QUANTIFICATION, ISOLATION, CHARACTERIZATION, AND BIOGEOGRAPHY OF THERMOACIDOPHILIC BACTERIA FROM MESOPHILIC SOILS

By

Margaux Karp

A Thesis Presented to

The Faculty of California State Polytechnic University, Humboldt

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Biology

Committee Membership

Dr. Mark Wilson, Committee Chair

Dr. Catalina Cuellar-Gempeler, Committee Member

Dr. Susan, Marshall, Committee Member

Dr. Karen Kiemnec-Tyburczy, Committee Member

Dr. Paul Bourdeau, Graduate Coordinator

May 2022

ABSTRACT

QUANTIFICATION, ISOLATION, CHARACTERIZATION, AND BIOGEOGRAPHY OF THERMOACIDOPHILIC BACTERIA FROM MESOPHILIC SOILS

Margaux Karp

Many thermoacidophilic bacterial species thrive in hot springs, fumaroles, and geothermally heated soils. These habitats are globally distributed but often small in size and located distantly from similar habitats, and therefore the biogeography of thermoacidophiles may be similar to island biogeography of plants and animals. Little is known about dispersal of thermoacidophiles. In this study I aimed to quantify, isolate, and characterize thermoacidophilic bacteria from non-permissive habitats. I also used genetic approaches to compare these isolates to thermoacidophilic bacteria from permissive habitats to provide insight to their biogeography.

To determine if I could enrich thermoacidophiles from non-permissive habitats, and if so, what taxa could be enriched, ten soil samples were incubated at five different temperatures between 30°C and 70°C in pH 2.5 PTYG medium. Thermoacidophiles were enriched from all ten soil samples, and at all temperatures from most samples. Forty-two isolates were obtained with at least 12 different species identified. Nine potentially novel species were isolated, sharing less than 97% 16S rRNA sequence identity with characterized isolates.

To determine what media would enrich the most thermoacidophiles, Most Probable Number (MPN) trials were carried out with three different media (PTYG, K medium, and Potato Dextrose medium [PD]) at pH 3 and 60°C. Five soil samples from the non-permissive Trinity Alps location produced growth up to at least the 10^{-2} dilution in all media. The PTYG medium produced the highest concentrations. Seven of the ten isolates from these experiments were novel species based on 16S rRNA gene analysis.

To quantify thermoacidophiles in non-permissive environments, twelve soil samples from each of six different locations were incubated at 60°C, pH 3 in PTYG medium in a MPN trial (one replicate by three ten-fold dilutions). Growth occurred from every location in at least eight of the 12 samples. The concentration of thermoacidophilic endospores was between 13-325/g. From the 72 samples, 50 isolates were collected with the most prevalent being *A. pomorum* (22) and *A. acidocaldarius* (8). Seventeen isolates were novel species based on 16S rRNA gene analysis.

To explore biogeographical distributions of *A. acidocaldarius* in non-permissive environments, I used comparative multilocus sequence typing (MLST) of six different highly conserved genes (*cpn60*, *eftu*, *emrB*, *gmp*, *gyrB*, and *rpoB*). These genes were PCR-amplified and sequenced from 29 *A. acidocaldarius* isolates from non-permissive locations. I found that isolates from different non-permissive locations had identical or nearly identical sequences. These isolates were compared to isolates from permissive habitats using MLST and individual gene trees which showed isolates from permissive and non-permissive habitats clustering together, with some isolates from these opposing habitats sharing identical or near identical alleles. Our findings support evidence for

broad, on-going dispersal of spores to the permissive and non-permissive regions sampled.

Surprisingly, thirty-three isolates failed to meet the 97% criterion for alignment of their 16S rRNA gene in BLASTn. Many of these isolates also did not share 97% identity with each other, suggesting that multiple potentially novel species were cultured from non-permissive habitats. Because so many studies have looked at thermoacidophiles in permissive environments, it seems unlikely that these novel species were not previously detected because of undersampling, and instead it suggests that non-permissive habitats may host a variety of novel thermoacidophiles. The source environments for these novel thermoacidophiles is unclear.

Collectively, these results suggest that massive numbers of thermoacidophilic spores are broadly dispersed in short timeframes across large scales and accumulate at high concentrations in diverse habitats. These spores come from diverse species, many of which have not yet been isolated or characterized.

ACKNOWLEDGEMENTS

I thank Dr. Mark Wilson for guidance and support, the committee members Dr. Catalina Cuellar-Gempeler, Dr. Susan, Marshall, and Dr. Karen Kiemnec-Tyburczy, Graduate student Terilyn Stoflet, the students in Bacteriology Lab—Spring 2020, Dr. Yoon Kim, Dr. Mark Rizzardi, and Dr. Oscar Mauricio Vargas.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICIES

CHAPTER 1: PROJECT GOALS

In a previous study, isolates of the genus *Alicyclobacillus* were obtained from both Hawai'i Volcanoes National Park (HVNP) and Lassen Volcanic National Park (LVNP) and the genotypes of *Alicyclobacillus tolerans* and *A. acidocaldarius* were analyzed using a multilocus sequence typing (MLST) approach [1]. The resulting MLST pattern was that *A. acidocaldarius* isolates from each site had unique genotypes and the genotypes found in HVNP were not found in LVNP. However, for *A. tolerans,* identical genotypes were found in isolates across all sites, including HVNP and LVNP [1]. Thus, there was endemism and biogeographical differentiation in one species of *Alicyclobacillus,* but another species of the same genus did not show this pattern. The *A. tolerans* results suggested that ongoing, rapid dispersal is taking place across thousands of km, at rates that are sufficient to erase traces of allopatric differentiation. The *A. acidocaldarius* results, however, suggest that either dispersal is very different for this species, or that dispersal is not always sufficient at displacing allopatric differentiation. For example, there may be periods where drift and selection create biogeographically unique communities, but periodic selective sweeps may erase that pattern of endemism. Because dispersal mechanisms are likely to be non-directed, i.e., spores will be dispersed from habitable sites to an array of uninhabitable sites, uninhabitable sites may have spores that represent pre- and post-sweep communities. Additionally, because spores persist for immense periods of times, non-permissive sites (sites where environmental conditions do not allow for vegetative growth) may contain collections of spores from

rare or unique lineages. Very few studies have focused on spores in uninhabitable environments [2]–[4]; my research was focused on spores of hyperthermophiles, specifically *A. tolerans* and *A. acidocaldarius*, from non-permissive environments.

In this study I aimed to investigate the abundance, taxonomy, and distribution of thermoacidophilic spores in habitats that are not permissible for growth. In addition, I aimed to assess the biogeography of *A. acidocaldarius* within non-permissive habitats and between non-permissive and permissive habitats by comparing allele patterns in an MLST analysis. To achieve this, first, I evaluated temperature and media to determine optimal spore recovery conditions. I developed a Most Probable Number (MPN) approach to quantify the spore concentrations in soil samples and applied it to six different soils from non-permissive habitats in Humboldt and Trinity counties in northern California. Isolates from the temperature and media experiments were obtained by single colony isolation on plates, and 16S rRNA genes of isolates were PCR-amplified and sequenced. Because *A. acidocaldarius* and *A. tolerans* had been used in the prior Maghfiroh study, I tried to identify enrichment conditions that targeted these species [1].

The *A. acidocaldarius* strains isolated from my experiments were used to create a MLST phylogeny of six highly conserved genes. The *A. acidocaldarius* samples isolated from non-permissive sites were then compared to the *A. acidocaldarius* isolates found by Maghfiroh (2015) in the permissive sites of LVNP and HVNP. Individual gene trees and MLST analysis were used to assess the diversity of the *Alicyclobacillus* species within and between these non-permissive and permissive sites. I hypothesized that shared alleles

will be present across permissive and non-permissive habitats suggesting long-range, ubiquitous dispersal within the regions sampled. Seeing no shared alleles between habitats would suggest that there is no shared dispersal between these sites.

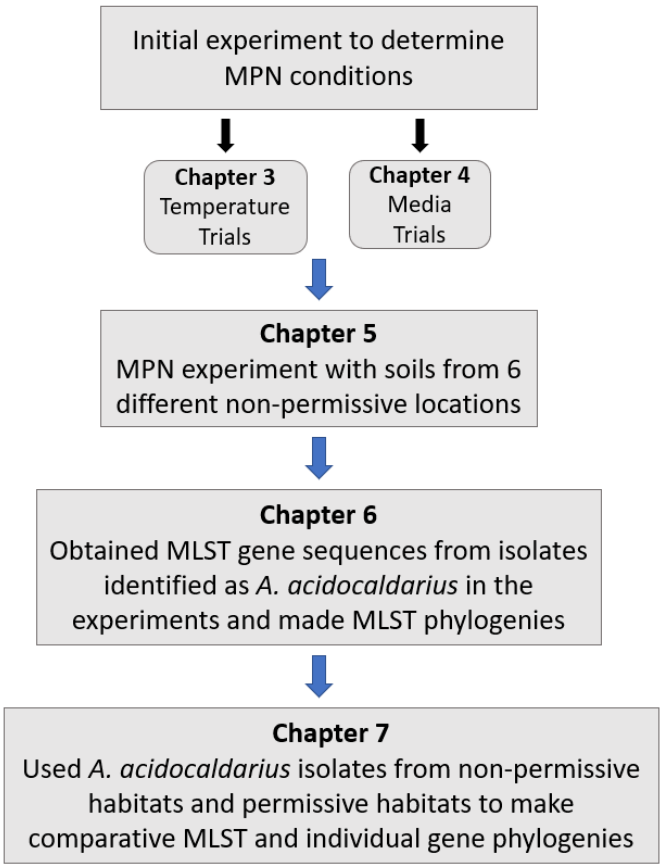

Figure 1. Flow chart of project tasks.

CHAPTER 2: LITERATURE REVIEW

Biogeography

Biogeography, the distribution of species throughout time and space, has been studied extensively in macroscopic species. Many of the resulting biogeographical theories have been widely assumed to affect microbes but limited research has been produced to validate these inferences. There is some evidence to suggest that biogeography and biodiversity are unified by similar mechanisms and patterns across all living systems [5], [6]. For example, microbes that are able to tolerate a range of environmental factors and disperse effectively should comprise wider ranges. One theory of microbial dispersal, Baas Becking, hypothesizes that "everything is everywhere but the environment selects" [7]. Rather than geographic limitations, this theory suggests that environmental factors are the key determinants in affecting spatial patterns of microbial biodiversity. However, several studies have found evidence for dispersal limitations of microbes representing a more complex system at play that likely includes geographical parameters [4], [8], [9]. A focus on multiple spatial scales in understanding microbial biogeography has been emphasized by Martiny et al. [10]. When looking across spatial distributions of microbes, past environmental and historical factors, such as glaciation [11], are thought to influence the present species biogeography. Some of the most important factors shaping and driving microbial biogeography are spatial variations and environmental heterogeneity. Spatial and environmental heterogeneity broadly include

the differences in geographical and ecological factors that can influence microbial biodiversity across locations.

Certain traits and characteristics of microbes, such as small sizes, vast populations, and rapid reproduction rates make microbial biogeographical patterns difficult to study [12]. While the existence of microbes was discovered over 300 years ago, our knowledge about bacteria is very limited in some respects, including our understanding of the dispersal and persistence of different bacteria. Many approaches have been introduced in order to culture and categorize bacteria and archaea but it is estimated that less than one percent of the microbes in existence have been identified and recorded [13]. With the rise of cheap and rapid sequencing, better insight is available into taxa assembly. Some researchers have called for studies to include more and more determining characteristics that define a microbe including traits, such as size, color, host-range, or motility [14] in order to provide increased certainty in defining microbial species. While our abilities to analyze microbes has improved drastically with advancements to technology, recent articles have suggested there is still much more to learn [15]. For example, uncharacterized cell structures and features were identified by electron cryotomography recently—study of which could illuminate further understanding of cell movement and function [16]. A better understanding of range area and biogeography of bacteria can help find and relate traits that may affect microbial dispersal.

Although traits that affect microbial dispersal have not been well researched, Choudoir et al. [17] took dust samples across the U.S., comparing phenotypic and

genotypic characteristics of bacterial and archaeal taxa that may influence range shape and size. They found, surprisingly, while somewhat unintuitive, found that within obligate aerobes, non-spore forming, pigmented, and gram-positive taxa had larger range sizes. In addition, they found that taxonomic identity, phenotypic traits, genomic attributes, and habitat preferences were key traits in determining range size of these more than 74,000 microbial taxa [17]. In similarity with macrobes, a majority of the diverse microbial taxa analyzed had small ranges, while few taxa had large ranges [17]. The pattern normally seen with macrobes is that organisms that are locally abundant, or higher in local density, often also have larger geographic ranges. Unlike macrobes, the authors of this study found that local density was not a good predictor of range size for these microbial taxa [17]. These findings suggest that not all patterns of range and dispersal seen with plants and animals may also be seen with microbes. However, it is possible some of the patterns may have been unique to the sampling and analysis methods such as the potential bias against spores due to their durability through DNA extractions. While insightful regarding characteristics of range sizes, this is still only a small piece of the puzzle when it comes to microbial dispersals, with many more environments left to be considered and studied. Research that focuses on relating descriptive taxa characteristics with dispersal will help expand our understanding of the microbial world.

There have been several studies of microbial diversity and the factors that govern microbial biogeography. Andam et al. [12] found a latitudinal diversity gradient in terrestrial *Streptomyces,* which they interpreted as a consequence of historical glaciation.

However, given the robust dispersal abilities of *Streptomyces*, and that glaciers receded from the area studied over 10,000 years ago, alternate forces may be responsible for the biogeographical distribution they described. Other studies have focused on the dispersal mechanism of thermophilic endospores located in cold bodies of water where growth would not occur consistently [2], [3]. Dispersal mechanisms such as currents, run-off, and industrial effluent were found to influence endospores in aqueous environments [2].

Much research on spore dispersal has been within the realm of plant pathogens due to the economic impacts they can inflict. The influence of environmental factors, including wind currents, in distributing plant pathogen spores such as coffee leaf rust, *Hemileia vastatrix*, or sugarcane rust, *Puccinia melanocephala*, spores throughout the globe have been investigated following the spread of these rusts from Africa to other continents [18], [19]. Geographically distinct regions of omnipresent fungal representation have been discovered demonstrating the potential for wide spore distributions of other species through environmental factors [19]. The role in dispersal on smaller scales like rain splash has even been investigated. Kim et al. [20] found that while minor dispersal capabilities of *Puccinia triticina* spores was shown from raindrops hitting an infected leaf, the spiraling air forces created from certain drops were found to have a larger impact potential on dispersal to other plants due to the greater heights these caused spores to go.

Aerial dispersal of bacterial has been widely documented. Diverse microbiomes have also been characterized in atmospheric samples as well as in samples of hail, snow, and rain. Bauer et al. [21] found the concentration of bacteria in cloud, aerosol, and snow

samples ranged from approximately $3.1x10^3$ to $2.0x10^4$ cell/ml. The range of fungal spores in these samples were found to be $6.2x10^2$ to $4.5x10^3$ spores/ml with the highest concentration found in cloud water samples [21]. Other studies have found that the concentration of microbes produced by aerosolization from sea spray have higher concentrations than the ocean body itself, a complex process that likely influences coastal dispersals [22], [23]. Variable diversity of aerosolized microbes are present throughout different land masses and marine regions [24]. Some researchers have found a greater diversity of microbes present in the air over land compared to over coastal areas, in addition to fluctuations throughout the time of day [25]. In general, there is a greater diversity but lower concentration of aerosolized bacteria in rural areas when compared to urban [24], [26], [27]. However, it is unclear how factors such as concentration or diversity would influence the dispersal of spores throughout an environment. To help understand the dispersal of microbes, some researchers have utilized a tracing method of species. For example, known soil- or leaf-associated bacteria discovered in something like an air sample, can be utilized to assess where, or what environment a species likely originated from [24], [28]. Challenges in method standardization such as with air sampling or culturing, pose difficulties for researchers as they may unknowingly affect the findings [24].

Island biogeography of plant and animal species has been heavily researched. Normally, the larger the area of land, the higher expected species diversity, but some islands comprise a disproportionate diversity to size [29]. Other factors such as distance from continental land mass influence the interactions between immigration and

emigration rates of an island and are thought to be the central factors of influencing island diversity [30], [31]. Island biogeography and the factors thought to influence island diversity may also apply to microbes from specific habitats such as hot springs and the thermoacidophilic bacteria that inhabit them. These hot spring 'islands' of permissive habitats may mimic similar patterns of island biogeography seen with plant and animals [8]. Certain hot springs may have vast distances between any other hot springs or comparable geothermal habitats meaning a lower chance of dispersal interactions. Like islands, the immigration and emigration rates of microbes may play key roles in the community composition of a geothermal region.

As extreme weather events increase across the globe due to climate change, researchers have begun to observe the effects this may have upon dispersal as a force of driving microbial diversity. Evans et al. [32] manipulated drought conditions, with filtered and unfiltered 'rain' upon soil, finding that aerial microbial dispersal rates can affect the community's response to stress. Models for extinction debt (the expected number of species in a population to likely become extinct), suggest that large areas for especially endemic macroscopic species will become unsuitable as a result of climate change, decreasing range sizes of already limited species [33]. While plants and animals may be affected more harshly by the climate change, microbes which have a faster generation time, may aid in the persistence of the macroscopic species by adapting faster to these pressures especially with plants and their associated microbiomes [34]. In addition, weather shifts will likely impact the persistence and dispersal of bacterial pathogens by supplying a larger permissive range, such as pathogens like *Escherichia*

coli and *Salmonella* sp. [35]. When applying models to predict future effects of microbial biodiversity and range to global heating it is important to include factors like aerial dispersal rates, as these shifting variables can affect the outcome of communities and dispersal.

Microhabitats

Soil is comprised of organisms and materials that includes a large biodiversity of microbes. The concurrent processes of feeding, degradation, and interacting introduces hurdles into separating or identifying components of this complex, such as microbial species [36]. While mesophilic environments may be non-permissive overall for the growth of species like *A. acidocaldarius*, microhabitats may be capable of providing habitable environments that allow for vegetative growth. Microenvironments are especially diverse in soils, which contain a variety of materials which can influence the types of organisms that grow in the various regions [37]. Minerals, particle charge, pore size, water, and oxygen availability all determine limiting factors for the microbial communities. Cells in soil matter can fit and grow in extremely small spaces that provide ample nutrients for growth. The specific physicochemical properties needed for growth may be met in microhabitats causing spores to produce vegetative cells. Vegetative cells are involved with ecological factors that can alter the genetic makeup of the cells or colonies. For this reason, microhabitats may produce a greater diversity of species in habitats that may be thought of as generally uninhabitable for thermoacidophilic species.

Evidence for permissible growth conditions of *Alicyclobacillus* species in compost environments have been found [38]. The breakdown of materials in composts provides high temperatures (from metabolic heat) and low pH (from acidic products of fermentation) for vegetative growth of *Alicyclobacillus* [39]. Habitats such as compost complicate the hypothesis that spores in non-permissive soil environments exclusively result from dispersal events.

Endospores and *Alicyclobacillus* Characteristics

Assessing microbial biogeography poses certain challenges. Endospores are produced by certain species of *Bacillus*, *Sporosarcina*, *Desulfotomaculum*, and *Heliobacterium*, for example, to facilitate survival [40]. However, the presence of endospores in an environment does not imply that the organisms are active in that environment because inactive spores may have dispersed to that site and persisted. For a particular study, biogeographical range may include habitable environments or both habitable and uninhabitable environments [41]. Studying endospores in non-permissive environments can provide valuable insight into biogeography, by helping to quantify the range and extent of spore dispersal.

The *Alicyclobacillus* genus is in the family *Alicyclobacillaceae* [42]. These obligately aerobic species thrive in high heat and low pH conditions. *Alicyclobacillus* are moderate thermoacidophiles that flourish in a variety of temperatures ranging from about 40 to 80°C [43]. Extreme thermophiles can grow above 80°C [44]. While some thermophiles can grow outside these ranges, an organism can only be considered a

thermophile if their growth spans encompass the moderate or extreme temperature ranges. Thermoacidophiles possess specialized traits that allow them to persist in such environments. Alicyclic fatty acids can be found in the cell walls of *Alicyclobacillus* sp. and are thought to play a role in heat and acid tolerance [44]. The ω- alicyclic fatty acids are thought to lower diffusion rates at higher temperatures, making the organism less susceptible to mechanistic problems associated with increased heat [45]. Some researchers are skeptical of this explanation because ω- alicyclic fatty acid can be found in other species that do not exhibit this heat or acid resistance [46], [47]. Other studies have suggested hopanoids, having a structure similar to cholesterol, are causing heat resistance in these bacteria [48]. Different *Alicyclobacillus* species tolerate different pH and temperature ranges. For example, *A. tolerans* prefers an optimal pH of 2.5-2.7 and temperature range 37-42°C [49] while *A. acidocaldarius* has an optimal range of 55-60°C at pH 4.5 [42]. Both are strictly aerobic, gram-positive species [42], [49].

In recent years, the *Alicyclobacillus* genus has been researched for the impact these species have had in the beverage industry. Fruit juices provide acidic environments and pasteurization involves incubation at high temperatures. Introduction of *Alicyclobacillus* species in juice typically occurs from soils which can then work its way down the production line, causing further contamination. If introduced during processing, these bacteria can alter the taste and cause economic losses [50]. Spores of this genus have been recorded surviving past the typical pasteurization of juice temperature treatments of 86-96°C for up to two minutes [51]. One of the main species responsible for spoilage is *A. acidoterristris* [44]. However, other *Alicyclobacillus* species such as *A.*

tolerans, *A. sendaiensis*, *A. pomorum*, *A. herbarius*, *A. hesperidum*, *A. disulfidoocidans*, and *A. vulcanalis* have been isolated from a range of habitats including soils, fruit juices, wastewater, dried hibiscus flowers, ores, and hot springs with much of the research surrounding these species consisting of characterization and isolation [42], [49], [52]–[57]. In addition, many thermoacidophiles have an important role in bioleaching (using microbes to help extract metals from minerals) due to the physiological strategies they obtain to survive in high heat and low pH conditions [58]. Researchers have looked at the potential to use specific genes from species like *A. acidocaldarius*, *Acidithiobacillus ferrooxidans*, and *A. caldus* to engineer cells for more optimal bioleaching [58].

Alicyclobacillus can be found readily in their vegetative state in hot spring-like habitats, but suitable habitats are often widely separated by inhospitable habitats. To combat this, *Alicyclobacillus* produces endospores that are able to survive adverse conditions when environments change or during dispersal. The production of dormant endospores by thermophilic and acidophilic bacteria, such as *Alicyclobacillus*, provides a unique advantage for studying dispersal patterns and biogeography. Bacterial endospores can survive at length under conditions that don't allow for vegetative growth, [40] which can create spore banks, or areas in which spores have compiled while lying dormant after dispersal. The species in these spore banks can be readily recovered by selective enrichment, and thus it is possible to obtain a record of the spores that dispersed over time to a given site. Dispersal limitations, or the lack thereof, might be assessed by studying the spatial distributions of spores and the associated spore-forming bacteria [4]. In addition, spore banks remain unaffected by many factors that drive evolution in

species. In non-permissive habitats, where spores exist, snapshots into the species diversity of the habitable environments where they were dispersed from can be observed.

Spore banks have been used in recent studies to investigate the dispersal methods of microbes upon aquatic environments [4], [59]. Wörmer et al. found that endospores become increasingly numerically dominant over vegetative cells as depth increased in subsurface marine sediments, because the concentration of vegetative cells declined with depth [59]. Evidence for some dispersal limitations were found in a study of isolated endospores from global seafloor sediment samples [4]. While the study found areas of shared abundancies of *Firmicutes* phylotypes (defined by shared similarity of 97% or greater) across large geographic distances, areas of low similarities, especially in the Artic regions were also identified [4]. These findings help demonstrate the influence that ocean currents and current connections have upon the biogeography of endospores, as the patterns seen support non-random dispersals [4]. However, few other studies have focused on dispersal of species using endospores from non-permissive habitats, thus little information is known.

Mutations, genetic drift, and selection provide mechanisms for diversity to be introduced into a microbial population [41]. Evolutionary factors that affect vegetative organisms, such as genetic drift and selection, are not able to affect dormant endospores. Genetic drift is caused by random factors that influence a population's gene frequencies over time [60]. Selection involves an increase in frequency over time of organisms that are better fit for an environment [61]. Within active populations, these elements can limit or increase diversity. When endospores are dispersed to non-permissive habitats, those

spores are essentially windows into the genetic variations in the organism they were produced by at the time of formation. Not undergoing the same genetic changes that would be seen in a growing population of vegetative cells, spore banks provide a unique system that can give insight to the biogeography of microbes.

Identification and Genetic Relation Analysis

In order to determine what taxa an isolate belongs to, housekeeping genes are sequenced and compared to those of characterized species. The 16S rRNA gene is currently the most utilized gene to determine bacterial identity and relations. However, many studies have focused on additional or alternate genes as a means of phylogenetic determinants, many which emphasize the limitation of only 16S rRNA analysis [62]–[69]. Housekeeping gene sequences other than the 16S rRNA gene are often used to compare subspecies microbial relationships. In most cases, housekeeping genes are highly conserved within species due to the fact that without them proper function is usually inhibited.

While qPCR is commonly used to calculate concentrations of an organism in a sample, this method is limited-especially when it comes to analyzing specific spores. For example, *Alicyclobacillus* spores are difficult to quantify using qPCR because the durable spores must be lysed without degradation of the DNA. A cultivation-based alternative to qPCR quantification is the MPN technique. Using MPN trials involve diluting environmental samples in permissive medium and incubating those samples at permissive temperatures. While there are assumptions about the ability to convert spores to

vegetative cells that can affect the accuracy of MPN trials, they may provide a more accurate method of acquiring concentration of viable spores for *Alicyclobacillus* species than qPCR [70].

The MLST analysis is a common method which uses a variety of (often housekeeping) genes to assess the relationships of gene sequences between isolates [71]. A shift to use MLST as a way to assess these genetic relations came about due to the fact that bacteria do not necessarily comprise clonal population structures [72], [73]. Portions of the genome can be altered by factors such as mutations or genetic transfers, meaning analyzing multiple areas along the chromosome is beneficial to determine more accurate phylogenies and subspecies identifications [72].

Phylogenies allows for visualization of evolutionary relationships based on genetic or even physical characteristics [74]. There are many methods capable of producing a phylogenetic tree including neighbor joining, unweighted pair group method with arithmetic mean, maximum parsimony, maximum likelihood, and Bayesian inference. Bayesian inference and maximum likelihood are two of the more common methods currently used for molecular phylogenies [75]–[77]. Bayesian inference is based on posterior probabilities, or the probability that the hypothesis is true based on the evidence provided [78]. This method can rely on previous data and complex analyses of evolution into the model—however, the prior data used within these models may introduce inaccuracies [78]. The maximum likelihood method is often used with sequencing data to create the tree with the highest probability by comparing the probability of each individual nucleotide position throughout the alignments [79]. While

some have argued gaps may pose a bias within maximum likelihood trees [80], others have determined that this method is sufficient in accounting for gaps in sequences [81]. Similar to the Bayesian inference method, maximum likelihood is dependent on the model of evolution used for the calculations but tends to perform well with violations made in these assumptions [82]. While a setback in using maximum likelihood is the long processing time, programs, over the years, are developing faster computation times for rapid maximum likelihood trees or phylogenetic analysis [82]–[86].

Species Abundance and Novel Species

While databases and comparative sequencing tools have made identifying isolates and determining relations to other species easier than ever, these databases only contain a fraction of the estimated 10^7 to 10^9 bacterial species worldwide [87], [88]. Currently, only around 30,000 species have been isolated into pure cultures and provided formal names [89]. Many researchers are studying isolates from extreme environments to try to find and culture novel species that make useful enzymes or compounds such as secondary metabolites. With the rise in antibiotic resistance, the search for novel species that may produce new antibiotics is prevalent. One study, focused on high salinity and high pH environments, found *Actinomycetes* strains with the potential to produce unique antibiotics [90]. As our technology and culturing methods improve, more novel species are being discovered and described. Especially given these improvements in culturing, extreme environments may provide many unstudied species.

Both thermophilic and acidophilic species have been sought out for their unique compounds and physiology. The strategies that extremophilic species use to combat harsh conditions often produce very heat- or acid-stable molecules. One of the most well-known compounds is the DNA polymerase from *Thermus aquaticus*, which is widely utilized in PCR [91]. Acidophilic species have long been utilized in bioleaching and often produce acid stable enzymes that can be used in a variety of processes including baking, fuel cells, and pharmaceuticals [92].

Horizontal Gene Transfer

Horizontal gene transfer (HGT) seems to occur readily between bacterial species and has even been found to occur across kingdoms in geothermal springs [93]. The most common mechanisms for HGT include conjugation, transformation, and transduction [94]. However, other gene transfer methods exist in some species such as through gene transfer agents or artificially induced methods [94]–[96]. Conjugation occurs using cell to cell contact via adhesins or pili to transfer DNA, encoded by plasmid genes or integrative and conjugative elements in the host genome [94], [97], [98]. Transformation is a method of genetic recombination where competent cells uptake extracellular DNA [94]. While species can be naturally able to transform [99], other are considered incompetent meaning under normal circumstances they cannot uptake extracellular DNA. However, it is possible to induce transformation artificially usually by increasing the permeability of the cell [100]. Bacteriophages can transfer genes to host cells in transduction [94]. Gene transfer agents are created by the host cell. These agents, similar to bacteriophages, can

transfer genetic content [94], [101]. HGT plays a large role in antibiotic resistance of cells and is a main focus of study for this process. For example, a population of *Helicobacter pylori* was found to pass an antibiotic resistance gene to other members of the population even in the absence of the antibiotic [102]. While few studies have been produced to study conjugation within *Alicyclobacillus* species, plasmids have been identified in *Alicyclobacillus* genomes on the National Center for Biotechnology Information (NCBI) database along with genes associated with conjugation present providing evidence of genetic transfer [1]. While HGT can provide ecological advantages, it introduces barriers to identifying isolates as it can integrate genes or fractions of genes into other species. For this reason, it is important to understand HGT when studying genetics as it proves a major way that species can evolve to new or even future environmental conditions or stressors that may be unexplained by tree—like approaches [103].

Evidence of HGT can be observed through various analyses which primarily involve phylogenies or composition comparisons [103], [104]. Genetic data can be utilized to observe patterns of HGT throughout a population or community. For example, a strain of *Neisseria meningitidis* was determined to have a gene originated from *Staphylococcus* plasmids using molecular phylogenetic trees and a horizontal transfer donor index [104]. A compositional based study consists of comparisons between compositional structures such as G+C content or codon usage biases [103], [105], [106]. Compositional studies seem to be performed using genomes of species whereas phylogenetics may more often use portions of the genome or genes. In procedures such as

MLST, genes that are more likely to vertically transfer are studied to lessen the impact of interspecific gene flow on phylogenetic analyses.

CHAPTER 3: ENRICHMENT TEMPERATURE EFFECTS ON TAXONOMIC IDENTITY OF ISOLATES

Introduction

Thermoacidophilic bacteria optimally grow in low pH conditions and high temperatures ranging from 40°C to greater than 80°C [43], [44]. Thermoacidophilic bacteria include many different genera, including *Alicyclobacillus*, *Clostridium*, and *Bacillus*, among others. *Alicyclobacillus* species, while primarily thriving in acidic and higher temperature environments, range in optimal growth preferences depending on the species or subspecies. *Alicyclobacillus tolerans*, while having documented optimal growth of 37-42°C has been cultured from higher temperature environments of up to 55°C [44]. *Alicyclobacillus acidocaldarius* grows optimally at 55-60°C but has been isolated from 70°C environments [44].

In this study, I aimed to isolate *A. acidocaldarius* and *A. tolerans* because an MLST analysis of hot spring isolates from these species had been previously performed [1]. I enriched soil samples in pH 2.5 medium at temperatures of spanning 10°C increments from 30°C to 70°C to determine the ability to enrich thermoacidophiles and specifically the ability to enrich *A. tolerans* and *A. acidocaldarius*. Enrichments were plated and streaked for single colony isolation, and the 16S rRNA gene of isolates was sequenced to determine if there was an optimal enrichment temperature to produce *A. tolerans* or *A. acidocaldarius*.

Materials and Methods

Ten soil samples were collected from a Spruce forest (SF) in Loleta, California. Samples were collected approximately two m away from each other at a depth of five cm. Approximately 20 g of each soil sample were collected using aseptic techniques and stored overnight at 4°C until inoculated in enrichment medium.

Culturing conditions to enrich for thermoacidophiles

Trials were run to determine how temperature affects which taxonomic groups are enriched. I incubated one g of each of ten soil samples in foil-covered 500 ml conical flasks with 100 ml of pH 2.5 PTYG medium. Replicate flasks were incubated at 30°C, 40°C, 50°C, 60°C, and 70°C without shaking. The broth for 30°C and 40°C temperature incubation contained 50 μg/ml of cycloheximide to inhibit fungal growth. Single colony isolates were obtained by streaking from flasks onto pH 3 PTYG plates and restreaking until putative pure cultures were obtained. Gelrite was used for all plates incubated at 50°C, 60°C, or 70°C to allow for solidification during incubation at these higher temperatures. Plates incubated at 30°C and 40°C were solidified with agar and included 50 μg/ml of cycloheximide to inhibit fungal growth. The plates were incubated at the respective temperatures their enrichments were initially incubated.

Genetic identification of thermoacidophiles

Colony PCR was performed on any putative pure colonies that grew to amplify 16S rRNA genes in order to determine the identity of the organisms grown. All colony

PCR was prepped in a PCR hood that had been UV light treated for 15 minutes and all workspace and materials wiped down with ethanol prior to the procedure.

Thirty-five µl PCR reactions were produced for each isolate. For each reaction 17.2 µl of sterile water, 17.5 µl of Lucigen EconoTaq Plus 2x Mastermix, 0.15 µl of each primer, 8F(5'-AGAGTTTGATCCTGGCTCAG-3') and 1525R (5'-AAGGAGGTGATCCAGCC-3') were used. The water was first added into each PCR tube. The needle was then sterilized in a Bacti-cinerator, and one single colony was touched with the needle and transferred into the respective PCR tube with the needle being resterilized in between samples. After the water and cells were added to the tubes the mastermix and primers were combined and mixed and 17.8 µl of was distributed into each PCR tube. The tubes were capped and put in the Thermal Cycler (MJ Research PTC-100 Thermal Cycler or Applied Biosystem - 2720 Thermal Cycler PCR). The cycling conditions consisted of 95°C for four minutes, 94°C for one minute, 57 °C for one minute (annealing temperature for 16S rRNA gene), 72°C for one minute, repeating steps two through four for 30 to 35 cycles, seven minutes at 72 °C, and 24+ hours of 4°C.

PCR products were then purified using Wizard SV Gel and PCR Clean-Up System (Promega) protocol to facilitate removal of extraneous components such as nucleotides, primers, buffers, and enzymes. After cleaning the PCR products, samples were analyzed via electrophoresis on 1% agarose gels to ensure the presence of appropriately sized PCR products.

A Nanodrop spectrophotometer was used to determine DNA concentrations. If concentrations were high enough (approximately 20 µg/µl or greater), they were diluted

for sequencing based on template size. Sequences were prepped following the manufacturer's protocol using SimpleSeq Premixed Kit tubes and shipped overnight to Eurofins Genomics (Louisville, KY).

16s rRNA sequence analysis

Returned sequencing files were cleaned meaning they were screened for quality including trimming the ends or editing nucleotides of low quality using CodonCode Aligner (Version 10.0.1). Where applicable forward and reverse sequence products were aligned and used to make contigs.

Sequences were identified using Basic Local Alignment Search Tool for nucleotides (BLASTn) by the National Center for Biotechnology Information (NCBI). Isolates with greater than 97% identity were considered to be the same species as the BLAST match.

The R package "lme4" was used to perform the Generalized Linear Model (GLM). "Growth" in this analysis was manipulated into a binomial variable where "1" denoted the identification of *A. acidocaldarius* and "0" denoted the identification of another species. N/A was used to represent the samples where no strains were able to be isolated meaning that the outcome of *A. acidocaldarius* growth could not be determined.

Phylogenies were created using the 16S rRNA sequences that were aligned using Multiple Alignment using Fast Fourier Transform (MAFFT) under the default settings (https://www.ebi.ac.uk/Tools/msa/mafft/) and clipped using AliView (Version 1.27). Maximum likelihood trees with rapid bootstrapping were produced using the Cyberinfrastructure for Phylogenetic Research (CIPRES) interface with the RAxML-

HPC2 (Version 8.2.12) task on Extreme Science and Engineering Discovery Environment (XSEDE) using all other default parameters [86]. The 16S rRNA gene sequence from a *Clostridioides difficile* isolate (AF072473.1), a distantly related thermoacidophilic species, was used to root the phylogeny.

Results

Each of ten soil specimens were incubated at five temperatures. From the 50 samples, 42 isolates were obtained. Seven of the ten 70°C treatments, and one of the 60°C treatments, had no growth. Twelve different characterized species were identified, and many isolates may belong to additional uncharacterized species (Table 1). Thirteen isolates of *A. acidocaldarius* were identified. Four were from 50°C treatments, seven were from 60°C, and two were from 70°C. Three *Burkholderia terricola* isolates were identified, all at 30°C. Five isolates were identified as *Alicyclobacillus acidoterrestris*, between 30-50°C. Three isolates were identified as *Alicyclobacillus contaminans*, all at 50°C. Two isolates of *Alicyclobacillus pomorum* and two isolates of *Paraburkholderia phenazinium* were identified. One isolate was identified for each of the following species: *Alicyclobacillus cyclohetanicus, Alicyclobacillus fastidiosus, Alicyclobacillus herbarius, Bacillus coagulans*, and *Burkholderia sediminicola*. No *A. tolerans* isolates were obtained.

Table 1. Basic Local Alignment Search Tool for nucleotides (BLASTn) return results of isolates obtained from Spruce forest (SF) soil by enrichment at temperatures 30°C, 40°C, 50°C, 60°C, and 70°C using PTYG medium.

Sample ID	Enrichment Temperature (°C)	Closest Cultured BLASTn match	Percent Identity with Closest Cultured BLASTn match
SF1.30	30	*Burkholderia terricola*	98.39%
SF1.40	40	*Alicyclobacillus acidoterrestris*	96.50%
SF1.50	50	*Alicyclobacillus herispum*	96.75%
SF1.60	60	*Alicyclobacillus acidocaldarius*	98.42%
SF2.30	30	*Burkholderia terricola*	97.98%
SF2.40	40	*Alicyclobacillus acidoterrestris*	99.72%
SF2.50	50	*Alicyclobacillus acidocaldarius*	98.19%
SF3.30	30	*Burkholderia terricola*	99.22%
SF3.40	40	*Alicyclobacillus pomorum*	100.00%
SF3.50	50	*Alicyclobacillus herbarius*	98.94%
SF3.60	60	*Alicyclobacillus acidocaldarius*	99.37%
SF4.30	30	*Paraburkholderia phenazinium*	99.59%
SF4.40	40	*Alicyclobacillus pomorum*	96.10%
SF4.50	50	*Alicyclobacillus contaminans*	99.45%
SF4.60	60	*Alicyclobacillus acidocaldarius*	99.58%
SF5.30	30	*Burkholderia caledonica*	93.47%
SF5.40	40	*Alicyclobacillus pomorum*	98.46%
SF5.50	50	*Alicyclobacillus acidocaldarius*	98.96%
SF5.60	60	*Alicyclobacillus acidocaldarius*	98.41%
SF5.70	70	*Alicyclobacillus acidocaldarius*	99.04%
SF6.30	30	*Alicyclobacillus acidoterrestris*	99.67%

Sample ID	Enrichment Temperature (°C)	Closest Cultured BLASTn match	Percent Identity with Closest Cultured BLASTn match
SF6.40	40	*Alicyclobacillus acidoterrestris*	98.96%
SF6.50	50	*Alicyclobacillus acidocaldarius*	99.37%
SF6.60	60	*Alicyclobacillus acidocaldarius*	99.28%
SF7.30	30	*Alicyclobacillus acidoterrestris*	99.68%
SF7.40	40	*Alicyclobacillus acidoterrestris*	99.69%
SF7.50	50	*Alicyclobacillus acidocaldarius*	99.38%
SF7.60	60	*Alicyclobacillus acidocaldarius*	99.39%
SF8.30	30	*Paraburkholderia phenazinium*	98.58%
SF8.40	40	*Alicyclobacillus acidocaldarius*	96.16%
SF8.50	50	*Alicyclobacillus cycloheptanicus*	99.07%
SF8.60	60	*Alicyclobacillus acidocaldarius*	98.88%
SF9.30	30	*Alicyclobacillus fastidiosus*	98.63%
SF9.40	40	*Alicyclobacillus acidoterrestris*	94.81%
SF9.50	50	*Bacillus coagulans*	98.40%
SF9.60	60	*Alicyclobacillus acidocaldarius*	96.60%
SF9.70	70	*Alicyclobacillus cellulosilyticus*	92.14%
SF10.30	30	*Burkholderia sediminicola*	98.38%
SF10.40	40	*Alicyclobacillus contaminans*	99.90%
SF10.50	50	*Alicyclobacillus contaminans*	99.17%
SF10.60	60	*Alicyclobacillus acidocaldarius*	92.83%
SF10.70	70	*Alicyclobacillus acidocaldarius*	99.19%

Nine of the 42 isolates shared less than 97% identity with the closest cultured BLASTn match. Of these, three were under 94% identity. These potentially novel species identified are labeled as “SF novel” in the phylogeny, along with an individual number for referral (i.e., SF novel 1). In the Figure 2 phylogeny, many of the novel isolates aligned nearest to the species which they were determined to be the highest match with from the BLASTn result. However, while one of the unidentified isolates (SF1.40), matched most closely with *A. acidoterrestris*, Figure 2 shows that this species has a more common ancestor with *A. fastidious*. The SF4.40 novel isolate appears to share common ancestors with *A. pomorum*, but this isolate falls on a separate node despite being most closely matched in BLASTn with *A. pomorum*. The *A. acidocaldarius* isolates and novel isolates most closely related to *A. acidocaldarius* (SF8.40, SF9.60, and SF10.60) all fall within the same clade split into several nodes. The SF10.60 novel isolate appears to be the most diverse when compared to the other *A. acidocaldarius* isolates, as its branch length is triple the others. The *B. coagulans* isolate, SF9.50, is fairly distinct within this community, diverging earlier than other groups, being separate from the distinct clades with a branch length around 0.14. The SF9.70 isolate most closely aligned with *A. cellulosilyticus* in the BLASTn query but aligned near *A. herbarius* and *A. contaminans* in the phylogeny. The SF5.30 novel isolate branches off a node shared with *B. sediminicola* and was matched most to this species from the BLASTn return. Despite sharing a node, this SF5.30 isolate is fairly diverse from its clade with a long branch, representing larger differences in the nucleotide sequence.

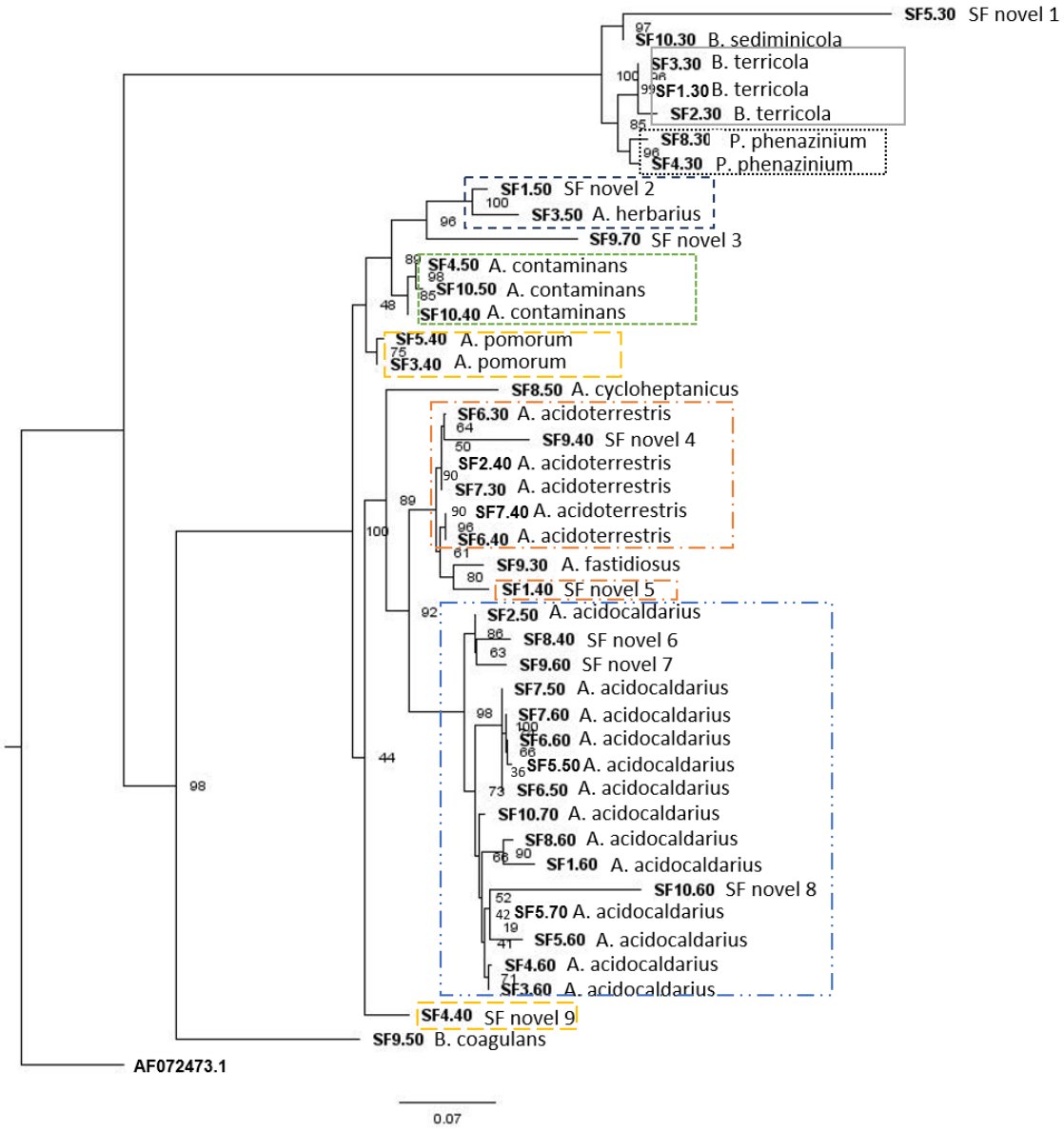

Figure 2. Maximum likelihood phylogeny of 16S rRNA sequences of isolates obtained from incubation at 30°C, 40°C, 50°C, 60°C, and 70°C with bootstrapping (n=100) across 1,486 nt. Colored boxes highlight distinct clades of species identified or most closely related to the same species.

The SF3.30 and SF1.30 isolates were identical, and grouped with SF2.30, all identified as *B. terricola*. The isolate SF5.30 was under the 97% criterion for identity, but most closely aligned with *B. terricola* (SF10.30) sharing 90.75% of the sequence

nucleotides. The potentially novel species SF1.50 was most closely related to the SF3.50 *A. herbarius* isolate in the phylogeny, sharing 96.46% nucleotide identity. The novel isolate SF4.40 was under the 97% identity criterion and most closely related to *A. pomorum* during BLASTn search. This isolate most closely aligned with the *A. pomorum* isolate SF5.40 at 97.18% shared nucleotides. Isolates SF9.40 and SF1.40 were under the 97% criterion but that most closely aligned with *A. acidoterrestris* in BLASTn results. In the Figure 2 phylogeny, both SF1.40 and SF9.40 grouped with the *A. acidoterrestris* isolates. The SF1.40 isolate aligned best with an *A. acidoterrestris* isolate, SF6.40, sharing 97.31%, whereas SF9.40 best matched with SF8.60, an *A. acidocaldarius* isolate at 95.48% despite grouping with *A. acidoterrestris* in the Figure 2 phylogeny. The three potentially novel isolates most closely related to *A. acidocaldarius* via BLASTn results also aligned most closely with *A. acidocaldarius* isolates in the Figure 2 phylogeny (SF10.60, SF8.40, and SF9.60).

A GLM with a binomial distribution was used to analyze the effectiveness of the five different temperatures in promoting the outcome of *A. acidocaldarius*. A generalized linear mixed model was also assessed for the analysis, but it was found that specifying a random-effects term was not necessary for this study as it produced the same output as our GLM model.

Due to the low growth outcome of any species at 70°C, this temperature was excluded from the analysis as it biased the GLM in a false manner and did not provide a reasonable capability to obtain isolates. The deviance statistics provided high significance for this model, finding that as temperature increases the odds and probability for growth

increase. The odds ratio was 1.27 meaning growth of *A. acidocaldarius* is estimated to increase 27 times per unit increase of temperature. Based on the graph of the data and model the highest growth odds appeared to be around 60°C (Figure 3). The function "predict" was used to estimate the probabilities of growth per temperature. From this 60°C was determined to have the highest probability of growth at (30°C =0.0038, 40°C =0.0395, 50°C =0.3096, **60°C =0.8301**).

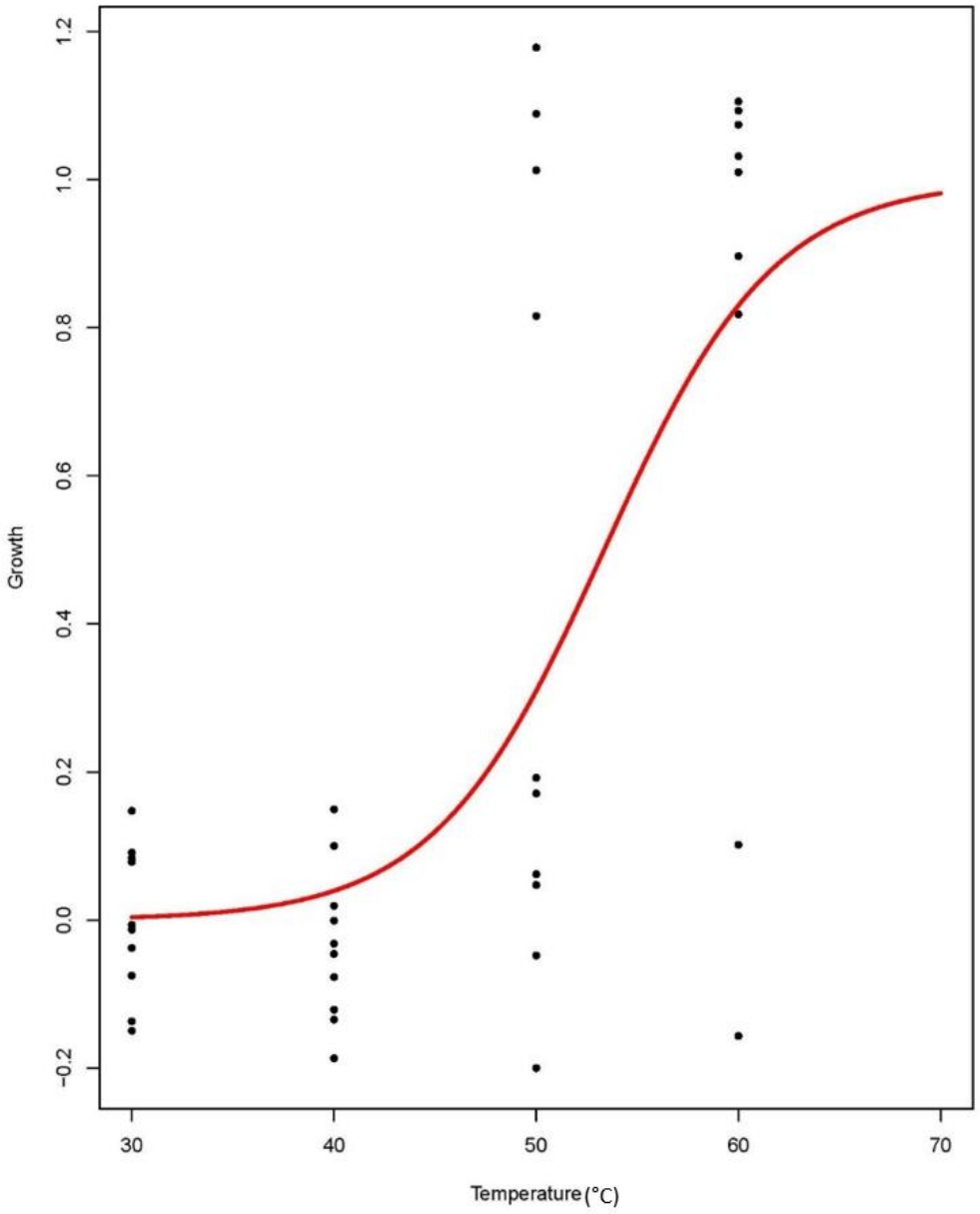

Figure 3. Results of the Generalized Linear Model analyzing growth concentration of spores by temperature of incubation (°C).

Six different species were isolated at 30°C including *A. acidoterrestris*, *A. fastidiosus B. terricola*, *B. sediminicola*, *P. phenazinium*, and one species below the 97% criterion via BLASTn. The 40°C treatment grew seven different species including four

isolates under 97% identity criterion, *A. acidoterrestris*, *A. contaminans*, and *A. pomorum*. Six species, including *A. acidocaldarius, A. contaminans, A. cycloheptanicus, A. herbarius, B. coagulans*, and one species under the 97% criterion, were obtained from the 50°C incubation. Only three species were isolated from the 60°C treatment including *A. acidocaldarius*, and two isolates under the 97% criterion. The 70°C incubation produced two species, *A. acidocaldarius* and a novel isolate. The 40°C temperature produced the highest species abundance (S) of seven, followed by 30°C and 50°C, both at six. The 40°C temperature also had the highest Shannon-Wiener Index at 1.8343 when compared to the other four temperatures tested. However, both 30°C and 40°C had similar evenness values of 0.9464 and 0.9427 respectively.

Table 2. Community structure of richness, Shannon-Weiner Index, and evenness by temperature from isolates obtained from soil enrichment.

Temperature (°C)	Richness (S)	S-W Index (H)	Evenness (J)
30	6	1.6957	0.9464
40	7	1.8343	0.9427
50	6	1.6094	0.8982
60	3	0.6837	0.6224
70	2	0.6365	0.9183

Discussion

Bacterial species were cultured and isolated at pH 2.5 from all five temperatures of the soil enrichments. At least 12 different species were identified from the temperature trials using samples from one location using PTYG medium. Two species were from the genus *Burkholderia*, eight from *Alicyclobacillus*, one from *Bacillus*, and one from

Paraburkholderia. Including all the species under 97% identity, 21 distinct species were identified. The lower temperatures of 30°C, 40°C, and 50°C had the highest richness—at least double the temperatures of 60°C and 70°C. These lower temperatures would be expected to have more total species abundance as these temperatures are more widespread throughout the environment [107]. The 70°C treatments had the lowest return of isolates. This may mean there is a lower number and density of species able to grow at this temperature present in soils. While soils have been documented to reach high temperatures, it is likely that the strains isolated came from endospores that were dispersed from a permissive location.

While many isolates were similar, sharing 99% or more nucleotide identity with at least one other isolate, only two isolates were identical, both being identified as *B. terricola* (SF3.30 and SF1.30). Along with the largest number of isolates from one species, *A. acidocaldarius* isolates shared the greatest diversity within nucleotide identity with each strain. This diversity may represent dispersal of spores from different regions or active sites that would have the potential to produce slight genetic variations. The diversity may also represent a dispersal from single region with many permissive habitats or high local strain diversity. The novel isolates of SF8.40 and SF9.60 may be the same species given their close relation in the phylogeny and sequence alignments.

In line with the literature values for the preferred temperature range for *A. acidocaldarius* of 55-60°C [42], I found that 60°C was the most likely to enrich for *A. acidocaldarius*. However, despite testing a wide range of temperatures encapsulating the reported optimal temperatures for *A. tolerans*, 37-42°C [49], this species was not found.

Since *A. tolerans* has been isolated from places near the areas sampled for this project [1], it is unlikely that there is truly no *A. tolerans* present, but rather that the preferred conditions for the species to grow were not met or that concentrations of this species are very low. Few studies have been published involving *A. tolerans* but it has been identified within samples from Russia [49], China [108], Japan [109], Germany [110], as well as the United States [1].

In addition, it is difficult to solidify plates at the lower pH optimal for *A. tolerans*, 2.5-2.7 [49]. Lower pH may have aided in the isolation of *A. tolerans*. Plate limitations may be why *A. acidocaldarius* was able to be isolated as this species prefers a higher pH at approximately 4.5 [42]. Further research is needed to understand the broader biogeography of this species.

Five *A. acidoterrestris* isolates were identified. Isolates within this species had low genetic diversity. Similar to the pattern seen with *A. tolerans* isolates from Hawai'i and Lassen, this may mean these isolates are dispersing from the same location at a single or multiple times or potentially a seeding event. However, further isolation of strains and analysis of alleles would provide more insight to their genetic relations. *Alicyclobacillus acidoterrestris* is a more widely studied species for its contamination involving juices [44]. Currently a wide variety of different growth media has been utilized in different studies including Potato Dextrose Broth [111], Orange Serum Broth [112], and K Broth [113] media, but PTYG medium may provide an additional suitable medium for further research on this species.

Unexpectedly, nine isolates were under 97% nucleotide identity with their closest matches in NCBI database. Nine potentially novel species out of 42 is a large return of novel species. The return of novel species from this experiment may represent the lack of studies under enrichments at high heat and low pH conditions. The PTYG medium may provide a more optimal nutrient environment that allowed these novel species to be cultured in comparison to past studies that have processed soil samples comparably. Utilizing new media in future studies might produce similar results of novel species from thermoacidophilic conditions. This large discovery of novel species may represent dispersal from understudied permissive habitats to the regions sampled. These findings also support the diversity and information that spore banks have to offer, which is currently underutilized in research.

CHAPTER 4: EFFECTS OF DIFFERENT ENRICHMENT MEDIA ON THE CONCENTRATION OF THERMOACIDOPHILES DETECTED USING MOST PROBABLE NUMBER APPROACH

Introduction

While some *Alicyclobacillus* species, such as *A. acidoterrestris,* are well characterized because they contaminate fruit juices, causing an economic impact on the juice industry, other species of this genus are less well studied [44]. An optimal growth medium for recovering *A. acidocaldarius* spores from environmental samples has not been identified [1], [111]–[113]. In addition, many of these studies that have looked at media were focused on isolation from juices or growing pure cultures of isolates. To aid in the isolation of *A. acidocaldarius* from soil, I tested several media that have been used in other published research projects to grow various *Alicyclobacillus* species. I assessed the comparative growth of thermoacidophilic spores in different media using a MPN approach. I then characterized isolates from the different media using 16S rRNA gene sequencing (Chapter 4). These trials helped to determine the medium that allowed for greater recovery of spores than other media, which I then used for quantifying thermoacidophilic spores in environmental samples, and for isolating *A. acidocaldarius*.

I aimed to identify a medium that promotes growth of the inoculated thermoacidophilic spores, especially *A. acidocaldarius*. The media tested were K medium, Potato Dextrose medium (PD), and PTYG medium as listed in

Table 3. Soil samples were tested with each medium through MPN trials, followed by pure culture isolation and identification by 16S rRNA gene sequencing.

Materials and Methods

Three media were prepared at 3 pH including K [113], PD [111], and PTYG [1] (

Table 3). Sulfuric acid was used in place of tartaric acid for PD broth. A three-replicate, ten-fold serial dilution MPN experimental design was used. This means three replicates were produced for each of three tenfold dilutions totaling to nine tubes for each soil sample.

Table 3. Media references and ingredients used to determine enrichment on the concentration of thermoacidophiles detected using Most Probable Number approach.

Media	Species grown in reference	Ingredients (g/L)
K [113]	*A. acidoterrestris*	Filtered citric acid 25% (w/v) (for adjusting pH)
		5.0 Peptone
		1.0 Glucose
		1.0 Tween 80
		2.5 Yeast extract
Potato Dextrose [111]	*A. acidoterrestris*	Adjusted to pH using tartaric acid[1]
		4.0 Potato Starch (from infusion)
		20.0 Dextrose
PTYG [1]	*A. tolerans, A. acidocaldarius*	Adjusted to pH using sulfuric acid
		0.25 Peptone
		0.25 Tryptone
		0.50 Yeast extract
		0.50 Dextrose
		0.6 $MgSO_4$ $7H_2O$
		0.07 $CaCl_2$ $2H_2O$

Processing soil samples for enrichment of thermoacidophiles

Thirty soil samples were collected from the Trinity Alps (TA) at approximately 20 m apart. The TA samples were farther from the coast and at higher elevation than the other sites, diversifying the sample locations. Five of these soil samples were randomly chosen and tested, each being inoculated into the three media types. Samples were

[1]Rozali et al. used tartaric acid, but this acid was not available so sulfuric was used in its place.

labeled with the collection number of the soil (i.e., 7.TA) in addition to the medium used (i.e., 7.TA.PTYG). Samples were then placed in sterile containers to dry for at least 14 days at room temperature (approximately 25 to 30°C). Once dry, samples were processed through a two-mm sieve and placed into sterile containers. The sieve was brushed of large debris and sprayed with ethanol in between samples and left to evaporate to sterilize the sieve and limit cross contamination. The dried, sieved soil was then used for inoculation in the media MPN trials.

Culturing condition to enrich for thermoacidophiles using a Most Probable Number process

Serial dilutions were made by measuring one gram of prepped soil into nine ml physiological saline (0.9%) in 16x125 mm screwcap tubes. One ml was pipetted into another tube of nine ml saline to produce the second tenfold dilution. This was repeated until 10^{-3} dilution was reached. Tubes were mixed well and one ml from each diluted saline sample was then pipetted into a screwcap 20x150 mm tube containing nine ml of the appropriate medium. Tubes were mixed well before drawing from. The MPN tubes were incubated at 60°C for 14 to 15 days. Turbidity was used to determine growth or no growth in each tube. The MPN amounts were determined based on Table 1 of the Bacteriological Analytical Manual (8^{th} edition), fitting the specific dilution and replicate set up used for this study [114].

Plates were then used to obtain single colony isolates from each sample of each media. However, the PD and K plates were unable to solidify under the low pH and high temperature conditions so PTYG plates at 3 pH were used for all of the single colony

isolations. Once putatively pure cultures were obtained, colony PCR was used to amplify 16S rRNA genes, which were then analyzed, sequenced, and identified using the same procedures as in Chapter 3. The 16S rRNA sequence of the SF22 isolate, identified as *A. acidocaldarius*, was used to root the phylogeny.

Results

Each media was able to enrich thermoacidophiles at 60°C from at least two of the five soil samples (Table 4). The K and PTYG media had growth from at least one tube from all soil samples tested. The 11.TA.K sample had growth in all three dilutions, but this only occurred in one of the three sets. Two samples from the PTYG medium had turbidity at 10^{-3}, but only in one set (11.TA.PTYG and 18.TA.PTYG). The PD medium had turbidity in only two samples, 23.TA.PD and 18.TA.PD. No turbidity occurred from either of these samples past 10^{-3}. The other three PD samples, while growth was present, lacked turbidity (7.TA.PD, 11.TA.PD, and 27.TA.PD).

Table 4. Growth results of the 3x3 Most Probable Number media experiment by turbidity (+) or no turbidity (-) using Potato Dextrose (PD), K. and PTYG media.

Sample ID	Medium	Set 1: 10^{-1}	Set 1: 10^{-2}	Set 1: 10^{-3}	Set 2: 10^{-1}	Set 2: 10^{-2}	Set 2: 10^{-3}	Set 2: 10^{-1}	Set 2: 10^{-2}	Set 2: 10^{-3}
7.TA.PD	PD	-	-	-	-	-	-	-	-	-
11.TA.PD	PD	-	-	-	-	-	-	-	-	-
23.TA.PD	PD	+	+	-	+	-	-	+	+	-
27.TA.PD	PD	-	-	-	-	-	-	-	-	-
18.TA.PD	PD	+	+	-	+	+	-	+	-	-
7.TA.K	K	+	-	-	-	-	-	+	-	-
11.TA.K	K	+	+	+	+	-	-	-	-	-
23.TA.K	K	+	+	-	+	-	-	+	-	-
27.TA.K	K	+	+	-	+	-	-	-	-	-
18.TA.K	K	+	+	-	+	+	-	-	+	-
7.TA.PTYG	PTYG	+	+	-	-	-	-	-	-	-
11.TA.PTYG	PTYG	+	+	-	+	+	-	+	+	+
23.TA.PTYG	PTYG	+	+	-	+	+	-	+	-	-
27.TA.PTYG	PTYG	+	+	-	+	+	-	+	+	-
18.TA.PTYG	PTYG	+	+	+	+	+	-	+	+	-

The highest concentration was from two samples grown in PTYG, 18TA.PTYG and 11.TA.PTYG, with a thermoacidophile concentrations of 460 MPN/g (Table 5). The highest concentration from inoculation in the K medium was 43 MPN/g from sample 23.TA.PTYG. The samples 18TA and 23TA were the only samples to produce growth by turbidity in all the media tested. The 7TA sample had the lowest thermoacidophile concentration out of the determinable values of the soil samples, at 7.4 MPN/g in PTYG and 3.6 MPN/g in K.

Table 5. Most Probable Number (MPN) thermoacidophile concentrations (MPN/g) and confidence intervals from media MPN trial using K, Potato Dextrose (PD), and PTYG media.

Sample ID	Medium	MPN/g	Confidence Interval Low	Confidence Interval High
7.TA.PD	PD	NA	NA	9.5
11.TA.PD	PD	NA	NA	9.5
18.TA.PD	PD	93	18	420
23.TA.PD	PD	93	18	420
27.TA.PD	PD	NA	NA	9.5
7.TA.K	K	3.6	0.17	18
11.TA.K	K	20	4.5	42
18TA.K	K	29	8.7	94
23.TA.K	K	43	9	180
27.TA.K	K	15	3.7	42
7.TA.PTYG	PTYG	7.4	1.3	18
11.TA.PTYG	PTYG	460	90	2,000
18.TA.PTYG	PTYG	460	90	2,000
23.TA.PTYG	PTYG	93	18	420
27.TA.PTYG	PTYG	240	42	1,000

The small sample size (n=5) for each medium of the media trials meant the data was non-normal. For this reason, a nonparametric, Kruskal Wallis, test was used to analyze the media trial data. Results of this test showed that there was no significant difference between the three media (p=0.08491) (Figure 4). However, the results gave the highest MPN values with PTYG, and thus I used PTYG for quantifying thermoacidophiles in environmental samples for studies. Although a p-value of 0.08 is above the normally acceptable p-value of 0.05, is still relatively low, and a larger sample

size might show significance. The PTYG medium was at least as good as the other media for allowing the growth of thermoacidophilic spores in these soil samples. While isolates were obtained from all PD samples, growth by turbidity was only observed in two isolates out of the five. No observable growth was present in the other three PD tube sets meaning no MPN concentration could be determined based on the reference table.

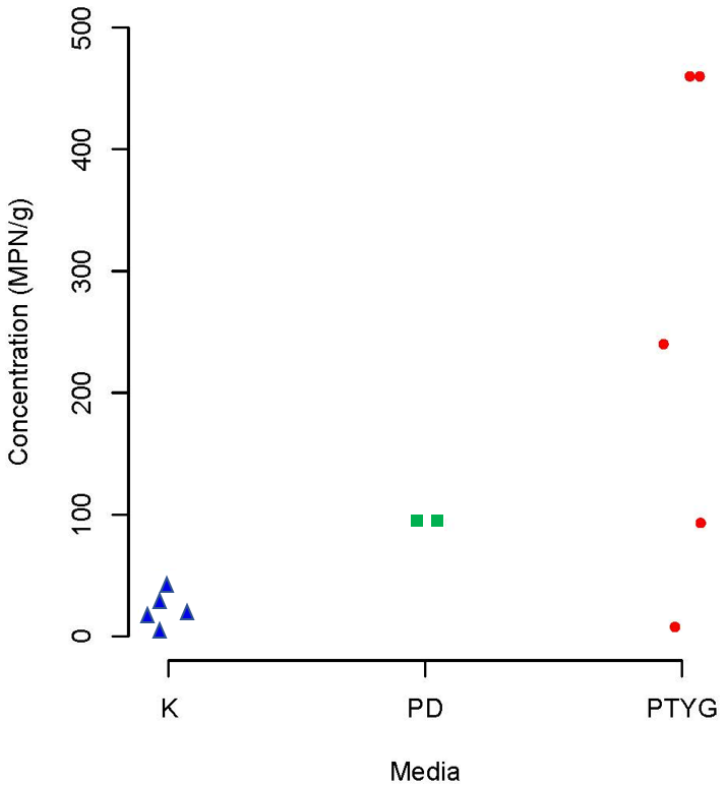

Figure 4. Concentration (MPN/g) results of media Most Probable Number (MPN) trials using K, Potato Dextrose (PD), and PTYG media.

I sequenced 16S rRNA genes to identify pure cultures derived from these enrichments (Table 6). None of the isolates from these samples were identified as either *A. acidocaldarius* or *A. tolerans*. No species were able to be isolated from the K samples. Three of the isolates from the PTYG samples were identified as *A. pomorum*

(11.TA.PTYG, 23.TA.PTYG, and 27.TA.PTYG). The two other isolates, 7.TA.PTYG and 18.TA.PTYG, most closely matched to *A. pomorum* but did not meet 97% identification criterion when processed through BLASTn. These potentially novel species identified from this experiment are labeled as "TA novel" in the phylogeny, along with an individual number for referral (i.e., TA novel 1). None of the five isolates obtained from the PD medium met the 97% identification limit, but were most closely matched to *Alicyclobacillus macrosporangiidus* (23.TA.PD), *A. contaminans* (11.TA.PD), or *A. herbarius* (7.TA.PD, 18.TA.PD, and 27.TA.PD).

Table 6. Basic Local Alignment Search Tool for nucleotides (BLASTn) return results of isolates obtained from Trinity Alps (TA) soil by enrichment at temperature 60°C using PTYG and Potato Dextrose (PD) media.

Sample ID	Medium	Closest Cultured BLASTn return	Percent Identity with Closest Cultured BLASTn return
11.TA.PD	PD	*Alicyclobacillus contaminans*	93.35%
27.TA.PD	PD	*Alicyclobacillus herbarius*	96.53%
18.TA.PD	PD	*Alicyclobacillus herbarius*	96.62%
7.TA.PD	PD	*Alicyclobacillus herbarius*	96.56%
23.TA.PD	PD	*Alicyclobacillus macrosporangiidus*	92.73%
7.TA.PTYG	PTYG	*Alicyclobacillus pomorum*	96.27%
11.TA.PTYG	PTYG	*Alicyclobacillus pomorum*	99.79%
18.TA.PTYG	PTYG	*Alicyclobacillus pomorum*	95.72%
23.TA.PTYG	PTYG	*Alicyclobacillus pomorum*	99.87%
27.TA.PTYG	PTYG	*Alicyclobacillus pomorum*	99.86%

The 16S rRNA phylogeny shows grouping by species identity. The isolates 18.TA.PTYG and 7.TA.PTYG, unidentified, but most closely matching *A. pomorum*,

share a node. The three isolates identified as *A. pomorum* all closely group within the same clade. The three unidentifiable species most closely related to *A. herbarius* all bunch together as well in the same clade. The last two isolates, identified most closely to *A. contaminans*, and *A. macrosporangiidus* are most closely related to each other in the phylogeny.

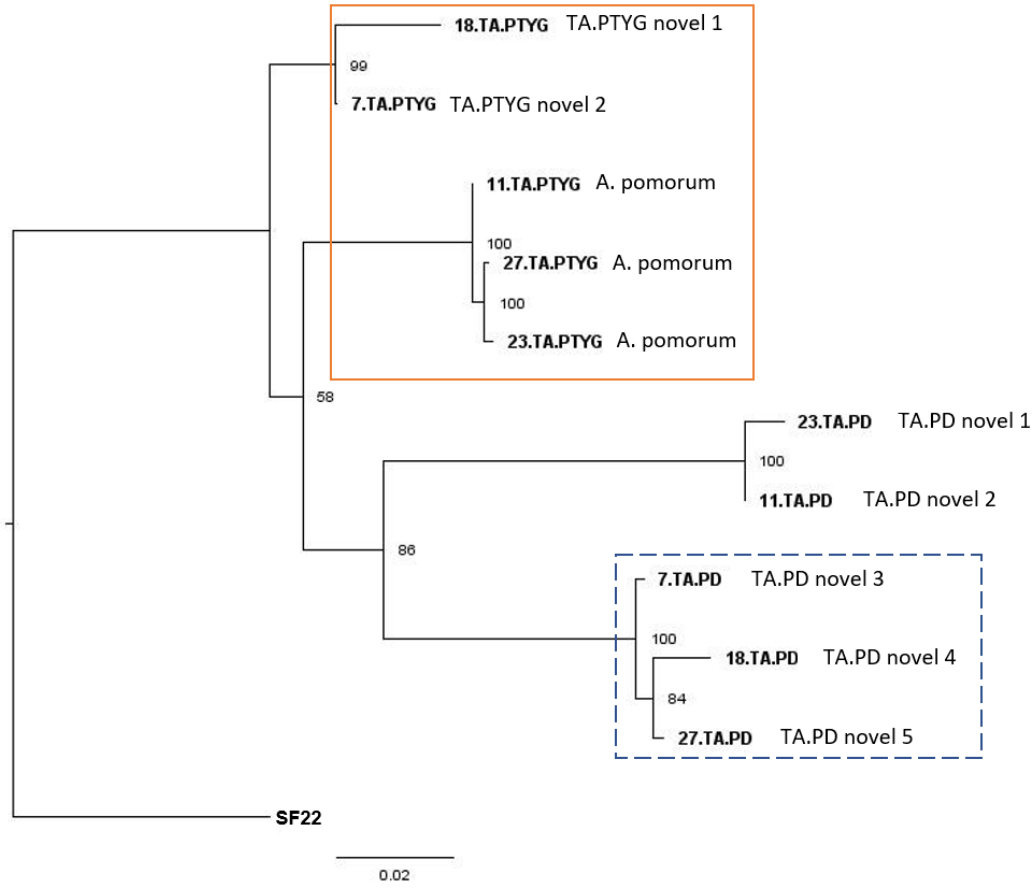

Figure 5. Maximum likelihood phylogeny of 16S rRNA sequences of isolates obtained from enrichment of soils in PTYG and Potato Dextrose (PD) media with bootstrapping (n=100) across 1,456 nt. Colored boxes highlight distinct clades of species identified or most closely related to the same species.

The three *A. pomorum* species, 11.TA.PTYG, 23.TA.PTYG, and 27.TA.PTYG all shared greater than 99% identity of the 16S rRNA gene (Table 6). The remaining isolates

in the phylogeny were under the 97% criterion for identification in the BLASTn search. Two of these species from PTYG medium grouped together in the phylogeny but were only similar across 94.15% nucleotides (7.TA.PTYG and 18.TA.PTYG) (Figure 5. Maximum likelihood phylogeny of 16S rRNA sequences of isolates obtained from enrichment of soils in PTYG and Potato Dextrose (PD) media with bootstrapping (n=100) across 1,456 nt.). The 23.TA.PD and 11.TA.PD isolates were nearly identical in alignment, falling close to each other on the phylogeny. The last three isolate of 7.TA.PD, 18.TA.PD, and 27.TA.PD, grouped closely on the phylogeny, with 7.TA.PD and 27.TA.PD sharing slightly more identity in nucleotide sequences.

Discussion

In this study I used an MPN approach to evaluate the relative abilities of three different media to recover thermoacidophilic endospores from soil samples. Isolates were obtained from two of the three media tested. I determined that thermoacidophilic spores were present at concentrations ranging from at least one to 460 per g in each of the soil samples.

The PTYG medium allowed for the highest MPN values, but these values were not significantly higher than the other two media as judged by a Kruskal Wallis test. The PTYG and K media both contain peptone, a sugar source, and yeast extract. In addition to these ingredients PTYG contains tryptone, magnesium sulfate heptahydrate, and calcium chloride dihydrate whereas K has tween 80. The PD medium contains a sugar source as well, along with potato starch. While two soil samples produced a higher MPN in the PD

medium than in the K medium, three other soil samples produced no growth in the PD medium despite these soil samples producing growth in both the K and PTYG media. K medium produced the lowest growth concentrations of thermoacidophilic spores but is most similar to PTYG medium in ingredients. The nutrients provided by PTYG may provide a set of more optimal nutrients for the spores to grow. Alternatively, the lower concentration of the nutrients in PTYG may allow for better recovery of the environmental isolates. More robust studies of the biochemical characteristics of each medium and their interactions with thermoacidophilic species could be investigated to discover a more optimized growing environment.

None of the isolates collected from these samples were identified as *A. acidocaldarius*. However, this does not mean this species is not present in the soils, but rather, that this species may be less abundant than others or that it was outcompeted for resources. While growth was evident in the MPN tubes for K, no isolates were obtained. This may be due to the use of PTYG plates for colony isolation, which was necessary given the instability of the K plates. More time would have to be devoted into developing a plate medium for K to see what diversity of species could be obtained. The isolates falling under the 97% match criterion may be novel species, and there was a large number of these isolates, as in the temperature trials. The high number of novel species may represent the diversity that spore banks contain, either due to recent or past dispersals of species that have been previously understudied. In addition, the lack of genetic and environmental changes like those that would affect vegetative cells may provide insight to past species genetics. However, because these novel species have been

previously under-studied, further research would be needed to understand where they might have originated from or other insights to their biogeography.

The phylogeny of the 16S rRNA sequences from the isolates obtained from the PTYG and PD media suggest that these media may promote the growth of different sets of bacteria (Figure 5). The potentially novel species 7.TA.PTYG and 11.TA.PTYG are closely related to *A. pomorum* given their close relation on the phylogeny. However, these isolates only matched across 94.15% of their nucleotide sequences which may suggest separate species entirely. Three of the five isolates from PD that did not meet the 97% criterion for identification (7.TA.PD, 18.TA.PD, and 27.TA.PD), may also belong to the same species, as they share a clade with greater than 98% identity. The isolates most closely related to *A. macrosporangiidus* (23.TA.PD) and *A. contaminans* (11.TA.PD) via the BLASTn query shared the same node and likely belong to the same species, sharing 99.37% identity between nucleotide sequences. All the isolates from the PD medium were below the 97% criterion for identity, suggesting that these species may be novel. Utilizing PD medium under these high temperature, low pH conditions may provide growth conditions for currently undescribed species or set of species.

CHAPTER 5: QUANTIFICATION AND CHARACTERIZATION OF THERMOACIDOPHILES FROM SIX DIFFERENT SOILS LOCATED IN HUMBOLDT AND TRINITY COUNTIES, CA

Introduction

Some bacteria have the ability to produce endospores when conditions become unfavorable. When conditions become favorable again, these spores will become vegetative cells. These durable spores can be dispersed from habitats that are permissive for vegetative growth by various factors and deposited in regions that are not permissive for vegetative growth. Analysis of spore banks in nonpermissive habitats may produce insight to dispersal patterns, origins, and abundance.

While dispersal methods are important to understand the distribution of a species, other characteristics can also influence the range of a species. Choudior et al. [17] found that factors such as taxonomic identity, phenotypic traits, and genetic elements were all significant in determining the range of microbial taxa of samples across the U.S. This means that common bacterial traits such as gram-positive vs gram-negative may affect distribution. While the ability to produce endospores may increase the range of a species, the vegetative range would only be a portion of the total occupied range, especially for species such as thermoacidophiles that require high temperature and low pH conditions. However, documenting where endospores are being dispersed can aid in the

understanding of biogeography and dispersal of thermoacidophiles, and characterization of spore banks adds an important element to understanding microbial diversity.

In this study I used an MPN approach to quantify thermoacidophile spores in soils from six geographically distinct locations. Ubiquitous presence of spores and species across sites suggests that there was a large event that seeded the areas with the same species and amounts of thermoacidophile spores. On the other hand, uneven presence of spore concentrations and species between sites suggests environmental factors, site characteristics, and/or methods of dispersal that distributes spores to each site distinctly. The shared effects would likely produce areas of homogeneity, or overlap of identical clones between sites, but the effects of unshared factors would be seen as the heterogeneity, or different species between sites.

Materials and Methods

Samples were collected from six different sites, TA (Trinity Alps), Maple Creek Douglas fir forest (DF), Maple Creek meadow (M), Mad River Douglas fir forest (MDF), Redwood forest (RED), and SF (Spruce forest [same site as Chapter 3 but resampled]) (Figure 6). Approximately five g of each soil sample was collected using aseptic techniques. Samples were collected at least two m away from each other. Soil samples were processed as in the media trials protocol (Chapter 4). Samples were placed in sterile containers to dry for at least 14 days at room temperature (approximately 25 to 30°C) and processed through a two-mm sieve.

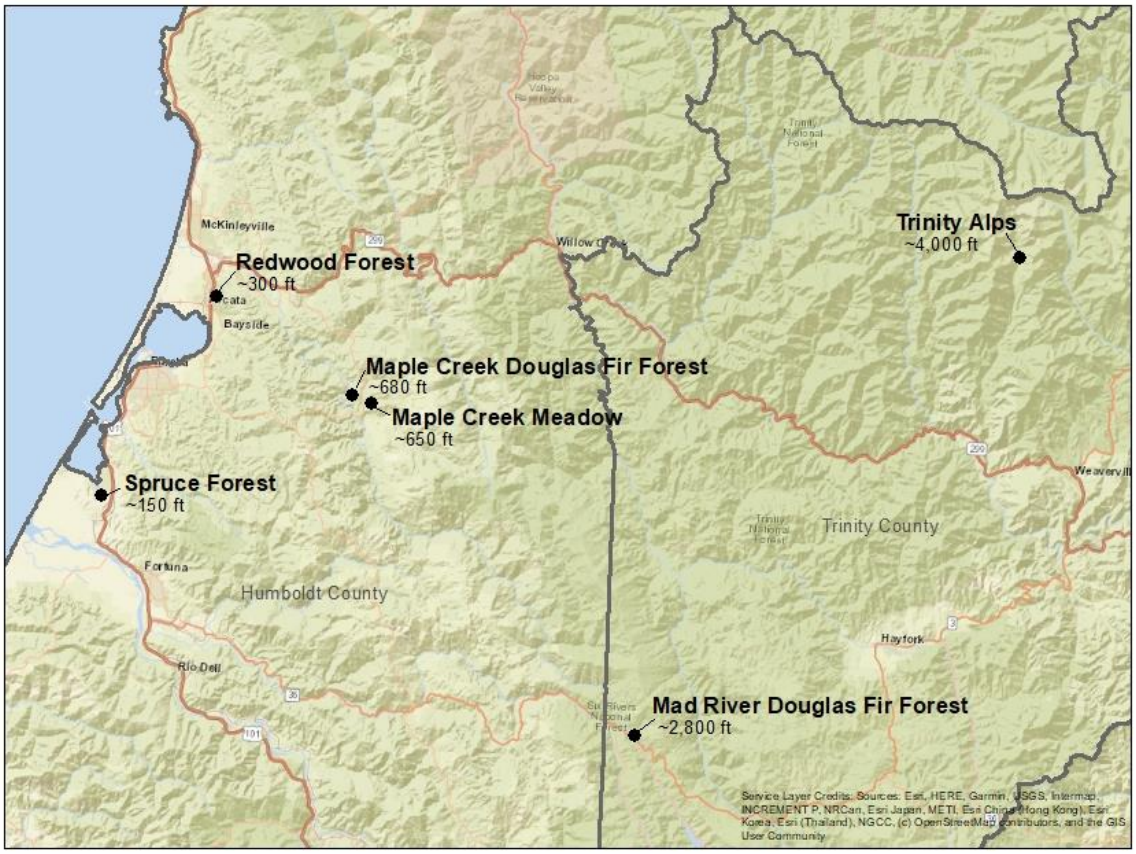

Figure 6. Map showing locations of Trinity Alps, Spruce forest, Maple Creek Douglas fir forest, Maple Creek meadow, Mad River Douglas fir forest, and Redwood forest throughout Humboldt and Trinity Counties in California.

Culturing conditions to enrich for thermoacidophiles using a Most Probable Number process

The location MPN experiment consisted of one set of three ten-fold dilutions for each soil sample. One g of soil was added to nine ml of physiological saline (0.9%) solution denoting the 10^{-1} dilution. One ml of this dilution was aseptically pipetted into a new nine ml tube of physiological saline. This was repeated until the 10^{-3} dilution. One ml from each serial dilution was then aseptically pipetted into a tube containing nine ml PTYG media. Tubes were mixed thoroughly for at least 20 seconds between pipetting.

The soil was prevented from fully settling to the bottom of the tube before the aliquots were drawn. The MPN tubes were incubated at 60°C for 14 to 15 days. The presence of turbidity determined growth and the absence of turbidity determined no growth in each tube. The Environmental Protection Agency's MPN Calculator was used to determine MPNs [115]. This calculator allows manipulation of number of dilutions, number of tubes per dilution, inoculum volume, and number of positive tubes.

One tube from each sample was streaked onto pH 3.5 PTYG plates to obtain single colony isolates from each sample. Once putatively pure cultures were obtained by serial streaking for single colony isolation, samples were processed by colony PCR in 35 µl reactions using 16S rRNA gene amplification protocols used in the temperature trials (Chapter 3), followed by PCR purification and Sanger sequencing as previously described. Once returned, sequences were checked for quality, edited for accuracy, and then used to construct a maximum likelihood phylogeny as in Chapter 3.

Results

Each location had growth as assessed by turbidity in at least eight of the 12 samples (Table 7. Growth results of the 3x3 Most Probable Number location trial by turbidity (+) or no turbidity (-) from Trinity Alps (TA), Spruce forest (SF), Mad River Douglas fir forest (MDF), Maple Creek Douglas fir forest (DF), Maple Creek meadow (M), and Redwood forest (RED) soils using PTYG medium.). The TA location had three samples with growth up to 10^{-2} dilution (TA12, TA9, and TA11), five samples with growth only at 10^{-1} (TA10, TA1, TA3, TA5, and TA6), and four samples with no visible

growth (TA2, TA4, TA8, and TA7). The MDF location had no growth in any samples at 10^{-3} dilution but six samples with growth up to 10^{-2} (MDF1, MDF2, MDF3, MDF5, MDF7, and MDF9) and five samples with growth at only 10^{-1} (MDF4, MDF6, MDF8, MDF10, and MDF12). The RED location had two samples with growth up to 10^{-3} dilution (RED5 and RED9) with the other 10 samples having growth up to 10^{-2} dilutions (RED1, RED2, RED3, RED4, RED6, RED7, RED8, RED10, RED11, and RED12). Four samples grew at 10^{-2} dilution (SF23, SF25, SF26, and SF27) and seven grew at only 10^{-1} (SF21, SF22, SF24, SF28, SF29, SF30, and SF31) from the SF location. The DF location had four samples with growth up to 10^{-2} (DF1, DF2, DF3, and DF8) and four sample with growth up to 10^{-1} only (DF4, DF6, DF7, and DF9). The M location had one sample with growth up to 10^{-3} (M1), two samples with growth up to 10^{-2} (M3 and M4), and eight samples with growth at only 10^{-1} (M2, M5, M6, M7, M8, M9, M11, and M12).

Table 7. Growth results of the 3x3 Most Probable Number location trial by turbidity (+) or no turbidity (-) from Trinity Alps (TA), Spruce forest (SF), Mad River Douglas fir forest (MDF), Maple Creek Douglas fir forest (DF), Maple Creek meadow (M), and Redwood forest (RED) soils using PTYG medium.

Sample ID	Location	10^{-1}	10^{-2}	10^{-3}
TA1	TA	+	-	-
TA2	TA	-	-	-
TA3	TA	+	-	-
TA4	TA	-	-	-
TA5	TA	+	-	-
TA6	TA	+	-	-
TA7	TA	-	-	-
TA8	TA	-	-	-
TA9	TA	+	+	-
TA10	TA	+	-	-
TA11	TA	+	+	-
TA12	TA	+	+	-
MDF1	MDF	+	+	-
MDF2	MDF	+	+	-
MDF3	MDF	+	+	-
MDF4	MDF	+	-	-
MDF5	MDF	+	+	-
MDF6	MDF	+	-	-
MDF7	MDF	+	+	-
MDF8	MDF	+	-	-
MDF9	MDF	+	+	-
MDF10	MDF	+	-	-
MDF11	MDF	-	-	-
MDF12	MDF	+	-	-
RED1	RED	+	+	-
RED2	RED	+	+	-

Sample ID	Location	10^{-1}	10^{-2}	10^{-3}
RED3	RED	+	+	-
RED4	RED	+	+	-
RED5	RED	+	+	+
RED6	RED	+	+	-
RED7	RED	+	+	-
RED8	RED	+	+	-
RED9	RED	+	+	+
RED10	RED	+	+	-
RED11	RED	+	+	-
RED12	RED	+	+	-
SF20	SF	-	-	-
SF21	SF	+	-	-
SF22	SF	+	-	-
SF23	SF	+	+	-
SF24	SF	+	-	-
SF25	SF	+	+	-
SF26	SF	+	+	-
SF27	SF	+	+	-
SF28	SF	+	-	-
SF29	SF	+	-	-
SF30	SF	+	-	-
SF31	SF	+	-	-
DF1	DF	+	+	-
DF2	DF	+	+	-
DF3	DF	+	+	-
DF4	DF	+	-	-
DF5	DF	-	-	-
DF6	DF	+	-	-
DF7	DF	+	-	-
DF8	DF	+	+	-

Sample ID	Location	10^{-1}	10^{-2}	10^{-3}
DF9	DF	+	-	-
DF10	DF	-	-	-
DF11	DF	-	-	-
DF12	DF	-	-	-
M1	M	+	+	+
M2	M	+	-	-
M3	M	+	+	-
M4	M	+	+	-
M5	M	+	-	-
M6	M	+	-	-
M7	M	+	-	-
M8	M	+	-	-
M9	M	+	-	-
M10	M	-	-	-
M11	M	+	-	-
M12	M	+	-	-

Since the location MPNs had no replicates as a one by three MPN structure, individual sample MPNs could not be determined. Instead, each location MPN was determined using the 12 samples from each location. For each calculation, MPN, MPN Corrected for Bias, Cornish & Fisher [116] and Loyer & Hamilton 95% confidence intervals [117], Spearman-Karber Estimate [118], [119], Thomas [120], and R Based MPN values are provided by the portal. The small number of dilutions (k=3) suggests that the exact Loyer & Hamilton confidence intervals would be an appropriate interval to use [117]. MPN values seemed acceptable to use over the other methods provided based on the calculations used [121]. The results of these calculations are listed in

Table 8. The Redwood site was determined to have the highest concentration of thermoacidophilic spores (MPN= 325.713/g).

Table 8. Results of the Most Probable Number (MPN) calculations of MPN concentration as MPN/g and Loyer & Hamilton 95% exact confidence intervals from the Environment Protection Agency's MPN calculator based on turbidity growth for each location tested: Trinity Alps, Mad River Douglas fir forest, Redwood forest, Spruce forest, Maple Creek Douglas fir forest, and Maple Creek meadow.

Location	MPN/g	Loyer & Hamilton Lower 95%/g	Loyer & Hamilton Upper 95%/g
Trinity Alps	13.239	6.588	25.021
Mad River Douglas fir forest	35.665	18.760	59.482
Redwood forest	325.713	128.967	691.331
Spruce forest	28.241	12.375	58.298
Maple Creek Douglas fir forest	14.689	9.062	22.841
Maple Creek meadow	27.865	13.576	50.550

From the 72 samples, 50 isolates were collected and their 16S rRNA genes were sequenced. Ten isolates were obtained from DF, 12 from M, six from MDF, 11 from RED, five from SF, and six from TA (Table 9). Twenty-two isolates were identified as *A. pomorum*, with each site producing either three or four isolates of this species. Eight isolates were identified as *A. acidocaldarius*. Of these, two were from the DF location (DF9 and DF10), five were from the RED location (RED1, RED5, RED8, RED9, and RED12), and one was from the SF location (SF22). One isolate of each of the following species was obtained, each from a different site: *Bacillus smithii* (TA3), *Tuberibacillus calidus* (M11), and *A. contaminants* (DF1).

Table 9. Basic Local Alignment Search Tool for nucleotides (BLASTn) results of isolates obtained from Spruce forest, Trinity Alps, Maple Creek Douglas fir forest, Maple Creek meadow, Mad River Douglas fir forest, and Redwood forest soils using PTYG medium.

Sample ID	Location	Closest Cultured BLASTn return	Percent Identity with Closest Cultured BLASTn return
DF9	Maple Creek Douglas fir forest	*Alicyclobacillus acidocaldarius*	99.70%
DF10	Maple Creek Douglas fir forest	*Alicyclobacillus acidocaldarius*	99.16%
RED1	Redwood forest	*Alicyclobacillus acidocaldarius*	99.16%
RED5	Redwood forest	*Alicyclobacillus acidocaldarius*	99.03.%
RED8	Redwood forest	*Alicyclobacillus acidocaldarius*	99.09%
RED9	Redwood forest	*Alicyclobacillus acidocaldarius*	98.78%
RED12	Redwood forest	*Alicyclobacillus acidocaldarius*	99.09%
SF22	Spruce forest	*Alicyclobacillus acidocaldarius*	99.65%
M12	Maple Creek meadow	*Alicyclobacillus herbarius*	92%
MDF2	Mad River Douglas fir forest	*Alicyclobacillus herbarius*	96.61%
RED4	Redwood forest	*Alicyclobacillus herbarius*	96.82%
RED7	Redwood forest	*Alicyclobacillus herbarius*	96.82%
DF3	Maple Creek Douglas fir forest	*Alicyclobacillus macrosporangiidus*	93.47%
DF4	Maple Creek Douglas fir forest	*Alicyclobacillus macrosporangiidus*	93.48%
DF5	Maple Creek Douglas fir forest	*Alicyclobacillus macrosporangiidus*	93.41%
M1	Maple Creek meadow	*Alicyclobacillus macrosporangiidus*	92%
M2	Maple Creek meadow	*Alicyclobacillus macrosporangiidus*	92%
M5	Maple Creek meadow	*Alicyclobacillus macrosporangiidus*	92%

Sample ID	Location	Closest Cultured BLASTn return	Percent Identity with Closest Cultured BLASTn return
M6	Maple Creek meadow	*Alicyclobacillus macrosporangiidus*	93.50%
RED6	Redwood forest	*Alicyclobacillus macrosporangiidus*	93.50%
TA5	Trinity Alps	*Alicyclobacillus macrosporangiidus*	95.11%
M9	Maple Creek meadow	*Alicyclobacillus microsporangiidus*	93.61%
M10	Maple Creek meadow	*Alicyclobacillus microsporangiidus*	92.55%
MDF8	Mad River Douglas fir forest	*Alicyclobacillus microsporangiidus*	92.98%
DF6	Maple Creek Douglas fir forest	*Alicyclobacillus pomorum*	100%
DF7	Maple Creek Douglas fir forest	*Alicyclobacillus pomorum*	99.58%
DF8	Maple Creek Douglas fir forest	*Alicyclobacillus pomorum*	100%
DF11	Maple Creek Douglas fir forest	*Alicyclobacillus pomorum*	99.93%
M3	Maple Creek meadow	*Alicyclobacillus pomorum*	99.93%
M4	Maple Creek meadow	*Alicyclobacillus pomorum*	99.93%
M7	Maple Creek meadow	*Alicyclobacillus pomorum*	99.93%
M8	Maple Creek meadow	*Alicyclobacillus pomorum*	99.93%
MDF1	Mad River Douglas fir forest	*Alicyclobacillus pomorum*	99.93%
MDF3	Mad River Douglas fir forest	*Alicyclobacillus pomorum*	99.93%
MDF4	Mad River Douglas fir forest	*Alicyclobacillus pomorum*	99.93%

Sample ID	Location	Closest Cultured BLASTn return	Percent Identity with Closest Cultured BLASTn return
MDF9	Mad River Douglas fir forest	*Alicyclobacillus pomorum*	99.39%
RED2	Redwood forest	*Alicyclobacillus pomorum*	99.65%
RED3	Redwood forest	*Alicyclobacillus pomorum*	99.65%
RED10	Redwood forest	*Alicyclobacillus pomorum*	99.65%
SF21	Spruce forest	*Alicyclobacillus pomorum*	99.93%
SF23	Spruce forest	*Alicyclobacillus pomorum*	100%
SF27	Spruce forest	*Alicyclobacillus pomorum*	99.93%
SF28	Spruce forest	*Alicyclobacillus pomorum*	99.93%
TA6	Trinity Alps	*Alicyclobacillus pomorum*	97.49%
TA9	Trinity Alps	*Alicyclobacillus pomorum*	99.90%
TA12	Trinity Alps	*Alicyclobacillus pomorum*	99.69%
DF1	Maple Creek Douglas fir forest	*Alicyclobacillus contaminans*	99.58%
TA3	Trinity Alps	*Bacillus smithii*	99.11%
M11	Maple Creek meadow	*Tuberibacillus calidus*	100%
TA10	Trinity Alps	*Tumebacillus flagellatus*	94.25%

Of the 50 isolates, 17 were below the 97% criterion for identity (Table 9). These potentially novel species identified are labeled with “LT” for location trial, the location abbreviation, followed by “novel” in the phylogeny, along with an individual number for referral (i.e., LT.M novel 1). Four of these isolates most closely related to *A. herbarius* but under the 97% criterion, were obtained from MDF, M, and RED locations (M12, MDF2, RED4, and RED7). Twelve isolates were most closely related to either *Alicyclobacillus macrosporangiidus* (DF3, DF4, DF5, M1, M2, M5, M6, RED6, and

TA5) or *A. microsporangiidus* (M9, M10, and MDF8) were obtained from locations M, DF, MDF, RED, and TA. All of these isolates were under 96% BLASTn identity, with 11 being under 94%. These isolates are grouped together on the phylogeny, belonging to the same clade, in addition to one novel isolate most closely related to *A. herbarius*, M12 (Figure 7).

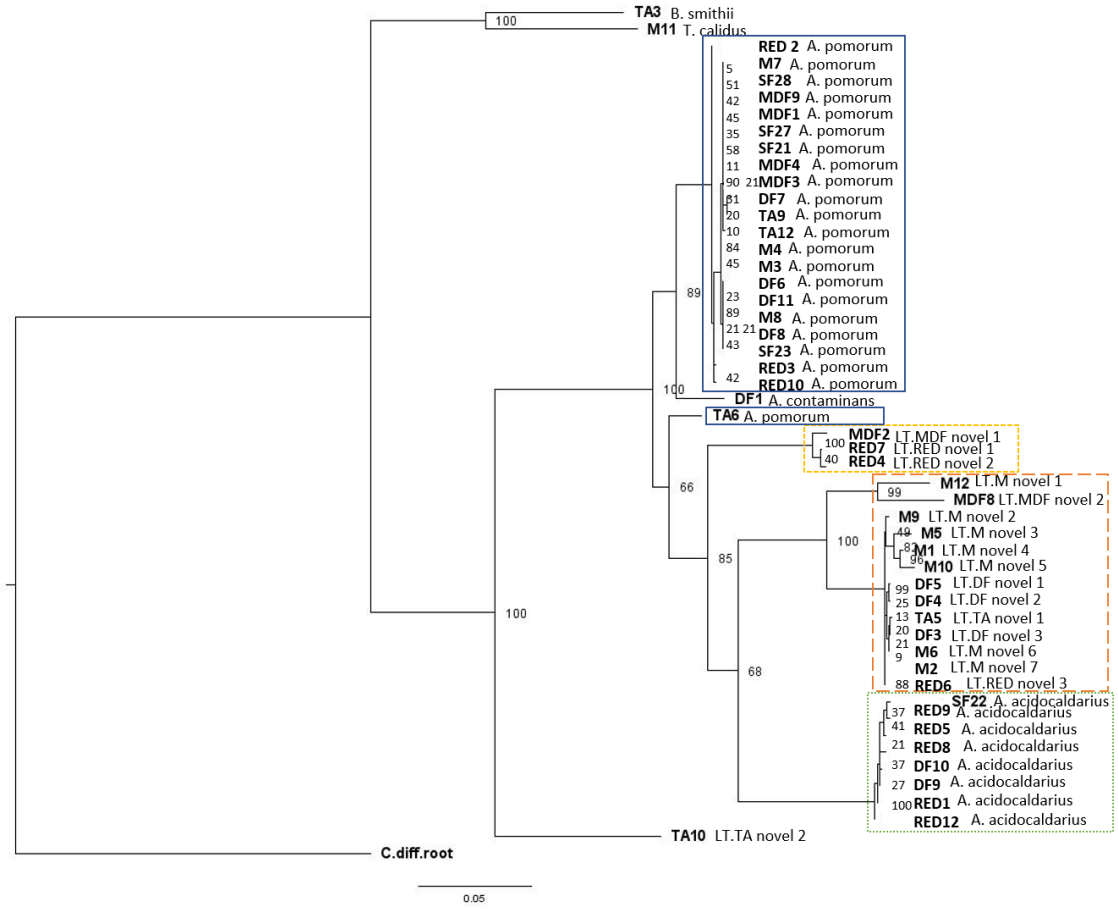

Figure 7. Maximum likelihood phylogeny of 16S rRNA sequences of isolates obtained from the Spruce forest (SF), Trinity Alps (TA), Maple Creek Douglas fir forest (DF), Maple Creek meadow (M), Mad River Douglas fir forest (MDF), and Redwood forest (RED) soils with bootstrapping (n=100) across 1,478 nt. Colored boxes highlight distinct clades of species identified or most closely related to the same species.

The 22 isolates identified as *A. pomorum* shared greater than 99% identity except for TA6. The TA6 isolate only shared 96.38% or greater sequence identity with the other *A. pomorum* isolates and fell on a different branchpoint than the other *A. pomorum* isolates in the Figure 7 phylogeny. The *A. pomorum* isolates MDF1, SF27, MDF9, SF28, M7, MDF3, MDF4, and SF21 were identical in sequence. In addition, the *A. pomorum* isolates of DF6, DF8, DF11, M8, and SF23 shared 100% identity, as well as the isolates DF6, DF8, M3, M4, and SF23. Three isolates that failed to meet the 97% identity criterion, MDF2, RED7, RED4, grouped closely together from a shared node with nearly identical nucleotide identity. Thirteen other isolates that failed to meet the 97% criterion grouped together in the Figure 7 phylogeny in the same clade. Two of these isolates from M and MDF locations were most closely related with only 95.28% alignment (M12 and MDF8). The other 11 of these isolates shared greater than 98.54% identity with M2 and RED6 being identical (DF3, DF4, DF5, M1, M2, M5, M6, M9, M10, RED6, and TA5). The isolate TA10 was lower than 97% identity via the BLAST query and most closely aligned with DF1 by sequence comparison.

The M location produced the highest species abundance (S) of 10, followed by DF at six (

Table 10). The M location also had the highest Shannon-Wiener Index at 2.138 when compared to the other five other locations tested. The DF location had the second highest Shannon-Wiener Index at 1.609 followed by RED at 1.418. The M location also had the highest evenness (J=0.9287). The evenness of DF, TA, and RED were all similar, within 0.02 units of each other. Both the MDF and SF locations had the same richness

(S=2), Shannon-Wiener Index (H=0.5), and evenness (J=0.7219) values. However, in these calculations all novel species were assumed to be separate. In addition, the small number of isolates sampled there resulted in a lack of precision in the calculations.

Table 10. Community structure of richness, Shannon-Weiner (S-W) Index, and evenness by location from isolates obtained from the Spruce forest, Trinity Alps, Maple Creek Douglas fir forest, Maple Creek meadow, Mad River Douglas fir forest, and Redwood forest soils.

Location	Richness (S)	S-W Index (H)	Evenness (J)
Maple Creek Douglas fir forest	6	1.609	0.8982
Maple Creek meadow	10	2.138	0.9287
Mad River Douglas fir forest	2	0.5	0.7219
Redwood forest	5	1.418	0.8814
Spruce forest	2	0.5	0.7219
Trinity Alps	4	1.2425	0.8962

Discussion

All locations had detectable spore concentrations with all sites having approximately equal concentrations except the RED which had about a ten-fold higher thermoacidophilic spore concentration. In comparison to the other locations sampled, the RED site may be exposed to factors that distribute a higher concentration of thermoacidophilic spores such as coastal fog, forest management, or increased foot traffic [21], [122], [123]. The nearly equal thermoacidophilic spore concentrations of the other five sites may represent more basal dispersal factors that are affecting these sites similarly. The differences in spore concentration gradients across the locations sampled

would suggest that even dispersal is not occurring. Distinct dispersal factors are likely playing a role in the abundance and diversity of the thermoacidophilic spores in the regions. Specific research would need to be performed to better understand what environmental factors most influence dispersal to each individual region such as comparing the microbial diversities found within the non-permissive locations to the microbial diversities found in different local samples of air, cloud, and/or rain near the non-permissive locations.

Thermoacidophilic isolates were obtained from all locations using soil enrichments with PTYG medium. Every location had detectable spore concentrations. Multiple of the same species were isolated from non-permissive soils from different locations. The M location produced the highest culturable number of isolates, with isolates obtained from 12 out of the 12 samples tested. Seventeen of the isolates collected from the location trial were potentially novel species, with the 16S rRNA gene sequences below 97% identity with the closest described isolate. These novel isolates were collectively produced from five of the six locations tested. The M location also produced the highest number of potentially novel isolates. The novel species from the M location likely prefer the conditions produced within this study causing a higher abundance over other species from the location. On the other hand, these potentially novel isolates may be present in higher concentrations compared to other thermoacidophilic spores in the M location.

Five of the eight *A. acidocaldarius* isolates were obtained from the RED location. The other *A. acidocaldarius* isolates were obtained from the SF and DF locations, with

none from MDF, M, or TA. This may be due to a higher abundance of *A. acidocaldarius* or that the experimental environment provided preferential growth conditions over other species present in these areas. The RED, DF, and SF locations are slightly more coastal than the other sites. Isolating *A. acidocaldarius* from these areas may suggest the influence of coastal environmental factors in the distribution of this species to these regions.

The species *A. pomorum* was returned from every location suggesting widespread dispersal and/or presence of these endospores throughout non-permissive soils. However, this species was not cultured from another study involving permissive habitats of thermoacidophiles which may mean this species is present in lower concentrations in these habitats of low pH, high temperature, geothermal formations [1]. The optimal temperature for this species is between 45-50°C at pH 4.5-5 [55]. These conditions are slightly less extreme than other thermoacidophiles, meaning there may be more suitable habitat for this species to occupy and disperse from.

Alicyclobacillus tolerans was not isolated from any of the sites. One theory may be that *A. tolerans* dispersal is less prevalent in these areas sampled. Since *A. tolerans* was isolated from geothermal regions in both Hawai'i and nearby in Northern California, having no dispersal to the regions sampled is unlikely [1]. Given that our MPN was conducted at 60°C, it is also possible that the conditions for *A. tolerans* to grow was not met.

Some of the potentially novel isolates, isolated from different locations, are likely to be within the same species. The MDF2, RED7, and RED4 novel isolates, for example,

clustered together in the Figure 7 phylogeny, likely being the same species from two different locations. The larger clustering of novel isolates from DF, M, MDF, RED, and TA in the Figure 7 phylogeny likely also represent a separate species isolated from five different locations. However, no novel isolates were found from the SF location which may provide insight to the dispersal or availability of these species in the regions sampled. Further research would need to be performed to assess if these isolates under 97% identity during 16S rRNA database search discovered during this study are truly novel. Given the large number of novel isolates, similar studies of soil enrichments, at low pH and high temperature environments may produce the isolation of other novel species, previously undescribed.

The high percentage of identity between the group of thirteen of the novel isolates (DF3, DF4, DF5, M1, M2, M5, M6, M9, M10, M12, MDF8, RED6, and TA5) suggests this likely represent one to three potentially novel species. The grouping of three of the novel isolates (MDF2, RED4, and RED7) with high percentages of identity, as well as the individually grouped TA10 isolate, may also represent new, undescribed species. The larger grouping of 13 novel species in the Figure 7 phylogeny was from five different locations, whereas the smaller grouping of three was from two different locations. These findings suggest that multiple potentially novel species are present with widespread dispersal to these geographically distinct sites. All of the bootstrap values for the major nodes were above 66. However, some of the closer related, node tips had lower values which likely means there were slightly different placements for these isolates on these tips in different replicate trees.

Based on the study, the M location appeared to have the most diversity from any location with the highest richness, Shannon-Wiener Index value, and evenness. However, these values were based on every species under 97% identity being individual species when there is a chance, they could be part of the same species. In addition, the small sample size for each location, these values are not the most precise, but may provide a general summary of diversity by location. The second most diverse location, based on richness, Shannon-Wiener Index, and evenness, was the DF site. This site is the closest geographically from the M location, being approximately two miles to the east. These results suggest that the Maple Creek region is diverse and abundant in thermoacidophilic endospores. However, the RED location had the highest MPN calculation suggesting that largest concentration of thermoacidophilic spores in this area. The RED location had a richness of five which, when compared with M or DF, is lower. The DF location had the second highest concentration of thermoacidophilic spores which was approximately 290 MPN/g less than the RED location.

CHAPTER 6: PHYLOGENETICS AND BIOGEOGRAPHY OF *ALICYCLOBACILLUS ACIDOCALDARIUS* WITHIN AND BETWEEN NON-PERMISSIVE ENVIRONMENTS

Introduction

Dispersal includes abiotic dispersal events such as glaciation or volcanic eruptions and climatic (wind, rain, etc.) dispersals [11], [17] derived from permissible habitats. Animals, insects, and human activities can also be a source of spore dispersal. However, non-natural permissive habitats can be suitable environments for growth of thermoacidophiles. Examples of non-natural permissive habitats include industrial processes that utilize high temperatures and low pH, such as rubber tire production, juice pasteurization, or composting. The pattern of dispersal discovered in this study may help identify factors that have influenced the dispersal of these species.

In this study I used MLST analysis to study the biogeography of *A. acidocaldarius* isolates from different locations in Humboldt and Trinity Counties, CA. The locations that the isolates came from are believed to be outside of the growth range for *A. acidocaldarius*, and thus these isolates were likely derived from spores that were dispersed to these sites. The sample sites may be affected by different dispersal mechanisms because they differ in elevation, location from the coast, and surrounding environments.

The dispersal mechanisms of thermophilic spores within cold bodies of water has been studied [2], [124]. One study found that currents, run-off, and effluents affected the dispersal of endospores within marine environments [2]. Other researchers have found nearly omnipresent representation of fungal spores in different environments throughout the globe due to wind currents [18], [19]. Across inclining elevations, a lower concentration of spores may be expected as the elevations incline due to the harsher conditions of higher elevation areas. Alternatively, these high elevations may increase dispersal due to snow, rain, wind, etc. resulting in a higher concentration of spores. Coastal environmental factors may also play an important role in the dispersal of these spores. Crashing waves, currents, and other marine activity can aerosolize materials potentially influencing marine-associated dispersal in these lower coast regions [124]. Seeing even abundance and identical genotypes of thermoacidophilic spores across gradients may represent consistent and ubiquitous dispersal, at least in the regional areas being sampled. If spore concentrations and genotypes vary significantly across sample sites, then environmental conditions in the sites may affect dispersal or persistence of the spores.

In this study, MLST analysis was used to assess genetic relationships between *A. acidocaldarius* isolates from various non-permissive habitats. Utilizing MLST can help provide better insight to genetic relations by comparing multiple portions of the genome. Genes were sequenced and combined to make one concatenated sequence for each individual isolate. These concatenated sequences were used to make phylogenies and percent identification matrices to compare nucleotide alignments and allele patterns of

the isolates between and within each location. Based on the patterns of shared or diverse genotypes I can infer whether environmental conditions of the sites affects dispersal.

The genes used in the MLST study were *emrB*, *gyrB*, *cpn60*, *gmp*, *rpoB*, and *eftu*. With the exception of *emrB*, these are housekeeping genes that are unlikely to undergo high rates of horizontal gene transfer due to their vital functions. The *emrB* gene encodes an efflux pump protein. The gene is associated with multidrug resistance mechanisms, most prominently studied in *Escherichia coli* [125]. The *emrB* locus has been known to protect cells from hydrophobic compounds and uncouplers of oxidative phosphorylation which interfere with the synthesis of ATP [125]. The *gyrB* gene encodes a Type II topoisomerase, known as DNA gyrase subunit B, which aids in DNA replication [126], [127]. The vital function of this gene means that it is widespread throughout bacterial species [127], [128]. Many studies have utilized this gene as a phylogenetic marker of species as the inferred rate of evolution is faster than that of 16S rRNA [63]–[65], [129]–[133]. The *cpn60* gene encodes a chaperonin, heat shock protein that can help prevent misfolding which is also known as *GroEL* or *HSP60* [66], [134]. This gene has been found within bacteria, eukaryotes, and even some archaea [135]. The *cpn60* gene has also been utilized to define phylogenies for identification of organisms similar to how the 16S rRNA gene is utilized [62], [66]–[68], [136]–[139]. The *gmp* gene encodes guanosine monophosphate synthetase which is involved in purine biosynthesis and the guanosine pathway [140]. The *rpoB* gene encodes the RNA polymerase B-subunit and is a housekeeping gene commonly used in phylogenetic studies [13], [69]. This gene has also

been studied for its involvement with rifampicin resistance due to mutations of the gene [141]. The *eftu* gene encodes the elongation factor-Tu which is a protein that aids in the binding of aminoacyl-tRNA to the ribosome during elongation in translation [142]. The role of *eftu* makes it a highly conserved gene within species [142]. The *keto* gene encodes the trans*keto*lase enzyme which is responsible for catalyzing the cleavage of carbon-carbon bonds from ketose sugars [143], [144]. The enzyme uses thiamine pyrophosphate as a cofactor [143].

Materials and Methods

Alicyclobacillus acidocaldarius strains were restreaked from previous plates (Chapter 3, Chapter 4, Chapter 5) or frozen stocks on 3.5 pH PTYG plates until single colony isolates were obtained. Single colonies were then used to inoculate 50 ml of sterile 3.5 pH PTYG broth in 250 ml flasks. Flasks were left at 50°C at 100 rpm in a shaking incubator for 12 to 48 hours, depending on turbidity. Once sufficient growth was reached, one to 1.5 ml were pipetted into sterile, labeled, two ml tubes and centrifuged at 13,000 x g for two to five minutes. This was repeated until sufficient cell pellets were obtained, and tubes were stored at -18°C.

Genetic sequencing of Alicyclobacillus acidocaldarius isolates

Cell pellets were used for DNA extractions, following the protocol from the Promega Wizard Genomic DNA Purification Kit for *Isolating Genomic DNA from Gram Positive and Gram Negative Bacteria*. The DNA concentration was then determined for each sample using a NanoDrop spectrophotometer. Gel electrophoresis was performed to

ensure high molecular weight DNA had been extracted. Thirty-five µl PCR reactions were performed using 16.5 µl of Lucigen EconoTaq Plus 2x Mastermix, 16.2 µl of sterile water, 0.15 µl of each forward and reverse primer and two µl of DNA template at three ng/µl. The procedure was performed in a sterile PCR hood with all equipment wiped down with 70% ethanol before use. The PCR procedure was run in a Thermal Cycler (MJ Research PTC-100 Thermal Cycler or Applied Biosystem - 2720 Thermal Cycler PCR). The steps included 95°C for four minutes, 94°C for one minute, the specific gene annealing temperature for one minute (Table 11), 72°C for one minute, run 30 to 35 cycles of steps two to four, seven minutes at 72°C, and then 24+ hours of 4°C.

Table 11. Primers, annealing temperatures, and PCR product lengths of genes used in the multilocus sequence typing analysis of *A. acidocaldarius* isolates.

Gene	Annealing Temp (°C)	Primers (5'-3')	Length (bp)
emrB	52	F: GAGCGAACTCACGAACT R: GGTCTTGGCGATGGGTGTTT	700
cpn60	48	F167: GACGGTGTGACCATCGC R1155: GCTTCTTCTCCTTSAGCTC	980
gyrB	50	F: TTCGTCCGCTGGCTAAATCA R: GTCACGTCCATCTTCTTGCT	527
eftu	50	F: GCKGACTAYGTGAARAACATGAT R: MGCACCGATGTTGTCACC	575
rpoB	48	F: GGGYGARGTSAARGARCAAG R: CRATRCGGAAYTGGTTYTGC	875
gmp	47	F: GGYATTTGYTACGGMATG R: GTGAARTACTGCCAGATYTC	980
16S rRNA	57	F: AGAGTTTGATCCTGGCTCAG R: AAGGAGGTGATCCAGCC	1,500

PCR products were then purified followed by Sanger sequencing as previously described (Chapter 3). Once returned, sequences were checked for quality, edited for accuracy, and then used to construct a maximum likelihood phylogeny as in Chapter 3.

Gene sequence analysis

Clipped gene sets were then combined into one file and used to concatenate to the other genes of the same isolate. For example, the isolate "SF1" in each gene set of *cpn60*, *eftu*, *gyrB*, *rpoB*, *emrB*, and *gmp* was called out and combined, respectively, to create one large sequence for the isolate. The R packages DECIPHER, Biostrings, and ape were used to perform this. These concatenations were then used to create phylogenetic trees using CIPRES [86]. The RAxML-HPC2 on XSEDE task was used to create a maximum likelihood tree with rapid bootstrapping. Default parameters were selected with data specific information added. A partition file was also added containing the separate gene location characters: *cpn60* from 1-875, *eftu* from 876-1,353, *gyrB* from 1,354-1,841, *rpoB* from 1,842-2,697, *emrB* from 2,698-3,313, and *gmp* from 3,314-4,286. A concatenation of the corresponding gene sequences of *cpn60*, *emrB*, *eftu*, *gmp*, *gyrB*, and *rpoB* from the *A. tolerans* isolate LTF1 from the study by Maghfiroh [1] was used to root the Figure 8 and Figure 9 phylogenies.

Results

Six MLST genes were sequenced for each of the 28 *A. acidocaldarius* isolates. The *emrB* gene from the 29th isolate, DF9, was unable to be obtained so this isolate was excluded in the six gene MLST phylogeny but included in the five gene MLST

phylogeny below. These 29 isolates were utilized to make the phylogenies for thermoacidophiles from non-permissive habitats. In addition to the 19 *A. acidocaldarius* strains obtained from the location and temperature MPN trials, 10 more strains previously isolated from other non-permissive locations including pasture soil (PS), football field rubber (FFR), and Spruce forest (SF), were utilized as well.

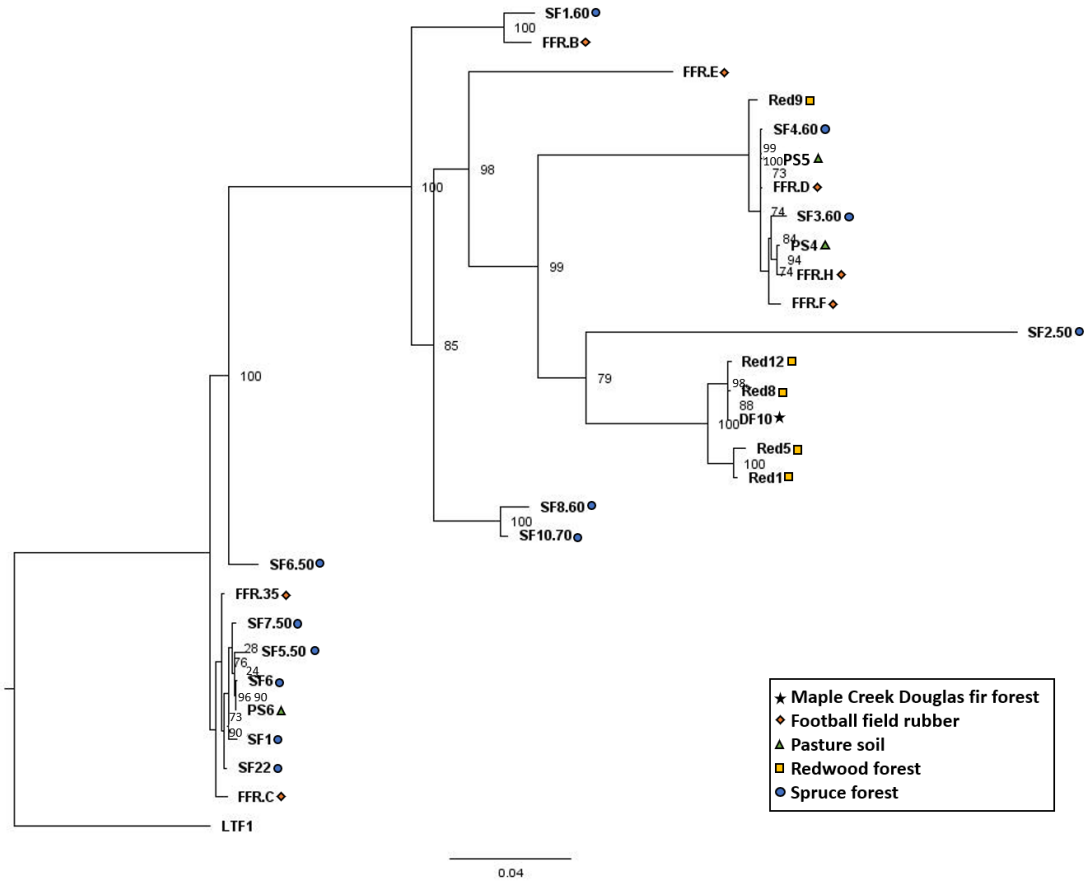

Figure 8. Multilocus sequence typing, rapid-maximum likelihood phylogeny of *A. acidocaldarius* isolates from soils using the *cpn60*, *gyrB*, *gmp*, *emrB*, *rpoB*, and *eftu* gene sequences with 100 bootstrap parameters across 4,648 nt.

The Figure 8 phylogeny shows several distinct nodes which provides evidence for isolates not grouping by their respective locations based on the concatenated sequences of *cpn60*, *gyrB*, *gmp*, *emrB*, *rpoB*, and *eftu*. Two areas on the phylogeny had several isolates from different locations, which while on multiple distinct nodes, clustered together with very low nucleotide diversity between the isolates (Group 1: FFR.35, SF7.50, SF5.50, SF6, PS6, SF1, SF22, and FFR.C; Group 2: RED9, SF4.60, PS5, FFR.D, SF3.60, PS4, FFR.H, and FFR.F). Four of the RED isolates (RED1, RED5, RED8, and RED12) shared a common node point, being closely related, but the fifth RED isolate, RED9, was distinct from the others being more similar to isolates from SF, PS, and FFR locations. The SF2.50 isolate was the most distinct from the other isolates but most closely aligned with DF10, 87.56% identical. The PS6 isolate did not highly match with either PS4 or PS5 but had almost identical alleles with SF6. A SF isolate and FFR isolate also shared a prominent, common node with each other (SF1.60 and FFR.B).

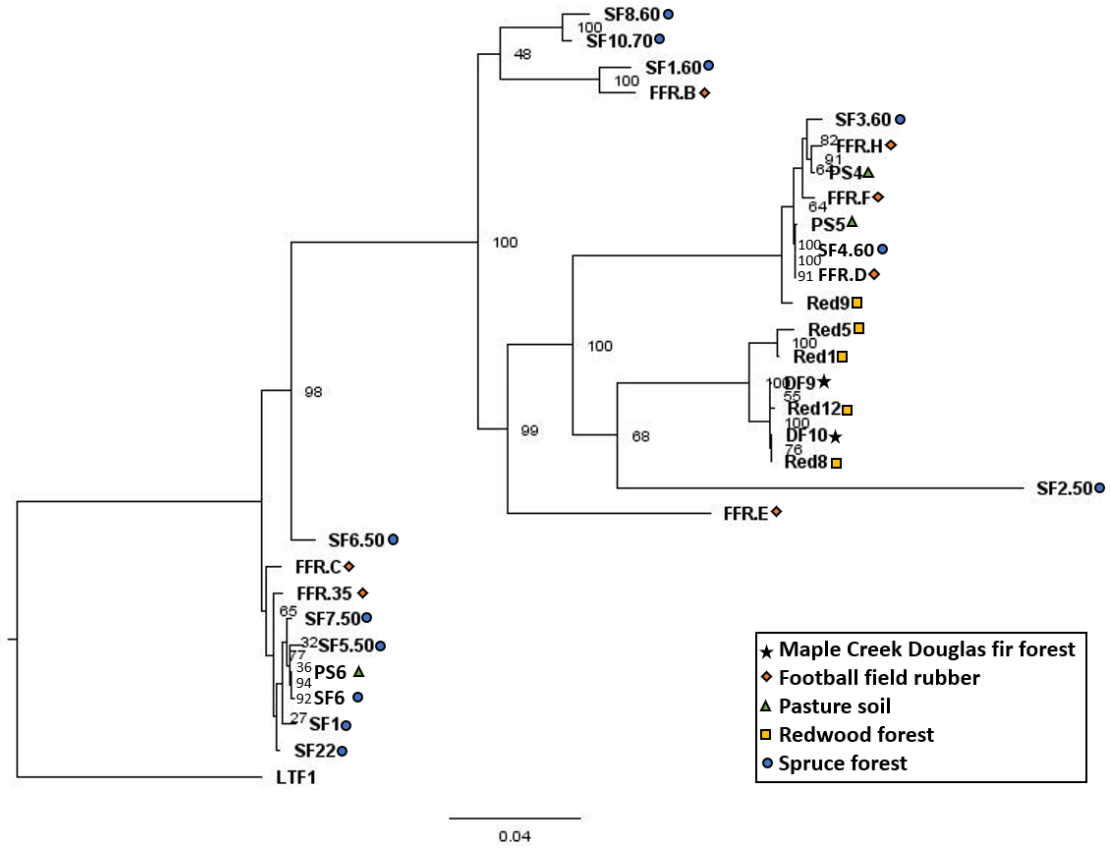

Figure 9. Multilocus sequence typing, rapid-maximum likelihood phylogeny of *A. acidocaldarius* isolates from soils using the *cpn60*, *gyrB*, *gmp*, *rpoB*, and *eftu* gene sequences including the DF9 isolate with 100 bootstrap parameters across 3,942 nt.

The Figure 9, MLST phylogeny with the DF9 isolate included the concatenated genes *cpn60*, *gyrB*, *gmp*, *rpoB*, and *eftu*. This phylogeny shows similar branch patterns to Figure 8, which included one additional gene, *emrB*. Notably, the DF9 isolate aligned closely to the other DF isolate in the group with RED12, DF10, and RED8. Without the *emrB* gene, the SF4.60 and FFR.D isolates had identical sequences to each other and nearly identical to PS5. Bootstrapping values for this phylogeny are, in general, slightly lower in comparison to Figure 8.

Discussion

In this study I aimed to study the biogeography of thermoacidophilic species by comparing six protein coding genes from *A. acidocaldarius* isolates obtained from geographically separate, non-permissive sites using MLST. Our results show that sampling origin of isolates does not determine the relationships in the phylogenies. Identical and nearly identical alleles from geographically separate locations were found. The phylogenies show distinct isolates from separate sampling sites grouping together providing evidence that these isolates have identical or nearly identical alleles despite geographical distance. Within an individual location there was both diversity and similarity, such as with the RED9 isolate sharing less than 91.00% identity to any other RED sample, but the RED12 and RED8 isolates sharing 99.78% identity between concatenated nucleotide sequences across the six genes.

The two phylogenies showed similar results of isolate relations despite comparing different gene groupings. The phylogeny without the *emrB* gene included a more closely shared clade between the SF1.60/FFR.B grouping and the SF8.60/SF10.70 grouping. However, this grouping had a low bootstrap value of 48 compared to the separate groupings from the six gene MLST phylogeny which had a value of 85. Additionally, some nodes had very low bootstrap support on both phylogenies, which is likely due to slight differences in the nucleotide sequences of these isolates to several others, which upon replication, cause subtle shifts on the phylogeny. Despite this, the major nodes of the six gene MLST phylogeny had strong bootstrapping values of 79 or greater. All other isolates appeared to fall on identical or nearly identical nodes and branchpoints within the

two phylogenies. The pairings of RED8/DF10, RED8/RED12, FFR.35/SF22, PS6/SF6/SF5.50, SF7.50/SF6, and SF4.60/FFR.D/PS5 isolates were nearly identical to each other across six different genes differing by less than one percent. Without the *emrB* gene the SF4.60 and FFR.D isolates were 100% identical. The majority of these pairings are from different locations but still comprise nearly identical alleles. All locations had isolates that both had low sequence nucleotide diversity between other isolates from the same location as well as between isolates from other locations.

Both phylogenetic trees show some distinct grouping between isolates. While some isolates from the same location appear to display patterns of grouping, many location isolates are dispersed throughout the tree with close relations to other location isolates. These results demonstrate dispersal of endospores from permissible locations to the regions sampled.

I hypothesized that there may be dispersal via unique mechanisms to distinct sampling sites. For example, coastal sites might be more affected by wave aerosols and high elevation, inland sites might be more affected by snowfall and wildfires. If source populations were distinct, then dispersal via unique mechanisms might result in a pattern of biogeographically defined diversity. However, I saw no evidence of this. The results are consistent with broad dispersal via mechanisms that are not significantly affected by local characteristics. These results are interesting because differing microbial species compositions have been found within various environmental and atmospheric samples [12], [17], [21], [25].

The RED location had the highest MPN concentration of thermoacidophilic spores by almost ten-fold more than the other locations. This high concentration would most likely be explained by isolated dispersal factors that are distinctly affecting the RED location. However, while the RED isolates were similar to each other, they were also highly similar to isolates from other locations suggesting that these locations have shared dispersal. It is unclear why the RED location would have a higher concentration of thermoacidophilic spores compared to the other locations but factors such as such as coastal fog, some historical anthropogenic activity, or other site-specific factors may play a role [19], [21], [22]. Alternatively, low pH microhabitats may be more prevalent in the Redwood forest.

I hypothesized that the isolates from the rubber granules of our campus football field would be genetically distinct to isolates from natural environments, because the rubber isolates might have originated from bacteria growing during industrial vulcanization processes. If these isolates did come from rubber processing activities, they might be distinct from isolates originating from geothermal environments. However, the rubber isolates were closely related to the isolates from natural environments, and the rubber isolates exhibited as much diversity as the isolates from natural environments. Thus, it may be that these isolates have accumulated from dispersal that occurred since the turf was last resurfaced in 2011.

In addition to the six other genes sequenced, I attempted to include the *keto* gene in the MLST analysis, however, I were unable to produce a useable PCR product for sequencing as multiple or no products were produced. This gene and primer pair required

gel purification from non-specific amplicons in the past study by Maghfiroh which studied species from the *Alicyclobacillus* genus [1].

While the *cpn60*, *gmp*, *eftu*, *gyrB*, and *rpoB*, PCR products and sequences were able to be obtained from the DF9 isolate, the *emrB* gene was unattainable. This likely means that this isolate is missing the *emrB* gene or that the primer binding site has mutated. The gene encodes an efflux pump protein associated with multidrug resistance, which is not a housekeeping function, and its absence would likely not be detrimental to the cell.

CHAPTER 7: PHYLOGENETICS AND BIOGEOGRAPHY WITHIN AND BETWEEN PERMISSIVE AND NON-PERMISSIVE ENVIRONMENTS

Introduction

This project aimed to assess if there are biogeographical patterns of *A. acidocaldarius* between habitats that are permissive for vegetative growth and non-permissive habitats. A prior study by Maghfiroh (2015) focused on *Alicyclobacillus* isolates from sites that are permissive for vegetative growth in LVNP and HVNP. That study found distinct alleles of *A. acidocaldarius* in distinct sites providing evidence of endemism and geographical differentiation [1]. However, in that same study, identical genotypes of *A. tolerans* isolates were discovered across all sites including between LVNP and HVNP suggesting on-going, large-scale dispersal of an isolate closely related to *A. acidocaldarius*. Collectively, these results suggest that *Alicyclobacillus* species are capable of widespread, on-going, rapid dispersal, but that such dispersal is not always capable of overcoming forces acting on local populations, such as genetic drift and natural selection. Alternatively, a recent, massive dispersal event of *A. tolerans* may have occurred, for example as might be expected from a volcanic eruption, but *A. acidocaldarius* was absent from the community involved in this event.

In permissive habitats where cells are actively growing and dividing, it would be expected that genetic drift would increase diversity within and across sites. Selection might increase or decrease diversity across these habitats but given the wide range of

environmental differences between hot spring habitats, it might be more likely that selection would increase diversity. However, the identical genotypic pattern seen with *A. tolerans* isolated from HVNP and LVNP suggests that selective sweeps across distant and environmentally disparate environments are possible.

I hypothesized that non-permissive sites might act as reservoirs of *Alicyclobacillus* spore dispersal events. Thus, by examining isolates from non-permissive sites I could determine if the genotypes from permissive sites had recently dispersed. In addition, I might be able to detect diverse thermoacidophiles from past dispersal events that are no longer prevalent in permissive sites because they have been displaced by other ecotypes. I hypothesized that spores from non-permissive sites might have a greater diversity of genotypes due to the lack of purifying selection upon accumulated spores.

I used MLST analysis to characterize *A. acidocaldarius* isolates from different non-permissive locations in Humboldt and Trinity Counties, CA and compared those sequences with data from *A. acidocaldarius* isolates from LVNP and HVNP. In addition, individual gene phylogenies of isolates from non-permissive and permissive locations were created to compare more of the isolates from LVNP and HVNP that lacked the collection of genes included in the MLST study.

Materials and Methods

The sequences from *A. acidocaldarius* isolates collected from Maghfiroh's research of permissive, hot-spring like habitats were used in this study [1]. For the MLST

phylogenies, 14 of these *A. acidocaldarius* isolates from permissive habitats were included in addition to the 29 *A. acidocaldarius* isolates. The gene sequences from the permissive habitats were limited by the availability from Maghfiroh's study [1]. The number of permissive isolates included in the individual gene trees also varied based on this gene sequence availability. Each set of gene sequences was aligned using MAFFT with the default settings and clipped using AliView. For the MLST analysis, clipped gene sets were then combined into one file and used to concatenate to the other genes of the same isolate. For example, the isolate "SF1" in each gene set of *emrB*, *eftu*, *gyrB*, and *rpoB* was called out and combined, respectively, to create one large sequence for the isolate. The R packages DECIPHER, Biostrings, and ape were used to perform this, and the example code is listed in the previous chapter.

Concatenations or gene files were used to create phylogenetic trees using CIPRES [86]. The RAxML-HPC2 on XSEDE task was used to create a maximum likelihood tree with rapid bootstrapping. Default parameters were selected with data specific information added. A partition file was also added for concatenated sequences containing the separate gene location characters: *rpoB* from 1-922, *eftu* from 923-1,477, *gyrB* from 1,478-1,989, and *emrB* from 1,990-2,687. The MLST tree was rooted with the *A. tolerans* isolate LTF1 from Maghfiroh's study. Isolates from the NCBI database, with shared alignment, were utilized to root the individual gene phylogenies: *rpoB, emrB,* and *eftu* with the respective portions of CP002902 isolate, *gyrB* with AB08984, *gmp* with CP001727, and *cpn60* with EF685187. Gene trees differ in permissive isolates as not all gene sequences were available for the isolates from permissive sites. Five of the individual gene trees for

cpn60, *eftu*, *gmp*, *gyrB*, and *rpoB* utilized the 29 non-permissive isolates collected from this study and from previous studies including locations SF, RED, DF, FFR, and PS. The *emrB* gene tree utilized only 28 non-permissive isolates because this gene was not available for DF9.

Results

Multilocus Sequence Typing Phylogeny

While none of the *A. acidocaldarius* isolates from non-permissive habitats shared 100% identity across all genes with *A. acidocaldarius* isolates from permissive habitats, several were nearly identical, sharing greater than 99% nucleotide identity. The a554b50H1 isolate from LVNP was nearly identical with greater than 99% shared identity with the non-permissive site isolates of PS6, SF6, SF22, and FFR.C. The SF1 and SF5.50 alleles were nearly identical to those from the LVNP permissive habitat, bu50H1, DKu50M1, and DKu50M2.

In other cases, isolates from non-permissive habitats shared a greater than 97% identity, but less than 99% identity, with isolates from permissive habitats. The isolate FFR.E shared greater than 97.16% identity with two LVNP isolates, cu50L1 and bu50L1. Both SF10.70 and SF8.60 shared greater than 98% identity with bu50M1 and bb50M1 isolates. The SF6.50 isolate shared at least 98.2% identity with five LVNP isolates (bb50H1, bu50M3T, a554b50H, bu50H1, and DKu50M1). The FFR.35 isolate shared greater than 98.6% identity with a554b50H, bu50H1, DKu50M1, and DKu50M2. The only two Hawai'i isolates, HVS1. HTF1 and HVS1.LTF1, shared identity across 2,673 of

the 2,687 nt used in the nucleotide alignment. The two Hawai'i isolates also shared greater than 98% identity with the isolates on the closest branchpoints in the phylogeny including the isolates from non-permissive habitats, SF8.60 and SF10.70, as well as the Lassen isolates bu50M1 and bb50M1.

Several of the clusters in the phylogeny were composed solely of isolates from the non-permissive habitats, but none of the clusters were composed solely of isolates from the permissive habitats. A large cluster of non-permissive isolates were distinctly grouped in the phylogeny including FFR.F, PS5, FFR.D, SF4.60, RED9, FFR.H, PS4, and SF3.60. Another included the isolates RED5, RED1, DF10, RED8, and RED12. The non-permissive isolate SF2.50 clustered separately from all other isolates.

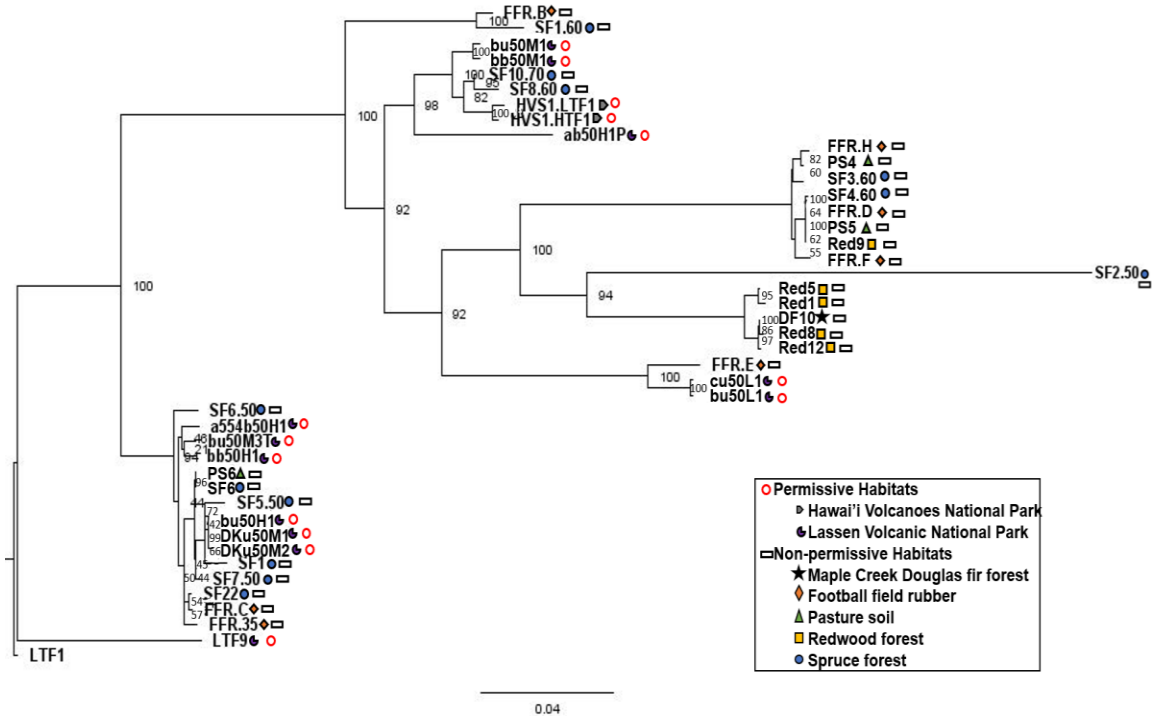

Figure 10. Multilocus sequence typing, rapid-maximum likelihood phylogeny of *A. acidocaldarius* isolates from permissive and non-permissive habitats using the *rpoB*, *gyrB*, *emrB*, and *eftu* gene sequences with bootstrapping (n=100) across 2,687 nt.

rpoB Gene Phylogeny

The *rpoB* gene phylogeny included 20 permissive isolates with two of these being from Hawai'i (HVNP) and 18 being from Lassen (LVNP) (Figure 11). Some of the *A. acidocaldarius* isolates from non-permissive habitats shared 100% identity across the *rpoB* gene with *A. acidocaldarius* isolates from permissive habitats. Notably, both the SF6 and PS6 *rpoB* genes were identical to three isolates from Lassen, bu50H1, Dku50M1, and Dku50M2.

In other cases, isolates from non-permissive habitats shared greater than 99% identity across the *rpoB* gene with isolates from permissive habitats. The two Hawai'i

isolates, HVS1.HTF1 and HVS1.LTF1 aligned closely with two SF isolates (SF10.70 and SF8.60) sharing greater than 99.11% identity. The Dku50H2 isolate clustered most closely with several DF and RED isolates (DF9, DF10, RED1, RED5, RED8, RED10, and RED12) sharing greater than 99% identity. The SF5.50, SF7.50, and SF22 isolates shared greater than 99% identity in their *rpoB* genes with six LVNP isolates (bu50H1, DKu50M1, DKu50M2, 554b50H1, bb50H1, and bu50M3T).

Several of the clusters in the *rpoB* phylogeny were composed solely of isolates from the non-permissive habitats, and other clusters were composed solely of isolates from the permissive habitats. The large cluster in the phylogeny contained isolates from three different non-permissive habitats, SF, PS, RED, and FFR. Another cluster included only isolates from the permissive LVNP sites (LTF9, B18, B19, and cu50L2).

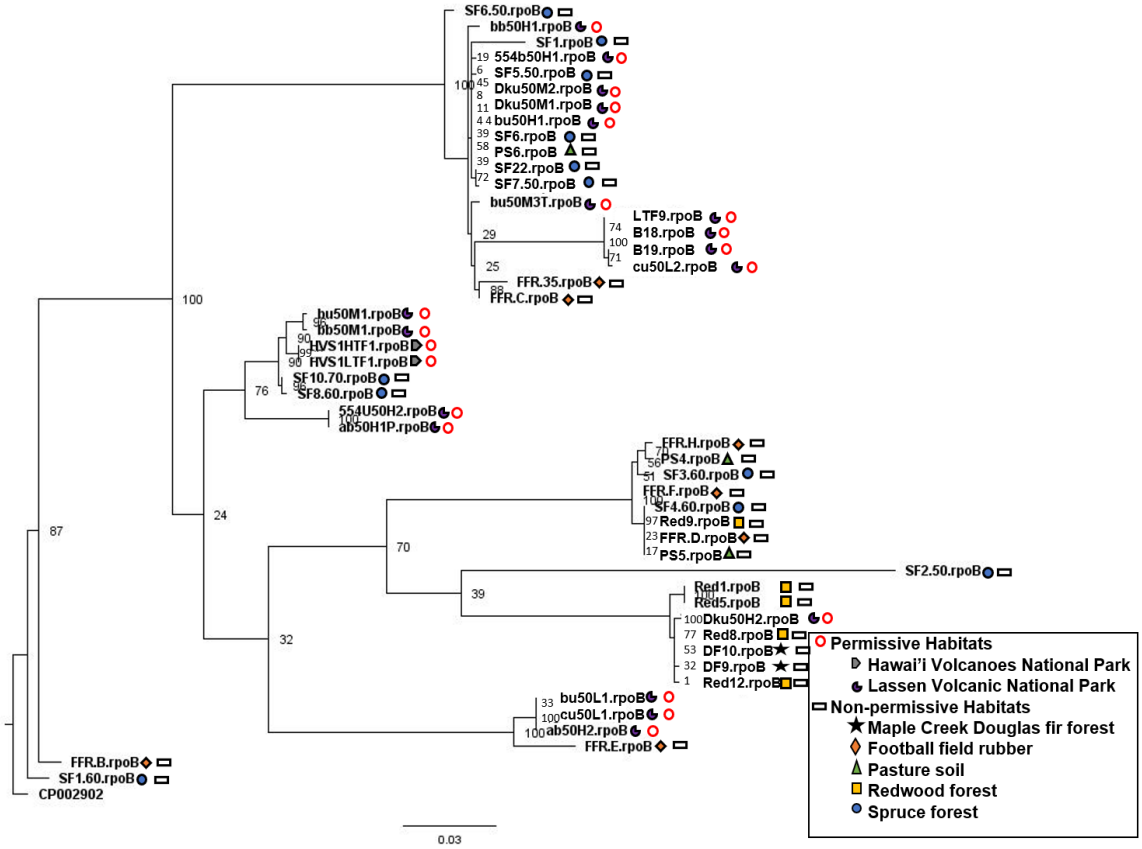

Figure 11. Maximum likelihood phylogeny using the *rpoB* gene sequences of *A. acidocaldarius* isolates from permissive and non-permissive habitats with bootstrapping (n=100) across 729 nt.

gyrB Gene Phylogeny

The *gyrB* phylogeny contained 21 isolates from permissive habitats with two from HVNP and 19 from LVNP (Figure 12). Some of the *A. acidocaldarius* isolates from non-permissive habitats shared 100% identity across the *gyrB* gene with *A. acidocaldarius* isolates from permissive habitats. The DKU50H2 isolate had identical *gyrB* alleles with DF9 and RED12. The SF7.50 isolate had an identical *gyrB* gene sequence to the two isolates from the LVNP permissive habitat (BU50H1 and DKU50M1).

In other cases, isolates from non-permissive habitats shared greater than 99% identity across the *gyrB* gene with isolates from permissive habitats. The SF7.50 isolate shared 99.55% identity with DKu50M2. The LVNP isolates BB50H1, BU50M3T, and 554b50H1 also aligned closely with isolates from non-permissive habitats at greater than 99.11% shared identity (SF22, FFR.C, FFR.35, and SF6.50).

Several of the clusters in the *gyrB* phylogeny were composed solely of isolates from the non-permissive habitats, and other clusters were composed solely of isolates from permissive habitats. Four isolates from LVNP clustered closely in the phylogeny (ab50H1P, 554b50L2, AU50M1, and 554U50H2).

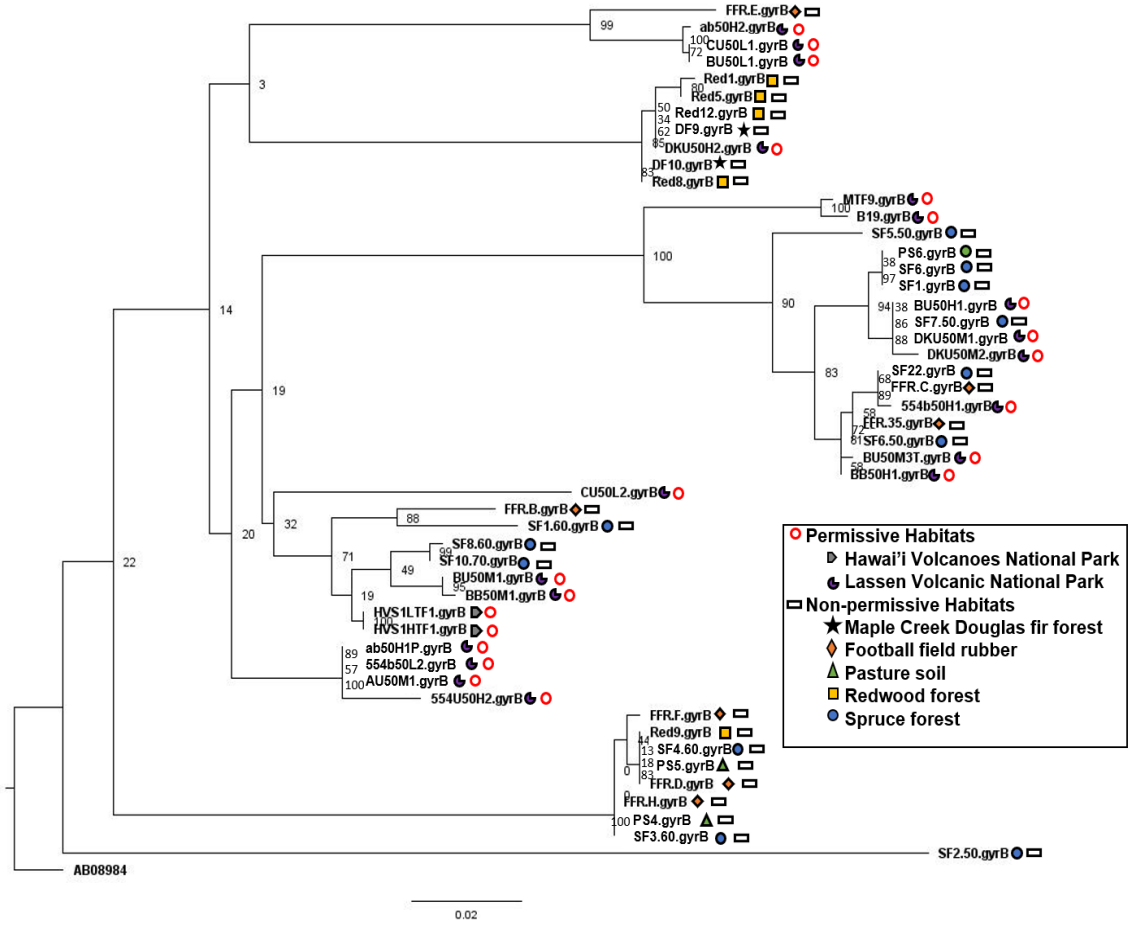

Figure 12. Maximum likelihood phylogeny using the *gyrB* gene sequences of *A. acidocaldarius* isolates from permissive and non-permissive habitats with bootstrapping (n=100) across 448 nt.

emrB Gene Phylogeny

The *emrB* phylogeny utilized 21 isolates from permissive habitats, two of which were from HVNP and the rest being from LVNP (Figure 13). Some of the *A. acidocaldarius* isolates from non-permissive habitats shared 100% identity across the *emrB* gene with *A. acidocaldarius* isolates from permissive habitats. The isolates SF1, SF5.50, BU50H1, DKU50M2, and DKU50M1 shared identical *emrB* alleles.

In other cases, isolates from non-permissive habitats shared greater than 99% identity across the *emrB* gene with isolates from permissive habitats. The two isolates from Hawai'i (HVS1.LTF1 and HVS1.LTF2) had high identity with SF8.60 and SF10.70 sharing greater than 99.33%. Six non-permissive isolates shared 99.33% identity with BB50H1 (FFR.35, FFR.C, PS6, SF6, SF7.50, and SF22).

Several of the clusters in the *emrB* phylogeny were composed solely of isolates from the non-permissive habitats, and other clusters were composed solely of isolates from the permissive habitats. Two distinct clusters of non-permissive isolates in the *emrB* phylogeny included one group involving SF3.50 and another group involving RED5. The group with the B18 isolate and the group with the 554b50L2 isolate were two clusters of only permissive habitat isolates.

The SF2.50 isolate was again distinct from the rest of the isolates, having the highest percent of nucleotide differences compared to the other isolates. Low bootstrapping values were present in the Figure 13 tree at the tips and for some major nodes throughout the phylogeny. However, high percent sequence identity between gene sequences was still present.

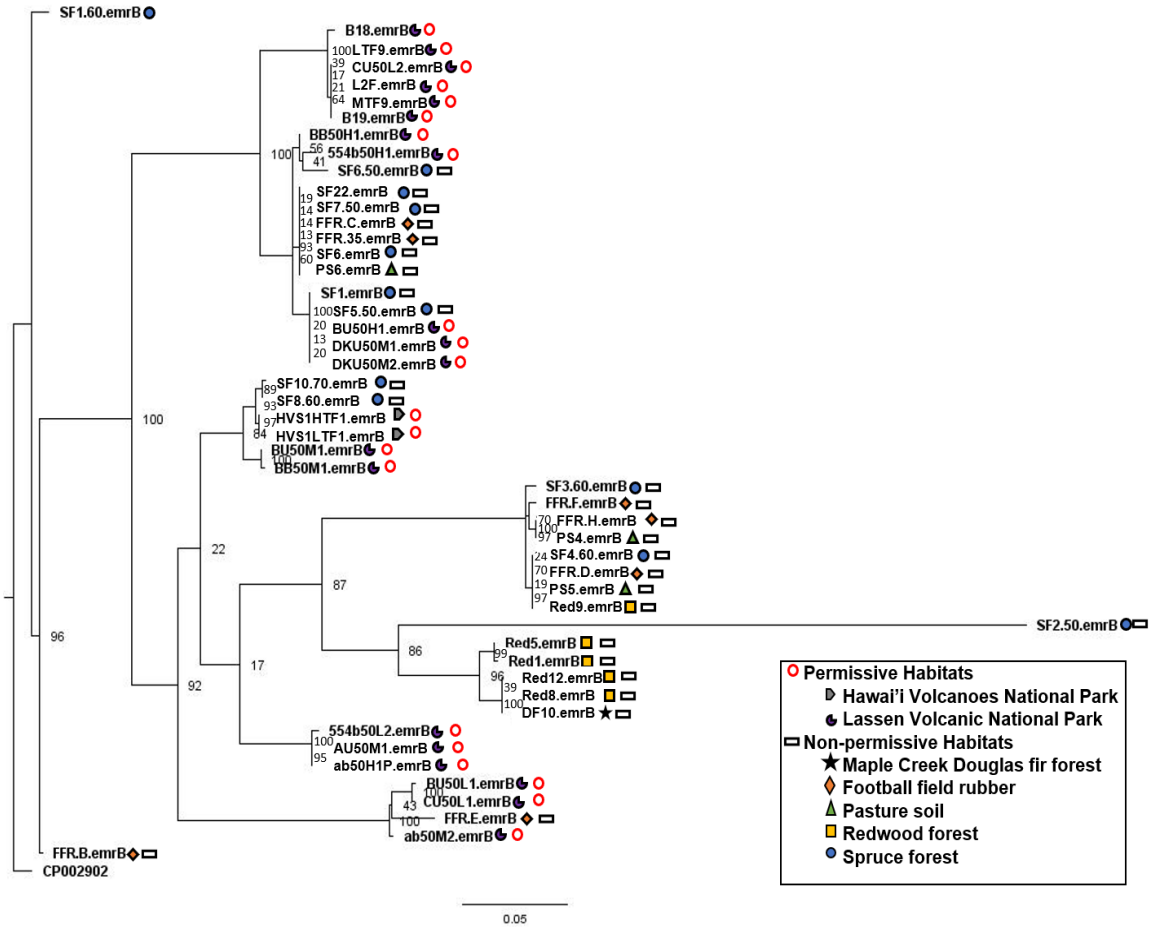

Figure 13. Maximum likelihood phylogeny using the *emrB* gene sequences of *A. acidocaldarius* isolates from permissive and non-permissive habitats with bootstrapping (n=100) across 599 nt.

eftu Gene Phylogeny

Twenty-one permissive isolates were used in the *eftu* gene tree with two from Hawai'i and 19 from Lassen sites (Figure 14). Some of the *A. acidocaldarius* isolates from non-permissive habitats shared 100% identity across the *eftu* gene with *A. acidocaldarius* isolates from permissive habitats. A large group of permissive and non-permissive isolates had identical alleles from locations SF, FFR, PS, and LVNP (SF1, FFR.C, DKU50M2, SF22, BU50H1, DKU50M1, FFR.35, SF7.50, SF5.50, PS6, and

SF6). The HVS1.LTF1 and HVS1.HTF1 isolates from HVNP had identical *eftu* alleles with SF8.60. The DKU50H2 isolate clustered closely with the non-permissive RED and DF isolates sharing identical *eftu* gene sequences (DF9, DF10, RED5, RED8, and RED1).

In other cases, isolates from non-permissive habitats shared greater than 99% identity across the *eftu* gene with isolates from permissive habitats. The SF8.60 and SF10.70 isolates were nearly identical to two isolates from LVNP (BU50M1 and BB50M1) across at least 474 of 477 nt.

Several of the clusters in the *eftu* phylogeny were composed solely of isolates from the non-permissive habitats, and other clusters were composed solely of isolates from the permissive habitats. A distinct cluster of isolates from permissive habitats is present in the phylogeny involving 554b50L2, 554U50H2, AU50M1, and AB50H1P. The FFR.B and SF1.60 isolates from non-permissive habitats cluster together in the phylogeny. A second grouping of non-permissive habitat isolates consisting of eight isolates is also present in the tree (FFR.D, FFR.F, FFR.H, PS4, PS5, RED9, SF3.60, and SF4.60).

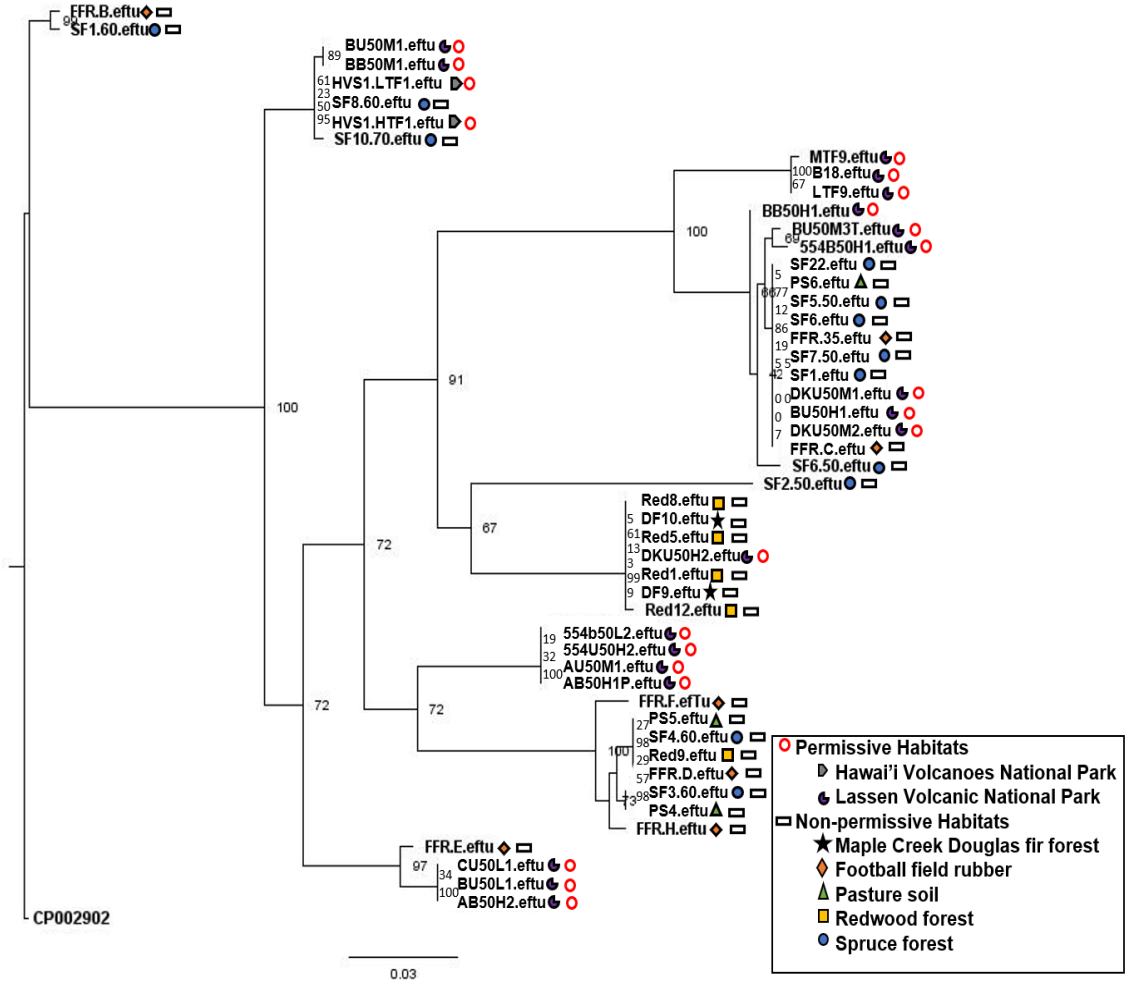

Figure 14. Maximum likelihood phylogeny using the *eftu* gene sequences of *A. acidocaldarius* isolates from permissive and non-permissive habitats with bootstrapping (n=100) across 477 nt.

gmp Gene Phylogeny

The *gmp* phylogeny utilized ten genes from isolates of permissive habitats with eight from Lassen and two from Hawai'i (Figure 15). While none of the *A. acidocaldarius* isolates from non-permissive habitats shared 100% identity across the *gmp* gene with *A. acidocaldarius* isolates from permissive sites, some shared greater than 99% identity. The HVS1.LTF1 and HVS1.HTF1 isolates shared high percent identity at

greater than 99% to SF6.50, SF8.60, and SF10.70, clustering closely on the phylogeny. The LVNP isolate B18 aligned closely with several non-permissive isolates from SF, PS, and FFR (FFR.C, FFR.35, SF22, SF1, SF7.50, SF6, PS6, and SF5.50) sharing greater than 408 across 414 nt.

Several of the clusters in the *gmp* phylogeny were composed solely of isolates from the non-permissive habitats, and other clusters were composed solely of isolates from the permissive habitats. Two clusters on the tree had only non-permissive isolates including the group with RED5 and the group with FFR.H. The L2F, B19, and MTF9 isolates from the permissive site of LVNP also clustered together on the phylogeny.

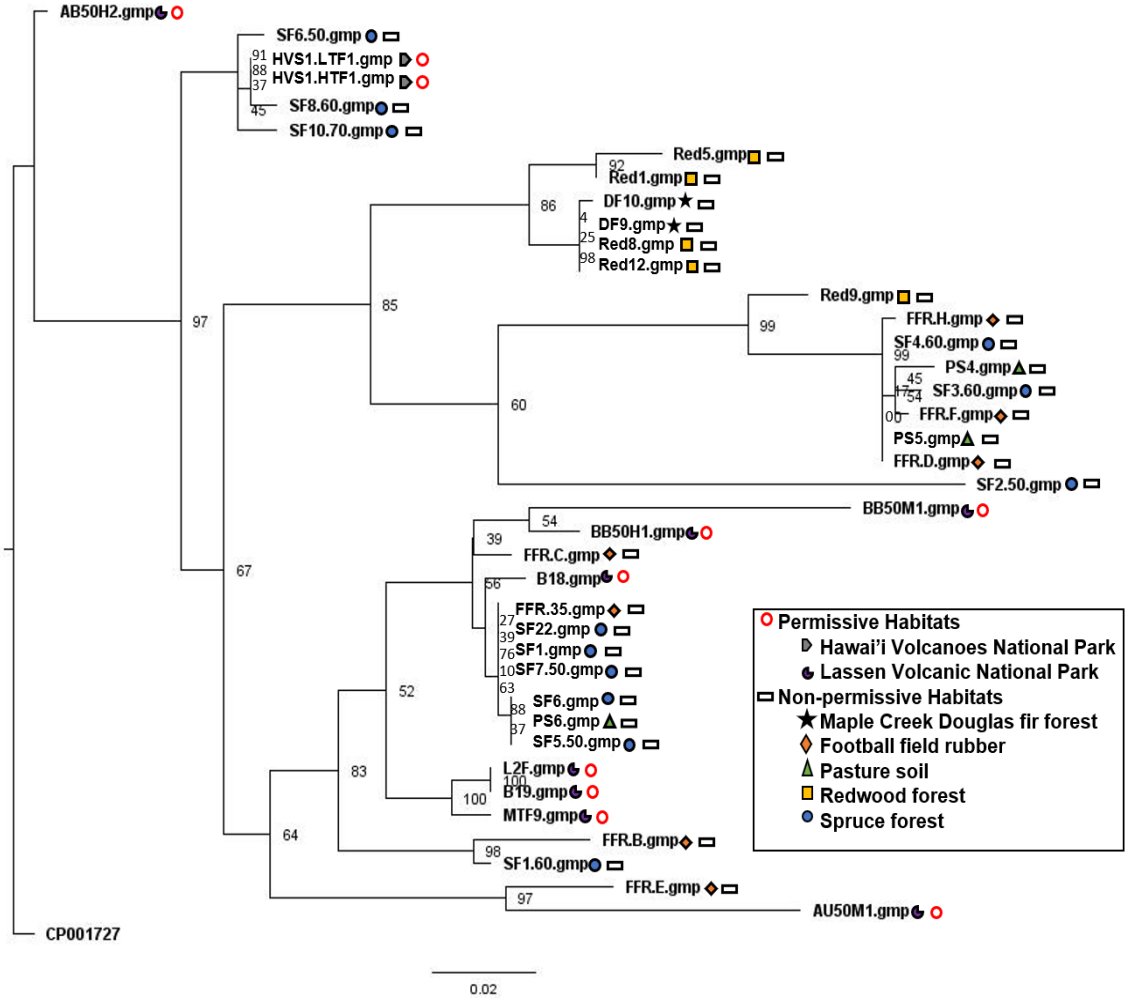

Figure 15. Maximum likelihood phylogeny using the *gmp* gene sequences of *A. acidocaldarius* isolates from permissive and non-permissive habitats with bootstrapping (n=100) across 414 nt.

cpn60 Gene Phylogeny

For the *cpn60* phylogeny only six sequences from permissive isolates were available with only one being from HVNP and the other five from LVNP (Figure 16). While none of the *A. acidocaldarius* isolates from non-permissive habitats shared 100% identity across the *cpn60* gene with *A. acidocaldarius* isolates from permissive sites, some shared greater than 99% identity. The HVS1.HTF1 isolate was closely aligned with

SF8.60 sharing 99% identity. The B18 isolate was most closely aligned with an isolate from a non-permissive site, FFR.C, sharing 492 of 494 nt.

Several of the clusters in the *cpn60* phylogeny were composed solely of isolates from the non-permissive habitats, and other clusters were composed solely of isolates from the permissive habitats. Two clusters of non-permissive isolates were present in the phylogeny involving the RED12 and FFR.H isolates. In addition, one group of only permissive isolates was present involving the B19 isolate.

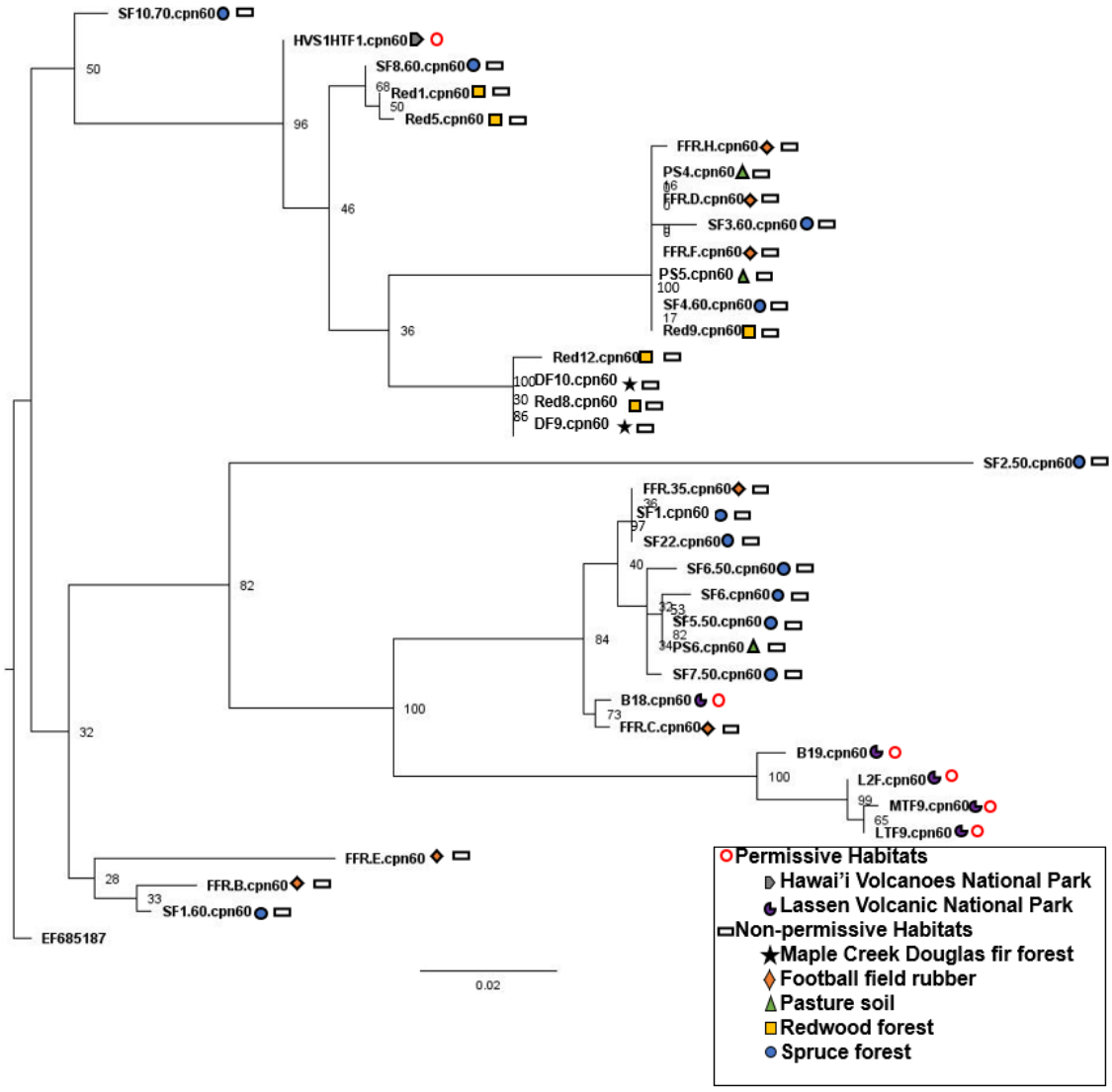

Figure 16. Maximum likelihood phylogeny using the *cpn60* gene sequences of *A. acidocaldarius* isolates from permissive and non-permissive habitats with bootstrapping (n=100) across 494 nt.

Discussion

None of the groups of isolates sharing 100% identity across all genes had members from both non-permissive and permissive habitats in the MLST analysis. The isolates from non-permissive habitats SF6 and PS6 isolates were identical as well as the

PS5, RED9, and FFR.D isolates. The DKu50M1 and bu50H1 were the only identical isolates from permissive habitats in the MLST analysis. However, many isolates from these opposing habitats shared greater than 99% identity across all genes. For example, the Lassen isolates including a554b50H1 shared greater than 99% identity with isolates from the SF and FFR while the SF1 isolate from the SF location shared high identity (>99%) with several LVNP isolates (bu50H1, DKu50M1, DKu50M2). For each individual gene that I examined, I found identical or nearly identical alleles in isolates from both non-permissive and permissive habitats. For example, the two Hawai'i isolates (HVS1.HTF1 and HVS1.LTF1) consistently shared high allele identity to SF10.70 and SF8.60, two isolates from a non-permissive location. In addition, the *eftu* genes of the Hawai'i isolates (HVS1.LTF1, HVS1.HTF2), two LVNP isolates (BU50M1 and BB50M1) and one SF isolate (SF10.70) were nearly identical across 475 of 477 nt. The individual gene trees, in general, provided evidence for identical alleles between permissive and non-permissive habitats. The Lassen site isolates shared identical gene sequences with isolates from the non-permissive locations SF, DF, FFR, RED, and PS. These results further suggest that there is dispersal between Hawai'i, Lassen, SF, DF, FFR, RED, and PS. The different sites of the non-permissive habitats are likely involved in similar dispersal patterns which may introduce both similar and different genotypes to the regions. This may explain why identical or highly similar alleles were seen between different non-permissive sites as well as within them. Selection against some genotypes may be occurring in the permissive sites, which may explain why some isolates from non-permissive habitats are not present in the permissive sites. The existence of identical

alleles and nearly identical isolates in both permissive and non-permissive environments indicates that these habitats are not genetically isolated from one another. There must have been recent dispersal that connected the non-permissive sites with both the HVNP and the LVNP sites. Because the HVNP site is approximately 4,000 km from LVNP, and because the non-permissive sites spanned a range of habitats that were separated by many kilometers, this dispersal must be taking place on a rapid and massive scale.

The genetic analysis also shows diversity within the non-permissive and permissive sites. Although all of the isolates were identified as *A. acidocaldarius* on the basis of 16S rRNA gene sequences, I detected significant diversity in protein-coding genes in isolates from a single site. Within permissive sites, this diversity might reflect ecotype diversification, or it might be neutral variation, or it may be a consequence of HGT. Within non-permissive sites, this diversity might arise if multiple distinct genotypes were dispersing from a given site at a given time, or if different genotypes were dispersing from a given site at multiple different times and accumulating in the non-permissive site, or if multiple distinct genotypes were dispersing from multiple sites at multiple times.

The genetic analysis also shows diversity between locations of both the non-permissive and permissive sites. That is, although all of the isolates were identified as *A. acidocaldarius* on the basis of 16S rRNA gene sequences, I detected significant diversity in protein-coding genes from non-permissive sites that was not present in permissive sites. To a lesser extent, there was some diversity in the genes from permissive sites that was not captured in the analysis of genes from non-permissive sites. The unique protein-

coding genes in permissive sites might be variation that has arisen between dispersal events, or it might represent diversity that would be identified in non-permissive sites if more sampling were to take place. The unique protein-coding genes in non-permissive sites might be variation from past dispersal events from populations that evolved prior to current populations in permissive sites. Alternatively, this diversity might reflect dispersal from multiple sites that exhibited local variations.

If there were distinct dispersal sources for each location, alleles from each different region would share high identity within each location, clustering together on the phylogenies by location. On the other hand, alleles may have low identity within locations as well as between locations, representing complex patterns of diversification and dispersal.

If allopatric effects or endemism were occurring, I would expect our permissive site alleles to be most highly related to other samples from the same location and not to any other permissive site locations or non-permissive site locations. The individual gene and MLST phylogenies had clusters with isolates from only non-permissive site and only permissive sites. For example, within the *cpn60* phylogeny two groups of only non-permissive habitat isolates were present, one involving RED and DF, and the other involving FFR, PS, SF, and RED. A group of only permissive habitat isolates was also present in the *cpn60* phylogeny involving different Lassen isolates. These results show that evolutionary factors such as genetic drift or mutations may be able to create endemic populations within the permissive sites.

While there were two groups of non-permissive isolates clustered within the MLST phylogeny, isolates from non-permissive sites also clustered closely with those from permissive habitats. However, the RED isolates clustered solely with other non-permissive samples in the MLST phylogeny. The RED region may be less involved in the dispersal paths from the permissive areas of LVNP and HVNP sampled, or this pattern may be due to a seeding event. Interestingly, many of the isolates from the FFR site, located approximately one km from the RED location, grouped with permissive habitat isolates. However, the individual gene trees demonstrate evidence that the RED site is included in the broad dispersal. For example, two RED isolates RED1, RED5, and RED8 shared identical *eftu* alleles to the Lassen DKU50H2 isolate.

Some of the phylogenies had low bootstrapping values at the tips and for some major branchpoints. However, high percent sequence identity between gene sequences in clusters was still present in these phylogenies. While portions of the trees may not be significantly supported, the high percent identity between isolates demonstrates genetic relations.

CHAPTER 8: SUMMARY

The purpose of our study was to assess the biogeography of thermoacidophilic bacterial spores in soils from different non-permissive locations. I developed and applied methods to enrich and quantify thermoacidophiles, and I isolated and characterized thermoacidophiles using genetic approaches. Isolates determined to be *A. acidocaldarius* on the basis of 16S rRNA genes were used in a comparative MLST study with *A. acidocaldarius* isolates from a previous study focused on permissive habitats [1].

I enriched acidophiles at 30°C, 40°C, 50°C, 60°C, and 70° and characterized the resulting isolates by sequencing their 16S rRNA genes. Twelve different bacterial species were isolated, and several other isolates shared less than 97% identity with characterized strains. Using GLM analysis I found that 60°C was the most likely to enrich for *A. acidocaldarius*.

I used an MPN approach to determine which of three media was most effective for enriching thermoacidophiles. Thermoacidophilic species grew in each of the three media tested—K, PTYG, and PD—at a concentration of at least one to 460 per gram. The PTYG medium was selected for further MPN enrichment based on the medium having the highest MPN values. Differences in the nutrients of the media may cater to specific species as all of the isolates from PD were novel. In addition, the lack of ability to grow the K isolates on the PTYG plates utilized suggests the potential for more novel isolates. When the PTYG-based MPN was used to assess the concentration of thermoacidophile spores in soils from five different regions in northern California, I detected spores in all

soils tested at a range of approximately 10-500 spores/g of soil. The Redwood forest location had the highest detectable concentration of thermoacidophilic endospores with approximately 10-fold that of other locations. The presence of easily enriched spores in all soils studied suggests that massive spore banks of thermoacidophiles are distributed across regions that are inhospitable to their vegetative growth. The uneven abundance and diversity of spores across locations suggests that uneven dispersal is occurring, perhaps due to site-specific environmental factors such as those researched in several studies [19], [21], [22].

To assess the pattern of alleles present, MLST analysis was first performed on *A. acidocaldarius* isolates of non-permissive habitats. Our results indicated that the sampling origin does not determine the relationships in the phylogenies, as isolates from different non-permissive locations grouped together having identical or nearly identical alleles. Both diversity and similarity were present within and between locations demonstrating some shared dispersal connecting the habitats sampled.

The alleles from permissive and non-permissive *A. acidocaldarius* isolates were compared through both MLST and individual gene tree analysis. Identical or near identical alleles were found between and within permissive and non-permissive habitats, providing evidence for shared dispersal between geographically distant, non-permissive habitats as well as between non-permissive and permissive habitats. For example, a Lassen isolate (a554b50H) shared greater than 99% identity with an isolate from the SF (SF1) non-permissive site in the MLST phylogeny. These patterns were also seen within the individual gene tress such as with the two Hawai'i isolates (HVS1.HTF1 and

HVS1.HTF2) sharing identical *emrB* alleles with a SF isolate (SF8.60). The mixed clustering of opposing habitats, only permissive sites, and only non-permissive sites throughout the phylogenies further represent broad ranged and on-going spore dispersal between these regions, as well as distinct dispersal factors that may involve only certain sites. Unlike several recent studies, evidence for geographical dispersal limitations were not identified in our study [4], [8], [9], [12], [17]. However, this field is evolving, and it is an open question whether dispersal limitations differ for different bacteria in different sites.

Additionally, several potentially novel isolates were collected from this study. The 16S rRNA sequences of these isolates shared less than 97% identity with sequences in Genbank from characterized isolates. This included nine of the 42 isolates from the temperature trial, seven of the ten isolates from the media trial, and 17 of the 50 isolates from the location trial. Of the total 33 novel isolates, five were under 94%. However, some of the novel isolates are likely to be a part of the same species, having high identity between sequences. Clustering of novel isolates from different locations was also present in the phylogenies, suggesting the potential for shared dispersal between these sites. Studies enriching soil for species from non-permissive habitats may provide conditions for the isolation of previously uncultured, novel species. The high return of novel species may expose a need for more studies of low pH and high temperature conditions. In addition, the large collection of novel species may represent dispersal from understudied permissive areas. Studying endospores from non-permissive locations may provide a greater diversity of species due to their lack of influence by evolutionary factors that may

provide the isolation of species that are now competitively subordinate, extinct, or survive at very low abundance in permissive habitats. Furthermore, moderately extreme environments have likely been under sampled relative to those that are more extreme leaving the potential for new insights when studying these habitats. Our results emphasize the potential for discovery and diversity that spore banks have to offer given their prevalence and lack of genetic alterations compared to their vegetative cells.

REFERENCES

[1] M. Maghfiroh, “Horizontal gene transfer and biogeography of isolated thermal acidic bacteria communities (Unpublished master’s thesis). Humboldt State University, Arcata, California,” 2015.

[2] E. Bell, L. I. Blake, A. Sherry, I. M. Head, and C. R. J. Hubert, “Distribution of thermophilic endospores in a temperate estuary indicate that dispersal history structures sediment microbial communities,” *Environ. Microbiol.*, vol. 20, no. 3, pp. 1134–1147, Mar. 2018.

[3] C. A. Hanson *et al.*, “Historical factors associated with past environments influence the biogeography of thermophilic endospores in arctic marine sediments,” *Front. Microbiol.*, vol. 10, no. FEB, 2019.

[4] A. L. Müller *et al.*, “Endospores of thermophilic bacteria as tracers of microbial dispersal by ocean currents,” *ISME J.*, vol. 8, no. 6, pp. 1153–1165, Jun. 2014.

[5] J. Green and B. J. M. Bohannan, “Spatial scaling of microbial biodiversity,” *Trends in Ecology and Evolution*, vol. 21, no. 9. pp. 501–507, Sep-2006.

[6] K. J. Locey and J. T. Lennon, “Scaling laws predict global microbial diversity,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 21, pp. 5970–5975, May 2016.

[7] L. G. M. B. Becking, “Geobiologie of inleideing tot de mileukunde,” *WP Van Stock. Zoon*, pp. 18–19, 1934.

[8] R. T. Papke, N. B. Ramsing, M. M. Bateson, and D. M. Ward, “Geographical isolation in hot spring cyanobacteria,” *Environ. Microbiol.*, vol. 5, no. 8, pp. 650–

659, Aug. 2003.

[9] W. J. Sul, T. A. Oliver, H. W. Ducklow, L. A. Amaral-Zettlera, and M. L. Sogin, "Marine bacteria exhibit a bipolar distribution," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 6, pp. 2342–2347, Feb. 2013.

[10] J. B. H. Martiny, J. A. Eisen, K. Penn, S. D. Allison, and M. C. Horner-Devine, "Drivers of bacterial β-diversity depend on spatial scale," 2011.

[11] M. J. Choudoir and D. H. Buckley, "Phylogenetic conservatism of thermal traits explains dispersal limitation and genomic differentiation of Streptomyces sister-taxa," *ISME J.*, vol. 12, no. 9, pp. 2176–2186, Sep. 2018.

[12] C. P. Andam, J. R. Doroghazi, A. N. Campbell, P. J. Kelly, M. J. Choudoir, and D. H. Buckley, "A latitudinal diversity gradient in terrestrial bacteria of the genus Streptomyces," *MBio*, vol. 7, no. 2, Apr. 2016.

[13] S. Das, H. R. Dash, N. Mangwani, J. Chakraborty, and S. Kumari, "Understanding molecular identification and polyphasic taxonomic approaches for genetic relatedness and phylogenetic relationships of microorganisms," *Journal of Microbiological Methods*, vol. 103. Elsevier, pp. 80–100, 2014.

[14] O. Prakash *et al.*, "Polyphasic approach of bacterial classification - An overview of recent advances," *Indian Journal of Microbiology*, vol. 47, no. 2. pp. 98–108, Jun-2007.

[15] J. D. Franke, J. A. Fuerst, and A. M. Poole, "Editorial: Structure, Function and Evolution of Complex Cellular Organization in Bacteria and Archaea," *Front. Microbiol.*, vol. 12, p. 751416, Aug. 2021.

[16] M. Dobro *et al.*, “Uncharacterized Bacterial Structures Revealed by Electron Cryotomography,” *J. Bacteriol.*, vol. 199, no. 17, Sep. 2017.

[17] M. J. Choudoir, A. Barberán, H. L. Menninger, R. R. Dunn, and N. Fierer, “Variation in range size and dispersal capabilities of microbial taxa.,” *Ecology*, vol. 99, no. 2, pp. 322–334, 2018.

[18] M. Huang and C. M. Hull, “Sporulation: How to survive on planet Earth (and beyond),” *Curr. Genet.*, vol. 63, no. 5, p. 831, Oct. 2017.

[19] J. K. M. Brown and M. S. Hovmøll, “Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease,” *Science*, vol. 297, no. 5581, pp. 537–541, Jul. 2002.

[20] S. Kim *et al.*, “Vortex-induced dispersal of a plant pathogen by raindrop impact,” *PNAS*, vol. 116, no. 11, pp. 4917–4922, 2019.

[21] H. Bauer *et al.*, “The contribution of bacteria and fungal spores to the organic carbon content of cloud water, precipitation and aerosols,” *Atmos. Res.*, vol. 64, no. 1–4, pp. 109–119, Sep. 2002.

[22] J. Aller, J. Y. Aller, M. R. Kuznetsova, C. J. Jahns, and P. F. Kemp, “The sea surface microlayer as a source of viral and bacterial enrichment in marine aerosols,” *Aerosol Sci.*, vol. 36, pp. 801–812, 2005.

[23] D. C. Blanchard and L. D. Syzdek, “Water-to-Air Transfer and Enrichment of Bacteria in Drops from Bursting Bubbles,” *Appl. Environ. Microbiol.*, vol. 43, no. 5, pp. 1001–1005, 1982.

[24] V. R. Després *et al.*, “Primary biological aerosol particles in the atmosphere: a

review," *http://dx.doi.org/10.3402/tellusb.v64i0.15598*, vol. 64, no. 1, 2012.

[25] B. T. Shaffer and B. Lighthart, "Survey of Culturable Airborne Bacteria at Four Diverse Locations in Oregon: Urban, Rural, Forest, and Coastal," *Microb. Ecol.*, vol. 34, no. 3, pp. 167–177, 1997.

[26] V. R. Després, J. F. Nowoisky, M. Klose, R. Conrad, M. O. Andreae, and U. Pöschl, "Characterization of primary biogenic aerosol particles in urban, rural, and high-alpine air by DNA sequence and restriction fragment analysis of ribosomal RNA genes," *Biogeosciences*, vol. 4, pp. 1127–1141, 2007.

[27] C. Fahlgren, Å. Hagström, D. Nilsson, and U. Li Zweifel, "Annual Variations in the Diversity, Viability, and Origin of Airborne Bacteria †," *Appl. Environ. Microbiol.*, vol. 76, no. 9, pp. 3015–3025, 2010.

[28] R. M. Bowers, S. McLetchie, R. Knight, and N. Fierer, "Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments," *ISME J. 2011 54*, vol. 5, no. 4, pp. 601–612, Nov. 2010.

[29] J. M. Costanzi and Ø. Steifetten, "Island biogeography theory explains the genetic diversity of a fragmented rock ptarmigan (Lagopus muta) population," *Ecol. Evol.*, vol. 9, no. 7, p. 3837, Apr. 2019.

[30] A. Kalmar, D. C.-G. E. and Biogeography, and undefined 2006, "A global model of island biogeography," *Glob. Ecol. Biogeogr.*, vol. 15, no. 1, pp. 72–81, Jan. 2006.

[31] D. M. Power, "Numbers of Bird Species on the California Islands," *Evolution (N.*

Y)., vol. 26, no. 3, p. 451, Sep. 1972.

[32] S. Evans, L. Bell-Dereske, K. Dougherty, and H. Kittredge, "Dispersal alters soil microbial community response to drought," *Environ. Microbiol.*, vol. 22, no. 3, pp. 905–916, Mar. 2020.

[33] S. Dullinger *et al.*, "Extinction debt of high-mountain plants under twenty-first-century climate change," *Nat. Clim. Chang.*, vol. 2, 2012.

[34] J. Ladau *et al.*, "Existing Climate Change Will Lead to Pronounced Shifts in the Diversity of Soil Prokaryotes," *mSystems*, vol. 3, no. 5, Oct. 2018.

[35] R. Hellberg and E. Chu, "Effects of climate change on the persistence and dispersal of foodborne bacterial pathogens in the outdoor environment: A review," *Crit. Rev. Microbiol.*, vol. 42, no. 4, pp. 548–572, Jul. 2016.

[36] E. E. Silvestri *et al.*, "Optimization of a sample processing protocol for recovery of Bacillus anthracis spores from soil," *J. Microbiol. Methods*, vol. 130, pp. 6–13, Nov. 2016.

[37] J. K. Carson, L. Campbell, D. Rooney, N. Clipson, and D. B. Gleeson, "Minerals in soil select distinct bacterial communities in their microhabitats," *FEMS Microbiol. Ecol.*, vol. 67, no. 3, pp. 381–388, Mar. 2009.

[38] L. F. Martins *et al.*, "Metagenomic Analysis of a Tropical Composting Operation at the São Paulo Zoo Park Reveals Diversity of Biomass Degradation Functions and Organisms," *PLoS One*, vol. 8, no. 4, Apr. 2013.

[39] L. P. Antunes *et al.*, "Microbial community structure and dynamics in thermophilic composting viewed through metagenomics and metatranscriptomics,"

Sci. Rep., vol. 6, no. 1, pp. 1–13, Dec. 2016.

[40] W. L. Nicholson, N. Munakata, G. Horneck, H. J. Melosh, and P. Setlow, "Resistance of Bacillus Endospores to Extreme Terrestrial and Extraterrestrial Environments," *Microbiol. Mol. Biol. Rev.*, vol. 64, no. 3, pp. 548–572, Sep. 2000.

[41] C. A. Hanson, J. A. Fuhrman, M. C. Horner-Devine, and J. B. H. Martiny, "Beyond biogeographic patterns: Processes shaping the microbial landscape," *Nature Reviews Microbiology*, vol. 10, no. 7. pp. 497–506, Jul-2012.

[42] J. D. Wisotzkey, P. Jurtshuk, G. E. Fox, G. Deinhard, and K. Poralla, "Comparative sequence analyses on the 16S rRNA (rDNA) of Bacillus acidocaldarius, Bacillus acidoterrestris, and Bacillus cycloheptanicus and proposal for creation of a new genus, Alicyclobacillus gen. nov.," *Int. J. Syst. Bacteriol.*, vol. 42, no. 2, pp. 263–269, 1992.

[43] B. M. Zeldes, M. W. Keller, A. J. Loder, C. T. Straub, M. W. W. Adams, and R. M. Kelly, "Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals," *Frontiers in Microbiology*, vol. 6, no. NOV. Frontiers Research Foundation, 2015.

[44] E. Ciuffreda, A. Bevilacqua, M. Sinigaglia, and M. Corbo, "Alicyclobacillus spp.: New Insights on Ecology and Preserving Food Quality through New Approaches," *Microorganisms*, vol. 3, no. 4, pp. 625–640, Oct. 2015.

[45] M. Oshima and T. Ariga, "Omega-cyclohexyl fatty acids in acidophilic thermophilic bacteria. Studies on their presence, structure, and biosynthesis using precursors labeled with stable isotopes and radioisotopes.," *J. Biol. Chem.*, vol.

250, no. 17, pp. 6963–8, Sep. 1975.

[46] K. I. Suzuki, K. Saito, A. Kawaguchi, S. Okuda, and K. Komagata, “Occurrence of ω-cyclohexyl fatty acids in curtobacterium pusillum strains,” *J. Gen. Appl. Microbiol.*, vol. 27, no. 3, pp. 261–266, 1981.

[47] K. Kusano, H. Yamada, M. Niwa, and K. Yamasato, “Propionibacterium cyclohexanicum sp. nov., a new acid-tolerant ω- cyclohexyl fatty acid-containing propionibacterium isolated from spoiled orange juice,” *Int. J. Syst. Bacteriol.*, vol. 47, no. 3, pp. 825–831, 1997.

[48] B. Hippchen, A. Röll, and K. Poralla, “Occurrence in soil of thermo-acidophilic bacilli possessing ω-cyclohexane fatty acids and hopanoids,” *Arch. Microbiol.*, vol. 129, no. 1, pp. 53–55, Mar. 1981.

[49] G. I. Karavaiko *et al.*, “Reclassification of ‘Sulfobacillus thermosulfidooxidans subsp. thermotolerans’ strain K1 as Alicyclobacillus tolerans sp. nov. and Sulfobacillus disulfidooxidans Dufresne et al. 1996 as Alicyclobacillus disulfidooxidans comb. nov., and emended description of the genus Alicyclobacillus,” *Int. J. Syst. Evol. Microbiol.*, vol. 55, no. 2, pp. 941–947, Mar. 2005.

[50] K. Goto *et al.*, “Proposal of six species of moderately thermophilic, acidophilic, endospore-forming bacteria: Alicyclobacillus contaminans sp. nov., Alicyclobacillus fastidiosus sp. nov., Alicyclobacillus kakegawensis sp. nov., Alicyclobacillus macrosporangiidus sp. nov., Alicyclobacillus sacchari sp. nov. and Alicyclobacillus shizuokensis sp. nov.,” *Int. J. Syst. Evol. Microbiol.*, vol. 57,

no. 6, pp. 1276–1285, Jun. 2007.

[51] S. Y. Lee, R. H. Dougherty, and D. H. Kang, “Inhibitory Effects of High Pressure and Heat on Alicyclobacillus acidoterrestris Spores in Apple Juice,” *Appl. Environ. Microbiol.*, vol. 68, no. 8, p. 4158, 2002.

[52] M. Walker and C. A. Phillips, “Alicyclobacillus acidoterrestris: an increasing threat to the fruit juice industry?,” *Int. J. Food Sci. Technol.*, vol. 43, no. 2, pp. 250–260, Feb. 2008.

[53] L. Albuquerque *et al.*, “Alicyclobacillus hesperidum sp. nov. and a related genomic species from solfataric soils of Sao Miguel in the Azores,” *Int. J. Syst. Evol. Microbiol.*, vol. 50, no. 2, pp. 451–457, Mar. 2000.

[54] K. Goto *et al.*, “Alicyclobacillus herbarius sp. nov., a novel bacterium containing ω-cycloheptane fatty acids, isolated from herbal tea,” *Int. J. Syst. Evol. Microbiol.*, vol. 52, no. 1, pp. 109–113, Jan. 2002.

[55] K. Goto, K. Mochida, M. Asahara, M. Suzuki, H. Kasai, and A. Yokota, “Alicyclobacillus pomorum sp. nov., a novel thermo-acidophilic, endospore-forming bacterium that does not possess ω-alicyclic fatty acids, and emended description of the genus Alicyclobacillus,” *Int. J. Syst. Evol. Microbiol.*, vol. 53, no. 5, pp. 1537–1544, Sep. 2003.

[56] N. Tsuruoka, Y. Isono, O. Shida, H. Hemmi, T. Nakayama, and T. Nishino, “Alicyclobacillus sendaiensis sp. nov., a novel acidophilic, slightly thermophilic species isolated from soil in Sendai, Japan,” *Int. J. Syst. Evol. Microbiol.*, vol. 53, no. 4, pp. 1081–1084, Jul. 2003.

[57] J. Simbahan, R. Drijber, and P. Blum, “Alicyclobacillus vulcanalis sp. nov., a thermophilic acidophilic bacterium isolated from Coso Hot Springs, California, USA,” *Int. J. Syst. Evol. Microbiol.*, vol. 54, no. 5, pp. 1703–1707, Sep. 2004.

[58] Y. Gumulya, N. J. Boxall, H. N. Khaleque, V. Santala, R. P. Carlson, and A. H. Kaksonen, “In a Quest for Engineering Acidophiles for Biomining Applications: Challenges and Opportunities,” *Genes (Basel).*, vol. 9, no. 2, Feb. 2018.

[59] L. Wörmer *et al.*, “Microbial dormancy in the marine subsurface: Global endospore abundance and response to burial,” *Sci. Adv.*, vol. 5, no. 2, p. eaav1024, Feb. 2019.

[60] J. Arnold, “Genetic Drift,” in *Encyclopedia of Genetics*, Elsevier, 2001, pp. 832–834.

[61] P. Bijma, “A general definition of the heritable variation that determines the potential of a population to respond to selection,” *Genetics*, vol. 189, no. 4, pp. 1347–1359, Dec. 2011.

[62] G. SH *et al.*, “Identification of Enterococcus species and phenotypically similar Lactococcus and Vagococcus species by reverse checkerboard hybridization to chaperonin 60 gene sequences,” *J. Clin. Microbiol.*, vol. 38, no. 11, pp. 3953–3959, 2000.

[63] S. Niemann, D. Harmsen, S. Rüsch-Gerdes, and E. Richter, “Differentiation of clinical Mycobacterium tuberculosis complex isolates by gyrB DNA sequence polymorphism analysis,” *J. Clin. Microbiol.*, vol. 38, no. 9, pp. 3231–3234, 2000.

[64] M. Fukushima, K. Kakinuma, and R. Kawaguchi, “Phylogenetic analysis of

Salmonella, Shigella, and Escherichia coli strains on the basis of the gyrB gene sequence," *J. Clin. Microbiol.*, vol. 40, no. 8, pp. 2779–2785, Aug. 2002.

[65] M. La Duc, M. Satomi, N. Agata, and K. Venkateswaran, "gyrB as a phylogenetic discriminator for members of the Bacillus anthracis-cereus-thuringiensis group," *J. Microbiol. Methods*, vol. 56, no. 3, pp. 383–394, 2004.

[66] J. E. Hill, S. L. Penny, K. G. Crowell, S. H. Goh, and S. M. Hemmingsen, "cpnDB: A chaperonin sequence database," *Genome Res.*, vol. 14, no. 8, pp. 1669–1675, Aug. 2004.

[67] S. Goh *et al.*, "Streptococcus iniae, a human and animal pathogen: specific identification by the chaperonin 60 gene identification method," *J. Clin. Microbiol.*, vol. 36, no. 7, pp. 2164–2166, 1998.

[68] A. Kwok, J. Wilson, M. Coulthart, L. Ng, L. Mutharia, and A. Chow, "Phylogenetic study and identification of human pathogenic Vibrio species based on partial hsp60 gene sequences," *Can. J. Microbiol.*, vol. 48, no. 10, pp. 903–910, Oct. 2002.

[69] A. Dekowska, J. Niezgoda, and B. Sokołowska, "Genetic Heterogeneity of Alicyclobacillus Strains Revealed by RFLP Analysis of vdc Region and rpoB Gene.," *Biomed Res. Int.*, vol. 2018, p. 9608756, 2018.

[70] B. E. Luedtke and J. M. Bosilevac, "Comparison of methods for the enumeration of enterohemorrhagic Escherichia coli from veal hides and carcasses," *Front. Microbiol.*, vol. 6, no. SEP, p. 1062, Sep. 2015.

[71] M. C. J. Maiden *et al.*, "Multilocus sequence typing: A portable approach to the

identification of clones within populations of pathogenic microorganisms," *Proc. Natl. Acad. Sci.*, vol. 95, no. 6, pp. 3140–3145, Mar. 1998.

[72] A. Belén, I. Pavón, and M. C. J. Maiden, "Multilocus Sequence Typing," *Methods Mol. Biol.*, vol. 551, p. 129, 2009.

[73] J. Smith, N. Smith, M. O'Rourke, and B. Spratt, "How clonal are bacteria?," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 10, pp. 4384–4388, 1993.

[74] N. R. Pace, J. Sapp, and N. Goldenfeld, "Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life," *Proc. Natl. Acad. Sci.*, vol. 109, no. 4, pp. 1011–1018, Jan. 2012.

[75] B. JP, "Bayesian model adequacy and choice in phylogenetics," *Mol. Biol. Evol.*, vol. 19, no. 7, pp. 1171–1180, 2002.

[76] A. Lemmon and E. Moriarty, "The importance of proper model assumption in bayesian phylogenetics," *Syst. Biol.*, vol. 53, no. 2, pp. 265–277, Apr. 2004.

[77] J. Ripplinger and J. Sullivan, "Does choice in model selection affect maximum likelihood analysis?," *Syst. Biol.*, vol. 57, no. 1, pp. 76–85, Feb. 2008.

[78] F. F. Nascimento, M. dos Reis, and Z. Yang, "A biologist's guide to Bayesian phylogenetic analysis," *Nat. Ecol. Evol.*, vol. 1, no. 10, p. 1446, Oct. 2017.

[79] W. Ludwig, F. Oliver Glöckner, and P. Yilmaz, "The Use of rRNA Gene Sequence Data in the Classification and Identification of Prokaryotes," *Methods Microbiol.*, vol. 38, pp. 349–384, 2011.

[80] T. Warnow, "Standard maximum likelihood analyses of alignments with gaps can be statistically inconsistent," *PLoS Curr.*, vol. 4, 2012.

[81] J. Truszkowski and N. Goldman, “Maximum Likelihood Phylogenetic Inference is Consistent on Multiple Sequence Alignments, with or without Gaps,” *Syst. Biol.*, vol. 65, no. 2, p. 328, Mar. 2016.

[82] A. Stamatakis, T. Ludwig, and H. Meier, “RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees,” *Bioinformatics*, vol. 21, no. 4, pp. 456–463, Feb. 2005.

[83] G. Olsen, H. Matsuda, R. Hagstrom, and R. Overbeek, “fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood,” *Comput. Appl. Biosci.*, vol. 10, no. 1, pp. 41–48, Feb. 1994.

[84] J. Huelsenbeck and F. Ronquist, “MRBAYES: Bayesian inference of phylogenetic trees,” *Bioinformatics*, vol. 17, no. 8, pp. 754–755, 2001.

[85] S. Guindon and O. Gascuel, “A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood,” *Syst. Biol.*, vol. 52, no. 5, pp. 696–704, Oct. 2003.

[86] M. A. Miller, W. Pfeiffer, and T. Schwartz, “‘Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees’ in Proceedings of the Gateway Computing Environments Workshop (GCE),” p. New Orleans, LA pp 1-8, 2010.

[87] P. D. Schloss and J. Handelsman, “Status of the Microbial Census,” *Microbiol. Mol. Biol. Rev.*, vol. 68, no. 4, p. 686, Dec. 2004.

[88] T. P. Curtis, W. T. Sloan, and J. W. Scannell, “Estimating prokaryotic diversity and its limits,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 16, pp. 10494–10499,

Aug. 2002.

[89] D. Dykhuizen, “Species Numbers in Bacteria,” *Proc. Calif. Acad. Sci.*, vol. 56, no. 6 Suppl 1, p. 62, Jun. 2005.

[90] L. Trenozhnikova and A. Azizan, “Discovery of Actinomycetes from Extreme Environments with Potential to Produce Novel Antibiotics,” *Cent. Asian J. Glob. Heal.*, vol. 7, no. 1, Dec. 2018.

[91] S. Ishino and Y. Ishino, “DNA polymerases as useful reagents for biotechnology – the history of developmental research in the field,” *Front. Microbiol.*, vol. 5, no. AUG, 2014.

[92] A. Sharma, Y. Kawarabayasi, and T. Satyanarayana, “Acidophilic bacteria and archaea: acid stable biocatalysts and their potential applications,” *Extremophiles*, vol. 16, no. 1, pp. 1–19, Jan. 2012.

[93] C. A. Fuchsman, R. E. Collins, G. Rocap, and W. J. Brazelton, “Effect of the environment on horizontal gene transfer between bacteria and archaea,” *PeerJ*, vol. 2017, no. 9, 2017.

[94] C. Wintersdorff *et al.*, “Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer,” *Front. Microbiol.*, vol. 7, no. FEB, p. 173, Feb. 2016.

[95] Y. Shi, Y. Sun, X. Qu, L. Zhou, T. Yue, and Y. Yuan, “Preparation of species-specific monoclonal antibody and development of fluorescence immunoassay based on fluorescence resonance energy transfer of carbon dots for accurate and sensitive detection of Alicyclobacillus acidoterrestris in apple juice,” *Food Chem.*,

vol. 347, p. 129069, Jan. 2021.

[96] D. Sun *et al.*, “Escherichia coli is naturally transformable in a novel transformation system,” *FEMS Microbiol. Lett.*, vol. 265, no. 2, pp. 249–255, Dec. 2006.

[97] C. Smillie, M. Garcillán-Barcia, M. Francia, E. Rocha, and F. de la Cruz, “Mobility of plasmids,” *Microbiol. Mol. Biol. Rev.*, vol. 74, no. 3, pp. 434–452, Sep. 2010.

[98] C. Johnson and A. Grossman, “Integrative and Conjugative Elements (ICEs): What They Do and How They Work,” *Annu. Rev. Genet.*, vol. 49, pp. 577–601, Nov. 2015.

[99] F. Griffith, “The Significance of Pneumococcal Types,” *J. Hyg. (Lond).*, vol. 27, no. 2, p. 113, 1928.

[100] P. F. Sparling, “Genetic Transformation of Neisseria gonorrhoeae to Streptomycin Resistance,” *J. Bacteriol.*, vol. 92, no. 5, p. 1364, 1966.

[101] B. Marrs, “Genetic Recombination in Rhodopseudomonas capsulata,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 71, no. 3, p. 971, 1974.

[102] L. C. Woods, R. J. Gorrell, F. Taylor, T. Connallon, T. Kwok, and M. J. McDonald, “Horizontal gene transfer potentiates adaptation by reducing selective constraints on the spread of genetic variation,” *Proc. Natl. Acad. Sci.*, vol. 117, no. 43, pp. 26868–26875, Oct. 2020.

[103] G. Sevillya, O. Adato, and S. Snir, “Detecting horizontal gene transfer: a probabilistic approach,” *BMC Genomics 2020 211*, vol. 21, no. 1, pp. 1–11, Mar. 2020.

[104] Y. Nakamura, T. Itoh, H. Matsuda, and T. Gojobori, “Biased biological functions of horizontally transferred genes in prokaryotic genomes,” *Nat. Genet. 2004 367*, vol. 36, no. 7, pp. 760–766, Jun. 2004.

[105] S. Garcia-Vallvé, A. Romeu, and J. Palau, “Horizontal Gene Transfer in Bacterial and Archaeal Complete Genomes,” *Genome Res.*, vol. 10, no. 11, pp. 1719–1725, Nov. 2000.

[106] R. Grantham, C. Gautier, M. Gouy, R. Mercier, and A. Pavé, “Codon catalog usage and the genome hypothesis,” *Nucleic Acids Res.*, vol. 8, no. 1, p. 197, Jan. 1980.

[107] H. Li, Q. Yang, J. Li, H. Gao, P. Li, and H. Zhou, “The impact of temperature on microbial diversity and AOA activity in the Tengchong Geothermal Field, China,” *Sci. Reports 2015 51*, vol. 5, no. 1, pp. 1–12, Nov. 2015.

[108] Z. Liang, “Isolation and phylogenetic characterization of acidophiles in acid mine drainage environment at a mining area of Mengi, Yunnan Province,” *Unpublished*, 2020.

[109] K. Goto and M. Asahara, “16S rDNA and gyrB gene sequences of Alicyclobacillus isolates,” *Unpublished*, 2007.

[110] S. Lu, S. Gischkat, M. Reiche, D. M. Akob, K. B. Hallberg, and K. Küsel, “Ecophysiology of Fe-Cycling Bacteria in Acidic Sediments,” *Appl. Environ. Microbiol.*, vol. 76, no. 24, p. 8174, Dec. 2010.

[111] S. N. M. Rozali, E. A. Milani, R. C. Deed, and F. V. M. Silva, “Bacteria, mould and yeast spore inactivation studies by scanning electron microscope

observations," *Int. J. Food Microbiol.*, vol. 263, pp. 17–25, Dec. 2017.

[112] R. C. Witthuhn, W. Duvenage, and P. A. Gouws, "Evaluation of different growth media for the recovery of the species of Alicyclobacillus," *Lett. Appl. Microbiol.*, vol. 45, no. 2, pp. 224–229, Aug. 2007.

[113] M. Kakagianni, K. Kalantzi, E. Beletsiotis, D. Ghikas, A. Lianou, and K. P. Koutsoumanis, "Development and validation of predictive models for the effect of storage temperature and pH on the growth boundaries and kinetics of Alicyclobacillus acidoterrestris ATCC 49025 in fruit drinks," *Food Microbiol.*, vol. 74, pp. 40–49, Sep. 2018.

[114] R. Blodgett, "BAM Appendix 2: Most Probable Number from Serial Dilutions | FDA," *U.S. Food & Drug Administration*, 2010. [Online]. Available: https://www.fda.gov/food/laboratory-methods-food/bam-appendix-2-most-probable-number-serial-dilutions. [Accessed: 28-Feb-2020].

[115] United States Environmental Protection Agency (EPA), "EPA MPN Calculator." [Online]. Available: https://mostprobablenumbercalculator.epa.gov/mpnForm. [Accessed: 27-Sep-2021].

[116] E. A. Cornish and R. A. Fisher, "Moments and Cumulants in the Specification of Distributions," *Rev. Int. Stat. Inst.*, vol. 5, no. 4, pp. 307–320, 1938.

[117] M. W. Loyer and M. A. Hamilton, "Interval Estimation of the Density of Organisms Using a Serial-Dilution Experiment," *Biometrics*, vol. 40, no. 4, p. 907, Dec. 1984.

[118] C. Spearman, "The Method of 'Right And Wrong Cases' ('Constant Stimuli')

Without Gauss's Formulae," *Br. J. Psychol. 1904-1920*, vol. 2, no. 3, pp. 227–242, 1908.

[119] G. Kärber, "Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche," *Naunyn-Schmiedebergs Arch. für Exp. Pathol. und Pharmakologie 1931 1624*, vol. 162, no. 4, pp. 480–483, Jul. 1931.

[120] H. A. Thomas, "Bacterial Densities From Fermentation Tube Tests," *J. Am. Water Works Assoc.*, vol. 34, no. 4, pp. 572–576, 1942.

[121] W. E. Garthright, "Bias in the Logarithm of Microbial Density Estimates from Serial Dilutions," *Biometrical J.*, vol. 35, no. 3, pp. 299–314, Jan. 1993.

[122] S. Fuzzi, P. Mandrioli, and A. Perfetto, "Fog droplets—an atmospheric source of secondary biological aerosol particles," *Atmos. Environ.*, vol. 31, no. 2, pp. 287–290, Jan. 1997.

[123] H. Nacke *et al.*, "Pyrosequencing-Based Assessment of Bacterial Community Structure Along Different Management Types in German Forest and Grassland Soils," *PLoS One*, vol. 6, no. 2, p. e17000, 2011.

[124] C. A. Hanson *et al.*, "Historical Factors Associated With Past Environments Influence the Biogeography of Thermophilic Endospores in Arctic Marine Sediments.," *Front. Microbiol.*, vol. 10, p. 245, 2019.

[125] O. Lomovskaya and K. Lewis, "emr, an Escherichia coli locus for multidrug resistance," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. 19, pp. 8938–8942, Oct. 1992.

[126] P. M. Watt and I. D. Hickson, "Structure and function of type II DNA

topoisomerases," *Biochemical Journal*, vol. 303, no. 3. Portland Press Ltd, pp. 681–695, 1994.

[127] L.-T. Wang, F.-L. Lee, C.-J. Tai, and H. Kasai, "Comparison of gyrB gene sequences, 16S rRNA gene sequences and DNA–DNA hybridization in the Bacillus subtilis group," *Int. J. Syst. Evol. Microbiol.*, vol. 57, no. 8, pp. 1846–1850, Aug. 2007.

[128] W. Huang, "Bacterial diversity based on type II DNA topoisomerase genes," *Annu. Rev. Genet.*, vol. 30, pp. 79–107, 1996.

[129] S. Yamamoto and S. Harayama, "PCR amplification and direct sequencing of gyrB genes with universal primers and their application to the detection and taxonomic analysis of Pseudomonas putida strains.," *Appl. Environ. Microbiol.*, vol. 61, no. 3, p. 1104, 1995.

[130] S. Yamamoto and S. Harayama, "Phylogenetic relationships of Pseudomonas putida strains deduced from the nucleotide sequences of gyrB, rpoD and 16S rRNA genes," *Int. J. Syst. Bacteriol.*, vol. 48 Pt 3, no. 3, pp. 813–819, 1998.

[131] S. Yamamoto, P. Bouvet, and S. Harayama, "Phylogenetic structures of the genus Acinetobacter based on gyrB sequences: comparison with the grouping by DNA-DNA hybridization," *Int. J. Syst. Bacteriol.*, vol. 49 Pt 1, no. 1, pp. 87–95, 1999.

[132] H. Kasai, T. Ezaki, and S. Harayama, "Differentiation of Phylogenetically Related Slowly Growing Mycobacteria by Their gyrB Sequences," *J. Clin. Microbiol.*, vol. 38, no. 1, p. 301, 2000.

[133] M. Yáñez, V. Catalán, D. Apráiz, M. Figueras, and A. Martínez-Murcia,

"Phylogenetic analysis of members of the genus Aeromonas based on gyrB gene sequences," *Int. J. Syst. Evol. Microbiol.*, vol. 53, no. Pt 3, pp. 875–883, May 2003.

[134] S. H. Gob, S. Potter, J. O. Wood, S. M. Hemmingsen, R. P. Reynolds, and A. W. Chow, "HSP60 gene sequences as universal targets for microbial species identification: Studies with coagulase-negative staphylococci," *J. Clin. Microbiol.*, vol. 34, no. 4, pp. 818–823, 1996.

[135] D. Klunker *et al.*, "Coexistence of group I and group II chaperonins in the archaeon Methanosarcina mazei," *J. Biol. Chem.*, vol. 278, no. 35, pp. 33256–33267, Aug. 2003.

[136] S. Goh *et al.*, "Identification of Staphylococcus species and subspecies by the chaperonin 60 gene identification method and reverse checkerboard hybridization," *J. Clin. Microbiol.*, vol. 35, no. 12, pp. 3116–3121, 1997.

[137] C. Dale *et al.*, "Identification and sequencing of the groE operon and flanking genes of Lawsonia intracellularis: use in phylogeny," *Microbiology*, vol. 144 (Pt 8, no. 8, pp. 2073–2084, 1998.

[138] A. Kwok and A. Chow, "Phylogenetic study of Staphylococcus and Macrococcus species based on partial hsp60 gene sequences," *Int. J. Syst. Evol. Microbiol.*, vol. 53, no. Pt 1, pp. 87–92, Jan. 2003.

[139] A. Lew, K. Gale, C. Minchin, V. Shkap, and D. de Waal, "Phylogenetic analysis of the erythrocytic Anaplasma species based on 16S rDNA and GroEL (HSP60) sequences of A. marginale, A. centrale, and A. ovis and the specific detection of A.

centrale vaccine strain," *Vet. Microbiol.*, vol. 92, no. 1–2, pp. 145–160, Mar. 2003.

[140] J. C. Oliver, R. S. Linger, S. V. Chittur, and V. J. Davisson, "Substrate Activation and Conformational Dynamics of GMP Synthetase," *Biochemistry*, vol. 52, no. 31, pp. 5225–5235, Aug. 2013.

[141] H. Taniguchi Al* *et al.*, "Rifampicin resistance and mutation of the rpoB gene in Mycobacterium tuberculosis," *FEMS Microbiol. Lett.*, vol. 144, pp. 103–108, 1996.

[142] A. Weijland, K. Harmark, R. H. Cool, P. H. Anborgh, and A. Parmeggiani, "Elongation factor Tu: a molecular switch in protein biosynthesis," *Mol. Microbiol.*, vol. 6, no. 6, pp. 683–688, Mar. 1992.

[143] E. Fullam, F. Pojer, T. Bergfors, T. A. Jones, and S. T. Cole, "Structure and function of the transketolase fromMycobacterium tuberculosis and comparison with the humanenzyme," *Open Biol.*, vol. 2, no. 1, 2012.

[144] G. Zhao and M. Winkler, "An Escherichia coli K-12 tktA tktB mutant deficient in transketolase activity requires pyridoxine (vitamin B6) as well as the aromatic amino acids and vitamins for growth," *J. Bacteriol.*, vol. 176, no. 19, pp. 6134–6138, 1994.

APPENDIX A: R CODE USED IN STATISTICAL MODELS AND SEQUENCE MANIPULATIONS

The R code for Generalized Linear Model and analysis is as follows:

```
#temperature trials glm
model1<-glm(growth~temperature, data=temp.trials.alicyclo.acido, family=binomial)
summary(model1)
#plot growth vs temperature
x <- seq(30,70,length=300)
y <- predict(model1,list(temperature=x),type="response")
plot(temperature,jitter(growth),pch=20,xlim=c(30,70),col=1,cex=1)
points(x,y,type="l",col=”red”,lwd=3)
#predicting from model for each temperature
predict(model1,list(temperature=30),type="response")
predict(model1,list(temperature=40),type="response")
predict(model1,list(temperature=50),type="response")
predict(model1,list(temperature=60),type="response")
```

The R code used for concatenating the genes used in the multilocus sequence typing analysis as follow:

```
#import gene file and read
All.genes.file<-readDNAStringSet("mlst.clipped.genes.txt", format="fasta")
```

#callout sequences by name from file and concatenate for each isolate

PS5<-xscat(All.genes.file$PS5.*cpn60*, All.genes.file$PS5.*eftu*, All.genes.file$PS5.*gyrB*, All.genes.file$PS5_*rpoB*, All.genes.file$PS5.*emrB*, All.genes.file$PS5.*gmp*)

#create a list of all the concatenated sequences

mlst <- list(PS5=PS5, FFR.D=FFR.D,.....Red5=Red5)

#save list to computer

write.dna(mlst, “mlst.concatenated.genes.fasta”, format=”fasta”)

APPENDIX B: ENVIRONMENTAL CHARACTERISTICS AND DESCRIPTIONS OF THE NON-PERMISSIVE LOCATIONS

Sample Name	Location	Date Sampled	GPS Coordinates	Approximate Elevation (ft)	Chapter Used
Football field rubber	Arcata, CA	N/R	N/R	150	6, 7
Mad River Douglas fir forest	Mad River, CA	2/6/21	40.302, -123.330	2,800	5, 6, 7
Maple Creek Douglas fir forest	Maple Creek, CA	2/6/21	40.786, -123.889	680	5, 6, 7
Maple Creek meadow	Maple Creek, CA	2/6/21	40.786, -123.881	650	5, 6, 7
Pasture soil	Humboldt County, CA	N/R	N/R	N/R	6, 7
Redwood forest	Arcata, CA	2/6/21	40.875, -124.074	300	5, 6, 7
Spruce forest	Loleta, CA	2/5/20	40.671, -124.215	150	3
Spruce forest	Loleta, CA	2/6/21	40.671, -124.215	150	5, 6, 7
Trinity Alps	Junction City, CA	10/3/20	40.929, -123.021	4,000	4, 5, 6, 7

*N/R=Not Recorded