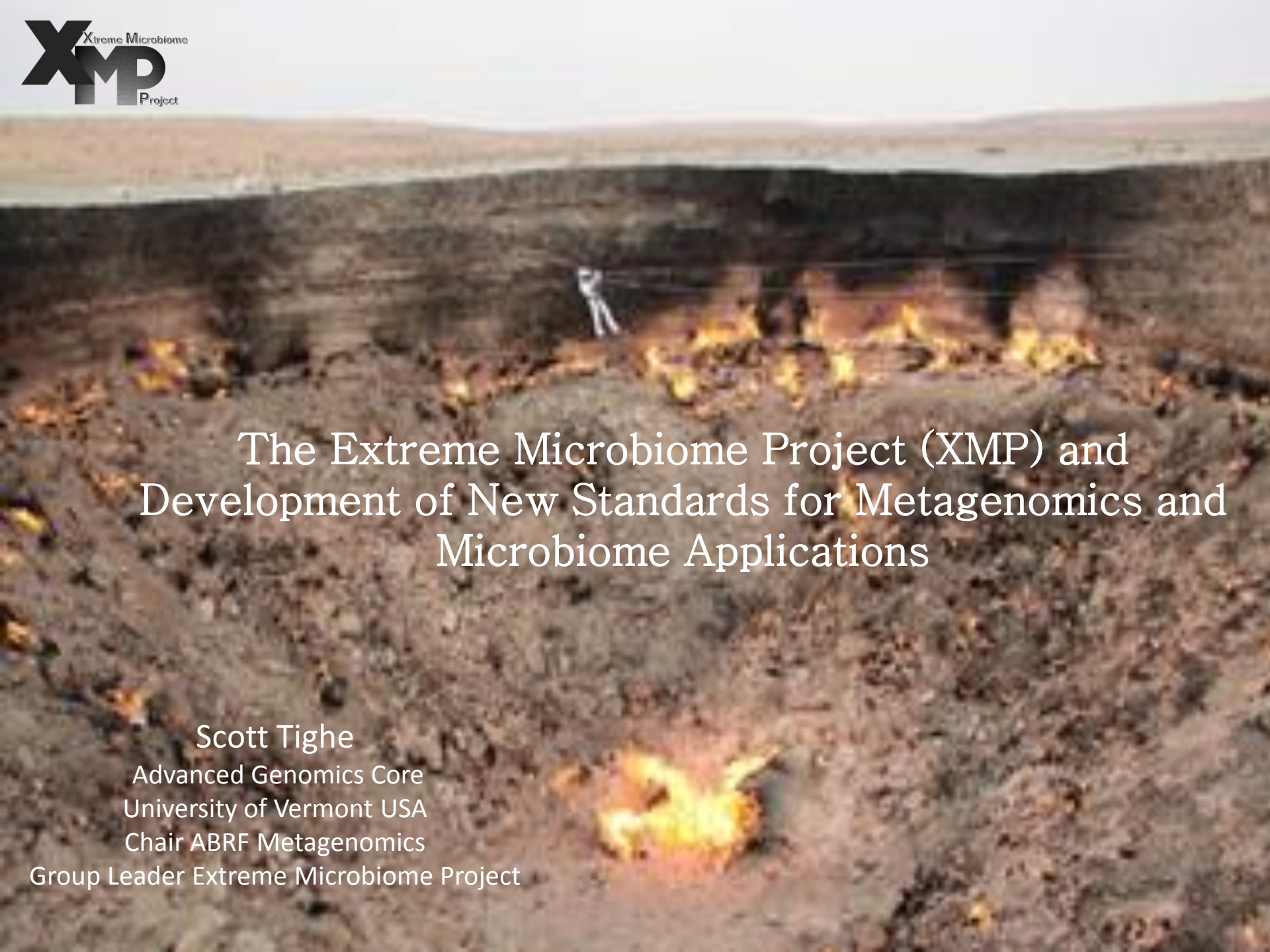


The Extreme Microbiome Project (XMP) and Development of New Standards for Metagenomics and Microbiome Applications

Scott Tighe
Advanced Genomics Core
University of Vermont USA
Chair ABRF Metagenomics
Group Leader Extreme Microbiome Project

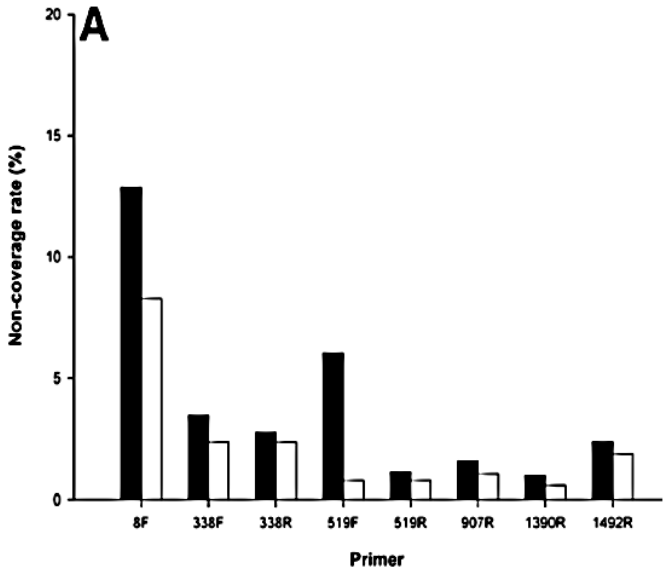
Metagenomics and the Microbiome

- A developing and very exciting field in all disciplines of biology that uses complex genetic analysis to detect and describe all organisms in a complex community.
  - Bacteria, viruses, fungi, insects, dinosaurs, anything with DNA and RNA
  - Includes environmental, clinical, commensal, parasitic, industrial, astronomical
  - Important roles in health and disease: Gut, Mouth, Vagina, Skin (diverse sites: Nasal epithelial, neonatal, lung, blood, liver, urine, Crohn's, Acne
  - Will, undoubtedly, expand into all areas of biology and medicine in the next 20 years
  - However, there are many challenges with detecting everyone in a mixed community

.

Challenges with Metagenomics

- Sampling (ANAF)- Aseptic Nucleic Acid Free
  - Filter Funnels
  - Swabs
  - Bottles
  - Reagents
- DNA/RNA extraction
  - Selective lysis
  - Non-lysis
- Library Amplification
    - GC bias
    - Sample loss due to clean up
    - Amplification Bias
        - 16s rDNA, ITS, or *gyrase* B (gyrB)
        - Primer Non-coverage can bias
- Software



Coverage evaluation of universal bacterial primers using the metagenomic datasets Dan-Ping Mao, Quan Zhou, Chong-Yu Chen and Zhe-Xue Quan BMC Microbiology 2012, 12:66

Aquatic Metagenomics

Collect samples in Sterile DNase, RNase, DNA, and RNA-free

Culture

Genomics

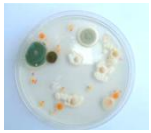
Microscopy

Growth/Culturing

Heterotrophic Plate Counts on R2A (aerobically/Anaerobically/22C/37C) for Bacteria

Culture Fungi , Culture algae (BG11)

Count and ID each morphotype



MICROSCOPICALLY IDENTIFIABLE ORGANISMS

Algae, Nemotodes, rotifers, some bacteria

I. Lugol's:

Transfer 50 mls of each sample to sterile 50 ml conical tube and add Lugols Iodine to a final of 1%

Store at 4 C

Highgate Springs VT	5833		5951		6058		6149		6196	
	Lugols	centri	Lugols	centri	Lugols	centri	Lugols	centri	Lugols	centri
Anabaena	1	0	5	24	45	73	16	16	0	2
Melosira	4	16	436	604	44	42	110	149	5	13
Sphaerocystis	7	5	0	0	0	0	0	0	0	0
Aphanocapsa	3	3	0	0	0	0	0	0	0	0
Coelastrum	2	3	0	0	0	0	0	4	0	4
Oocystis	1	0	0	0	0	0	0	11	0	0
Pediastrum	1	0	2	0	2	6	2	2	1	0
Fragilaria	0	40	1	0	2	0	0	2	0	4
Microcystis	0	0	2	0	0	25	0	13	0	0
Aphanizomenon	0	0	0	0	27	95	35	109	3	19
Ceratium	0	0	0	0	1	0	0	0	0	0
Staurastrum	0	0	0	0	0	0	1	4	0	0
Dinobryon	0	0	0	0	0	0	0	0	2	0
UGC	0	227	0	910	0	1127	0	998	0	56
UCBG	0	91	0	490	0	66	0	368	0	14
Gloeocystis	0	8	0	0	0	0	0	0	0	0
Crucigenia	0	3	0	0	0	0	0	0	0	0
Pennate Diatom	0	3	0	6	0	67	0	58	0	83

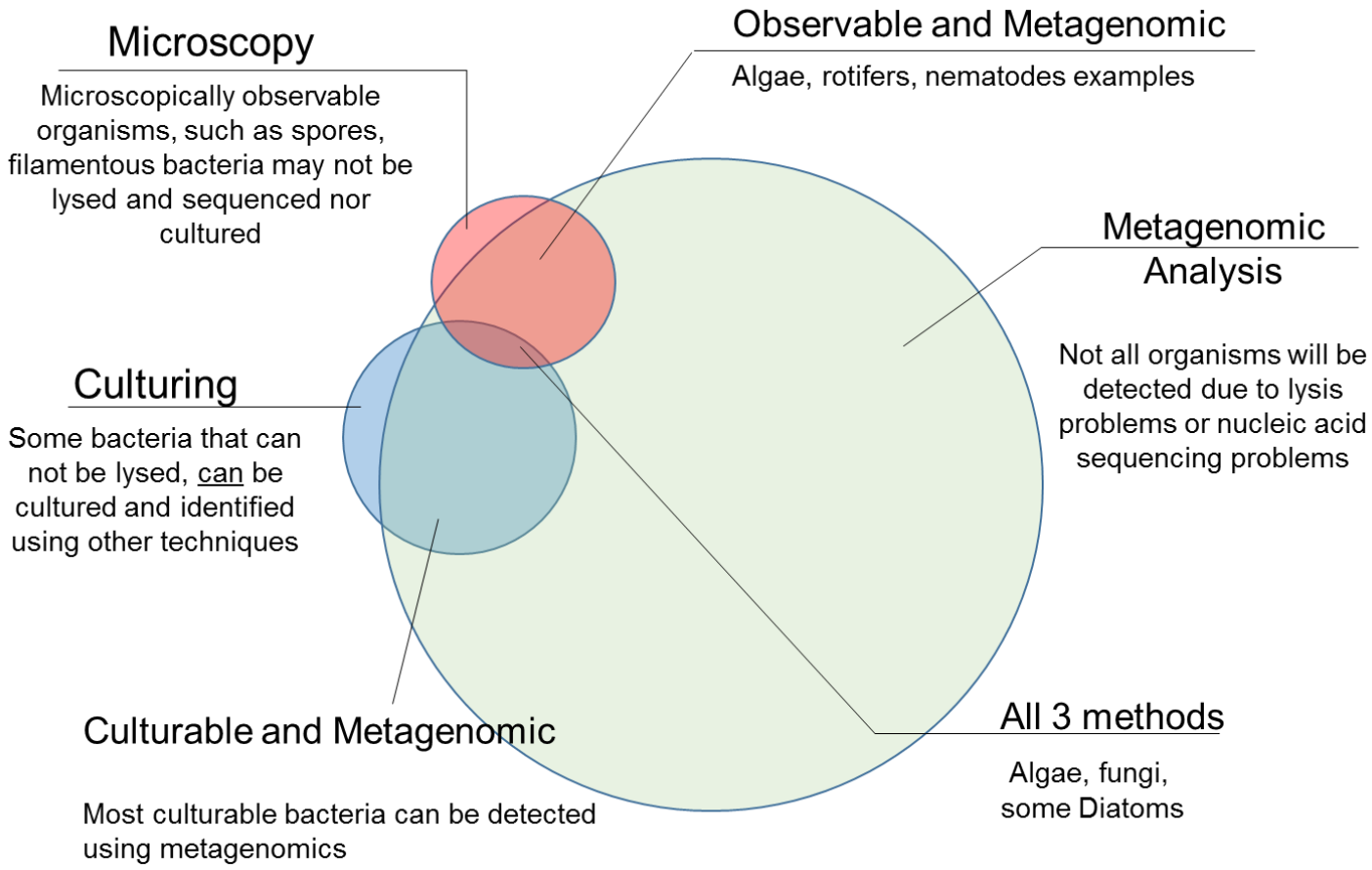
DNA SAMPLE

DNA free filter funnel with NTE-PC 0.2um membrane installed. Filter and transfer in a 50 ml Filter 1-3 membranes worth. FastPrep with Scalpel blade. Centrifuge to concentrate

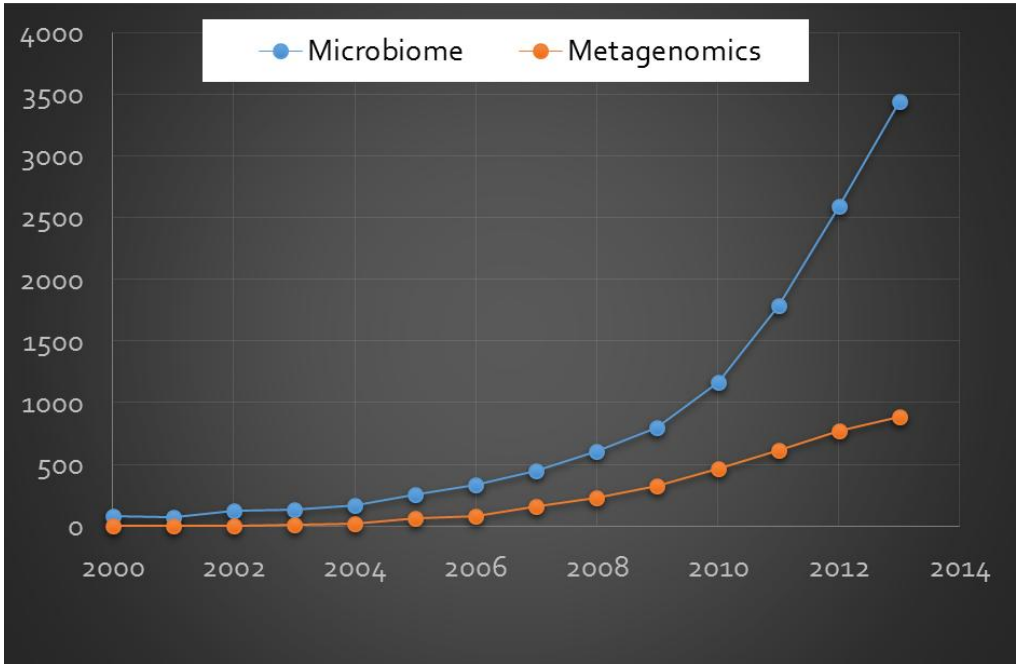
Enzymes, abrasive, Bead beater,

Enzymes→CTAB→ SDS → AlO3 abrasive, →double phenol→ Column

What you see is not always what you get



Metagenomic analysis is rapidly growing



12623

3549

Search terms in End Note on PubMed 5/15/2014: Microbiome, Metagenomic, Metagenomics, Metagenetic]

- In 2013 The Establishment of the open access ***Journal of Microbiome***
  Journal ID ***: Microbiome***

Projects in Metagenomics

PathoMAP

NIH HUMAN MICROBIOME PROJECT

Xtreme Microbiome Project

Alkek Center for Metagenomics and Microbiome Research

JGI JOINT GENOME INSTITUTE DEPARTMENT OF ENERGY

the kitten microbiome project.
the cutest, fluffiest microbiome project ever, coming soon to a litterbox near you.

earth microbiome project

BCM Baylor College of Medicine

BROAD INSTITUTE

Hospital Microbiome

J. Craig Venter INSTITUTE

Home microbiome study

Xtreme Microbiome Project

What makes the XMP project different for other microbiome projects?

- Sample coming from extreme environment
  - Volcano's, Craters, extreme ocean depths, ect
- Whole genome shotgun sequencing PE 2x250 Illumina
- Life Tech Ion Torrent 400bp
- Some PacBio and Oxford Nanopore
- Both RNA and DNA
- Potential for discovery-new Species and new biomolecules
- XMP Samples are perfect for optimization studies

ABRF
Metagenomics
Research Group

Background of XMP and the ABRF Metagenomics Consortium

- Research group established under the Association of Biomolecule Research Facilities (ABRF) in 2014
- Includes an Analytics group, Bioinformatics, and "Earth sciences"
  - Microbiologists, geneticists, computational biologist, geochemists, oceanographers, ecologists.
- Development of new standardized techniques, protocols, products
- Working with corporates partners for developing new kits and reagents
- Study extreme and unique environments with shotgun sequencing
  - Test new protocol on worst case samples
  - Develop new approaches to difficult sample
  - Discovery potential

The MGRG XMP team

• Scott Tighe (SCIENTIFIC LEAD)	University of Vermont
• Christopher Mason (COMPUTATIONAL LEAD)	Weill Cornell Medical College
• Ebrahim Afshinnekoo	Weill Cornell Medical College
• Noah Alexander	Weill Cornell Medical College
• Samantha Joye	University of Georgia
• Rita Colwell	University of Maryland
• Nadim Ajami	Baylor COM
• Russ Carmical	Baylor COM
• Don Baldwin	Signal Technologies
• Nathan Bivens	University of Missouri
• Stefan J Green	University of Illinois at Chicago
• Jodie Lee	American Type Culture Collection (ATCC)
• Shawn Levy	HudsonAlpha Institute for Biotechnology
• Ken McGrath	Australian Genome Research Facility
• Natalia G. Reyero Vinas	MIssissippi State University
• Matthew L Settles	University of Idaho
• Kelley Thomas	University of New Hampshire
• Sarah Johnson	Georgetown University
• Ian Charold Herriott	Univ of Alaska Fairbanks
• Mohamed Donia	Princeton University
• Tim C Hunter (ABRF EB Liaison)	University of Vermont

The XMP team

• Diana Krawczyk	Greenland Institute of Natural Resources
• Jill Mikucki	University of Tennessee/Middlebury
• Svein-Ole Mikalsen	University of Faroe Islands
• John M Lizamore and Don Cater	Western Australia Government

• TBD	Illumina Corp
• Anjal Shah, Ashley Wilson	Life Technologies
• Mike Farrell	Omega Bio-Tek
• Fiona Stewart	New England Biolabs
• Adam Morris	BioO Scientific
• Aaron Sin, Bob Gates	Sigma Chemical
• Liz Kerrigan	American Type Culture Collection
• Tauni Wright and Jason Struthers	Steritech Filtration

ABRF Metagenomics Research Group

Current Projects

- ✓Microbial Reference Standard
  - Whole Cell
  - Quantitative gDNA
  - Distributed through the ATCC

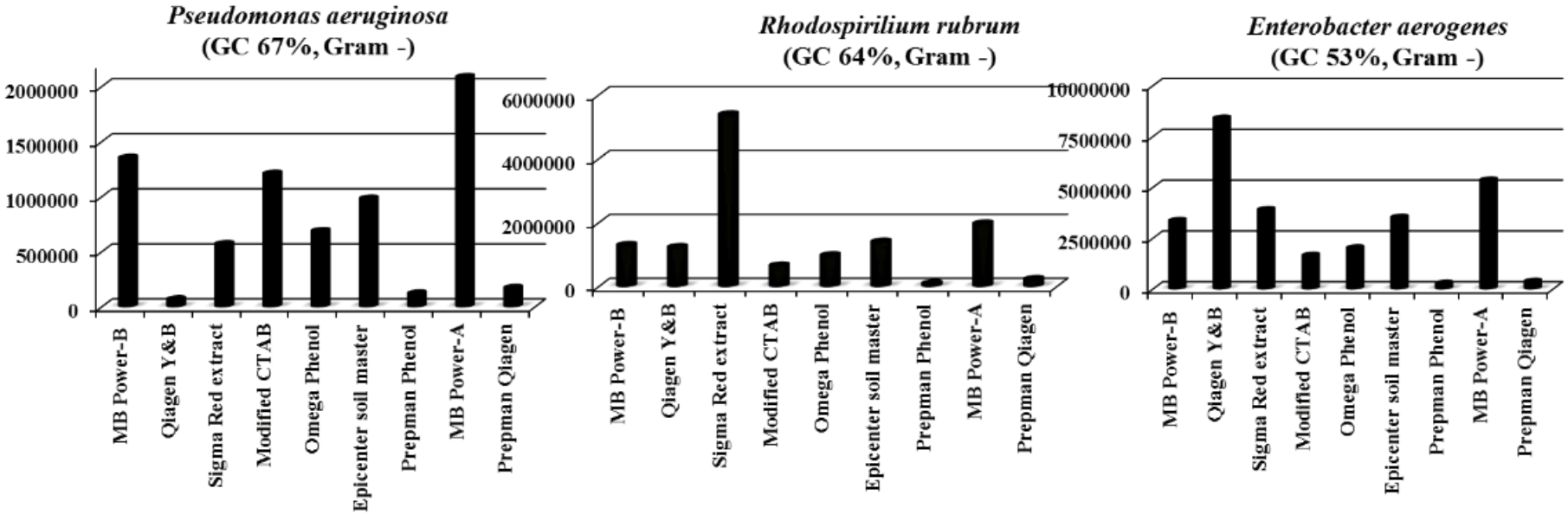
- ✓DNA and RNA extraction protocols

- ✓Sigma PolyZyme-Mulilytic enzyme mix
  - 7 enzymes mix

- ✓Current data from the Extreme Microbiome Study (XMP)
  - Penguin Microbiome
  - The Doors to Hell Gas Crater
  - Lake Hillier Western Australia

DNA extraction efficiency is highly variable

- Believe it or not, a bacteria or fungus that you can grow, might not release any DNA or RNA and will be missed
- No Standard method
- Different techniques have varying yields
- No standard Enzyme Mix…MACL4



ABRF Nucleic Acids Research Group 2012-2013 Study

Evaluating DNA Extraction Methods for Metagenomic Analysis

V. Nadella[1], J. Holbrook[2], R. Carmical[3], M. Robinson[4], C. Rosato[5], H. Auer[6], N. Beckloff[7], Z. Herbert[8], S. Chittur[9], A. Perera[10] , W. Trimble[11], S. Tighe[12]

[1]Ohio University, [2]Nemours/A.I. DuPont Hospital for Children, [3]University of Texas Medical Branch, [4] University of Zurich, Switzerland, [5]Oregon State University, [6]Institute for Research in Biomedicine, Barcellona, Spain, [7]Case Western Reserve University, [8]Dana Farber Cancer Institute, [9]University at Albany-SUNY, [10]Stowers Institute for Medical Research, [11]Argonne National Laboratory, [12]University of Vermont.

ABRF
Metagenomics
Research Group

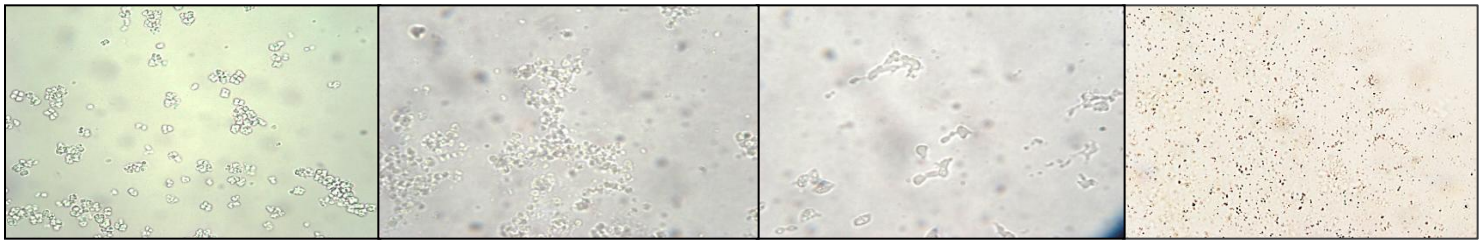
Enzymes increase DNA Extraction Efficiency

- Current Extraction Reagents need higher efficiency
- Multi-lytic Enzyme Mix for Digestion of Cell Walls

The MRGR is collaborating with Sigma Chemical develop a multi- enzyme product that may include

Metagenomics Polyzyme

- *Achromopeptidase*
- *Mutanolysin*
- *Chitinase*
- *Lyticase*
- *Lysozyme*
- *Lysostaphin*
- *Labiase*
- *Others???*



Sphearoplasting vs Exposure time and 0.1% SDS

But how do you study DNA extraction Efficiency without a control?

Assembly of Class I and Class I + Microbial Reference Standards

XMP, ABRF, ATCC, NIST

Class I: Contains few repetitive sequences except for the ribosomal operons (5-7 kbp); can be reliably sequenced using short reads

Class II: Contains many repetitive sequences, such as insertion elements, but none greater than 7 kbp; a PacBio can provide a complete assembly, but short reads will offer fragmented contigs

Class III: Contains large repetitive sequences of >7 kb PacBio will offer a higher quality but will not be able to provide a complete genome

Class I Standard Selections

Class	Repeats	Max Repeats	Genome	Gram	M/O	GC Content	Growth Methods
Class I	55	5110	2564615	+	*Staphylococcus epidermidis* ATCC 12228	32.8	Standard
Class I	91	5260	4170008	+	*Halobacillus halophilus* ATCC 35676	46.8	Marine Broth Agar 2216
Class I	65	4153	2501097	+	*Micrococcus luteus* NCTC 2665 ATCC 4698	72	Standard
Class I	28	5821	3850272	-	*Pseudoalteromonas haloplanktis* TAC125 ATCC 35231	40.1	Marine Broth Agar 2216
Class I	77	5463	4639675	-	*Escherichia coli* str. K-12 substr. MG1655/ATCC 700926	50.8	Standard
Class I	35	5825	6845832	-	*Pseudomonas fluorescens* F113 ATCC 13525	61.4	Standard

Class I+ Standard Additions

Class	Repeats	Max Repeats	Genome	Gram	M/O		Growth Methods
Class I	19	5399	2739625	+	*Enterococcus faecalis* OG1RF ATCC 47077	37.2	Standard
Class I	39	6625	2008345	-	*Zymomonas mobilis* subsp. *mobilis* ATCC 29191	46	Standard
Class I	43	5750	4751080	-	*Chromobacterium violaceum* ATCC 12472	64.8	Standard
Class I	27	5837	4215606	+	*Bacillus subtilis* subsp. *subtilis* str. 168 ATCC 23857	43.5	Standard
Class I	90	5547	4012900	N/A (Archea)	*Haloferax volcanii* DS2 ATCC 29605	65.5	Halobacterium medium 974

http://genomebiology.com/2013/14/9/R101

ABRF Metagenomics Research Group

Oxford Nanopore Minion Data (MAP)

- DNA extracted from each organism and mixed equal nMoles based on Qubit, 1 ug run in duplicate on Version 7.3 flow cells, Standard Shearing as per MAP program
- Analysis of "2D" read
  - 534 reads for 612,038bp. Mean length was 1,146bp. Ranged from 200 to 8,299bp
- MetaPhlan produced no matches. Data still too noisy.
- BWA-MEM was used to force align to the six species.

Ont2d

- Pseudomonas flourescens: 39/179 = 21.79%
- E. coli: 37/178 = 20.79%
- Halobacillus halophilus: 26/179 = 14.61%
- Pseuoalteromonas haloplanktis: 23/179 = 12.57%
- Staphylococcus epidermidis: 4/179 = 2.25%
- Micrococcus luteus: 1/179 = 0.56%

Yogesh

- Pseudomonas flourescens: 23
- E. coli: 76
- Halobacillus halophilus: 47
- Pseuoalteromonas haloplanktis: 20
- Staphylococcus epidermidis: 4
- Micrococcus luteus: 0

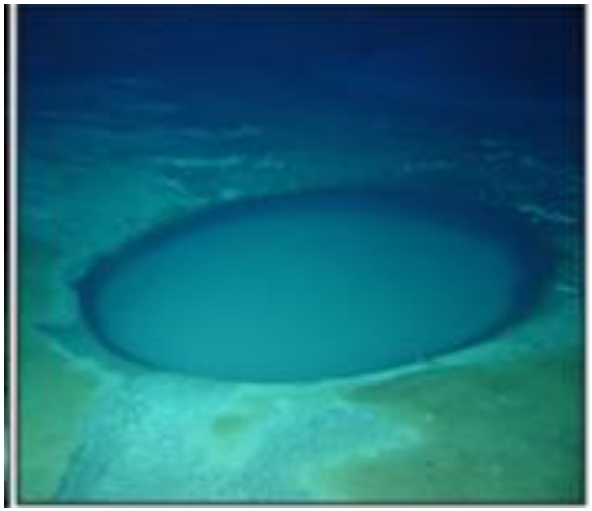
lastz

- **Pseudomonas flourescens: 43**
- **E. coli: 82**
- **Halobacillus halophilus: 71**
- **Pseudoalteromonas haloplanktis: 20**
- **Staphylococcus epidermidis: 3**
- **Micrococcus luteus: n/a**

What’s in the pink lake...

Study Sites of XMP

Deep Ocean Brine Lakes (Joye Lab)



Deep Ocean Brine Lakes (Joye Lab)
Brine lake are located at a depth of 15000.
Collected using the Alvin by Samantha Joye

High Acidity Saline lakes (Johnson Lab)

Located in Western Australia, Sarah Johnson collects sample for these pH1.5 20% saline ponds.

Lake Hillier (McGrath Lab)

With support from the Western Australia government, XMP members collected samples from LH. Located in Western Australia off the coast in the archipelago, Lake Hillier sits in the middle of a small remote island. pH 7.6 salinity 28%

Study Sites of XMP

Greenland deep ocean silt

Diana Krawczyk from Greenland Institute of Natural Resources focuses on diatom research and collects sample from deep ocean of Western Greenland to study popukation shifts on a geological time scale

Alaskan Permafrost

Several members of the XMP team are focused on the geological shift and microbial populations of borings form permafrost.

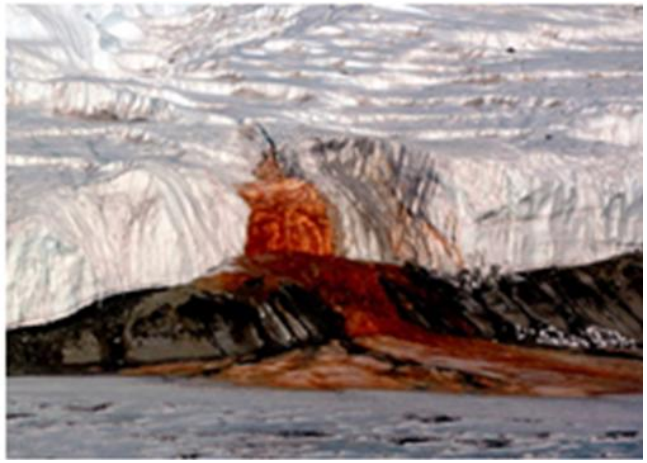
Doors to Hell Gas Crater (Greene Lab)

A recent trip by XMP member Stefan Green to the Doors from Hell gas crater included metagenomic sampling. Culturing, 16s, and Shotgun sequencing were completed

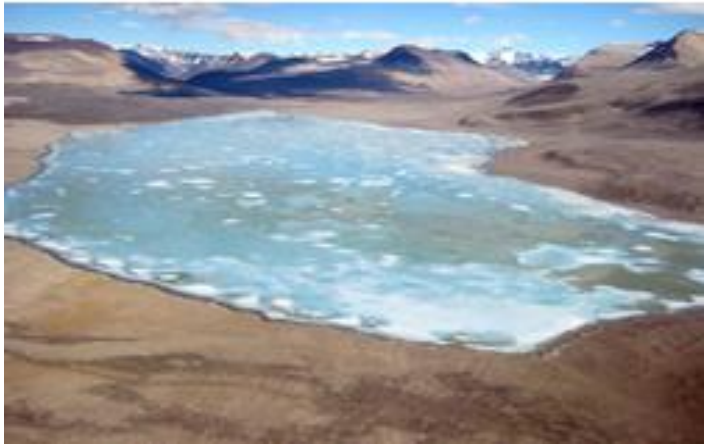
Study Sites of XMP

Sites in Antarctica include the hyper-oxide saline rich Blood Falls and Hyper-saline lakes.

Expeditions of Mandy Joye (MGRG member) and Jill Mikucki (University of Tenn/Middlebury College)



Blood Falls of the Taylor Valley
McMurdo Dry Valleys in Victoria Land, East Antarctica.



Hyper-saline lakes of Antarctica-Don Juan Pond, Lake Vanda

Fecal Microbiomes

Comparative microbiome studies of low fat vs high fat storage

Emperor Penguin Samples collected by Vladimir Samarkin of Samantha Joyes lab.

Hummingbird
(Costa Rica)
Samples to be collected by Ian Herriott July 2015 using NAF apparatus

Extreme Microbiome Project (XMP)

XMP Bioinformatics tools

- BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi)
- MetaPhlAn (https://bitbucket.org/biobakery/metaphlan2)
- Kraken (https://ccb.jhu.edu/software/kraken/)
- PhyloSift (https://phylosift.wordpress.com/)
- GOTTCHA (https://github.com/poeli/GOTTCHA)
- CosmosID
- Mg-RAST

RESULTS

Xtreme Microbiome Project

"*Doors to Hell" Gas Crater*

Darvaza, Karakum Desert, Turkmenistan

Sampled by: Stefan Green (ABRF MGRG)

DNA Extracted: 10 Grams yield 438 pg

DNASeq Library: Rubicon ThruPlex 20 cycles

Sequencing: Natalia Reyero MGRG
MiSeq PE 2x250 MSU

Data Analysis: Ebrahim Afshinnekoo MGRG
MetaPHan, MegaBlast

```
Nocardioides sp. JS614 ........................ 1 hits 1 orgs
Pimelobacter simplex .......................... 1 hits 1 orgs
Propionibacterium avidum 44067 ................ 1 hits 1 orgs
Catenulispora acidiphila DSM 44928 ............ 1 hits 1 orgs
Stackebrandtia nassauensis DSM 44728 .......... 1 hits 1 orgs
Streptosporangium roseum DSM 43021 ............ 1 hits 1 orgs
Leifsonia xyli ................................ 2 hits 2 orgs
Streptomyces cattleya DSM 46488 ..............2 hits 1 orgs
Kitasatospora setae KM-6054 .................. 1 hits 1 orgs
```

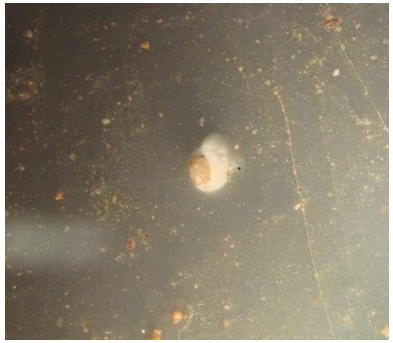
The Door to Hell is noted for its natural gas fire which has been burning continuously since it was lit by Soviet petroleum engineers in 1971.[1] The fire is fed by the rich natural gas deposits in the area. The pungent smell of burning sulfur pervades the area for some distance

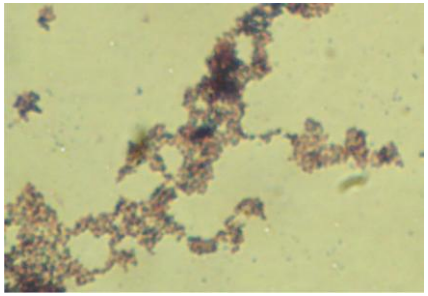
Culturing the Crater Material

40mg plated to TSA 28C for 50 days
4 cfu (2 morphotypes)

Arthrobacter tumbae strain NIOT-Ch-19 bacteria isolated from deep sea sediments of Bay of Bengal and Andaman Sea

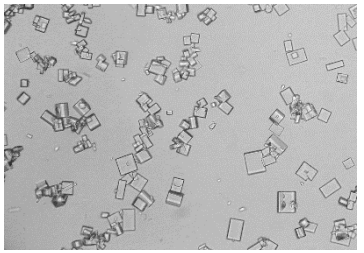
Arthrobacter subterraneus sp. nov., isolated from deep subsurface water of the South Coast of Korea



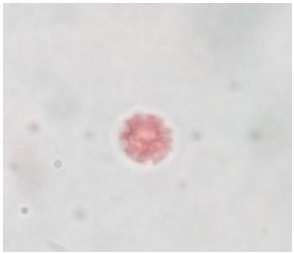


Lake Hillier

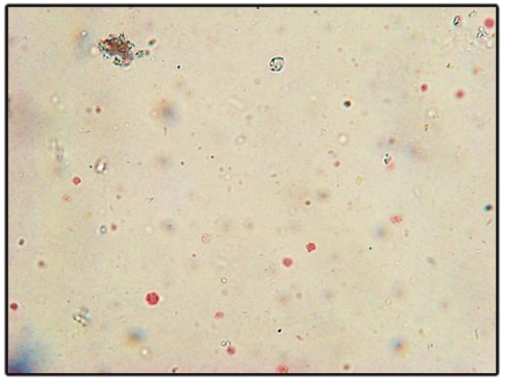
Australia's Recherche Archipelago



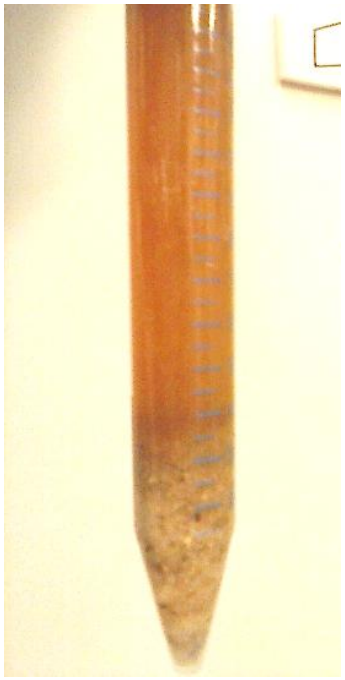
- Extreme Hyper saline shallow lake- 25% during sampling
- Salt precipitates out of solution instantly
- pH 7.4 at 26C



Dunaliella sp ??

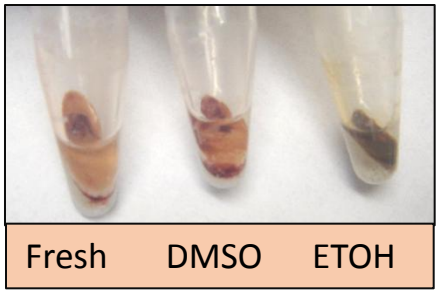


Lake Hillier
Australia's Recherche Archipelago



- Tested three collection Preservatives
  - ETOH, DMSO. Fresh (cold)
- Extracted RNA (Trizol LS) DNA (MAC4L-Omega)
- Tested two processing protocols

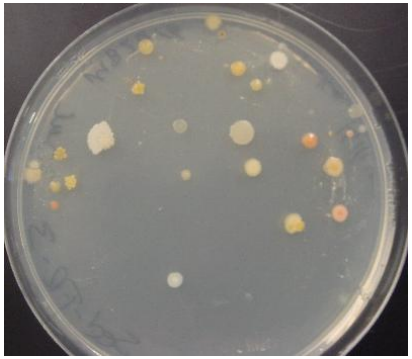
  Diluted and Filtered
  Diluted and Centrifuged



Method	Sample	Volume (mL)	Total RNA (25ul)	Total DNA (25ul)
Filter Process	Sed-Fresh-filtered	0.5	ND	7.75
	Sed-ETOH-filtered	1.7	50.75	192.5
	Sed-DMSO-filtered	1.7	35	327.5
	Water-Mid-fresh-filtered	7.5	27.5	23.3
	Water-Mid-ETOH-filtered	7.5	ND	10.0
	Water-Mid-DMSO-filtered	7.5	ND	105.0
Direct	Sed-Fresh-Direct	0.2	55	55.0
	Sed-ETOH-Direct	0.2	37.5	15.0
	Sed-DMSO-Direct	0.2	37.5	97.5
	Bank-Fresh-Direct	0.2	NA	627.5
	Bank-ETOH-Direct	0.2	950	520.0
	Bank-DMSO-Direct	0.3	NA	560.0

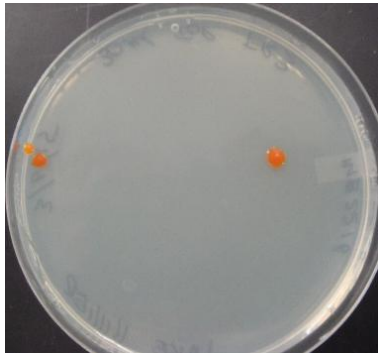
Lake Hillier -Culturing

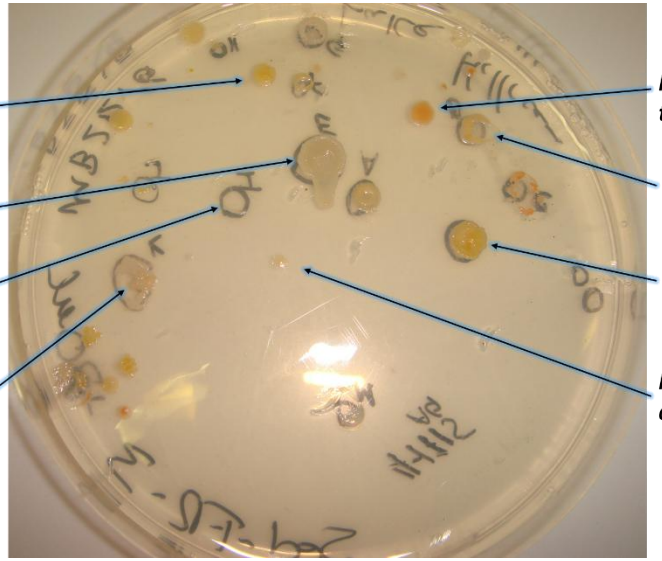
Australia's Recherche Archipelago

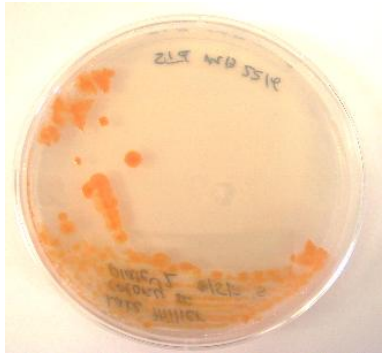


20ul Sediment

20ul Water

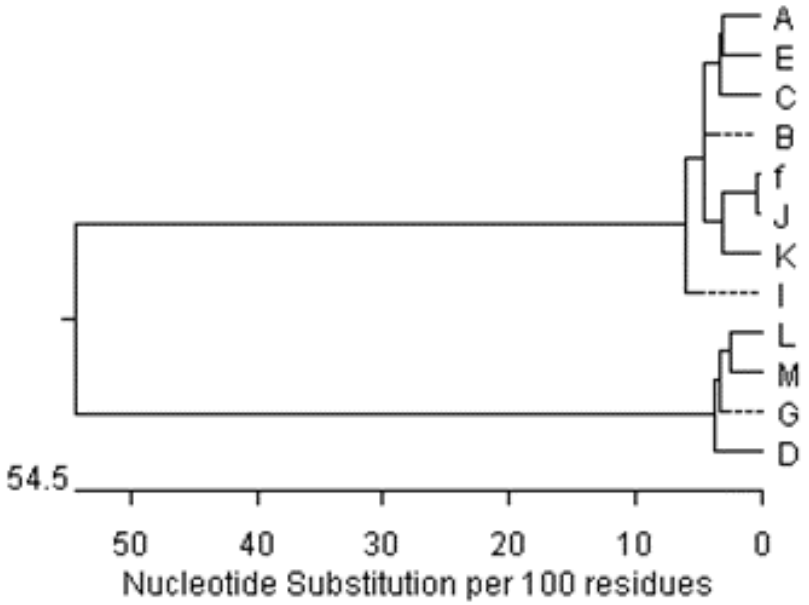






Psychroflexus tropicus

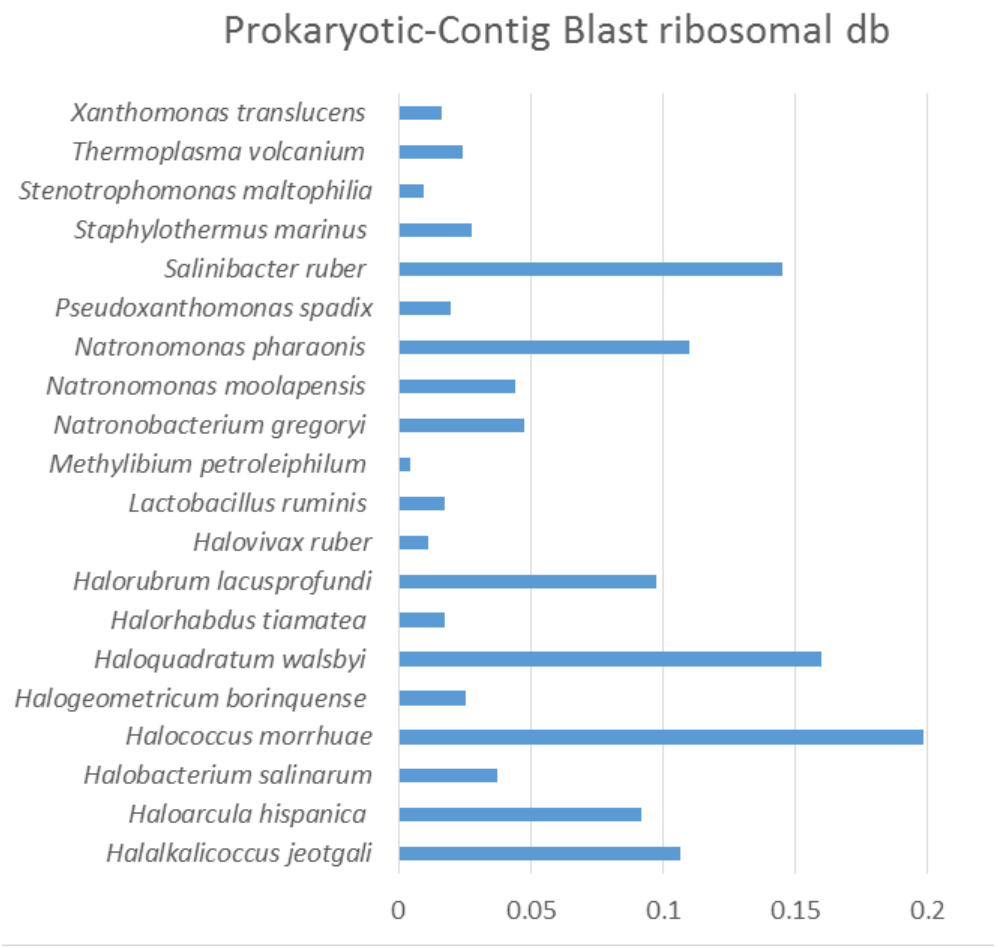
Full length 16s Sanger

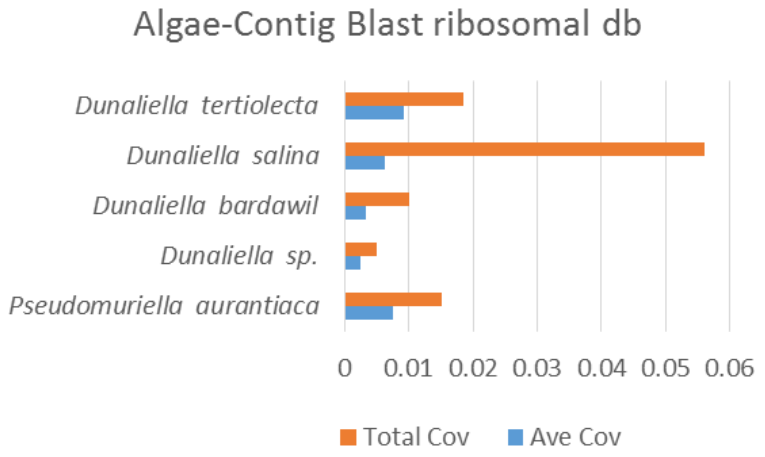


Colony	1st BLAST Result		2nd BLAST result	
A	Bacillus halmapalus	1400/1418	Bacillus zhanjiangensis	1393/1418
B	Bacillus aquamaris	1377/1407	Bacillus marisflavi	1373/1406
C	Bacilllus luteolus	1399/1423	Bacillus humi	1370/1426
D	Parabacillus quinghaiensis	710/710	Paraliobacillus ryukyuensis	697/710
E	Geotgalibacillus campisalis	1408/1424	Geotgalibacillus marinus	1403/1426
F	Halobacillus halophilus	1437/1439	Same	Same
G	Bacillus korlensis	907/915	Bacillus koreensis	901/916
I	Bacillus foramenis	1396/1424	Bacillus subterraneus	1388/1424
J	Halobacillus alkaliphilus	1432/1438	Halobacillus halophilus	1431/1438
K	Virgibacillus marismortuis	1442/1443	Virgibacillus salarius	1442/1443
L	Bacillus halmapalus	989/997	Bacillus zhanjiangensis	984/997
M	Sediminibacillus albus	1256/1280	Virgibacillus koreensis	1221/1239

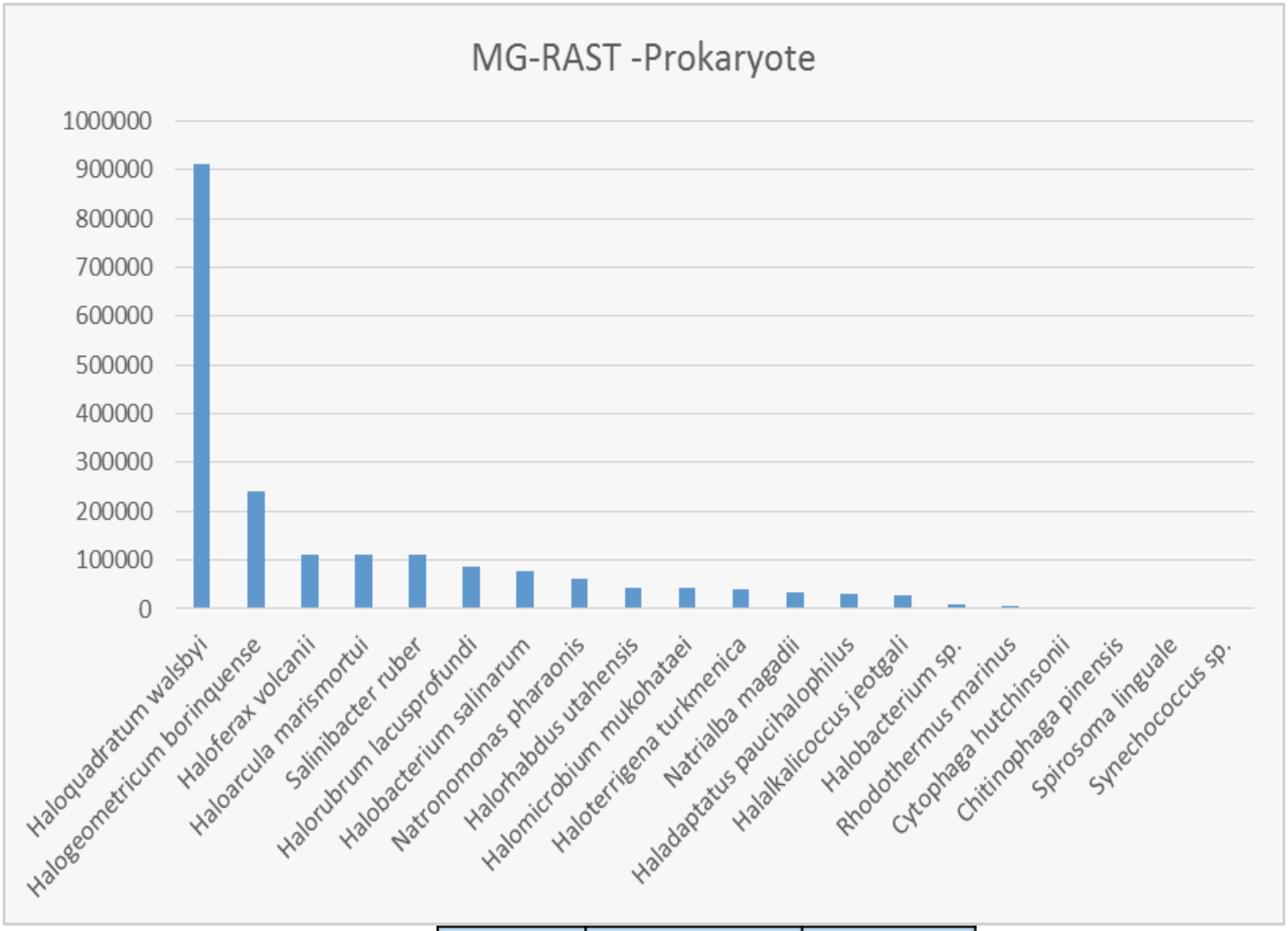
Metagenomics of Lake Hillier

- Contig assembly and Megablast Kelley Thomas UNH
- Mostly Halophilic archaea and *Dunaliella salina*





Cosmos Genius: Fungi:Melampsora_pinitorqua



Genus	Abundance	Avg eValue
Dunaliella	919	-68.79

Summary

- DNA/RNA extraction is the gatekeeper
- Shotgun metagenomics requires special software processing. One size does not fit all [yet].
- Results should be examined by a microbiologist when pathogens are detected.
- There are many biases in both laboratory processing and data analysis.
- Databases are less than perfect.
- Microscopy and culturing should be considered

www.extrememicrobiome.org