Novel, Deep-Branching Heterotrophic Bacterial Populations Recovered from Thermal Spring Metagenomes

Daniel R. Colman1†, Zackary J. Jay2, William P. Inskeep2*, Ryan deM. Jennings2†, Kendra R. Maas1†, Douglas B. Rusch3 and Cristina D. Takacs-Vesbach1*

1 Department of Biology, University of New Mexico, Albuquerque, NM, USA, 2 Thermal Biology Institute and Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT, USA, 3 Center for Genomics and Bioinformatics, Indiana University, Bloomington, IN, USA

Thermal spring ecosystems are a valuable resource for the discovery of novel hyperthermophilic Bacteria and Archaea, and harbor deeply-branching lineages that provide insight regarding the nature of early microbial life. We characterized bacterial populations in two circumneutral (pH ∼8) Yellowstone National Park thermal (T ∼80°C) spring filamentous “streamer” communities using random metagenomic DNA sequence to investigate the metabolic potential of these novel populations. Four de novo assemblies representing three abundant, deeply-branching bacterial phylotypes were recovered. Analysis of conserved phylogenetic marker genes indicated that two of the phylotypes represent separate groups of an uncharacterized phylum (for which we propose the candidate phylum name “Pyropristinus”). The third new phylotype falls within the proposed Calescamantes phylum. Metabolic reconstructions of the “Pyropristinus” and Calescamantes populations showed that these organisms appear to be chemoorganoheterotrophs and have the genomic potential for aerobic respiration and oxidative phosphorylation via archaeal-like V-type, and bacterial F-type ATPases, respectively. A survey of similar phylotypes (>97% nt identity) within 16S rRNA gene datasets suggest that the newly described organisms are restricted to terrestrial thermal springs ranging from 70 to 90°C and pH values of ∼7–9. The characterization of these lineages is important for understanding the diversity of deeply-branching bacterial phyla, and their functional role in high-temperature circumneutral “streamer” communities.

Keywords: Aquificales, hot springs, Thermotogae, Calescamantes, Pyropristinus, hyperthermophiles, Yellowstone National Park

INTRODUCTION

The discovery and characterization of early-branching lineages of Bacteria and Archaea has been crucial for studying the origin and evolution of life on Earth. There is considerable evidence for the hypothesis that life originated in environments similar to modern hydrothermal settings, although other scenarios are also proposed (e.g., cold origins; Price, 2009). Hyperthermophiles inhabit geothermal environments that are analogous to those of early Earth (Baross and Hoffman, 1985), and are generally the deepest branching representatives of the tree of Life (Di Giulio, 2003; Stetter, 2006). The well-characterized and largely hyperthermophilic bacterial phyla Aquificae
and Thermotogae have been considered the most basal bacterial lineages on the basis of phylogenetic evidence (Barion et al., 2007; Zhaxybayeva et al., 2009). More recently, an uncultured bacterium from subsurface thermal fluids, Candidatus “Acetothermum autotrophicum,” has also been posited as one of the deep lineages in the Bacteria based on phylogenetic analysis of genome sequence (Takami et al., 2012). Consequently, discovery and characterization of new and uncultured lineages of thermophilic microorganisms are extremely useful toward the broader goal of understanding genomic and metabolic attributes of deep-branching phyla, which inhabit modern-day environments that may be analogs to those potentially important in the origin(s) of life.

The characterization of uncultured microorganisms from thermal environments has been integral for expanding the scope of known microbial diversity. Early phylogenetic surveys based on 16S rRNA gene analysis revealed a significant diversity of uncultivated microorganisms in various hydrothermal settings, including numerous candidate phyla (Barns et al., 1994; Reysenbach et al., 1994; Hugenholtz et al., 1998; Takai and Horikoshi, 1999). However, due to the difficulty of cultivating environmentally relevant microorganisms (particularly extremophiles), the physiological diversity of many of these phyla has remained largely unknown since their discovery. Environmental genomics (e.g., metagenomics and single-cell genomics) has provided valuable tools for assessing the metabolic capabilities and phylogenetic diversity of thermophiles and other extremophilic Bacteria and Archaea (Baker et al., 2010; Nunoura et al., 2011; Takami et al., 2012; Dodsworth et al., 2013; Inskeep et al., 2013; Kantor et al., 2013; Kozubal et al., 2013; Rinke et al., 2013; Hedlund et al., 2014; Wrighton et al., 2014; Castelle et al., 2015). However, numerous microbial phyla remain uncharacterized, and continued studies in high-temperature habitats hold promise for dissecting the functional role of early-branching lineages in less-complex microbial communities.

Filamentous “streamer” communities containing members of the Aquificales are common in geothermal spring outflow channels and hydrothermal vents in marine systems globally (Ferrera et al., 2007). We recently described and characterized metagenomes from six filamentous “streamer” communities from geochemically distinct habitat types from Yellowstone National Park (YNP; Inskeep et al., 2013; Takacs-Vesbach et al., 2013). Three primary genera of Aquificales dominate different streamer communities based on geochemical conditions (e.g., pH, sulfide), and each habitat type supports different co-occurring heterotrophic community members. Two non-sulfidic, slightly alkaline (~7.8–8) streamer communities (Octopus and Bechler springs) contained abundant Thermocrinis spp. (Aquificales) and representatives of several uncultured bacterial lineages. A novel member of the Aigarchaeota (Ca. Caldentenuis aerorheumensis) was recently characterized from the Octopus Spring (OS) streamer communities (Beam et al., 2016); however, several novel and abundant bacteria in these communities have remained uncharacterized. Consequently, the objectives of this study were to (1) obtain and curate de novo sequence assemblies from these two streamer communities corresponding to three uncharacterized bacterial phylotypes, (2) assess the phylogenetic position and functional potential of the three phylotypes, and (3) determine the distribution of these populations in YNP and other thermal environments. Here we describe three new phylotypes curated from random shotgun Sanger sequencing of two slightly alkaline (pH ~8) filamentous “streamer” communities (temperature ~80°C) from Octopus and Bechler springs (Yellowstone National Park). These aerobic chemoorganoheterotrophs are representatives of two distinct and deeply-branching, phylum-level lineages in the domain Bacteria. “Pyropristinus” is proposed here as a newly described lineage containing two distinct phylotypes, while the other novel phylotype belongs to the proposed Calescamantes phylum (Rinke et al., 2013; Hedlund et al., 2014). The discovery and characterization of these early-branching bacteria are critical for dissecting microbial community structure and function in modern-day high-temperature habitats, and provides significant opportunities for understanding the evolution of deeply-branching hyperthermal bacterial lineages.

MATERIALS AND METHODS

Recovery of Uncharacterized Bacterial Populations from Hot-Spring Metagenomes

Details of site sampling, metagenome sequencing, assembly, and geochemical analyses have been described previously (Inskeep et al., 2013; Takacs-Vesbach et al., 2013). Briefly, filamentous microbial community samples were collected near the top of the spring runoff channels from a spring in the Bechler Three Rivers Junction region of YNP (T = 80–82°C, pH = 7.8; 44.2858 N, −110.8784 E) and Octopus Spring in the Lower Geyser Basin of YNP (T = 80–82°C, pH = 7.9; 44.53408 N, −110.7979 E). A phenol/chloroform extraction method was used to extract community DNA (Inskeep et al., 2010), which was then used to construct a small-insert clone library. Sanger sequencing was used for random shotgun sequencing of the inserts (~40 Mb total DNA sequence for each site). Metagenomes were assembled using the Celera assembler; automated tools in the Integrated Microbial Genomes server (IMG; Markowitz et al., 2012) were used to predict and annotate genes. Nucleotide word frequency–principal components analysis (NWF-PCA) was used to identify predominant populations in the metagenomic contigs (>3 kbp) as described previously (Takacs-Vesbach et al., 2013). The contigs were further analyzed and screened using G+C content (%) and phylogenetic analysis (most useful for phylotypes exhibiting closest neighbors above 80% nt ID) to obtain four de novo sequence assemblies corresponding to abundant and uncharacterized members of these communities (T1.1, T1.2, T2.1, T3.1; contig coverage > 1). Genome completeness was estimated using three metrics: tRNA synthetase complement was estimated by the presence of at minimum, one partial annotated gene for each of 21 prokaryotic genes coding for tRNA synthetases, the presence of 178 “conserved” bacterial housekeeping genes (Garcia Martin et al., 2006) and 40 “conserved” prokaryotic universal markers (Wu et al., 2013).
Amino acid identity (AAI) was used to assess taxonomic relationships among assemblies and other closely related genomes. AAI was calculated as the mean percentage of differing amino acid residues between homologous protein coding genes in pairwise comparisons of assemblies using blastp after filtering of low quality alignments (Konstantinidis and Tiedje, 2005b). Protein coding gene alignments were used that (1) shared at least 30% a.a. similarity, (2) were alignable up to 70% of the length of the subject sequence, and (3) had an alignment length of at least 100 residues. The T3.1 (Calescambantes-like) phylogeny from OS was also compared to the recently described Ca. Calescibacterium nevadense (EM19-like) genome (IMG taxon ID: 2527291514). Average nucleotide identity (ANI) was calculated between scaffolds of the two closely related assemblies, T1.1 and T1.2 using default parameters (minimum length = 700 bp, window size = 100 bp, min. identity = 70%, min alignment number = 50, window size = 100 bp and step size = 200 bp) with the online ANI calculator (http://enve-omics.ce.gatech.edu/ani/index; Goris et al., 2007).

**Phylogenetic Analyses**

Phylogenetic analyses were conducted by surveying all three lineages (T1, T2, and T3) for homologous single-copy housekeeping genes (at least partial copies shared amongst all three lineages) that were previously identified as bacterial-specific or universal (Wu et al., 2013). Genomic references were chosen based on blastp searches of ribosomal proteins against publically available genomes and curated such that every bacterial reference (Supplementary Table 1) contained ≥ 16 of the 18 total genes (5 universal: pheT, rplE, rplO, rpsK, rpsM, and 13 bacterial-specific marker genes: clpX, leuS, ligA, murD, pnp, pth, pyrG, rplL, rpoA, rpoB, rpoC, secY, serS; with the exception of Ca. C. nevadense: 3 genes missing). Each gene was aligned individually with Clustal Omega (Sievers et al., 2011), and alignment positions were confidence weighted using Zorro (Wu et al., 2012) to reduce the influence of ambiguously aligned positions. An evolutionary substitution model was chosen for each individual gene alignment using ProtTest v. 3.4 (Darriba et al., 2011). The concatenated gene alignment (8928 informative amino acid positions) was used in a maximum likelihood (ML) analysis in RAxML v. 8.2.3 (Stamatakis, 2006) using alignment weights, and partitioning the concatenation so that each gene was modeled separately by the appropriate substitution model (primarily the LG substitution model; with a gamma distribution of rates and proportion of invariant sites). Archaeal outgroups (n = 27) were used to root bacterial phylogenies using the five universal single copy housekeeping genes of the dataset that were common to both Bacteria and Archaeae. Phylogenies were bootstrapped with 1000 ML replicates using the RAxML rapid bootstrapping algorithm.

Phylogenetic analysis was also conducted using near full-length 16S rRNA genes (>1300 bp). The T3.1 (EM19-like) and T1.1 assemblies were omitted from this analysis because they did not contain full-length 16S rRNA genes, although conspecific (>97% nt identity) relatives of these lineages serve as proxies for their phylogenetic placement. A 953 bp 16S rRNA gene present in the Octopus Spring metagenome (but not included in the T3 de novo assembly due to length) was 99% identical (nucleotide) to a nearly full length EM19-like 16S rRNA gene sequence obtained from Octopus Spring (OS_clone_YNP11_11_1). Further, the ~850 bp T1.1 16S rRNA gene from the T1.1 Octopus Spring assembly was 98% identical to the nearly full-length T1.2 16S rRNA gene from the Bechler assembly, and thus the more complete T1.2 16S rRNA gene sequence from Bechler was used to represent both the OS T1.1 and the Bechler T1.2 population. Genes were aligned using PyNAST (Caporaso et al., 2010) with the Greengenes reference dataset (DeSantis et al., 2006). The DNA substitution model for the alignment was selected using Modeltest v. 3.7 (Posada and Crandall, 1998) and the Akaike Information Criterion (AIC) model metric. ML analysis was conducted in MEGA v.6 (Tamura et al., 2013) using the General Time Reversible model with a proportion of invariant sites and a gamma distribution of rates.

Conserved signature indel (CSI) analyses were also used to assess if the newly described lineages belonged to closely related phyla, such as the Thermotogae and Aquificae. CSIs specific to the Thermotogae (18 total; Gupta and Bhandari, 2011) or Aquificae (4 total; Gupta and Lali, 2013), relative to the rest of Bacteria, were used by referencing the de novo sequence assemblies against available Thermotogae and Aquificae genomes available in IMG. A total of 22 genes (encompassing 22 CSIs) were aligned with Clustal, as described in the original study (Gupta and Bhandari, 2011), and inspected for the characteristic CSIs.

**Metabolic Reconstruction**

Functional similarity of the “Pyropristinus” and Calescambantes-OS types relative to the Thermotogae, Aquificae, Ca. Calescibacterium nevadense, and Thermodesulfolobacteria were statistically assessed with a non-metric multidimensional scaling (NMDS) ordination of the presence-absence of COGs
in each genome (using Euclidean distances) in R. Annotated genes were used to assess the presence of metabolic pathways in all three lineages. The conspecific-level relatedness between T1.1 and T1.2 assemblies allowed the use of the less complete T1.1 assembly to augment the genes not found in T1.2. Where possible, genome sequence of *Ca. Calescibacterium nevadense* (Rinke et al., 2013) was used as a reference for the presence of pathways in T3, which was related to *Ca. C. nevadense*. Genomic data for the four assemblies produced here is available under the NCBI Bioproject ID PRJNA280379.

**Ecological Distribution**

Full-length 16S rRNA genes of the three lineages were used in BLASTn searches against available datasets to determine the habitat distribution of these newly described populations. Because a full-length 16S rRNA gene was not present in the T3.1 assembly, a representative sequence from the 16S rRNA gene library of the same Octopus Spring metagenome sample was used (Takacs-Vesbach et al., 2013). This clone group (EM19) was also described in pink-streamer communities of the same spring (Reysenbach et al., 1994; Blank et al., 2002).
Ca. C. nevadense is also closely related to the EM19 clone from Octopus Spring (Rinke et al., 2013), and was the closest genome sequence available for comparison to the Calescamantes-like assembly from Octopus Spring (Calescamantes-OS; T3.1 used here). Searches were conducted against 16S rRNA gene datasets including Genbank, IMG metagenomes, Greengenes (DeSantis et al., 2006), the Ribosomal Database Project (Cole et al., 2014), as well as YNP-specific surveys (including 454 pyrosequencing datasets) of 49 YNP springs spanning a wide range of temperature and pH values (Takacs-Vesbach et al., unpublished data) and clone-libraries of 82 YNP springs (Mitchell, 2009). 16S rRNA gene matches with >97% nucleotide identity to each of the three lineages were considered a positive occurrence. Metadata for each reference sample (temperature, pH, and geographic location) were collected from the publishing reports, where available, and augmented with data from the YNP Research Coordination Network database (http://www.rcn.montana.edu; Supplementary Table 2). Mean values for sample temperature are used where ranges were reported. Statistical differences of temperature and pH distributions among groups were tested using a Kruskal–Wallis rank sum analysis of variance test in R (R Core Team, 2014).

RESULTS AND DISCUSSION

Recovery of Uncharacterized Bacterial Populations from Hot-Spring Metagenomes

The assembled metagenome sequence from Octopus and Bechler Springs was analyzed using nucleotide word frequency–principal components analysis (NWF-PCA) to obtain contigs and scaffolds (>3 kbp length only) sharing similar sequence character (Figure 1). These scaffolds and contigs were further separated using G+C content (%), coverage and phylogenetic analysis to obtain de novo sequence assemblies corresponding to each of the predominant phylotypes in these communities (Supplemental Figure 1). The Octopus Spring community contained at least eight predominant phylotypes (Desulfurococcaceae, not shown), while Bechler spring contained three abundant phylotypes (Figure 1A). Each of the two streamer communities contained highly-related populations of Thermocrinis spp. (Aquificales), Pyrobulcum spp. (Thermoproteales), and a novel population referred to here as Type 1 (T1) of candidate phylum “Pyropristinus.” The streamer community from Octopus Spring also contained abundant populations of a Type 2 (T2) “Pyropristinus” population, a relative of the proposed bacterial phylum Calescamantes (EM19 candidate division; Rinke et al., 2013; Hedlund et al., 2014), an uncharacterized member of the Firmicutes, and a member of the candidate archaeal phylum Aigarchaeota (Beam et al., 2016). The average coverage, G+C content (%), and cumulative sequence plots of contigs corresponding to the three “Pyropristinus” (T1.1, T1.2, T2), and the Calescamantes populations are provided in supplemental information (Supplemental Figure 1). To assess the contribution of these assemblies to population abundances in the metagenomic data, the de novo assemblies compiled from these sites were used to bin the original random metagenome sequence reads (Figure 2; Table 1). A G+C (%) frequency plot of random metagenome sequence reads (average read length = 820 bp) that were phylogenetically assigned (>90% nucleotide identity) to the de novo assemblies compiled from these sites showed that “Pyropristinus” T1 and T2, as well as the Calescamantes-like phylotypes were significant members (~7–8% of all reads) of the more even microbial community in Octopus Spring, and that the “Pyropristinus” T1 was also abundant (~12%) in Bechler spring (Figure 2, Table 1).

Average estimates of genome completeness based on housekeeping genes present in the de novo sequence assemblies were 65, 72, and 63% for the T1 “Pyropristinus” (T1.2), T2 Pyropristitus (T2.1), and the Calescamantes-like populations, respectively (Table 2). Estimates by tRNA synthetase complement were higher (86% for both T1.2 and T2) than those based on the presence of “conserved” housekeeping genes involved in many cellular processes (50–60% and 59–73% for T1.2 and T2, respectively). The lower estimates based on the presence of housekeeping genes involved in a variety of cellular processes may be conservative due to the lack of appropriate references for identifying these genes in phylogenetically novel, deep-branching organisms (discussed further below).

| TABLE 2 | Genome assembly statistics for “Pyropristinus” Type 1 and Type 2, and Calescamantes populations from Octopus and/or Bechler springs. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Spring Name     | “Pyropristinus” T1 | “Pyropristinus” T1 | “Pyropristinus” T1 | “Pyropristinus” T1 | “Pyropristinus” T1 |
|                 | Octopus          | Bechler          | Octopus          | Calescamantes-OS |
| Phyla Abbrev.    | T1.1             | T1.2             | T2.1             | T3.1             |
| Size (in Mbp)   | 1.02             | 1.24             | 1.1              | 1.29             |
| Completenessa   | 47.0             | 64.9             | 72.2             | 62.9             |
| G+C (%)         | 44.3             | 44.3             | 28.9             | 35.2             |
| Contigs         | 116              | 72               | 117              | 164              |
| Coding genes    | 1249             | 1464             | 1376             | 1569             |
| Longest contig (kbp) | 34.4             | 59.2             | 32.5             | 26.1             |

*a*Estimated based on the average of three completeness estimation methods: (1) tRNA aaRS complement; (2) 40 conserved universal prokaryotic housekeeping genes (Wu et al., 2013); (3) 178 conserved universal bacterial housekeeping genes (García Martín et al., 2009).
The cumulative sequence and contig coverage plots, coupled with genome coverage of ~2–3.5x and robustness of Sanger sequencing methodologies indicate that these genomes were adequately sampled (Supplemental Figure 1).

**Phylogenetic Analyses**

The “Pyropristinus” T1 assemblies from Octopus and Bechler Springs (T1.1 and T1.2, respectively) were highly-related to one another independent of the comparison method: they exhibited ANIs of 96 ± 1.3% (n = 2410 comparisons), AAIIs of 94 ± 10.1%, (homologous alignment n = 743; Figure 3A), and 16S rRNA gene identities of 98%. The high nucleotide and amino acid identities of the two “Pyropristinus” T1 assemblies indicate that these populations belong to the same genus and likely to the same species, based on empirical species delineations (Konstantinidis and Tiedje, 2005a,b). In contrast, the “Pyropristinus” T1.2 and T2 assemblies were substantially different from one another (AAI = 46.6 ± 12.3%, n = 442; Figure 3A), and both are considerably different than the Calescamantes-like population from Octopus Spring (average AAI ~42%; Figure 3B). While taxonomic rank delineations using AAI do not follow discrete cutoffs, an AAI of only 47% between T1 and T2 is consistent with phylum- or class-level differentiation (Konstantinidis and Tiedje, 2005b). Further, 16S rRNA genes from T1.2 and T2 only shared 84% similarity, which is consistent with empirical phylum- or class-level delineations (Yarza et al., 2014). Until more closely related genomic or isolate representatives are obtained to resolve the taxonomic differences between the T1 and T2-inclusive groups, we tentatively assign both phylotypes to the “Pyropristinus” division that we propose here. The Calescamantes population from OS (Calescamantes-OS) was more closely related to the recently described candidate species *Ca.* Calescibacterium nevadense from Great Boiling Spring, Nevada (78.0 ± 18.1% mean AAI, n = 1053; Figure 3B). The Calescamantes-OS and *Ca.* C. nevadense are different enough to suggest that these two populations represent different genera or even families within the proposed Calescamantes division (Rinke et al., 2013).

Phylogenetic analyses using concatenations of 18 universal or bacterial-specific single-copy genes common to all three phylotypes (comprising 8928 informative amino acid positions) suggest that the “Pyropristinus” T1 and T2 lineages belong to a deeply-branching phylum, distinct from all currently characterized Bacteria (Figure 4). Analysis of only ribosomal proteins common to all three phylotypes (*rplE, rplL, rplO, rpsK, rpsM*; 788 informative amino acid alignment positions) confirmed that the “Pyropristinus” lineage is a well-supported monophyletic group, and is phylogenetically basal relative to the rest of Bacteria (Supplemental Figure 2). The phylogenomic analysis also confirmed that the “streamer” community from Octopus Spring contained a population related to the recently proposed Calescamantes phylum (formerly EM19; Rinke et al., 2013; Hedlund et al., 2014). Phylogenetic analysis without archaeal outgroups confirmed that the “Pyropristinus” and Calescamantes are highly supported groups separate from the Aquificae/Thermodesulfobacteria clade, and which together are distinct from all other Bacteria (Supplemental Figure 3).

Phylogenetic analyses of these organisms using long-fragment 16S rRNA gene sequences also showed that both “Pyropristinus” T1 and T2 populations, along with other uncultured clones, form a deep-branching group near the Thermotogae (Figure 5). The “Pyropristinus” T1 population is closely-related (98% 16S rRNA gene identity) to the uncharacterized EM3 bacterium originally discovered in Octopus Spring (Reysenbach et al., 1994). Partial genome sequence for this “Pyropristinus” lineage was recovered from a single-cell from Great Boiling Spring, NV (only 14% estimated completeness by tRNA synthetase complement, IMG taxon ID: 2264867090; Rinke et al., 2013), but was not sufficiently complete for phylogenomic comparisons (e.g., Figure 4). The T1 and T2 lineages belonged to separate 16S rRNA gene clades, which is consistent with results from the phylogenomic
comparisons. “Pyropristinus” T1 and T2 formed a cohesive group with other uncultured organisms from the same and/or similar types of hydrothermal systems (mean 16S rRNA gene distance within the group = 16%).

The “Pyropristinus” and Calescamantes populations lacked conserved signature indels (CSIs) typical of Thermotogae and Aquificae genomes, which is consistent with all other phylogenetic analysis showing that they are clearly separate from either of these characterized phyla (Supplementary Table 3). T1 shared two CSIs with the Aquificae, but neither were present in the “Pyropristinus” T2 or the Calescamantes-OS assemblies. However, the highly supported monophyletic grouping of the T1, T2, and Calescamantes-OS, which together were separate from the Aquificae in phylogenetic analyses, suggests that neither the “Pyropristinus” or Calescamantes groups are members of the Aquificae. Moreover, a comparison of homologous gene AAI of the “Pyropristinus” T1, T2, and Calescamantes-OS (T3) to several members of the Thermotogae (average AAI: T1/T2: 40.2%, T3: 39.8%, n = 5), Aquificae (T1/T2: 42.4%, T3: 42.6%, n = 4), and Thermodesulfobacteria (T1/T2: 42.0%, T3: 42.5%, n = 3) confirmed that the “Pyropristinus” lineages did not belong to the Thermotogae, the Aquificae, or the Thermodesulfobacteria, as was indicated in the phylogenetic analysis.

The relationship of the “Pyropristinus” and Calescamantes lineages to the recently described, deep-branching bacterium Ca. Acetothermum autotrophicum (Takami et al., 2012) was also attempted. However, due to a lack of universal housekeeping marker genes in the available sequence for Ca. A. autotrophicum” (only three universal markers were shared among Ca. A. autotrophicum: IMG taxon ID: 2540341180, T1, and T2), consistent and well-supported placement of Ca. A. autotrophicum relative to the “Pyropristinus,” Calescamantes, Thermotogae, and Aquificae lineages could not be adequately assessed. A more robust set of universal marker genes from additional “Acetothermia” genome references will be necessary to confidently confirm the phylogenetic placement of “Acetothermia”-like populations.

**Metabolic Reconstruction and Potential Community Interactions**

Metabolic reconstruction of the “Pyropristinus” T1 and T2 populations showed that these organisms shared nearly all major biochemical attributes, despite their substantial phylogenetic dissimilarity. Statistical analysis of the COG distributions from the “Pyropristinus” (T1 and T2) and Calescamantes populations indicated that the functional content of the T2 assembly was highly similar to the two T1 assemblies, and that the “Pyropristinus” assemblies were distinct from the Aquificae, Thermodesulfobacteria, and Thermotogae (Figure 6). The Calescamantes-OS population was also functionally distinct from the “Pyropristinus” assemblies, and was clearly separate from the Ca. C. nevadense assembly from Great Boiling Spring, Nevada.

**Central Carbon Metabolism**

No evidence for inorganic C fixation (Fuchs, 2011) was found in either the “Pyropristinus” (T1 and T2) or Calescamantes-OS
populations, which suggests that these organisms are heterotrophic (Figure 7). The lack of CO₂ fixation pathways in the Calescamantes-OS is consistent with analysis of the related Ca. C. nevadense (Hedlund et al., 2014). The metabolism of polysaccharides was indicated in the “Pyropristinus” (T1, T2) and Calescamantes lineages by the presence of β-glucosidases and α-amylases, as well as other important protein-coding genes in polysaccharide degradation (cellulase in T1; α-glucosidase and starch synthase in Calescamantes). An oligosaccharide transporter present in T1 also suggests that they may utilize exogenous saccharides produced by autotrophic streamer community members, such as Thermocrinis spp. (Aquificales), and/or Aigarchaeota that are also present in these communities (Takacs-Vesbach et al., 2013; Beam et al., 2016). All genes necessary for Embden-Meyerhoff glycolysis were present in “Pyropristinus” T1 (Figure 7), and most were also present in the Calescamantes population (and Ca. C. nevadense) indicating the potential to oxidize glucose. An archaeal-like fructose 1,6-bisphosphatase (fbp) also indicated that gluconeogenesis may occur via a bacterial variant of the bifunctional enzyme that is conserved in Archaea and early-branching bacterial lineages such as the Aquificae (Say and Fuchs, 2010). A nearly complete oxidative TCA cycle was also present in T1 (exclusive of idh) and both Calescamantes populations (Calescamantes-OS and Ca. C. nevadense). Both the “Pyropristinus” and Calescamantes groups contained protein-coding genes involved in the oxidation of fatty acids to acetyl-CoA (β-oxidation pathway; Figure 7). Moreover, long-chain fatty acid transporters present in “Pyropristinus” assemblies may indicate heterotrophic dependence on fatty acids from other streamer community members. The “Pyropristinus” T1 and T2 populations did not contain any evidence of anaerobic fermentation, such as alcohol dehydrogenases, acetate kinases, formate dehydrogenases, and/or [FeFe] or [NiFe] hydrogenases. However, alcohol dehydrogenases were present in the Calescamantes group (both the OS and Ca. C. nevadense assemblies), which suggests possible fermentation in these phylotypes. The “Pyropristinus” and Calescamantes populations contained several amino acid/peptide transporters, peptidases and proteases, which further suggests the ability to import oligopeptides and/or amino acids that may be present in the streamer microenvironment for heterotrophic metabolism. The Thermocrinis spp., that are abundant in both Octopus and Bechler springs are capable of autotrophy via the reverse TCA carbon fixation pathway (Takacs-Vesbach et al., 2013) and are thought to be the primary producers in streamer filament communities (Takacs-Vesbach et al., 2013). Accordingly, the other co-occurring streamer community
members, including a *Pyrobaculum* sp. and *Ca.* Caldithenuis aerorheunensis, are predominantly, or strictly heterotrophic (Jay et al., 2015; Beam et al., 2016). The presence of heterotrophic “Pyropristinus” and Calescamantes-like organisms in these streamer communities provides further evidence for the hypothesis that the dominant *Thermocrinis* spp. support a diversity of co-occurring heterotrophic streamer community members. Moreover, C isotope studies have shown a mixture of both autotrophy and heterotrophy in streamer communities (including OS spring), and that these community members are capable of responding to transient organic carbon pulses (Schubotz et al., 2013; Jennings, 2015; Urschel et al., 2015). Thus, the “Pyropristinus” and Calescamantes-like organisms present in the streamer communities are likely utilizing both endogenous and exogenous organic carbon sources for heterotrophic metabolism.

**Energy Conservation**

Nearly complete respiratory complexes including subunit I heme Cu oxidases were recovered in the “Pyropristinus” T1 and T2 populations as well as the Calescamantes representatives, which indicate that these organisms likely utilize oxygen for respiration and conduct oxidative phosphorylation (Figure 7). The “Pyropristinus” T1 lineage and Calescamantes (OS/ *Ca.* C. nevadense) assemblies contained nearly complete NADH:quinone oxidoreductase (*nuo*) complexes necessary for NADH-mediated oxidative phosphorylation (Figure 7), but differed significantly in key energy conservation mechanisms. The “Pyropristinus” (T1/T2) assemblies contain archaeal V (vacuolar)-type ATPases, while the Calescamantes assembly contains a nearly complete F~0~F~1~ F-type ATPase. Only a small number of bacteria contain archaeal V-type ATPases; the F-type ATPase is ubiquitous and phylogenetically conserved among *Bacteria*, and is thought to be the ancestral bacterial ATPase (Mulkidjanian et al., 2007). The *Thermotogae* variously contain V-type or F-type ATPases (Nelson et al., 1999; Iida et al., 2002; Nesbo et al., 2002), whereas the *Aquificae* contain F-type ATPases (Koumandou and Kossida, 2014). The recently described deep-branching bacterium, *Ca.* Acetothermum autotrophicum, also contains an archaeal V-type ATPase (Takami et al., 2012). A BLAST search of an ATPase subunit I protein (614 aa) of the T1.2 assembly against available genomes in the IMG database showed limited homology (<33% aa id) to 11 genomes largely within the Synergistetes phylum, and to a lesser extent, the Firmicutes, Deltaproteobacteria and Actinobacteria. A BLAST search of an ATPase subunit I fragment present in the T2.1 assembly (233 bp; 40% homology to T1.2), showed similarly low homology to the above bacteria in addition to methanogenic archaea, Thermoplasmatales, and Archaeoglobi (<35% aa id). The disparity in archaeal-like and bacterial-like ATPase complexes between the “Pyropristinus,” *Thermotogae*, and *Ca.* A. autotrophicum lineages relative to the Calescamantes, *Aquificae*, and other bacteria suggests a major divergence in energy conserving mechanisms among these lineages, and warrants further investigation.

Key genes involved in NH~3~ oxidation (*amo*), sulfur oxidation (*sqr, hdr, tgo*), sulfur/sulfate reduction (*psr, dsr*), H~2~ oxidation (*hyn*), methanotrophy (*pmo*), arsenate/arsenite metabolism (*arr, aox*), and NO~3~ reduction (*nar, nap*) were not present in either of the "Pyropristinus" or the Calescamantes-OS populations. A *sqr*-like gene in the *Ca.* C. nevadense assembly suggests that HS~−~ may serve as an electron donor in that phylotype, however, a homologous *sqr* was not found in the Calescamantes-OS assembly. A nitrite reductase (*nirS*') present in *Ca.* C. nevadense with high homology (70%) to the cytochrome *cd* nitrite reductase from *Hydrogenobacter thermophilus* TK-6 (*Aquificales*; Suzuki et al., 2006) along with nosZ nitrous oxide reductase genes in the OS and *Ca.* C. nevadense assemblies suggests the potential for dissimilatory nitrite reduction in these organisms. However, as is common in the assembly of *de novo* genomes from environmental communities, the assemblies were not entirely complete (discussed above) and further analyses are needed to confirm the absence of genes necessary for lithotrophy or anaerobic respiration.

**Secondary Metabolites**

Several differences among the “Pyropristinus” and Calescamantes populations were also noted in pathways used for the synthesis of secondary metabolites. For example, vitamin B~12~ is a necessary cofactor for methylmalonyl-CoA mutase (*mcm*), which is present in “Pyropristinus” and Calescamantes, and is important in the degradation of amino acids and fatty acids into succinyl-CoA (Martens et al., 2002). “Pyropristinus” populations lacked all genes necessary for the synthesis of cobalamin (Vitamin B~12~), whereas the Calescamantes-OS populations contained only 33% of the ~30 genes necessary for *de novo* synthesis (similar to *Ca.* C. nevadense, which contained ~53% of
These genes). “Pyropristinus” (T1 and T2) populations contained genes coding for outer-membrane cobalamin receptor proteins and a permease involved in cobalamin transport, which suggests that they import this cofactor from the environment. The presence of biotin synthase (bioB) as well as bioADF suggest that the Calescamantes populations are capable of synthesizing biotin; conversely, the presence of only one biotin synthesis gene (bioH) from the “Pyropristinus” lineages suggests biotin auxotrophy.

All of the assemblies contain acetyl-CoA carboxylases, which require biotin for the synthesis of malonyl-CoA from acetyl-CoA in fatty acid synthesis (Streit and Entcheva, 2003), and further supports a divergence in secondary metabolite acquisition between “Pyropristinus” and the Calescamantes.

The “Pyropristinus” assemblies lacked all genes necessary for flagellar synthesis, whereas the Calescamantes populations contained numerous flagellar biosynthesis genes including flhA, flIM, flN, flIE, and flgC. The Ca. C. nevadense genome contained many of the flagellar biosynthesis genes not observed in the Calescamantes-OS population, and suggests that they are both capable of flagellar-mediated motility. Chemotaxis genes cheY and cheD were present in the Calescamantes-OS population, whereas “Pyropristinus” T1 contained cheB, cheY, and cheC. Both the Calescamantes and “Pyropristinus” populations are likely gram negative based on the presence of the essential outer-membrane protein assembly gene yfIO in Calescamantes, the yaeT outer-membrane assembly gene in “Pyropristinus,” and
several other outer-membrane associated proteins (Bos et al., 2007; Sutcliffe, 2011) in both candidate phyla.

**Ecological Distribution**

Previously compiled datasets of 16S rRNA gene diversity in YNP and public 16S rRNA gene databases were queried for the presence of “Pyropristinus” Types 1 and 2, and Calescamantes-OS populations. The temperature and pH range observed for thermal springs within the entire YNP ecosystem (Pearson’s $r = 0.66$, $P < 0.05$; Supplemental Figure 4), which indicates that this dataset was appropriate for inferring the presence or absence of these three populations with respect to temperature and pH within YNP. The observed temperature and pH ranges for “Pyropristinus” (T1 and T2) and the Calescamantes-OS phylotypes were not significantly different from one another ($P \geq 0.05$), which suggests that they all occupy similar physicochemical niches. These results are also consistent with earlier observations of closely related EM3 and EM19-like populations in high-temperature, circumneutral communities dominated by *Thermocrinis* spp. (Aquificae) (Reysenbach et al., 1994; Blank et al., 2002; Meyer-Dombard et al., 2011).

**CONCLUSIONS**

Results from phylogenetic analyses showed that two of the new phylotypes characterized here (Type 1 and Type 2) represent different groups of a phylum distinct from other characterized bacterial phyla, and for which we propose the candidate genera epithets “Candidatus Caldipriscus” sp. T1” (Cal’di.prisc.us. L. masc. adj. cal’dus, hot; L. masc. n. priscus, ancient or primitive; ancient thermophile) and “Candidatus Thermoproactor” sp. T2” (Ther.mo.pro.auct’or. Gr. fem. n. therme, heat; L. masc. n. pro.auctor, ancestor/founder; thermophilic ancestor), respectively. Further, on the basis of phylogenetic evidence, we propose the candidate phylum-level name “Pyropristinus” (Pyr.o’pris.tin.us. Gr. neutr. n. pyr, fire; L. masc. adj. pristinus, former/early; early thermophile) to include the Calescamantes, *Ca*. Thermoproactor, and other closely related uncultured organisms, inclusive of the formerly identified EM3. The “Pyropristinus” and Calescamantes populations described here appear to use reduced sources of organic C to respire aerobically, and likely rely on C sources from other autotrophic members of the “streamer” communities. The consistency with which the “Pyropristinus” and Calescamantes lineages co-occur with *Thermocrinis* spp. (Aquificae) in streamer environments suggests that these early branching bacteria may have co-evolved in circumneutral high-temperature environments. Further, metabolic reconstruction suggests these organisms play an important role in C cycling of these high temperature ecosystems. Differences in energy conservation mechanisms between the “Pyropristinus” and Calescamantes lineages (e.g., ATPases, potential to respire anaerobically) suggests that they occupy different microenvironments across oxygen gradients. Importantly, these newly-described phylotypes provide increased resolution of the metabolic attributes associated with deep-branching thermophilic bacterial lineages.

**AUTHOR CONTRIBUTIONS**

ZJ, WI, KM, CT, and DR either participated in sample acquisition, DNA extraction, and/or metagenome sequence analysis. DC, ZJ, WI, DR, RJ, and CT analyzed and interpreted the assembled sequence. DC, WI, ZJ, RJ, and CT contributed to manuscript preparation; DC, WI, and CT wrote and reviewed the manuscript.
ACKNOWLEDGMENTS

We thank Christie Hendrix and Stacey Gunther (Center for Resources, Yellowstone National Park) for research permitting and field assistance. Authors from MSU appreciate support from the Department of Energy (DOE)-Joint Genome Institute Community Sequencing Program (CSP 787081), the DOE-Pacific Northwest National Laboratory (Foundational Science Focus Area) (MSU subcontract 112443) and the NSF IGERT Program for support to ZJ and RJ (DGE 065436). The work conducted by the Joint Genome Institute (DE-AC02-05CH11231) and the Pacific Northwest National Laboratory (Foundational Scientific Focus Area) is supported by the Genomic Science Program, Office of Biological and Environmental Research, US DOE. The UNM authors appreciate support from an NSF Biotic Surveys and Inventories grant (02-06773), the New Mexico Space Grant Consortium, the Louis Stokes Alliance for Minority Participation grant (NSF HRD 0832947), a UNM Office of Graduate Studies Graduate Student Success Scholarship, a UNM Office of Graduate Studies Research Project and Travel Grant and the UNM Research Allocation Committee.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.00304

REFERENCES

Baker, B. J., Comolli, L. R., Dick, G. J., Hauser, L. J., Hyatt, D., Dill, B. D., et al. (2010). Enigmatic, ultrasmall, uncultivated Archaea. Proc. Natl. Acad. Sci. U.S.A. 107, 8806–8811. doi: 10.1073/pnas.0914701
Baron, S., Franchi, M., Gallori, E., and Di Giulo, M. (2007). The first lines of divergence in the Bacteria domain were the hyperthermophilic organisms, the thermogasales and the aquaicales, and not the mesophilic planctomycales. Biosystems 87, 13–19. doi: 10.1016/j.biosystems.2006.02.011
Barns, S. M., Fundyga, R. E., Jeffrey, M. W., and Pace, N. R. (1994). Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proc. Natl. Acad. Sci. U.S.A. 91, 1609–1613. doi: 10.1073/pnas.91.3.1609
Baross, J. A., and Hoffman, S. E. (1985). Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. Origins Life Evol. 15, 327–345. doi: 10.1007/BF01808177
Beam, J. P., Jay, Z. J., Schmid, M. C., Rusch, D. B., Romine, M. F., Jennings, R. D. M., et al. (2016). Ecophysiology of an uncultivated lineage of Aigarchaeota from anoxic, hot spring filamentous ‘streamer’ community. ISME J. 10, 210–224. doi: 10.1038/ismej.2015.83
Blank, C. E., Cady, S. L., and Pace, N. R. (2002). Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park. Appl. Environ. Microbiol. 68, 5123–5135. doi: 10.1128/AEM.68.10.5123–5135.2002
Bos, M. P., Robert, V., and Tommassen, J. (2007). Biogenesis of the gram-negative bacterial outer membrane. Annu. Rev. Microbiol. 61, 191–214. doi: 10.1146/annurev.micro.61.080706.093245
Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., and Knight, R. (2010). PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26, 266–267. doi: 10.1093/bioinformatics/btp536
Castelle, C. J., Wrighton, K. C., Thomas, B. C., Hug, L. A., Brown, C. T., Wilkins, M. J., et al. (2015). Genomic expansion of domain Archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr. Biol. 25, 690–701. doi: 10.1016/j.cub.2015.01.014
Cole, J. R., Wang, Q., Fish, J. A., Bao, B., Sun, Y., and et al. (2014). Ribosomal Database Project: data and tools for high throughputs RNA analysis. Nucleic Acids Res. 42, D633–D642. doi: 10.1093/nar/gkt1244
Darrabia, D., Taboada, G. L., Deaño, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165. doi: 10.1093/bioinformatics/btr088
DeSantis, T. Z., Hugenholtz, P., Keller, K., Brodie, E. L., Larsen, N., Piceno, Y. M., et al. (2006). NAST: a multiple sequence alignment server for comparative analysis of 16S RNA genes. Nucleic Acids Res. 34, W394–W399. doi: 10.1093/nar/gkl244
Di Giulo, M. (2003). The universal ancestor and the ancestor of Bacteria were hyperthermophiles. J. Mol. Evol. 57, 721–730. doi: 10.1007/s00239-003-2522-6
Dodsworth, J. A., Blainey, P. C., Murugapiran, S. K., Swingley, W. D., Ross, C. A., Tringe, S. G., et al. (2013). Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage. Nat. Commun. 4:1854. doi: 10.1038/ncomms2884
Dodsworth, J. A., Hungate, B. A., and Hedlund, B. P. (2011). Ammonia oxidation, denitrification and dissimilatory nitrate reduction to ammonium in two US Great Basin hot springs with abundant ammonia-oxidizing archaea. Environ. Microbiol. 13, 2371–2386. doi: 10.1111/j.1462-2920.2011.02508.x
Ferrera, I., Longhorn, S., Banta, A. B., Liu, Y., Preston, D., and Reysenbach, A. L. (2007). Diversity of 16S rRNA gene, ITS region and acbC gene of the Aquificales. Extremophiles 11, 57–64. doi: 10.1007/s00792-006-0009-2
Fuchs, G. (2011). Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? Annu. Rev. Microbiol. 65, 631–658. doi: 10.1146/annurev-micro-090110-102801
García Martín, H., Ivanova, N., Kunin, V., Warnecke, F., Barry, K. W., McHardy, A. C., et al. (2006). Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat. Biotechnol. 24, 1263–1269. doi: 10.1038/nbt1247
Goris, J., Constanitidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81–91. doi: 10.1099/ijs.0.64483-0
Gupta, R. S., and Bhandhari, V. (2011). Phylogeny and molecular signatures for the phylum Thermotogae and its subgroups. Antonie Van Leeuwenhoek 100, 1–34. doi: 10.1007/s10482-011-9576-z
Gupta, R. S., and Lali, R. (2013). Molecular signatures for the phylum Aquifaxis and its different clades: proposal for division of the phylum Aquificae into the emended order Aquificales, containing the families Aquificaceae and Hydrogenothermaceae, and a new order Desulfurobacteriales ord. nov., containing the family Desulfurobacteriaceae. Antonie Van Leeuwenhoek 104, 349–368. doi: 10.1007/s10482-013-9957-6
Hedlund, B. P., Dodsworth, J. A., Murugapiran, S. K., Rinke, C., and Woyke, T. (2014). Impact of single-cell genomics and metagenomics on the emerging view of extremophile “microbial dark matter”. Extremophiles 18, 865–875. doi: 10.1007/s00792-014-0664-7
Hugenholtz, P., Pitulle, C., Hershberger, K. L., and Pace, N. R. (1998). Novel division level bacterial diversity in a Yellowstone hot spring. J. Bacteriol. 180, 366–376.
Iida, T., Inatomi, K., Kamagata, Y., and Maruyama, T. (2002). F- and V-type ATPases in the hyperthermophilic bacterium Thermotoga maritima. Extremophiles 6, 369–375. doi: 10.1007/s00792-002-0266-7
Inskeep, W. P., Jay, Z. J., Tringe, S. G., Herrgard, M. J., Rusch, D. B., and YNP Metagenome Project Steering Committee et al. (2013). The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. Front. Microbiol. 4:67. doi: 10.3389/fmicb.2013.00067
Inskoek, W. P., Rusch, D. B., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Richardson, T. H., et al. (2010). Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS ONE* 5:e6973. doi: 10.1371/journal.pone.0006973

Jay, Z. J., Bean, J. P., Dohnalkova, A., Lohmayer, R., Bodle, B., Planer-Friedrich, B., et al. (2015). Pyrocubalamus yellowstonensis strain WP30 respires on elemental sulfur and/or arsenate in circumneutral sulfidic geothermal sediments of Yellowstone National Park. *Appl. Environ. Microbiol.* 85, 5907–5916. doi: 10.1128/AEM.01905-15

Jennings, R. (2015). *Inorganic Carbon Fixation and Trophic Interactions in High-Temperature Geothermal Springs of Yellowstone National Park*, WY, USA. Ph.D. Dissertation, Montana State University, Bozeman.

Kantor, R. S., Wrighton, K. C., Handley, K. M., Sharon, I., Hug, L. A., Castelle, C. J., et al. (2015). Evidence for lateral gene transfer between Archaea and bacteria from four candidate phyla. *MBio* 4, e00708-00713. doi: 10.1128/mBio.00708-13

Konstantinidis, K. T., and Tiedje, J. M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Konstantinidis, K. T., and Tiedje, J. M. (2005b). Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187, 6258–6264. doi: 10.1128/JB.187.18.6258-6264.2005

Koumandou, V. L., and Kossida, S. (2014). Evolution of the F0F1 ATP synthase complex in light of the patchy distribution of different bioenergetic pathways across prokaryotes. *PLoS Comput. Biol.* 10:e1003821. doi: 10.1371/journal.pcbi.1003821

Kozubal, M. A., Romine, M., Jennings, R., Jay, Z. J., Tringe, S. G., Rusch, D. B., et al. (2013). Geaorchaetia: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *ISME J.* 7, 622–634. doi: 10.1038/ismej.2012.132

Markowitz, V. M., Chen, I. M., Chu, E., Palaniappan, K., Grechkin, Y., et al. (2012). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res.* 40, D123–D129. doi: 10.1093/nar/gkr975

Martens, J. H., Rarg, H., Warren, M. J., and Jahn, D. (2002). Microbial production of vitamin B12. *Appl. Microbiol. Biotechnol.* 58, 275–285. doi: 10.1007/s00253-001-0902-7

Meyer-Dombard, D. R., Handley, K. M., Sharon, I., Hug, L. A., Castelle, C. J., et al. (2013). Small genomes and sparse metabolisms of sediment-associated Bacteria from four candidate phyla. *Science* 340, 275–279. doi: 10.1126/science.1233705

Meyer-Dombard, D. R., Handley, K. M., Sharon, I., Hug, L. A., Planer-Friedrich, B., et al. (2013). Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three Aquificales lineages. *Front. Microbiol.* 4:84. doi: 10.3389/fmicb.2013.00084

Mulkidjanian, A. Y., Makarova, K. S., Galperin, M. Y., and Koonin, E. V. (2007). *Systematic identification of gene families based taxonomy for prokaryotes.* Nat. Genet. 39, 3204–3223. doi: 10.1038/nrg1228

Posada, D., and Crandall, K. A. (1998). *MODELTEST* testing the model of DNA substitution. *Bioinformatics* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817

Price, P. B. (2009). Microbial genesis, life and death in glacial ice. *Can. J. Microbiol.* 55, 1–11. doi: 10.1139/W08-117

R Core Team (2014). *R: A language and Environment for Statistical Computing*, v. 3.1.0. *R Foundation for Statistical Computing*. Vienna. Available online at: http://www.R-project.org/

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi: 10.1038/nature12352

Say, R. F., and Fuchs, G. (2010). Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconoeogenic enzyme. *Nature* 464, 1077–1081. doi: 10.1038/nature08884

Schubotz, F., Meyer-Dombard, D. R., Bradley, A. S., Fredricks, H. F., Hinrichs, K.-U., Shock, E. L., et al. (2013). Spatial and temporal variability of biomarkers and microbial diversity reveal metabolic and community flexibility in Streamer Biofilm Communities in the Lower Geyser Basin, Yellowstone National Park. *Geobiology* 11, 54–569. doi: 10.1111/gbi.12051

Siers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

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

Stetter, K. O. (2006). Hyperthermophiles in the history of life. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 1837–1842; discussion: 1842–1833. doi: 10.1128/AEM.00511-15

Takai, K., and Horikoshi, K. (1999). Genetic diversity of Archaea in deep-sea hydrothermal vent environments. *Genetics* 152, 1285–1297.

Takami, H., Noguchi, H., Takagi, Y., Uchiyama, I., Toyoda, A., Nishi, S., et al. (2012). A deeply branching thermophilic bacterium with an ancient acetyl-CoA pathway dominates a subsurface ecosystem. *PLoS ONE* 7:e30559. doi: 10.1371/journal.pone.0030559

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/ms3197

Urschel, M. R., Kubo, M. D., Hoehler, T. M., Peters, J. W., and Boyd, E. S. (2015). Carbon source preference in chemosynthetic hot spring communities. *Appl. Environ. Microbiol.* 81, 3834–3847. doi: 10.1128/AEM.00511-15

Wrighton, K. C., Castelle, C. J., Wilkins, M. J., Hug, L. A., Sharon, I., Thomas, B. C., et al. (2013). Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal gradients in hydrothermal sediments at the Yonaguni Knoll IV hydrothermal field in the Southern Okinawa Trough. *Appl. Environ. Microbiol.* 76, 1198–1211. doi: 10.1128/AEM.00924-09

Wu, D., Jospin, G., and Eisen, J. A. (2013). Systematic identification of gene families for use as “Markers” for phylogenetic and phylogeny-driven ecological studies of Bacteria and Archaea and their major subgroups. *PLoS ONE* 8:e77033. doi: 10.1371/journal.pone.0077033

Colman et al.
Wu, M., Chatterji, S., and Eisen, J. A. (2012). Accounting for alignment uncertainty in phylogenomics. *PLoS ONE* 7:e30288. doi: 10.1371/journal.pone.0030288

Yarza, P., Yilmaz, P., Pruesse, E., Glockner, F. O., Ludwig, W., Schleifer, K. H., et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 12, 635–645. doi: 10.1038/nrmicro3330

Zhaxybayeva, O., Swithers, K. S., Lapierre, P., Fournier, G. P., Bickhart, D. M., DeBoy, R. T., et al. (2009). On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5865–5870. doi: 10.1073/pnas.0901260106

**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 © 2016 Colman, Jay, Inskeep, Jennings, Maas, Rusch and Takacs-Vesbach. 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.