

Story Overview

What we know:

- Piccard (PD) and Von Damm (VD) are very different in term of location/surrounding geology. Thus, they likely differ in geochemistry as well.
- There are differences in relative abundances of bacteria and archaea between PD and VD

What we suspect:

- PD and VD have similar community-level abundance of metabolic genes based on previous bubble plots with fewer genes and samples (Reveillaud et al., 2016)

What we don't know:

- Does the expression pattern of MAGs mirror their relative abundance? Are MAGs that are abundant at a sample also overexpressed at a sample?
- Are PD and VD as metabolically similar at the MAG-resolved level as they are at the gene-resolved level?
- Why are PD and VD similar metabolically at community level but not at the MAG level?

What our answers are:

- Gene-resolved abundance and expression of metabolic pathways is more similar than expected at VD/PD given different geochemistry
- PD and VD are quite different at the MAG-resolved level. Distinct sets of MAGs are active at VD and PD
- When looking at high-quality MAGs, VD / PD are enriched for certain metabolic pathways
- Most likely right now:
 1. PD / VD same in gene expression because our gene-level analyses aren't detailed enough. We need to look at whole modules/pathways and more modules overall. I currently find it improbable that we will continue to find similarities between VD and PD.
 2. PD / VD are different in MAG expression because of geochemical differences. When MAGs with the same metabolism show different expression patterns, it's probably just niche partitioning. No evidence pointing to nutrient cycling so far

Introduction

Quick Primer on hydrothermal vent biology

Geochemistry of Von Damm and Piccard

Previous Results and Gaps in knowledge

- Previous results in (Anderson et al., 2017; Reveillaud et al., 2016) regarding MAG evolution, MAG relative abundance, and gene-resolved functional differences between VD and PD
- What data do we have?
 1. MAGs
 2. Metagenomes
 3. Metatranscriptomes
- For this paper...
 1. Confirm previous results on functional differences / similarities
 2. Metatranscriptomes: gene- and MAG-resolved expression across samples

Functional metagenomics overview

3. Pangenomics-type analysis: comparison of MAG functional potential
4. Merging of MAG-resolved metatranscriptomics and functional data: Does metabolism predict expression? If not, why not? What's the role of biotic factors like nutrient cycling?

Results

General setup

- 73 high-quality MAGs from (Reveillaud et al., 2016) and Rika. >70% complete, <10% redundant
- Because of poor assembly, excluded FS841 metagenome for binning and mapping

Gene Abundance Across Samples

- **Fig. 1** shows abundance of key metabolic genes at the community level across samples
- No significant differences between VD and PD in community-level abundance of key metabolic genes
- PD / VD also had similar variance in expression within sample sites. While some sites at both PD and VD had particularly high relative abundance for some genes and 0 for others, most sites expressed a majority of genes to at least some extent.

Gene expression across samples

- **Fig. 2** shows few differences between VD and PD in gene-resolved metabolic expression. VD has higher expression of methanogenesis (*mcr*) gene and some hydrogenases
- Hot Chimlet, Shrimp Canyon show very low / 0 expression for a large majority of metabolic genes
- VD has higher diversity of gene expression, with more richness in number of genes expressed and higher evenness in genes

MAG expression across samples

- **Fig. 3** shows clear differences between VD and PD in MAG expression
- Defining “active” and “elevated activity” – currently, these are somewhat qualitative. Since Fig. 3 shows row-based (ie. MAG-based) z-scores, the phrase “MAG X is active/has elevated activity at sample(s) Y” means that X’s expression at Y is approximately 1 standard deviation higher than the average expression of MAG X at all samples.
- MAGs can be clustered into 3 broad groups: those active at Shrimp Hole 2012 and/or Shrimp Hole 2013, those active at Piccard, and those active at Von Damm, excluding Shrimp Hole (i.e. Ginger Castle, Hot Cracks #2 and Old Man Tree (2013))
- All MAGs have elevated activity at at least one site. No MAGs have average / close to average activity across all samples
- Noticeable differences between Shrimp Hole 2012 and Shrimp Hole 2013. MAGs that were active in 2012 have decreased activity in 2013, and several MAGs that weren’t active in 2012 are active in 2013
- Although MAGs with same taxonomy frequently active at same samples, some taxa (Sulfurovum, Methanococcus, Flavobacteriales, Nanohaloarchaea, Desulfobacterales, Thiotrichales, Aquificales) are active at more than one of the 3 clusters
- Most MAGs only active at 1-2 samples. Sulfurovum, Other Campylobacterales, Nautiales MAGs are unique for being active at up to 4 samples (mainly in Piccard)

MAG-resolved metabolic potential

- **Fig. 4** shows module completion ratio (MCR) of all high-quality MAGs across 73 KEGG Modules selected for carbohydrate metabolism, carbon fixation, methanogenesis and *aerobic* methane oxidation, sulfur/nitrogen redox, and various membrane proteins such as cytochromes and Mn/Zn/Fe/S/N transporter. We also included several hemolysin transporters as a potential marker for ectosymbiosis as described in Anantharaman et al. (2016) for Nanohaloarchaea
- Module completion ratio is a number between 0 and 1 describing the average completeness in the MAG for each reaction step in the module
- MAGs clustered by taxonomy, w/ exception of Thiotrichales, Aquificales, Flavobacteriales, Methanomicrobia. Some pseudomonas cluster separately as well
- There are 6 broad metabolic clusters that emerge, with the following *broad* characterizations:
 1. Core carbon metabolisms: At least 3-4 modules consistently present with high MCRs (>0.8) in all but 3 MAGs
 - glycolysis, gluconeogenesis, generic pentose phosphate pathway, citrate cycle, pyruvate oxidation
 - F-type ATPase (ATP synthase), iron complex transport, NADH quinone oxidoreductase
 2. Non-core carbon metabolism + N,S reduction : At least 3-4 modules consistently present with medium MCRs (~0.5) in all but 3 MAGs
 - Wood-Ljungdahl, formaldehyde assimilation, nitrate reduction, dissimilatory sulfate reduction, 3-hydroxypropionate bi-cycle, dicarboxylate-hydroxylbutyrate cycle, methylaspartate cycle, semi-phosphorylative Entner-Doudoroff
 - Cytochrome bd, acetate=>CH₄ methanogenesis
 3. Carbohydrate transport, cytochromes, N reduction, S oxidation: present w/ high MCRs in Sulfurovum, Alteromonas, Pseudomonas, Sphingomonas
 - PTS transporters
 - Cytochromes c, c-cbb3, bc1
 - Thiosulfate oxidation (sox)
 - Assimilatory nitrate reduction
 4. Rare metabolisms:Present with high MCR in 2-4 MAGs / module
 - Fumarate reductase
 - Energy-coupling factor transport
 - Nitrate/nitrite transport
 - Mn/Zn/Fe transport
 5. Carbon metabolism, cytochromes, sulfate transport: Present w/ high MCR in Pseudomonas, Alteromonas, Sphingomonas
 - Pentose phosphate, phosphate acetyltransferase kinases, Entner-Doudoroff, D-glucuronate / D-galacturonate degradation
 - Cytochrome c, bc1, o ubiquinol
 - Sulfate transport system
 6. Methanogenesis: Present with high MCR in Methanococci, Methanomicrobia
 - Pentose phosphate (archaea), 2-oxocarboxylic acid chain extension, CO₂, methanol, trimethylamine methanogenesis
 - Acetyl-CoA pathway

David Galambos

Functional metagenomics overview

- N fixation
- V/A type ATPase (H^+ transporter)
- Coenzyme M, F420 biosynthesis

Merging MAG expression and MAG metabolism

- No nutrient cycling relationships are currently obvious
- Nanohaloarchaea have low / 0 MCR for all modules except gluconeogenesis, glycolysis. 3 Nanohaloarchaea MAGs are active at 3 different samples and have an expression patterns associated with 3 different other MAGs:
 1. 43 associated w/ Methanococci_13b @ Ginger Castle
 2. 54 associated w/ Thiotrichales_3 @ Shrimp Hole
 3. 18 loosely associated w/ Sphingomonas, Pseudomonas @ Shrimp Hole 2013 and Hot Cracks
- There are several MAG groups with similar metabolisms but very different expression patterns. The following is a list of those taxa followed by a list of where bin numbers in that taxon are active
 1. Sulfurovum – 43 @ Shrimp Hole, 31 @ Ginger Castle, Hot Cracks, 43b, 9, 99, 37,13 @ Piccard
 2. Methanogens
 - Methanococci – 7 @ Shrimp Gulley, 13, 13b, 24 @ Ginger Castle, 69b @ Hot Chimlet, Shrimp Canyon (active everywhere except Shrimp Hole)
 - Methanomicrobia – 41,84,45 @ Shrimp Hole 2012, 78 @ Shrimp Hole 2013
 3. Desulfobacterales
 - 16,26 @ Ginger Castle, Hot Cracks, 29 @ Shrimp Gulley

Can MAG metabolism predict activity at a certain sample site?

- Since MAGs cluster by taxonomy for both expression and metabolism, some metabolisms seem to be associated with higher MAG expression at a sample site
- Conducted two-tailed Mann-Whitney U Test on a bin set and compared distribution of MCRs for a certain module to distribution of MCRs for that same module in all other MAGs (ex. can compare whether distribution of MCRs for glycolysis in MAGs active in Piccard is different from distribution for MAGs active at all other sites)
- Modules that were significantly enriched/depleted ($p < 0.05$) in MAGs active at each sample are shown in **Tables 1-3** (Piccard, Shrimp Hole, and Von Damm minus Shrimp Hole, respectively)

Discussion

Reiterate motivation/methods

1. Confirm previous results on functional differences / similarities
2. Metatranscriptomes: gene- and MAG-resolved expression across samples
3. Pangenomics-type analysis: comparison of MAG functional potential
4. Merging of MAG-resolved metatranscriptomics and functional data: Does metabolism predict expression? If not, why not? What's the role of biotic factors like nutrient cycling?

Limitation of Geochemical Data

- Lack of geochemical data beyond temperate and pH is a major limitation

David Galambos

Functional metagenomics overview

- While we know that VD and PD have different depths/ surrounding geology, detailed data for metabolites like methane, sulfur, manganese, iron, carbohydrates makes metabolic analysis difficult
- Harder to assess magnitude of biotic factors on MAG expression without knowledge of abiotic factors

Confirmation of previous results

- Results indicating little difference in gene-resolved metabolism between VD / PD consistent with Reveillaud et al. (2016). **Fig. 1** is a more detailed version of bubble plot in (Reveillaud et al., 2016) with more genes and sample-resolved (rather than vent field-resolved) data
- Despite (more than likely) different geochem between VD / PD, the abundance of metabolic genes at two fields is identical
- Sets up the possibility that expression between the two sites is different. It's possible that both sites possess all key metabolic genes in about equal abundance, but they are expressed selectively. This could make sense given the different geochemistry

Metabolic gene expression

- Selective overexpression is observed for some genes like kor at Old Man Tree (kor is identically abundant at all samples but is expressed more at certain samples)
- Still, in **Fig. 2**, only trends differentiating VD / PD are higher expression of methanogenesis (mcr) and certain hydrogenases at VD
- Not sure how to rationalize the above – VD doesn't have more H₂ or CO₂, and methanogenesis isn't that much more thermodynamically favorable at VD
- Higher diversity of expression at VD is consistent with higher MAG diversity and higher energy availability from all metabolic pathways at VD (Reveillaud et al., 2016)
- Still, given geochem differences, it's surprising that this is the only gene expression difference. C, N, S, Fe metabolisms are similar between VD / PD
- This analysis is still limited to a small range of metabolic pathways, and those pathways are represented by only one or two genes. Studies involving expression of larger-scale modules across samples is needed to get a more complete gene-level metabolic picture

MAG-resolved expression

- **Fig. 3** shows clear differences in MAG expression between PD/ VD. Piccard and VD cluster separately in the hierarchical clustering of samples.
 1. Piccard samples are more closely related to each other than VD samples are – consistent with greater diversity in metabolism / community composition in VD
 2. Shrimp Hole clusters completely separately. Ginger Castle, Hot Cracks #2, Old Man Tree 2013 are all more similar to Piccard than to Shrimp Hole. This is consistent with Shrimp Hole being more of a cold seep than a vent. We would expect that cold seep conditions would lead to different community composition
- The hierarchical clustering by MAG expression *somewhat* similar to MDS clustering based on bacterial 96% OTU abundance (Reveillaud et al., 2016). Piccard sites closely related to one another in both. However, in OTU clustering, Shrimp Hole more closely related to Ginger Castle, Hot Cracks, Old Man Tree than those sites are to Piccard. Differences are to be expected, since these are two different clustering based on different data. In sum, the expression method emphasizes the dissimilarity of Shrimp Hole from all other samples. This is reflected in cluster diagram in Reveillaud et al. (2016) based on archaeal OTUs, which places Shrimp Hole separately from all other samples

MAG-level Metabolism

- In **Fig. 4**, Most MAGs cluster by taxonomy, but some taxa have diverse metabolisms
 1. MAGs with similar metabolisms are classified at the genus level. MAGs diverse metabolisms are mainly those with higher classifications (class, phylum, order, etc), Aquificales, Thiotrichales, Methanomicrobia. – this alone can explain the greater diversity. We would need to get more detailed taxonomic classifications for these MAGs to better assess metabolic differences
 2. Bacteria and archaea cluster separately, showing (as expected) that bacteria and archaea possess very different carbohydrate metabolisms.
 3. For archaea, cursory evaluation shows that carbon fixation metabolisms match what we expect (Berg et al., 2010)
 4. So far, no evidence of unusual, unexpected metabolisms that could've arisen due to HGT
 5. Clusterings of modules makes sense given what we know about which pathways are found together. Ex. methanogenesis found together with coenzyme M synthesis. These clusterings also allow us to find “hidden” connections between pathways that may seem unrelated but nevertheless occur together, such as the phosphate acetyltransferase-kinase pathway and sulfate transporters
 - Cytochromes are associated with N / S reduction, which makes sense, but it is unexpected that no cytochromes are associated with core carbon metabolisms. We would expect more cytochromes to be associated with carbohydrate metabolisms that could lead to an electron transport chain
 6. Metabolic clusters provide for a more robust qualitative analysis of whether a MAG contains a metabolic pathway. If metabolism X is frequently associated with other metabolisms, the presence or absence of these “other” metabolisms can strengthen or weaken the case for X itself being present in a MAG.
 - For small/short modules, this clustering could be confounded by metabolic genes / pathways being on the same contig. In that situation, those metabolisms would always occur together
 - Still, given that clusters are large, it's unlikely that 4+ modules routinely occur together just by virtue of being on the same contig
 7. In Tables 1-3, tests suggest that when looking at MAGs, there are still metabolic differences among the samples. Although a plot of module expression across samples may still be useful, these tables can be analyzed to see which modules may be associated with higher MAG expression at a certain sample site.

Connecting MAG-level Expression and Metabolism

- Nanohaloarchaea seem to have very little metabolism, but nevertheless have expression patterns that follow MAGs with much more established metabolism. Scenario of two MAGs with radically different metabolisms but identical expression patterns points to some sort of interdependence/dependence. Reduced metabolism and the nanohaloarchaea classification point to possible ectosymbiosis as described for nanoarchaea in Huber et al. (2002)
 1. 43 associated w/ Methanococci_13b @ Ginger Castle
 2. 54 associated w/ Thiotrichales_3 @ Shrimp Hole
- MAG sizes/lengths for nanohaloarchaea provide mixed evidence. MAG 43 is 0.6 mbp, but 54 is 1.3 mbp. Nevertheless, further study can investigate the composition of the 54 genome given the lack of metabolic genes. If there are few metabolic genes, what is taking up 1.3 mbp of space?

David Galambos

Functional metagenomics overview

- In determining biotic factors, MAGs that have a) different metabolisms but similar expression patterns or b)similar metabolisms and different expression patterns are the best candidates for probing the influence of biotic factors.
- A) is a case where these two MAGs might be engaging in some type of nutrient cycling that yields a similar expression pattern across samples (and across different geochemical conditions) despite different metabolism
- So far, with the possible exception of the Nanohaloarchaea, I haven't found promising leads for case a). Again, an analysis of module expression across samples would be helpful in assessing which metabolisms are frequently active together and might in turn indicate cycling
- B) illustrates cases where metabolism of that MAG alone isn't influencing MAG expression, which means that the availability of nutrients isn't the main factor in expression. If nutrients were the factor, then MAGs with the same metabolism would likely have similar expression patterns across all samples. Instead, these MAGs might be differentiated by condition preferences (an abiotic factor) that are too subtle to be detected by the current metabolic analysis. However, this could also be an indication of MAGs with similar metabolisms facing different evolutionary pressures, ecological competition from other MAGs, or just simple niche partitioning where one species in low-level taxonomic group happens to become abundant at one site while a closely related, functionally identical species becomes abundant at a different site.
- Methanogens are an example of case B). Despite both being methanogens, Methanomicrobia are expressed exclusively at Shrimp Hole, while Methanococci are expressed everywhere else. Why the stark division?

Connecting functional data to an evolutionary context

- In the midst of functional data, important not to forget longer-term advantages conferred through well-adapted metabolisms, along with possibility of evolutionary adaptations that aren't covered by nutrient-focused metabolic analysis
- An example of case B) above are the Sulfurovum. These MAGs have essentially identical metabolisms, but they have very different expression patterns. 6 MAGs active at Piccard, one at Shrimp Hole, and two at Von Damm (minus Shrimp Hole). Moreover, the Piccard MAGs are all active at all 4 Piccard sites, while the VD MAGs are active at 1-2 sites each. Although it's clear that Sulfurovum is able to thrive in a diverse set of environments, Thus, assessing the evolutionary context of each helps shed light on what is causing this division
- Given that Sulfurovum MAGs in PD have more widespread activity across PD samples despite harsher conditions and more competition from PD MAGs, it is notable that there is no difference in SAAV/SNV ratio between Piccard/VD MAGs, and there aren't other evolutionary metrics that explain the division of separate species in the three main expression clusters (VD minus Shrimp Hole, PD and Shrimp Hole) (Anderson et al., 2017).

General limitations

- “Absence of evidence =/= evidence of absence”, and other challenges of MAG analysis, brought to you by Rika Anderson
- While MAGs high quality, not 100% complete
- Similarity within taxa is reassuring, but are differences within taxa truly reflective of differences or just binning error?
- An ortholog can catalyze different rxns in different organisms, as shown by archaea example

David Galambos

Functional metagenomics overview

- Need *in vivo* study to confirm metabolism of certain nutrients, either in the metagenome or in MAGs
- Taxonomic assignment is challenging

Summary of future work

- Analyze module expression across pathways to better assess nutrient cycling
- Look at syntrophy within modules as another possibility
- Make trees of the MAGs and of mcr genes (to assess ANME)
- Given that some MAGs are active at multiple sites, attempt to recover a “version” of the MAG at each site to see if this situation actually represents two different species present at the samples. Or, is it really the same species being active across multiple samples?
- Attempt to assign more detailed taxonomy to MAGs with class/phylum level classification
- Create more rigorous definitions for elevated expression and metabolic potential – what’s the cutoff for a MAG having a pathway or no?

Materials and Methods

From (Anderson et al., 2017; Reveillaud et al., 2016)

Sample Collection

DNA/RNA Extraction, Sequencing

Stuff we did w/ Rika:

Merge/filter reads

- Filter out low-quality reads?
- Merge overlapping paired-end reads

Metagenome assembly

- IDBA
- Each sample assembled individually — no coassembly
- FS841 metagenome discarded — 5 contigs
- Assembly statistics? Contigs, total length, N50, etc?

Genome/ORF annotation

- IMG/JGI pipeline
- Prodigal
- KEGG/ pfam / COG annotations

Mapping

- Bowtie2 w/ regular parameters and samtools

Binning and Bin Evaluation

- Anvio’s CONCOCT (which version of anvi’o?)
- Bin stats for completeness/ redundancy/total length, etc
- Taxonomic assignment based on manual assessment of phylosift output (which version of the phylosift database?)
- MAGs Methanococcus_69b, Methanococci_24 were refined from the original based on initial analysis of potential for methanogenesis

Calculation of abundance and expression values

- Bins

David Galambos

Functional metagenomics overview

1. Average coverage across whole MAG as calculated by anvi'o, divided by a sample-specific normalization factors
 2. Abundance: normalization factor = avg of avg coverages of 35 single-copy genes
 3. Important to note that we chose to look at just RNA to compare relative magnitude of metabolic processes across samples. NOT differential expression.
 4. Expression: # of metatranscriptomic reads in the sample
- Genes
 1. Sum of average coverages of every open reading frame with that annotation divided by a normalization factor. When assessing enzymes, coverage of all subunits was summed.
 2. Important to note that we chose to look at just RNA to compare relative magnitude of metabolic processes across samples. NOT differential expression.
 3. Abundance: normalization factor = avg of avg coverages of 35 single-copy genes
 4. Expression: # of metatranscriptomic reads in the sample

Module Completion Ratio (MCR)

MCR is a real number between 0 and 1. It is calculated by calculating the proportion of necessary KEGG orthologs possessed by the MAG for each reaction step. The completeness proportions for all reactions in the module are then averaged to obtain a single MCR for a MAG for a specific module.

Calculation of MCR Enrichment in MAGs active at PD/VD

- Used two-tailed Mann-Whitney U test as implemented in R `wilcox.test()`
- Null hypothesis is that two samples being compared have the same distribution of values
- Like t-test, but doesn't assume normal dist. in samples
- Can be used on samples with different sample sizes. More robust to outliers than t-test as well.

Supplementary Figures (Later)

- MAG data
- Assembly data
- Tree of MAGs
- Our limited geochem data
- The full tables from the MCR enrichment calculations with non-significant values + U-values/more stats include

David Galambos

Functional metagenomics overview

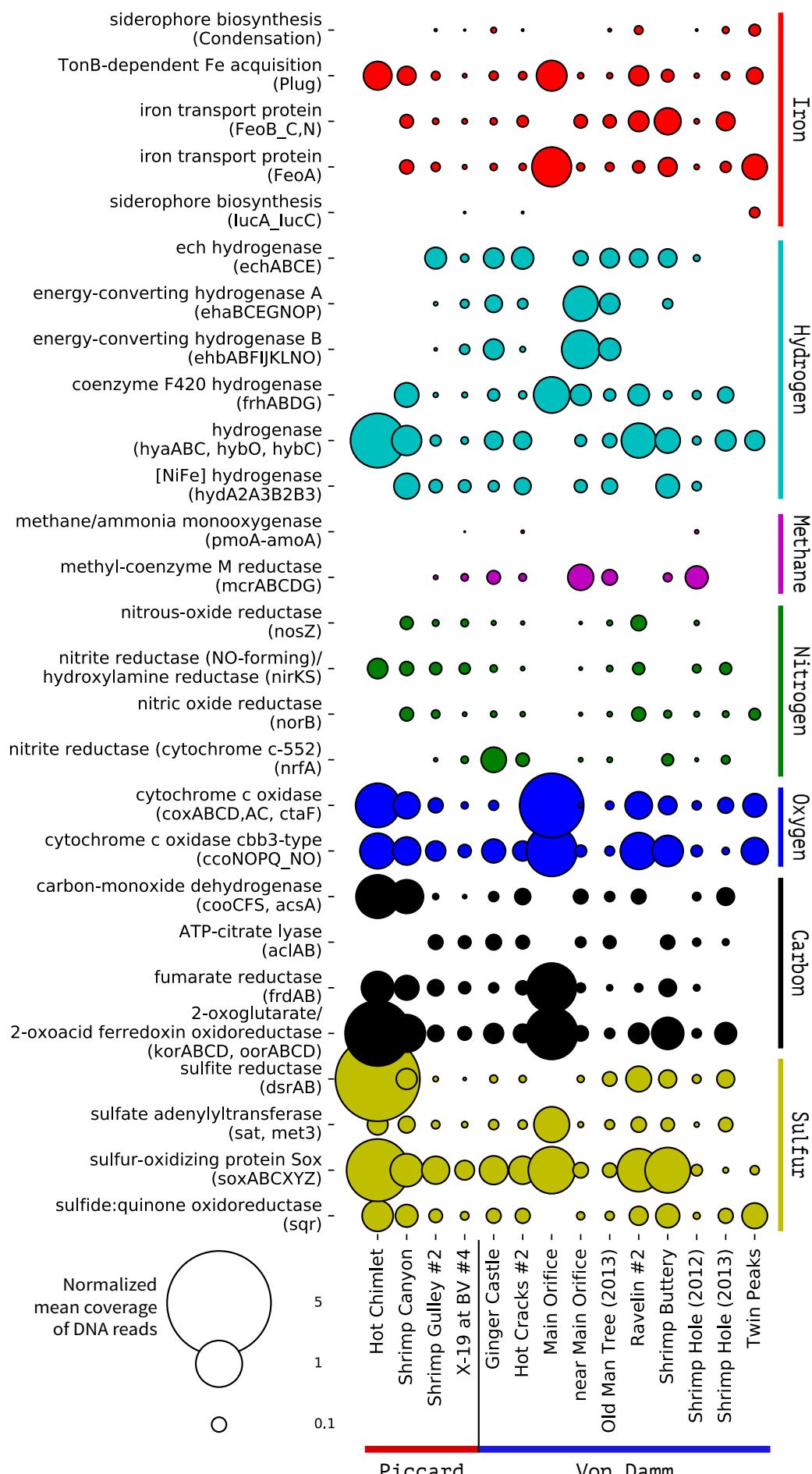


Figure 1. Normalized relative abundance of key metabolic genes. The x axis shows samples at Piccard and Von Damm; the y axis shows key genes. Bubble size represents normalized average coverage by DNA reads. Coverage values were normalized using coverage of 35 single-copy genes.

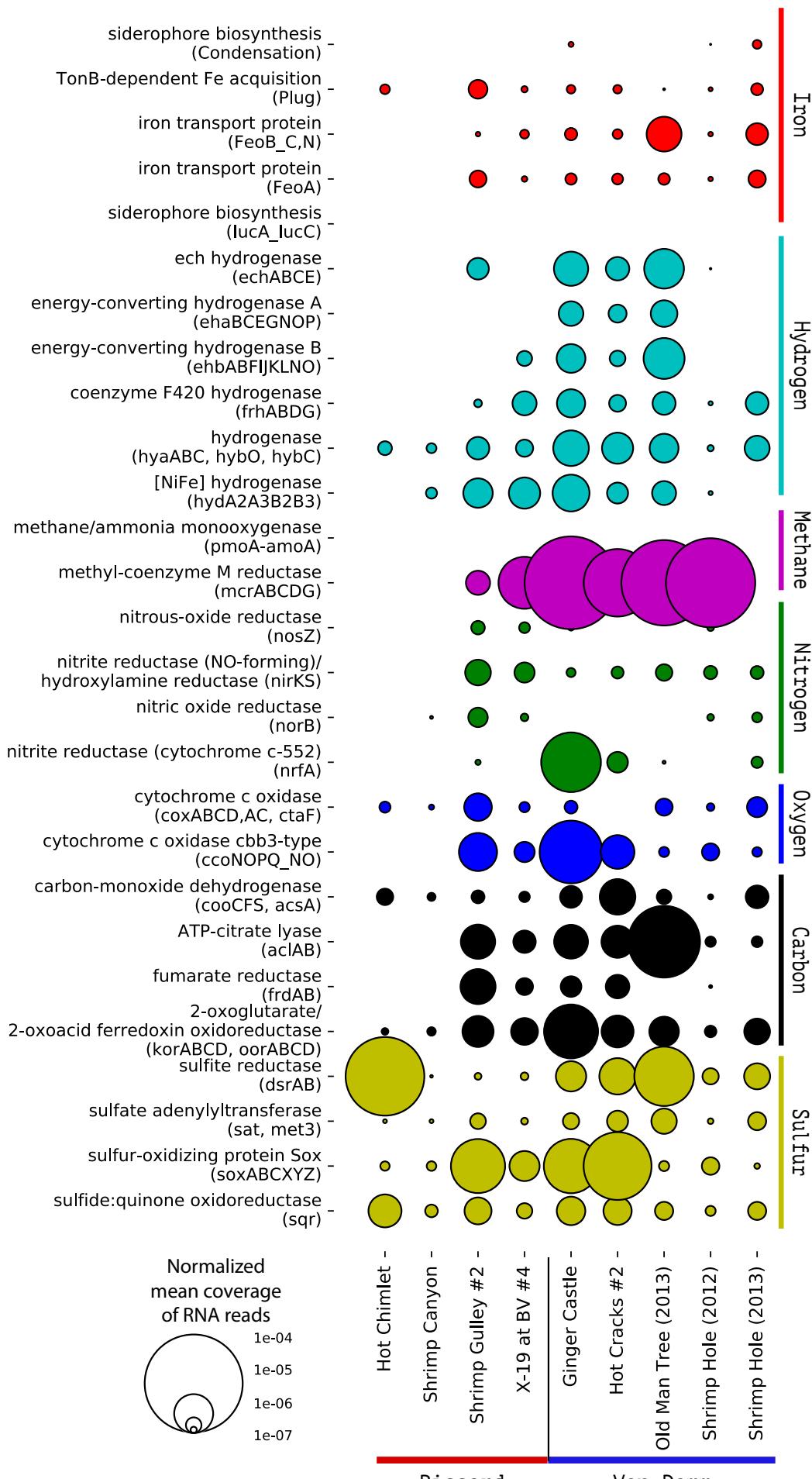


Figure 2. Normalized relative expression of key metabolic genes. The x axis shows samples at Piccard and Von Damm; the y axis shows key genes. Bubble size represents normalized average coverage by RNA reads. Coverage values were normalized by dividing by the number of metatranscriptomic reads in the sample, excluding

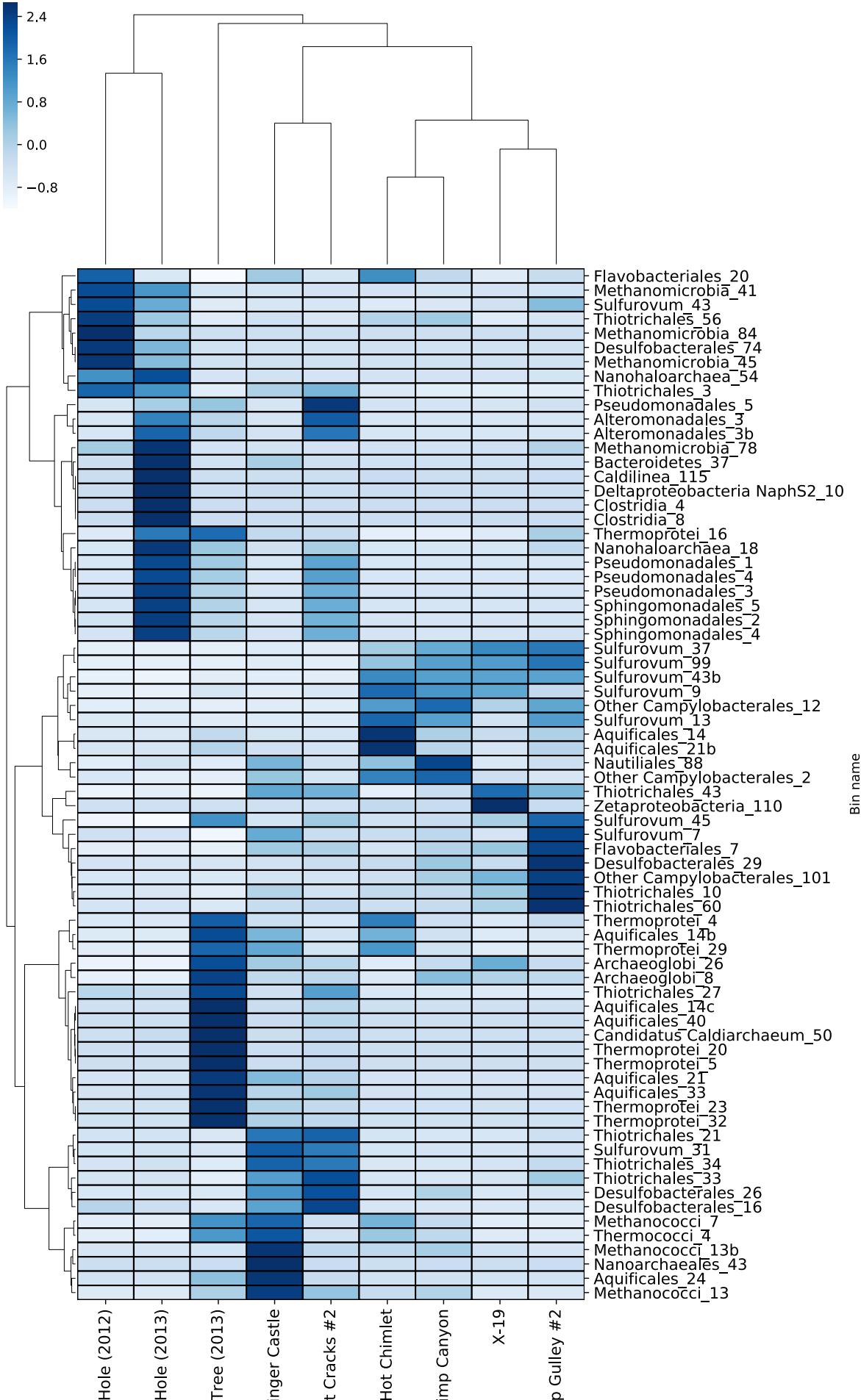


Figure 3. Normalized MAG Expression Across Samples. The x axis shows samples at Picard and Von Damm; the y axis shows high-quality MAGs. MAG coverage values were normalized using the number of metatranscriptomic reads in each sample. A z-score transformation was applied to each row; the legend indicates the z-score for a cell relative to the mean for all values in that row. Dendograms indicate hierarchical clustering of samples and MAGs.

Functional metagenomics overview

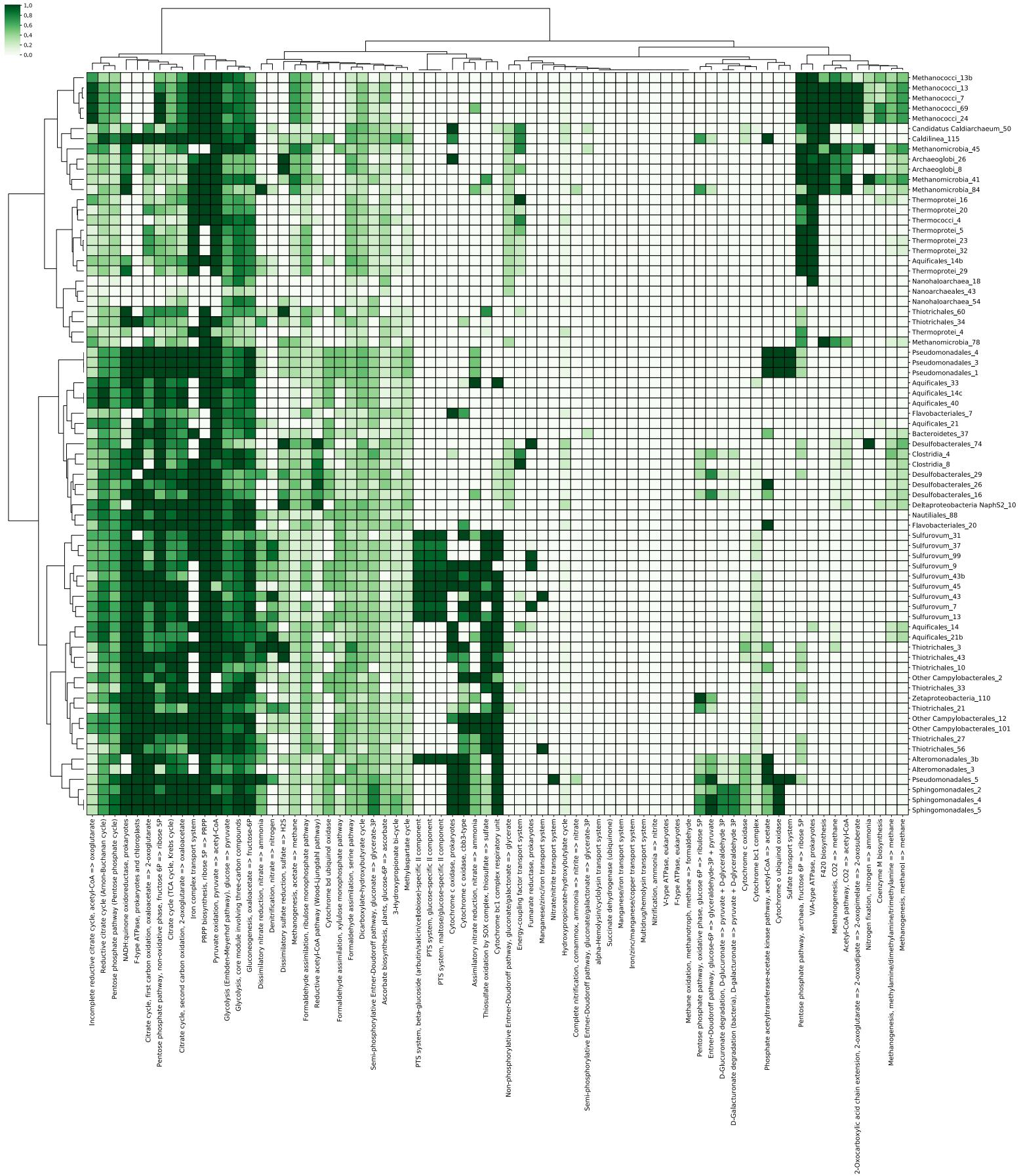


Figure 4. Metabolic potential of MAGs for key metabolic modules. The x axis shows metabolic pathways selected from the KEGG Modules database; the y axis shows high-quality MAGs. The legend shows the module completion ratio (MCR) for each MAG in each module. Module completion was based on the presence or absence of a given

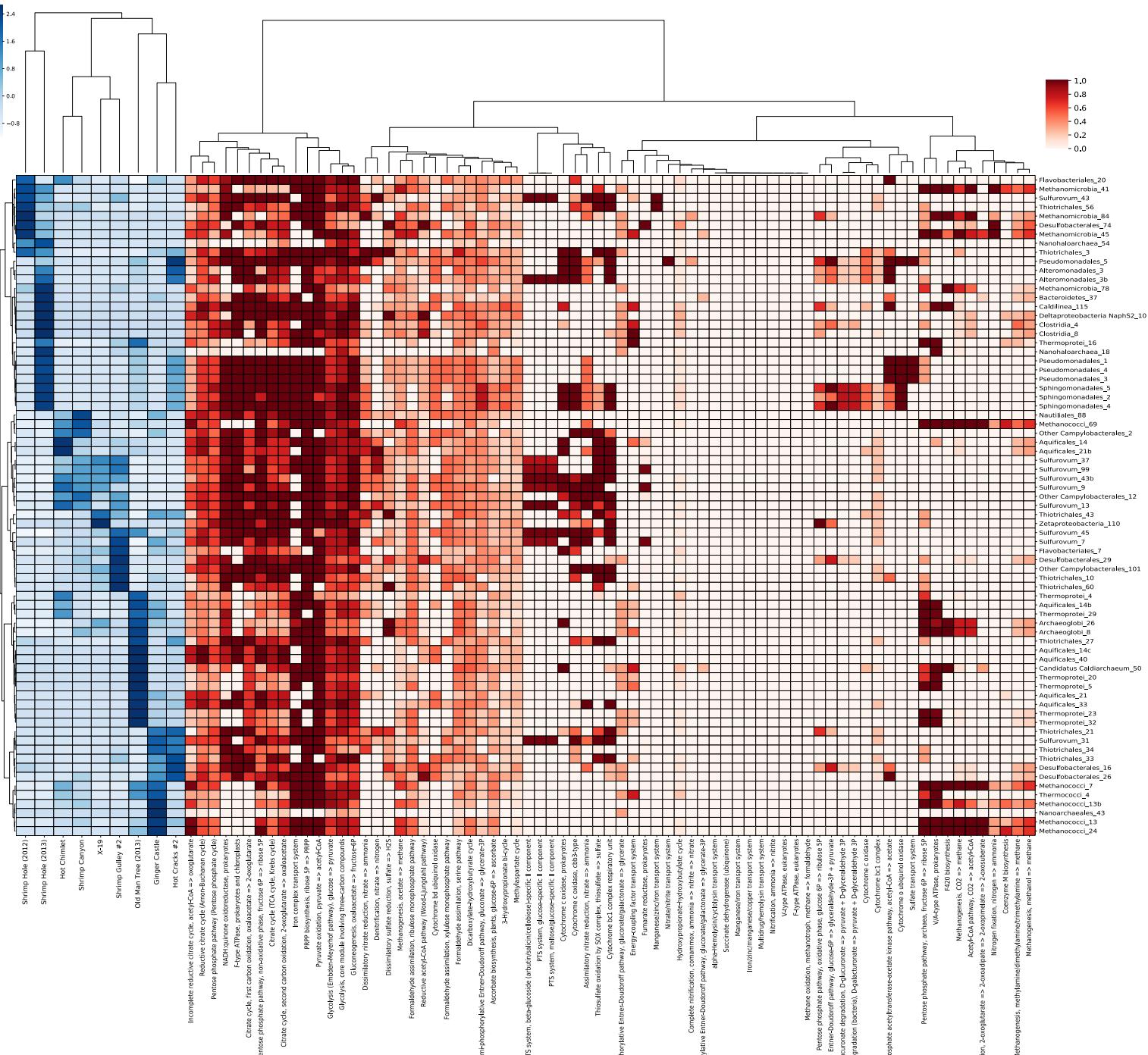


Figure 5. Combination plot of normalized MAG expression across samples and metabolic potential at modules. This plot combines figures 3 and 4. The blue legend indicates the z-score of MAG expression relative to the average expression across all samples. The red legend indicates module completion ratio for MAGs across modules. Dendrograms at the top indicate hierarchical clustering of samples and modules, respectively. Dendrogram at the left is based on hierarchical clustering of MAGs based on expression; thus, the order of MAGs on the y-axis is based on

David Galambos

Functional metagenomics overview

Table 1. Modules that are significantly enriched or depleted for MAGs highly expressed at Piccard.

Module	Mean MCR (Piccard)	Mean MCR (all others)	p-value
Pentose.phosphate.pathway..non.oxidative.phase..fructose.6P....ribose.5P	0.9375	0.64622642	0.00140501
Gluconeogenesis..oxaloacetate....fructose.6P	0.8875	0.74292453	0.00679898
NADH.quinone.oxidoreductase..prokaryotes	0.87592593	0.56785465	0.00878524
Citrate.cycle..second.carbon.oxidation..2.oxoglutarate....oxaloacetate	0.85916667	0.6745283	0.02437087
Citrate.cycle..TCA.cycle..Krebs.cycle.	0.81822917	0.64563679	0.02217609
F.type.ATPase..prokaryotes.and.chloroplasts	0.80116377	0.54675868	0.00923484
Reductive.citrate.cycle..Arnon.Buchanan.cycle.	0.6875	0.54363208	0.01084097
Cytochrome.bc1.complex.respiratory.unit	0.675	0.28301887	0.00182766
Thiosulfate.oxidation.by.SOX.complex..thiosulfate....sulfate	0.65333333	0.16855346	8.44E-05
Pentose.phosphate.pathway..Pentose.phosphate.cycle.	0.6	0.46698113	0.0099358
Cytochrome.c.oxidase..cbb3.type	0.5	0.22012579	0.01484015
Incomplete.reductive.citrate.cycle..acetyl.CoA....oxoglutarate	0.49821429	0.37884097	0.04758455
Assimilatory.nitrate.reduction..nitrate....ammonia	0.475	0.16981132	0.00792914
Formaldehyde.assimilation..ribulose.monophosphate.pathway	0.45	0.31132075	0.00470391
Formaldehyde.assimilation..xylulose.monophosphate.pathway	0.4125	0.16037736	1.26E-05
Dissimilatory.nitrate.reduction..nitrate....ammonia	0.39375	0.20125786	0.00342561
Denitrification..nitrate....nitrogen	0.371875	0.09669811	3.63E-06
Semi.phosphorylative.Entner.Doudoroff.pathway..gluconate....glycerate.3P	0.37	0.29811321	0.04045895
PTS.system..glucose.specific.II.component	0.32490577	0.05539511	0.00197756
PTS.system..maltose.glucose.specific.II.component	0.32490577	0.05539511	0.00197756
PTS.system..beta.glycoside..arbutin.salicin.celllobiose..specific.II.component	0.32490577	0.05539511	0.00197756
Methanogenesis..acetate....methane	0.21840278	0.31142558	0.04566785
Fumarate.reductase..prokaryotes	0.1702362	0.01958252	2.12E-06
Cytochrome.bc1.complex	0.15444444	0.06624738	0.00182766
Pentose.phosphate.pathway..archaea..fructose.6P....ribose.5P	0.13333333	0.41509434	0.00576488
Methanogenesis..CO2....methane	0.06796875	0.191111635	0.04435017
V.A.type.ATPase..prokaryotes	0.065	0.37305975	0.01306738
Hydroxypropionate.hydroxybutylate.cycle	0.03571429	0.08692722	0.00634321
Phosphate.acetyltransferase.acetate.kinase.pathway..acetyl.CoA....acetate	0.025	0.22641509	0.02894734
Non.phosphorylative.Entner.Doudoroff.pathway..gluconate.galactonate....glycerate	0.01666667	0.16352201	0.00056856

Table 2. Modules that are significantly enriched or depleted for MAGs highly expressed at Shrimp Hole.

Module	Mean MCR (Shrimp Hole)	Mean MCR (all others)	p-value
Phosphate.acetyltransferase.acetate.kinase.pathway..acetyl.CoA....acetate	0.365384615	0.063829787	0.00022109
Methanogenesis..acetate....methane	0.35042735	0.250265957	0.0078381
Reductive.acetyl.CoA.pathway..Wood.Ljungdahl.pathway.	0.305860806	0.172239108	0.00940031
X3.Hydroxypropionate.bi.cycle	0.270299145	0.172379827	0.0022866
Cytochrome.o.ubiquinol.oxidase	0.269230769	0	0.00021214
Formaldehyde.assimilation..ribulose.monophosphate.pathway	0.259615385	0.39893617	0.00154824
Methylaspartate.cycle	0.258741259	0.205029014	0.04330201
Entner.Doudoroff.pathway..glucose.6P....glyceraldehyde.3P...pyruvate	0.211538462	0.04787234	0.0056925
Pentose.phosphate.pathway..oxidative.phase..glucose.6P....ribulose.5P	0.192307692	0.04964539	0.00250864
Sulfate.transport.system	0.153846154	0	0.00629067
Cytochrome.c.oxidase	0.144230769	0.015957447	0.00321252
D.Galacturonate.degradation..bacteria...D.galacturonate....pyruvate...D.glyceraldehyde.3P	0.126923077	0.008510638	0.00415935
D.Glucuronate.degradation..D.glucuronate....pyruvate...D.glyceraldehyde.3P	0.123076923	0.008510638	0.00441824
Denitrification..nitrate....nitrogen	0.117788462	0.20212766	0.02487076
Hydroxypropionate.hydroxybutylate.cycle	0.107142857	0.053951368	0.00540892
Fumarate.reductase..prokaryotes	0.038245462	0.073366286	0.00927101
Succinate.dehydrogenase..ubiquinone.	0.000798118	0	0.01905661

David Galambos

Functional metagenomics overview

Table 3. Modules that are significantly enriched or depleted for MAGs highly expressed at Ginger Castle, Hot Cracks #2, and Old Man Tree 2013.

Module	Mean MCR (Von Damm - Shrimp Hole)	Mean MCR (all others)	p-value
Gluconeogenesis..oxaloacetate....fructose.6P	0.696428571	0.83423913	0.00192961
Glycolysis..Embden.Meyerhof.pathway...glucose....pyruvate	0.664285714	0.760869565	0.02154107
Citrate.cycle..second.carbon.oxidation..2.oxoglutarate....oxaloacetate	0.60297619	0.795289855	0.004214
Pentose.phosphate.pathway..non.oxidative.phase..fructose.6P....ribose.5P	0.598214286	0.804347826	0.00410285
Citrate.cycle..TCA.cycle..Krebs.cycle.	0.568824405	0.766077899	0.00080255
Reductive.citrate.cycle..Arnon.Buchanan.cycle.	0.525297619	0.613677536	0.04940226
Pentose.phosphate.pathway..archaea..fructose.6P....ribose.5P	0.523809524	0.217391304	0.00606449
Citrate.cycle..first.carbon.oxidation..oxaloacetate....2.oxoglutarate	0.511904762	0.717391304	0.01414645
NADH.quinone.oxidoreductase..prokaryotes	0.504144621	0.728234031	0.0042037
V.A.type.ATPase..prokaryotes	0.4933994	0.157760513	0.00186344
Pentose.phosphate.pathway..Pentose.phosphate.cycle.	0.419642857	0.554347826	0.00302847
F.type.ATPase..prokaryotes.and.chloroplasts	0.392453786	0.739594048	0.00080549
Semi.phosphorylative.Entner.Doudoroff.pathway..gluconate....glycerate.3P	0.228571429	0.373913043	0.0002512
Cytochrome.bc1.complex.respiratory.unit	0.205357143	0.494565217	0.01693873
Methylaspartate.cycle	0.185064935	0.249011858	0.01124989
Dissimilatory.sulfate.reduction..sulfate....H2S	0.166666667	0.322463768	0.00164285
Reductive.acetyl.CoA.pathway..Wood.Ljungdahl.pathway.	0.149659864	0.260869565	0.00131524
X3.Hydroxypropionate.bi.cycle	0.143849206	0.248792271	0.00150049
Thiosulfate.oxidation.by.SOX.complex..thiosulfate....sulfate	0.133333333	0.397101449	0.00849912
Cytochrome.c.oxidase..cbb3.type	0.130952381	0.405797101	0.01310544
Formaldehyde.assimilation..xylulose.monophosphate.pathway	0.116071429	0.298913043	0.00043549
Dissimilatory.nitrate.reduction..nitrate....ammonia	0.111607143	0.335144928	0.00065824
Assimilatory.nitrate.reduction..nitrate....ammonia	0.089285714	0.347826087	0.00206994
Cytochrome.c.oxidase..prokaryotes	0.080357143	0.358646469	0.00869751
Denitrification..nitrate....nitrogen	0.073660714	0.22826087	0.03178531
Cytochrome.bc1.complex	0.050396825	0.112801932	0.01693873
Pentose.phosphate.pathway..oxidative.phase..glucose.6P....ribulose.5P	0.035714286	0.137681159	0.04630267
Cytochrome.c.oxidase	0.008928571	0.092391304	0.03046795
Cytochrome.o.ubiquinol.oxidase	0	0.152173913	0.03206181

Literature Cited

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic Local Alignment Search Tool. *J Mol Biol* *215*, 403-410.

Anantharaman, K., Breier, J.A., Sheik, C.S., and Dick, G.J. (2013). Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *P Natl Acad Sci USA* *110*, 330-335.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., *et al.* (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* *7*.

Anderson, R.E., Reveillaud, J., Reddington, E., Delmont, T.O., Eren, A.M., McDermott, J.M., Seewald, J.S., and Huber, J.A. (2017). Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. *Nat Commun* *8*, 1114.

Berg, I.A., Kockelkorn, D., Ramos-Vera, W.H., Say, R.F., Zarzycki, J., Hugler, M., Alber, B.E., and Fuchs, G. (2010). Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* *8*, 447-460.

Darling, A.E., Jospin, G., Lowe, E., Matsen, F.A.t., Bik, H.M., and Eisen, J.A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* *2*, e243.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* *30*, 1575-1584.

Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* *3*, e1319.

Evans, P.N., Parks, D.H., Chadwick, G.L., Robbins, S.J., Orphan, V.J., Golding, S.D., and Tyson, G.W. (2015). Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* *350*, 434-438.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., *et al.* (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* *44*, D279-D285.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., and Stetter, K.O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* *417*, 63-67.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. *Comput Sci Eng* *9*, 90-95.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* *11*, 119-119.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* *28*, 27-30.

David Galambos

Functional metagenomics overview

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-U354.

Meier, D.V., Pjevac, P., Bach, W., Hourdez, S., Girguis, P.R., Vidoudez, C., Amann, R., and Meyerdierks, A. (2017). Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. *ISME J.*

Orcutt, B.N., Sylvan, J.B., Knab, N.J., and Edwards, K.J. (2011). Microbial Ecology of the Dark Ocean above, at, and below the Seafloor. *Microbiol Mol Biol R* 75, 361-+.

Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2010). IDBA - A Practical Iterative de Bruijn Graph De Novo Assembler. *Lect N Bioinformat* 6044, 426-440.

Reveillaud, J., Reddington, E., McDermott, J., Algar, C., Meyer, J.L., Sylva, S., Seewald, J., German, C.R., and Huber, J.A. (2016). Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environ Microbiol* 18, 1970-1987.

Stewart, F.J., Ulloa, O., and DeLong, E.F. (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* 14, 23-40.

Tully, B.J., Wheat, C.G., Glazer, B.T., and Huber, J.A. (2017a). A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. *ISME J.*

Tully, B.J., Wheat, C.G., Glazer, B.T., and Huber, J.A. (2017b). A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. *The ISME Journal*, ismej2017187.