Ecology of Uncultivated Freshwater Actinobacteria Revealed via Reverse Ecology

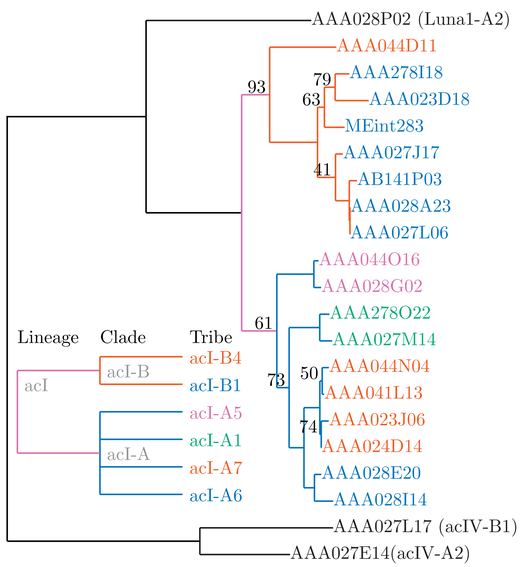
Joshua J. Hamilton[1], Brittany S. Brown[1], Matthew Bendall[2], Tijana Glavina del Rio[2], Stefan Bertilsson[3], Rex Malmstrom[2], Ramunas Stepanauskas[4], Susannah Tringe[2], Tanja Woyke[2], and Katherine D. McMahon[1]

[1]University of Wisconsin-Madison, [2]Department of Energy Joint Genome Institute, [3]Uppsala University, and [4]Bigelow Laboratory for Ocean Science

Introduction

Uncultured microbes are critical players in all ecosystems, where they support essential ecosystem functions such as nutrient cycling. While metagenomic approaches are increasingly being used to study the genetics and physiology of these organisms, such approaches would benefit from techniques which link specific biochemical transformations to individual community members. We have developed a computational pipeline which combines metabolic reconstructions of Actinobacterial genomes with a computational "reverse ecology" approach to identify biochemical requirements (e.g., carbon and nitrogen sources) and potential niches for members of the freshwater Actinobacteria.

Phylogeny and Metadata



We have sequenced and assembled 60 Actinobacterial single-cell genomes (SAGs) and genomes from metagenomes (GFMs). SAGs were classified using 16S rRNA sequences and a controlled taxonomic nomenclature. Phylogeny of all samples was determined using a concatenated alignment of 37 marker genes. Only genomes with classification to the 'tribe' level are shown. Genome completeness was estimated using the presence/absence of 37 marker genes.

Tribe	Num. of Genomes	GC Content	Avg. GC Content	Est. Completeness	Avg. Est. Completeness	Genome Size (MB)	Avg. Genome Size (MB)	Est. Size (MB)	Avg. Est. Size (MB)
Luna1-A2	1	0.49	0.49	0.84	0.84	0.99	0.99	1.18	1.18
acI-B4	1	0.44	0.44	0.95	0.95	1.15	1.15	1.22	1.22
acI-B1	7	0.4 to 0.42	0.41 + 0.01	0.14 to 0.97	0.57 + 0.31	0.66 to 1.16	0.87 + 0.17	1.03 to 4.89	2.06 + 1.35
acI-A5	2	0.46 to 0.46	0.46 + 0	0.95 to 0.97	0.96 + 0.02	1.35 to 1.41	1.38 + 0.05	1.42 to 1.45	1.44 + 0.02
acI-A1	2	0.47 to 0.48	0.48 + 0.01	0.05 to 0.97	0.51 + 0.65	0.82 to 1.14	0.98 + 0.22	1.17 to 15.21	8.19 + 9.93
acI-A7	4	0.44 to 0.46	0.45 + 0.01	0.11 to 0.76	0.36 + 0.3	0.7 to 1.38	1.04 + 0.35	1.03 to 8.52	4.74 + 3.37
acI-A6	2	0.45 to 0.46	0.46 + 0.01	0.14 to 0.84	0.49 + 0.5	0.78 to 0.88	0.83 + 0.07	1.06 to 5.8	3.43 + 3.35
Iluma-B1	1	0.50	0.50	0.78	0.78	1.03	1.03	1.31	1.31
Iluma-A2	1	0.53	0.53	0.59	0.59	0.77	0.77	1.29	1.29

Network Reconstruction

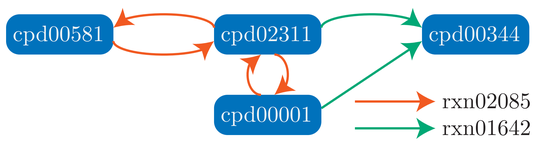
>AAA027L06_contig00001
TTTGAATTAAGTGATGATCAGATTGACTGGACAGTTG
ATGCCTATACCAGGGGCGTAATTGCAGATGAACAAATG
TCAGCACTTTTGATGGCTATTTTGTTAAATGGAATG ...

Step 1. Annotate microbial contigs using the RAST toolkit. Build metabolic network reconstruction using the Model SEED framework.

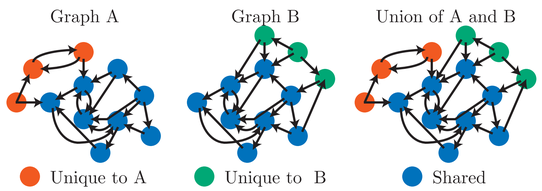
Reaction	Genes	Equation
rxn02085	CDS.1611	cpd02311 <-> cpd00001 + cpd00581
rxn01642	CDS.1612	cpd00001 + cpd02311 -> cpd00344
...		

Compound	Name	Formula	Charge
cpd00001	H_2O	H_2O	0
cpd00344	N-Formimino-L-glutamate	$C_6H_9N_2O_4$	-1
cpd00581	Urocanate	$C_6H_5N_2O_2$	-1
cpd02311	4-Imidazolone-5-propanoate	$C_6H_7N_2O_3$	-1
...			

Step 2. Convert reconstruction to a metabolic network graph. Metabolites are represented as nodes, and reactions as directed edges.

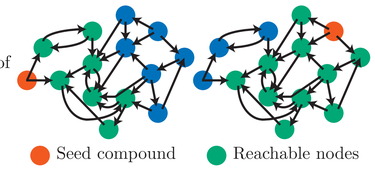


Step 3. Create network graphs for each tribe by joining graphs for all genomes from that tribe.



Resource Utilization

The **seed set** of a metabolic network is the minimal set of compounds from which all other compounds in the network can be synthesized. For a metabolic network represented as a directed graph, the seed set is the minimal set of nodes from which all other nodes can be reached. These nodes approximate a tribe's essential metabolites.



We classified 49 genomes as belonging to one of three major lineages: acI, acIV, and acV. Patterns of resource utilization reveal distinct niches for each of these lineages.

	Seed Compound	acI	acIV	acV
Branched-chain amino acids	Leucine	0.97	0.67	1.00
	Iso-Leucine	0.97	0.67	1.00
	Valine	0.97	0.67	1.00
Diamines	Ala-Leu	1.00	1.00	0.88
	Gly-Pro	1.00	1.00	0.88
Polyamines	Agmatine	0.56	0.50	0.88
	Putrescine	0.69	0.78	0.38
	Spermidine	0.69	0.78	0.38
Polysaccharides	Stachyose	0.28	1.00	0.13
	Manninotriose	0.28	1.00	0.13
Nucleotides	5-Methylcytosine	0.16	0.44	0.88
	OMP	0.97	0.78	0.75
	XMP	0.06	0.00	0.88
Sulfur compounds	Sulfide	0.91	0.78	0.50
	Thiosulfate	0.75	0.89	0.38
	Total Genomes	32	9	8

Fraction of Genomes: 1, 0

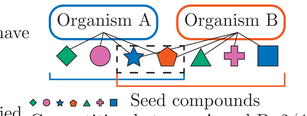
The acI lineage comprises three clades, each with a distinct substrate utilization pattern.

	Seed Compound	acI-A	acI-B	acI-C
Amino acids	Ornithine	0.50	0.78	0.00
	5-Aminolevulinate	0.83	0.72	0.50
Nucleotides	Deoxycytidine	0.25	0.78	0.00
	Cytidine	0.25	0.78	0.00
Polysaccharides	Maltose	0.33	0.17	1.00
	Stachyose	0.42	0.11	1.00
	Total genomes	12	18	2

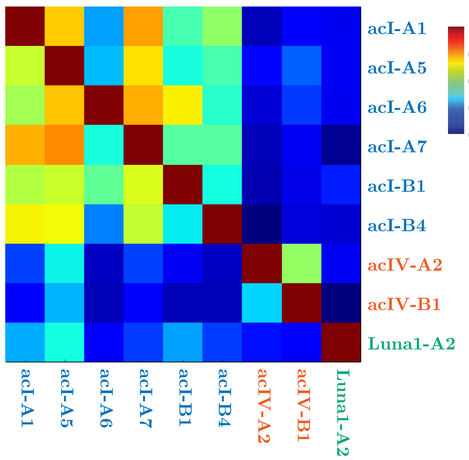
Fraction of Genomes: 1, 0

Metabolic Competition

Metabolic competition can arise when two organisms have similar nutritional profiles, and the potential for competition can be quantified using the metabolic competition index. For two organisms A and B, the **metabolic competition index** is calculated as the fraction of compounds in A's seed set that are also present in B's seed set.



Competition is strongest within—rather than between—lineages.



Unique seed compounds suggest niche specialization based on carbohydrate (acI lineage), amino acid (acI and acIV lineages) and polyamine (acIV) acquisition.

Tribe	Unique Seed Compounds
acI-A1	arabinose, citrate, glucose
acI-A5	chitobiose
acI-A6	altronate, prephenate
acI-B1	sarcosine
acI-B4	glucarate
acIV-A2	arginine, putrescine, spermidine
acIV-B1	agmatine, prephenate

Conclusions

Reverse ecology represents a powerful approach for inferring the ecological traits of an organism directly from its metabolic network. The predicted niche for acI tribes agrees with our experimental understanding, and predicted niches for acIV tribes have begun to emerge. These predictions will be enhanced through the sequencing of under-sampled tribes and the calculation of additional metrics, such as potential for cooperation. Finally, reverse ecology predictions are sensitive to genome completeness, and we are developing ANI- and coverage-based approaches to associate unclassified GFMs with specific tribes.

Acknowledgements

Applied Microbial Ecology C.N.P. McMahon Lab · NSF · LTER · JGI Joint Genome Institute · Bigelow Laboratory for Ocean Sciences · Uppsala Universitet