

Prelim Figures:

Example classification for a set of GFMs

GFM Name	domain	phylum	class	order	family	genus	species
MEint.metabat.1091	Bacteria	Actinobacteria	Actinobacteria	Acidimicrobiales	Acidimicrobiaceae	Ilumatobacter	
MEint.metabat.2538	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Candidatus Aquiluna	
MEint.metabat.112	Bacteria	Cyanobacteria					
MEint.metabat.7672	Bacteria	Planctomycetes	Planctomycetia	Planctomycetales	Planctomycetaceae	Rhodopirellula	
MEint.metabat.3080	Bacteria	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	Methylotenera	Methylotenera Versatilis
TBepi.metabat.3838	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Polynucleobacter	Polynucleobacter Necessarius

Table 1: Each GFM was classified by Phylosift (Darling et al., 2014) and custom script. Shown above is a subset of GFMs from Lake Mendota. The cyanobacterial GFM (MEint.metabat.112) is an example of one that does not follow Linnaean classification, and thus does not get labeled past phylum. The others stop getting classified when the next level does not match above the matching threshold specified.

Number of GFMs classified

	Mendota	TroutBog_epi	TroutBog_hypo
# of genome bins	102	36	64
classified at phylum level	98	36	60
phyla represented	9	6	8
classified at class level	78	35	56
classes represented	14	11	15

Table 2: The first row in this table gives the number of genomes binned, after filtering based on completeness and uniqueness of single copy genes. The second and forth rows show the number of GFMs classified to the phylum and class level, respectively. The third and fifth rows give the number of phyla and classes represented in the GFMs by each lake/layer.

GFM phyla distribution between lakes/layers

Phylum	Mendota	Trout Bog Epi	Trout Bog Hypo
ACIDOBACTERIA	0	2	3
ACTINOBACTERIA	17	9	9
BACTEROIDETES	33	3	9
CHLAMYDIAE	1	0	0
CHLOROBII	0	2	2
CHLOROFLEXI	1	0	0
CYANOBACTERIA	11	0	0
ELUSIMICROBIA	0	0	1
IGNAVIBACTERIA	0	0	2
PLANCTOMYCETES	13	0	0
PROTEOBACTERIA	12	17	26
ALPHAPROTEOBACTERIA	2	3	4
BETAPROTEOBACTERIA	7	9	11
DELTAPROTEOBACTERIA	1	1	4
EPSILONPROTEOBACTERIA	0	0	1
GAMMAPROTEOBACTERIA	2	4	4
TENERICUTES	2	0	0
VERRUCOMICROBIA	8	3	8

Table 3: The table above contains the distributions by phyla for the GFMs from each lake/layer. There are several phyla with only a few in one lake/layer and none in the others. However, cyanobacteria and planctomycetes have many GFMs represented in Lake Mendota and none in Trout Bog.

14 acI SAG, 400 gene tree

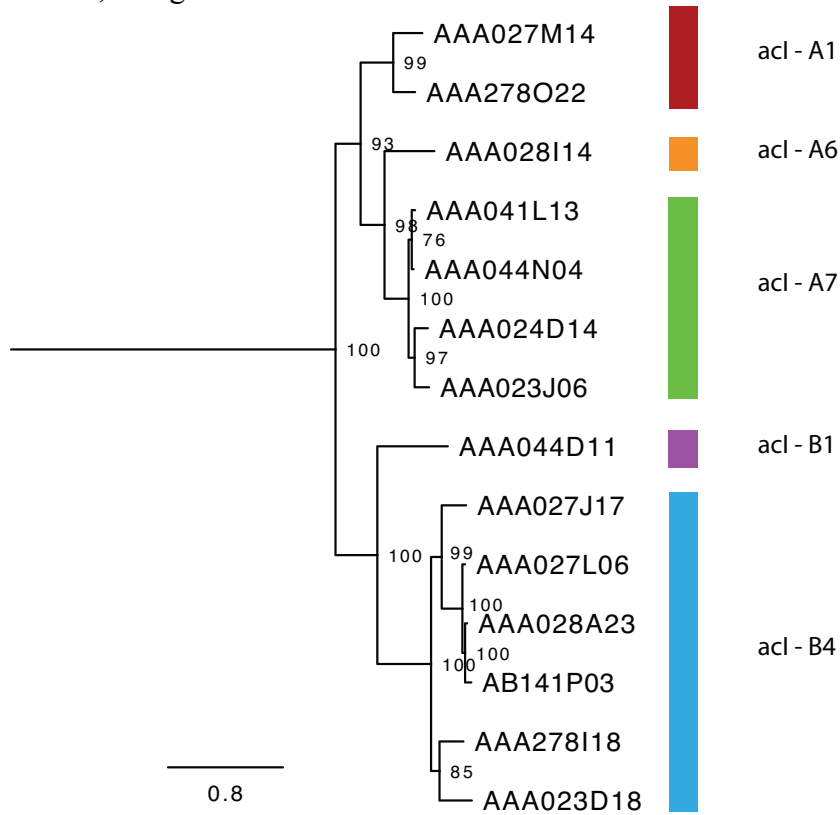


Figure 1: Using the alignment of 400 marker genes from PhyloPhlan (Segata, Börnigen, Morgan, & Huttenhower, 2013), a maximum likelihood tree was made with RAxML (Stamatakis, 2014). For this tree, 100 bootstrap replicates were run. The branching order for this tree, matches that of the 16S tree.

Heatmap showing relative abundance and ANI of acI SAGs in Lake Mendota

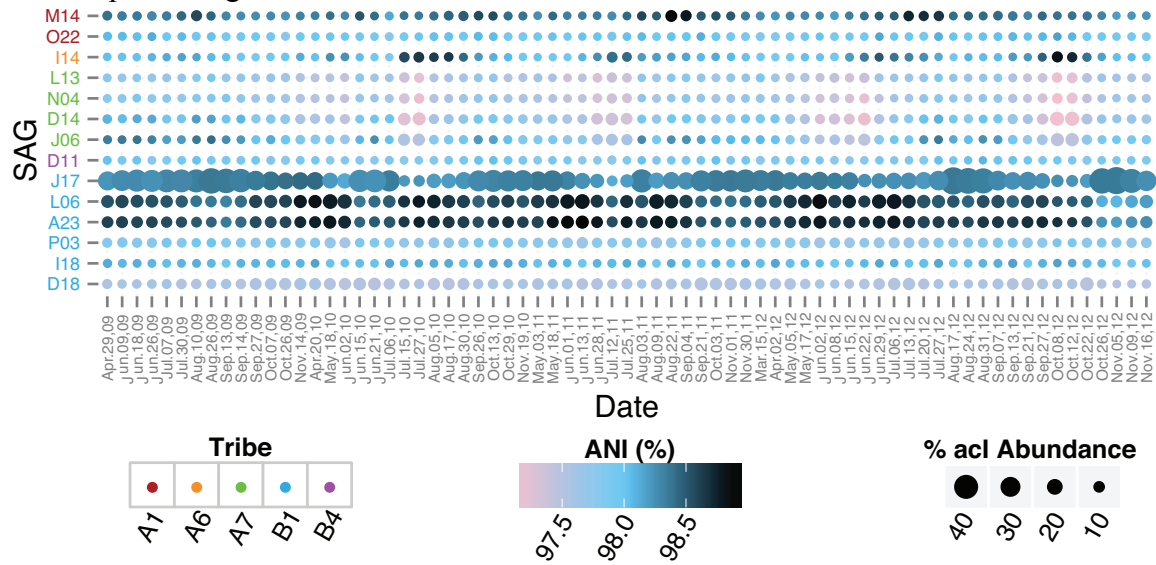


Figure 2: Relative abundance and ANI for each SAG in each metagenome. Reads were mapped using BLAST. Only hits longer than 200 bp and higher than 95% ID were kept.

AAI vs 16S identity

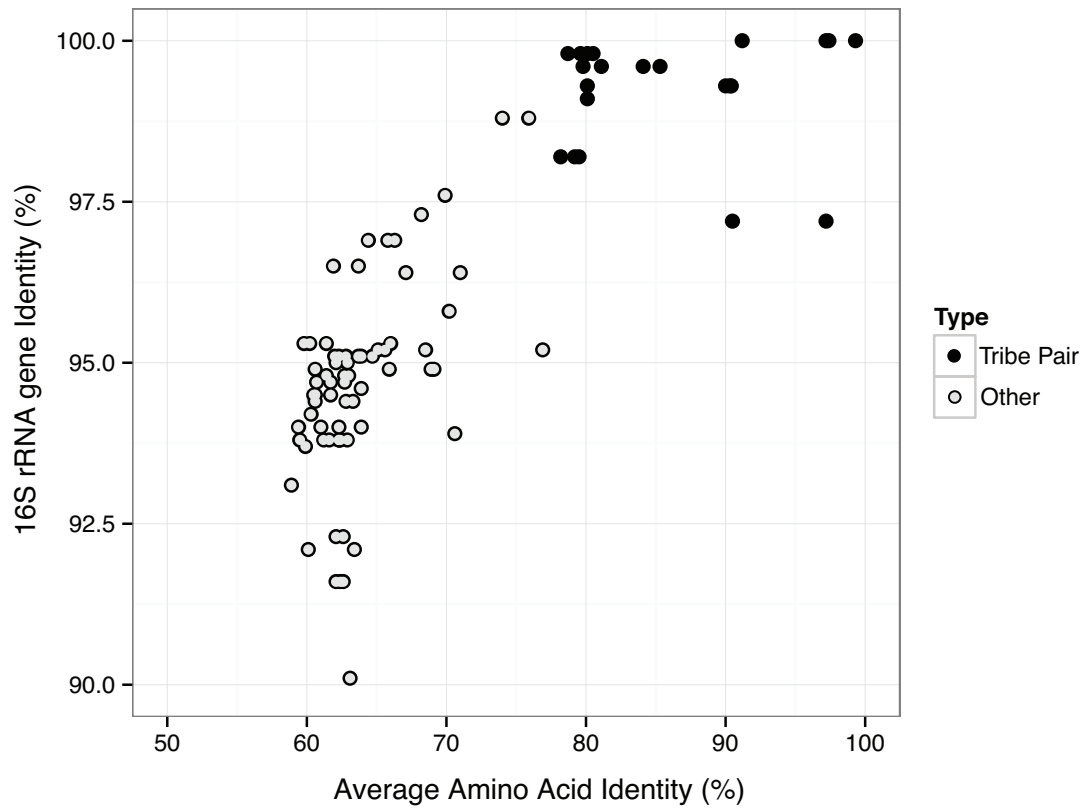


Figure 3: For every pair of acI SAGs, the AAI is plotted verses the 16S rRNA gene identity. Black indicates that both SAGs were from the same tribe. The 16S identity is not always a good predictor of AAI.

Experimental Plan Figures:

Mock Plot of Mean vs. CV

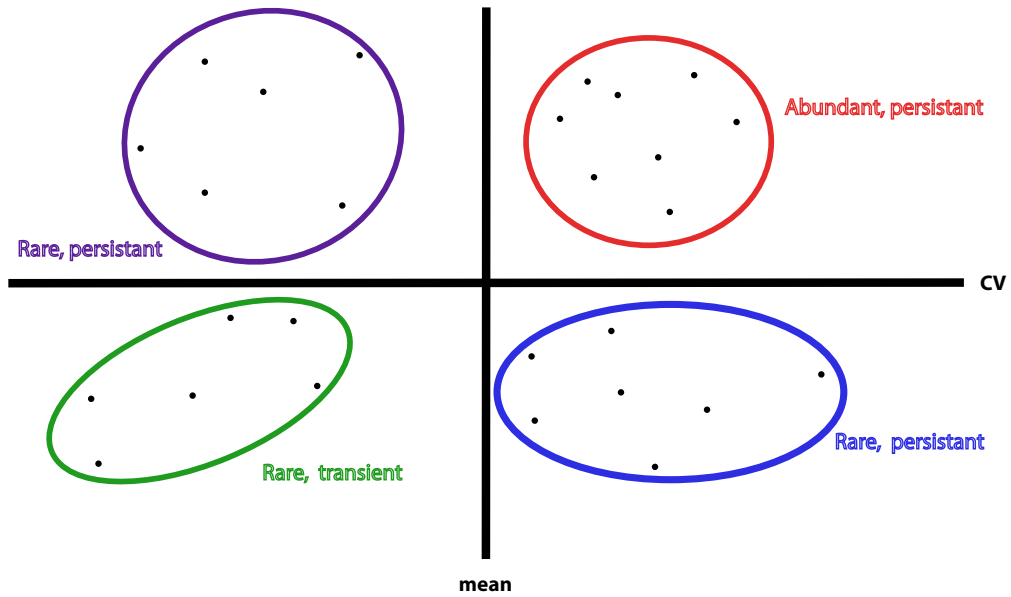


Figure 4: With mean relative abundance for each genome plotted against coefficient of variation of relative abundance, we should see a trend for which samples are abundant or rare and persistent or transient.