Association of Biomolecular Resource Facilities

Metagenomics Research Group

Life at the Extreme: The ABRF Metagenomics Research Group

Implementing New Standards in Metagenomics and the Extreme Microbiome Project

Alexa McIntyre, Ali Kusuma, Anto Budiharjo , Caryn Evilia, Christopher Mason, Daniela Bezdan, Dev Mittar, Diana Krawczyk, Don Baldwin, Ebrahim Afshinnekoo, Elena Zaikova, Ellen Jorgensen, Felipe Gómez, George Yeh, Ian Charold Herriott, Jessica Hoffman, Jodie Lee, Jon Penterman, Jonathan Foox, Joshua Hyman, Ken McGrath, Lynn M. Schriml, Matthew Settles, Mike Farrell, Mohamed Donia, Nadim Ajami, Natàlia Vinas, Nathan J. Bivens, Neon Jung, Nick Greenfield, Nikos Kyrpides, Noah Alexander, Rachid Ounit, Rita Colwell, Robyn Barbato, Russ Carmical, Samantha Joye, Sarah Johnson, Shawn Levy, Sofia Ahsanuddin, Stefan Green, Tara Rock, Timothy Hunter, W. Kelley Thomas, William Hendrickson, Scott Tighe

Xtreme Microbiome Project

The Mission

The goals 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) 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 improved bioinformatics pipelines. Several MGRG projects are now complete.

Cellular and DNA bacterial standards have been produced which include 10 biosafety level I bacteria with Class I genomes (minimal repetitive DNA) and a range of GC content. Stocks of preserved cells have been enumerated for precise cell counts, digital PCR was used to measure genomic copy numbers, pooled genomic DNA has been sequenced, and the standards have been submitted to ATCC for distribution. The bacteria are also being fabricated into whole cell reference standards which will be developed by 2019.

The multi-lytic Polyzyme enzyme is now complete and distributed through Millipore Sigma as Metapolyzyme for cell wall digestion and increased cellular lysis.

A modular DNA extraction kit has been developed with Omega Biotek and tested in Antarctica by Sarah Johnson to extract exotic soil systems of ancient microbial biofilms.

All these innovations are being test in the eXtreme Microbiome Project (XMP, www.extrememicrobiome.org) which uses shotgun metagenomic sequencing for characterizing extremophilic and unique environments from around the world. Data collection for XMP includes DNAseq, RNAseq, Culturing, Shotgun, 16s, ITS, 18s, and searching for biosynthetic gene clusters.

Activities

- Reference Standards:
  - DNA Standard
  - Whole Cell
  - Synthetic G-Block Standards (Don Baldwin and Rachid Ounit)-Pending
- Multi-Lytic Enzyme Mix (MetaPolyZyme)
- Modular DNA extraction kits for high molecular weight DNA (Omega BioTek)
- 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 and hummingbird
  - Blood Falls, Antarctica
  - New York subway
  - Blue Lagoon Iceland
  - Ethiopian Toxic Hot springs
  - Rio Tinto

Corporate Partners

- Illumina
- Bioo Scientific
- Omega Bio-tek
- One Codex
- Logos Biosystems
- New England BioLabs
- Promega
- ATCC
- MilliporeSigma

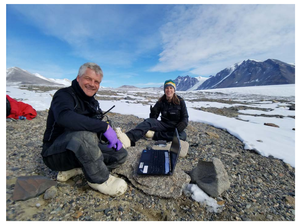
eXtreme Microbiome Project (XMP)

This 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.

Extreme Environments

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*.

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.



Scott Tighe and Dr.Sarah Johnson (both MGRG/XMP members) Test the Oxford Nanopore for remote field sequencing in the Victoria Valley of Antarctica

Fecal Microbiomes

Comparative microbiome studies of low fat vs high fat storage

Hummingbird
(Costa Rica)
Samples to be collected by
Ian Herriott from the Univ of Alaska Fairbanks July 2015 using NAF apparatus

Emperor Penguin Samples collected by Vladimir Samarkin in Antarctica.

(Samantha Joyes lab)

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. Useful tools include:

BLAST blast.ncbi.nlm.nih.gov/Blast.cgi
MetaPhlAn bitbucket.org/biobakery/metaphlan2
Kraken ccb.jhu.edu/software/kraken/
PhyloSift phylosift.wordpress.com/
GOTTCHA github.com/poeli/GOTTCHA

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: MetaPhlAn and MegaBlast

- Culturing: 40 mg plated on TSA at 28 ºC for 50 days

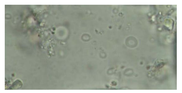
sediment culture subcultured colonies

Colony A is 99% *Arthrobacter tumbae*, a bacterial species isolated from deep sea sediments of the Bay of Bengal and Andaman Sea.

Adélie penguin fecal microbiome

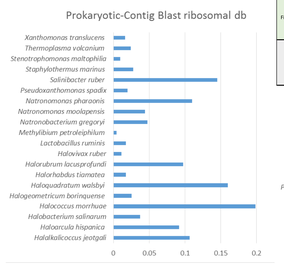
- DNA extracted: 0.1 g at 36 ng/ul in 30 ul
- MAC4L and ALO3 enzyme mixes, Omega extraction kit
- DNA library: Rubicon ThruPlex 8 cycles
- Sequencing: Illumina MiSeq 2x250

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

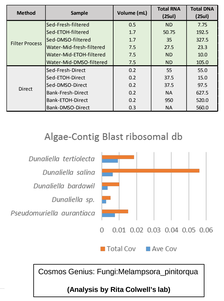


Lake Hillier

- Tested three collection preservatives (EtOH, DMSO, Fresh (cold))
- Extracted RNA (Trizol LS) and DNA (MAC4L-Omega)
- Tested two processing protocols: diluted and filtered; diluted and centrifuged



Method	Sample	Volume (mL)	Total RNA (Dbs)	Total DNA (Dbs)
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



Cosmos Genius: Fungi:Melampsora_pinitorqua

(Analysis by Rita Colwell's lab)

Microbial Reference Standards

Three types of Microbial Reference Standards (even or staggered genomic DNA) have been completed and will be distributed through our corporate partner, the ATCC.

Organism	ATCC® No.	Gram	ul for 100,000,000 genomes Qubit	ul for 100,000,000 genomes based on qPCR-16s (ave of Dupl)	Combine the following volume for Class 1+ based on Qubit (EVEN)	Copies added to mix Qubit	copies/ul Qubit	Copies added to mix qPCR	copies/ul qPCR
Staphylococcus epidermidis PCI 1200	12228™	+	3.50	1.67	1400	4.00E+10	2.14E+06	8.40E+10	4.49E+06
Chromobacter violaceum NCTC 9757	12472™	-	2.75	5.98	1099	4.00E+10	2.14E+06	1.84E+10	9.82E+05
Micrococcus luteus NCTC 2665	4698™	+	3.17	1.05	1266	4.00E+10	2.14E+06	1.21E+11	6.47E+06
Pseudoalteromonas haloplanktis TAC125	35231™	-	4.04	4.27	1616	4.00E+10	2.14E+06	3.78E+10	2.02E+06
Haloferax volcanii DS2	29605™	+	4.11	31.61	1645	4.00E+10	2.14E+06	5.20E+09	2.78E+05
Bacillus subtilis subsp. Subtilis str. 168	23857™	+	5.81	7.95	2326	4.00E+10	2.14E+06	2.93E+10	1.56E+06
Halobacillus halophilus DSM 2266	35676™	+	5.55	3.22	2221	4.00E+10	2.14E+06	6.89E+10	3.68E+06
Escherichia coli K-12 substr. MG1655	700926™	-	5.58	7.67	2233	4.00E+10	2.14E+06	2.91E+10	1.56E+06
Enterococcus faecalis OG1RF	47077™	+	3.57	2.61	1427	4.00E+10	2.14E+06	5.46E+10	2.92E+06
Pseudomonas fluorescens F113	13525™	-	8.68	10.86	3471	4.00E+10	2.14E+06	3.19E+10	1.71E+06

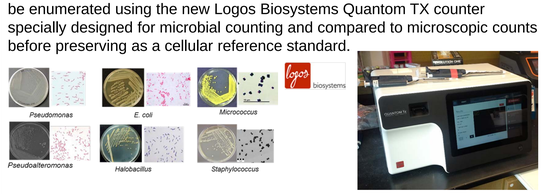
10 Species available as even or staggered copy number (ATCC® MSA-3001, ATCC® MSA-3002)

Organism	ATCC® No.	Gram	ul for 100,000,000 genomes Qubit	ul for 100,000,000 genomes based on qPCR-16s (ave of Dupl)	Combine the following volume for Class 1+ based on Qubit (EVEN)	Copies added to mix Qubit	copies/ul Qubit	Copies added to mix qPCR	copies/ul qPCR
Staphylococcus epidermidis PCI 1200	12228™	+	3.50	1.67	700	2.00E+10	3.60E+06	4.20E+10	7.56E+06
Chromobacter violaceum NCTC 9757	12472™	-	2.75	5.98	549	2.00E+10	3.60E+06	9.19E+09	1.65E+06
Micrococcus luteus NCTC 2665	4698™	+	3.17	1.05	633	2.00E+10	3.60E+06	6.05E+10	1.09E+07
Haloferax volcanii DS2	29605™	+	4.11	31.61	822	2.00E+10	3.60E+06	2.60E+09	4.68E+05
Escherichia coli K-12 substr. MG1655	700926™	-	5.58	7.67	1117	2.00E+10	3.60E+06	1.46E+10	2.62E+06
Pseudomonas fluorescens F113	13525™	-	8.68	10.86	1735	2.00E+10	3.60E+06	1.60E+10	2.87E+06

6 Species available as even copy number (ATCC® MSA-3000)

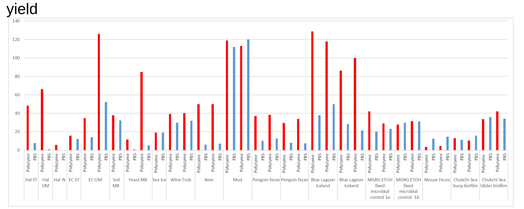
Whole Cell Microbial Standards

Microbes from above will also be fabricated as a preserved whole cell standard which can be used for DNA extraction efficiency and related studies. Samples be enumerated using the new Logos Biosystems Quantom TX counter specially designed for microbial counting and compared to microscopic counts before preserving as a cellular reference standard.



Polyzyme Enzyme Mix (Metapolyzyme)

In collaboration with Millipore Sigma, we have developed 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, but has since been modified to be a proprietary combination of enzymes. Extensive test has show this to be a great pretreatment to increase DNA yield



DNA extraction results for samples with and without Polyzyme treatment preformed by the MGRG special DNA extraction team

Acknowledgments

Many thanks to the ABRF executive board for supporting the MGRG and XMP. Special Thanks to our partners and dear friends that have supported these endeavors including One Codex, CosmosID, Logos Biosystems, BioO, Illumina, NEBL, Omega Bio-Tek, Millipore Sigma, Promega, and the ATCC for helping with experiment designs and supplying the necessary reagents and equipment needed for the MGRG and XMP projects. A special thanks to George Kourounis for his work on the Doors to Hell crater, Lynn Schriml, Folker Meyer, and Nikos Kyrpides of the Genomic Standard Consortium for help launching part of the XMP projects as well as contributions to manuscripts. A special thanks to John Lizmore and Don Cater from the Australian government for arranging Lake Hillier sampling.