Uncultivated thermophiles: current status and spotlight on ‘Aigarchaeota’
Brian P Hedlund1,2, Senthil K Murugapiran1, Timothy W Alba1, Asaf Levy3, Jeremy A Dodsworth4, Gisele B Goertz1, Natalia Ivanova3 and Tanja Woyke3

Meta-analysis of cultivation-independent sequence data shows that geothermal systems host an abundance of novel organisms, representing a vast unexplored phylogenetic and functional diversity among yet-uncultivated thermophiles. A number of thermophiles have recently been interrogated using metagenomic and/or single-cell genomic approaches, including members of taxonomic groups that inhabit both thermal and non-thermal environments, such as ‘Acetothermia’ (OP1) and ‘Atribacteria’ (OP9/JS1), as well as the exclusively thermophilic lineages ‘Korarchaeota’, ‘Calecamanetes’ (EM19), ‘Fervidibacteria’ (OctSpA1-106), and ‘Aigarchaeota’ (HWCG-I). The ‘Aigarchaeota’, a sister lineage to the Thaumarchaeota, likely includes both hyperthermophiles and moderate thermophiles. They inhabit terrestrial, marine, and subsurface thermal environments and comprise at least nine genus-level lineages, several of which are globally distributed.

Addresses
1 School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA
2 Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA
3 DOE Joint Genome Institute, Walnut Creek, CA 94598, USA
4 Department of Biology, California State University, San Bernardino, CA 92407, USA

Corresponding author: Hedlund, Brian P (brian.hedlund@unlv.edu)

Current Opinion in Microbiology 2015, 25:136–145
This review comes from a themed issue on Extremophiles
Edited by Haruyuki Atomi and Elizaveta Bonch-Osoiskaya

http://dx.doi.org/10.1016/j.mib.2015.06.008
1369-5274/© 2015 Elsevier Ltd. All rights reserved.

Introduction
Geothermal systems hosting thermophiles (optimum growth 45–80 °C) and hyperthermophiles (>80 °C) were some of the first environments to be explored by cultivation-independent approaches using 16S rRNAs and 16S rRNA gene amplicons [1,2]. These studies and many others revealed a vast diversity that filled in and expanded the tree of life, with yet-uncultivated organisms spanning the phylum to species level. Cultivation-independent investigations of Obsidian Pool in Yellowstone National Park were particularly ground-breaking, revealing deeply branching Archaea, including ‘Korarchaeota’ [3,4], and twelve novel lineages of Bacteria, provisionally named OP1–OP12 [5]. Many of these were subsequently shown to also inhabit non-thermal environments and have come into focus in a variety of habitats (e.g. [6–8]). Over the last ten years, a few thermophilic members of these groups and others have been isolated and described as axenic cultures (e.g. [9–13]); however, most of the diversity within them remains to be explored. Recent advancements in metagenomics and single-cell genomics have provided access to the genomes of yet-uncultivated thermophiles, enabling insight into their evolution, cell biology, and possible metabolic and ecological functions. Continued genomic exploration of these enigmatic organisms, along with experiments to test their functions in situ and cultivate them in the laboratory, will be an exciting frontier in biology for decades to come.

Extent of yet-uncultivated thermophile diversity
Currently there are over 4.3 million 16S rRNA gene sequences in public databases [14]. Despite exponential increases in the number of entries per year, the accumulation of new operational taxonomic units at all levels is approaching an asymptote, with total bacterial species richness predicted to be ~4 × 105 [15**]. At present, it is difficult to estimate the taxonomic richness of thermophiles and hyperthermophiles, although well-established negative correlations between microbial diversity and temperature [16,17•] suggest it may be a small fraction of the total taxonomic richness. Nevertheless, with very few species of thermophiles and hyperthermophiles formally described (<1000, estimated from ~12 000 total species on the List of Prokaryotic Names with Standing in Nomenclature [18]), the vast majority remains to be carefully studied. More intriguingly, many deep branches of thermophiles also remain poorly represented, or unrepresented by axenic cultures. Estimates of the total number of phyla of Bacteria and Archaea commonly range from 50 to 100 [19,20]; however, a conservative, unbiased approach recently estimated ~1600 bacterial lineages equivalent to the existing bacterial phyla [15**]. As with species, the number of phylum-level lineages containing some or many thermophiles is currently difficult to determine.
Cultivation-independent studies of geothermal systems provide a window to assess the abundance and distribution of yet-uncultivated thermophile diversity (Figure 1; Tables S1 and S2). The most comprehensive view currently available is gained by meta-analysis of 16S rRNA gene amplicon data. Parallel analysis of data from eight regions, comprising 372 samples from terrestrial geothermal systems [16,17,21–27] suggests a high abundance of uncultivated high-level taxa in geothermal systems globally, including phyla (\(\bar{x} = 16.8\%\)), classes (\(\bar{x} = 34.0\%\)), orders (\(\bar{x} = 42.1\%\)), and families (\(\bar{x} = 46.9\%\)). Most regions show a somewhat even distribution of uncultivated taxa among samples; however, an exception is Tengchong, China, in which the Rehai Geothermal Field springs contain a very low apparent abundance of uncultivated organisms and the Ruidian (Dientan) springs contain a very high abundance of uncultivated organisms [24,27]. Hotspots of uncharacterized biodiversity like the Ruidian springs may be obscured in other regions by greater sampling effort or different sampling strategies.

Examining metagenomes from geothermal systems in the Integrated Microbial Genomes-Microbiomes (IMG-M) database [28,29] provides another view of the extent of yet-uncultivated thermophile biodiversity. Analysis of proteins extracted from assemblies for 62 datasets reveals a very broad range of apparent abundance of unexplored genomic diversity, with 10–92% of predicted coding sequences taxonomically unassigned at 60% amino acid identity (\(\bar{x} = 55.1\%, n = 62\)) (Table S2; calculated using UCHIME in IMG-M [29]). This level of amino acid identity is comparable with the family level since bacterial genus boundaries correspond to \(~65\%\) average amino acid identity (AAI) [30]; however, unassigned genes may

Figure 1

Summary of unclassified reads in 16S rRNA gene amplicon datasets. Raw 16S rRNA gene pyrotags or illumina tags from [16,17,21–27] were processed using QiIME [31] and mothur [32], respectively, using default parameters (97% OTU clustering) and identified using the Greengenes taxonomy, accounting for unclassified phyla, classes, orders, and families. \(n\) represents the number of rRNA gene tag datasets per each site. Violin plots depict the range and data model (colored area), 25th and 75th percentile (box), mean (dark bar), and median (light bar). Individual data points are statistical outliers. Data are detailed in Table S1. Newly published data from the Great Basin were submitted to the NCBI under SRA# SRP059341.
also represent elements of the accessory genome residing within known taxa. Although these springs span a wide range of temperature (50–90 °C) and pH (2.1–9.3), no relationship between the percent of taxonomically unassigned sequences and either temperature (r² = 0.07) or pH (r² = 0.02) is evident (Table S2). Lastly, 16S rRNA genes extracted from these datasets represent a source of information, although read depth, which is necessary for abundance estimation, is only available for a subset of metagenomes. This approach also shows a wide variation in the abundance of yet-uncultivated microorganisms, with 0–82% of 16S rRNA gene sequences being taxonomically unassigned at the phylum level (x = 32.1%, n = 24) with a higher abundance of unassigned sequences at the class (x = 37.8%), order (x = 40.1%), and family (x = 44.1%) levels (Table S2; assigned using Green genes). Despite many caveats deriving from differential DNA extraction efficiency, primer bias, uneven taxonomic/phylogenetic structure, and differences in phylogenetic resolution among different regions of the 16S rRNA gene (e.g. [15**]), these data suggest that uncultivated taxa are abundant in most geothermal systems.

Progress on thermophile metagenomics and single-cell genomics

A few studies have explicitly targeted yet-uncultivated, deep lineages of thermophiles by using metagenomics or single-cell genomics approaches in geothermal environments. Studies of a thermal spring in a Japanese gold mine (70 °C, pH 5.1) involved screening a fosmid library for 16S rRNA genes of interest, followed by both targeted and random fosmid clone sequencing, leading to a complete genomic assembly for Candidatus Caldarchaeum subterraneum, a member of a deeply branching archaeal group, Hot Water Crenarchaeotal Group I (HWCG I), or Aigarchaeota [33**], and a nearly complete genomic assembly for Candidatus Acetothermum autotrophicum, belonging to candidate bacterial phylum OP1, Acetothermia [34**]. A shotgun metagenome study of iron oxide mats in One Hundred Springs Plain Spring (60–78 °C, pH 3.5) in Norris Geyser Basin, Yellowstone, yielded a near-complete consensus genome from a deeply branching archaeon, named NAG1, or Geoarchaeota [35*]. Another study purified Korarchaeota cells from an enrichment culture (85 °C, pH 6.5; Obsidian Pool) by SDS treatment followed by filtration (0.45 μm), enabling sequencing and assembly of a complete genome for Candidatus Korarchaeum cryptofilum [36**]. In the U.S. Great Basin, one study combined single-cell genome sequencing of 15 single cells sorted and amplified using an optofluidic approach from Little Hot Creek, California (80 °C, pH ~7) and shotgun metagenomic sequencing of a cellulolytic enrichment from Great Boiling Spring, Nevada (77 °C, pH ~7) [37**], yielding two fragmented but nearly complete genome assemblies for members of candidate bacterial phylum OP9 ‘Attribacteria’, provisionally named Candidatus Caldtribacterium californiense and Candidatus Caldtribacterium saccharofermentans’. Samples from Great Basin springs were also part of the largest single-cell genomics study to date, the Genomic Exploration of Bacteria and Archaea-Microbial Dark Matter (GBA-MDM) project, which used fluorescence activated cell sorting (FACS) followed by whole-genome amplification and sequencing [38**]. More than 30 single cell genomes were obtained from natural sediments in Great Boiling Spring, including assemblies ranging from 50 to 95% estimated genome coverage for two lineages, EM19 (‘Calescammentes’; Candidatus Calescibacterium nevadense; 10 single-amplified genomes (SAGs)), and OctSpA1-106 (Fervidibacteria, Candidatus Fervidibacter sacchari; 5 SAGs), along with large assemblies for several unnamed ‘Aigarchaeota’ species (14 SAGs) and smaller datasets for additional deep lineages.

In aggregate, these studies and others (e.g. [39,40]) have significantly expanded the genomic coverage of thermophilic Bacteria and Archaea. The major metabolic features of these organisms seem to be diverse (Figure 2). Based on genome-enabled predictions, the lineages described above potentially encode systems for autotrophy (‘Aceto-thermum’, Caldtribacterium), aerobic respiration (‘Aceto-thermum’, Caldtribacterium, Calescibacterium, Fervidibacterium), and various forms of anaerobic respiration (‘Aceto-thermum’, Caldtribacterium, Calescibacterium, Fervidibacter). Caldtribacterium and Korarchaeum are predicted to be obligate fermenters. Most of these representative organisms are rod-shaped or filamentous, based on either direct observation [36**,37**] or the presence of genes for MreB or crenactin, and all of the bacteria are probably diermer (i.e. Gram negative), based on genes encoding components for outer membrane synthesis and transport [41]. The archaia are presumed to be monoderm. None of these lineages appear to be symbionts, based on genome size and predicted biosynthetic potential, distinguishing these organisms from a multitude of deep lineages of Bacteria and Archaea that have reduced genome sizes (<1.3 Mbp) and predicted symbiotic lifestyles (e.g. [20,38**,42–49]).

It is noteworthy that the provisional taxonomy of these organisms is uncertain on several grounds, including: the requirement in microbial systematics for well-described axenic cultures; current limitations on higher-order taxonomy, which does not cover classifications above class; and disagreements within the scientific community about these unofficial designations and their applications to various lineages (e.g. [50**,51–54]). Some members of the systematics community have recently encouraged discussion to emend Candidatus status to better take advantage of advancements in environmental genomics [55,56]; however, no recent changes have been ratified by the International Committee on the Systematics of Bacteria and Archaea.
Metagenomic datasets provide a means to assess the distribution and abundance of yet-uncultivated taxa for which reference genomes exist. We used a conservative approach of >70% nucleotide identity and >100 bp long hits to recruit reads from 59 metagenomes from terrestrial and marine geothermal systems [28, 57-60] to seven reference genomes [37, 38, 49, 58, 34**]. This analysis roughly targets genus-level recruitment [30] and suggests that members of the ‘Aigarchaeota’ (‘Ca. Caldarchaeum’), ‘Korarchaeota’ (‘Ca. Korarchaeum’), ‘Nanobarchaeota’ (‘Ca. Nanobsidians’), ‘Acetothermum’ (‘Ca. Acetothermum’), ‘Atribacteria’ (‘Ca. Caldtribacterium’), ‘Calescibacteria’ (‘Ca. Calescibacterium’), and ‘Fervidibacteria’ (‘Ca. Fervidibacter’) are present in a majority of geothermal systems (Figure 2; Table S3). ‘Ca. Caldtribacterium’ and ‘Ca. Calescibacterium’ were particularly widespread, being detected in >80% of metagenomes, including metagenomes from all terrestrial systems analyzed, as well as marine hydrothermal systems. Close relatives of ‘Candidatus Nanobsidians stetteni’ were particularly widespread in Yellowstone metagenomes, with >0.1% estimated abundance in 29% of Yellowstone metagenomes, all of which are from acid sulfate systems hosting the probable host, Sulfolobales [49]. Some springs hosted many of these candidate taxa, including Great Boiling Spring and Little Hot Creek (Great Basin), Octopus Spring, OSP Spring, and Beowulf Spring (Yellowstone), as well as Gongxiaoshe and Gumingquan (Tengchong) (Figure 2). These candidate taxa were poorly represented in marine hydrothermal samples; however, this is not surprising since the fragment recruitment strategy was stringent. This high stringency also explains the low recruitment of candidate phyla in springs in which they are known to be abundant and diverse (e.g. [16*]).

**Aigarchaeota**: genomic insights, diversity, and distribution

One deep lineage that has received considerable attention in the literature, but is still poorly studied is the HWCG I group, or ‘Aigarchaeota’. The ‘Aigarchaeota’ was first highlighted as part of an investigation of a mildly
Phylogenetic diversity of 'Aigarchaeota' lineages. 16S rRNA gene sequences with BLASTN identity >80% to 'Ca. Caldiarchaeum subterraneum' were mined from GenBank, along with sequences from single-cell genomes and metagenomes in IMG-M [29]. Potential 'Aigarchaeota' 16S rRNA gene sequences >1200 nucleotides (n = 82; Table S4) were aligned with the 16S rRNA reference alignment in mothur, hard filtered using Lane mask, soft filtered (positions with >50% gaps), and processed in RAxML with 1000 pseudoreplicates [14,32,70]. The resulting phylogenetic tree was delineated into nine genus-level groups (G1-G9) according to >95% nucleotide identity [15*]. Taxon jackknifing was performed by randomly removing 30% of the sequences (10×) and repeating the bootstrapping analysis (Table S6), further supporting the monophyly of all groups. A similar phylogenetic analysis with the Lane mask but without a soft filter yielded similar results.
acidic spring within a Japanese gold mine as a single fosmid clone and later an entire composite genome of ‘Candidatus Caldiarchaeum subterraneum’ [33**]. Analysis of predicted features led the authors to suggest that this organism may couple hydrogen or carbon monoxide oxidation to aerobic or anaerobic respiration using nitrate or nitrite as electron acceptors. However, the hydrogenases in ‘Ca. Caldiarchaeum subterraneum’ belong to Group 3B and Group 4 NiFe hydrogenases [61], characterized members of which regulate redox homeostasis rather than hydrogenotrophy. ‘Ca. Caldiarchaeum subterraneum’ may be autotrophic via the dicarboxylate/4-hydroxybutyrate pathway, but lacks a canonical 4-hydroxybutyryl-CoA dehydratase. Preliminary analysis of single-cell genomes of several ‘Aigarchaeota’ lineages from Great Boiling Spring has yielded additional insights, including potential capacity for heterotrophic utilization of proteins and sugars and, in some cases, anaerobic respiration using sulfite or nitrous oxide [38**]. A more detailed interpretation and comparative analysis of these and other genomes from a diversity of ‘Aigarchaeota’ lineages and thermal environments is necessary to reveal potential core traits and the extent of metabolic plasticity within the ‘Aigarchaeota’.

Phylogenomic and comparative genomic studies have consistently revealed a deep relationship between Thaumarchaeota, ‘Aigarchaeota’, Crenarchaeota, and ‘Korarchaeota’ in the ‘TACK superphylum’ [38**,52,62,63], yet there is some uncertainty about whether ‘Aigarchaeota’ should be considered an independent phylum or a deep branch within the Thaumarchaeota [50**,51,52]. This uncertainty is symptomatic of unrest among microbiologists seeking consensus on guidelines for higher-order taxonomy, particularly Archaea [51,54]. The 16S rRNA gene identity between ‘Aigarchaeota’ and Thaumarchaeota (77–79%) is lower than the median sequence identity delimiting bacterial phyla (83.68% [15**]) and close to the minimum distance between bacterial phyla (77.43% [15**]), but has been reported to be similar to that delimiting euryarchaean orders [50**]. Regardless, it is clear that the cell biology of ‘Aigarchaeota’ shares many features with Thaumarchaeota [50**,64], although their ecology appears to be distinct. The evolution of metabolically specialized Thaumarchaeota from an apparently more flexible ‘Aigarchaeota’ ancestor [33**,38**] may be one of many examples of streamlining in archaeal evolution [63].

The ‘Aigarchaeota’ are phylogenetically diverse and inhabit a wide array of thermal ecosystems, including terrestrial and marine geothermal systems, subsurface aquifers, and mine fracture waters (Figures 3 and 4; Table S4). Clustering and phylogenetic analysis of ‘Aigarchaeota’ 16S rRNA gene sequences from 16S rRNA gene surveys, single-cell genomes, and metagenomes, shows at least nine genus-level groups (<95% 16S rRNA gene identity [15**]), although some are not strongly supported and should be

Figure 4

Distribution of ‘Aigarchaeota’ based on 16S rRNA gene sequences. (a) Geographic distribution of ‘Aigarchaeota’ genus-level groups, including short sequences (<1200 bp) in GenBank that could be confidently mapped to groups (Table S4); open dots are not confidently placed in a genus-level group. (b) Environments in which ‘Aigarchaeota’ groups are found. Colors correspond to genus-level groups in Figure 3. Optimum growth temperature was predicted based on 16S rRNA gene G+C content (Table S4).
reevaluated as additional sequences become available (Figure 3; Tables S5 and S6). All groups are found in terrestrial geothermal systems, including deeply branching ‘Aigarchaeota’, suggesting a continental origin (Figure 4). A few ‘Aigarchaeota’ 16S rRNA gene sequences have been recovered from moderate-temperature or cold-temperature samples; however, the vast majority are from high-temperature environments and all have a 16S rRNA gene G + C content consistent with a thermophilic or hyperthermophilic lifestyle, with calculated optimal growth temperatures of 40.8–101.9 °C (x = 78.4 °C, n = 357) (Table S3) [65]. A general trend toward lower 16S rRNA gene G + C content in newer and longer branches suggests a high-temperature origin, with niche invasion to lower temperature systems. ‘Aigarchaeota’ are predominantly found in habitats with circumneutral pH. The U.S. Great Basin appears to be somewhat of a hotspot for ‘Aigarchaeota’ diversity and abundance, with several ‘Aigarchaeota’ lineages prominent in Great Boiling Spring [167,66], Mud Hot Springs [66], Grass Valley Spring [67], Little Hot Creek [68], and an aquifer at the Nevada National Security Test Site. However, sampling is currently highly biased and the full extent of ‘Aigarchaeota’ diversity is yet to be uncovered.

Outlook
Both taxonomic and genomic diversity are poorly represented by thermophile laboratory cultures. Understanding thermophile diversity and ecology necessitates a polyphasic approach to biodiversity exploration, including microbial cultivation and characterization, cultivation-independent genomic exploration, and in situ activity measurements. Exploration of thermophile biodiversity will remain an exciting and important field for decades to come.

Acknowledgements
This work was supported by NASA Exobiology grant EXO-NNX11AR78G; U.S. National Science Foundation grant OISE 0968421; U.S. Department of Energy (DOE) grant DE-EE-0000716; and the U.S. Department of Energy Joint Genome Institute (CSP-182), a DOE Office of Science User Facility, supported under Contract No. DE-AC02-05CH12111. B.P.H. acknowledges generous support from Greg Fullmer through the UNLV Foundation. T.W.A. was supported by the Richard and Beverly Hermsen Fellowship. G.B.G. was supported by the Robert E. McNair Foundation and NSF DBI-1358896. We acknowledge Steve Quake, Paul Blaney, Jai Kanbar, and Iwijn De Vlaminck for collaboration with single-cell genomics and metagenomics and HaiIang Dong and Brandon Briggs for collaboration with metagenomics. William Inskeep and Jake Beamer provided an ‘Aigarchaeota’ 16S rRNA gene sequence from Hot Creek. Duane Moser and Scott Hamilton-Beach provided 16S rRNA gene sequences from the Nevada National Security Site (NNSS). Scott Miller, Ken Jones, Eric Boyd, Trinity Hamilton, Dan Colman, and Cristina Takacs-Vesbach provided raw 16S rRNA gene pyrotag reads. Fengping Wang, William Brazelton, and Matthew Schrenk provided metagenomic datasets.

Appendix A. Supplementary data
Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.mib.2015.06.008.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


13. Podosokorskaya OA, Kadinokov VV, Gavrilov SN, Mardanov AV, Merkel AY, Karnachuk OV, Ravin NV, Bonch-Osmolovskaya EA, Kublanov IV: Characterization of Melioribacter roseus gen. nov., sp. nov., a novel facultatively anaerobic thermophilic celluloxytic bacterium from the class Ignnavibacteria, and a


Although not focusing on thermophiles and focused only on Bacteria, this paper provides a great perspective on the yet-uncultivated microbial world. The authors use all taxa on the List of Prokaryotic Names with Standing in Nomenclature to calibrate 16S rRNA gene distance to the taxonomic structure and then use the calibration to expand the taxonomic structure to 16S rRNA genes from cultivation-independent studies, generating estimates of biodiversity and an up-to-date view on 16S rRNA gene exploration.


This study focuses on Great Boiling Spring, in the U.S. Great Basin, a hotspot for ‘Aigarchaeota’ diversity and abundance. This spring contains at least four genus-level groups of ‘Aigarchaeota’. In the highest sampling temperature, ‘Aigarchaeota’ comprise >5% of 16S rRNA gene pyrotags. Pyrotag-based estimates of ‘Aigarchaeota’ abundance in this spring are similar to metagenomic estimates (IMG #205556014).


This study examines 16S rRNA gene pyrotag data from ~170 geothermal systems in Canada, New Zealand, United States, and China and shows richness and diversity maximum at ~24°C, with both measurements declining as temperature increases.


This paper summarizes the Yellowstone National Park Metagenome Project, which included twenty sites, roughly divided into phototrophic mat communities, Aquificales-dominate ‘streamer’ communities, and Archaea-dominated communities. Although not focused on major, yet-uncultivated groups, the YNP Metagenome Project is the largest coordinated effort to obtain metagenomes from thermal environments.


This study describes a complete genomic assembly for the founding member of the ‘Aigarchaeota’, ‘Candidatus Caldiarchaeum subterraneum’, from a subterranean gold mine in Japan. The paper describes possible metabolic features, including hydrogenotrophy, autotrophy, aerobic respiration, and nitrate reduction, as well as some eukaryotic features such as a possible ubiquitin protein modification system.


This paper describes a complete genomic assembly for the founding member of the ‘Acotothermia’ (OP1), ‘Candidatus Acetothermus autotrophicum’, from a subterranean gold mine in Japan. This organism is likely to be an acetogen capable of autotrophy via the Wood-Ljungdahl pathway. The paper also speculates on the deep evolutionary origin of H2-dependent acetogenesis.


The ‘Geoaarchaeota’ (NAG1) is described here as a candidate phylum of Archaea, represented by a composite metagenome assembled from One Hundred Springs Plain Spring in Yellowstone; however, this group is monophyletic with the Thermoprotei so the candidate phylum status is in question.


This review describes the landscape of archaeal genomes, including a summary of genes predicted to encode various components of cell division, DNA maintenance, protein ubiquitylation, and transcription systems, including systems in 'Archaea', 'Korarchaeota', and 'Nanoarchaeota'. The paper also presents a cogent discussion of archaean phylogeny and taxonomy from a phylogenetic and genomic perspective.


