

2013

Biogeography and evolution of *Thermococcus* isolates from hydrothermal vent systems of the Pacific

Mark T. (Mark Thomas) Price
Western Washington University

Follow this and additional works at: <https://cedar.wvu.edu/wwuet>



Part of the [Biology Commons](#)

Recommended Citation

Price, Mark T. (Mark Thomas), "Biogeography and evolution of *Thermococcus* isolates from hydrothermal vent systems of the Pacific" (2013). *WWU Graduate School Collection*. 293.
<https://cedar.wvu.edu/wwuet/293>

This Masters Thesis is brought to you for free and open access by the WWU Graduate and Undergraduate Scholarship at Western CEDAR. It has been accepted for inclusion in WWU Graduate School Collection by an authorized administrator of Western CEDAR. For more information, please contact westerncedar@wwu.edu.

**BIOGEOGRAPHY AND EVOLUTION OF *THERMOCOCCUS* ISOLATES FROM
HYDROTHERMAL VENT SYSTEMS OF THE PACIFIC**

By

Mark T. Price

Accepted in Partial Completion
Of the Requirements for the Degree
Master of Science

Kathleen L. Kitto, Dean of the Graduate School

ADVISORY COMMITTEE

Chair, Dr. Craig Moyer

Dr. David Leaf

Dr. Jeff Young

Dr. Dietmar Schwarz

MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Mark Price

July 22, 2013

**BIOGEOGRAPHY AND EVOLUTION OF *THERMOCOCCUS* ISOLATES FROM
HYDROTHERMAL VENT SYSTEMS OF THE PACIFIC**

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

By
Mark T. Price
July 2013

ABSTRACT

Thermococcus are an Archaeal genus of hyperthermophilic microorganisms found to be ubiquitously present in hydrothermal habitats. DNA analyses of *Thermococcus* isolates focusing primarily on isolates from the Juan de Fuca Ridge, Gorda Ridge, and South East Pacific Rise, were applied in order to determine the relationship between geographic distribution and relatedness. Amplified fragment length polymorphism (AFLP) analysis and multilocus sequence typing (MLST) were used to resolve genomic differences at the species and strain level in 90 isolates of *Thermococcus*, allowing for the detection of biogeographic patterns and evolutionary relationships within this genus. Isolates were differentiated at regional levels and into distinct lineages within regions. Although no correlation was found between environmental characteristics and genotype, the presence of distinct lineages within the same vent site suggests the utilization of different ecological niches by distinct *Thermococcus* species. A unique group of *Thermococcus* identified, that lacked geographic genetic structure, contained highly similar isolates from disparate regions. The anomalous nature of this group may be explained by a recent population divergence, having high dispersal potential or possibly in being a niche specific ecotype. In addition to the resolution of biogeographic patterns and evolution in *Thermococcus* that this study has provided, new questions have been raised about the closely related *Pyrococcus* genus. Analysis of loci GC ratios and the placement of *Pyrococcus* type strains in phylogenetic trees have demonstrated the close association between *Thermococcus* and *Pyrococcus* and the unresolved divergence between these two genera.

ACKNOWLEDGEMENTS

I would like to thank John A. Baross's lab for donation of the "Zoo" culture collection and Debbie Whitley for sparking my interest in working with these cultures. I would also like to thank the NSF-funded REU students working in Craig Moyer's lab for their efforts towards maintaining the culture collection, in particular Kelsey Jesser, Kevin Hager and Kyle Hager. I would like to acknowledge Rick Davis and Sean McAllister for their help with phylogenetic trees. For their help and support in editing my thesis I would like to acknowledge Dietmar Schwarz, Heather Fullerton, and Craig Moyer. In addition I would like to thank my coworkers Peter Thut, Kendra Bradford and Jeannie Gilbert for their support, and family members Michael Price, Stephen Price, Alisa Sachs, and Ezra and Liza Price, for their backing and support.

TABLE OF CONTENTS

Abstract	iv
Acknowledgements	v
List of Figures	ix
List of Tables	x
Introduction	1
Materials and Methods	7
Thermococcales Isolates	7
Culturing	8
DNA Extraction	9
AFLP Reactions	9
Cluster Analysis	10
MLST Primer Design	10
Loci PCR and Sequencing	11
Loci Test for Selection	12
Phylogenetic Analysis	12

MLST Clade Analysis	13
ANI	13
%GC	13
Linkage Analysis	13
Codon GC Ratios	13
Mantel Test	14
Analysis of Molecular Variance	14
Principal Component Analysis	15
Results	16
Cluster Analysis	16
Loci Test for Selection	16
Phylogenetic Analysis	17
MLST Clade Analysis	18
ANI	18
%GC	18
Linkage Analysis	19
Codon GC Ratios	19

Mantel Test	19
Analysis of Molecular Variance	20
Principal Component Analysis	20
Discussion	21
Literature Cited	28
Figures	36
Tables	44
Supplementary Figures	46
Supplementary Tables	54

LIST OF FIGURES

Figure 1. Maps of the two main sample sites and associated vent segments	36
Figure 2. Cluster analysis of AFLP data	37
Figure 3. Maximum likelihood phylogenetic tree of concatenated MLST loci	39
Figure 4. Maximum likelihood phylogenetic tree of the SSU rRNA	41
Figure 5. Average %G+C for first and third codon position	42
Figure 6. Principal Component Analysis of AFLP data	43
Figure S1. Gene loci maps for six reference genomes	46
Figure S2. Unrooted Maximum likelihood tree for elongation factor 1 alpha	47
Figure S3. Unrooted Maximum likelihood tree for DNA polymerase II large subunit	48
Figure S4. Unrooted Maximum likelihood tree for DNA topoisomerase VI alpha subunit	49
Figure S5. Unrooted Maximum likelihood tree for pyruvate ferredoxin oxidoreductase beta subunit	50
Figure S6. Unrooted Maximum likelihood tree for histone acetyl transferase	51
Figure S7. Unrooted Maximum likelihood tree for threonyl tRNA synthetase	52
Figure S8. Average %G+C for first and third codon of six loci	53

LIST OF TABLES

Table 1. Isolates analyzed through AFLP and MLST and vent segments they were isolated from	44
Table 2. Clade analysis for ANI, %GC and test for linkage	45
Table S1. <i>Thermococcus</i> type strains analyzed by AFLP and MLST	54
Table S2. MLST primer sequences	55
Table S3. Loci dN/dS ratios	56
Table S4. Mantel test	57

INTRODUCTION

Microorganisms constitute the majority of known life forms making up greater than two-thirds of the metabolic and genetic diversity of the planet (1). The evolutionary forces responsible for shaping this vast amount of microbial diversity though have been a topic of debate. Historically microorganisms have been viewed as having panmictic distributions, being constrained only by environmental conditions (2), but in recent times there has been growing evidence for the divergence of microbial populations through isolation and genetic drift (3). The significance of physical isolation in shaping microbial diversity has been demonstrated in hyperthermophiles associated with terrestrial hot springs (1, 4). This raises the question of whether similar patterns can be observed in the island-like extreme environments of marine hydrothermal vents.

In terrestrial environments, islands and the organisms associated with them can be geographically isolated leading to the divergence of populations through local adaptation and genetic drift. In marine environments hydrothermal vent systems and their biota exist as island-like ecosystems dispersed along plate boundaries and separated from one another over vast stretches of the seafloor. The unique nature of hydrothermal vents with their extreme chemical, nutrient and temperature gradients, and the discontinuous nature of these habitats makes them ideal for the study of biogeography and evolution where questions about dispersal, distance effects and ecologically influenced divergence can be investigated (5–7). Biogeographic patterns and population divergence have been well documented in hyperthermophilic microorganisms that inhabit terrestrial hot springs (8, 9), but for hyperthermophilic microorganisms inhabiting marine hydrothermal vents descriptions of

biogeography have been uncommon (10–12). The limited descriptions of microbial biogeography at hydrothermal vents are in part due to the difficulty in culturing these organisms, a limiting step in most microorganisms. The isolation in culture of clonal organisms greatly enhances the ability to address questions of biogeography and evolution by making individual microbial genomes and the information they contain easily accessible (7, 13, 14).

The Archaeal order Thermococcales is a unique group of hyperthermophilic microorganisms found at hydrothermal vents that can serve as model organisms for the study of biogeography and evolution within these island-like habitats. Thermococcales belong to the Euryarchaeota phylum and consists of the three genera *Pyrococcus* (15), *Thermococcus* (16), and *Paleococcus* (17). The genera *Thermococcus* and *Pyrococcus* are commonly found inhabiting hydrothermal vents and are conducive to being studied as they are readily isolated in culture allowing for comparisons between individual isolates (18–21). Genome content differences between *Thermococcus* and *Pyrococcus* and a higher optimum growth temperature in *Pyrococcus* delineate these two closely related genera (22). *Thermococcus* though are the most frequently isolated and have the highest number of characterized isolates (18, 22–24).

Although *Thermococcus* can grow at temperatures as low as 60°C, their ability to grow at 90°C results in their classification as hyperthermophiles (25, 26). *Thermococcus* genera consist in general of anaerobic heterotrophs that utilize a sulfur reduction pathway to fix organic carbon (26). However some species of *Thermococcus* require sulfur in order to grow while others do not having instead the ability to utilize other terminal electron acceptors

(18). Lithotrophic metabolic pathways have also been demonstrated in *Thermococcus*. The oxidation of carbon monoxide to CO₂, a metabolism known as carboxydrotrophy using carbon monoxide dehydrogenases (CODHs), has been described in the type strain *Thermococcus onnurines* as well as in other type strains, allowing for lithotrophic growth from carbon monoxide (24, 27–29). *Thermococcus* have also demonstrated the ability to grow through formate oxidation and H₂ production, representing one of the simplest forms of anaerobic respiration (30). The ability to fix carbon or produce ATP through these low energy yielding metabolic pathways has been suggested to have important implications on survival in environments where energy supplies are at times transient or low (27, 30). In addition to having metabolic versatility, a number of putative hot and cold chaperonins and oxygen detoxifying enzymes have been identified in *Thermococcus* through whole-genome analyses (24, 31). These enzymes are hypothesized to aid in survival by maintaining protein function both within active vents as well as in low temperature and oxygenated environments. Thermococcales in particular are known for their ability to survive for extended periods in low temperature, oxygenated conditions (32), however the effect this resilience has on dispersal potential and biogeography has been difficult to ascertain.

The historical premise of microorganisms having panmictic distribution and being constrained only by environmental factors (2), has been challenged by the alternate view that through clonal reproduction and with low levels of genetic exchange or geographic isolation, genetic drift and selection will lead to endemic populations (7). In the island-like extreme environments of terrestrial hot springs a correlation between geographic isolation and population divergence has been demonstrated in both Bacteria and Archaea. In the cyanobacteria *Synechococcus* and Archaeal *Sulfolobus* populations were shown to cluster by

geographic locale (8, 9, 33). With no significant correlation between genotype and environmental characteristics in these studies, population divergence was attributed to local adaptation and genetic drift. This raises the question of whether similar patterns can be observed in hyperthermophilic microorganisms inhabiting the island-like extreme environments of marine hydrothermal vents and if there is greater potential for dispersal in these marine environments.

In the Thermococcales evidence of genetic drift and biogeographic patterns have been observed in populations of *Pyrococcus* through Multilocus Sequence Typing (MLST) and the analysis of genetic mobile elements (12). Populations of *Pyrococcus* from different regions were shown to be genetically differentiated through MLST and in a geographically isolated population from Volcano Island Italy genetic mobile elements were found to be at a high frequency. For *Thermococcus*, descriptions of population divergence and biogeography have proven enigmatic. This genus has been described in more general terms as being widespread and ubiquitous in hydrothermal habitats (23, 34). Studies of *Thermococcus* involving DNA analysis of small subunit ribosomal RNA (SSU rRNA) and the intergenic transcribed spacer region (ITS) have grouped environmental clones and isolates with reference strains from around the world (23, 34, 35). Although there has been evidence suggestive of a correlation between *Thermococcus* diversity, environmental conditions, and geography (18, 23, 34, 36), no strong biogeographic pattern has emerged. Analysis of *Thermococcus* by random amplified polymorphic DNA (RAPD) has identified a wide diversity of profiles from the same sample site and similar profiles from different sites, illustrating the biodiversity and dispersal potential detected within a hydrothermal vent field (35). Although it is accepted that *Thermococcus* are commonly found in hydrothermal systems, the lack of evidence for

biogeographic patterns has made it unclear whether populations are panmictic in their distribution or if they show levels of endemism at varying geographic scales.

With the exception of RAPD analysis, the markers used in previous studies of *Thermococcus* have not had a high degree of genetic resolution. Two different and yet complementary DNA typing methods that utilize multiple sites from across the genome are MLST and Amplified Fragment Length Polymorphism (AFLP) analysis. MLST is a well-established DNA typing method that uses the sequences of from five to seven conserved housekeeping genes spanning the genome, to construct robust phylogenetic relationships between related microorganisms (37). When applied in environmental microbiology, MLST analysis delineated biogeographic patterns in *Sulfolobus* isolates sharing 99.8% sequence similarity for the SSU rRNA gene (9). MLST data have also provided a basis for defining species level boundaries through the comparison of loci and their average nucleotide identity (ANI) (38). ANI comparisons of MLST loci, with sequence similarities of 95% or greater, have been correlated with species level ANI values determined from whole-genome comparisons (38–40). By applying MLST analysis to *Thermococcus* isolates, the evolutionary relationships between isolates as well as their biodiversity can be described more accurately. In comparison to MLST analysis which looks at a limited set of gene loci, AFLP analysis utilizes genome wide restriction fragment lengths. Having high discriminatory power AFLP analysis allows for the typing of microorganisms down to the strain level (41–43). AFLP genome fingerprints may vary as a result of nucleotide sequence divergence as well as the movement of transposable elements, insertions or deletions, or genome rearrangements. Due to some of these changes occurring at rates more rapidly than

nucleotide sequence divergence, AFLP fingerprints have the potential to resolve more recent genome differentiation (44).

Comparisons between fully sequenced Thermococcales genomes looking at gene synteny, have made apparent the significance that rearrangements have on genomic variation in this order (24, 27, 31). While there is evidence that genomic rearrangements in the Thermococcales may be the result of DNA damage from the extreme environments that these organisms inhabit (45–47), there is also evidence in support of genomic rearrangements resulting from recombination and transposition related events (12, 31, 48–50). By looking at genomic variation in *Thermococcus* within the context of a microbial species pan-genome, variation associated with rearrangements, in the core and dispensable genome, can be addressed more thoroughly (51, 52). Conserved loci used for MLST analysis are representative of a microbial species core genome while the genome wide restriction fragments used in AFLP analysis are representative of more variable regions associated with the dispensable genome where gene acquisition and loss occur more frequently.

To address questions of biogeography, biodiversity and evolution within the island-like extreme environments of marine hydrothermal vents, isolates of *Thermococcus* from different geographic regions were analyzed through the high resolution typing methods of MLST and AFLP. *Thermococcus* isolates were derived from the Juan De Fuca Ridge, Gorda Ridge, East Pacific Rise, Mid-Atlantic Ridge, Loihi Seamount, and the Mariana Arc. The primary objective of this study was the comparison among isolates from the Juan De Fuca Ridge, Gorda Ridge and East Pacific Rise, as the majority of the isolates analyzed were derived from these hydrothermal vent systems.

MATERIALS AND METHODS

Thermococcales Isolates

Sample material was collected from hydrothermal vent sites during research cruises between the years of 1988-2008. Both submersible and ROVs were used to collect a diversity of sample material that included plume samples, hot fluids, diffuse fluids, chimney walls, sulfide muds, and Alvinellid polychaete tissue samples. In this collection there are 86 *Thermococcus* isolates cultured from the Juan de Fuca Ridge (JdF), Gorda Ridge, East Pacific Rise (EPR), Mid Atlantic Ridge (MAR) and Loihi Seamount (Table 1). Sample sites within the JdF and EPR are at similar spatial distances providing nested sampling within these two regions (Fig. 1). Distances between vents within regions range from ~65 km to ~450 km, with distances between the two main regions in this study, the JdF and EPR, up to ~7000 km. Study sites and sampling are as previously described (23, 34, 36, 53). Isolates in this culture collection were previously characterized through analysis of the SSU rRNA for genus level associations (23, 34, 36). Four additional isolates were cultured from hydrothermal vent samples collected at Loihi Seamount (Marker 36 and Marker 39) and the Mariana Arc (Bubble Bath and Champagne). Table S1 lists *Thermococcus* type strains included in AFLP and MLST analysis. Cultures of the type strains *Thermococcus kodakarensis* (JCM 12380) and *Thermococcus onnurines* (JCM 13517) were acquired through the Riken BioResource Center (Ibaraki, Japan) and were cultured and treated in the same manner as the other isolates in the collection. Genomic DNA for the type strains *Thermococcus barophilus* and *Thermococcus peptonophilus* were acquired from the American Type Culture Collection (Manassas, VA).

Culturing Conditions

Collected sample material was used to inoculate liquid media for the enrichment of Thermococcales. Media formulations were as previously described (23). All culture tubes and stoppers were pre-sterilized before use. Sterile media was dispensed into Balch tubes with the addition of approximately 0.5 g of sterile elemental sulfur. Balch tubes were sealed with stoppers (Bellco: butyl rubber septum) and crimped with aluminum seal (Bellco: aluminum seals) before the head space of each tube was exchanged with argon gas to remove oxygen, using a gassing manifold with 0.2 μ m filter, 1ml syringe and 28 gauge needle. Positive pressure was removed from culture tubes during gas exchange and prior to inoculating with a 60cc syringe with 0.2 μ m filter and 28 gauge needle. Prepared media were left under positive pressure until inoculated. Media was reduced by the addition of filter sterilized 2.5% sodium sulfide to a final concentration of 0.05 %. Sterile 1ml louver lock syringes and 28 gauge needles were used for adding reducing agent and inoculating media with 0.2ml of active culture when refreshing, or up to 0.5ml of sample material (fluid sample with fine sediment). Inoculated culture tubes were placed in dry sand baths in a drying oven and incubated at temperatures between 78 and 82°C for two to five days. Cultures were checked for growth by epifluorescence microscopy through the addition of 0.25mM SYTO 13 (Molecular Probes, Eugene, OR) in 1X phosphate buffered saline. Cultures were stored at room temperature between inoculations. New isolates were obtained through dilution to extinction, with enriched cultures serial diluted out to 10⁻¹¹. The most dilute culture with growth preceding a dilution with no growth was used in the subsequent dilution series with the process repeated four times in order to isolate a single organism.

DNA Extraction

Genomic DNA (gDNA) extractions were made from freshly cultured cells. Cultured isolates were initially centrifuged at 750 x g for 5 minutes to remove sulfur with the supernatant removed for further centrifugation. Supernatant was centrifuged at 11,000 x g for 10 minutes in a chilled rotor (4°C), with the cell pellet saved for gDNA extraction. The DNeasy Tissue Kit (Qiagen, Valencia, CA) was used following standard protocols with the gDNA being eluted in 10mM Tris with 0.1 mM EDTA at pH 8. A NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, DE) was used to check DNA concentration and purity.

AFLP Reactions

AFLP analysis was performed using the Applied Biosystems (ABI) AFLP Microbial Fingerprinting kit (Applied Biosystems, Carlsbad, CA). Reactions and PCR conditions were as described in kit protocols using the restriction enzymes EcoRI and MseI. Primers for the selective amplification of restriction fragments of between 50 and 500 base pairs were initially designed through an in silico analysis of *Pyrococcus abyssii*, *Pyrococcus furiosus*, and *Pyrococcus horikoshii* genomes. Two selective primer sets were used in separate reactions to obtain two individual AFLP profiles (EcoRI-0 and MseI-CT, and EcoRI-C and MseI-G). AFLP reactions were purified across Sephadex G-75 columns and dried down in a 96 well plate before being resuspended in 15µl of a 1:30 dilution of Liz-500 (ABI) size standard in formamide. Fragment lengths were analyzed using an ABI Prism 3130XL Genetic Analyzer. Electropherograms were optimized with samples rerun or diluted when necessary to obtain relatively similar peak heights among isolates. Electropherogram data

was checked for quality using ABI Genemapper software, and imported into BioNumerics version 4.61 (Applied Maths, Sint-Martens-Latem, Belgium) for further analysis.

Cluster Analysis

Cluster analysis of electropherogram data was performed in BioNumerics version 4.61 using the Pearson product-moment correlation coefficient (54). Cluster analysis of individual primer sets were combined and averaged to construct a dendrogram, with cophenetic correlation coefficients calculated at nodes.

MLST Primer Design

Primers for MLST analysis were designed using nucleotide alignments from the following six annotated genomes: *Thermococcus kodakaraensis* KOD1 (31), *Thermococcus onnurineus* NA1 (27), *Thermococcus gammatolerans* EJ3 (24), *Thermococcus* sp. AM4 (28), *Thermococcus barophilus* MP (55) and *Thermococcus sibiricus* (49). Gene sequences retrieved through GenBank were aligned using the CLUSTALW multiple sequence aligner (56). Candidate genes were screened for conserved regions suitable for the amplification of between 300 and 500 base pairs. Loci associated with information processing and metabolism were the focus in identifying MLST candidate genes, along with the distribution of loci across the genome. Gene loci positions determined from GenBank were plotted for the six reference genomes to check distribution and compare gene synteny (FIG. S1). A portion of the following seven loci were picked for MLST analysis: SSU rRNA, Elongation Factor 1 Alpha subunit, DNA topoisomerase VI Alpha subunit, DNA polymerase II large subunit, threonyl-tRNA synthetase, pyruvate ferredoxin oxidoreductase Beta subunit, and histone acetyltransferase. Primer3 was used to design primers (57). Degenerate primers were

designed for all loci, with the exception of the SSU rRNA gene, in order to account for the diversity present in the six reference genomes. For primer sequences see supplementary material (Table S2). Primer pairs designed for the seven loci resulted in PCR amplicons with the following approximate sizes in base pairs (bp): SSU rRNA 480bp, EF1 α 534bp, DNA polymerase II 540bp, DNA topoisomerase VI 530bp, Histone Acetyltransferase 383bp, Pyruvate ferredoxin 335bp, and Threonyl tRNA synthetase 458bp. The concatenated alignment of all seven loci following the trimming of sequence data and removal of gaps resulted in ~2,650bp of nucleotide sequence for analysis.

Loci PCR and Sequencing

The following PCR reaction mix was used for the SSU rRNA and EF1 α loci: 50ng of gDNA template, 0.5 μ l JS Taq (2.5U/ μ l; Sigma) per 50 μ l rxn, 1X taq PCR buffer, 1.5mM MgCl₂, 0.2mM each deoxynucleoside triphosphate, 0.2 μ M each forward and reverse primer, and molecular grade water to a total volume of 50 μ l. For the SSU rRNA the following PCR cycle was used: an initial 10 min. hot start at 95°C, followed by 30 cycles of denaturation (95°C for 30 sec.), annealing (58°C 30 sec.), and elongation (72°C for 30 sec.). This was followed by a final elongation step at 72°C for 7 min. For the EF1 α loci the PCR cycle was the same as the above with the exception of a 56°C annealing temperature. For the remaining six loci the PCR mix and PCR conditions were the same as the above with the exception of a 0.8 μ M concentration for forward and reverse primers, and 55°C annealing temperature. PCR amplicons were verified and checked for size through gel electrophoresis. Loci were sequenced through Sanger sequencing using ABI BigDye Terminator v3.1. Nucleotide sequences were contiguously assembled using the BioNumerics sequence aligner.

Contiguous nucleotide sequences were verified as having a closest match to *Thermococcus* sequences through an NCBI BLAST search. The ExPASy web site (Swiss Institute of Bioinformatics) was used to determine the correct reading frame for protein coding genes before nucleotide sequences were translated into amino acid sequences using MEGA version 5 (MEGA5) (58). Sequences of individual loci were aligned by ClustalW in MEGA5 with all gaps removed.

Loci Test For Selection

A sequence based test for selection was performed on the entire sequence length of each protein coding locus using the Datamonkey online server (59). Non-synonymous versus synonymous (dN/dS) ratios were calculated on the in-frame nucleotide alignments of individual loci.

Phylogenetic Analysis

Maximum likelihood (ML) phylogenetic trees were constructed using the RAxML black box online server (60). A mixed/partition model was applied to the concatenated alignment of nucleotide (SSUrRNA) and amino acid sequences, using the default nucleotide CAT model (60, 61) and the WAG (62) protein model, with per gene optimization of branch lengths and 100 bootstraps. Tree files were run ten times with the lowest log likelihood selected. The concatenated ML tree was rooted using the Crenarchaeota *Staphylothermus marinus* as an outgroup. Homologs in *Staphylothermus marinus* for six of the seven loci (no homolog was found for DNA polymerase II) were independently aligned with isolates and type strains before being concatenated for phylogenetic analysis and rooting of tree using the parameters described above. Clades were assigned to distinct groupings of three or more

isolates. Bootstrap values of twenty and above were reported. ML trees were constructed for individual loci as well as the concatenated alignment of all seven loci. The individual SSU rRNA tree was rooted using *Staphylothermus marinus* as an outgroup and with the addition of *Palaeococcus ferriphilus* and *Palaeococcus helgesonii* sequences (17, 63). Unrooted trees were constructed from amino acid sequences for the remaining six loci. Isolate groupings between loci were checked for congruence, through visual comparisons between individual loci trees, to determine if isolate gene histories were in agreement.

MLST Clade Analysis

Average Nucleotide Identity (ANI). The 2,648bp nucleotide sequence of concatenated loci was used for ANI analysis of clades (38–40). ANI values were calculated through BLAST (bl2seq; NCBI BLAST) with the lowest similarity value in comparisons between isolates recorded as the ANI for a particular clade.

%GC. The six protein coding loci were used for the analysis of %GC of clades using the Datamonkey online server (59).

Linkage Analysis. Linkage between loci was tested using the Non-redundant database (NRDB; PubMLST) and Linkage Analysis (LIAN) version 3.5 (64). Isolate loci were coded through NRDB and the null hypothesis of linkage equilibrium for clades was tested using LIAN 3.5 with the standardized index of association (I_A) reported.

Codon GC Ratios

Ratios of %GC for first and third codon positions were analyzed as a measure of gene history (65). The average %GC for first and third codon positions of the six protein coding

loci were calculated for individual clades, and individual isolates and type strains when not included in Clades I-X, using MEGA5 (58). Two dimensional plots of %GC were constructed for individual loci as well as a combined plot of the average %GC for all six loci. First and third codon %GC were calculated by taking the average of a codon position for all isolates of a clade, or the codon %GC for single isolates, and plotting on X and Y axes. The averages for first and third codon position were plotted in a similar manner previously shown to differentiate genomic variation in bacteria (66).

Mantel Test

A Mantel test comparing geographic distance and genetic distance was performed using the statistical software zt (67). Pairwise genetic distances were calculated in MEGA5 using the Maximum Composite Likelihood model (58). Geographic distances between vents were recorded in kilometers and calculated from hydrothermal vent latitude and longitude. Matrices of isolate genetic and geographic distance were compared to test the null hypothesis of independence between matrices. A simple Mantel test with 10,000 randomizations was performed on all isolates, on isolates from the two main regions being investigated (JdF and EPR) and on phylogenetically related clades.

Analysis of Molecular Variance

Analysis of Molecular Variance (AMOVA) was calculated in Arlequin version 3.11 (68) on the concatenated nucleotide sequences of all seven loci. AMOVA was used to test for correlations between sample type and sample site. Isolates were grouped by the sample type that they were isolated from, and by the hydrothermal vent site from which they were isolated, with the p value significance test for variance carried out using 10,000 permutations.

Principal Component Analysis

Principal Component Analysis (PCA) was performed on AFLP band calling data using BioNumerics version 4.61. AFLP band calling data were collected through the automated selection of bands from both primer sets using the following parameters: a minimum profiling of 5% for primer set 1, a minimum profiling of 10% for primer set 2, and for both primer sets the optimization and position tolerances for selecting bands were set to 0.10 %. Band calling resulted in an average of 14 bands per isolate for primer set 1 and an average of 15 bands per isolate for primer set 2. Band calling data for both primer sets were combined and converted into a binary presence absence table for PCA. Default settings were applied for PCA, subtracting the average for characters.

RESULTS

Cluster Analysis

The dendrogram topology for cluster analysis of AFLP data is well supported by cophenetic correlation coefficients (Fig. 2). Through cluster analysis isolates are grouped by similarity with a general pattern of isolates grouped by region. Isolate diversity is exemplified by the number of distinct clusters found within the two main regions, with up to six distinct clusters identified from the JdF and five from the EPR. Many of the clusters consist of isolates from more than one vent site within a region, more than one sample type, and isolates from different sampling years. Isolate groupings also illustrate the dispersal potential between vent sites within a region with clusters containing isolates from vents spread throughout a region. Isolates previously identified through SSU rRNA analysis as *Pyrococcus* cluster together with the exception of isolate MV7. Type strains analyzed by AFLP as well as isolates from regions other than the JdF, Gorda Ridge, or EPR had low similarity with other isolates in this study with the exception of the type strain *Thermococcus onnurines* isolated from the Papua New Guinea-Australia-Canada-Manus (PACMANUS) field (19). The cluster containing *T.onnurines*, the Gorda Ridge isolates, and isolate CX3 showed high similarity ($\geq 68\%$) between isolates from different regions.

Loci Test for Selection

Analysis of the individual alignments for the six protein coding loci (EF1 α , DNA topoisomerase VI, DNA polymerase II, threonyl-tRNA synthetase, pyruvate ferredoxin oxidoreductase and histone acetyltransferase) did not show evidence for any of the loci

undergoing strong selection (Table S3), making these loci suitable for MLST analysis. Ratios of dN/dS were consistent with dN/dS ratios reported for conserved genes under purifying selection (69).

Phylogenetic Analysis

The ML tree for the concatenated alignment of MLST loci differentiated isolates into clades that are in agreement with isolate groupings through cluster analysis (Fig. 2 & 3). Isolates from the two main regions investigated, the JdF and EPR, were differentiated into distinct lineages that are phylogenetically related across these two regions (Fig. 3). There is a general pattern of isolates differentiating by region with individual clades made up of isolates from the same region. Regional groupings are most apparent in the isolates associated with the JdF and EPR, but regional groupings are also observed in the two isolates from the Mariana, the type strains *T. peptonophilus* and *T. kodakaraensis* both from Japan, and the type strains *T. sp.* 4557 and *T. sp.* AM4 from the EPR, grouping with two different lineages that are both from the EPR. Exceptions to regional groupings are seen in Clade I and Clade X. Clade I contains isolates from the JdF Ridge and Loihi, and the type strain *T. barophilus* from the Mid Atlantic Ridge (MAR). Clade X contains isolates from the Gorda Ridge, an isolate from the CoAxial Segment of the JdF and the type strain *T. onnurines* from the PACMANUS Basin, with all of these isolates having high sequence similarity. *Pyrococcus* type strains along with Clade I, the type strain *T. sibiricus* and isolate MV5 are placed in a basal position in the phylogenetic tree, ancestral to Clades II-X. Tree distances for Clade I and *Pyrococcus* type strains to Clades II - X are comparable, with the type strain *P. yayanosii* having the shortest tree distance to Clades II-X from this basal group. The divergence

between *Pyrococcus* type strains and *Thermococcus* isolates in the ML tree of concatenated loci, rooted with *Staphylothermus marinus*, is unresolved. The ML tree for the SSU rRNA gene rooted with *Staphylothermus marinus* and with the inclusion of *Palaeococcus* species, places Clade I in an ancestral position to both the *Pyrococcus* type strains as well as the other *Thermococcus* isolates found in Clades II-X (Fig. 4). Unrooted ML trees for individual loci show congruence in isolate grouping between the six protein coding loci with the same isolates grouped across trees (Fig. S2-S7). Clade X showed the greatest variation in placement within trees (see placement in the pyruvate ferredoxin oxidoreductase tree in comparison to other loci trees).

MLST Clade Analysis

ANI. Analysis of clade ANI allowed for clade diversity to be expressed in the context of species level boundaries ($ANI \geq 95\%$) (38–40). Clades I, II, and III have ANI values below 95% reflecting diversity beyond the individual species level (Table 2). Clades IV-X have ANI values at or above 95% and can therefore be theoretically categorized as individual species. Pairwise comparisons between all clades (data not shown) result in ANI values <95%, therefore individual clades can be considered to be distinct species.

%GC. Analysis of %GC was applied in order to look for differing gene histories of clades, as GC content has been correlated with phylogeny (65). Variation in %GC was observed between all clades (Table 2) with Clade I having the lowest %GC in comparison to other *Thermococcus* clades in this study, and with a %GC closer to what has been reported for *Pyrococcus* genera (70). Clade X has the next lowest %GC in comparison to clades II-IX.

Linkage Analysis. Linkage analysis was applied to test for linkage between MLST loci. In clonal organisms linkage between loci (linkage disequilibria) is expected with evidence of unlinked loci (linkage equilibria) associated with gene transfer or recombination events. Linkage analysis for individual clades rejected the null hypothesis of linkage equilibria, with the measure of linkage (I_A) significantly different from zero and or not significant when approaching zero. This was in agreement with individual loci phylogenetic trees where congruence in isolate groupings was observed between individual trees.

Codon GC Ratios

Plots of the average %GC for individual protein coding loci (Fig. S8) and the combined plot (Fig. 5), illustrate the differences in %GC at first and third codon positions for clades and isolates or type strains in the phylogenetic tree. Clade I along with the type strain *Thermococcus sibiricus* and isolate MV5 have %GC ratios that are closest to *Pyrococcus abyssi*, *Pyrococcus sp.* NA2, *Pyrococcus horikoshi*, and *Pyrococcus furiosus*, in agreement with the phylogenetic tree. *Pyrococcus yayanosii*, which has been shown to have a higher GC content in comparison to other *Pyrococcus* (71), has %GC ratios closer to those calculated for the *Thermococcus* isolates in Clades II-X.

Mantel Test

Mantel's analysis testing for correlations between genetic distance and geographic distance did not find a significant correlation when looking at all isolates or isolates from the JdF or EPR (Table S4). The lack of a correlation at these scales is in part due to the degree of genetic diversity present within regions and within vent sites. Analysis of phylogenetically related clades from the two main regions (JdF and EPR) did find a significant correlation

between genetic distance and geographic distance. An exception to this is the analysis of Clades IX and X, where geographic distance and genetic distance are not correlated, reflecting the unique nature of these isolates (in particular Clade X) with isolates not differentiated at regional levels.

Analysis of Molecular Variance

AMOVA Analysis of MLST nucleotide data did not show a significant correlation between isolate groupings and environmental characteristics. Variation in isolates grouped by sample type (-1.1% of variation $p = 0.324$) and by sample site (16.2 % of variation $p = 0.121$) was not significant. This suggests that isolate groupings are not correlated with sample site or the descriptions for sample type that the isolates were collected and cultured from.

Principal Component Analysis

Principal component analysis of AFLP band calling data described 16.8% of the total variation in the first three principal components (Fig. 6). Principal components 1 and 2 plotted on the X and Y axes respectively differentiate isolates from Clade I and Clade X from other isolates in this study. Principal component 3 plotted on the Z axis differentiates isolates into larger regionally related groups.

DISCUSSION

A comparison of *Thermococcus* isolates through AFLP and MLST analysis have made biogeographic patterns evident. The strongest biogeographic patterns observed are between northern isolates from the Juan de Fuca Ridge and southern isolates from the East Pacific Rise. These two regional populations, at the JdF Ridge stretching from the Middle Valley segment to the Cleft segment and at the EPR from the Guaymas Basin to 21° south, have diverged at least three times (Fig. 4). Phylogenetic analysis suggests that lineages shared between these two regions colonized and then diverged from one another; however these data do not suggest a simple model of isolation by distance. Lineages from the JdF Ridge consist of isolates with high similarity from vent segments spread throughout this vent system, and lineages from the South EPR share similarity with isolates and type strains from the North EPR and Guaymas Basin suggesting that there is high dispersal in the *Thermococcus* genus. Isolation and subsequent divergence among these lineages may be due to the discontinuous nature of the plate boundary and its associated hydrothermal vents, as the North American Plate overrides the Pacific plate creating a physical barrier to dispersal. Analysis of *Thermococcus* isolates from a more contiguous series of hydrothermal venting regions, like the Mid Atlantic Ridge, would be useful in further resolving the relationship between dispersal and genetic distance in this genus.

The distribution of *Thermococcus* and dispersal observed within regional populations may result from an ability to survive outside of the hydrothermal vent environment for extended periods of time, allowing for dispersal between active vents and the potential for leapfrogging from one vent to another. Dispersal over great distances is a possibility that

must be considered particularly when looking at isolates in Clade X, where high sequence and AFLP profile similarity is apparent in isolates from vent sites ~9500km apart. Although Clade X was anomalous in these analyses it poses an interesting dilemma for developing theories to explain the relatedness of these isolates. Some possible explanations for the high similarity observed in these isolates may include: a recent divergence, the acquisition of advantageous genes allowing for extended survival outside of the vent environment or an increased metabolic potential resulting in the generation of a niche specific ecotype. Isolates in Clade X collected from the Gorda Ridge have already been shown to grow over an atypically broad range of temperatures (45-90°C) for hyperthermophiles (72). The discovery of the CODH gene cassette in the type strain *T. onnurines*, found in Clade X, and the occurrence of these genes in other distantly related type strains (Fig. 4) may be indicative of a gene transfer event providing a selective advantage. With five of the seven isolates from Clade X collected from a plume event it is also intriguing to consider this population representative of a unique subsurface ecotype (72). During a plume event buoyant hydrothermal fluids are released from a subseafloor reservoir (73). Having residence times of from months to years and detectable thousands of kilometers from their origin (74, 75), hydrothermal plumes could be vehicles for the dispersal of microorganisms from hydrothermal vents over great distances. Long distance dispersal of thermophiles either through continuous venting or plume events has been proposed with the detection of thermophilic activity in enrichments of deep-sea sediment collected far from any known hydrothermal venting (76). The true nature of Clade X will require further investigating but in having the full-genome sequence of the type strain *T. onnurines*, comparisons can be made

against other fully sequenced *Thermococcus* genomes to address questions regarding the anomalous nature of this group.

Clade I represents a unique group in comparison to the other clades in this study as it also contains isolates from different regions, although with lower sequence and AFLP profile similarity between isolates there may be inadequate sampling to represent biogeographic patterns in this group. The most notable characteristic of Clade I is the phylogenetic distance between this clade and the other *Thermococcus* isolates in this study and the close proximity of Clade I to *Pyrococcus* type strains (Fig. 3). Both Clade I and *Pyrococcus* type strains are placed in an ancestral position to *Thermococcus* isolates from Clades II-X (Fig. 3 and 4). Clade I also has GC ratios that are closer to those reported for most *Pyrococcus* (Fig. 5 and Table 2), illustrating the divergence of the Clade I lineage and the *Pyrococcus* genus from the other *Thermococcus* isolates examined in this study. The SSU rRNA gene tree rooted with the Crenarchaeota *Staphylothermus marinus* and including *Palaeococcus* species (Fig. 4) places Clade I in an ancestral position to both *Pyrococcus* type strains and *Thermococcus* isolates from Clades II-X, resulting in paraphyletic groupings for *Thermococcus* and *Pyrococcus* genera. These data suggest that Clade I is a close sister lineage to *Pyrococcus* requiring reclassification.

The significant biodiversity present in the *Thermococcus* genus has been made evident in this study with examples of three or more species level groups co-occurring within the same hydrothermal vent. Lineages observed through cluster and phylogenetic analyses are representative of distinct species that are prevalent throughout a region. Divergence between lineages encompassing broader levels of diversity, for example Clades I and II from

Clades VII and VIII, can also be observed in the phylogenetic analysis (Fig. 3). The conserved nature of these two larger lineage groupings is evident in the fact that they are present in the two main regions investigated (JdF and EPR) and have been maintained spatially and temporally, diverging at regional levels due to isolated populations. These distinct lineages have likely evolved through ecological adaptations which have allowed them to coexist at the same vent site through the utilization of different niches. A correlation between environmental habitat and phylogeny has been described in a previous study which examined many of the same isolates used in this study. The hydrothermal vent habitats of sulfide chimneys and seafloor zones were proposed as the differing selective environments involved in the maintenance of diversity and correlated with phylogeny in the *Thermococcus* isolates investigated (36). Although no significant correlation between isolate lineages and environmental characteristics were found in our study, the larger divergence between lineages observed in the phylogenetic tree may well be due to the differing source environments previously described (36). Defining the source habitat is extremely difficult though particularly when considering the continuous mixing of fluids within subsurface hydrothermal conduits and the limitations imposed by current sampling techniques. While it may not be possible to improve greatly on sampling techniques in the near future, the biodiversity described can be investigated more thoroughly through metagenomic and metatranscriptomic analyses of the metabolic characteristics associated with different lineages with an emphasis on how these metabolic characteristics relate to the utilization of different habitats.

Being cognizant of the genetic diversity present in the *Thermococcus* genus, AFLP and MLST data were investigated for evidence of genomic rearrangements and gene transfer

events. The conserved nature of MLST loci, evident in dN/dS ratios (Table S3), linkage analysis (Table 2), and in the congruence in isolate groupings between individual loci phylogenetic trees (Fig. S2-S7) supports the theory that conserved housekeeping genes within a species pangenome are unlikely to undergo gene transfer or recombination (51, 52). AFLP analysis also tended to support the relative low impact of genome rearrangements on individual lineages, with the same lineages identified between both typing methods (Fig 2 and 3). Although AFLP analysis detected greater genomic variation in lineages as compared to MLST, the effects of genome rearrangements and the subsequent genomic variation is not extreme enough as to obscure isolate groupings altogether. In contrast, the analysis of AFLP band calling data through PCA (Fig. 6) provides evidence suggestive of regionally related recombination detected in restriction fragment length variation, having a cohesive effect in isolates at regional levels.

Horizontal gene transfer (HGT) and recombination are believed to play a significant role in the mixing of populations and shaping of microbial diversity (77–79). The significance of HGT and recombination have been brought into question however (e.g., extrinsic and intrinsic barriers to gene transfer and recombination), with a decrease in both of these events observed as geographic distance and phylogenetic distance between microorganisms increase (33, 80, 81). While MLST loci in this study do not provide evidence for gene transfer between *Thermococcus* isolates, there is evidence from the analysis of AFLP data by PCA (Fig. 6) suggestive of geographically related recombination events. Through PCA of phylogenetically distinct lineages present in the two main regions studied, lineages are differentiated into larger regionally related groups (observed in the plot of isolates on the Z axis). This is in contrast to what is depicted in the ML phylogenetic tree

constructed from MLST data (Fig 3) where isolates from a shared region are differentiated into distinct clades. Shared characteristics at regional levels, between divergent lineages, may be detected in AFLP data through PCA as the genome wide restriction fragment lengths used for AFLP are more representative of variable regions of the genome where gene acquisition and loss occur more frequently.

In studies from both geothermal hot springs and marine hydrothermal vents biogeographic patterns have been demonstrated for mobile genetic elements and viruses. Through the analysis of viral genomes associated with *Sulfolobus* geographically distinct viral populations have been identified, with viruses shown to be associated with geothermal regions (82). In the same study analysis of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) regions in *Sulfolobus*, which provide a record of host-virus interactions, also demonstrated biogeographic patterns. In another study looking at integron-integrase genes at marine hydrothermal vents, phylogenetically distinct lineages were identified from different sites suggesting endemism (83). This study proposed that gene transfer associated with integron genes may take place between populations within a vent habitat, but not between populations from distant vents. Biogeography in integron families has also been described in a marine *Vibrio*, with different species of *Vibrio* from a shared geographic region having higher similarity in integron genes in comparison to the same species of *Vibrio* from different geographic regions (84). In the Thermococcales, geographically associated recombination has been reported through the identification of a shared 16kb region that is flanked by IS elements, found in *Pyrococcus furiosus* and *Thermococcus litoralis*, both of which were isolated from Vulcano Island Italy (50). Mobile genetic elements like insertion sequence elements and transposons, identified in a number of

Thermococcales genomes (12, 31, 48–50), along with CRISPR regions and other viral related elements (24, 31, 49, 55, 85), can be investigated through genome comparisons between *Thermococcus* populations from varying geographic locations for additional evidence in support of regionally related recombination. When considering the biogeographic patterns that have already been demonstrated for mobile genetic elements in both terrestrial and marine environments, there is growing evidence for a biogeographic component to gene transfer and recombination involved in the shaping of microbial diversity.

The ability to observe divergence in microorganism is largely dependent on the geographic scale investigated and level of genetic resolution applied, by identifying the appropriate scale and resolution the impact that geography and barriers to dispersal have on the overall diversity of microorganisms are now being realized. Biogeographic patterns illustrated in the Archaeal genus *Thermococcus* show that even in microorganisms with dispersal over thousands of kilometers, divergence can occur when populations are isolated from one another. This study begins to elucidate the level of genomic resolution required to track population divergence in isolates of *Thermococcus* from hydrothermal habitats.

LITERATURE CITED

1. **Whitaker RJ.** 2006. Allopatric origins of microbial species. *Phil. Trans. R. Soc. B Biol. Sci.* **361**:1975–84.
2. **Baas-Becking LGM.** 1934. *Geobiologie of Inleiding Tot De Milieukunde.* Van Stockum & Zoon, The Hague.
3. **Cho JC, Tiedje JM.** 2000. Biogeography and degree of endemism of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* **66**:5448–56.
4. **Papke RT, Ward DM.** 2004. The importance of physical isolation to microbial diversification. *FEMS Microbiol. Ecol.* **48**:293–303.
5. **Van Dover CL.** 2000. *The ecology of deep-sea hydrothermal vents.* Princeton University Press.
6. **Reysenbach A, Shock E.** 2002. Merging genomes with geochemistry in hydrothermal ecosystems. *Science* **296**:1077–1082.
7. **Ramette A, Tiedje JM.** 2007. Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb. Ecol.* **53**:197–207.
8. **Papke RT, Ramsing NB, Bateson MM, Ward DM.** 2003. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**:650–9.
9. **Whitaker RJ, Grogan DW, Taylor JW.** 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**:2002–2004.
10. **DeChaine EG, Bates AE, Shank TM, Cavanaugh CM.** 2006. Off-axis symbiosis found: characterization and biogeography of bacterial symbionts of *Bathymodiulus* mussels from Lost City hydrothermal vents. *Environ. Microbiol.* **8**:1902–12.
11. **McAllister SM, Davis RE, McBeth JM, Tebo BM, Emerson D, Moyer CL.** 2011. Biodiversity and emerging biogeography of the neutrophilic iron-oxidizing *Zetaproteobacteria*. *Appl. Environ. Microbiol.* **77**:5445–57.
12. **Escobar-Páramo P, Ghosh S, DiRuggiero J.** 2005. Evidence for genetic drift in the diversification of a geographically isolated population of the hyperthermophilic archaeon *Pyrococcus*. *Molec. Biol. Evol.* **22**:2297–303.

13. **Button DK, Schut F, Quang P, Martin R, Robertson BR.** 1993. Viability and isolation of marine bacteria by dilution culture: theory, procedures, and initial results. *Appl. Environ. Microbiol.* **59**:881–91.
14. **Kaeberlein T, Lewis K, Epstein S.** 2002. Isolating “uncultivable” microorganisms in pure culture in a simulated natural environment. *Science* **296**:1127–9.
15. **Fiala G, Stetter K.** 1986. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C. *Arch. Microbiol.* **145**:56–61.
16. **Zillig W, Holz I, Janekovic D, Schäfer W, Reiter WD.** 1983. The Archaeobacterium *Thermococcus celer* represents, a novel genus within the thermophilic branch of the Archaeobacteria. *Syst. Appl. Microbiol.* **4**:88–94.
17. **Takai K, Sugai A, Itoh T, Horikoshi K.** 2000. *Palaeococcus ferrophilus* gen. nov., sp. nov., a barophilic, hyperthermophilic archaeon from a deep-sea hydrothermal vent chimney. *Int. J. Syst. Evol. Microbiol.* **50**:489–500.
18. **Teske A, Edgcomb V, Rivers AR, Thompson JR, De Vera Gomez A, Molyneux SJ, Wirsén CO.** 2009. A molecular and physiological survey of a diverse collection of hydrothermal vent *Thermococcus* and *Pyrococcus* isolates. *Extremophiles* **13**:905–15.
19. **Bae SS, Kim YJ, Yang SH, Lim JK, Jeon JH, Lee HS, Kang SG, Kim S-J, Lee J-H.** 2006. *Thermococcus onnurineus* sp. nov., a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent area at the PACMANUS field. *J. Microbiol. Biotechnol.* **16**:1826–1831.
20. **Erauso G, Reysenbach AL, Godfroy A, Meunier JR, Crump B, Partensky F, Baross JA, Marteinsson V, Barbier G, Pace NR, Prieur D.** 1993. *Pyrococcus abyssi* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch. Microbiol.* **160**:338–349.
21. **Atomi H, Fukui T, Kanai T, Morikawa M, Imanaka T.** 2004. Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. *Archaea* **1**:263–267.
22. **Garrity GM, Boone DR, Castenholz RW.** 2001. *Bergey’s Manual of Systematic Bacteriology*, Vol. 1: The Archea and the deeply branching and phototrophic bacteria. Springer, New York.
23. **Holden JF, Takai K, Summit M, Bolton S, Zyskowski J, Baross JA.** 2001. Diversity among three novel groups of hyperthermophilic deep-sea *Thermococcus* species from three sites in the northeastern Pacific Ocean. *FEMS Microbiol. Ecol.* **36**:51–60.

24. **Zivanovic Y, Armengaud J, Lagorce A, Leplat C, Guérin P, Dutertre M, Anthouard V, Forterre P, Wincker P, Confalonieri F.** 2009. Genome analysis and genome-wide proteomics of *Thermococcus gammatolerans*, the most radioresistant organism known amongst the Archaea. *Genome Biol.* **10**:R70.
25. **Adams MW.** 1993. Enzymes and proteins from organisms that grow near and above 100 degrees C. *Ann. Rev. Microbiol.* **47**:627–658.
26. **Robb FT.** 1995. *Archaea: a laboratory manual.* Cold Spring Harbor Laboratory Press.
27. **Lee HS, Kang SG, Bae SS, Lim JK, Cho Y, Kim YJ, Jeon JH, Cha S-S, Kwon KK, Kim H-T, Park C-J, Lee H-W, Kim S Il, Chun J, Colwell RR, Kim S-J, Lee J-H.** 2008. The complete genome sequence of *Thermococcus onnurineus* NA1 reveals a mixed heterotrophic and carboxydrotrophic metabolism. *J. Bacteriol.* **190**:7491–9.
28. **Oger P, Sokolova TG, Kozhevnikova DA, Chernyh NA, Bartlett DH, Bonch-Osmolovskaya EA, Lebedinsky AV.** 2011. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus* sp. strain AM4, capable of organotrophic growth and growth at the expense of hydrogenogenic or sulfidogenic oxidation of carbon monoxide. *J. Bacteriol.* **193**:7019–20.
29. **Sokolova TG, Henstra A-M, Sipma J, Parshina SN, Stams AJM, Lebedinsky AV.** 2009. Diversity and ecophysiological features of thermophilic carboxydrotrophic anaerobes. *FEMS Microbiol. Ecol.* **68**:131–41.
30. **Kim YJ, Lee HS, Kim ES, Bae SS, Lim JK, Matsumi R, Lebedinsky AV, Sokolova TG, Kozhevnikova DA, Cha S-S, Kim S-J, Kwon KK, Imanaka T, Atomi H, Bonch-Osmolovskaya EA, Lee J-H, Kang SG.** 2010. Formate-driven growth coupled with H₂ production. *Nature* **467**:352–5.
31. **Fukui T, Atomi H, Kanai T, Matsumi R, Fujiwara S, Imanaka T.** 2005. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* **15**:352–363.
32. **Jannasch HW, Wirsén CO, Molyneux SJ, Langworthy TA.** 1992. Comparative physiological studies on hyperthermophilic Archaea isolated from deep-sea hot vents with emphasis on *Pyrococcus* strain GB-D. *Appl. Environ. Microbiol.* **58**:3472–81.
33. **Reno ML, Held NL, Fields CJ, Burke PV, Whitaker RJ.** 2009. Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc. Natl. Acad. Sci. U.S.A.* **106**:8605–10.
34. **Huber JA, Butterfield DA, Baross JA.** 2006. Diversity and distribution of seafloor Thermococcales populations in diffuse hydrothermal vents at an active deep-sea volcano in the northeast Pacific Ocean. *J. Geophysic. Res.* **111**.

35. **Lepage E, Forterre P, Marguet E, Matte-Tailliez O, Geslin C, Zillig W, Tailliez P.** 2004. Molecular diversity of new *Thermococcus* isolates from a single area of hydrothermal deep-sea vents as revealed by randomly amplified polymorphic DNA fingerprinting and 16S rRNA gene sequence analysis. *Appl. Environ. Microbiol.* **70**:1277–1286.
36. **Summit M, Baross JA.** 2001. A novel microbial habitat in the mid-ocean ridge subseafloor. *Proc. Natl. Acad. Sci. U.S.A.* **98**:2158–63.
37. **Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* **95**:3140–3145.
38. **Konstantinidis KT, Tiedje JM.** 2005. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **102**:2567–72.
39. **Konstantinidis KT, Ramette A, Tiedje JM.** 2006. The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**:1929–40.
40. **Konstantinidis KT, Ramette A, Tiedje JM.** 2006. Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Appl. Environ. Microbiol.* **72**:7286–93.
41. **Olive DM, Bean P.** 1999. Principles and applications of methods for DNA-based typing of microbial organisms. *J. Clin. Microbiol.* **37**:1661–1669.
42. **Lin JJ, Kuo J, Ma J.** 1996. A PCR-based DNA fingerprinting technique: AFLP for molecular typing of bacteria. *Nucleic Acids Res.* **24**:3649–50.
43. **Savelkoul PHM, Aarts HJM, Haas JDE, Dijkshoorn L, Duim B.** 1999. Amplified-Fragment Length Polymorphism Analysis: the state of an art. *J. Clin. Microbiol.* **37**:3083–3091.
44. **Rademaker JL, Hoste B, Louws FJ, Kersters K, Swings J, Vauterin L, Vauterin P, De Bruijn FJ.** 2000. Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *Internat. J. Syst. Evol. Microbiol.* **50**:665–77.
45. **Maeder DL, Weiss RB, Dunn DM, Cherry JL, González JM, DiRuggiero J, Robb FT.** 1999. Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* **152**:1299–1305.

46. **DiRuggiero J, Santangelo N, Nackerdien Z, Ravel J, Robb FT.** 1997. Repair of extensive ionizing-radiation DNA damage at 95 degrees C in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* **179**:4643–5.
47. **Tapias A, Leplat C, Confalonieri F.** 2009. Recovery of ionizing-radiation damage after high doses of gamma ray in the hyperthermophilic archaeon *Thermococcus gammatolerans*. *Extremophiles* **13**:333–43.
48. **Zivanovic Y, Lopez P, Philippe H, Forterre P.** 2002. *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. *Nucleic Acids Res.* **30**:1902–1910.
49. **Mardanov AV, Ravin NV, Svetlitchnyi VA, Beletsky AV, Miroshnichenko ML, Bonch-Osmolovskaya EA, Skryabin KG.** 2009. Metabolic versatility and indigenous origin of the archaeon *Thermococcus sibiricus*, isolated from a siberian oil reservoir, as revealed by genome analysis. *Appl. Environ. Microbiol.* **75**:4580–8.
50. **Diruggiero J, Dunn D, Maeder DL, Holley-Shanks R, Chatard J, Horlacher R, Robb FT, Boos W, Weiss RB.** 2000. Evidence of recent lateral gene transfer among hyperthermophilic archaea. *Molec. Microbiol.* **38**:684–93.
51. **Tettelin H, Riley D, Cattuto C, Medini D.** 2008. Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* **11**:472–477.
52. **Medini D, Donati C, Tettelin H, Massignani V, Rappuoli R.** 2005. The microbial pan-genome. *Curr. Opin. Genet. Develop.* **15**:589–94.
53. **Davis RE, Moyer CL.** 2008. Extreme spatial and temporal variability of hydrothermal microbial mat communities along the Mariana Island Arc and southern Mariana back-arc system. *J. Geophys. Res.* **113**:1–17.
54. **Häne BG, Jäger K, Drexler HG.** 1993. The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprint profiles than band matching algorithms. *Electrophoresis* **14**:967–972.
55. **Vannier P, Marteinson VT, Fridjonsson OH, Oger P, Jebbar M.** 2011. Complete genome sequence of the hyperthermophilic, piezophilic, heterotrophic, and carboxydophilic archaeon *Thermococcus barophilus* MP. *J. Bacteriol* **193**:1481–2.
56. **Thompson JD, Higgins DG, Gibson TJ.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–80.

57. **Rozen S, Skaletsky H.** 2000. Primer3 on the WWW for general users and for biologist programmers. *Meth. Molec. Biol.* **132**:365–386.
58. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molec. Biol. Evol.* **28**:2731–9.
59. **Pond SLK, Frost SDW.** 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**:2531–3.
60. **Stamatakis A, Hoover P, Rougemont J.** 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* **57**:758–71.
61. **Stamatakis A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**:2688–90.
62. **Whelan S, Goldman N.** 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molec. Biol. Evol.* **18**:691–9.
63. **Amend JP, Meyer-Dombard DR, Sheth SN, Zolotova N, Amend AC.** 2003. *Palaeococcus helgesonii* sp. nov., a facultatively anaerobic, hyperthermophilic archaeon from a geothermal well on Vulcano Island, Italy. *Arch. Microbiol.* **179**:394–401.
64. **Haubold B, Hudson RR.** 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Bioinformatics* **16**:847–848.
65. **Muto A, Osawa S.** 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc.Natl. Acad. Sci. U.S.A.* **84**:166–9.
66. **Kaplan JB, Fine DH.** 1998. Codon usage in *Actinobacillus actinomycetemcomitans*. *FEMS Microbiol.* **163**:31–6.
67. **Bonnet E, Peer Y Van De.** 2002. zt: a software tool for simple and partial Mantel tests. *Compute* **7**:1–12.
68. **Excoffier L, Laval G, Schneider S.** 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinfo. Online* **1**:47–50.
69. **Kuhn G, Francioli P, Blanc D.** 2006. Evidence for clonal evolution among highly polymorphic genes in methicillin-resistant *Staphylococcus aureus*. *J. Bacteriol.* **188**:169–178.

70. **Lecompte O, Ripp R, Puzos-Barbe V, Duprat S, Heilig R, Dietrich J, Thierry JC, Poch O.** 2001. Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic Archaea. *Genome Res.* **11**:981–93.
71. **Jun X, Lupeng L, Minjuan X, Oger P, Fengping W, Jebbar M, Xiang X.** 2011. Complete genome sequence of the obligate piezophilic hyperthermophilic archaeon *Pyrococcus yayanosii* CH1. *J. Bacteriol.* **193**:4297–4298.
72. **Summit M, Baross JA.** 1998. Thermophilic seafloor microorganisms from the 1996 north Gorda Ridge eruption. *Deep Sea Res. Part II: Topical Studies in Oceanography* **45**:2751–2766.
73. **Lupton JE, Baker ET, Massoth GJ.** 1999. Helium, heat, and the generation of hydrothermal event plumes at mid-ocean ridges. *Earth Planet. Sci.* **171**:343–350.
74. **Lupton JE.** 1996. A far-field hydrothermal plume from Loihi Seamount. *Science* **272**:976–979.
75. **Lupton J, Baker E, Garfield N, Massoth G, Feely R, Cowen J, Greene R, Rago T.** 1998. Tracking the evolution of a hydrothermal event plume with a RAFOS neutrally buoyant drifter. *Science* **280**:1052–5.
76. **Dobbs FC, Selph KA.** 1997. Thermophilic bacterial activity in a deep-sea sediment from the Pacific Ocean. *Aquat. Microbial Ecol.* **13**:209–212.
77. **Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT.** 2006. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* **16**:1099–108.
78. **Nesbø CL, Dlutek M, Doolittle WF.** 2006. Recombination in *Thermotoga*: implications for species concepts and biogeography. *Genetics* **172**:759–69.
79. **Gogarten JP, Doolittle WF, Lawrence JG.** 2002. Prokaryotic evolution in light of gene transfer. *Molec. Biol. Evol.* **19**:2226–38.
80. **Whitaker RJ, Grogan DW, Taylor JW.** 2005. Recombination shapes the natural population structure of the hyperthermophilic archaeon *Sulfolobus islandicus*. *Molec. Biol. Evol.* **22**:2354–61.
81. **Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, Krause DJ, Whitaker RJ.** 2012. Patterns of gene flow define species of thermophilic Archaea. *PLoS Biol.* **10**:e1001265.
82. **Held NL, Whitaker RJ.** 2009. Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environ. Microbiol.* **11**:457–66.

83. **Elsaied H, Stokes HW, Nakamura T, Kitamura K, Fuse H, Maruyama A.** 2007. Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. *Environ. Microbiol.* **9**:2298–312.
84. **Boucher Y, Cordero OX, Takemura A, Hunt DE, Schliep K, Baptiste E, Lopez P, Tarr CL, Polz MF.** 2011. Local mobile gene pools rapidly cross species boundaries to create endemism within global *Vibrio cholerae* populations. *mBio* **2**:e00335-10
85. **Portillo MC, Gonzalez JM.** 2009. CRISPR elements in the Thermococcales: evidence for associated horizontal gene transfer in *Pyrococcus furiosus*. *J. Appl Genet.* **50**:421–30.

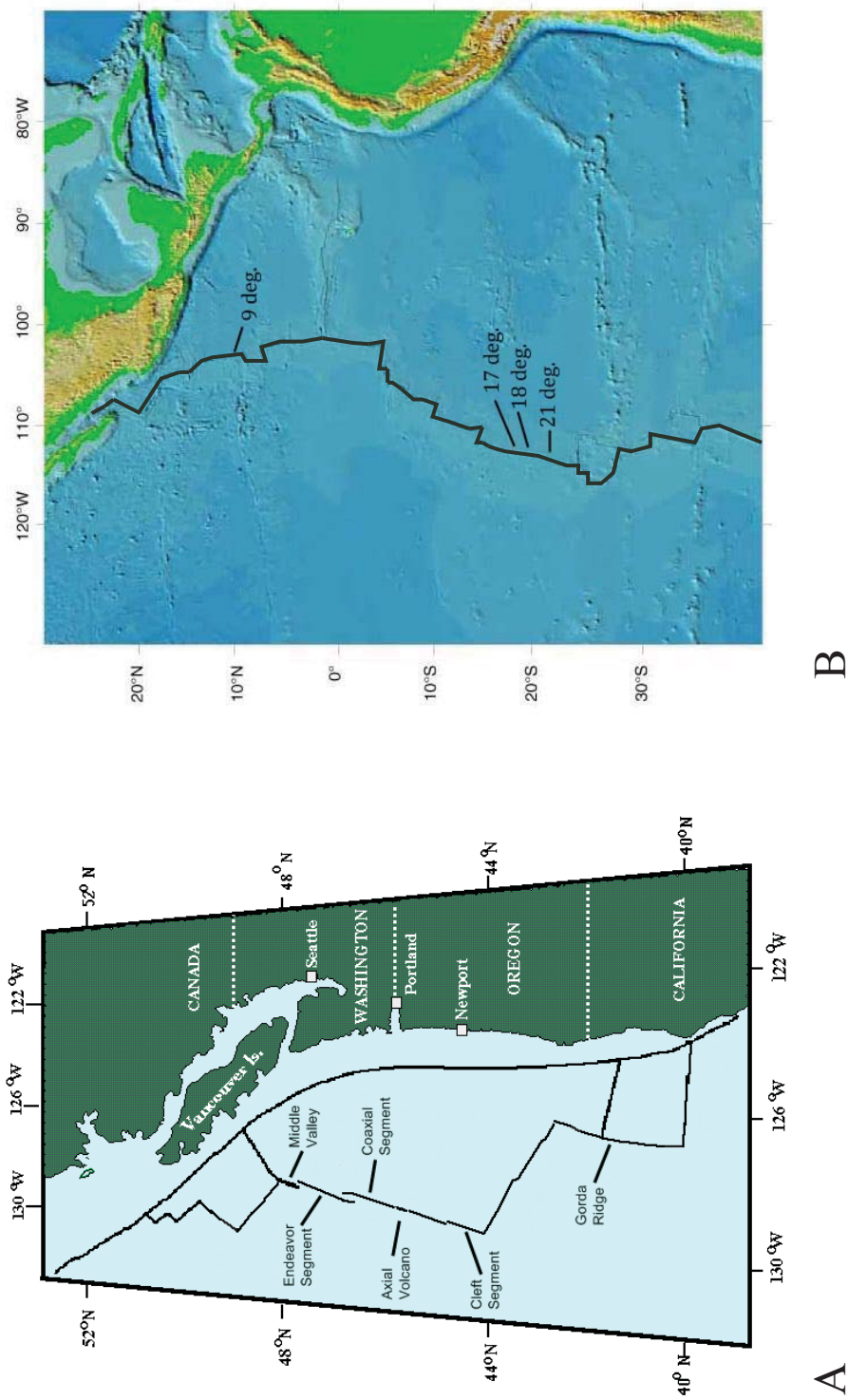


FIG. 1. Location of vents within the two main sample sites. A) Juan de Fuca Ridge and associated vent segments with the Gorda Ridge to the south. B) East Pacific Rise with the vents 9 deg. North and 17, 18, and 21 deg. South. Sample sites within these two regions are at comparable distances from one another and provide nested sampling within regions.

FIG. 2. Cluster analysis of AFLP data through the Pearson product-moment correlation coefficient and UPGMA methods. Isolates and type strains are clustered by percent similarity with the general pattern of isolates clustered into regionally related clusters consisting of isolates from varying sample types and sample sites.

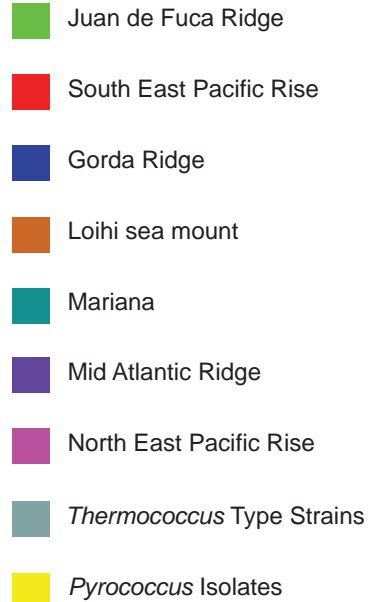
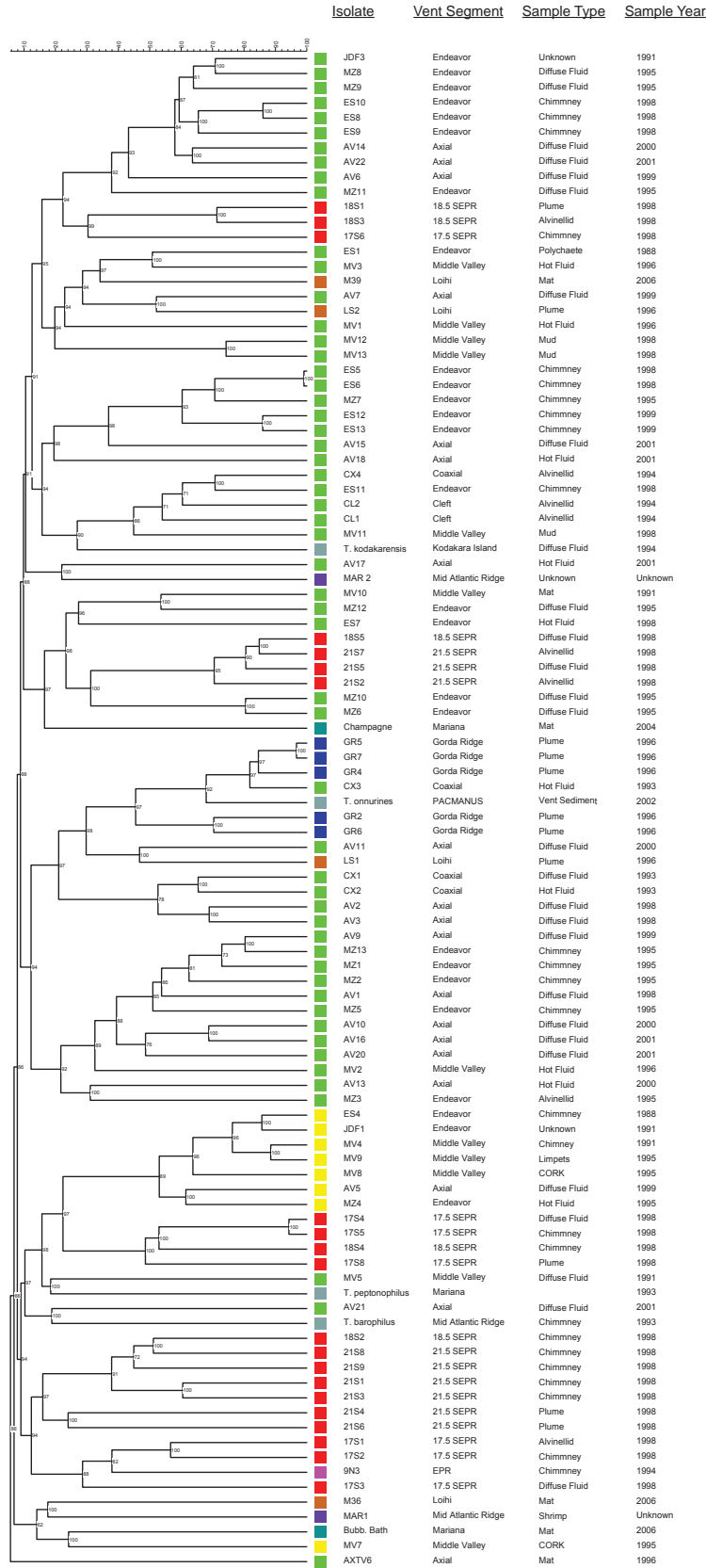
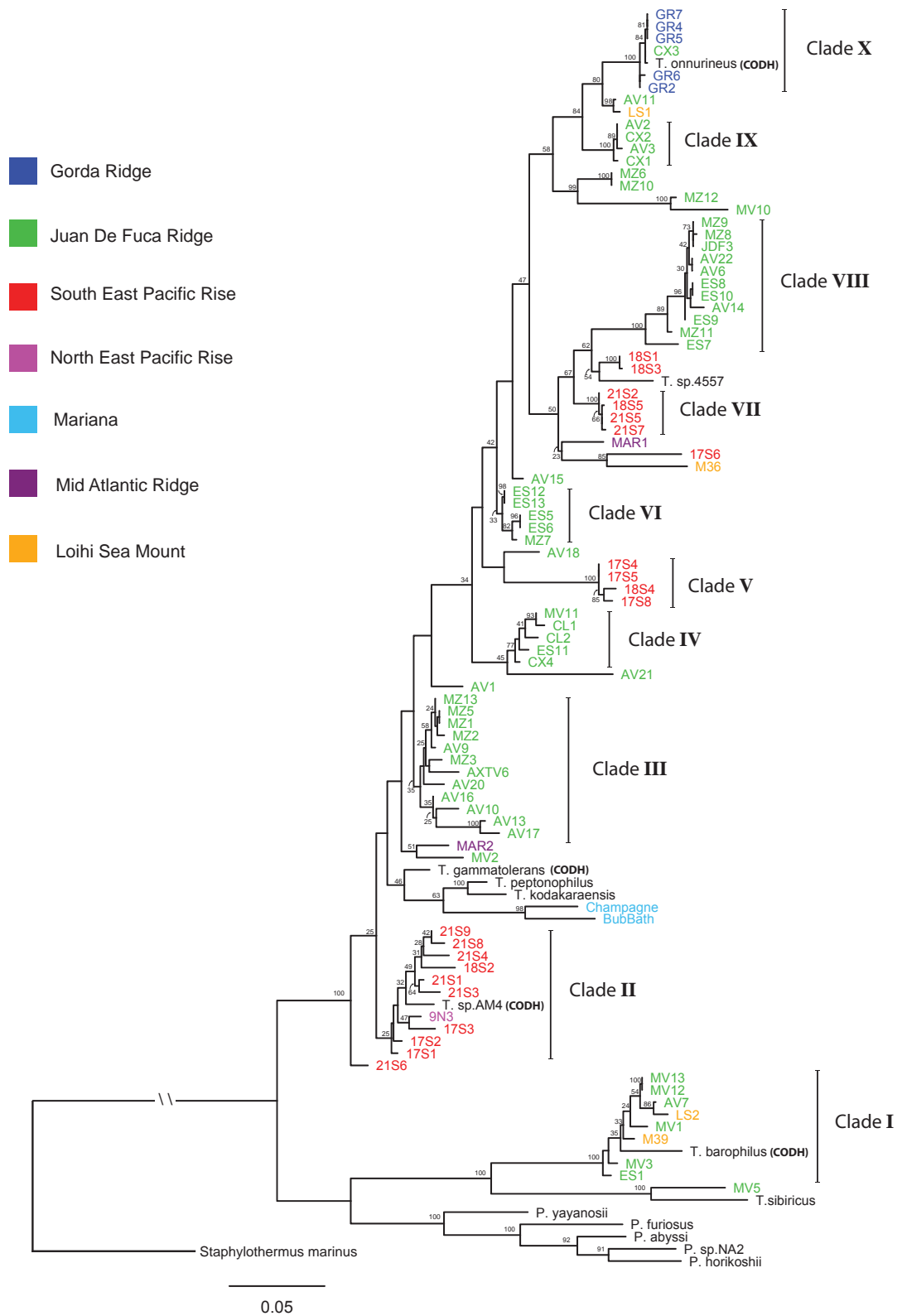


FIG. 3. Maximum likelihood (ML) phylogenetic tree of the concatenated amino acid and SSU rRNA sequences constructed using RAxML with a mixed/partitioned model and per gene optimization of branch lengths with 100 boot straps. Boot strap values of twenty and above are reported. The Crenarchaeota *Staphylothermus marinus* was used as an outgroup to root tree using homologs for six of the seven loci. Genomes containing carbon monoxide dehydrogenase genes are labeled CODH.



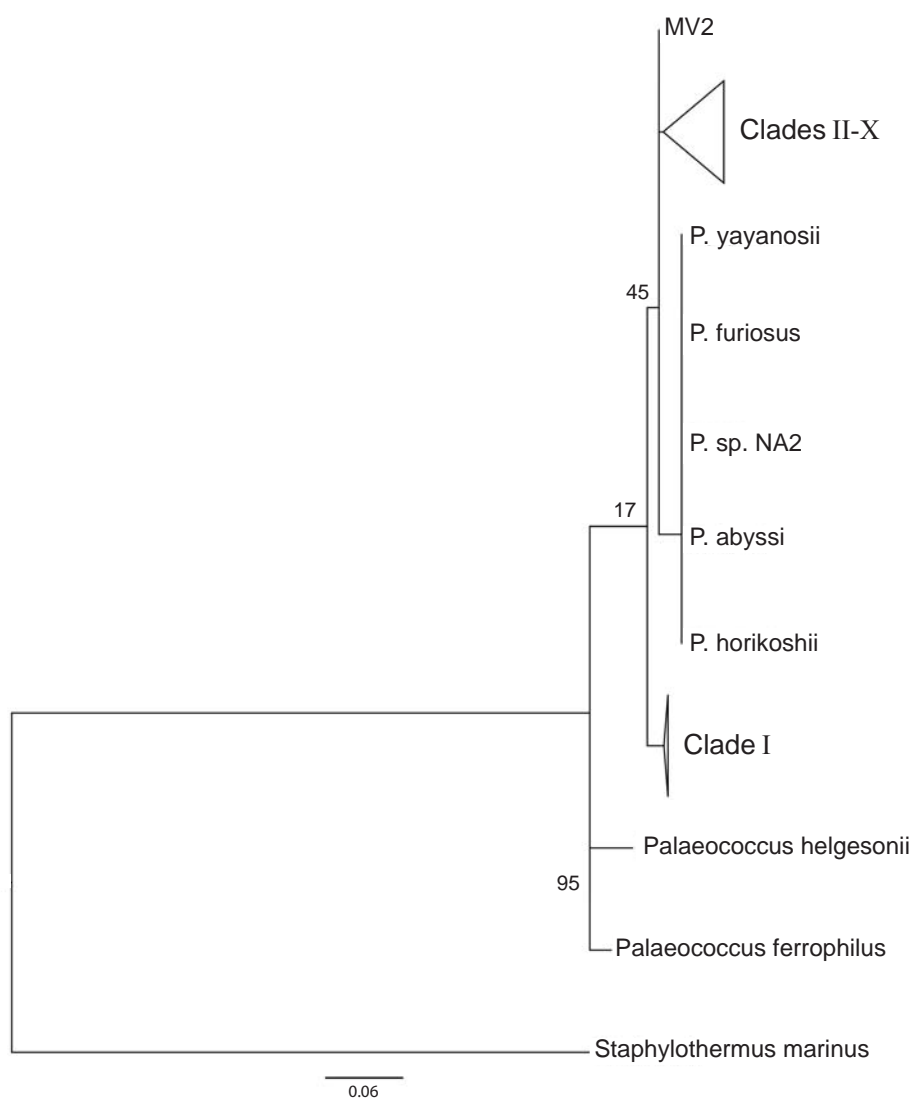


FIG. 4. Maximum likelihood phylogenetic tree of the SSU rRNA with the inclusion of *Palaeococcus* species and rooted with *Staphylothermus marinus*. Clade I is placed in an ancestral position to both *Pyrococcus* type strains and *Thermococcus* isolates in clades II-X.

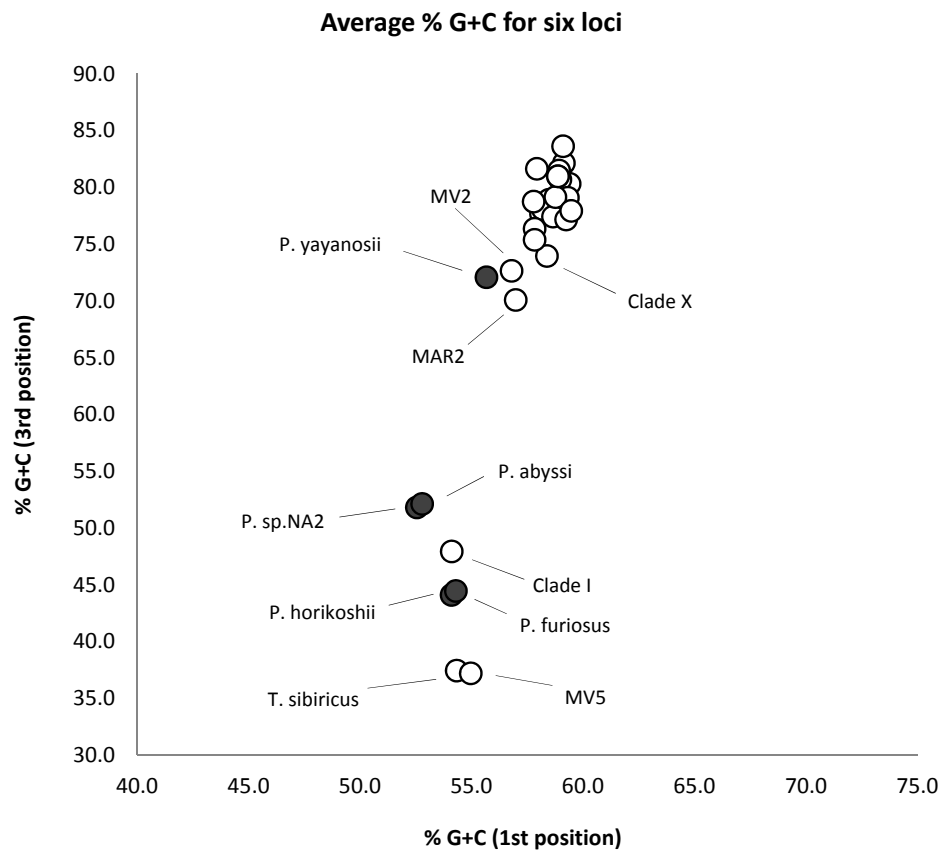


FIG. 5. Average % G+C for first and third codon positions of six protein coding loci. Averages were taken for clades and for individual isolates or type strains when not associated with clades I - X. Closed circles are *Pyrococcus* and open circles are *Thermococcus*.

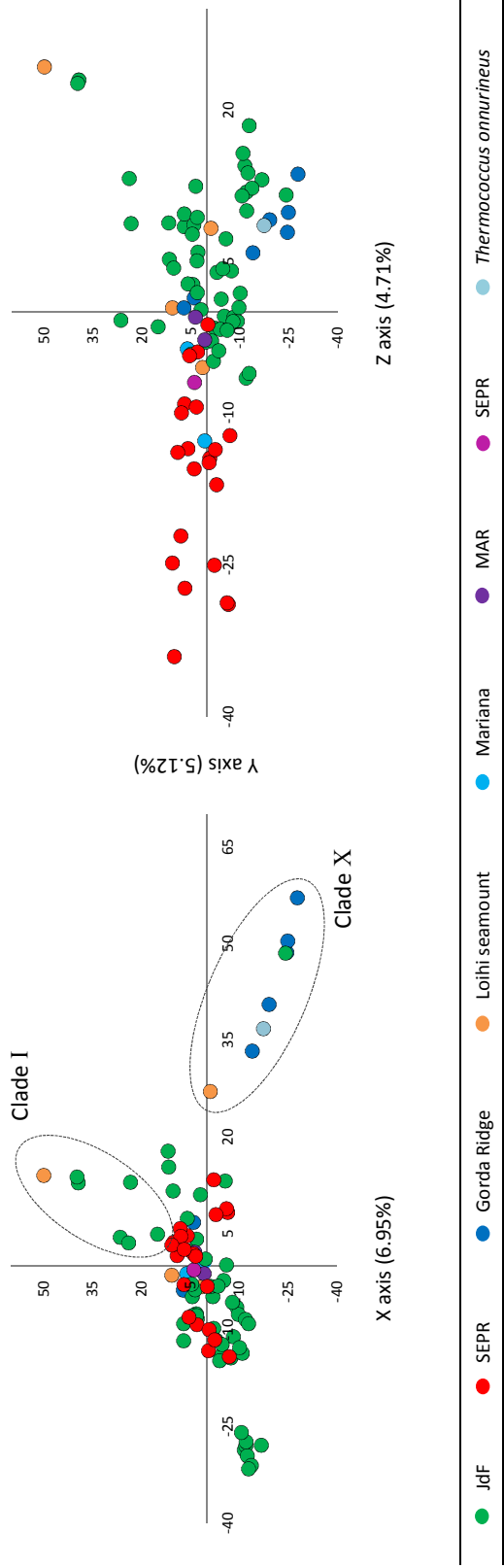


FIG. 6. Principal Component Analysis of AFLP band calling data with the first three principal components describing the greatest variation plotted on the X, Y and Z axes. Principal components 1 and 2 plotted on the X and Y axes differentiate isolates in Clades I and Clade X from other isolates in this study. Principal component 3 plotted on the Z axis differentiates isolates into larger regionally related groups.

TABLE 1. Isolates analyzed through AFLP and MLST and their corresponding vent segments.

<i>Thermococcus</i> Isolates analyzed by AFLP and MLST (n = 90)					
Juan de Fuca Ridge	<u>55</u>	East Pacific Rise	<u>22</u>	Gorda Ridge	5
Middle Valley	8	9 deg. North	1	Loihi Seamount	4
Endeavour Segment	23	17.5 deg. South	7	Mid Atlantic Ridge	2
Coaxial Segment	4	18.5 deg. South	5	Mariana	2
Axial Volcano	18	21.5 deg. South	9		
Cleft Segment	2				

TABLE 2. Clade analysis for Average Nucleotide Identity (ANI), %GC and test for linkage between loci with the Index of association (I_A) and probability value reported.

Clade	ANI	% GC	I_A	p value
I	91%	45.50	0.3667	0.001
II	94%	57.65	0.0189	0.394
III	93%	57.97	0.4845	0.001
IV	98%	59.05	-0.0352	0.605
V	97%	57.43	0.6179	0.001
VI	99%	58.49	0.5873	0.001
VII	99%	59.03	0.3644	0.001
VIII	95%	59.64	0.0328	0.248
IX	99%	57.03	0.1061	0.251
X	99%	55.93	0.6356	0.001

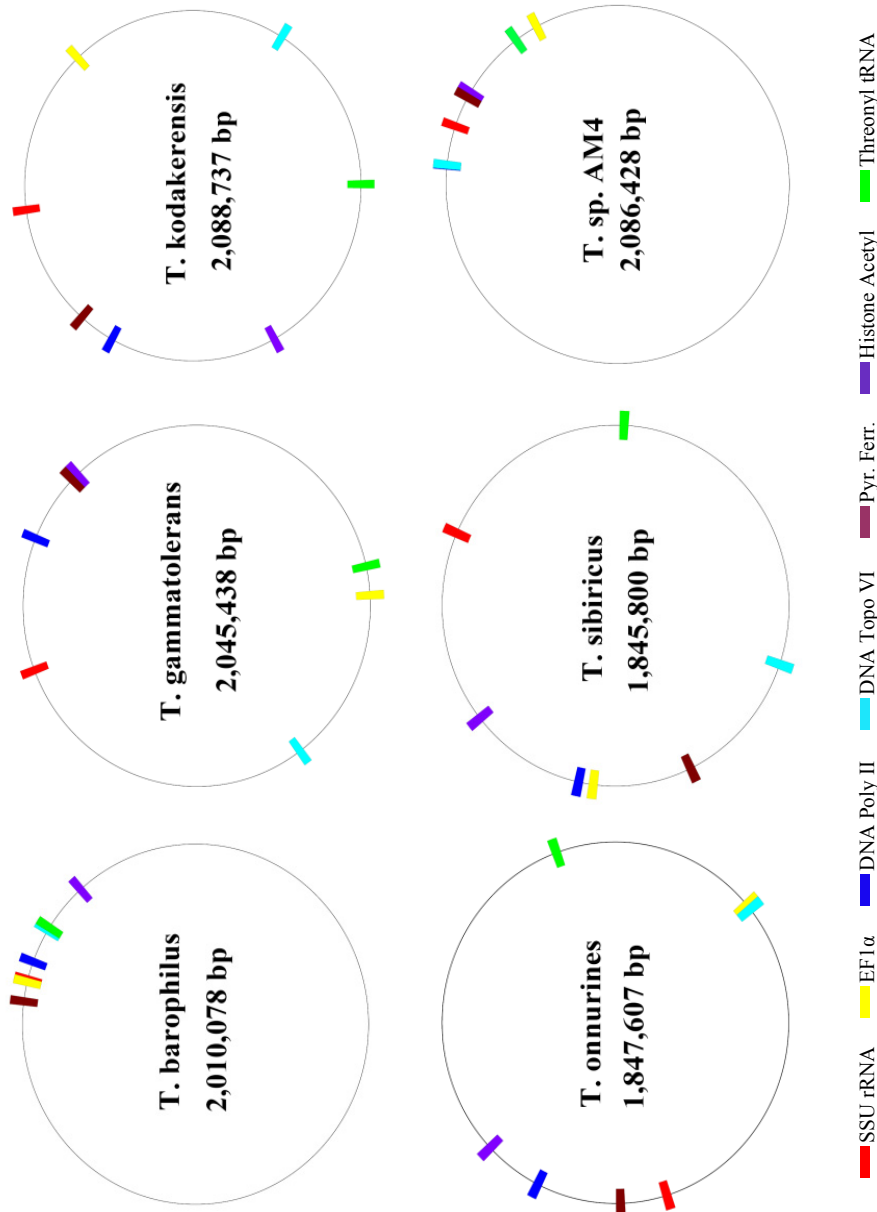


FIG. S1. Gene loci maps for the six *Thermococcus* reference genomes used to pick MLST loci and design primers. Loci maps illustrate the genomic rearrangements and reduced gene synteny between *Thermococcus* species.

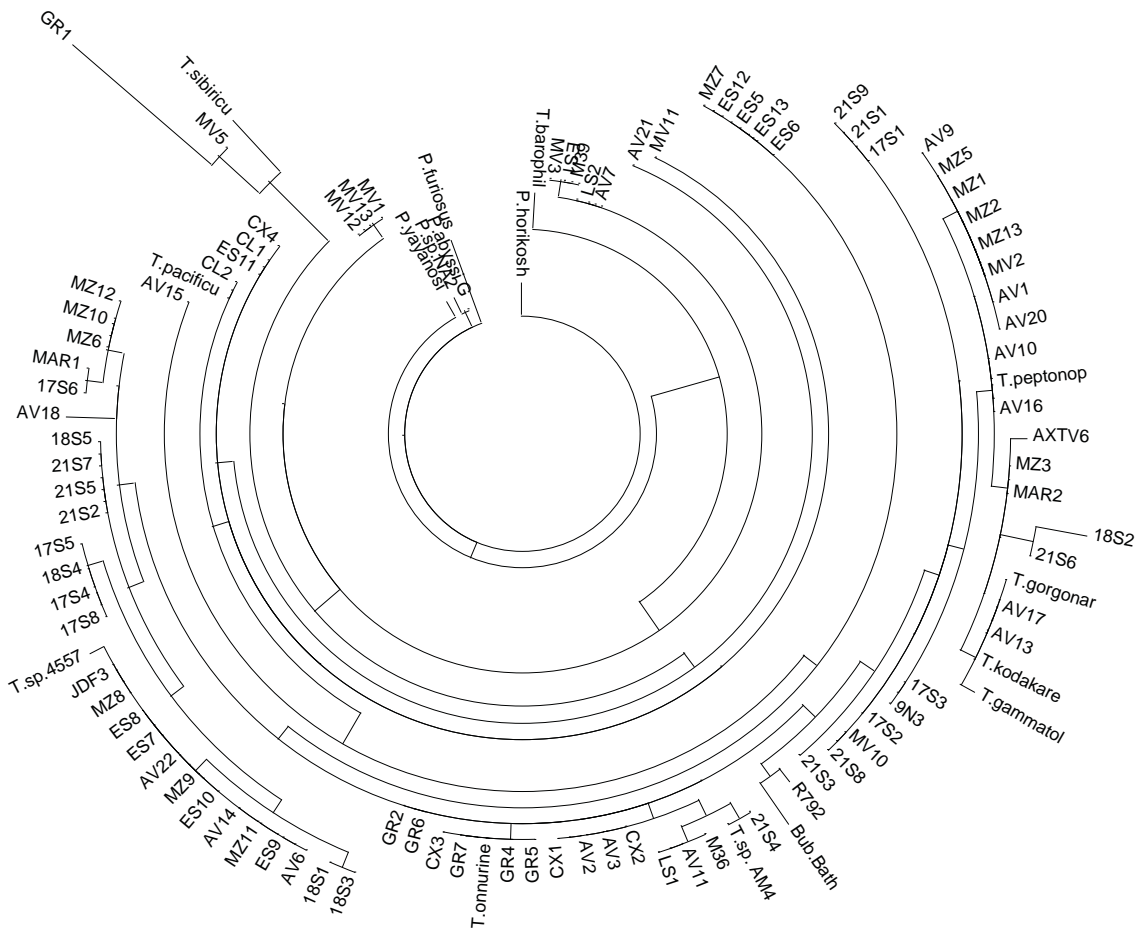


FIG. S2. Unrooted ML tree for elongation factor 1 alpha subunit

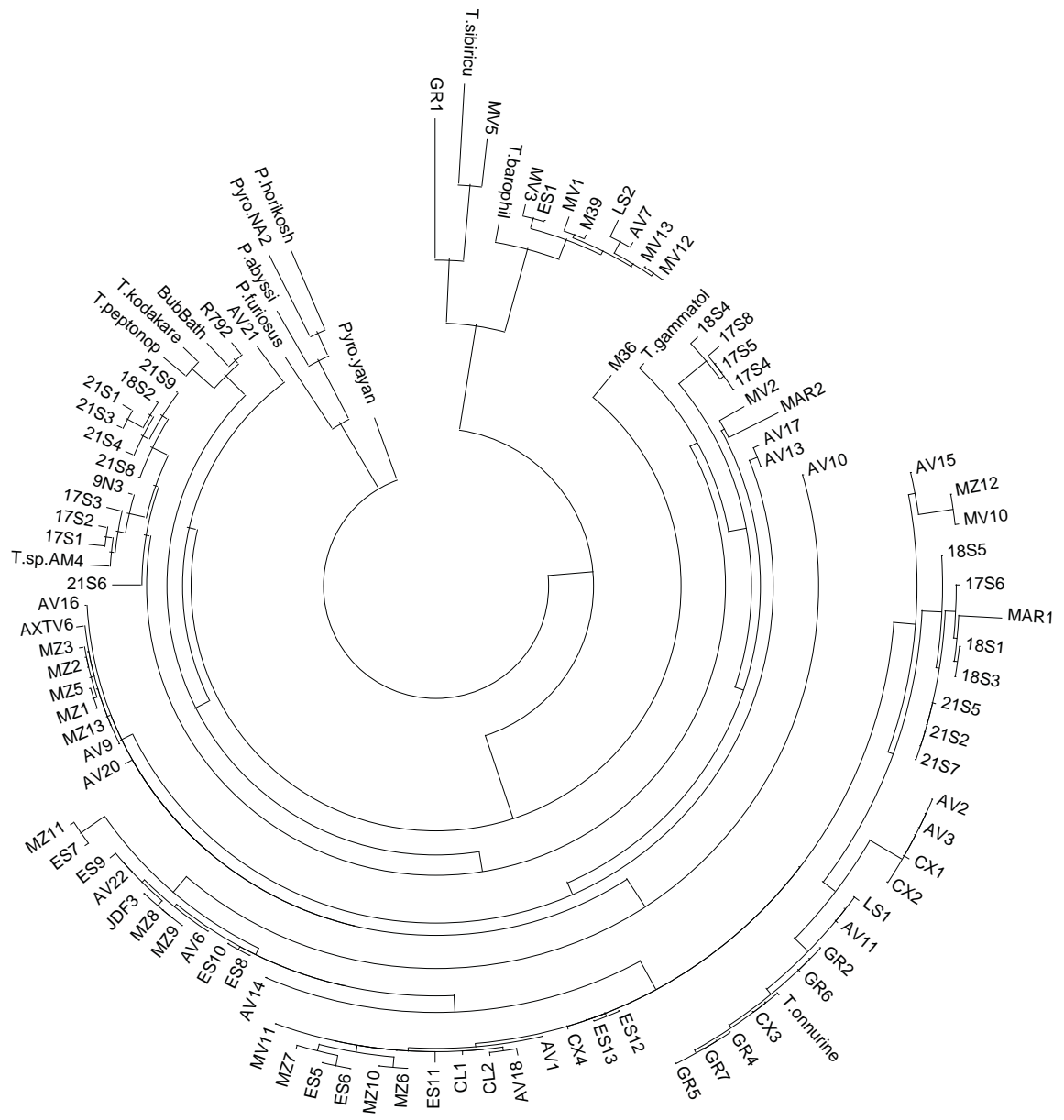


FIG. S3 Unrooted ML tree for DNA polymerase II large subunit

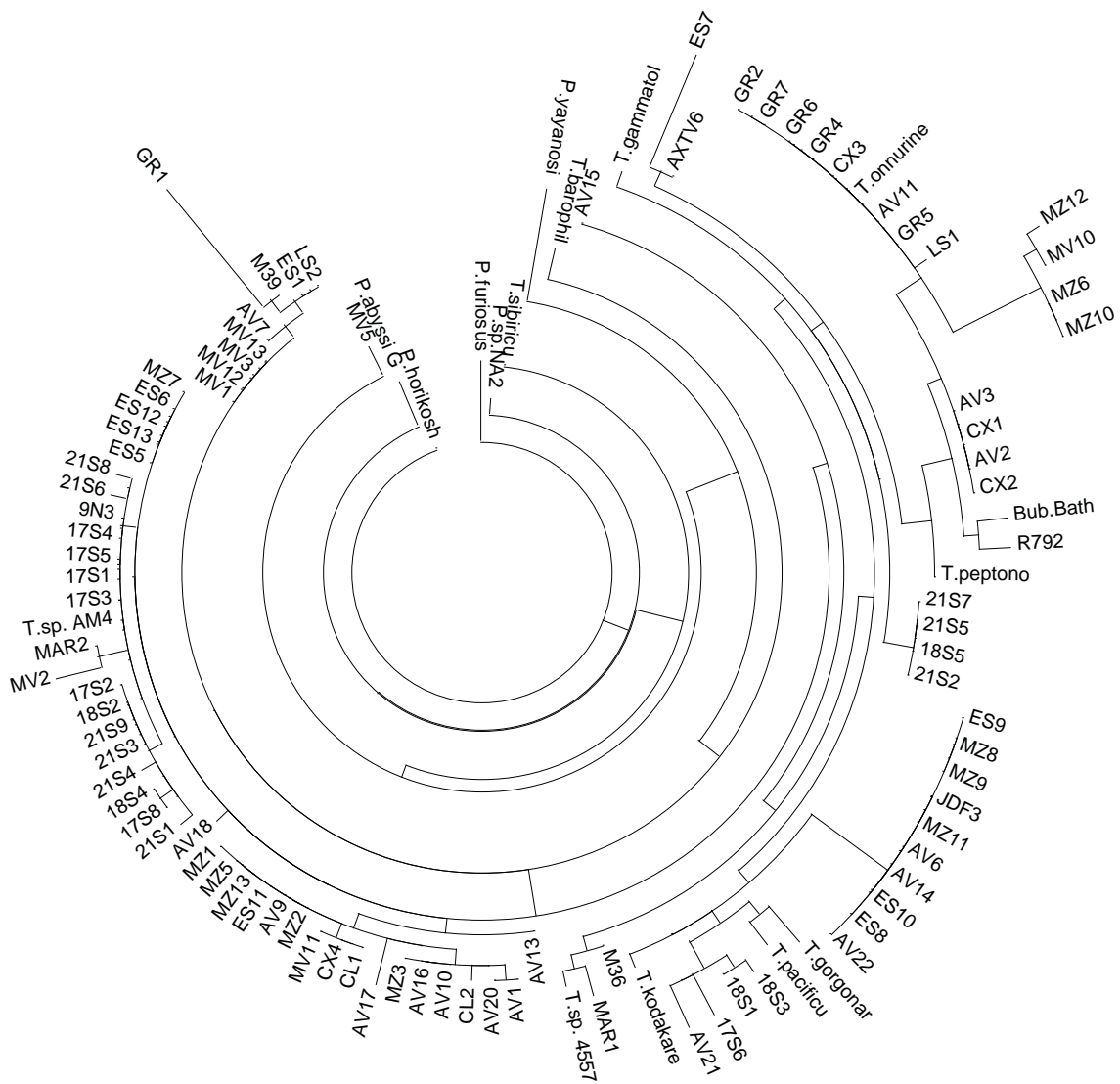


FIG S6. Unrooted ML tree for histone acetyltransferase

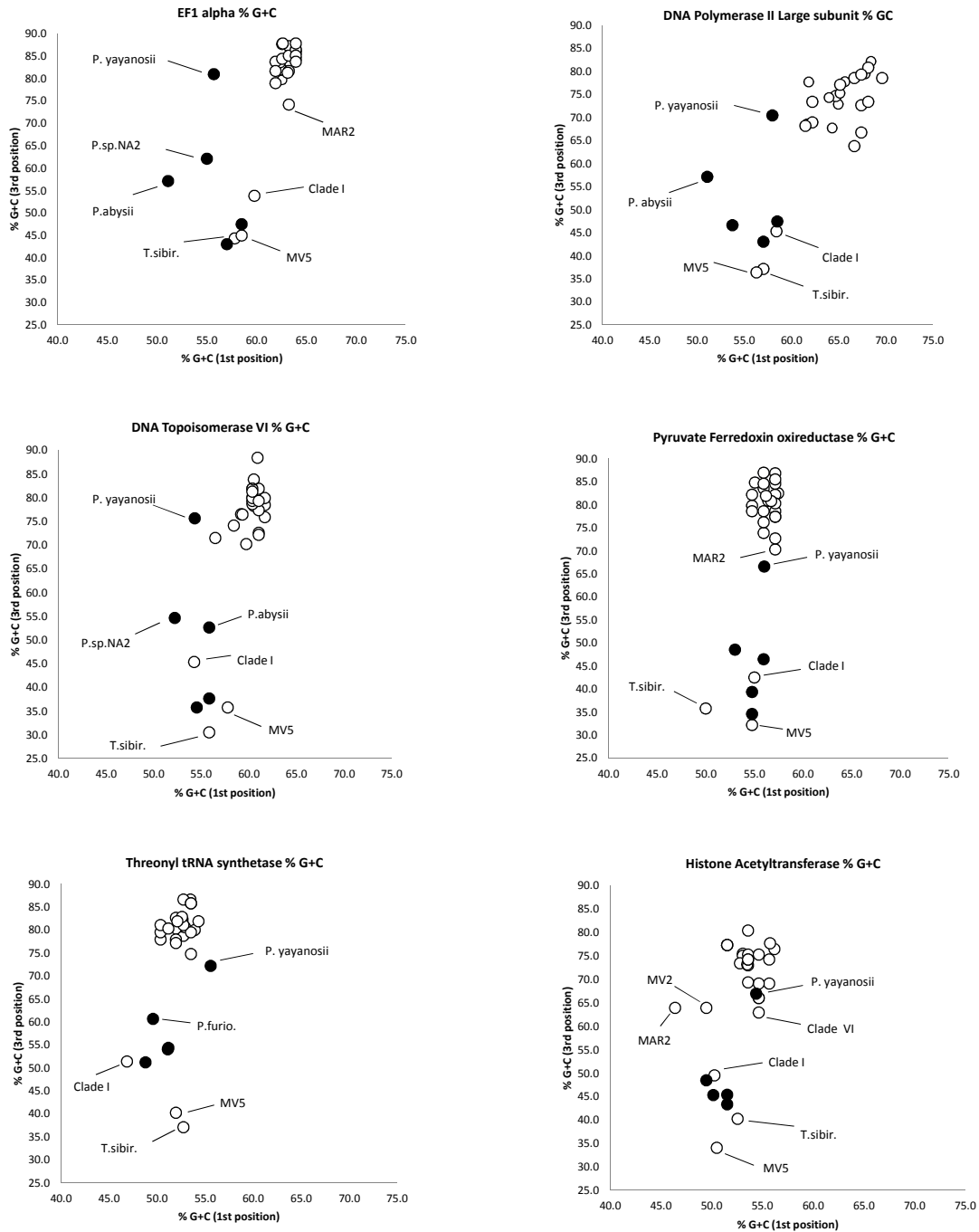


FIG. S8. Average % G+C for first and third codon positions of individual protein coding loci. Averages were taken for clades and for individual isolates or type strains when not associated with clades I - X. Closed circles are *Pyrococcus* and open circles are *Thermococcus*.

TABLE S1. *Thermococcus* type strains analyzed by AFLP and MLST.

Type Strains (n = 8)	AFLP	MLST
<i>Thermococcus kodakerensis</i>	x	x
<i>Thermococcus peptonophilus</i>	x	x
<i>Thermococcus onnurines</i>	x	x
<i>Thermococcus barophilus</i>	x	x
<i>Thermococcus gammatolerans</i>		*
<i>Thermococcus sibiricus</i>		*
<i>Thermococcus</i> sp. AM4		*
<i>Thermococcus</i> sp. 4557		*

* sequence data retrieved from GenBank

TABLE S2. MLST primer sequences

Sequence Name	Bases	Sequence	Gc Cont.	Tm
DNA polymerase II large subunit Forward	20	GAR GAR TGG TGG RTT CAG GA	52.50	55.31
DNA polymerase II large subunit Reverse	20	ATC CAR CTT ATH CCC CKR TC	49.17	53.39
Pyruvate ferredoxin oxidoreductase beta subunit Forward	20	GAR GAC AAG CCV AAG AAG TG	50.83	54.02
Pyruvate ferredoxin oxidoreductase beta subunit Reverse	22	CTT GAA DAG GTG CTT RAA YCT K	40.15	52.17
Threonyl tRNA synthetase Forward	20	GCC MGA TAT GCA YAC SGT HG	56.67	56.74
Threonyl tRNA synthetase Reverse	20	TGW ATB GGR CTG AGC CAK AG	53.33	55.83
DNA topoisomerase VI alpha subunit Forward	21	AAG CMT AYT AYG CSA ACA AGC	45.24	54.63
DNA topoisomerase VI alpha subunit Reverse	20	ATR TCR TCC ATN GTC ATK CC	45.00	52.23
Histone acetyltransferase Forward	21	ACT GGG TTM GSG THA TGA GRA	49.21	56.48
Histone acetyltransferase Reverse	19	CVC KRT GCT GCC ACT CRT A	58.77	57.88
Elongation factor 1 alpha subunit Forward	23	MGT YAA GAA GAG CGA CAA GAT GC	47.80	56.70
Elongation factor 1 alpha subunit Reverse	21	CAC CGG TCT TGA TGA AYT GYG	52.30	56.30
SSU rRNA Forward	20	GGG GTC CGA CTA AGC CAT GC	65.00	60.60
SSU rRNA Reverse	21	GAT TTC GCC AGG GAC TTA CCG	57.10	58.00

TABLE S3. dN/dS ratios calculated for MLST protein coding loci as a measure of selection. Loci dN/dS values reflect the conservation of these genes under purifying selection.

Gene Locus	dN/dS
DNA Polymerase II	0.1106
DNA Topoisomerase VI	0.0199
Elongation Factor 1 α	0.0623
Histone Acetyltransferase	0.0483
Pyruvate Ferredoxin Reductase	0.0464
Threonyl tRNA Synthetase	0.0371

TABLE. S4. Mantel Test for selected isolates. Phylogenetically related clades from different regions show a significant correlation between genetic distance and geographic distance with the exception of Clades IX and X.

Mantel Test	r²	p ≤ 0.05
All Isolates	0.016	0.0229
Juan De Fuca Isolates	0.007	0.0425
South East Pacific Rise Isolates	0.006	0.1186
Clade I	0.312	0.0511
Clades II & III	0.719	0.0001
Clades IV, V & VI	0.603	0.0003
Clades VII & VIII	0.699	0.0001
Clades IX & X	0.008	0.3579