Variability In Bacterial Biological Processes At the Tica Vent (9° N) On The Eastern Pacific Rise Using Peptide Analysis

Senior Thesis

School of Oceanography

University of Washington

Seattle WA 98195-5351

nkrutina@uw.edu

Abstract

The purpose of this study is to observe and determine how proteins vary near hydrothermal vents over a period of 9 months. Microbial communities found at hydrothermal vents thrive off the chemicals released from the plumes and hydrothermal fluid, using chemosynthesis to convert those chemicals into energy. The white smoker bacterial community located at the Tica vent (9°50.4 N, 104°17.5 W) on the East Pacific Rise, was observed for its protein variability over a 9-month period in 2004 between February and November. The dataset used in this study was acquired by Sievert et al. between 2004-2006 and compares the change of concentration of proteins over a 9-month period in 2004. Protein sequences were created from 16s rRNA gene analysis of biofilms on basalt panels located at the Tica vent site where a colony of tubeworms and mussels resided (experiment) and 2.5 m from the colony (control). Conversion of RNA data to peptide sequences allowed for evaluation of the biochemical processes occurring in the microbes living on the basalt. Over the course of the 9-month period, the samples became more microbially diverse in their proteins over time. The biological process GO-terms had an increase over the 9-month period with some GO terms having an inversely proportional trend to the GO terms that were increasing.

Plain Language Summary

My research focuses on the microbial diversity at volcanically active areas at the mid-ocean ridge known as hydrothermal vents. Microbes can thrive at hydrothermal vents because the vents release nutrient rich fluids and there are breakdowns of the rock at hydrothermal vents that microbes can use as energy. As they use those nutrients as energy, the microbes can undergo a series of biological processes and functions to maintain the microbe's life and build a larger and/or more diverse community. This research investigates a dataset that was taken throughout a

nine-month period at a hydrothermal vent, known as the Tica vent, and a site that was a few meters from the Tica vent at the East Pacific Rise (EPR). The dataset is comprised of a series of unique rRNA, which is a molecule that facilitates in the protein making to form ribosomes. Ribosomes are incredibly important to a cell because it helps with biological processes such as translation, move a messenger RNA molecule, and catalyze the construction of amino acids into protein. Each rRNA sequence from the data is converted into a unique code for a protein. That protein data is analyzed to determine what is going on with the microbial community at the hydrothermal vent and a couple meters away from the hydrothermal vent. I hypothesized that over a nine-month period, the community will become more diverse. With more diversity at the Tica vent, there will be a change in the biological processes. My results showed that there was in fact more diversity over the nine-month period at the Tica vent, but the trend of biological processes over a nine-month period were still unclear. Not knowing the biological process trends of the microbial communities may be due to the complexity of the communities. While there was a lack of outside sources about the function of biological processes in microbial communities, this research was the start to what is hoping, a new branch of research.

Introduction

Hydrothermal vents were discovered in 1977 (Corliss et al., 1979). The discovery unveiled an integrative world of geology, microbiology, and chemistry. Hydrothermal vents give insight into the beginning of life on Earth. Microorganisms in hydrothermal fluids reflect a subsurface biosphere (Deming, Baross,1993). Microbial communities live in extreme conditions, thriving at an upper-temperature limit of 115°C.

At and around hydrothermal vents, microbial communities can thrive from nutrients released from the vents. The microorganisms that live there are known as thermophile bacteria.

Thermophiles are classified by rRNA analysis as archaea, one of the branches for the earliest organisms to live on Earth (Zierenberg et al., 2000). The thermophile communities of focus are those that were observed thriving on basaltic rock. Basaltic rock has high permeability and allows for circulation of seawater where they undergo weathering and release solutes (Fisher, 1998). With an abundant source of potential energy from the release of different nutrients from hydrothermal vents and basaltic rock weathered down, life can thrive. Obtaining information on the microorganisms provides an evolutionary understanding of the origin of life.

The biological productivity depends on the chemical energy from the chemicals released at the hydrothermal vents (McCollom, 2000). Hydrothermal fluid is rich in silica, calcium, manganese, and iron, and microbes can undergo a series of metabolic reactions. Near hydrothermal vents, anaerobic processes dominate while aerobic processes can still occur (Martin et al., 2008). Microbes use CO_2, H_2, CH_4, S, and SO_4^{2-} (Martin et al., 2008) to undergo a reduction-oxidation reaction. Methanogenesis, sulfate reduction and Fe (II) oxidation are among the many redox reactions that occur. These microbes are referred to as chemoautotrophs because they use chemistry as energy to synthesize and survive. At the data collection site, sulfate reduction and nitrogen cycling are due to the microbial communities (Table 1). Zhang et al. concluded that microbial communities play an important role in nitrogen and sulfur cycling and metal metabolism at inactive hydrothermal vents (Table 1).

TABLE 2. Common redox reactions and associated standard free energies of reaction that occur in the dark ocean and can be exploited for metabolic energy

Pathway	Reaction	ΔG° (kJ/mol)[a]
Oxic respiration	$CH_2O + O_2 \rightarrow CO_2 + H_2O$	−770
Denitrification	$CH_2O + 4/5NO_3^- \rightarrow 1/5CO_2 + 2/5N_2 + 4/5HCO_3^- + 3/5H_2O$	−463
MnO_2 reduction	$CH_2O + 3CO_2 + H_2O + 2MnO_2 \rightarrow 2Mn^{2+} + 4HCO_3^-$	−557
Fe(III) oxide reduction	$CH_2O + 7CO_2 + 4Fe(OH)_3 \rightarrow 4Fe^{3+} + 8HCO_3^- + 3H_2O$	−697
Sulfate reduction	$CH_2O + 1/2SO_4^{2-} \rightarrow HCO_3^- + 1/2H_2S$	−98
Sulfate reduction (from methane)	$CH_4 + SO_4^{2-} \rightarrow HCO_3^- + HS^- + H_2O$	−33
Methanogenesis (from acetate)	$CH_3COOH \rightarrow CH_4 + CO_2$	−24
Methanogenesis (from H_2/CO_2)	$H_2 + 1/4HCO_3^- + 1/4H^+ \rightarrow 1/4CH_4 + 3/4H_2O$	−57
Fermentation (from ethanol)	$CH_3CH_2OH + H_2O \rightarrow CH_3COOH + 2H_2$	−181
Fermentation (from lactate)	$CH_3CH_2COO^- + 3H_2O \rightarrow CH_3COOH + HCO_3^- + 3H_2$	−1,075
Acetogenesis	$H_2 + 1/2CO_3^- + 1/4H^+ \rightarrow 1/4CH_3COO^- + H_2O$	−90
Hydrogen oxidation	$H_2 + 1/2O_2 \rightarrow H_2O$	−263
Methane oxidation	$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$	−859
Sulfide oxidation	$H_2S + 2O_2 \rightarrow SO_4^{2-} + 2H^+$	−750
	$H_2S + 8/5NO_3^- \rightarrow SO_4^{2-} + 4/5N_2 + 4/5H_2O + 2/5H^+$	−714
Fe(II) oxidation	$Fe^{2+} + 1/4O_2 + H^+ \rightarrow Fe^{3+} + 1/2H_2O$	−48
	$Fe^{2+} + 1/5NO_3^- + 6/5H^+ \rightarrow Fe^{3+} + 3/5H_2O + 1/10N_2$	−44
	$Fe^{2+} + MnO_2 + 2H^+ \rightarrow Fe^{3+} + MnO + H_2O$	ND
Mn(II) oxidation	$Mn^{2+} + O_2 \rightarrow MnO_2$	−149
	$Mn^{2+} + 2/5NO_3^- + 4/5H_2O \rightarrow MnO_2 + 1/5N_2 + 8/5H^+$	−79
Nitrification	$NH_4^+ + 2O_2 \rightarrow NO_3^- + 2H^+ + H_2O$	−302
Anammox	$NH_4^+ + NO_2^- \rightarrow N_2 + 2H_2O$	−345

[a] Values for ΔG° calculations were taken from reference 15. ND, not determined.

Table 1. Common reduction-oxidation reactions that microorganisms undergo around hydrothermal environments (Orcutt, 2011)

The data that was collected was at the Tica hydrothermal vent (9˚50.4 N, 104˚17.5 W). The Tica vent is a high temperature area and is located along the center of the axial summit trough (AST) on the eastern side of the AST. It is surrounded by collapsed structures of the AST and is approximately 20 x 10 m (Ferrini et al, 2007). This is an area of constant change due to the aggregation and destruction of sulfide chimneys but specifically at the Tica vent, there has been a build up of chimney structures on sheet lava surfaces that are adjacent to the primary fissure in the axial summit trough (AST). This vent is known to be dominated by Riftia tubeworms but the continuously diffuse flow in 2004 has changed the major organisms that dominate that vent, which is now mussels. The higher temperature environment also results in more sulfide deposits which was measured by K. L. Von Damm in 2004. The AST next to the Tica Vent is unique because it has a wide floor (45-60 m) with a 10 m wide fissure that extends 8 m down. The depth

difference between the Tica Vent and the AST is about 18 m, meaning that there is almost a vertical wall between the Tica Vent and the AST (Figure 1).

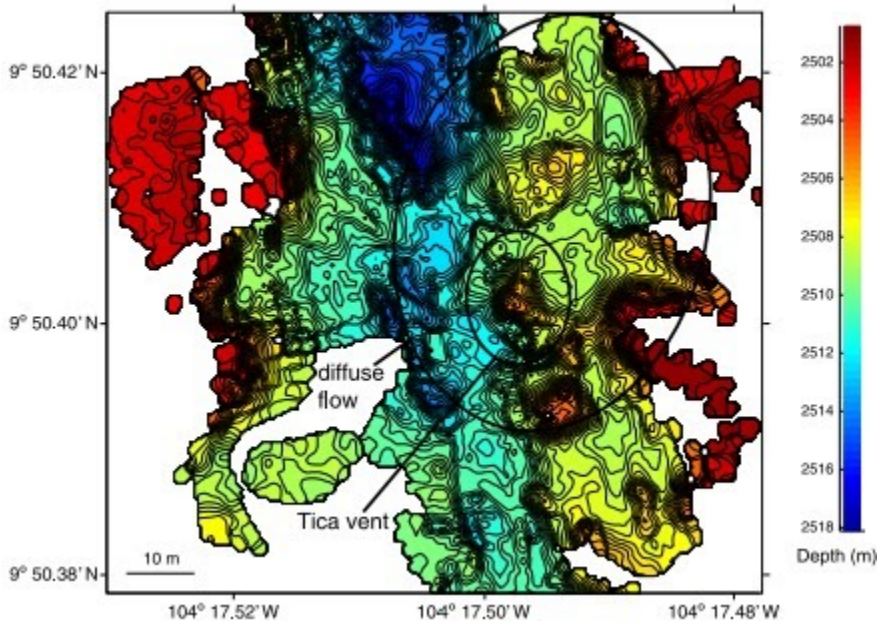


Figure 1. A composite map of the Tica vent and the Axial Summit Trough (AST) (Ferrini et al, 2007).

This dataset was processed as 16s rRNA sequences which is the microbial community composition of the samples. The data was collected off of various basalt panels during Sievert's research cruises AT11-07, AT11-10, AT11-20 and AT15-06 at the 9ºN deep-sea hydrothermal Tica vent on the East Pacific Rise, Pacific Ocean between Jan. 2004 and June 2006. The dataset contains links to full-length 16S rRNA sequence data from the NCBI GenBank with accession numbers KT257735-, KT257859, etc. The collected data from Sievert's basalt panel experiments have resulted in relatively few publications, none of which focus on changes in proteins over the course of the time allocated in the data collection. It is important to analyze this dataset because it gives a sense of protein functions over the course of nine months at the EPR.

While the 16S rRNA Sievert data is an effective method to determine the genus and its abundance at hydrothermal vents, proteomics give a more comprehensive look into the class level of organisms that reside at hydrothermal vents. It is important to investigate proteins rather than genes because proteins provide a more detailed look into how bacteria respond to different factors such as temperature and chemical composition in order to adapt to a particular environment. For example, in higher temperatures, organisms are put under stress, which changes the protein that is synthesized and may even result in degradation, while anaerobic environments may result in a reduction of energy metabolism proteins (Tomanek, 2011). Mikan reported that functional changes at a class level include 85% of the peptide evidence, while at the genus level, it would contain 53% of the peptide evidence (Mikan et al, 2019). Proteins are directly related to the molecular function of the microbe which will give information about the abundance of these proteins, biological processes, and characterization of the protein (Mikan et al, 2019).

The peptide sequences are matched up with Gene Ontology (GO) terms. GO terms are unique labels that help describe the biological knowledge in a community. These GO terms are important because they help people who are analyzing a dataset to understand the different processes that occur at a site, the location of the gene products that are involved, and describe the correlation between normal functions. These terms are used to describe the genes in any organism and to share that information between scientists and laypeople (Blake, 2001). There are three categories, which are common to all living organisms, that GO terms will fall into: Biological Process, molecular function, and cellular component. The biological processes are processes are necessary for an organism to live and occur by multiple molecular activities. Molecular functions describe the activities that happen at a molecular level but do not specify

where or when the action took place. Cellular components are the locations relative to the cellular structures where a gene product preforms a function. While the analyzation of GO terms can be complex, the focus of this research will be on the biological processes (Figure 2).

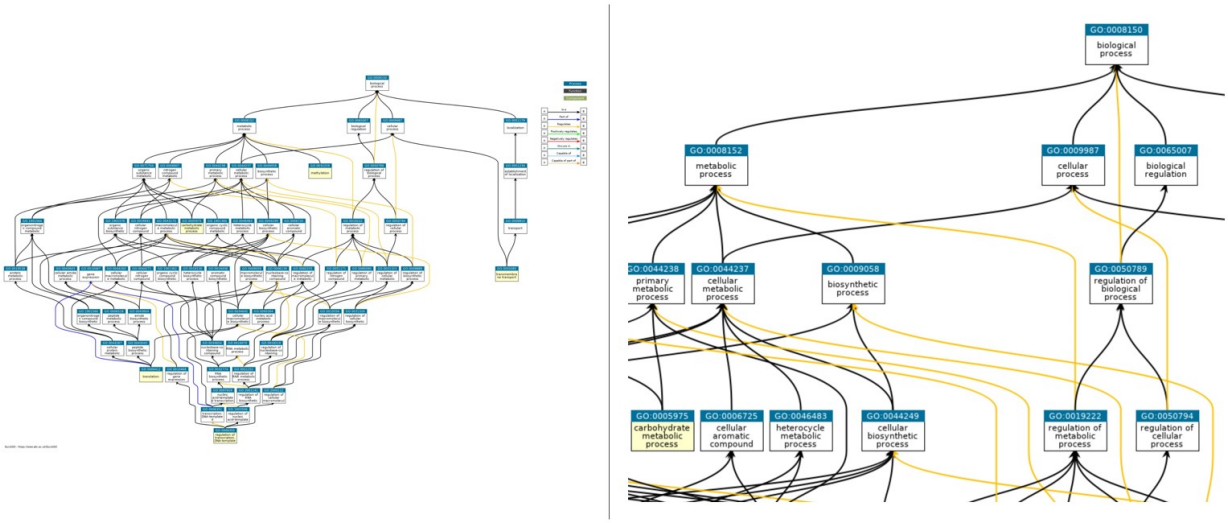


Figure 2 – A GO graph of the biological processes, where GO is roughly hierarchical. The 'child' terms are more specialized than the 'parent' terms. 'Child' terms can have multiple 'parent' terms. The biological process (at the top) is the parent term of all the terms below it. The black line means that the child term is a parent term. The yellow line means that the child term regulates the parent term. The blue line means that the child term is a part of the parent term. The left side shows the complexity of the biological process GO terms. The right side shows a zoomed in snapshot of the GO graph.

This project will aim to see how proteins diversify at hydrothermal vents. Determination of the proteins used by the microbial communities provide a more detailed look into how bacteria in an extreme environment respond to different conditions such as chemical composition and temperature. Over the course of 293 days at the Tica Vent on the East Pacific Rise (9˚50.4 N, 104˚17.5 W), the proteomic data will diversify and specific proteins that are utilized in biological processes will increase in concentration. The limitations of the research are the data that was lost on day 76 and the inconclusive data from day 4.

Methods

Stefan M. Sievert, a principal investigator, deployed basalt panels at the location known as Tica Vent on the EPR (9˚50.4 N, 104˚17.5 W with a depth of 2513 m) during a research cruise and gathered data during three different time periods throughout 2004. The purpose was to observe bacterial diversity on basalt surfaces at hydrothermal vents. There were two basalt panels deployed; the experimental site was located at a diffuse-flow vent site near an area where tubeworms and mussels resided while the control basalt panel was located 2.5 m from the experimental site, ½ m from the edge of the colonized vent. The experimental site's temperature was 15 °C while the control site was 1.8 °C. Both the experimental and control sites are considered pre-eruption samples since the samples were collected 19-28 months before the 2005 eruption at the EPR.

The basalt panels were 10.2 cm x 10.2 cm x 2.5 cm and were originally collected on the EPR. The panels were sealed in aluminum foil and sterilized using an autoclave. The panels were filtered through with 0.2 μm filtered seawater to prevent contamination of the surface seawater. The experimental and control basalt panels were exposed for 5 different time intervals: 4 days, 9 days, 13 days, 76 days, and 9 months (the control had a time interval of 283 days while the experimental had a time interval of 293 days). The basalt panel for day 76 was lost. The DNA was extracted from the basalt panels using a large volume CTAB extraction. The DNA was precipitated by using a 0.1 volume of 3M sodium acetate and 2.5 volume of cold 100% ethanol was added and placed into a closed container at -20C for 3 hours. After the three hours, the sample was centrifuged and then carefully isolating the supernatant DNA.

The DNA was amplified at the 16s rRNA region and combined with replicates of 10 PCR amplifications. The PCR product of the 16s rRNA gene was cloned for sequencing purposes where it was then analyzed to match with a sequence from the ARB software package. The data

was uploaded on the BCO-DMO database where a downloadable CSV file was accessible to the public (http://dmoserv3.bco-dmo.org/jg/serv/BCO-DMO/Microbe_Vent_Communities/16S_sequences.html1?sample_descrip%20eq%2016S_tags).

Each individual 16s rRNA sequence was converted to a protein sequence through the NCBI database. the 16S rRNA sequences found in the dataset were added into the GenBank database with accession numbers KT257735-KT257859. Once converted into a protein sequence, the EXPASY Peptide Cutter was able to process each protein sequence and cleaved each cleavage site using trypsin (in silico digest) to form a table of cleavage sites, sorted sequentially by amino acid number. The digested peptide sequences are then ready to go through a metaproteomic analysis using Unipept (http://unipept.ugent.be/).

Unipept processes the peptides and outputs various information such as taxonomic levels that were associated with the list of peptides and Gene Ontology (GO) annotations. The list of proteins can be converted into GO terms where these terms will capture the biological processes, molecular functions, and cellular components of the proteins. The GO terms are used to access gene product functional information, analyze the proteomic dataset, and get an overview of proteomics at the experimental site and the control site.

Results

Sievert's 16s rRNA gene data was converted into peptides using a series of software's. EXPASY Peptide Cutter converts the RNA sequence into a protein sequence. Once all the protein sequences have been outputted, the protein sequences are then processed through

UniPept metaproteomic analysis which outputs a list of peptides which is then used to build a treeview diagram and analyze GO terms.

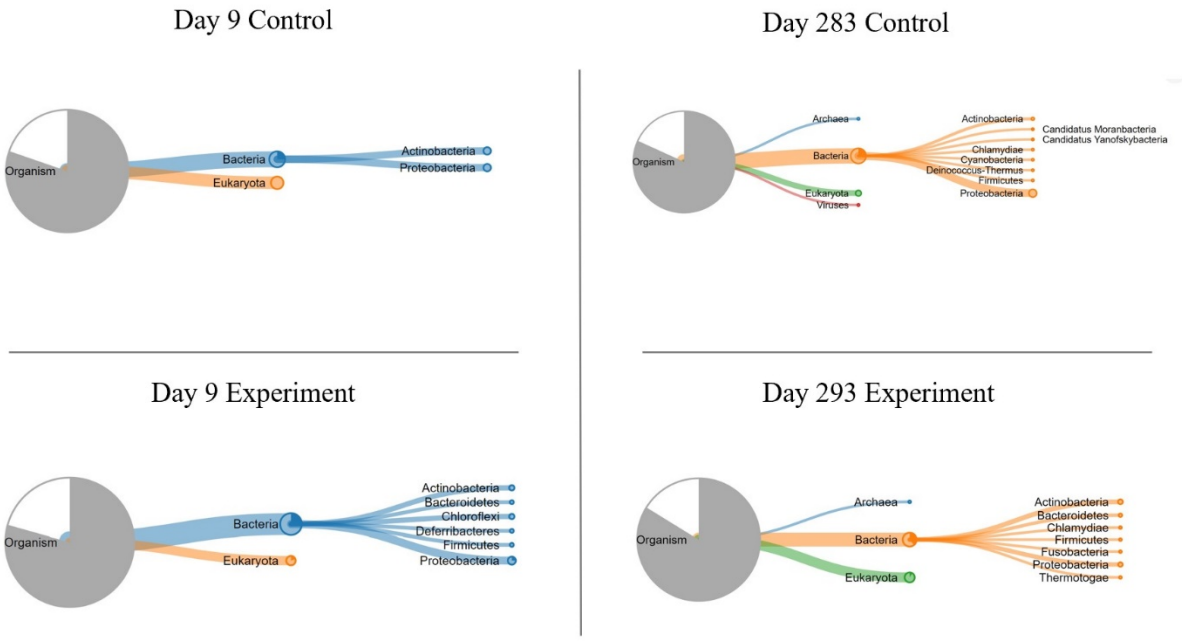


Figure 3. A set of tree view diagrams that link the peptide sequences to different taxonomic levels. Those in the largest circle are the number of peptide sequences that are specific to the organism level or lower. The next branch point shows the number of peptide sequences that are specific to the domain level or lower. The last branches are the number of peptide sequences that are specific to the phylum level or lower. The upper tree view diagrams are for the control site – 2.5 m from the experimental site – with day 9 on the left and day 283 on the right. The lower tree view diagrams are for the experimental site – at a diffuse flow vent site – with day 9 on the left and day 293 on the right.

The day nine experiment data matched 92 peptides out of the total 212 peptides while day 9 at the control site matched 71 peptides out of the total 187 total peptide sequences. All the peptide sequences in the tree view diagrams are linked to various taxonomic levels including super kingdom, kingdom, phylum, and class (Figure 2). Day 283 (control) had 187 out of 431 peptide sequences matched up at a taxonomic level. Day 293 (experiment) had 346 out of 891 peptide sequences match up to a specific taxonomic level. The relative contribution percent of

bacteria at the experiment site ranges between 9.54% and 15.22% (Appendix table 1) with day 9 having a relative percent contribution of 15.2% and day 293 with a relative percent contribution of 9.54%. The relative percent contribution of bacteria at the control site between day 9 and day 283 ranges between 11.27% and 13.37%. Peptide sequences related to eukaryotes ranged from 5.88% to 8.16% in the experiment site and 3.75% to 8.45% in the control site. Archaea and virus relative percent contribution were not included because the data was too limited for analysis.

Of the GO terms at the experimental and control site, there is a spike of the biological processes on day 4, a lower relative percent contribution on day 9, and will increase, decrease, or stay constant over both time periods (Figure 4, Figure 5). The GO terms that had the highest relative contribution at the experimental site were the carbohydrate metabolic process, regulation of transcription, and the transmembrane transport. A plot of the relative percent contribution of those GO terms at the control site was produced to compare how those biological processes vary at both sites (Figure 7).

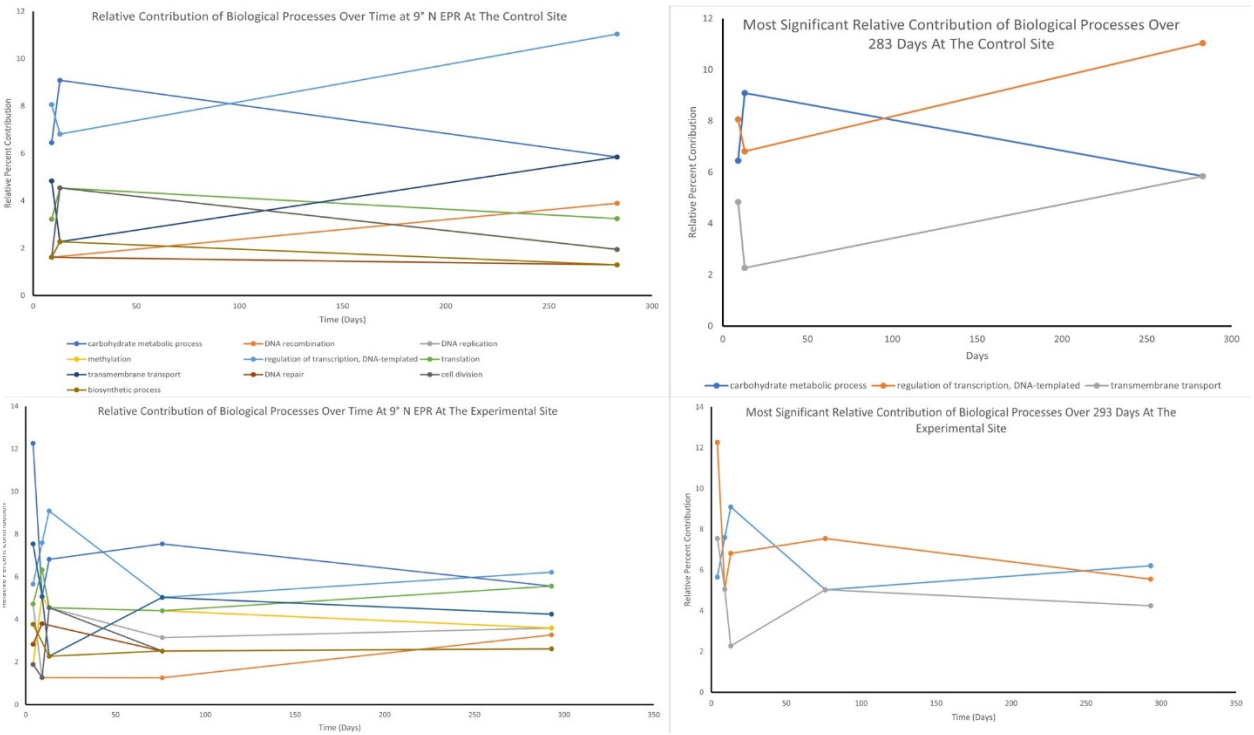


Figure 4. A plot of the relative contributions of the ten most abundant GO-terms for the biological process over the course of 293 days at the Tica vent. A plot of the relative contributions of the ten most abundant GO-terms for the biological process over the course of 283 days at the control site. The most significant relative contribution of the GO terms for the experimental site over 293 days. The most significant GO terms were the carbohydrate metabolic process, regulation of transcription, and transmembrane transport. The most significant relative contribution of the GO terms for the control site was plotted. The most significant GO terms were the carbohydrate metabolic process, regulation of transcription, and transmembrane transport.

At the Tica vent (the experimental site) the carbohydrate metabolic process increases until day 76, where it reaches a relative percent contribution of 5.03% but then increases to 6.21% at day 293. The regulation of transcription is initially 12.26%, reaches a low of 5.06% on day 9, and steadily increases to 7.54% by day 76. The transmembrane transport at the Tica vent ranges from 7.55% on day 4 to 4.25% on day 293. The carbohydrate metabolic process at the Tica vent overall increases over the period while the regulation of transcription and transmembrane transport decreases over time. At the control site, the carbohydrate metabolic process is 6.45% on day 4, 9.09% on day 13, and 5.84% on day 283. The regulation of transcription is 8.06% on day 4, 6.81% on day 13, and 11.03% on day 283. Lastly, the

transmembrane transport is 4.84% on day 4, 2.28% on day 13, and 5.84% on day 283. At the control site, the carbohydrate metabolic process decreases over the period while the regulation of transcription and the transmembrane transport increase over time.

Discussion

While the RNA data gave an assessment of the taxonomic diversity of the community at a deep-sea vent, by converting the RNA into proteins, one can get a complex look of the functions and processes of an active community. The limitation of using peptide sequences rather than RNA generalizes the results but it allows for the proteins to be identified and evaluated for their gene ontology. It is useful to see the different taxonomic levels associated with the peptides because not only does it show the composition of the sample, but it shows what peptides are common enough in the sample that they are identified in the second taxonomic rank, kingdom. This information means that all the peptides that were broken down into each taxonomic level tells us whether the peptides are so common that it is seen within the entire kingdom level or that the peptide is so specific that it is only seen for one species. Using peptide sequences allows one to observe biological processes, molecular function, and cellular components. By converting the RNA into peptides sequences, the Keil lab was able to provide extensive knowledge on how to evaluate peptides and protein sequences.

Over the course of 9 months, the tree view diagrams show more diversity on the basalt panels that were there for the 9-month period than the basalt panels that were there for 9 days (Figure 3). The 9-month experiment was slightly less diverse than the control. The first few weeks show many similarities of organisms between the Tica vent and the control site. The similarity in diversity may be due the fact that organisms were transported from the vent site by currents and deposited at the control site (Gulman et al., 2015). The general trend is as time

increases, so does the diversity of the microbial communities at each site. The diversity shows in the tree view of the control site and the experimental site by having peptide sequences that matched up with similar taxonomic levels such as proteobacteria, actinobacteria, and bacteroidetes. The highest relative percent contribution was the proteobacteria then actinobacteria. The proteobacteria are essential to the nitrogen, carbon, and sulfur cycling at hydrothermal vents so one could expect to see a higher concentration of proteobacteria in the data (Campbell, B., et al, 2006). Those bacteria in the proteobacteria phyla mainly comprised of epsilonproteobacteria, a class that undergo reduction of sulfur and hydrogen and oxidize these compounds, were more dominant during the beginning of the time but over the nine months, the diversity of the peptide sequences increased drastically (Sievert et al., 2015). The peptide sequences at the control site on day 283, although more diverse, were much lower in concentration and vastly different than any of the other basalt panels on the other days for control and experiment. Some of the peptide sequences that matched up to different taxonomic levels in the 283-day control site that were not present in the experiment site were the following bacteria: Candidatus Moranbacteria, Candidatus Yanofskybacteria, and Deinococcus-Thermus. While the Candidatus phyla is still understudied, the Deiococcus-Thermus is known as a chemoheterotroph that thick cell wall giving them a gram-positive stain, but the second membrane resembles a gram-negative stain (Steinsbu et al., 2011). It is worth noting that the epsilonproteobacteria and bacteroidetes dominate around inactive hydrothermal though, this research did not investigate the dominating bacteria at active hydrothermal vents (Zhang et al., 2016).

At both the experimental site, the highest relative abundance GO terms were the carbohydrate metabolic process, the regulation of transcription, and the transmembrane transport. The carbohydrate metabolic process is a series of pathways and chemical reactions that produces

energy to cells by providing the cell with carbohydrates. The regulation of transcription controls the rate of DNA-templated transcription. The transmembrane transport is the process where a solute is transported across a lipid bilayer, from one side of a membrane to the other (not including the nuclear membrane). Day 4 has a wide range of various organisms that were present at the experiment site, with less diversity at the control site. By day 4, one can observe that the GO terms facilitate in the initial rapid rate of higher density communities at the experimental site, and to a lesser extent at the control site (Figure 4). Such processes that would facilitate in rapid rate of a high-density microbial community are regulation of transcription, translation, carbohydrate metabolic process, and transmembrane transport. At the experimental site, the carbohydrate metabolic process increases while the regulation of transcription and the transmembrane transport decrease. At the control site, the carbohydrate metabolic process decreases while the regulation of transcription and the transmembrane transport both increases. The control site has the opposite effect from the experimental site. Both the experimental and the control site show that the carbohydrate metabolic process is directly, inversely proportional to the transmembrane transport.

The hypothesis was confirmed; the proteins diversified, and certain proteins would become more significant over time. The ability of the microbial community to grow at a rapid rate in the first couple days leads more diversity on the basalt panel, due to the organic matter present where other groups could thrive (Gulman, 2015). Well established biofilm enables a wide range of conditions for different bacteria, archaea, eukaryotes, and even viruses to grow and live (Davey, O'toole, 2000). The limitation to these results lies in the GO terms – the trend of all the biological processes is unclear and one cannot confidently say that there is an obvious trend. There are still not enough literature values quantifying the GO terms of microbial

communities at hydrothermal vents. New questions to be considered are, how do certain GO terms affect other GO terms in microbial communities and what trends may lie in all the GO term domains.

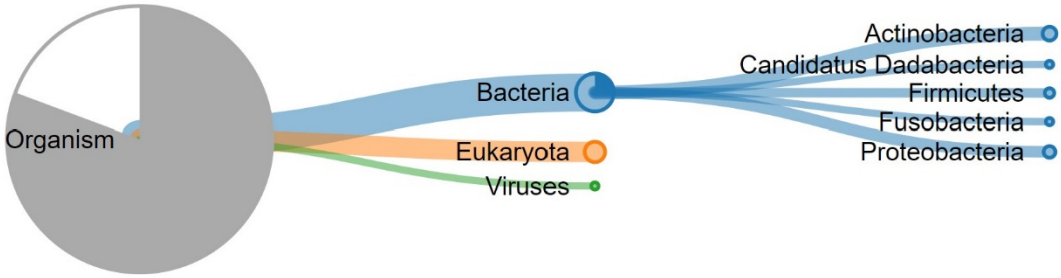
Conclusions

The variance in diversity of proteins is important to determine how a microbial community functions and uses their energy to thrive at hydrothermal vents. While there was a lot of research done on the analysis of 16S rRNA at hydrothermal vents, little was known about the change over time in specific processes and quantifying those processes. Over the 9-month period in 2004, there was an increase in diversity of proteins at the Tica Vent on the East Pacific Rise while there was no clear trend of all the biological process GO terms. This indicates that a microbial community becomes more complex over time at hydrothermal vents. While the findings of the change of proteins was clear, the trend of the GO terms is still uncertain. A potential follow up study would be to quantify the change of GO terms over time at hydrothermal vents to determine what important biochemical processes occur there.

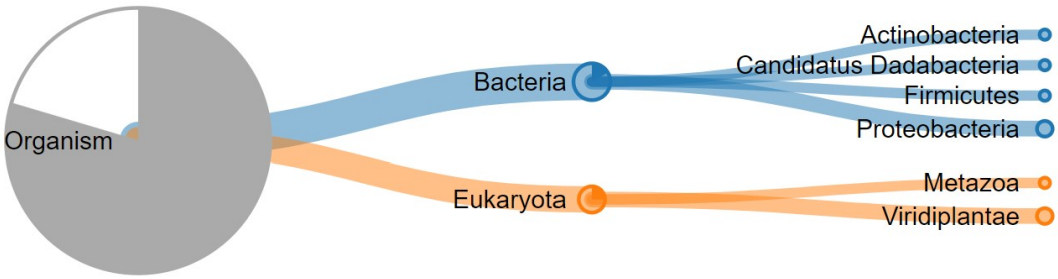
Acknowledgements

Dedicated to my cat and my grandparents. Thank you to the wonderful Keil Lab. I could not have done this without the immense support and help from Jaqui Neibauer and Rick Keil. Thanks to the Sievert lab for their work and for posting their data to BCO-DMO so the public could access it and use it for future research.

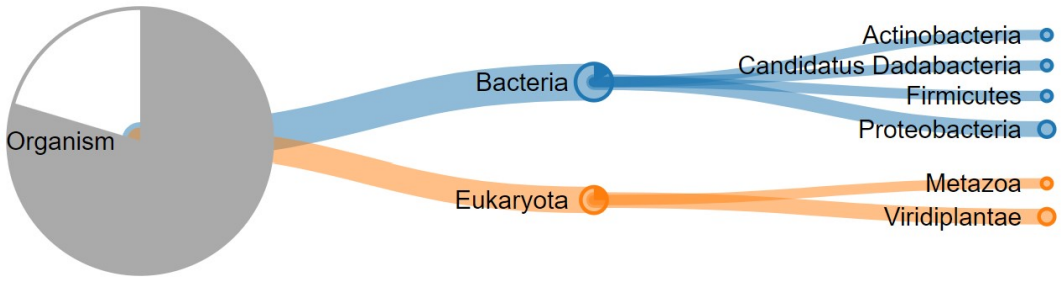
Appendix



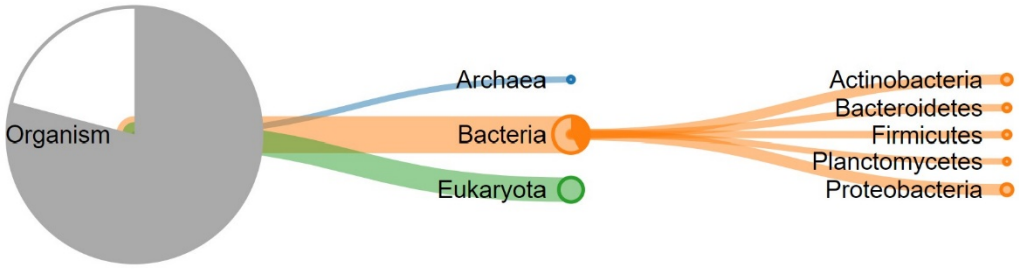
Appendix Figure 1 - A set of tree view diagrams that link the peptide sequences to different taxonomic levels. Those in the largest circle are the number of peptide sequences that are specific to the organism level or lower. The next branch point shows the number of peptide sequences that are specific to the domain level or lower. The last branches are the number of peptide sequences that are specific to the phylum level or lower. This tree view corresponds to the day 4 basalt panel from the Tica vent (experimental site).



Appendix Figure 2 – A set of tree view diagrams that link the peptide sequences to different taxonomic levels. Those in the largest circle are the number of peptide sequences that are specific to the organism level or lower. The next branch point shows the number of peptide sequences that are specific to the domain level or lower. The last branches are the number of peptide sequences that are specific to the phylum level or lower. This tree view corresponds to the day 13 basalt panel from the control site.



Appendix Figure 3 - A set of tree view diagrams that link the peptide sequences to different taxonomic levels. Those in the largest circle are the number of peptide sequences that are specific to the organism level or lower. The next branch point shows the number of peptide sequences that are specific to the domain level or lower. The last branches are the number of peptide sequences that are specific to the phylum level or lower. This tree view corresponds to the day 13 basalt panel from the Tica vent (experimental site).



Appendix Figure 4 - A set of tree view diagrams that link the peptide sequences to different taxonomic levels. Those in the largest circle are the number of peptide sequences that are specific to the organism level or lower. The next branch point shows the number of peptide sequences that are specific to the domain level or lower. The last branches are the number of peptide sequences that are specific to the phylum level or lower. This tree view corresponds to the day 76 basalt panel from the Tica vent (experimental site).

Control Site			
Day	**9**	**13**	**283**
Total peptides	187	109	431
Organism	71	49	187
Archaea	0	0	1
Viruses	0	0	1
Eukayota	6	4	7
relative contribution (%)	8.450704225	8.163265306	3.743315508
Bacteria	8	6	25
relative contribution (%)	11.26760563	12.24489796	13.36898396
Actinobacteria	3	1	2
Candidatus Dadabacteria	0	1	0
Candidatus Moranbacteria	0	0	1
Candidatus Yanofskybacteria	0	0	1
Chlamydiae	0	0	1
Cyanobacteria	0	0	1
Deinococcus-Thermus	0	0	1
Firmicutes	0	1	1
Proteobacteria	3	2	11

Appendix Table 1 – The data sets from the control site converted from 16s rRNA to peptide sequences. The first row is the total peptides that were able to be matched in a database. The number of peptides for each domain level is determined: Archaea, Viruses, Eukaryota, and Bacteria. The number of peptides for the phylum level of only the bacteria is determined since the other phyla of the different domains are limited. The peptide sequences of the eukaryote and bacteria domains are then calculated to determine its relative contribution at a taxonomic level.

Experimental Site					
Day	**4**	**9**	**13**	**76**	**293**
Total peptides	297	212	109	464	891
Organism	119	92	49	191	346
Archaea	0	0	0	1	1
Viruses	1	0	0	0	0
Eukayota	7	5	4	15	22
relative contribution (%)	5.882352941	5.434782609	8.163265306	7.853403141	6.358381503
Bacteria	15	14	6	24	33

relative contribution (%)	12.60504202	15.2173913	12.24489796	12.56544503	9.537572254
Actinobacteria	4	2	1	4	7
Bacteroidetes	0	1	0	2	3
Chloroflexi	0	2	0	0	0
Deferribacteres	0	1	0	0	0
Candidatus Dadabacteria	1	0	1	0	0
Chlamydiae	0	0	0	0	1
Firmicutes	2	1	1	2	1
Fusobacteria	1	0	0	0	1
Planctomycetes	0	0	0	1	0
Proteobacteria	3	4	2	5	6
Thermotogae	0	0	0	0	1

Appendix Table 2 – The data sets from the experimental site converted from 16s rRNA to peptide sequences. The first row is the total peptides that were able to be matched in a database. The number of peptides for each domain level is determined: Archaea, Viruses, Eukaryota, and Bacteria. The number of peptides for the phylum level of only the bacteria is determined since the other phyla of the different domains are limited. The peptide sequences of the eukaryote and bacteria domains are then calculated to determine its relative contribution at a taxonomic level.

References

Blake, J. 2001. Creating the Gene Ontology Resource: Design and Implementation. Genome Research.

Campbell, B., Engel, A., Porter, M, Takai, K. 2006. The versatile ε-proteobacteria: key players in sulphidic habitats. Nat Rev Microbiol **4**: 4458–468. Doi: 10.1038/nrmicro1414

Corliss, John & Dymond, J.G. & Gordon, Louis & Edmond, John & Herzen, Richard & Ballard, Robert & Green, Kenneth & Williams, D. & Bainbridge, Arnold & Crane, Kathy & van Andel, Tj. 1979. Submarine Thermal Springs on the Galapagos Rift. Science (New York, N.Y.) **203**: 1073-83. Doi: 10.1126/science.203.4385.1073.

Davey, M.E., O'toole, G.A. 2000. Microbial biofilms: from ecology to molecular genetics. Microbiol Mol Biol Rev. **64**(4): 847-67. Doi: 10.1128/mmbr.64.4.847-867.

Deming, Jody W., Baross, John A. 1993. Deep-sea smokers: Windows to a subsurface biosphere? Geochimica et Cosmochimica Acta **57**: 3219-3230.

Fuchida, Shigeshi, Mizuno, Yuki, Masuda, Harue, Toki, Tomohiro, Makita, Hiroko. 2014. Concentrations and distributions of amino acids in black and white smoker fluids at temperatures over 200°C. Organic Geochemistry **66**: 98-106. Doi: 10.1016/j.orggeochem.2013.11.008.

Gulmann Lara, Beaulieu Stace, Shank Timothy, Ding Kang, Seyfried William, Sievert Stefan. 2015. Bacterial diversity and successional patterns during biofilm formation on freshly exposed basalt surfaces at diffuse-flow deep-sea vents. Frontiers in Microbiology **6**: 901. Doi: 10.3389/fmicb.2015.00901

Haymon, Rachel M., Kastner, Miriam. 1981. Hot spring deposits on the East Pacific Rise at 21°N: preliminary description of mineralogy and genesis. Earth and Planetary Science Letters **53**: Issue 3. Doi: 10.1016/0012-821X(81)90041-8.

Karson, J. A., Klein, E. M., Hurst, S. D., Lee, C. E., Rivizzigno, P. A., Curewitz, D., Morris, A. R., and Hess Deep '99 Scientific Party .2002. Structure of uppermost fast-spread oceanic crust exposed at the Hess Deep Rift: Implications for subaxial processes at the East Pacific Rise, Geochem. Geophys. Geosyst. **3**: 1002. Doi:10.1029/2001GC000155.

Karson, J.A., Kelley, D.S., Fornari, D.J., Perfit, M.R., and T.M.Shank. 2015. Discovering the Deep: A Photographic Atlas of the Seafloor and Ocean Crust. Cambridge University Press. ISBN:978113905024.

Kelley, D.S., Baross, J.A., Delaney, J.R. 2002. Volcanoes, Fluids, and Life At The Mid-Ocean Ridge Spreading Centers. Annu. Rev. Earth Planet. Sci. **30**: 385-491 Doi: 10.1146/annurev.earth.30.091201.141331

Kormas, K.A., Tivey, M.K., Von Damm, K, Teske, A. 2006. Bacterial and archaeal phylotypes associated with distinct mineralogical layers of a white smoker spire from a deep-sea hydrothermal vent site (9°N, East Pacific Rise).

Martin, W., Baross, J., Kelley, D., and M.J. Russell. 2008. Hydrothermal vents and the origin of life. Nat Rev Microbio. **6**: 805-814.

McCollom, Thomas M. 2000. Geochemical constraints on primary productivity in submarine hydrothermal vent plumes. Deep Sea Research Part I: Oceanographic Research Papers **47**, Issue 1: 85-101. Doi: 10.1016/S0967-0637(99)00048-5.

Mikan, M.P., Harvey, H.R., Timmins-Schiffman, E., Riffle, M., May, D.H., Salter, I., Noble, W.S., Nunn, B.L. 2019. Metaproteomics reveal that rapid perturbations in organic matter prioritize functional restructuring over taxonomy in western Arctic Ocean microbiomes. ISME J **14**: 39–52. Doi: 10.1038/s41396-019-0503-z

Mincheol Kim, Jongsik Chun. 2014. Chapter 4 - 16S rRNA Gene-Based Identification of Bacteria and Archaea using the EzTaxon Server. Methods in Microbiology. Doi: 10.1016/bs.mim.2014.08.001.

Orcutt, Beth N., Sylvan, Jason B., Knab, Nina J., and Edwards, Katrina J. 2011. Microbial Ecology of the Dark Ocean above, at, and below the Seafloor. American Society for Microbiology. Doi:10.1128/MMBR.00039-10.

Romain, Causse & Biscoito, Manuel & Patrick, Briand. 2005. First record of the deep-sea eel Ilyophis saldanhai (Synaphobranchidae, Anguilliformes) from the Pacific Ocean. CYBIUM (0399-0974) (SFI) **29**, Issue 4: 13-416. 29.

Schulz, H. D., & Zabel, M. 2006. Marine geochemistry (2nd ed.). Berlin: Springer. doi:10.1007/3-540-32144-6

Steinsbu, Bjorn O., Tindall, Brian J., Torsbik, Vigdis L., Thorseth, Ingunn H., Daae, Frida L., Pedersen, Rolf B. 2011. Rhabdothermus arcticus gen. nov., sp. nov., a member of the family Thermaceae isolated from a hydrothermal vent chimney in the Soria Moria vent field on the Arctic Mid-Ocean Ridge. International Journal of Systematic and Evolutionary Microbiology. Doi: 10.1099/ijs.0.027839-0

Tomanek L. 2011. Environmental proteomics: changes in the proteome of marine organisms in response to environmental stress, pollutants, infection, symbiosis, and development. Ann Rev Mar Sci. **3**: 373-99. doi: 10.1146/annurev-marine-120709-142729

Webber, Alexander P., Roberts, Stephen, Murton, Bramley J., Mills, Rachel A., Hodgkinson, Matthew R. S. 2017. The formation of gold-rich seafloor sulfide deposits: Evidence from the Beebe hydrothermal vent field, Cayman Trough. Geochemistry, Geophysics, Geosystems June 2017 **18**, Issue 6: 2011-2027. Doi: 10.1002/2017GC006922

Zhang, L., Kang, M., Xu, J., Shuai, Y., Zhou, X., Yang, Z., Ma, K. 2016 Bacterial and archaeal communities in the deep-sea sediments of inactive hydrothermal vents in the Southwest India Ridge. Sci Rep **6**, Article Number: 25982. Doi: 10.1038/srep25982

Zierenberg, Robert A., Adams, Michael W. W., Arp, Alissa J. 2000. Life in extreme environments: Hydrothermal vents. Proceedings of the National Academy of Sciences Nov 2000, **97** (24): 12961-12962; Doi: 10.1073/pnas.210395997