Using 15 metagenomes and 10 metatranscriptomes from across the Mid-Cayman Rise, we obtained 73 high-quality MAGs with >70% completeness and <10% redundancy. Due to poor assembly characteristics such as low N50, sequence data from sample FS841 were excluded from the analysis (see Table S\*\*). **Add more things about cell counts, vent geochem, sequencing stats, etc). A couple supplementary tables?**

Metabolic Gene Abundance and Expression Across Samples

Our first aim was to ascertain the relative abundance of key metabolic genes at the community level across different vent sites. Metabolic genes were chosen as those that catalyze reactions involving key hydrothermal metabolites such as sulfur, oxygen, nitrogen, hydrogen, iron, and organic carbon (incl. methane) (Reveillaud et al., 2016). There were no notable differences between Von Damm and Piccard vent fields; vents in both locations displayed similar abundances of selected metabolic genes (Fig. 1). We also observed similar variation in abundances within sample sites. Although both vent fields have sample sites where some genes are highly abundant compared to others, genes are uniformly abundant within most sites.

In addition to community-level gene abundance, we also studied community-level gene expression. The gene expression data mirror gene abundance, as Von Damm and Piccard display similar expression profiles across the selected genes (Fig. 2). However, some selected differences begin to emerge. Compared to Piccard, Von Damm displays higher expression of some hydrogenases and the methanogenesis gene *mcr.* Meanwhile, the Piccard sites Hot Chimlet and Shrimp Canyon have extremely low relative expression levels of most of the metabolic genes studied.

MAG-resolved Expression Across Samples

In contrast to the similarities in gene expression and abundance between the two vent fields, Von Damm and Piccard display very different patterns in the expression of the 73 high-quality MAGs selected for this analysis. We performed a hierarchical clustering of the expression z-scores of each MAG. A MAG has an elevated z-score if its expression at that site is higher than the average for that MAG across all sites. In a hierarchical clustering of the average expression of each MAG at each sample, we identify three broad groups: Shrimp Hole; Ginger Castle and Von Damm; and all four Piccard sites (Fig. 3). Interestingly, despite being in Von Damm, Ginger Castle and Hot Cracks #2 are more closely related to all Piccard sites than to Shrimp Hole. There is also lower diversity in MAG expression patterns within Von Damm, as those four sites are more closely related than the Von Damm sites are to each other.

None of the MAGs display uniform expression across samples. In fact, for most MAGs, there is at least one sample where expression is ~1 z-score above average, with all other samples having below-average expression of that MAG. The MAGs that show elevated activity in one Piccard sample, such as *Sulfurovum\_99*, tend to have elevated activity in all of Piccard. Meanwhile, most MAGs that are active at a Von Damm site display elevated activity only at that site.

Most lineages with expected methanogenic or ANME capabilities such as *Methanomicrobia* and *Methanococci* are more active at Von Damm than at Piccard, with the notable exception of *Methanococci\_69*, which is active at Shrimp Canyon. There are also notable differences between Shrimp Hole 2012 and 2013. There is a repertoire of MAGs that are active at both time points, but Shrimp Hole 2013 hosts additional MAG activity that is not seen in 2012. It is important to note that these samples may differ not only in time, but spatially as well, since it is difficult to sample the same point of a hydrothermal vent non-consecutively.

Functional potential of MAGs on the KEGG pathway level

We assessed the functional potential of each MAG by defining a module completion score (MCS) for each MAG (0≤m≤1) that describes the functional capability of a MAG while taking into account that there may exist several different reactions for each step of a pathway, each of which is sufficient on its own (see Supplementary Material for code and a description of the algorithm). We selected 73 KEGG modules by searching for pathways with known microbial origin and relation to key metabolites described in Figs. 1, 2. These in include carbohydrate metabolism, carbon fixation, methanogenesis and aerobic methane oxidation, sulfur and nitrogen redox, and various membrane proteins such as cytochromes and Mn/Zn/Fe/S/N transporters. We also included several hemolysin transporters as a potential marker for Nanohaloarchaea ectosymbiosis as described in Anantharaman et al. (2016).

With this analysis, 5 distinct metabolic clusters emerged (Fig. 4). The first category consists of core carbon cycle metabolisms that are highly conserved in most MAGs with MCS>0.8. These include modules related to glycolysis, gluconeogenesis, the TCA cycle, and a generic pentose phosphate pathway. The next cluster can be broadly defined as related to methane metabolism. The most prominent lineages with capabilities in this cluster include *Archaeaglobi*, *Methanomicrobia*, and *Methanocci*. We observe 3 methanogenesis modules in the cluster along with pathways for F420 biosynthesis, coenzyme M biosynthesis, acetyl-CoA to CO2, and nitrogen fixation (see Supplementary Material for KEGG module accession numbers associated with each pathway). The third group consists of the variable genome where MAGs mostly show MCS≥0.9 or MCS≤0.3. This cluster reveals the presence of both thiosulfate oxidation and assimilatory nitrate reduction in several *Sulfurovum* and other Campylobacterial MAGs, along with the PTS sugar uptake systems used by those MAGs. The fourth group is highly conserved across MAGs, but most of the MCS values are only about 0.5. However, the Wood-Ljungdahl pathway, denitrification, dissimilatory nitrate reduction and dissimilatory sulfate reduction are present with high MCS in 2-6 MAGs each.

Questions

How is style so far?

How to compare to Julie’s paper?

Do I need more details about other things, or talk about things that are already there in more detail?

Combined heatmap: is the one on slack clustered by expression?

Goal: Have a nice results draft by 5th week, a nice discussion draft by 7th week. Look over methods 8th week. Revise all three by the end of the term.

Other:

How to read and remember papers in new field? Things I’ve seen so far are 1. Read early papers in the area 2. Look up the author’s photo 3. Take notes

References

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*.

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. Env. Microbiol *18*, 1970–1987.