Life at the Extreme: The ABRF Metagenomics Research Group

Implementing New Standards in Metagenomics and the Extreme Microbiome Project

Don A. Baldwin (Signal Biology Inc.), Natàlia G. Reyero Vinas (Mississippi State U.), Ebrahim Afshinnekoo (Cornell U.), Nadim Ajami (Baylor College of Medicine), Noah Alexander (Cornell U.), Nathan Bivens (U. Missouri), Russ Carmical (Baylor College of Medicine), Stefan Green (U. Illinois - Chicago), Tim Hunter (Vermont Cancer Ctr.), Samantha Joye (U. Georgia), Diana Krawczyk (Greenland Inst. for Climate Research), Jodie Lee (ATCC), Shawn Levy (HudsonAlpha Institute for Biotechnology), Christopher Mason (Cornell U.), Ken McGrath (Australian Genome Research Facility), Darryl Reeves (Cornell U.), Matt Settles (U. Idaho), Kelley Thomas (Hubbard Ctr. Genome Studies), and Scott Tighe (U.Vermont)

The goal of the Metagenomics Research Group is to evaluate, develop, and refine methodologies for metagenomics and microbiome studies – including study design, controls, detection methods, and bioinformatics pipelines – to standardize methods and increase detection efficiencies.

Abstract

The Metagenomics Research Group (MGRG) is the latest addition to the ABRF research group family and focuses on evaluating, studying, and refining methodologies for analyzing all genomes in a complex population of microorganisms. This includes developing standardized methods, microbial controls and refined bioinformatics pipelines. Currently, two projects are underway. The first will create a cellular-based bacterial standard that can be used as a reference control, and the second is the Extreme Microbiome Project (XMP). The development of the bacterial standard builds on previous work by the Nucleic Acids Research Group in 2013 which assembled an ethanol-fixed intact microbial community reference standard with known numbers and types of bacteria used for DNA extraction efficiency studies. Although the MGRG is assembling a similar standard, it is unique because it is designed to meet very specific parameters. The standard will include six biosafety level I bacteria which possess Class I genomes with varying GC content, enumerated using two high resolution volumetric flow cytometry techniques. Future standards with Class II and Class III microbial genomes, and representatives from other kingdoms, will be developed at a later time. The XMP is conducting whole genome shotgun sequencing that focuses on extremophilic and unique environments. Samples will be analyzed from several sites such as the pink Lake Hillier in Western Australia, the "Door to Hell" crater in Turkmenistan, deep ocean brine lakes in the Gulf of Mexico, deep Arctic Ocean sites of western Greenland, permafrost tunnels in Alaska, and the International Space Station, among others. The goals of the XMP study are multifaceted: to develop refined techniques for the detection and characterization of novel microbes, evaluate DNA and RNA techniques optimized for extremophile samples, and evaluate bioinformatics pipelines, with an overall aim of discovering new genes, metabolic pathways, and life forms.

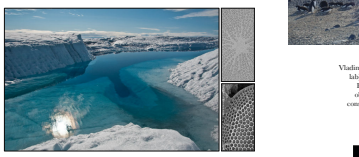
Activities

- Reference Standards:
  - DNA Standard
  - Whole Cell Bacterial Standards
- Multi-Lytic Enzyme Mix
- Preservatives for Microbiome Samples
- Extreme Microbiome Project, sample and assay:
  - Greenland
  - Antarctica
  - Door to Hell crater
  - Deep ocean brine lakes
  - International Space Station
  - Lake Hillier, Australia
  - Permafrost tunnel
  - Penguin microbiome
  - Blood Falls, Antarctica

Extreme Microbiome Project (XMP)

This new metagenomics project focuses on developing and evaluating methods for the recovery of DNA and RNA from unique sample types containing complex mixtures of microorganisms, and is creating bioinformatics tools for *de novo* assembly of deep sequencing data generated from these XMP samples.

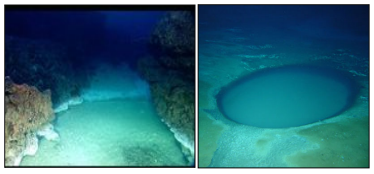
Locations and Samples



XMP collaborator Diana Krawczyk samples the unique waters and sediments along the West Greenland coast. Core samples represent specific centuries in the past and give insight into environmental changes. They also contain novel diatoms!

Cute penguins in Antarctica!

Vladimir Samarkin (Joye lab) collects Emperor Penguin feces after obtaining informed consent from a donor.



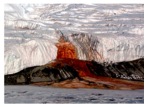
Mandy Joye uses the deep-sea submersible *Alvin* to visit hydrothermal vents and deep ocean brine lakes. On the seafloor of the Gulf of Mexico, salt tectonics has formed supersaturated lakes of water that are denser than the surrounding seawater. These underwater lakes have waves and shorelines just like those on land, but are rich in unusual microbial life.

Lake Hillier, on Middle Island in an archipelago near Western Australia, has a permanent pink hue and high salt content (38%). The color may be due to the micro-alga *Dunaliella salina* or halophilic Archaea such as *Halobacterium*.

Chris Mason and his lab are participating in the NASA Twins Study. Astronauts Scott and Mark Kelly are identical twins; Scott will collect gut and environmental samples during his yearlong mission aboard the International Space Station, while Mark provides earthbound comparison samples.

The Door to Hell crater is located in a natural gas field in central Turkmenistan. Its gas fire has been burning continuously since it was ignited by Soviet petroleum engineers in 1971. Pictured at right is explorer George Kourounis descending into the crater to collect samples.

The US Army ERDC Permafrost Tunnel near Fairbanks Alaska offers access to unique frozen environments over 40,000 years old. The permafrost contains bones of the Pleistocene bison, mammoth, and horse as well as a variety of plants. It's a real-life Time Tunnel!



Future sample collection expeditions to East Antarctica will visit the hypersaline Blood Falls (tinted by iron oxide) and lakes of the McMurdo Dry Valleys.

Methods

Several sample extraction techniques will be compared to recover both DNA and RNA for shotgun sequencing by long- and short-read technologies. RNA-Seq, DNA-Seq, and Methyl-Seq assays will be performed. Library synthesis techniques and reagents will be evaluated for suitability with high (and highly variable) GC content. Bioinformatics approaches are a strong interest of the XMP, including evaluation of currently available software and creating new assembly and analysis pipelines.

Results

- **Door to Hell gas crater**
  - DNA extracted: 10 g at 438 pg/ul in 20 ul
  - DNA library: Rubicon ThruPlex 20 cycles
  - Sequencing: Illumina MiSeq 2x250
  - Data analysis: MetaPHan and 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
```

- **Emperor Penguin fecal microbiome**
  - DNA extracted: 0.1 g at 36 ng/ul in 30 ul
  - MAC4L and ALO3 enzyme mixes, Omega extraction kit
  - DNA library: Rubicon ThruPlex 8 cycles
  - Sequencing: Illumina MiSeq 2x250

Taxa	Abundance (%)
Gillisia (unclassified)	76.9
Geobacillus kaustophilus	5.2
Clostridium perfringens	5.1
Marinobacter (unclassified)	4.8
Geobacillus (unclassified)	4.3
Thermus (unclassified)	1.7
Anoxybacillus flavithermus	1.5
Psychrobacter cryohalolentis	0.6

- **Lake Hillier**
  - Compared three sample preservatives
  - Extracted RNA with TRIzol LS
  - Extracted DNA with MAC4L and Omega kit
  - Compared two processing protocols

Processing Method	Sample - Preservative	Volume Ext (mL)	RNA Yield (ng in 25ul)	DNA Yield (ng in 25ul)
Filtered	sediment - fresh	0.5	ND	7.75
	sediment - ethanol	1.7	50.75	192.5
	sediment - DMSO	1.7	35	327.5
	mid water - fresh	7.5	27.5	23.3
	mid water - ethanol	7.5	ND	10
	mid water - DMSO	7.5	ND	105
Centrifuged	sediment - fresh	0.2	55	55
	sediment - ethanol	0.2	37.5	15
	sediment - DMSO	0.2	37.5	97.5
	bank - fresh	0.2		627.5
	bank - ethanol	0.2	950	520
	bank - DMSO	0.3		560

Reference Standards

We are developing two metagenomics standard samples. The bacterial cell standard contains six microbes as fixed cells with known quantities, and can be added to an experimental sample or matrix as a spike-in to evaluate recovery efficiencies. Selected in collaboration with NIST and ATCC, the species represent a range of characteristics:

Gram Positive	% GC Content	Growth Methods	Group	DNA Repeats	Total Repetitive	Genome Length
Staphylococcus epidermidis, ATCC 12228	32.8	Standard	Firmi	55	5110	2,564,615
Halobacillus halophilus, 35676 ATCC	46.8	MBA2216	Firmi	91	5260	4,170,008
Micrococcus luteus, NCTC 2665 ATCC 4698	72.0	Standard	Actino	65	4153	2,501,097
Gram Negative						
Escherichia coli str. K-12 substr. MG1655, ATCC 700926	50.8	Standard	Gamma	77	5463	4,639,675
Pseudomonas fluorescens F113, ATCC 13525	61.4	Standard	Gamma	35	5825	6,845,832
Pseudoalteromonas haloplanktis, ATCC 14393	40.1	MBA2216	Gamma	28	5821	3,850,272

Cells will be enumerated by microscopy and automated counters to precisely quantify the contents of the cellular reference standard.

The second standard sample will contain genomic DNA extracted from the cultured cells. The reference will contain known copy numbers of each genome as determined by digital PCR. An Oxford Nanopore MinION sequencing run was recently completed for a trial version of the DNA standard.

Multi-Lytic Enzyme Mix

In collaboration with Sigma Aldrich, we are developing the MAC4L Polyzyme mix for digestion of cell walls from the range of species present in metagenomic samples. MAC4L initially contains mutanolysin, achromopeptidase, chitinase, lysozyme, lysostaphin, lyticase, and labiase.

Microbiome Preservatives

DNA and RNA preservatives from Norgen Biosciences, Polysciences, and DNA Genotek are being evaluated using the bacterial reference standards, as are custom mixes containing DMSO or ethanol.

Acknowledgments

Special thanks to **Vladimir Samarkin** (Joye Lab-U. Georgia) and **Jill Mikucki** (U. Tennessee) for Antarctica and penguin sampling, and **John Lizmore** and **Don Cater** from the Australian government for arranging Lake Hillier sampling.

Many thanks to the ABRF executive board for supporting the formation of MGRG, and especially to EB liaison **Tim Hunter** for supporting our unique and extreme study plan.

Significant contributions of reagents are appreciated from our industry partners, Norgen Biosciences (**Nezar Rghei**), Rubicon, DNA Genotek (**Carlos Merino** and **Aaron Del Duca**) and Sigma Chemical (**Aaron Sin** and **Bob Gates**).

Many thanks to Illumina and **Craig Rowell** who is a major supporter of the XMP project and **Chris Streck** for arranging our Illumina partnership.

We greatly appreciate the lab work of our colleagues at the Mississippi State University and Weill Cornell sequencing centers.