

Diversity and Abundance of Arsenic Resistant Genes in Soil Microbiomes

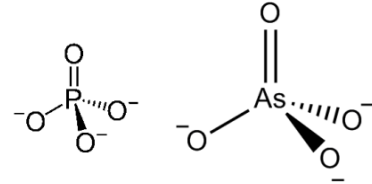
Susanna Yeh
iCER ACRES Presentation, 7/27/17

Arsenic is the most Ubiquitous Toxic Element

Arsenic (As) terrorizes the metabolism process

Arsenate (As(V)) is similar to phosphate and can substitute in pathways

Exposure is related to cancer, cardiovascular disease, diabetes, neuropathy



As is present in groundwater and soil

Bacteria influence the As cycle in the environment

As toxicity and bioavailability depends on oxidation and methylation states

Arsenate (As(V)): less mobile

Arsenite (As(III)): more mobile and toxic

Methylated species: volatile

Not Much is Known about AsRG Diversity

Arsenic resistant genes (AsRG) are well-studied in cultivated bacteria
There is an increasing number of soil metagenomes available; it is possible to identify phylogenetic distribution and diversity
Better understanding of this diversity leads to improvement in As bioremediation process and risk assessment

Objectives:

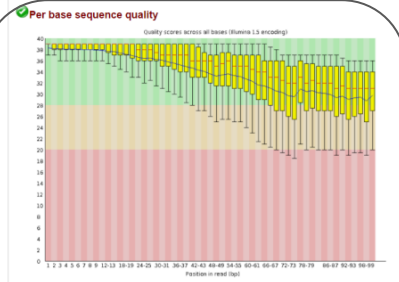
- Compare AsRG in non-contaminated soil microbiomes
 - 12 sites, 2-3 samples per site, 5 countries
- Examine phylogeny of AsRG
- Expand the known diversity of AsRG
- Analyze abundance of AsRG in diverse soils

Workflow

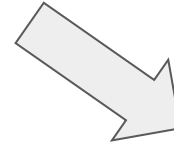
MG-RAST

metagenomics analysis server

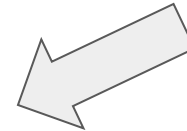
Download metagenomes
from MG-RAST



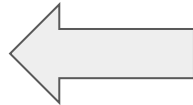
Quality filter data



Xander: gene-targeted
assembler



Post-assembly
assessment



- Phylogenetic analysis
- Alpha diversity
- Gene Abundance

Diverse Datasets were Chosen

Sample ID	Coordinates	Sampling Date	Average Genome Size	Giga Base Pairs	MG-RAST Sample ID
Iowa_corn22.3	41, -92	2009-06-28	5,248,659	8.19	4539522.3
Iowa_corn23.3			5,208,064	8.12	4539523.3
Iowa_agricultural00.3	42, -93	2012-07-19	6,203,490	24.98	4509400.3
Mangrove02.3	4, 100	2014-10-02	5,979,725	24.38	4603402.3
Mangrove70.3			6,005,221	24.54	4603270.3
Permafrost_Russia12.3	68, 161	2007-08-05	5,135,678	20.05	4546812.3
Permafrost_Russia13.3			5,131,959	19.2	4546813.3
Iowa_prairie75.3	42, -93	2009-06-28	6,994,965	18.79	4539575.3
Iowa_prairie72.3			7,169,200	17.58	4539572.3
Iowa_prairie76.3			7,088,957	17.43	4539576.3
Brazilian_forest95.3	10, 63	2010-01-01	5,787,131	13.27	4546395.3
Brazilian_forest39.3			8,789,768	9.04	4536139.3
Brazilian_forest54.3			8,371,924	9.69	4535554.3
Minnesota_creek46.3	45, -93	2013-06-05	6,504,862	10.65	4541646.3
Minnesota_creek45.3			6,057,590	9.77	4541645.3
Disney_preserve18.3	28, -81	2013-11-14	6,888,257	11.2	4664918.3
Disney_preserve25.3			7,230,169	4.14	4664925.3
California_grassland15.3	34, -118	2010-12-17	7,414,208	6.5	4511115.3
California_grassland62.3			12,981,156	5.77	4511062.3
Illinois_soil91.3	42, -88	2010-11-19	7,441,858	7.95	4653791.3
Illinois_soil88.3			6,954,025	7.14	4653788.3
Wyoming_soil22.3	48 -111	2013-07-16	5,354,164	0.82	4670122.3
Wyoming_soil20.3			5,350,858	6.41	4670120.3
Permafrost_Canada23.3	79, 91	2011-05-01	3,932,079	6.52	4523023.3
Permafrost_Canada45.3			4,719,216	5.52	4523145.3

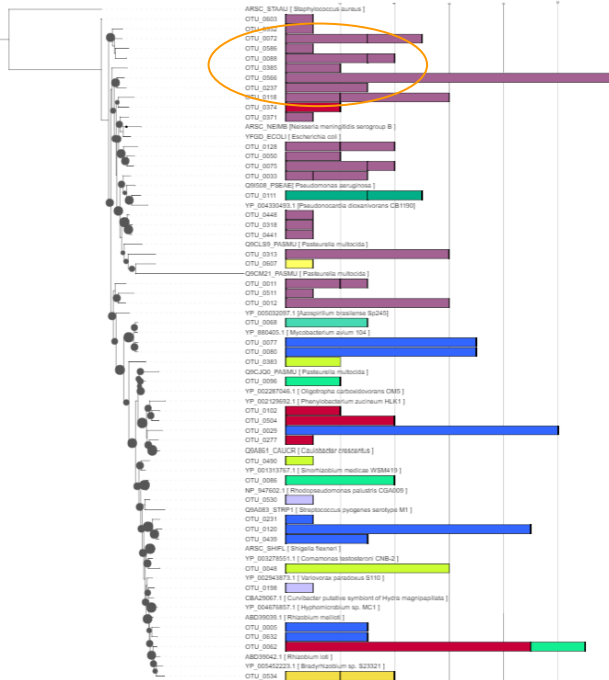
Full-length, High-Quality Sequences Assembled

Gene	Protein Function	Xander Max Length Assembled (aa)	Xander Model Length (aa)	Xander Max Percent Identity
<i>aioA</i>	Arsenite oxidase	815	837	97.619
<i>arxA</i>	Arsenite oxidase	508	831	95.652
<i>arsM</i>	Arsenite methyltransferase	217	269	96.154
<i>acr3</i>	Arsenite efflux pump	351	363	100
<i>arsB</i>	Arsenite efflux pump	266	430	100
<i>arsC glut</i>	Cytoplasmic arsenate reductase	116	117	100
<i>arsC thio</i>	Cytoplasmic arsenate reductase	83	133	100
<i>arrA</i>	Dissimilatory arsenate reductase	842	846	96.629
<i>arsD</i>	Arsenite metallochaperone	124	123	100
<i>rplB</i>	50S ribosomal protein L2	279	277	98.909

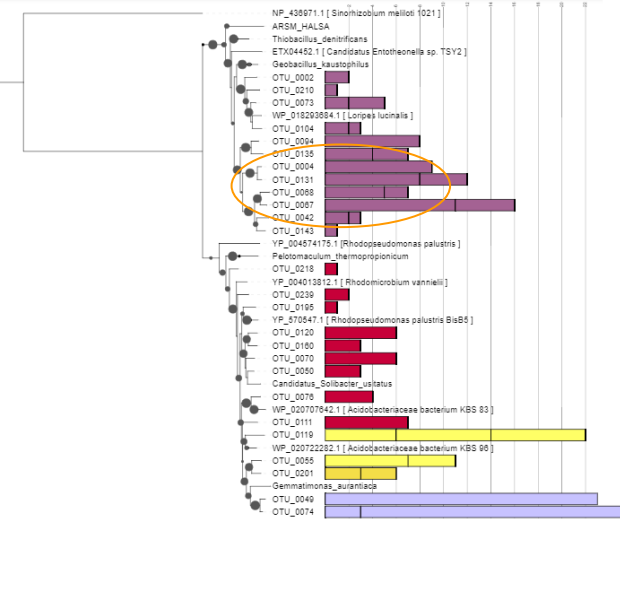
- *aioA*, *acr3*, *arsC glut*, *arrA*, *arsD*, *rplB* have sequences $\geq 90\%$ of the HMM model length
- Each gene has a high maximum percent identity, giving more confidence to the model
- Xander can assemble near full-length sequences for each gene of interest

Potential Novel Diversity of AsRG

ArsC glut (arsenate reductase)



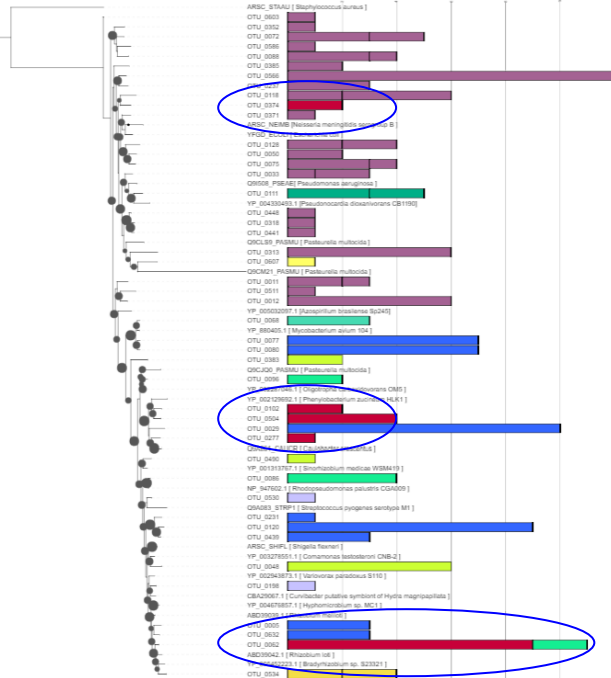
ArsM (arsenite methyltransferase)



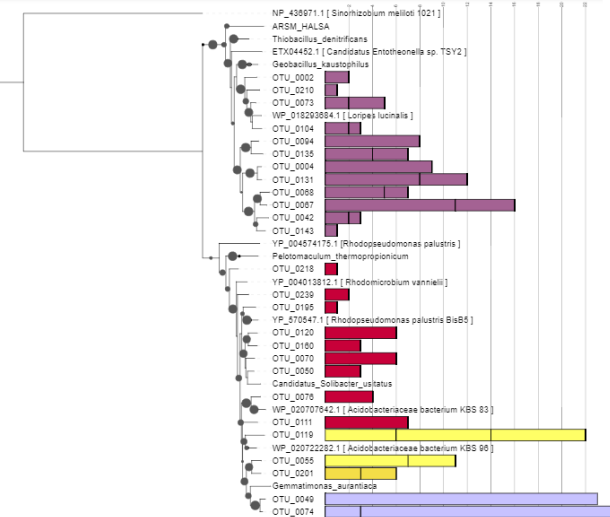
- Sequences $\geq 90\%$ of model length
- OTUs are assembled from metagenomes, clustered at 90% ID
- Accession numbers are from seed sequences and blast nr database
- Phylogeny is diverse within and across microbiomes

Phylogeny is diverse

ArsC glut (arsenate reductase)

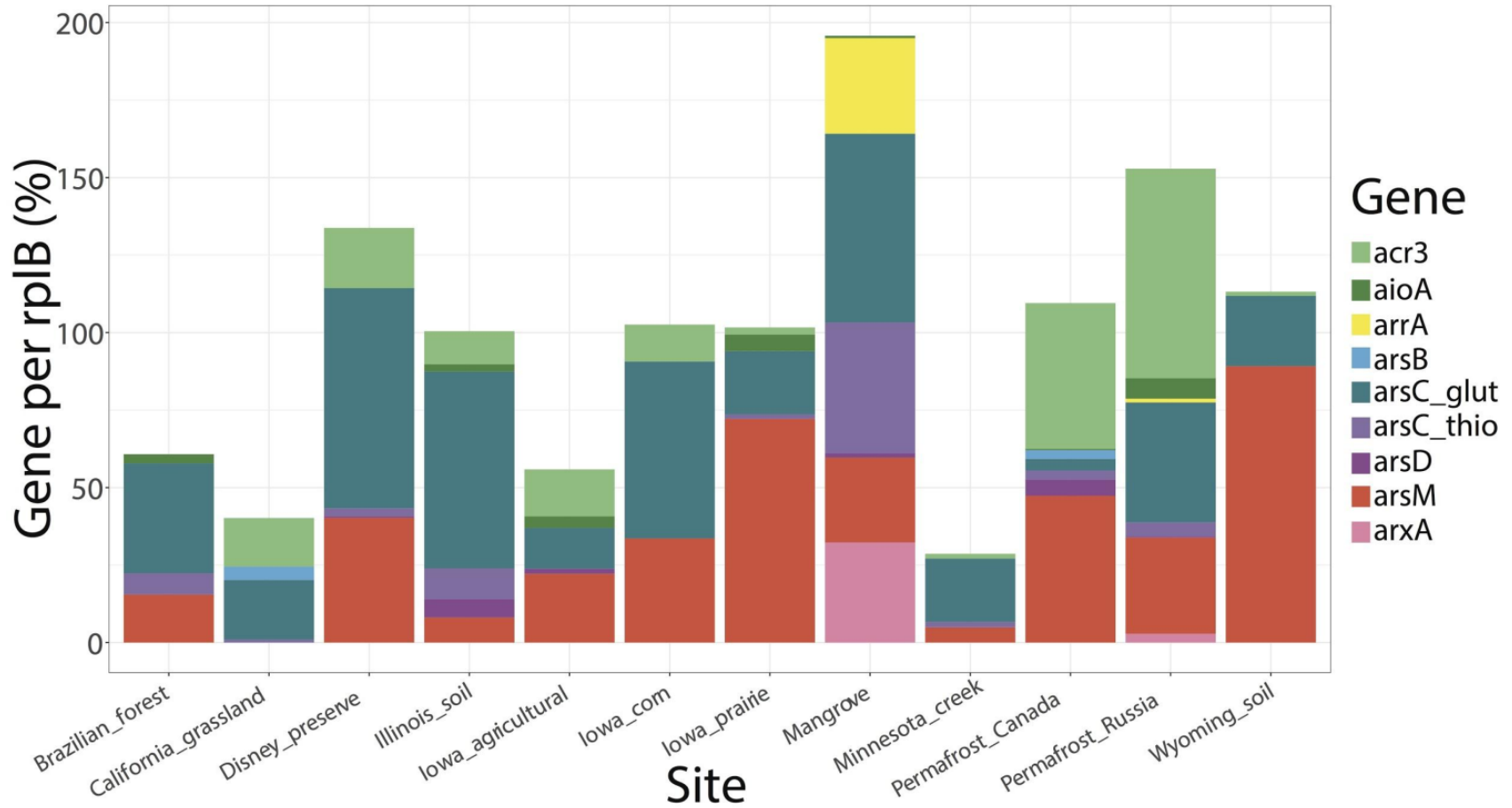


ArsM (arsenite methyltransferase)



- Sequences $\geq 90\%$ of model length
- OTUs are assembled from metagenomes, clustered at 90% ID
- Accession numbers are from seed sequences and blast nr database
- Phylogeny is diverse within and across microbiomes

Normalized AsRG Abundance Varies



Conclusions

Abundance and phylogeny of AsRG are diverse within and across microbiomes

Diverse microbiomes have the genetic potential to transform arsenic

Xander can assemble near full-length, high-quality contigs for AsRG

Future Direction

Include contaminated site for comparison

Acknowledgements

Shade Lab

Dr. Ashley Shade

Taylor Dunivin

Institute for Cyber-Enabled Research

Camille Archer

Brian O'Shea

HPCC

NSF

BBB