



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

Mark T. Price, Heather Fullerton and Craig L. Moyer\*

Department of Biology, Western Washington University, Bellingham, WA, USA

## OPEN ACCESS

### Edited by:

Beth Orcutt,  
Bigelow Laboratory for Ocean  
Sciences, USA

### Reviewed by:

Julle L. Meyer,  
University of Florida, USA  
Sean Patrick Jungbluth,  
University of Hawaii, USA

### \*Correspondence:

Craig L. Moyer,  
Department of Biology, Western  
Washington University, 516 High  
Street, MS #9160, Bellingham,  
WA 98225, USA  
cmoyer@hydro.biol.wvu.edu

### Specialty section:

This article was submitted to  
Extreme Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 27 May 2015

Accepted: 01 September 2015

Published: 24 September 2015

### Citation:

Price MT, Fullerton H and Moyer CL  
(2015) Biogeography and evolution of  
*Thermococcus* isolates from  
hydrothermal vent systems of the  
Pacific. *Front. Microbiol.* 6:968.  
doi: 10.3389/fmicb.2015.00968

*Thermococcus* is a genus of hyperthermophilic archaea that is ubiquitous in marine hydrothermal environments growing in anaerobic subsurface habitats but able to survive in cold oxygenated seawater. DNA analyses of *Thermococcus* isolates were applied to determine the relationship between geographic distribution and relatedness focusing primarily on isolates from the Juan de Fuca Ridge and South East Pacific Rise. Amplified fragment length polymorphism (AFLP) analysis and multilocus sequence typing (MLST) were used to resolve genomic differences in 90 isolates of *Thermococcus*, making biogeographic patterns and evolutionary relationships apparent. Isolates were differentiated into regionally endemic populations; however, there was also evidence in some lineages of cosmopolitan distribution. The biodiversity identified in *Thermococcus* isolates and presence of distinct lineages within the same vent site suggests the utilization of varying ecological niches in this genus. In addition to resolving biogeographic patterns in *Thermococcus*, this study has raised new questions about the closely related *Pyrococcus* genus. The phylogenetic placement of *Pyrococcus* type strains shows the close relationship between *Thermococcus* and *Pyrococcus* and the unresolved divergence of these two genera.

**Keywords:** *Thermococcus*, hydrothermal vents, subsurface microbiology, biogeography, evolution

## Introduction

Microorganisms constitute the majority of known life forms making up greater than two-thirds of the metabolic and genetic diversity of the planet (Whitaker, 2006). The evolutionary forces responsible for shaping this vast amount of microbial diversity have been a topic of debate. Historically, microorganisms were viewed as having panmictic distributions, being constrained only by environmental conditions (Baas-Becking, 1934). Contemporary research has shown that limited dispersal, geographic isolation, selection and genetic drift can lead to divergence in microbial populations (Cho and Tiedje, 2000). The significance of physical isolation on the population divergence of terrestrial hyperthermophiles has been well documented (Papke et al., 2003; Whitaker et al., 2003; Papke and Ward, 2004; Whitaker, 2006; Reno et al., 2009); however, it is unknown whether similar patterns of isolation can be observed in hyperthermophiles at marine hydrothermal vents or if there is greater dispersal in these marine environments.

Marine hydrothermal vents are unique environments with extreme chemical, nutrient and temperature gradients. The organisms inhabiting hydrothermal vents are dispersed along plate

boundaries and can be separated from other vent systems by thousands of kilometers creating island-like ecosystems. While there has been evidence in support of endemic community structure at hydrothermal vents (Opatkiewicz et al., 2009; Flores et al., 2012), descriptions of biogeography for individual microbial taxa have been uncommon (Escobar-Páramo et al., 2005; DeChaine et al., 2006; McAllister et al., 2011; Mino et al., 2013). The limited descriptions of microbial biogeography at hydrothermal vents can be mitigated through the isolation of microorganisms in culture and sequencing of individual genomes (Ramette and Tiedje, 2007; Meyer and Huber, 2014).

The hyperthermophilic archaea of the *Thermococcales* order are found at hydrothermal vents and can serve as model organisms for the study of biogeography and evolution. *Thermococcales* belong to the euryarchaeota phylum and consist of the three genera *Pyrococcus* (Fiala and Stetter, 1986), *Thermococcus* (Zillig et al., 1983), and *Paleococcus* (Takai et al., 2000). The genera *Thermococcus* and *Pyrococcus* are commonly found at hydrothermal vent systems and are readily isolated in culture allowing for comparisons among individual isolates (Jannasch et al., 1992; Erauso et al., 1993; Atomi et al., 2004; Bae et al., 2006; Teske et al., 2009). Genomic and phenotypic differences between *Thermococcus* and *Pyrococcus* delineate these two closely related genera, e.g., *Pyrococcus* has a higher optimum growth temperature (Garrity et al., 2001). Known *Thermococcus* isolates have a maximum growth temperature of 90°C, but some species have been shown to grow at temperatures as low as 45°C (Adams, 1993; Robb and Place, 1995; Summit and Baross, 1998). However, between these two genera, *Thermococcus* has the highest number of characterized isolates (Garrity et al., 2001; Holden et al., 2001; Teske et al., 2009; Zivanovic et al., 2009).

*Thermococcus* exhibit a variety of metabolic strategies; however, the majority are anaerobic heterotrophs that ferment organic compounds (Robb and Place, 1995; Teske et al., 2009). The addition of elemental sulfur enhances growth in *Thermococcus* and is required in some strains (Teske et al., 2009). Lithotrophic metabolic pathways have been shown as well, e.g., *T. onnurines* and other type strains, can oxidize carbon monoxide to CO<sub>2</sub> using carbon monoxide dehydrogenases or CODHs (Lee et al., 2008; Sokolova et al., 2009; Zivanovic et al., 2009; Oger et al., 2011). Growth via formate oxidation and H<sub>2</sub> production, representing one of the simplest forms of anaerobic respiration, has also been demonstrated in *Thermococcus* (Kim et al., 2010). The ability to use 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 unavailable (Lee et al., 2008; Kim et al., 2010).

Biogeographic patterns have been observed in *Thermococcales* using Multilocus Sequence Typing (MLST) with populations of *Pyrococcus* from different regions shown to be genetically distinct (Escobar-Páramo et al., 2005). However, descriptions of population divergence and biogeography for *Thermococcus* have been inconclusive, therefore *Thermococcus* are described as being widespread and ubiquitous in hydrothermal habitats (Holden et al., 2001; Huber et al., 2006). Although there has been evidence suggestive of a correlation among *Thermococcus* diversity, environmental conditions, and geography (Holden

et al., 2001; Summit and Baross, 2001; Huber et al., 2006; Teske et al., 2009), no strong biogeographic pattern has emerged. Analysis of *Thermococcus* by random amplified polymorphic DNA (RAPD) has identified diverse genomic profiles from the same sample site and similar profiles from different sites, illustrating both the biodiversity present and dispersal potential within a hydrothermal vent field (Lepage et al., 2004). While *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 whether there is population structure at varying geographic scales.

With the exception of RAPD analysis, the genetic markers used in previous *Thermococcus* studies have not had a high degree of 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 using multiple loci spanning the genome to construct robust phylogenetic relationships among related microorganisms (Maiden et al., 1998). By applying MLST analysis to *Thermococcus* isolates their evolutionary relationships and biodiversity can be described more accurately, e.g., MLST analysis has delineated biogeographic patterns in *Sulfolobus* isolates sharing 99.8% sequence similarity for the SSU rRNA gene (Whitaker et al., 2003). MLST data have also provided a basis for defining species level boundaries through the comparison of loci and their average nucleotide identity or ANI (Konstantinidis and Tiedje, 2005). 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 in bacteria (Konstantinidis and Tiedje, 2005; Konstantinidis et al., 2006a,b).

In contrast to MLST analysis, AFLP analysis utilizes genome-wide restriction fragment lengths giving AFLP a higher discriminatory power allowing for the typing of microorganisms to the strain level (Lin et al., 1996; Olive and Bean, 1999; Savelkoul et al., 1999). 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 these large scale genetic changes occurring at higher rates in comparison to nucleotide sequence divergence, AFLP fingerprints have the potential to resolve more recent genomic differentiation (Rademaker et al., 2000). Conserved loci used for MLST analysis are representative of the core genome while the genome-wide restriction fragments used in AFLP analysis may be more representative of variable regions associated with the dispensable or flexible genome, where gene acquisition and loss occurs more frequently (Medini et al., 2005; Tettelin et al., 2008; Cordero and Polz, 2014).

To address questions of biogeography, biodiversity and evolution in the oceanic crust, *Thermococcus* isolates from different hydrothermal regions were analyzed using MLST and AFLP. *Thermococcus* isolates from the Juan de Fuca Ridge (JdF) and East Pacific Rise (EPR) were the primary focus of this study with type strains and isolates from other hydrothermal habitats included for comparison.

## Materials and Methods

### *Thermococcales* Isolates and Type Strains

Sample material was collected from hydrothermal vent sites during research cruises between the years of 1988–2008. Study sites and sampling are as previously described (Holden et al., 2001; Summit and Baross, 2001; Huber et al., 2006; Davis and Moyer, 2008). Both submersible and remotely operated vehicles (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. Sample material was used to inoculate liquid media for the enrichment of *Thermococcales*. Media formulations and methods are as previously described (Holden et al., 2001). Isolates in this culture collection were previously characterized through analysis of the SSU rRNA for genus level associations (Holden et al., 2001; Summit and Baross, 2001; Huber et al., 2006). The collection of 90 *Thermococcus* isolates contains representatives from the Juan de Fuca Ridge (JdF), Gorda Ridge, East Pacific Rise (EPR), Mid Atlantic Ridge (MAR), Mariana Arc and Loihi Seamount (Table 1). Sample sites within the JdF and EPR are at similar spatial distances providing nested sampling within these two regions (Figure 1). Distances among vents within regions range from ~65 to ~450 km, with distances between the main study sites, the JdF and EPR, up to ~7000 km. *Thermococcus* isolates and type strains from other regions, as well as *Pyrococcus* type strains, were included for comparison. 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 in the same manner as 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).

### DNA Extraction

Genomic DNA (gDNA) was extracted from freshly cultured cells. Cultured isolates were initially centrifuged at  $750 \times g$  for 5 min to pellet the sulfur. Supernatant was removed and centrifuged at  $11,000 \times g$  for 10 min in a chilled rotor (4°C). The resulting cell pellet was used for gDNA extraction with the DNeasy Tissue Kit (Qiagen, Valencia, CA) following manufacturer's protocols. A NanoDrop ND-1000 spectrophotometer (Thermo Scientific,

Wilmington, DE) was used to check DNA concentration and purity.

### AFLP Reactions and Cluster Analysis

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 were designed for the selective amplification of restriction fragments of between 50 and 500 base pairs. Two selective primer sets were used in separate reactions to obtain two individual AFLP profiles (EcoRI-0 and MseI-CT; 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 15  $\mu$ l 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 for comparable peak heights among isolates. Electropherogram data were 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 of AFLP electropherogram data was performed using the Pearson product-moment correlation coefficient in BioNumerics (Häne et al., 1993). Cluster analyses of AFLP primer sets were combined and averaged to construct a dendrogram with cophenetic correlation coefficients.

### Multi Locus Sequence Typing

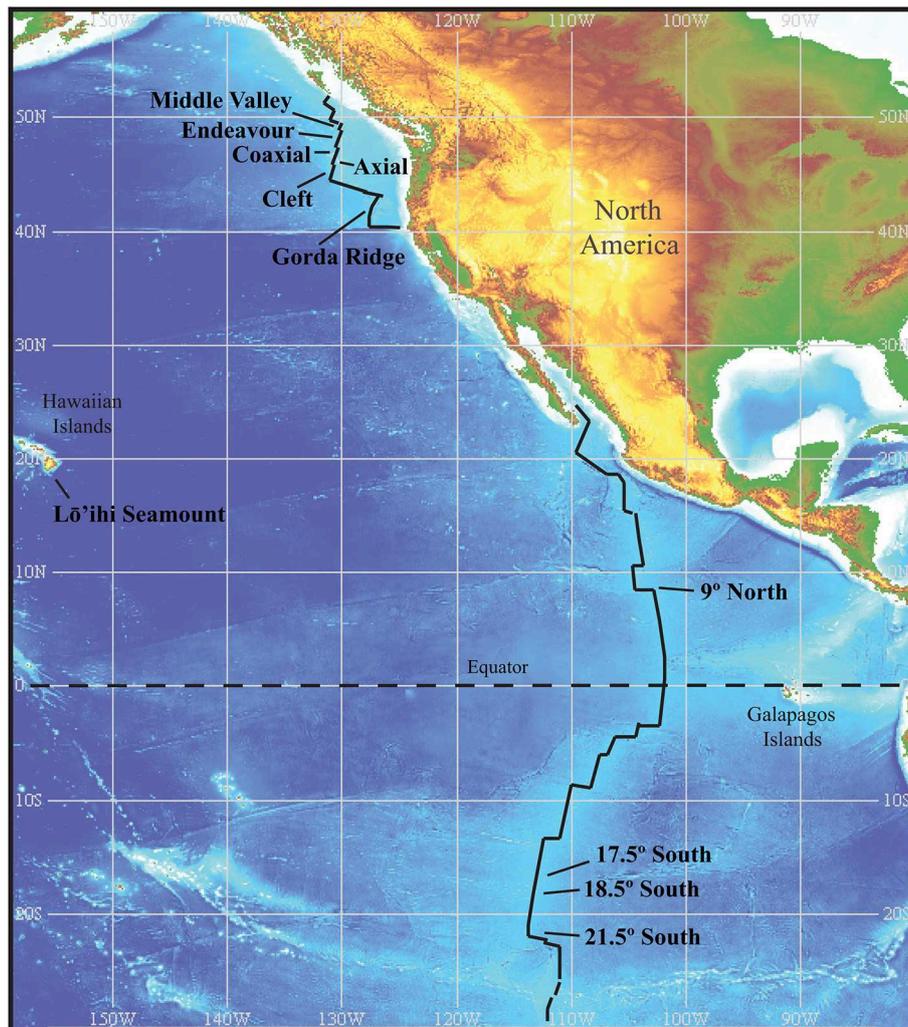
Primers for MLST analysis were designed using nucleotide alignments from the following six annotated genomes: *Thermococcus kodakaraensis* KOD1 (Fukui et al., 2005), *Thermococcus onnurineus* NA1 (Lee et al., 2008), *Thermococcus gammatolerans* EJ3 (Zivanovic et al., 2009), *Thermococcus* sp. AM4 (Oger et al., 2011), *Thermococcus barophilus* MP (Vannier et al., 2011), and *Thermococcus sibiricus* (Mardanov et al., 2009). Gene sequences retrieved through GenBank were aligned with ClustalW multiple sequence aligner using default parameters (Thompson et al., 1994). 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 in the six reference genomes were plotted to check distribution and compare gene synteny (Figure 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. The primer design software Primer3 was used to design primers (Rozen and Skaletsky, 2000). 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 (Table S2).

The following PCR mix was used for the SSU rRNA and EF1 $\alpha$  genes: 50 ng of gDNA template, 0.5  $\mu$ l JumpStart Taq (2.5 U/ $\mu$ l;

**TABLE 1 | *Thermococcus* isolates analyzed through AFLP and MLST and their corresponding vent segments (n = 90).**

Juan de Fuca Ridge	55	East Pacific rise	22	Gorda Ridge	5
Middle valley	8	9° North	1	Lō'ihi Seamount	4
Endeavor segment	23	17.5° South	7	Mid Atlantic Ridge	2
Coaxial segment	4	18.5° South	5	Mariana Arc	2
Axial volcano	18	21.5° South	9		
Cleft segment	2				

Vent regions are in bold, vent segments are included as subsets for the two primary regions under study.



**FIGURE 1 | Location of vents within the two main regions sampled.** The Juan de Fuca Ridge and associated vent segments with the Gorda Ridge to the south, and the East Pacific Rise. Sample sites within these two regions are at comparable distances from one another and provide nested sampling within regions. Image reproduced from the GEBCO world map 2014, www.gebco.net.

Sigma, St. Louis, Mo), 1X Taq PCR buffer, 1.5 mM MgCl<sub>2</sub>, 0.2 mM each deoxynucleoside triphosphate, 0.2 μM each forward and reverse primer, and molecular grade water to a total volume of 50 μl for each reaction. For the SSU rRNA genes 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 s), annealing (58°C for 30 s), and elongation (72°C for 30 s). This was followed by a final elongation step at 72°C for 7 min. For the EF1α genes the PCR cycle was the same as the above with the exception of a 56°C annealing temperature. For the remaining five genes the mixture and conditions were the same as the above with the exception of a 0.8 μM concentration for forward and reverse primers, and 55°C annealing temperature.

PCR amplicons were verified and checked for size through gel electrophoresis. Amplicons were sequenced using ABI BigDye Terminator v3.1, using an ABI Prism 3130XL Genetic Analyzer. Nucleotide sequences were contiguously

assembled using BioNumerics and were verified via BLASTn (Altschul et al., 1990). The concatenated alignment of all seven amplicons following the masking of sequence data and removal of non-homologous positions resulted in 2648 bp of nucleotide sequence for analysis. Partial sequences for all loci were submitted to GenBank and assigned accession numbers (SSU rRNA: KP187908-KP187997, DNA polymerase II large subunit: KP187998-KP188087, DNA topoisomerase VI subunit A: KP188088-KP188177, elongation factor-1 alpha: KP188178-KP188267, histone acetyltransferase: KP188268-KP188357, pyruvate:ferredoxin oxidoreductase subunit beta: KP188358-KP188447, threonyl-tRNA synthetase: KP188448-KP188537).

The ExpASy translate tool (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 v5 (Tamura et al., 2011).

Sequences of individual loci were aligned by ClustalW in MEGA v5 with all gaps removed.

A sequence based test for selection was performed on each protein coding locus using the Datamonkey online server using the following parameters: Data type was codon, genetic code set to universal code, and the method used was SLAC (Pond and Frost, 2005). Non-synonymous versus synonymous (dN/dS) ratios were calculated on the in-frame nucleotide alignments of individual loci to determine if loci were under strong selection.

### Phylogenetic Analysis

Maximum likelihood (ML) phylogenetic trees were constructed using the RAxML BlackBox online server (Stamatakis et al., 2008). A mixed/partition model was applied to the concatenated alignment of nucleotide (SSU rRNA) using the CAT model (Stamatakis, 2006; Stamatakis et al., 2008) and amino acid sequences, using the WAG (Whelan and Goldman, 2001) protein model, with per gene optimization of branch lengths and 100 bootstraps. The tree with the lowest log likelihood score was selected from 10 replicates. The concatenated ML tree was rooted using the *Crenarchaea*, *Staphylothermus marinus* as an outgroup. Homologs in *Staphylothermus marinus* for six of the seven loci (no suitable homolog was found for DNA polymerase II) were used for phylogenetic analysis and rooting of the tree using the parameters described above. Clades were assigned to groupings of three or more isolates. Bootstrap values of 20 and above were reported. The ML tree for the SSU rRNA locus, with the inclusion of *Palaeococcus ferriphilus* and *Palaeococcus helgesonii* sequences (Takai et al., 2000; Amend et al., 2003), was also rooted using *Staphylothermus marinus* as an outgroup. For both trees, the choice of *Staphylothermus marinus* as an outgroup was made as it required minimal masking relative to the several other *Crenarchaeota* tested without changing position of the root.

### MLST Clade Analysis

The 2648 bp concatenated nucleotide sequence was used for analysis of the Average Nucleotide Identity (ANI) of clades, as an estimate of species level boundaries (Konstantinidis and Tiedje, 2005; Konstantinidis et al., 2006a,b). 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.

The six protein-coding loci were used for the analysis of GC-content of clades, an estimate of evolutionary history, using the Datamonkey online server as described above (Pond and Frost, 2005).

Linkage analysis was performed to test for evidence of recombination in MLST loci. In clonal organisms linkage among loci (linkage disequilibria) is expected with evidence of unlinked loci (linkage equilibria) associated with gene transfer or recombination events. Linkage among loci was tested using the Non-redundant database (NRDB; PubMLST) and Linkage Analysis (LIAN) v3.5 set to default parameters, e.g., Monte Carlo with 100 iterations. (Haubold and Hudson, 2000). Isolate loci were coded through NRDB and the null hypothesis of linkage equilibrium for clades was tested using LIAN v3.5 with the standardized index of association ( $I_A$ ) reported.

### Codon GC-content

Ratios of the GC-content for first and third codon positions were analyzed as a measure of gene history (Muto and Osawa, 1987). A two dimensional plot of codon ratios was constructed from the average GC-content of all six protein coding loci. The first and third codon positions of loci were calculated for individual clades and individual isolates and type strains when not associated with Clades I-X using MEGA v5 (Tamura et al., 2011). The first and third codon positions were plotted in a similar manner previously shown to differentiate genomic variation in bacteria (Kaplan and Fine, 1998).

### Mantel Test

A Mantel test comparing geographic distance and genetic distance was performed using the statistical software zt (Bonnet and de Peer, 2002). Pairwise genetic distances for nucleotide sequence data were calculated using the Maximum Composite Likelihood model in MEGA v5 (Tamura et al., 2011). Geographic distances among vents were 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 closely associated phylogenetic clades.

### Analysis of Molecular Variance

Analysis of Molecular Variance (AMOVA) was calculated in Arlequin version 3.11 (Excoffier et al., 2005) 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 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.

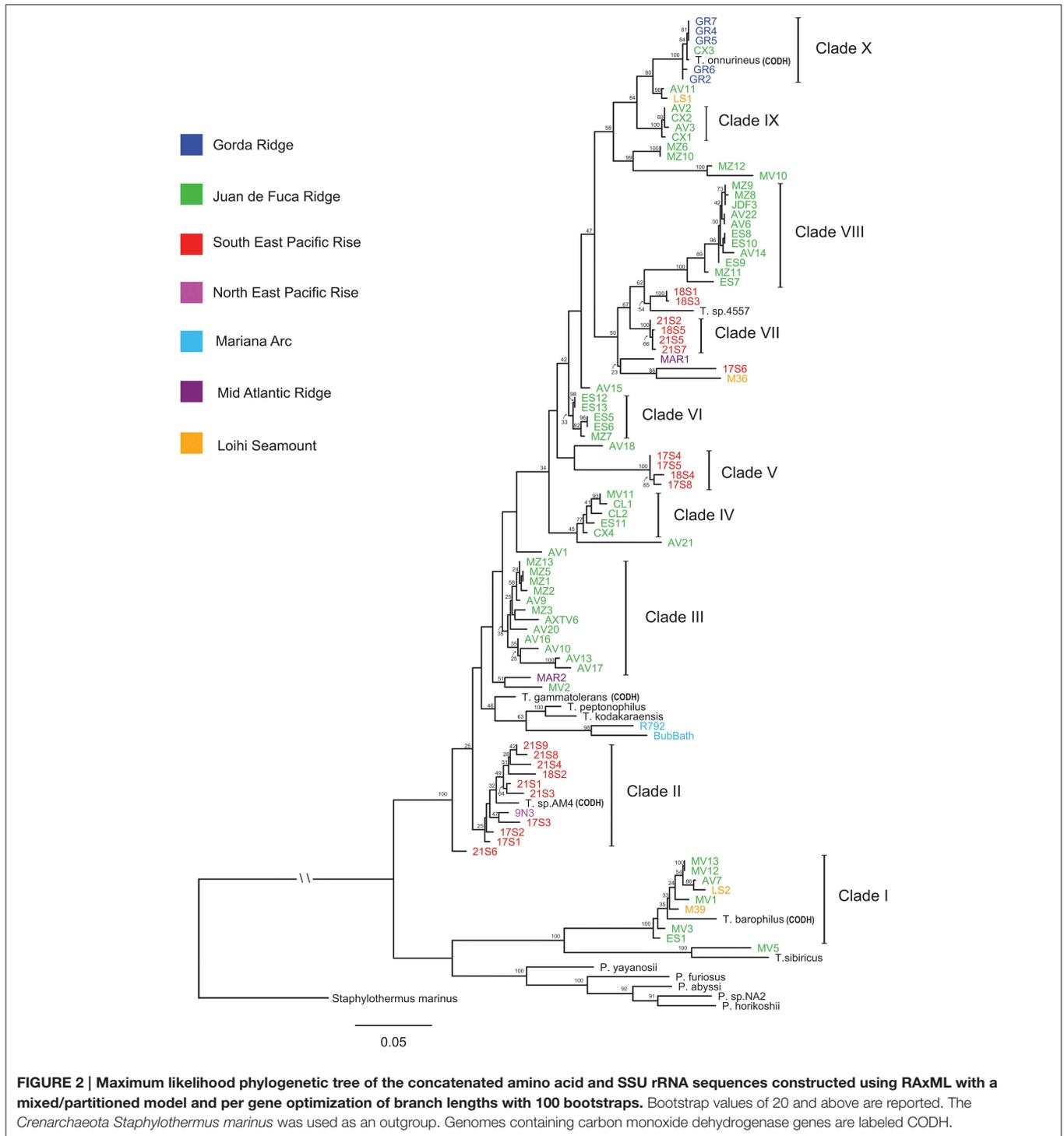
### Principal Component Analysis

Principal Component Analysis (PCA) was performed on AFLP band calling data using BioNumerics v4.61. 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, with the optimization and position tolerances for selecting bands set to 0.10% for both primer sets. 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 from both primer sets were combined and converted into a binary presence absence matrix. Default settings were applied for PCA in BioNumerics, subtracting the average for characters.

## Results

### AFLP Cluster Analysis

Through cluster analysis, isolate AFLP profiles are grouped by similarity. The cluster analysis dendrogram topology is well supported by cophenetic correlation coefficients (Figure S2). Isolates from the JdF and EPR are differentiated at regional levels



with no clustering of isolates between these two regions. Isolate diversity is exemplified by the number of clusters found within the JdF and EPR regions with up to five clusters identified for each. These clusters are made up of isolates from more than one vent site within a region, more than one sample type and for JdF isolates, from different sampling years. Isolate groupings 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* clustered together with the exception of isolate MV7. Type strains and isolates from regions other than the JdF or EPR had low AFLP profile similarity with isolates in this study, with the exception of the type strain *Thermococcus onnurines*. This type strain (*T. onnurines*) was isolated from the Papua New

Guinea-Australia-Canada-Manus (PACMANUS) field (Bae et al., 2006) and had high similarity with the cluster containing Gorda Ridge isolates and isolate CX3 from the coaxial segment, with  $\geq 68\%$  AFLP profile similarity among these isolates.

### MLST and Phylogenetic Analysis

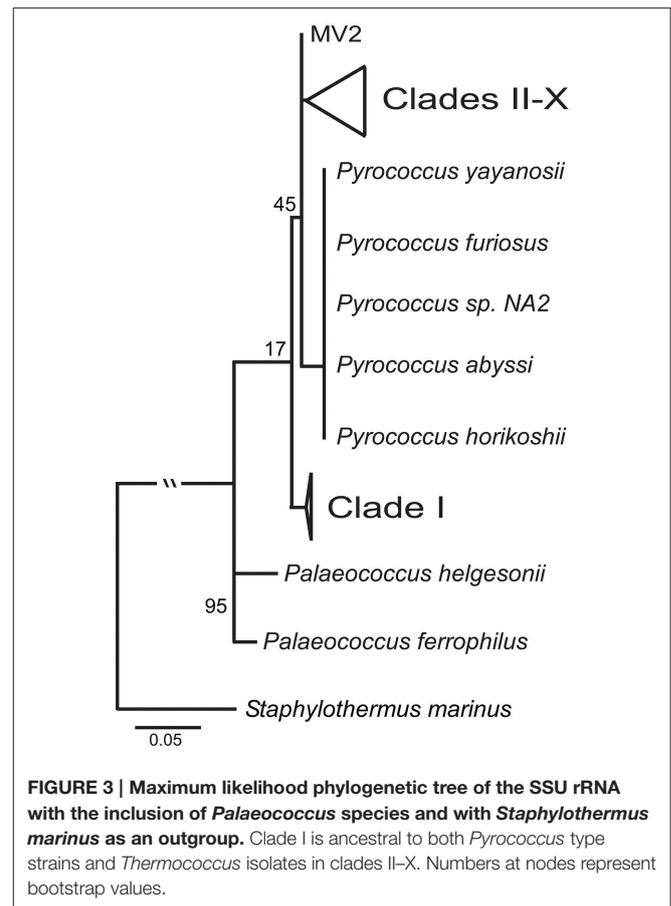
Analysis of the individual alignments for the six protein coding loci did not show significant evidence for any of the loci undergoing strong selection (Table S3), making these loci suitable for MLST analysis. Ratios of dN/dS were consistent with dN/dS ratios reported for conserved genes under purifying selection (Kuhn et al., 2006).

The ML phylogeny of concatenated MLST loci differentiated isolates into clades that are in general agreement with isolate groupings through AFLP cluster analysis (Figure 2 and Figure S2). Isolates from the JdF and EPR are differentiated into lineages that are phylogenetically related across these two regions (Figure 2). The regional divergence among phylogenetically related isolates is most apparent in the paraphyly observed between Clades VII and VIII. A general pattern of isolates differentiated by region was observed with individual clades made up of isolates from that same region. Regional groupings were most apparent in the isolates associated with the JdF and EPR; however, regional groupings were also observed in the two isolates from the Mariana Arc, the type strains *T. peptonophilus* and *T. kodakaraensis* both from Japan, and the type strains *T. sp.* 4557 and *T. sp.* AM4 isolated from the EPR (which both group with lineages also from the EPR). Exceptions to regional groupings are seen in Clades I and X, a pattern also observed through AFLP cluster analysis. Clade I contains isolates from the JdF Ridge, Loihi Seamount, 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 sharing 99% sequence identity. *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 through X. The ancestral relationship between *Pyrococcus* type strains and *Thermococcus* isolates in the ML phylogeny of concatenated MLST loci remains unresolved.

The ML tree for the SSU rRNA gene, which includes *Palaeococcus* isolates and is rooted with *Staphylothermus marinus*, places the Clade I *Thermococcus* isolates in an ancestral position to *Pyrococcus* type strains and the *Thermococcus* isolates found in Clades II through X (Figure 3). This illustrates the paraphyletic associations among members of the *Thermococcus* and *Pyrococcus* genera. Several other *Crenarchaeota* were tested as an outgroup including: *Cenarchaeum symbiosum*, *Sulfolobus tokodaii*, *Nitrosopumilus maritimus*, and *Pyrobaculum aerophilum*. None of these other taxa changed the position of the root (data not shown).

### MLST Clade Analysis

Analysis of clade ANI allowed for clade diversity to be estimated based on species level boundaries observed in bacteria (ANI  $\geq$



**FIGURE 3 | Maximum likelihood phylogenetic tree of the SSU rRNA with the inclusion of *Palaeococcus* species and with *Staphylothermus marinus* as an outgroup.** Clade I is ancestral to both *Pyrococcus* type strains and *Thermococcus* isolates in clades II–X. Numbers at nodes represent bootstrap values.

95%). Clades I, II, and III have ANI values below 95% reflecting diversity that is likely beyond the individual species level (Table 2). Clades IV through X have ANI values at or above 95% suggesting that these clades may represent individual species. Pairwise comparisons among all clades (data not shown) result in ANI values  $<95\%$ , suggesting that individual clades represent distinct species.

Variation in GC-content was observed among all clades (Table 2). Clade I at 45.50% had the lowest GC-content in comparison to other *Thermococcus* clades and was closer to GC values reported for *Pyrococcus* genera (Lecompte et al., 2001). Clade X with a GC-content of 55.93% had the next lowest percentage with clades II through IX having GC-content ratios between 57.03 and 59.64%.

Linkage analysis was applied to test for linkage among the MLST loci. Linkage analysis of individual clades rejected the null hypothesis of linkage equilibria, with the measure of linkage ( $I_A$ ) significantly different from zero and/or not significant. Therefore, clade analysis did not show evidence of gene transfer or recombination in MLST loci examined.

### Codon GC-content Ratios

A two dimensional plot of average GC-content at the first and third codon position illustrates differences in codon GC ratios for clades, isolates and type strains from the phylogenetic tree (Figure S3). *Thermococcus* isolates from Clade I along with the type strain *Thermococcus sibiricus* and isolate MV5 have codon

**TABLE 2 | Clade analysis for Average Nucleotide Identity (ANI), GC-content and test for linkage among loci with the Index of association ( $I_A$ ) and probability value reported.**

Clade	ANI%	GC ratio	$I_A$	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

GC ratios that are closer to *Pyrococcus* isolates and type strains, in agreement with our MLST phylogenetic tree (Figure 2). *Pyrococcus yayanosii*, which has been shown to have a higher GC content in comparison to other *Pyrococcus* (Jun et al., 2011), has a codon GC ratio closer to those calculated for the *Thermococcus* isolates found in Clades II through X.

### Mantel Test

Mantel's analysis testing the correlation between genetic distance and geographic distance did not find a significant correlation when looking at all isolates or isolates from either the JdF or EPR regions separately (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 closely associated phylogenetic 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. In particular, Clade X contains isolates from disparate regions which share high sequence similarity.

### Analysis of Molecular Variance

Variation in isolates grouped by sample type (1.1% of variation  $P = 0.324$ ) and by sample site (16.2% of variation  $P = 0.121$ ) showed no significant correlation using an AMOVA of MLST nucleotide data.

### Principal Component Analysis

PCA revealed that the first three principal components described 16.8% of the variation in the AFLP data (Figure 4). 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, particularly the JdF isolates (northern-most) from the EPR isolates (southern-most).

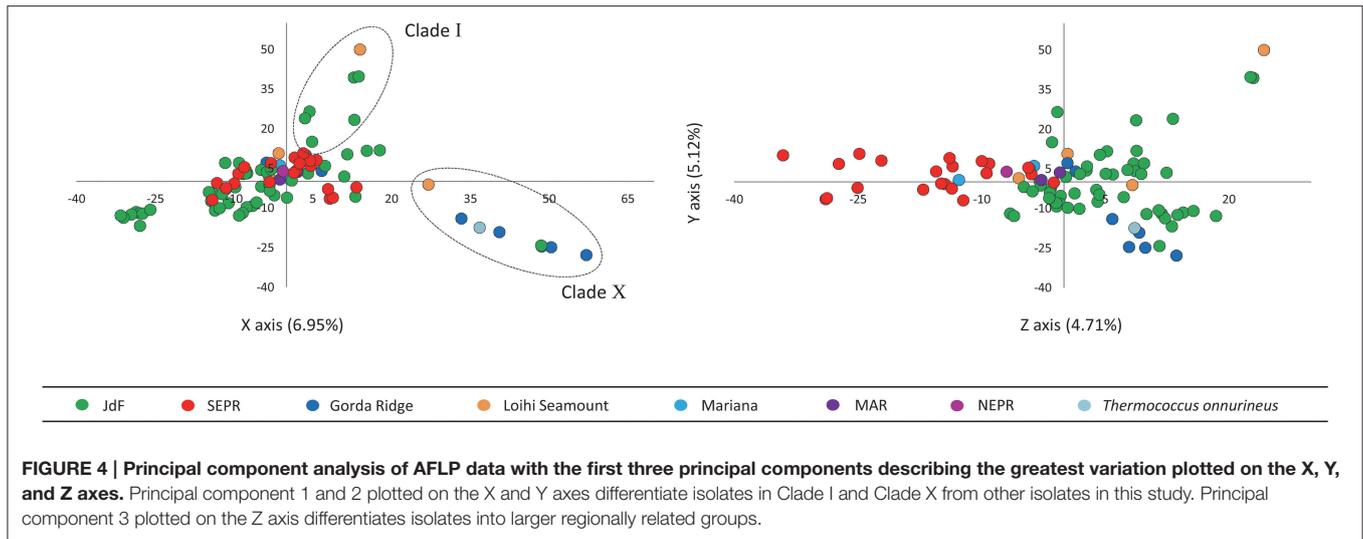
## Discussion

A comparison of *Thermococcus* isolates through AFLP and MLST analysis have made biogeographic patterns evident; however,

phylogenetic analysis does not suggest a simple model of allopatric speciation. This suggests that there is high dispersal in the *Thermococcus* genus. The strongest biogeographic patterns were observed between northern isolates from the JdF and southern isolates from the EPR, most likely due to these regions having the highest representation (Table 1).

A barrier to dispersal between northern and southern hydrothermal regions developed as the North American Plate overrode the Pacific Plate. The discontinuous nature of the plate boundary and associated hydrothermal vents allowed for endemic populations to develop over time between the northern JdF and southern EPR. The paraphyletic association between Clades VII and VIII demonstrates the shared ancestry between these two lineages and their divergence when populations became isolated from one another. The limited mixing of water masses through circulating ocean currents may have also contributed to the isolation of these populations (Müller et al., 2014). 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 high dispersal observed in *Thermococcus* may result from an ability to survive outside of the hydrothermal vent environment for extended periods of time. Thermophilic activity has been measured in equatorial deep-sea sediments hundreds of km away from any known hydrothermal activity, thereby hypothesizing long distance transport and survival from the EPR (Dobbs and Selph, 1997). Dispersal over great distances must be considered particularly when looking at the isolates found in Clade X, where high sequence and AFLP profile similarity was found in isolates from vent sites ~9500 km apart (i.e., Gorda Ridge and Manus Basin). Explanations for the high similarity observed in these isolates may include: a population bottleneck or recent population divergence, increased dispersal potential resulting from the ability to survive outside of the vent environment, or possibly a niche-specific ecotype due to unique metabolic characteristics. Isolates in Clade X that were collected from the Gorda Ridge have already been shown to grow over an atypically broad range of temperatures (45–90°C) for hyperthermophiles (Summit and Baross, 1998). 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 may be indicative of a gene transfer event providing a selective advantage. Isolates found in Clade X may also be representative of a subsurface ecotype, with five of the seven isolates in this clade isolated from a plume event (Summit and Baross, 1998). Hydrothermal fluids released from subsurface reservoirs during a plume event are believed to have residence times of from months to years and hydrothermal fluids have been detected thousands of kilometers from their origin (Lupton, 1996; Lupton et al., 1998, 1999). This makes hydrothermal plumes a potential vehicle for the dispersal of microorganisms associated with subsurface hydrothermal habitats over great distances. Long distance dispersal has also been demonstrated in dormant thermophilic endospores through the circulation of ocean currents (Müller et al., 2014). Further investigation of Clade X will require full-genome comparisons against other



*Thermococcus* genomes, to look for genes unique to this group that may allow for greater survival and dispersal outside the hydrothermal environment.

Clade I represents a group of *Thermococcus* that was unique in comparison to the other clades identified; however, Clade I like Clade X contained isolates from different regions (i.e., Loihi Seamount and Juan de Fuca Ridge). The most striking characteristic of the Clade I group was the phylogenetic distance between this group from other *Thermococcus* isolates as compared to their close proximity to several *Pyrococcus* type strains. This pattern was also observed in the GC-content analyses illustrating the divergence of the Clade I lineage with the *Pyrococcus* genus from the other *Thermococcus* isolates examined. In the SSU rRNA gene tree (Figure 3), Clade I was placed in an ancestral position relative to both *Pyrococcus* type strains and *Thermococcus* isolates from Clades II through X resulting in paraphyletic groupings of the *Thermococcus* and *Pyrococcus* genera. This topology has also been demonstrated in other phylogenetic analyses of the *Thermococcales* (Teske et al., 2009; Yilmaz et al., 2014). These data suggest that Clade I is a close sister lineage to *Pyrococcus* requiring reclassification.

Biodiversity in the *Thermococcus* genus has been made evident in this study through the identification of three or more lineages co-occurring within the same hydrothermal vent system. These divergent lineages have likely adapted to different ecological niches, allowing them to coexist at the same vent site. Previous research has shown a correlation between environmental habitat and phylogeny with the hydrothermal vent habitats of sulfide chimneys and seafloor zones proposed as the differing environments involved in the maintenance of diversity (Summit and Baross, 2001). Although no significant correlation was found in our study between lineages and the environmental samples from which they were isolated, the broader divergence observed in the phylogenetic tree may well be due to the differing source habitats as previously described. These broader lineage groups, Clades II and III vs. Clades VII and VIII, may represent lineages adapted to different niches and present in both

regions, which have diverged over time into regionally endemic populations. Determining the source environment for any of the lineages identified will be extremely difficult 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 the genomic analysis of metabolic characteristics associated with different lineages and how these relate to the utilization of different habitats.

Horizontal gene transfer (HGT) and recombination are believed to play an important role in the mixing of microbial populations and shaping of diversity (Gogarten et al., 2002; Nesbø et al., 2006; Zhaxybayeva et al., 2006), however the role of HGT is unclear since there is a decrease in these events as geographic distance and genetic distance among microorganisms increase, e.g., extrinsic and intrinsic barriers to gene transfer and recombination (Whitaker et al., 2005; Reno et al., 2009; Cadillo-Quiroz et al., 2012). Genomic diversity in *Thermococcus* isolates was investigated through the analysis of AFLP and MLST data for evidence of genomic rearrangements and gene transfer events. The conserved nature of MLST loci evident in dN/dS ratios and through linkage analysis support the theory that conserved housekeeping genes making up the core genome are unlikely to undergo gene transfer or recombination (Medini et al., 2005; Tettelin et al., 2008; Cordero and Polz, 2014). The same lineages identified through MLST were also identified through AFLP, although AFLP analysis detected greater genomic variation. Analysis of AFLP data through PCA (Figure 4) provides evidence suggestive of regionally related recombination, with phylogenetically distinct lineages clustered into larger regionally related groups. This is in contrast to what is shown in the ML phylogenetic tree constructed from MLST data (Figure 2) where isolates from a shared region are often differentiated into distinct clades. Shared genomic characteristics within regions, between divergent lineages, may be detected in AFLP data as the genome

wide restriction fragment lengths are more representative of variable regions of the genome where gene acquisition and loss occur more frequently.

Biogeographic patterns in viruses and mobile genetic elements have been identified in both terrestrial and marine environments as exemplified by *Sulfolobus* and *Vibrio*, respectively (Held and Whitaker, 2009; Boucher et al., 2011). In the *Thermococcales*, geographically associated recombination has been reported with the identification of a shared 16 kb region flanked by insertion sequence (IS) elements, found in *Pyrococcus furiosus* and *Thermococcus litoralis* both of which were isolated from Vulcano Island, Italy (Diruggiero et al., 2000). To address regionally related recombination in *Thermococcus*, mobile genetic elements like insertion sequence elements and transposons (Diruggiero et al., 2000; Zivanovic et al., 2002; Escobar-Páramo et al., 2005; Fukui et al., 2005; Mardanov et al., 2009) along with viral related elements (Fukui et al., 2005; Mardanov et al., 2009; Portillo and Gonzalez, 2009; Zivanovic et al., 2009; Vannier et al., 2011) and extracellular vesicles (Choi et al., 2015) can be investigated within populations at regional levels and across regions, for further evidence in support of regionally related recombination.

This study begins to elucidate the level of genomic resolution required to track population divergence in isolates of *Thermococcus* at marine hydrothermal systems. These systems are thermodynamically energized by strong gradients (e.g., temperature, oxygen, etc.) fueling multiple potential habitats within the subsurface crustal environment (Edwards et al., 2011; Orcutt et al., 2011). The ability to observe divergence in microorganisms is largely dependent on the

geographic scale investigated and level of genetic resolution applied. Biogeographic patterns shown in the archaeal genus *Thermococcus* demonstrate that even in microorganisms with dispersal over thousands of kilometers, divergence can occur when populations are isolated from one another. By identifying the appropriate scale and level of resolution required, the impact that geography and barriers to dispersal have on shaping microbial diversity are now being realized. Identifying the role and significance of mobile genetic elements in the shaping of microbial populations is the next step to understanding the evolutionary forces behind the planet's vast microbial diversity.

## Acknowledgments

Many thanks to John Baross and his lab for allowing us access to this amazing collection and to the many undergraduate researchers in the Moyer lab who helped maintain the Zoo Collection of *Thermococcus*, with special thanks to Debbie Whitley for getting us started during her summer internship. This work was funded in part by WWU's Office of Research and Sponsored Programs and by the Center for Dark Energy Biosphere Investigations. This is C-DEBI Contribution 279.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.00968>

## References

- Adams, M. W. (1993). Enzymes and proteins from organisms that grow near and above 100°C. *Annu. Rev. Microbiol.* 47, 627–658. doi: 10.1146/annurev.mi.47.100193.003211
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Amend, J. P., Meyer-Dombard, D. R., Sheth, S. N., Zolotova, N., and Amend, A. C. (2003). *Palaeococcus helgesonii* sp. nov., a facultatively anaerobic, hyperthermophilic archaeon from a geothermal well on Vulcano Island, Italy. *Arch. Microbiol.* 179, 394–401. doi: 10.1007/s00203-003-0542-7
- Atomi, H., Fukui, T., Kanai, T., Morikawa, M., and Imanaka, T. (2004). Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. *Archaea* 1, 263–267. doi: 10.1155/2004/204953
- Baas-Becking, L. G. M. (1934). *Geobiologie of Inleiding tot de Milieukunde*. The Hague: Van Stockum, W. P., and N. V. Zoon.
- Bae, S. S., Kim, Y. J., Yang, S. H., Lim, J. K., Jeon, J. H., Lee, H. S., et al. (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.
- Bonnet, E., and de Peer, Y. V. (2002). zt: a software tool for simple and partial Mantel tests. *J. Stat. Softw.* 7, 1–12.
- Boucher, Y., Cordero, O. X., Takemura, A., Hunt, D. E., Schliep, K., Bapteste, E., et al. (2011). Local mobile gene pools rapidly cross species boundaries to create endemism within global *Vibrio cholerae* populations. *mBio* 2, e00335–e00310. doi: 10.1128/mBio.00335-10
- Cadillo-Quiroz, H., Didelot, X., Held, N. L., Herrera, A., Darling, A., Reno, M. L., et al. (2012). Patterns of gene flow define species of thermophilic Archaea. *PLoS Biol.* 10:e1001265. doi: 10.1371/journal.pbio.1001265
- Cho, J. -C., and Tiedje, J. M. (2000). Biogeography and degree of endemism of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* 66, 5448–5456. doi: 10.1128/AEM.66.12.5448-5456.2000
- Choi, D. H., Kwon, Y. M., Chiura, H. X., Yang, E. C., Bae, S. S., Kang, S. G., et al. (2015). Extracellular vesicles of the hyperthermophilic archaeon *Thermococcus onnurineus* NA1<sup>T</sup>. *Appl. Environ. Microbiol.* 81, 4591–4599. doi: 10.1128/AEM.00428-15
- Cordero, O. X., and Polz, M. F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* 12, 263–273. doi: 10.1038/nrmicro3218
- Davis, R. E., and Moyer, C. L. (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, B08S15. doi: 10.1029/2007JB005413
- DeChaine, E. G., Bates, A. E., Shank, T. M., and Cavanaugh, C. M. (2006). Off-axis symbiosis found: characterization and biogeography of bacterial symbionts of *Bathymodiolus* mussels from Lost City hydrothermal vents. *Environ. Microbiol.* 8, 1902–1912. doi: 10.1111/j.1462-2920.2005.01113.x
- Diruggiero, J., Dunn, D., Maeder, D. L., Holley-Shanks, R., Chatard, J., Horlacher, R., et al. (2000). Evidence of recent lateral gene transfer among hyperthermophilic Archaea. *Mol. Microbiol.* 38, 684–693. doi: 10.1046/j.1365-2958.2000.02161.x
- Dobbs, F. C., and Selph, K. A. (1997). Thermophilic bacterial activity in a deep-sea sediment from the Pacific Ocean. *Aquatic Microbiol. Ecol.* 13, 209–212. doi: 10.3354/ame013209
- Edwards, K. J., Wheat, C. G., and Sylvan, J. B. (2011). Under the sea: microbial life in volcanic oceanic crust. *Nat. Rev. Microbiol.* 9, 703–712. doi: 10.1038/nrmicro2647

- Erauso, G., Reysenbach, A. -L., Godfroy, A., Meunier, J. -R., Crump, B., Partensky, F., et al. (1993). *Pyrococcus abyssi* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch. Microbiol.* 160, 338–349. doi: 10.1007/BF00252219
- Escobar-Páramo, P., Ghosh, S., and DiRuggiero, J. (2005). Evidence for genetic drift in the diversification of a geographically isolated population of the hyperthermophilic archaeon *Pyrococcus*. *Mol. Biol. Evol.* 22, 2297–2303. doi: 10.1093/molbev/msi227
- Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1, 47–50.
- Fiala, G., and 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. doi: 10.1007/BF00413027
- Flores, G. E., Shakya, M., Meneghin, J., Yang, Z. K., Seewald, J. S., Wheat, C. G., et al. (2012). Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* 10, 333–346. doi: 10.1111/j.1472-4669.2012.00325.x
- Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S., and Imanaka, T. (2005). Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* 15, 352–363. doi: 10.1101/gr.3003105
- Garrity, G. M., Boone, D. R., and Castenholz, R. W. (2001). *Bergey's Manual of Systematic Bacteriology, Vol. 1: The Archaea and the Deeply Branching and Phototrophic Bacteria*. New York, NY: Springer.
- Gogarten, J. P., Doolittle, W. F., and Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238. doi: 10.1093/oxfordjournals.molbev.a004046
- Häne, B. G., Jäger, K., and Drexler, H. G. (1993). The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprint profiles than band matching algorithms. *Electrophoresis* 14, 967–972. doi: 10.1002/elps.11501401154
- Haubold, B., and Hudson, R. R. (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Bioinformatics* 16, 847–848. doi: 10.1093/bioinformatics/16.9.847
- Held, N. L., and Whitaker, R. J. (2009). Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environ. Microbiol.* 11, 457–466. doi: 10.1111/j.1462-2920.2008.01784.x
- Holden, J. F., Takai, K., Summit, M., Bolton, S., Zyskowski, J., and Baross, J. A. (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. doi: 10.1111/j.1574-6941.2001.tb00825.x
- Huber, J. A., Butterfield, D. A., and Baross, J. A. (2006). Diversity and distribution of subseafloor Thermococcales populations in diffuse hydrothermal vents at an active deep-sea volcano in the northeast Pacific Ocean. *J. Geophys. Res.* 111, G04016. doi: 10.1029/2005JG000097
- Jannasch, H. W., Wirsén, C. O., Molyneux, S. J., and Langworthy, T. A. (1992). Comparative physiological studies on hyperthermophilic Archaea isolated from deep-sea hot vents with emphasis on *Pyrococcus* strain GB-D. *Appl. Environ. Microbiol.* 58, 3472–3481.
- Jun, X., Lupeng, L., Minjuan, X., Oger, P., Fengping, W., Jebbar, M., et al. (2011). Complete genome sequence of the obligate piezophilic hyperthermophilic archaeon *Pyrococcus yayanosii* CH1. *J. Bacteriol.* 193, 4297–4298. doi: 10.1128/JB.05345-11
- Kaplan, J. B., and Fine, D. H. (1998). Codon usage in *Actinobacillus actinomycetemcomitans*. *FEMS Microbiol. Lett.* 163, 31–36. doi: 10.1111/j.1574-6968.1998.tb13022.x
- Kim, Y. J., Lee, H. S., Kim, E. S., Bae, S. S., Lim, J. K., Matsumi, R., et al. (2010). Formate-driven growth coupled with H<sub>2</sub> production. *Nature* 467, 352–355. doi: 10.1038/nature09375
- Konstantinidis, K. T., Ramette, A., and Tiedje, J. M. (2006a). Toward a more robust assessment of intraspecific diversity, using fewer genetic markers. *Appl. Environ. Microbiol.* 72, 7286–7293. doi: 10.1128/AEM.01398-06
- Konstantinidis, K. T., Ramette, A., and Tiedje, J. M. (2006b). The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1929–1940. doi: 10.1098/rstb.2006.1920
- Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Kuhn, G., Francioli, P., and Blanc, D. S. (2006). Evidence for clonal evolution among highly polymorphic genes in methicillin-resistant *Staphylococcus aureus*. *J. Bacteriol.* 188, 169–178. doi: 10.1128/JB.188.1.169-178.2006
- Lecompte, O., Ripp, R., Puzos-Barbe, V., Duprat, S., Heilig, R., Dietrich, J., et al. (2001). Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. *Genome Res.* 11, 981–993. doi: 10.1101/gr.GR1653R
- Lee, H. S., Kang, S. G., Bae, S. S., Lim, J. K., Cho, Y., Kim, Y. J., et al. (2008). The complete genome sequence of *Thermococcus onnurineus* NAI reveals a mixed heterotrophic and carboxydrotrophic metabolism. *J. Bacteriol.* 190, 7491–7499. doi: 10.1128/JB.00746-08
- Lepage, E., Marguet, E., Gesline, C., Matte-Tailliez, O., Zillig, W., Forterre, P., et al. (2004). Molecular diversity of new *Thermococcales* 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. doi: 10.1128/AEM.70.3.1277-1286.2004
- Lin, J. -J., Kuo, J., and Ma, J. (1996). A PCR-based DNA fingerprinting technique: AFLP for molecular typing of bacteria. *Nucleic Acids Res.* 24, 3649–3650. doi: 10.1093/nar/24.18.3649
- Lupton, J. E. (1996). A far-field hydrothermal plume from Loihi Seamount. *Science* 272, 976–979. doi: 10.1126/science.272.5264.976
- Lupton, J. E., Baker, E. T., Garfield, N., Massoth, G. J., Feely, R. A., Cowen, J. P., et al. (1998). Tracking the evolution of a hydrothermal event plume with a RAFOS neutrally buoyant drifter. *Science* 280, 1052–1055. doi: 10.1126/science.280.5366.1052
- Lupton, J. E., Baker, E. T., and Massoth, G. J. (1999). Helium, heat, and the generation of hydrothermal event plumes at mid-ocean ridges. *Earth Planet. Sci. Lett.* 171, 343–350. doi: 10.1016/S0012-821X(99)00149-1
- Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Microbiology* 95, 3140–3145. doi: 10.1073/pnas.95.6.3140
- Mardanov, A. V., Ravin, N. V., Svetlitchnyi, V. A., Beletsky, A. V., Miroshnichenko, M. L., Bonch-Osmolovskaya, E. A., et al. (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, 4580–4588. doi: 10.1128/AEM.00718-09
- McAllister, S. M., Davis, R. E., McBeth, J. M., Tebo, B. M., Emerson, D., and Moyer, C. L. (2011). Biodiversity and emerging biogeography of the neutrophilic iron-oxidizing *Zetaproteobacteria*. *Appl. Environ. Microbiol.* 77, 5445–5457. doi: 10.1128/AEM.00533-11
- Medini, D., Donati, C., Tettelin, H., Massignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006
- Meyer, J. L., and Huber, J. A. (2014). Strain-level genomic variation in natural populations of *Lebetimonas* from an erupting deep-sea volcano. *ISME J.* 8, 867–880. doi: 10.1038/ismej.2013.206
- Mino, S., Makita, H., Toki, T., Miyazaki, J., Kato, S., Watanabe, H., et al. (2013). Biogeography of *Persephonella* in deep-sea hydrothermal vents of the Western Pacific. *Front. Microbiol.* 4:107. doi: 10.3389/fmicb.2013.00107
- Müller, A. L., Rosa de Rezende, J. R., Hubert, C. R., Kjeldsen, K. U., Lagkouvardos, I., Berry, D., et al. (2014). Endospores of thermophilic bacteria as tracers of microbial dispersal by ocean currents. *ISME J.* 8, 1153–1165. doi: 10.1038/ismej.2013.225
- Muto, A., and Osawa, S. (1987). The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci. U.S.A.* 84, 166–169. doi: 10.1073/pnas.84.1.166
- Nesbo, C. L., Dlutek, M., and Doolittle, W. F. (2006). Recombination in *Thermotoga*: implications for species concepts and biogeography. *Genetics* 172, 759–769. doi: 10.1534/genetics.105.049312
- Oger, P., Sokolova, T. G., Kozhevnikova, D. A., Chernykh, N. A., Bartlett, D. H., Bonch-Osmolovskaya, E. A., et al. (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, 7019–7020. doi: 10.1128/JB.06259-11
- Olive, D. M., and Bean, P. (1999). Principles and applications of methods for DNA-based typing of microbial organisms. *J. Clin. Microbiol.* 37, 1661–1669.
- Opatkiewicz, A. D., Butterfield, D. A., and Baross, J. A. (2009). Individual hydrothermal vents at Axial Seamount harbor distinct seafloor microbial communities. *FEMS Microbiol. Ecol.* 70, 413–424. doi: 10.1111/j.1574-6941.2009.00747.x
- Orcutt, B. N., Sylvan, J. B., Knab, N. J., and Edwards, K. J. (2011). Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol. Mol. Biol. Rev.* 75, 361–422. doi: 10.1128/MMBR.00039-10
- Papke, R. T., Ramsing, N. B., Bateson, M. M., and Ward, D. M. (2003). Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* 5, 650–659. doi: 10.1046/j.1462-2920.2003.00460.x
- Papke, R. T., and Ward, D. M. (2004). The importance of physical isolation to microbial diversification. *FEMS Microbiol. Ecol.* 48, 293–303. doi: 10.1016/j.femsec.2004.03.013
- Pond, S. L., and Frost, S. D. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531–2533. doi: 10.1093/bioinformatics/bt1320
- Portillo, M. C., and Gonzalez, J. M. (2009). CRISPR elements in the Thermococcales: evidence for associated horizontal gene transfer in *Pyrococcus furiosus*. *J. Appl. Genet.* 50, 421–430. doi: 10.1007/BF03195703
- Rademaker, J. L., Hoste, B., Louws, F. J., Kersters, K., Swings, J., Vauterin, L., et al. (2000). Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *Int. J. Syst. Evol. Microbiol.* 50, 665–677. doi: 10.1099/00207713-50-2-665
- Ramette, A., and Tiedje, J. M. (2007). Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb. Ecol.* 53, 197–207. doi: 10.1007/s00248-005-5010-2
- Reno, M. L., Held, N. L., Fields, C. J., Burke, P. V., and Whitaker, R. J. (2009). Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8605–8610. doi: 10.1073/pnas.0808945106
- Robb, F. T., and Place, A. R., (eds.) (1995). *Thermophiles: Archaea a Laboratory Manual*. Plainview, NY: Cold Spring Harbor Laboratory Press.
- Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132, 365–386. doi: 10.1385/1-59259-192-2:365
- Savelkoul, P. H., Aarts, H. J., de Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., et al. (1999). Amplified-fragment length polymorphism analysis: the state of an art. *J. Clin. Microbiol.* 37, 3083–3091.
- Sokolova, T. G., Henstra, A. M., Sipma, J., Parshina, S. N., Stams, A. J., and Lebedinsky, A. V. (2009). Diversity and ecophysiological features of thermophilic carboxydrotrophic anaerobes. *FEMS Microbiol. Ecol.* 68, 131–141. doi: 10.1111/j.1574-6941.2009.00663
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446
- Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* 57, 758–771. doi: 10.1080/10635150802429642
- Summit, M., and Baross, J. A. (1998). Thermophilic seafloor microorganisms from the 1996 north Gorda Ridge eruption. *Deep Sea Res. II* 45, 2751–2766. doi: 10.1016/S0967-0645(98)00092-7
- Summit, M., and Baross, J. A. (2001). A novel microbial habitat in the mid-ocean ridge seafloor. *Proc. Natl. Acad. Sci. U.S.A.* 98, 2158–2163. doi: 10.1073/pnas.051516098
- Takai, K., Sugai, A., Itoh, T., and 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. doi: 10.1099/00207713-50-2-489
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Teske, A., Edgcomb, V., Rivers, A. R., Thompson, J. R., de Vera Gomez, A., Molyneux, S. J., et al. (2009). A molecular and physiological survey of a diverse collection of hydrothermal vent *Thermococcus* and *Pyrococcus* isolates. *Extremophiles* 13, 905–915. doi: 10.1007/s00792-009-0278-7
- Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 12, 472–477. doi: 10.1016/j.mib.2008.09.006
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (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, 4673–4680. doi: 10.1093/nar/22.22.4673
- Vannier, P., Marteinsson, V. T., Fridjonsson, O. H., Oger, P., and Jebbar, M. (2011). Complete genome sequence of the hyperthermophilic, piezophilic, heterotrophic, and carboxydrotrophic archaeon *Thermococcus barophilus* MP. *J. Bacteriol.* 193, 1481–1482. doi: 10.1128/JB.01490-10
- Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699. doi: 10.1093/oxfordjournals.molbev.a003851
- Whitaker, R. (2006). Allopatric origins of microbial species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1975–1984. doi: 10.1098/rstb.2006.1927
- Whitaker, R. J., Grogan, D. W., and Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301, 976–978. doi: 10.1126/science.1086909
- Whitaker, R. J., Grogan, D. W., and Taylor, J. W. (2005). Recombination shapes the natural population structure of the hyperthermophilic archaeon *Sulfolobus islandicus*. *Mol. Biol. Evol.* 22, 2354–2361. doi: 10.1093/molbev/msi233
- Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and “All-species living tree project (LTP)” taxonomic frameworks. *Nucleic Acids Res.* 42, D643–D648. doi: 10.1093/nar/gkt1209
- Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F., and Papke, R. T. (2006). Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* 16, 1099–1108. doi: 10.1101/gr.5322306
- Zillig, W., Holz, I., Janekovic, D., Schäfer, W., and Reiter, W. (1983). The archaeobacterium *Thermococcus celer* represents, a novel genus within the thermophilic branch of the archaeobacteria. *Syst. Appl. Microbiol.* 4, 88–94. doi: 10.1016/S0723-2020(83)80036-8
- Zivanovic, Y., Armengaud, J., Lagorce, A., Leplat, C., Guérin, P., Dutertre, M., et al. (2009). Genome analysis and genome-wide proteomics of *Thermococcus gammatolerans*, the most radioresistant organism known amongst the Archaea. *Genome Biol.* 10:R70. doi: 10.1186/gb-2009-10-6-r70
- Zivanovic, Y., Lopez, P., Philippe, H., and Forterre, P. (2002). *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. *Nucleic Acids Res.* 30, 1902–1910. doi: 10.1093/nar/30.9.1902

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Price, Fullerton and Moyer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.