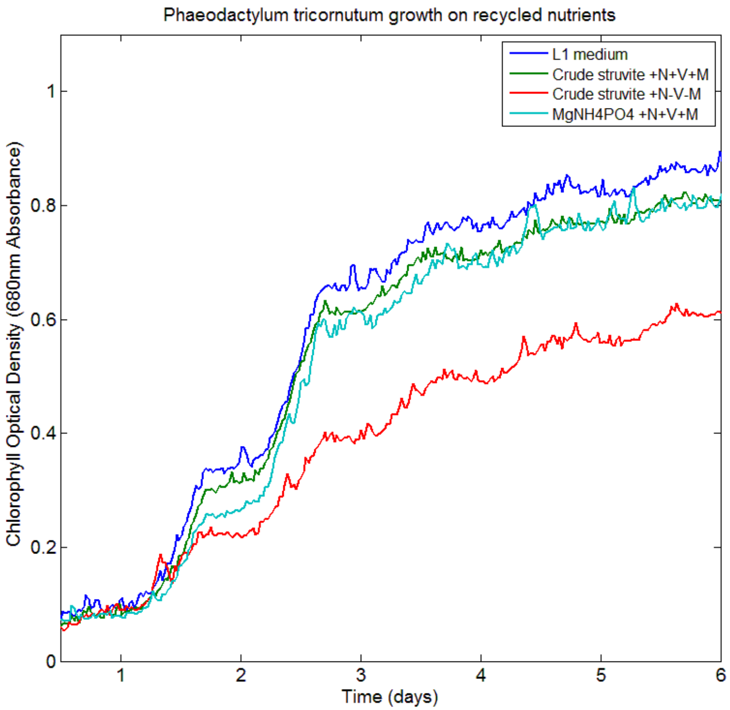
<u>Results</u>

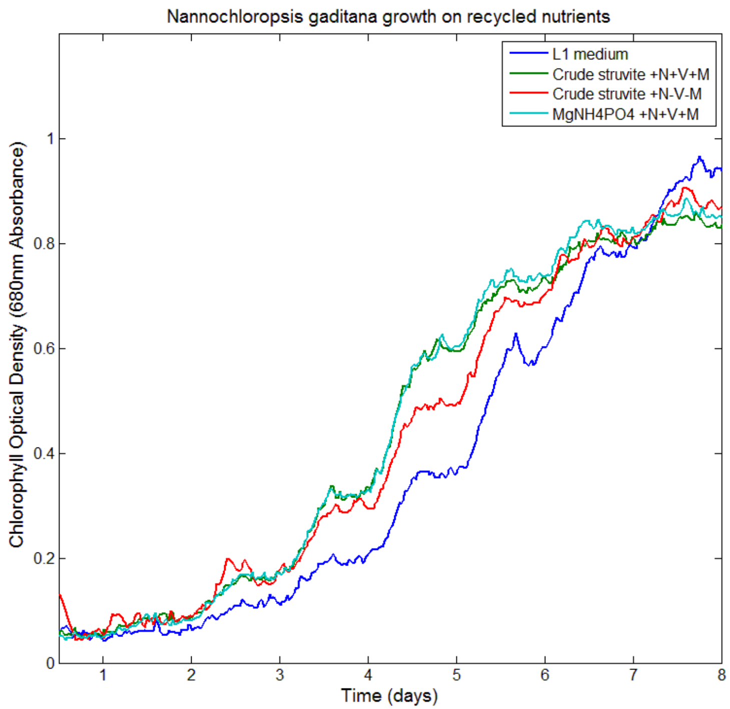
1) Mixed alcohols = 50% protein yield using 5-step process:
dilute acid pretreatment -> ethanolic fermentation -> distillation ->
enzymatic digestion (proteins) -> microaerobic fermentation (37°C, 96-120 hrs)
2) Alcohol components do not significantly vary with biomass type
3) Accumulation of alcohols proceeds in distinct temporal phases:
isobutanol -> n-methyl-butanol -> phenylethanol and n-butanol
4) <u>Ammonium is accumulated in fermentation liquor as amino acid breakdown product</u>





Struvite can replace "new" nutrients in microalgal culture





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

Spare slides

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.

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.

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

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

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

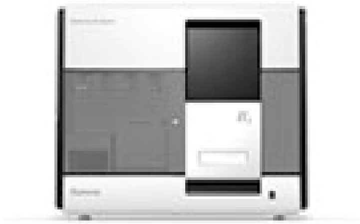
- Development of probe sets for persistent pest
  - Additional sequencing of molecular barcode regions
    - Full SSU, LSU, ITS regions, mtCO etc.
  - Informatic analysis
    - Define the most informative regions
  - Probe development, validation, optimization
    - qPCR
    - Specificity, cross reactivity,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

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

SANDIA BIOSCIENCE

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

Instead of fractionating biomass extensively we create targeted libraries for sequencing

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

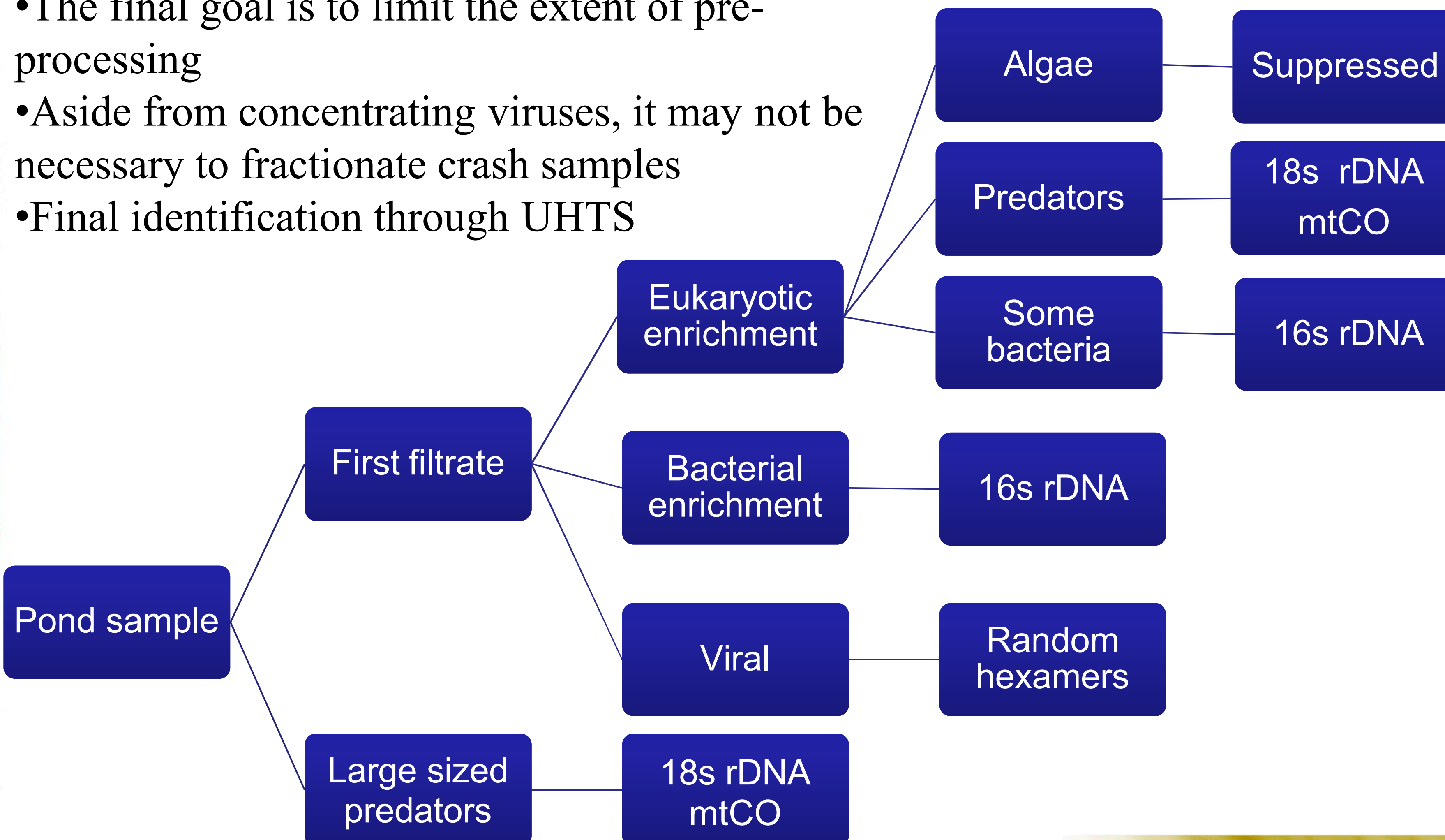
Subtraction:
rDNA analysis or metagenomic analysis
Physically removes unwanted sequences

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

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



Sandia has extensive experience with developing and fielding CB defense solutions

Detection Systems

BioWatch Gen 2
BioWatch Gen 3
Tenix Water Monitor
RDCDS
μGasAnalyzer
PROTECT
RapTOR

ConOps Development & Support

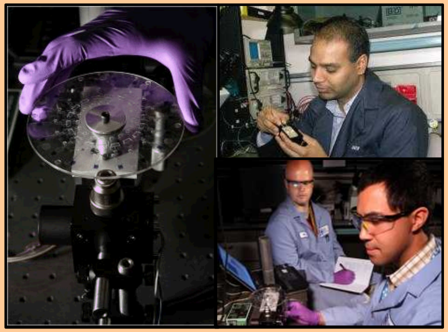
BioNet
BioWatch Incident Characterization (BWIC)
BioWatch Indoor Reachback Center (BIRC)
PROACT

Decontamination & Restoration

Sandia Decon Foam
DF-200
IBRD
WARRP
Validated Sampling Procedures
BROOM
AWARE

Medical Diagnostics

Toxin Diagnostics
Oral Diagnostics
Radiation Biodosimetry
Rickettsial Diagnostics
NASA astronaut diagnostics



Sandia Bioscience