Antibiosis helps Aquificae surmount competition by archaeal thermophiles, and crowding by bacterial mesophiles, to dominate the boiling vent-water of a Trans-Himalayan sulfur-borax spring

Nibendu Mondal[1,$], Subhajit Dutta[1], Sumit Chatterjee[1], Jagannath Sarkar[1], Mahamadul Mondal[1], Chayan Roy[2], Ranadhir Chakraborty[3] and Wriddhiman Ghosh[1,*]

Author Addresses:

[1] Department of Biological Sciences, Bose Institute, Kolkata, India
[2] Department of Plant and Environmental Sciences, University of Copenhagen, Copenhagen, Denmark
[3] Department of Biotechnology, University of North Bengal, Siliguri, India

Present Address:

[$] Division of Molecular Biology and Genomics, International Institute of Innovation and Technology, Kolkata, India

* **Correspondence:** Wriddhiman Ghosh (wriman@jcbose.ac.in / Wriman@rediffmail.com)

Keywords: Trans-Himalayan sulfur-borax hot springs, vent-water microbiome, hot spring mesophiles, Proteobacteria, Aquificae, antibiosis

---

Abstract

Trans-Himalayan hot spring waters rich in sulfur, boron, sodium and chlorine (but poor in calcium and silicon) are known based on PCR-amplified 16S rRNA gene sequence data to harbor high diversities of infiltrating bacterial mesophiles. Yet, little is known about the population dynamics, primary productivity, mutual interactions, and thermal adaptations of the microorganisms present in the steaming waters discharged by these geochemically peculiar spring systems. We revealed these aspects of a bacteria-dominated microbiome thriving in the boiling (85°C) fluid vented by a sulfur-borax spring called Lotus Pond, situated at 4436 m above the mean sea-level, in the Puga valley of eastern Ladakh, on the Changthang plateau. Assembly, annotation, and population-binning of >15-GB metagenomic sequence illuminated the numeral predominance of Aquificae (particularly, the genera *Hydrogenobacter* and *Sulfurihydrogenibium*), even as Proteobacteria, and Firmicutes, Cyanobacteria, Chloroflexi and Bacteroidetes, contributed most of the genetic diversity of the microbiome; prevalence and diversity of Deinococcus-Thermus, Thermotogae, Thermodesulfobacteria, Crenarchaeota and Euryarchaeota were far less. Perpetual presence of bacterial mesophiles in Lotus Pond's vent-water was evident from the detection of 51 such genera that had been discovered in previous 16S-amplicon-based investigations of the habitat, but have no member with laboratory growth at >45°C. The metagenome was replete with genes crucial for hydrothermal, and other stress, adaptations, and a majority of them were attributed to phylogenetic relatives of mesophilic bacteria; examples included genes encoding heat shock proteins, molecular chaperones, and chaperonin complexes; proteins controlling/modulating/inhibiting DNA gyrase; universal stress proteins; methionine sulfoxide reductases; fatty acid desaturases; different toxin–antitoxin systems; enzymes protecting against oxidative damage; proteins

conferring flagellar structure/function, chemotaxis, cell adhesion/aggregation, biofilm formation, and quorum sensing. The Lotus Pond Aquificae not only dominated the microbiome numerically but also acted potentially as the main primary producers of the ecosystem, with chemolithotrophic sulfur oxidation (Sox) being the fundamental bioenergetic mechanism, and reductive tricarboxylic acid (rTCA) cycle the predominant carbon fixation pathway. Furthermore, the Aquificae members present *in situ* contained copious genes for virulence functions, and biosynthesis of the antibiotics monobactams, streptomycin, prodigiosin, acarbose and validamycin, novobiocin, carbapenems, and vancomycin (a few genes were there for phenazine, and neomycin/kanamycin/gentamicin as well). The Lotus Pond Aquificae also contained copious genes for resistance against beta-lactam antibiotics (penicillins, cephalosporins, monobactams, and carbapenems) cationic antimicrobial peptides, vancomycin, aminoglycosides (streptomycin and neomycin/kanamycin/gentamicin), and fluoroquinolones; they also had several genes for multidrug resistance efflux pumps. Among the other thermophiles present, only Thermodesulfobacteria and Deinococcus-Thermus had genes for all these functions, albeit at much lesser frequencies; the Archaea and Thermotogae were either devoid of, or had very few genes for, the aforesaid functions. The Lotus Pond mesophiles, led by Alphaproteobacteria and Gammaproteobacteria, possessed a slightly wider variety, and higher abundance, of antibiotic synthesis/resistance genes, compared with the Aquificae, besides having a more uniform gene count across the antibiotics synthesized/resisted. We hypothesize that antibiosis helped Aquificae avoid niche overlap with other thermophiles/hyperthermophiles, especially the Archaea, besides exacerbating the bioenergetic costs of thermal endurance for the mesophiles infiltrating the system, thereby keeping the potential proliferation of these intruders at bay.

Introduction

Hydrothermal environments, considered to be the nurseries of early life on Earth [1-3], are characterized by the entropic disordering of biomacromolecular systems, which in turn restricts their habitability [4, 5]. Thermophilic and hyperthermophilic microorganisms [6] are the habitual natives of hydrothermal vents, and for most taxa abundant at the 30-40°C sites of these ecosystems, habitability is progressively constrained in the 50°C to 100°C trajectory [5, 7-10]. Conversely, as the thermodynamic/bioenergetic cost of cell system maintenance get relaxed down the temperature gradient, habitability, and therefore *in situ* diversity, increases for most microbial taxa [11, 12].

Idiosyncratic to the above community structure axiomatic for hydrothermal ecosystems, a geochemically unusual category of Trans-Himalayan hot springs (poor in total dissolved solids including silicon, but rich in boron, sodium and diverse sulfur species), located on the either side of the tectonically active Indus Tsangpo Suture Zone (ITSZ, the collision intersection between the Asian and Indian continental crusts engaged in Himalayan orogeny) in eastern Ladakh, India [13-16], have been reported to feature extra-ordinarily high habitability [5, 10, 17, 18]. Concurrent to such reports of exceedingly bio-diverse microbiomes from the springs of the Puga and Chumathang geothermal areas situated on the Changthang plateau in the south and north of the ITSZ respectively, studies of microbial diversity within the vent-waters, and vent-

adjacent niches, of a number of hot springs located across discrete Himalayan and Trans-Himalayan geothermal regions have corroborated that hydrothermal environments were not “Thermophiles Only Territories” [19-26].

As for the well-studied Puga hydrothermal system, PCR-amplified 16S rRNA gene sequence-based investigations of microbial communities within the boiling vent-waters of a number of springs have revealed the presence of several such genera whose members are incapable of laboratory-growth at >45°C [10, 17, 18]. Diurnal fluctuations in the biodiversity and chemistry of the vent-waters, in conjunction with circumstantial geological evidences, indicated that the copious tectonic faults of this ITSZ-adjoining area serve as potential conduits for the infiltration of meteoritic water, and therewith mesophilic microorganisms, to the shallow geothermal reservoir [10]. Furthermore, shotgun metagenome sequencing-based investigation of microbial mat communities along the spring-water transit/dissipation channels of a hydrothermal microbialite of Puga has identified several genus-level populations to occupy such sites within the hydrothermal gradients which have temperatures tens of degree-Celsius above the *in vitro* growth-limits of those taxa [5, 27]. While the overall geomicrobiology of the Puga hot springs indicated the system’s peculiar chemistry (mineralogy) to be the key facilitator of the high habitability of the environment [5, 18], physiological attributes of a vent-water isolate identified dissolved solutes such as lithium and boron, characteristic to the Puga hydrothermal system [15, 16], as the key enablers of high temperature endurance by the mesophilic bacteria present *in situ* [28].

Despite a considerable progress in our basic awareness about the peculiar geobiology of Trans-Himalayan sulfur-borax hot springs, we still do not have any quantitative idea about the structural and functional dynamics of the vent-water microbiomes of these astrobiologically significant hydrothermal systems [16, 29, 30]. For instance, we have a fair measure of the microbial diversity in terms of the variety of taxa present, but we are completely uninformed about the relative abundance of the individual taxa constituting the microbiome, the thermal adaptations of the native microorganisms (especially the phylogenetic relatives of mesophilic bacteria present), and their mutual interactions. We also know nothing about the pathways of primary productivity that sustain the ecosystem in these boiling spring waters. To address these shortcomings in our knowledge on the ecology of Trans-Himalayan sulfur–borax spring systems, this study comprehensively reveals the vent-water microbiome of the Lotus Pond hot spring, which is one of the most conspicuous hydrothermal manifestations within the Puga valley [10, 16, 17]. While the bacterial diversity of Lotus Pond’s vent-water has been investigated previously by means of 16S amplicon sequencing [10, 17], the present exploration, for the first time, unravelled the qualitative as well as quantitative architecture of the microbiome via whole metagenome shotgun sequencing, and fluorescence microscopy. Besides elucidating the dominance and equitability of taxa within the microbiome, the >15 giga base metagenomic sequence data generated was analyzed to illuminate the aforesaid aspects of microbiome functioning.

Materials and Methods

Study site and sampling

Lotus Pond, located at 33°13’46.3”N and 78°21’19.7”E, is one of the most prominent hot springs or geysers within the Puga valley (Fig. 1A), which in turn is one of the most vigorous geothermal areas of the Indian Trans-Himalayas; notably Puga Valley is situated on the Changthang plateau of eastern Ladakh, at an altitude of 4436 m, where the boiling point of water remains at around 85°C [10, 16, 31]. The venting point (Fig. 1B) is embedded in a hot water pool that has been formed by its own discharges and lies within the old crater (Fig 1A) formed from the same hydrothermal eruption that gave rise to the geyser itself. From one side of the hot water pool, the spring-water flows over thin unstable terraces of geothermal sinters, into the Rulang rivulet, which meanders through the valley of Puga. On the other flank of the hot water pool, a wall of travertine, sulfur, and boron mineral deposits [10, 16] guard the crater and debars the spring-water from escaping the pool (Fig 1C).

Fig 1. Topography of the study site: (**A**) the Lotus Pond hot spring in the context of the river Rulang and the valley of Puga, (**B**) the boiling vent-water explored for microbiome structure and function, (**C**) boratic sinters, and fresh condensates sulfur and boron minerals covering the broken wall of the old crater within which Lotus Pond is embedded.

Water (temperature 85°C, pH 7.5) was collected from Lotus Pond’s vent at 13:00 h of 29 October 2021, when ebullition was strong and fumarolic activity copious. For microscopic studies, 500 mL water was passed through a sterile 0.22 µm mixed cellulose ester membrane filter (Merck Life Science Private Limited, India) having a diameter of 47 mm, and inserted into a cryovial containing 0.9% (w/v) NaCl solution supplemented with 15% (v/v) glycerol. Simultaneously, for the extraction of total environmental DNA (metagenome), a total of 5 L water was passed through five individual sterile 0.22 µm filters, each of which was inserted into a cryovial containing 5 mL of 50 mM Tris:EDTA (pH 7.8). All the cryovials were sealed with parafilm (Tarsons Products Limited, India), packed in polyethylene bags, put into dry ice, and air lifted to the laboratory at Bose Institute in Kolkata.

Enumeration of microbial cells

Total microbial cell count per unit volume of vent-water was determined alongside the number of viable and non-viable cells present. The microbial cells that were filtered out from 500 mL vent-water were first detached from the adherence matrix by shredding the filter with sterile scissors, and then vortexing for 30 min, within the NaCl-glycerol containing cryovial in which the filter was inserted on field. The vial was spun at 1000 *g* for 5 seconds to allow the filter shreds to settle down. Finally, the supernatant was collected (without disturbing the bottom debris), its exact volume measured, and then analyzed as described previously [32] using a

hemocytometer (Paul Marienfeld GmbH & Co. KG, Germany) and an upright fluorescence microscope (Olympus BX53 Digital, Olympus Corporation, Japan), with the axiom that it contained all the microbial cells that were present in the 500 mL vent-water filtered on field. The cell suspension was stained with 4′6-diamidino-2-phenylindole (DAPI), fluorescein diacetate (FDA) or propidium iodide (PI) solution, and cell numbers were counted as described previously [33].

Metagenomics

Metagenome was extracted from the microbial cells that were filtered out from Lotus Pond's vent-water in 0.22 µm mixed cellulose ester membrane filter, as described previously [5, 10, 18]. Paired end (2 × 250 nucleotides) whole metagenome shotgun sequencing was performed on Novaseq 6000 platform (Illumina Inc., USA). Clipping of adapters and quality-filtering of the trimmed sequences (for average Phred score ≥20) were carried out using Trim Galore v0.6.4. *De novo* assembly of the metagenomic sequence was performed using Megahit v1.2.9 [34]. Open reading frames (ORFs) or genes were first identified within the assembled contigs using Prodigal v2.6.3 [35], and then annotated as potential protein coding sequences (CDSs) by searching against the eggNOG database v5.0 [36] using eggNOG-mapper v2.1.9 [37].

Quality-filtered raw reads were searched against the RDP rrnDB database v5.8 [38] to extract sequences matching prokaryotic 16S rRNA genes, using Bowtie2 v2.2.5 [39]. The extracted reads were annotated taxonomically using RDP Classifier (http://rdp.cme.msu.edu/classifier/classifier.jsp). The quality-filtered read-set was also searched against the SILVA database v138.1 [40] to extract sequences matching eukaryotic 18S rRNA genes, using the Basic Rapid Ribosomal RNA Predictor tool Barrnap v0.9 [41]. The extracted reads were annotated taxonomically using Silva Alignment, Classification and Tree Service [42].

Construction of metagenome-assembled genomes (MAGs)

Metagenomic contigs were first binned into putative population genomes or MAGs using Metabat2 v2.12.1 [43], MaxBin2 v2.2.4 [44], and CONCOCT v1.1.0 [45]. Best quality MAGs were then selected using DASTool v1.1.6 [46] and checked for their completeness and contamination levels using CheckM v1.2.2 [47].

Taxonomic identity of the selected MAGs was first predicted independently via searches for genom-genome relatedness conducted using Genome Taxonomy Database Toolkit (GTDB-Tk) v1.7.0 [48], Rapid Annotations using Subsystems Technology (RAST) [49], and Type Strain Genome Server (TYGS) in default mode [50]. As and when required, MAG versus genome/MAG pairs were tested for dDDH using the Genome-to-Genome Distance Calculator v3.0 [50]. Orthologous genes-based average nucleotide identity (ANI) between a pair of MAG and genome/MAG was calculated using ChunLab's online ANI Calculator [51].

All MAGs were annotated using Prokaryotic Genome Annotation Pipeline (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/), and phylogeny of bacterial MAGs, in relation to their closest genomic relatives, was reconstructed based on conserved marker gene sequences, using the Up-to-date Bacterial Core Gene Pipeline v3.0 [52]. The best fit nucleotide substitution model was selected using ModelTest-NG v0.1.7 [53], following which a maximum likelihood tree was generated using RAxML

v8.2.12 [54]. The Interactive Tree of Life v6.7.5 [55] was used to visualize the phylogram. The KEGG (Kyoto Encyclopedia of Genes and Genomes) automatic annotation server (KAAS) was used for ortholog assignment and pathway mapping within the MAGs [56]. Relative abundance of the individual MAGs within the metagenome was determined by mapping raw read-pairs onto the different MAG sequences using Bowtie2 v2.2.5 [39].

16S amplicon sequencing and analysis

PCR-amplified 16S rRNA gene sequence-based taxonomic diversity was revealed using the fusion primer protocol [10, 17, 22] on the Ion S5 DNA sequencing platform (Thermo Fisher Scientific, USA), which in turn involved PCR amplification, sequencing, and analysis of the V3 regions of all Bacteria-specific 16S rRNA genes and V4-V5 regions of all Archaea-specific 16S rRNA genes [18] present in the Lotus Pond metagenome. Reads present in an amplicon sequence dataset (725025 and 589009 reads were there in the bacterial and archaeal sets respectively) were trimmed for their adapters, filtered for average Phred score ≥20 and length >100 nucleotides, and then clustered into operational taxonomic units (OTUs), i.e. species-level entities united by ≥97% 16S rRNA gene sequence similarities, by Usearch v10.0.259 [57]. While singleton sequences were eliminated before clustering, OTUs generated were classified taxonomically based on their consensus sequences using the RDP Classifier. Rarefaction curves (Fig S1) were generated via Usearch.

Details of microscopy and microbiome analyses are given in Supplementary Methods.

Results

Density of microbial cells in Lotus Pond

Gross density of microbial cells in the vent-water of Lotus Pond was found to approximately 8.5×10^4 mL^{-1}, via staining with DAPI, followed by hemocytometry using a fluorescence microscope (Fig 2). On the other hand, the density of metabolically viable cells was estimated to be approximately 5.4×10^4 mL^{-1}, via staining with FDA, while that of apparently dead cells was 3.2×10^4 cells mL^{-1}, according to hemocytometry via PI staining. Thus, the ratio between the densities of metabolically active and dead cells in Lotus Pond's vent water was almost 1.7.

Fig 2. Representative picture of the fluorescence microscopic fields based on which microbial cell density was calculated in the vent-water sample of Lotus Pond: (**A**) sample stained with DAPI, (**B**) sample stained with FDA, (**C**) sample stained with PI. Scale bars in all the three micrographs indicate 50 µm length.

Microbiome architecture reconstructed from the assembled metagenome

Quality filtration of the native metagenomic sequence dataset having 31353866 read-pairs resulted in the retention of 31342225 read-pairs. Assembly of the quality-filtered dataset yielded 286918 contigs having a

minimum and an average length of 200 and 563 nucleotides respectively (96% of all quality-filtered reads participated in the assembly, at an average coverage of ~85X). While the largest contig obtained was 197270-nucleotide long, the consensus sequence of the assembled metagenome was ~162 million nucleotides, with a total of 379412, complete or partial, ORFs or genes identified therein. Out of the 379412 ORFs identified, 220687 could be annotated as CDSs, of which again 220635 could be ascribed to potential taxonomic sources (Table S1). Approximately 97.3%, 1%, 1.5% and 0.2% of the taxonomically classifiable CDSs, i.e. 214652, 2309, 3276 and 398 CDSs, belonged to Bacteria, Archaea, Eukarya and Viruses respectively.

The 214652 bacterial CDSs identified were found to be distributed over 23 phyla, with most CDSs being affiliated to Proteobacteria / Pseudomonadota (90793), followed by Aquificae / Aquificota (54151), Firmicutes / Bacillota (18670), Deinococcus-Thermus / Deinococcota (10591), Cyanobacteria (5827), Chloroflexi / Chloroflexota (4254), Bacteroidetes / Bacteroidota (2753), Thermotogae / Thermotogota (2671), Thermodesulfobacteria / Thermodesulfobacteriota (2547), Actinobacteria / Actinomycetota (2357). Unclassified Bacteria accounted for 18386 CDSs, while 13 other (minor) phyla, namely Planctomycetes / Planctomycetota (355), Synergistetes / Synergistota (200), Verrucomicrobia / Verrucomicrobiota (160), Acidobacteria / Acidobacteriota (134), Fusobacteria / Fusobacteriota (133), Nitrospirae / Nitrospirota (126), Chlorobi / Chlorobiota (93), Spirochaetes / Spirochaetota (77), Thermomicrobia / Thermomicrobiota (63), Deferribacteres / Deferribacterota (43), Tenericutes / Mycoplasmatota (25), Gemmatimonadetes / Gemmatimonadota (21) and Chlamydiae / Chlamydiota (20) were represented sparsely in the CDS catalog (Tables S1 and S2; Fig 3A).

Among the different classes of Proteobacteria, Alphaproteobacteria, followed by Gammaproteobacteria, Betaproteobacteria, Deltaproteobacteria and Epsilonproteobacteria accounted for most CDSs in the assembled metagenome (41306, 31865, 14394, 1145 and 466 CDSs respectively). The other proteobacterial classes Hydrogenophilia, Acidithiobacillia and Oligoflexia were represented by 118, 106 and 101 CDSs respectively, while unclassified Proteobacteria accounted for 1046 CDSs (Fig 3A). Furthermore, within the proteobacterial CDS catalog, Xanthomonadales/Lysobacterales, Vibrionales, Oceanospirillales (all three belonged to the Gammaproteobacteria), and Rhodocyclales (belonging to Betaproteobacteria) were the maximally represented orders, with 10039, 6025, 3,940 and 3582 CDSs affiliated to them respectively (Tables S1 and S2).

The small archaeal genetic diversity was contributed by Crenarchaeota/Thermoproteota, followed by Euryarchaeota (1479 and 647 CDSs respectively); unclassified Archaea accounted for 183 CDSs (Tables S1 and S2). While the Crenarchaeota related CDSs were not classifiable further, the Euryarchaeota-affiliated CDSs were distributed over the classes Halobacteria (177), Thermococci (169), Archaeoglobi (83), Methanococci (60), Methanomicrobia (30), Methanobacteria (19) and Thermoplasmata (9); unclassified Euryarchaeota accounted for 100 CDSs (Tables S1 and S2). Most of the putatively eukaryotic CDSs detected exhibited highest sequence similarities with homologs from Opisthokonta, affiliated to either Fungi

or Metazoa (Table S1). The few viral CDSs detected were mostly contributed by Podoviridae, Myoviridae, Siphoviridae and Caudovirales (Table S1).

Fig 3. Taxonomic distribution of (**A**) the CDSs identified in the assembled whole metagenome of Lotus Pond, and (**B**) the metagenomic reads which corresponded to 16S rRNA genes (actual CDS count or metagenomic read count recorded for each taxon is given in parenthesis). In (**A**) “minor phyla” included the following: Acidobacteria, Chlamydiae, Chlorobi, Deferribacteres, Fusobacteria, Gemmatimonadetes, Nitrospirae, Planctomycetes, Spirochaetes, Synergistetes, Tenericutes, Thermomicrobia and Verrucomicrobia. In (**B**) “other phyla”, also marginal in representation, included the following: Cyanobacteria, Thermotogae, Bacteroidetes, Chloroflexi, Actinobacteria, Armatimonadetes and Synergistetes.

Genus-level constitution of the microbiome

Taxonomic classification of CDSs predicted within the assembled metagenome was not extended up to the genus level. Instead, generic components of the microbiome were delineated via taxonomic classification of (i) raw metagenomic reads corresponding to 16S rRNA genes, and (ii) OTUs clustered from PCR-amplified 16S rRNA gene sequences. Additionally, the taxonomic diversity data obtained from the current analysis of PCR-amplified 16S rRNA gene sequences were utilized to evaluate the consistency of the Lotus Pond vent-water microbiome by comparing with equivalent data available from previous explorations of the habitat.

Out of the 62684450 high quality reads analyzed for the Lotus Pond metagenome, 81686 represented parts of prokaryotic 16S rRNA genes: 81223 bacterial, 396 archaeal and 67 unclassifiable (Table S3). Notably, no read having any meaningful similarity (homology) with any eukaryotic 18S rRNA gene was detected within the entire metagenomic sequence dataset.

80348 out of the total 81223 metagenomic reads corresponding to bacterial 16S rRNA genes were classifiable under 12 phyla (the remaining 875 reads belonged to unclassified Bacteria), of which Aquificae accounted for an overwhelming majority (65853) of reads (Table S3; Fig 3B). Substantive number of reads were also ascribable to Proteobacteria (11639), Deinococcus-Thermus (1517) and Firmicutes (567), while Thermodesulfobacteria (308), Cyanobacteria (131), Thermotogae (196), Bacteroidetes (56), Chloroflexi (56), Actinobacteria (21), Armatimonadetes / Armatimonadota (2), and Synergistetes (2) contributed much fewer reads. The proteobacterial reads, again, were mostly from Gammaproteobacteria (9553), followed by Betaproteobacteria (1345) and Alphaproteobacteria (595). Furthermore, 70729 out of the total 81223 bacterial 16S rRNA gene sequence related reads got ascribed to 92 genera (Table S3) led by *Hydrogenobacter* (44547), *Sulfurihydrogenibium* (16371), *Halomonas* (3075), *Vibrio* (2549), *Thermus* (1454), *Tepidimonas* (965) and *Paracoccus* (196). Whereas the two Aquificales genera *Hydrogenobacter* and *Sulfurihydrogenibium*, affiliated to the families Aquificaceae and Hydrogenothermaceae respectively, accounted for 80% of all 16S rRNA gene related reads present in the raw metagenomic sequence dataset;

the other Aquificae members detected, namely *Thermocrinis*, *Venenivibrio*, *Persephonella* and *Hydrogenothermus*, accounted for only 180, 48, 6 and 2 such reads respectively (Table S3).

As for the 396 reads coresponding to archaeal 16S rRNA genes, all were attributable to the phylum Crenarchaeota and only 9 were classifiable up to the level of the genus, detecting *Aeropyrum* (6), *Stetteria* (2) and *Pyrobaculum* (1) in the process (Table S3).

A corroborative experiment involving PCR amplification, seuencing and analysis of the V3 regions of all Bacteria-specific 16S rRNA genes present in the metagenome extracted from the hot water discharged by Lotus Pond revealed 602 bacterial OTUs that in turn were classifiable under 16 different phyla: Proteobacteria (352), Firmicutes (63), Bacteroidetes (52), Actinobacteria (40), Cyanobacteria (20), Deinococcus-Thermus (16), Thermotogae (8), Aquificae (3), Acidobacteria (2), Armatimonadetes (2), Chloroflexi (2), Spirochaetes / Spirochaetota (2), Thermodesulfobacteria (2), Dictyoglomi / Dictyoglomota (1), Rhodothermaeota / Rhodothermota (1) and Synergistetes (1); unclassified Bacteria accounted for 35 OTUs (Table S4). Of the total 602 bacterial OTUs identified, 249 were classifiable upto the level of genus. Overall, 121 genera were identified in that way (Table S4), with maximum number of OTUs getting affiliated to *Vibrio* (21), followed by *Halomonas* (16) and *Thermus* (13).

Concurrent to the above experiment, PCR amplification, seuencing and analysis of the V4-V5 regions of all Archaea-specific 16S rRNA genes present in the Lotus Pond vent-water metagenome were carried out to reveal the *in situ* archaeal diversity (Table S4). Only 30 OTUs were detected in this way, of which, 10 were ascribable to Crenarchaeota, 6 each belonged to *Euryarchaeota* and Nitrososphaerota, and 8 were not classifiable below the domain level (Table S4). Furthermore, none of the Crenarchaeota OTUs were classifiable at the genus level, whereas out of the 6 OTUs affiliated to Euryarchaeota, 1 and 3 were classifiable as species of *Methanomassiliicoccus* and *Methanospirillum* respectively; all the 6 Nitrososphaerota OTUs were classified as members of *Nitrososphaera* (Table S4).

Predominant generic entities identified via population genome binning

When all the >2500 nucleotide contigs present in the assembled metagenome were binned using Metabat2 , MaxBin2 , and CONCOCT via separate *in silico* experiments, 19, 14 and 30 MAGs were obtained respectively. Refinement and optimization of all these population genome bins using DASTool yielded 14 bacterial (Tables 1 and S5; Fig 4) and one archaeal (Tables 1 and S5) for subsequent consideration and analysis. Each of these draft genomes had a contamination level of <5%, except the one designated as LotusPond_MAG_Paracoccus_sp._1. Likewise, most of the short-listed MAGs (9 out of 15) had >90% completeness; two, identified as Paracoccus_sp._2 and Unclassified_Thermodesulfobacteriaceae, had <50% completeness; and four, identified as Paracoccus_sp._1, Fervidobacterium_sp., Thermus_sp. and Unclassified_Bacteria, possessed approximately 51.2%, 70.2%, 70.3% and 81.9% completeness respectively. In terms of consensus sequence length, the population genomes classified as *Vibrio metschnikovii* and *Halomonas* sp. were the largest (approximately 3.5 and 3.4 mb respectively) whereas the one classified as a novel member of the family Thermodesulfobacteriaceae was the smallest (approximately

0.4 mb only). Furthermore, when the 31342225 high quality read-pairs used for metagenome assembly were searched against the 15 population genomes, a sum total of 68.08% read-pairs mapped concordantly on to the MAGs as a whole. Individually, the 15 MAGs accounted for 0.06% to 55.9% of the metagenomic read-pairs analyzed.

Fig 4. Phylogenetic tree (number of bootstrap tests carried out = 10000), based on 92 universal bacterial core genes, showing the evolutionary relationships among the 14 bacterial MAGs reconstructed from the Lotus Pond vent-water metagenome and their closest relatives identified based on dDDH and/or rRNA gene sequence relationships. Scale bar denotes a distance equivalent to 10% nucleotide substitution. In the phylogeny reconstructed, nucleotide substitutions were interpreted using the generalized time-reversible model, which considers the proportion of invariable sites and/or the rate of variation across the sites.

The genome bin LotusPond_MAG_Unclassified_Aquificaceae accounted for an overwhelming proportion of the metagenome (55.9% of all read-pairs assembled mapped back to this MAG), while another Aquificales bin identified as *Sulfurihydrogenibium azorense*, a member of Hydrogenothermaceae, accounted for 3.5% of all read-pairs. Four more bins individually represented >1% of the metagenome; these were identifed as (i) an unknown species of the genus *Halomonas*, (ii) an unclassified member of the order Desulfurococcales within the archaeal class Thermoprotei, (iii) the gammaproteobacteium *Vibrio metschnikovii*, and (iv) an unknown species of the genus *Thermus*. These MAGs accounted for 2.7%, 1.6%, 1.5% and 1% of all metagenomic read-pairs respectively. Seven other MAGs, ascribed to *Tepidimonas taiwanensis*, *Tepidimonas* sp., *Fervidobacterium* sp., *Thermosynechococcus* sp., *Paracoccus* sp., an unclassified member of the gammaproteobacterial order Chromatiales, and an unclassified taxon of Bacteria, individually accounted for >0.1% but <1% of the metagenome. The remaining two Lotus Pond MAGs individually represented <0.1% of the metagenome; these were identified as one unclassified member each from the family Thermodesulfobacteriaceae of the monotypic phylum Thermodesulfobacteria, and the genus *Paracoccus*. Details of the phylogenomic relationships, and the consequent basis of taxonomic identification, of the individual MAGs are given in Supplementary Results.

Table 1. Key features of the population genome bins (MAGs) constructed from the assembled metagenome of Lotus Pond's vent-water.

Name of the population genome bin obtained (genome accession number)	% of quality-filtered metagenomic read-pairs matching sequences of the MAG concordantly	No. of contigs	Genome size (nucleotides)	G+C content (%)	Completeness (%)	Contamination (%)	Number of ORFs (genes) / CDSs detected in the MAG
LotusPond_MAG_Unclassified_Aquificaceae (JARJNB000000000)	55.92	131	1489711	43.32	95.12	0.61	1754 / 1707
LotusPond_MAG_Sulfurihydrogenibium_azorense (JARJNA000000000)	3.48	47	1516769	32.77	98.98	0.00	1680 / 1639
LotusPond_MAG_Halomonas_sp. (JARJMQ000000000)	2.65	97	3349510	60.21	99.46	0.94	3245 / 3176
LotusPond_MAG_Unclassified_Desulfurococcales (JARJMZ000000000)	1.60	37	1553981	49.38	99.63	0.00	1621 / 1559
LotusPond_MAG_Vibrio_metschnikovii (JARJMP000000000)	1.54	249	3487509	44.28	94.07	1.85	3222 / 3172
LotusPond_MAG_Thermus_sp. (JARJMW000000000)	1.01	372	1887946	62.80	70.34	4.24	2206 / 2157
LotusPond_MAG_Tepidimonas_sp. (JARJMT000000000)	0.46	200	2150705	66.30	94.16	1.29	2200 / 2155
LotusPond_MAG_Thermosynechococcus_sp. (JARJMS000000000)	0.35	104	2328058	52.52	97.05	0.12	2367 / 2319
LotusPond_MAG_Tepidimonas_taiwanensis (JARJMR000000000)	0.28	245	2568138	69.51	90.53	3.56	2567 / 2525
LotusPond_MAG_Fervidobacterium_sp. (JARJMY000000000)	0.24	293	1525481	39.93	70.18	0.88	1639 / 1611
LotusPond_MAG_Paracoccus_sp._1 (JARJMU000000000)	0.17	528	2138329	62.94	51.22	15.08	2383 / 2345
LotusPond_MAG_Unclassified_Bacteria (JARJMV000000000)	0.11	335	2025372	61.84	81.86	0.31	2082 / 2039
LotusPond_MAG_Unclassified_Chromatiales (JARJMX000000000)	0.10	214	1644354	62.57	91.04	2.01	1796 / 1758
LotusPond_MAG_Unclassified_Thermodesulfobacteriaceae (JARJND000000000)	0.07	114	439694	36.23	30.15	0.31	477 / 467
LotusPond_MAG_Paracoccus_sp._2 (JARJNC000000000)	0.06	284	1087115	64.14	34.31	1.10	1214 / 1201

Environmentally advantageous functions encoded by the Lotus Pond MAGs

Apart from house-keeping genes, all the MAGs encompassed genes for heat stress management. Irrespective of their completeness level, MAGs affiliated to truly thermophilic taxa had less than 10 such genes each, whereas almost all the MAGs affiliated to mesophilic taxa had ≥10 heat stress management genes each. Almost all the MAGs contained genes for managing oxidative and periplasmic stress; a half of them had genes for withstanding carbon starvation and nitrosative stress. Eight MAGs encoded genes conferring high frequency of lysogeny as a means of stress response. Every bacterial MAG, but not the archaeal one, encompassed genes for stringent response to environmental stimuli via (p)ppGpp metabolism. Most of the MAGs encoded versatile sporulation-associated proteins, while those affiliated to mesophilic or moderately thermophilic taxa encoded a universal stress protein. Four of the 15 MAGs possessed 12-34 genes for different toxin-antitoxin systems, eight had 1-9 such genes, whereas three (belonging to Aquificaceae, Desulfurococcales and Thermodesulfobacteriaceae) had none. All but one MAG contained genes for cation transport; most had genes for twin-arginine motif-based protein translocation and Ton and Tol transport systems, while some encoded ABC transporters, uni- sym- and antiporters, TRAP transporters, and various secretion systems. Half of the MAGs harbored genes for flagellar motility or chemotaxis; protection against metals, toxins, and antibiotics; and CRISPR systems.

Details of the adaptationally useful genes encompassed by the different MAGs are given in Supplementary Results.

Microbiome functions predicted from the assembled metagenome

The assembled Lotus Pond metagenome encompassed 220687 complete or partial protein coding sequences (Table S1), out of which 210761 were ascribable to clusters of orthologous genes (COGs) that again were distributed across 23 COG categories (Table S6). Of these, energy production and conversion, and DNA metabolism, had the maximum number of CDSs ascribed to them (23053 and 22454 respectively). Next in attendance were the COG categories translation, ribosomal structure and biogenesis (18543 CDSs), amino acid transport and metabolism (17442 CDSs), cell wall/membrane/envelope biogenesis (14204 CDSs), coenzyme transport and metabolism (13812 CDSs), and inorganic ion transport and metabolism (13507 CDSs). Notably, the lowest numbers of CDSs, 42, 28 and 24, were allocated to cytoskeleton, extracellular structures, and RNA processing and modification respectively. The COG categories nuclear structure and mobilome (prophages and transposons) had no gene assigned to them at all (Table S6).

For energy production and conversion, maximum homologs were detected for the genes encoding NADH-quinone oxidoreductase and cytochrome *c* oxidase. While oxidative phosphorylation was the most represented pathway under energy production and conversion, NADH:quinone oxidoreductase, the reductive tricarboxylic acid (rTCA) cycle or the reductive citrate cycle, the TCA or citrate cycle, and cytochrome *c* oxidase were the biochemical modules with most CDSs.

For DNA metabolism, maximum number of homologs was detected for a transposase encoding gene. Mismatch repair, homologous recombination, replication, and nucleotide excision repair were the most represented pathways, while DNA polymerase III complex and fluoroquinolone-resistance gyrase-protecting protein Qnr, were the most represented biochemical modules. For translation and ribosomal structure and biogenesis, maximum homologs were detected for the elongation factor Tu. Aminoacyl-tRNA biosynthesis was the most represented pathway, while ribosome (bacterial followed by archaeal) was the maximally represented biochemical module.

Regarding amino acid transport and metabolism, maximum homologs were there for glycine cleavage system protein H and cysteine desulfurase. Biosynthesis of secondary metabolites was the most represented pathway, while maximum CDSs were attributed to the biochemical module photorespiration.

As for cell wall/membrane/envelope biogenesis, maximum homologs were detected for the transaminase which isomerizes glutamine:fructose-6-phosphate, and the GTP-binding protein LepA, which acts as a potential fidelity factor of translation (for accurate and efficient protein synthesis) under certain stress conditions [58, 59]. Under this category, amino sugar and nucleotide sugar metabolism, antibiotics biosynthesis, and peptidoglycan biosynthesis were the most represented pathways. At the module level, nucleotide sugar biosynthesis, dTDP-L-rhamnose biosynthesis, lipoprotein-releasing system, lipopolysaccharide biosynthesis (KDO2-lipid A), and multidrug resistance related efflux pumps AcrAB-TolC/SmeDEF, AcrAD-TolC, MexAB-OprM and AmeABC were maximally represented.

For coenzyme transport and metabolism, maximum homologs were detected for uroporphyrin-III C-methyltransferase. Secondary metabolites biosynthesis, porphyrin metabolism, pantothenate and CoA biosynthesis, and antibiotics biosynthesis were the most represented pathways, while heme biosynthesis was the most represented biochemical module.

Regarding inorganic ion transport and metabolism, maximum homologs were detected for NADH-quinone oxidoreductase subunit L. Under this category, ABC transporters, oxidative phosphorylation, quorum sensing, and two-component system, were the most represented pathways, wheras at the module level, NADH:quinone oxidoreductase, peptides/nickel transport system, phosphate transport system, iron complex transport system, nickel transport system, NitT/TauT family transport system, nitrate/nitrite transport system, and iron(III) transport system were represented maximally.

Chemolithotrophic sulfur oxidation (Sox) is the primary bioenergetic mechanism of the Lotus Pond vent-water ecosystem

Out of all the mechanisms of chemolithotrophy known, only sulfur oxidation (Sox) was found to have substantial numbers of coding sequences, partial or complete, in the assembled metagenome for most of its constituent biochemical reactions (enzymatic steps). Overall, 333 CDSs were there for the different components of the Sox multienzyme complex: 21 for the L-cysteine S-thiosulfotransferase protein SoxX (K17223), 53 for the SoxA subunit of L-cysteine S-thiosulfotransferase (K17222), 112 and 57 for the sulfur

compounds chelating/binding protein subunits SoxY (K17226) and SoxZ (K17227) respectively, 79 for the S-sulfosulfanyl-L-cysteine sulfohydrolase SoxB (K17224), and 11 for the sulfane dehydrogenase protein's molybdopterin-containing subunit SoxC (K17225) and diheme *c*-type cytochrome subunit SoxD (K22622), the CDSs of which occur universally as overlapping ORFs. In terms of taxonomic distribution, maximum number of Sox protein encoding sequences (~40% of the total 333) was ascribed to Aquificae, followed by Alphaproteobacteria (~25%), Betaproteobacteria (14%), and Deinococcus-Thermus (9%). Unclassified Bacteria, Gammaproteobacteria, and other Proteobacteria individually accounted for approximately 8, 4 and 1 percent of alll the Sox CDSs annotated (Table S1).

As for the other forms of chemolithotrophy known in the microbial world, only a small number of CDSs could be detected for the key proteins of arsenite and hydrogen oxidation. Whereas five CDSs each were detected for the small and large subunits AoxA (K08355) and AoxB (K08356) of arsenite oxidase respectively, only a single CDS was identifiable for the small subunit of ferredoxin hydrogenase (K00534), even as no CDS could be detected for the hydrogenase large subunit (K00533). Notably, no CDS was also there for any of the other hydrogenase variants known in the literature.

No CDS was present in the Lotus Pond microbiome for the chemolithotrophic iron oxidation protein iron:rusticyanin reductase (K20150). Likewise, there was also no CDS for the enzymes concerned with the chemolithotrophic oxidation of ammonia to nitrite. However, despite the absence of genes encoding the different subunits of methane/ammonia monooxygenase (*pmoABC* - K10944, K10945 and K10946), or for that matter the nonexistence of any CDS for the other ammonia-oxidizing enzyme hydroxylamine dehydrogenase (K10535), the Lotus Pond metagenome did encompass several genes for the alpha (K00370) and beta (K00371) subunits of nitrite to nitrate oxidizing enzyme nitrate reductase / nitrite oxidoreductase (269 and 90 CDSs respectively).

So far as the population genome bins are concenrned, the one designated as LotusPond_MAG_Unclassified_Aquificaceae was found to have genes coding for the proteins SoxX, SoxA, SoxB, SoxY, and SoxZ, plus the accessory protein SoxW, which is a thioredoxin serving the purpose of reducing the different subunits of the Sox multienzyme complex to their functionally active states. The genome bin designated as LotusPond_MAG_Thermus_sp. encoded the proteins SoxX, SoxA, SoxB, SoxC, and SoxD, while the one entitled as LotusPond_MAG_Unclassified_Chromatiales contained genes for SoxX, SoxA, SoxB, and SoxC. LotusPond_MAG_Paracoccus_sp._1 contained gene encoding SoxB, SoxCD, and the additional sulfur oxidation proteins SoxF (a periplasmic flavoprotein sulfide dehydrogenase) and SoxG (a thiosulfate-induced periplasmic zinc metallohydrolase that can act as a potential thiol esterase), while the largely incomplete genome bin designated as LotusPond_MAG_Paracoccus_sp._2 contained genes for only SoxY, SoxZ, and the transcriptional regulator of the *sox* operon called SoxR.

The population genome bin LotusPond_MAG_Sulfurihydrogenibium_azorense encompassed genes for all the four subunits - alpha (K17993), delta (K17994), gamma (K17995) and beta (K17996) – of the

signature protein of the genus called sulfhydrogenase (HydADGB), which is H_2:polysulfide oxidoreductase catalyzing hydrogen oxidation via reduction of elemental sulfur or polysulfide to hydrogen sulfide.

Centrality of the rTCA cycle, and Aquificae, in Lotus Pond's primary productivity

In the assembled metagenome of Lotus Pond, only rTCA (Arnon-Buchanan) cycle and reductive pentose phosphate (Calvin-Benson-Bassham) cycle had substantial numbers of CDSs, whether complete or partial, for all the enzymatic steps necessary to execute the pathway. For each of the other four pathways of carbon assimilation (autotrophic carbon fixation) known thus far, genes concerned with a number of intermediate transformations were either missing from the metagenome or were represented by just one or two CDSs (Fig 5). In terms of taxonomic distribution (Table S1), most of the rTCA cycle related CDSs (~53% of the total 2910) identified in the assembled metagenome were attributed to Aquificae, whereas most of the reductive pentose phosphate cycle related CDSs identified were attributed to Alphaproteobacteria and Firmicutes (each of the two taxa accounted for ~20% of the total 1076 CDSs detected for Calvin cycle proteins). Cyanobacteria, notably, accounted for only ~2% of the total Calvin cycle CDSs that were there.

Fig 5. Carbon assimilation (autotrophic) pathways [93] identified in the Lotus Pond vent-water microbiome based on the annotation of CDSs in the assembled metagenome with reference to the modules curated under KEGG Pathway Maps database located at https://www.genome.jp/kegg/pathway.html: (**A**) reductive citrate cycle, (**B**) reductive pentose phosphate cycle, (**C**) dicarboxylate/4-hydroxybutyrate cycle, (**D**) 3-hydroxypropionate/4-hydroxybutyrate cycle, (**E**) 3-hydroxypropionate bi-cycle, and (**F**) reductive acetyl Co-A pathway. The active inorganic carbon substrates assimilated by the different pathways are highlighted in grey. The biochemical conversions (steps) of the different pathways for which CDSs could be identified in the assembled metagenome are indicated by the roman numerals I though L. While the figures given in parentheses denote the numbers of CDSs identified for the individual enzymatic steps, the structural (enzyme-coding) genes actually detected in the metagenome for the biochemical steps I through L are enumerated below by their KEGG orthology identifiers. I - K00024; II - K01676, K01677, K01678, K01679; III - K00239, K00240, K00241, K00242, K00244, K00245, K00246, K00247, K18556, K18557, K18558, K18559, K18560; IV - K01902, K01903; V - K00174, K00175, K00176, K00177; VI - K00031; VII - K01681, K01682; VIII - K15232, K15233, K15234; IX - K15230, K15231, K15232, K15233, K15234; X - K00169, K00170, K00171, K00172, K03737; XI - K01006, K01007; XII - K01595; XIII - K01958, K01959, K01960; XIV - K00855; XV - K01601, K01602; XVI - K00927; XVII - K00134, K00150; XVIII - K01623, K01624; XIX - K01086, K02446, K03841, K11532; XX - K00615; XXI - K01086, K11532; XXII - K01807, K01808; XXIII - K15038; XXIV - K14534; XXV - K15016; XXVI - K00626; XXVII - K00169, K00170, K00171, K00172; XXVIII - K01007; XXIX - K01677, K01678; XXX - K00239, K00240, K00241, K18860; XXXI - K01902, K01903; XXXII - K15037; XXXIII - K15020; XXXIV - K05606, K01848, K01849; XXXV - K01961, K01962, K01963, K02160; XXXVI - K14468; XXXVII - K14469; XXXVIII - K08691; XXXIX - K14449; XL - K14470; XLI - K09709; XLII -

K15052; XLIII - K05606, K01847, K01848, K01849; XLIV - K14471, K14472; XLV - K00239, K00240, K00241, K01679; XLVI - K05299, K22015; XLVII - K00198, K14138, K00194, K00197; XLVIII - K01938; XLIX - K01491; L - K00297.

So far as the key carbon fixation steps of the different autotrophic pathways are concerned, rTCA cycle had 411, 34, 57 and 274 CDSs within the metagenome for the conversion of (i) acetyl-CoA to pyruvate (with CO_2 addition), (ii) phosphoenolpyruvate to oxaloacetate (with HCO_3^- addition), (iii) succinyl-CoA to 2-oxoglutarate, and (iv) 2-oxoglutarate to isocitrate (the last two conversions involve CO_2 addition) respectively (Fig 5A). Likewise, 40 CDSs were detectable for the sole carbon fixation step (ribulose 1,5-bisphosphate to 3-phosphoglycerate with the addition of CO_2) of the reductive pentose phosphate cycle (Fig 5B).

Dicarboxylate/4-hydroxybutyrate cycle had 400 and 34 CDSs in the metagenome for its two carbon fixation steps converting (i) acetyl-CoA to pyruvate, and (ii) phosphoenolpyruvate to oxaloacetate, via addition of CO_2 and HCO_3^- respectively. Yet, the numbers of CDSs detected for the reactions of the hydroxybutyrate half of the cycle were extremely low (Fig 5C). In this context it is noteworthy that the carbon fixation reactions of the dicarboxylate/4-hydroxybutyrate cycle are same as the first two carbon fixation steps of the rTCA cycle: phosphoenolpyruvate to oxaloacetate conversion involves the same enzyme phosphoenolpyruvate carboxylase (KEGG orthology: K01595) in either pathway, while there is only one additional enzyme [pyruvate-ferredoxin/flavodoxin oxidoreductase Por (K03737)] for acetyl-CoA to pyruvate conversion by the rTCA cycle, and that precisely accounted for the extra 11 CDSs ascribed to the corresponding step of the rTCA cycle (Fig 5). Evidently, therefore, the high CDS count for the carbon fixation reactions of dicarboxylate/4-hydroxybutyrate cycle were only a reflection of the overall preponderance of the rTCA cycle, and dicarboxylate/4-hydroxybutyrate cycle is actually not prevalent in the microbiome.

In relation to carbon fixation steps, the 3-hydroxypropionate/4-hydroxybutyrate cycle had only one CDS in the assembled metagenome for its signature protein acetyl-CoA/propionyl-CoA carboxylase (K15037), which by its three subunits structured in an $\alpha_4\beta_4\gamma_4$ configuration (and having functional domains for carboxyltransferase, biotin carboxylase, and biotin carboxyl carrier protein, respectively) catalyzes (i) the conversion of acetyl-CoA to malonyl-CoA, as well as (ii) the conversion of propanoyl-CoA to (S)-methylmalonyl-CoA, using $HCO3^-$ as the active inorganic carbon source in both the reactions.

The 3-hydroxypropionate bi-cycle encompasses similar carbon fixation steps as those mentioned for the 3-hydroxypropionate/4-hydroxybutyrate cycle, but the two $HCO3^-$ assimilating reactions here are catalyzed by distinct types of enzyme systems. Acetyl-CoA to malonyl-CoA conversion in the bi-cycle is rendered by the acetyl-CoA carboxylase complex AccABCD made up of the biotin carboxylase subunit AccC (K01961), the carboxyl transferase subunits alpha (AccA; K01962) and beta (AccD; K01963), and the biotin carboxyl carrier AccB (K02160). Propanoyl-CoA, in the bi-cycle, is converted to (S)-methylmalonyl-CoA by an

independent propionyl-CoA carboxylase (K15052). The Lotus Pond metagenome encompassed 479 CDSs for the AccABCD complex, whereas only eight CDSs could be identified for propionyl-CoA carboxylase.

For the two CO_2 fixation steps of the reductive acetyl-CoA pathway, which generate formate and tetrahydrofolate, the following CDSs were identified in the Lotus Pond vent-water metagenome. Total four CDSs were there to encode the alpha subunit FdhA (K05299) of the formate dehydrogenase [$NADP^+$] FdhAB, and the other formate dehydrogenase (hydrogenase) homolog which has the KEGG orthology K22015. On the other hand, a total of ten CDSs could be detected for the following components of the tetrahydrofolate generating enzyme system: (i) anaerobic carbon-monoxide dehydrogenase catalytic subunit (K00198), (ii) acetyl-CoA synthase (K14138), (iii) acetyl-CoA decarbonylase/synthase, CODH/ACS complex subunits delta and gamma (K00194 and K00197 respectively).

The most predominant Lotus Pond MAG, identified as an unclassified Aquificaceae member closely related to *Hydrogenobacter*, was found to possess at least six rTCA cycle genes: (i) aconitate hydratase *acnA* (K01681), (ii) isocitrate dehydrogenase *icd* (K00031), (iii) pyruvate, water dikinase *ppsA* (K01007), (iv) pyruvate carboxylase *pyc* (K01958), (v) pyruvate ferredoxin oxidoreductase alpha subunit *porA* (K00169), and (vi) succinyl-CoA synthetase alpha subunit *sucD* (K01902). Other MAGs encompassing considerable numbers of rTCA cycle genes were as follows.

- LotusPond_MAG_Thermus_sp., which encoded the proteins (i) AcnA (K01681), (ii) pyruvate, orthophosphate dikinase PpdK (K01006), (iii) PpsA (K01007), (iv) phosphoenolpyruvate carboxylase Ppc (K01595), (v) Por (K03737), (vi) succinate dehydrogenase flavoprotein subunit SdhA (K00239), and (vii) succinyl-CoA synthetase beta subunit SucC (K01903).
- LotusPond_MAG_Tepidimonas_sp., which encoded (i) AcnA, (ii) Icd (K00031), (iii) PpsA, (iv) Ppc, and (v) SdhA.
- LotusPond_MAG_Tepidimonas_taiwanensis, which encoded (i) AcnA, (ii) Icd, (iii) PpsA, (iv) Ppc, (v) SucC
- LotusPond_MAG_Unclassified_Bacteria, which encoded (i) AcnA, (ii) PpdK, (iii) SdhA, (iv) SucC, and (v) SucD (K01902).
- LotusPond_MAG_Unclassified_Desulfurococcales, which encoded (i) AcnA, (ii) 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase alpha subunit KorA (K00174), (iii) PpsA, and (iv) SdhA.
- LotusPond_MAG_Fervidobacterium_sp., which encoded (i) 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunits alpha (OorA, K00174) and beta (OorB, K00175), (ii) PpdK, (iii) fumarate hydratase alpha subunit FumA (K01677), and (iv) PorA (K00169).
- LotusPond_MAG_Paracoccus_sp._1, which encoded (i) AcnA, (ii) Icd, (iii) class II fumarate hydratase FumC (K01679), (iv) SdhA.
- LotusPond_MAG_Vibrio_metschnikovii, which encoded (i) AcnA, (ii) Icd, (iii) Ppc, and (iv) Por.

The putatively autotrophic entity LotusPond_MAG_Thermosynechococcus_sp. encompassed the following genes for the reductive pentose phosphate cycle proteins: (i) fructose-1,6-bisphosphatase II /

sedoheptulose-1,7-bisphosphatase (K11532), (ii) ribose 5-phosphate isomerase A RpiA (K01807), (iii) fructose-bisphosphate aldolase, class I (K01624), (iv) fructose-1,6-bisphosphatase I Fbp (K03841), (v) transketolase TktA (K00615), and (vi) ribulose bisphosphate carboxylase, both the small (RbcS) and large (RbcL) chains (K01602 and K01601). Additionally, this MAG had at least nine more genes for CO_2 uptake/fixation and carboxysome; these encoded (i) the RuBisCo chaperonin RbcX, (ii) the NADH dehydrogenase subunit 4 (K03948), (iii) the NADH dehydrogenase subunit 5 (K03949), (iv) a sensory subunit of low CO_2-induced protein complex, (v) the low-affinity CO_2 hydration protein CphX, (vi) the carboxysome protein CcmL (K08697), (vii) the CO_2-concentrating carboxysome protein CcmM (K08698), (viii) the CO_2-concentrating carboxysome protein CcmN (K08699), and (ix) the CO_2-concentrating mechanism protein CcmO (K08700).

LotusPond_MAG_Unclassified_Chromatiales, the other putatively autotrophic genome bin detected, encoded the reductive pentose phosphate cycle proteins (i) transketolase TktA (K00615), (ii) ribose 5-phosphate isomerase A RpiA (K01807), and (iii) ribulose bisphosphate carboxylase, both the small and large chains (K01602 and K01601); plus (i) the RuBisCo activation proteins CbbO and CbbQ, and (ii) the RuBisCO operon transcriptional regulator CbbR. Other MAGs encompassing considerable numbers of reductive pentose phosphate cycle genes included the following.

- LotusPond_MAG_Tepidimonas_taiwanensis, which encoded (i) RbcS, (ii) RbcL, (iii) TktA, (iv) Fbp, (v) CbbO, and (vi) CbbR.
- LotusPond_MAG_Tepidimonas_sp., which encoded (i) RbcS, (ii) RbcL, (iii) TktA, and (iv) CbbR.

Copious invasion and virulence related genes in the Lotus Pond metagenome

While there were several complete or partial CDSs (a total of 76) in the Lotus Pond vent-water's assembled metagenome for proteins concerned with biofilm formation and cellular adhesion, quite a few CDSs attributed to hemagglutinin and hemolysin biosynthesis (86), and stationary phase metabolisms including the formation of persister cells (43) were also there (Table S1). Furthermore, the metagenome encompassed 1084 complete or partial CDSs encoding the different components of Type-I, Type-II, Type-III, Type-IV, Type-VI, Type-VIII and Type-IX secretion systems (SSs). Of these, most were parts of the Type-II, and Type-III and Type-IV SSs. In terms of taxonomic distribution (Table S1), most of the SS related CDSs identified (57% of the total 1084) were attributed to Aquificae, followed by Betaproteobacteria (~10%). Gammaproteobacteria and Firmicutes each accounted for 8%, while Alphaproteobacteria accounted for ~5% of all these CDSs. Approximately 1-3% of all SS related CDSs were ascribable to Deinococcus-Thermus, Unclassified Proteobacteria, Unclassified Bacteria, Thermodesulfobacteria, Cyanobacteria, Actinobacteria, and Archaea. Thermotogae accounted for ~0.6% of the SS related CDSs identified in the assembled metagenome. The Lotus Pond MAGs also encompassed quite a few CDSs for biofilm formation, cellular adhesion, invasion and virulence, with the one identified as *Vibrio metschnikovii* having the maximum number and diversity of CDSs for these functions.

Antibiotic synthesis and resistance genes detected in the Lotus Pond metagenome

Within the Lotus Pond metagenome, 3164 complete or partial CDSs (Tables S1 and S7) could be identified as involved in the biosynthesis of antibiotics such as the monobactams (816), streptomycin (788), prodigiosin (419), acarbose and validamycin (329), novobiocin (288), vancomycin (177), phenazine (105), carbapenems (96), penicillins and cephalosporins (85), neomycin/kanamycin/gentamicin (55), and staurosporine (6).

In terms of taxonomic distribution (Fig 6 and Table S7), maximum number of these antibiotic synthesis related CDSs (~29% of the total 3164) was attributed to Aquificae, followed by Alphaproteobacteria (~21%). Thermodesulfobacteria and Gammaproteobacteria each accounted for ~11% of all antibiotic synthesis related CDSs (ASCs) identified, while Betaproteobacteria and Firmicutes accounted for ~6% each. Approximately 4.5% ASCs were ascribable to Deinococcus-Thermus, whereas Chloroflexi, the entire domain Archaea, unclassified Bacteria, and Cyanobacteria contributed 2-2.5% ASCs each (out of the 68 archaeal ASCs detected, 46 and 22 were ascribed to Euryarchaeota and Crenarchaeota respectively). Furthermore, the phyla Actinobacteria, Bacteroidetes and Thermotogae each accounted for ~1% of the total ASCs present in the Lotus Pond metagenome (Fig 6 and Table S7).

Fig 6. Heat maps showing the quantitative distribution of antibiotic (**A**) synthesis and (**B**) resistance related CDSs (ASCs and ARCs respectively) across microbial phyla in the Lotus Pond vent-water. For a given antibiotic, the numbers of ASCs or ARCs detected across the microbial phyla (Table S7) were first normalized by calculating the geometric mean of the CDS counts in question and then dividing each CDS count by that mean. Subsequently, for heat map construction, Z-score normalization was performed on the normalized CDS counts across phyla for each antibiotic. Z-scores were computed on an antibiotic-by-antibiotic (column-by-column) basis by subtracting the mean of the normalized CDS counts from each normalized CDS count and then dividing the individual remainders by the standard deviation of the normalized CDS counts. The columns (antibiotics) being Z-Score scaled, cell colors represent how ASC counts or ARC counts vary for a single antibiotic across the phyla, with white representing a column Z-Score of 0, while increasing positive values are represented by intensifying shades of green and increasing negative values are represented by intensifying shades of red.

The taxonomic groups possessing the widest variety of ASCs within the Lotus Pond microbiome were the alpha, beta, and gamma sub-classes of Proteobacteria, which possessed genes for all the antibiotics mentioned above. Among the native thermophilic taxa, Aquificae followed by Thermodesulfobacteria were found to have the most number of ASCs, which in turn were concerned with the synthesis of all the antibiotics mentioned above except penicillins and cephalosporins, and staurosporine. However, the Thermodesulfobacteria were far less enriched than the Aquificae in terms of ASCs for carbapenems, monobactams, novobiocin, phenazine, and prodigiosin. Deinococcus-Thermus members present in the

microbiome possessed ASCs for the same antibiotics as those synthesized by Aquificae and Thermodesulfobacteria, plus penicillins and cephalosporins. Thermotogae had very few CDSs each for all the antibiotics synthesized by Aquificae and Thermodesulfobacteria, except carbapenems and phenazine.

The Lotus Pond Archaea had no CDS for the biosynthesis of carbapenems, staurosporine, and phenazine; they had only one or two (and in a few cases three or four) CDS(s) each for vancomycin, penicillins and cephalosporins, neomycin/kanamycin/gentamicin, acarbose and validamycin, novobiocin and prodigiosin; and just 46 and 10 CDSs for monobactams and streptomycin respectively (Fig 6 and Table S7).

In tandem with the copious ASCs, 5899 such complete or partial CDSs (Tables S1 and S7) were identified in the assembled metagenome of Lotus Pond's vent-water that were involved in the resistance of beta-lactam antiibiotics (penicillins, cephalosporins, monobactams, and carbapenems) (1122); vancomycin (469); cationic antimicrobial peptides or CAMPs (840); aminoglycoside antibiotics such as streptomycin, neomycin, kanamycin and gentamicin (396); fluoroquinolones (175); tetracycline (38); imipenem (34); polymyxin antibiotics (7); bacitracin (4); nisin (4); and a wide variety of antimicrobial drugs by means of efflux pumping for which there were 2810 CDSs.

In terms of taxonomic distribution (Fig 6 and Table S7), maximum number of these antibiotic resistance related CDSs (30% of the total 5899) was attributed to Aquificae, followed by Gammaproteobacteria (~28%), Alphaproteobacteria (19%), and Betaproteobacteria (~10%). Unclassified Proteobacteria, Cyanobacteria, Bacteroidetes, Deinococcus-Thermus, unclassified Bacteria, and Firmicutes each accounted for approximately 1-3.5% of all the antibiotic resistance related CDSs (ARCs) identified, while Actinobacteria, Thermotogae, Thermodesulfobacteria, and Chloroflexi accounted for 0.2-0.8% each. Only approximately 0.02-0.03% ARCs each were ascribable to Spirochaetes, Deferribacteres, the entire domain Archaea, Verrucomicrobia, and Synergistetes. Notably, the whole of Archaea accounted for only one ARC, and the same was meant to protect against aminoglycosides.

The taxonomic group possessing the widest variety of ARC (Fig 6 and Table S7) was Gammaproteobacteria, which possessed CDSs for the resistance of aminoglycosides, beta-lactam antiibiotics, CAMPs, fluoroquinolones, imipenem, polymyxin antibiotics, tetracycline, vancomycin, and a wide variety of drugs by means of efflux pumping. Although the Aquificae had the most number of ARCs, the CDSs possessed by the group conferred resistance only against aminoglycosides, beta-lactam antibiotics, CAMPs, fluoroquinolones, and vancomycin; they however had the highest number of CDSs for multidrug resistance efflux pumps. Thermodesulfobacteria and Deinococcus-Thermus were found to have only a few CDSs each for resistance against the same set of antibiotics as the ones resisted by the Aquificae. Thermotogae too had very few CDSs each for all the antibiotics synthesized by Aquificae, Thermodesulfobacteria and Deinococcus-Thermus, except CAMPs.

The Lotus Pond MAGs also encompassed large repertoires of ASCs and ARCs, with the ones identified as *Thermus* sp., unclassified Bacteria, and *Paracoccus* sp. 2 having the maximum number and diversity of CDSs for antibiotic synthesis and resistance.

Discussion

Comparing microbial diversities revealed by shotgun metagenomics and 16S amplicon sequencing

The extents of Lotus Pond's microbial diversity revealed by the different approaches of molecular genetics had their own differences, as well as concurence. For instance, two archaeal phyla, Crenarchaeota and Euryarchaeota, were identified by taxonomic annotation of the CDSs predicted within the assembled metagenome (Table S2), but searching of raw metagenomic reads for 16S rRNA-encoding sequences detected only Crenarchaeota, whereas PCR amplified 16S rRNA gene sequence based OTU analysis identified Crenarchaeota, Euryarchaeota and Nitrososphaerota. Concurrently, the archaeal genera *Aeropyrum*, *Pyrobaculum* and *Stetteria*, identified by direct annotation of raw metagenomic reads went undetected in 16S amplicon-based analysis, whereas *Methanomassiliicoccus*, *Methanospirillum* and *Nitrososphaera*, detected via 16S amplicon sequencing, were undetectable in the direct annotation of raw metagenomic reads (Tables S3 and S4).

For Bacteria, 23 phyla were identified via taxonomic classification of CDSs: of these, 10 (Actinobacteria, Aquificae, Bacteroidetes, Chloroflexi, Cyanobacteria, Deinococcus-Thermus, Firmicutes, Proteobacteria, Thermodesulfobacteria, and Thermotogae) were major phyla embracing >2000 CDSs each, and 13 (Acidobacteria, Chlamydiae, Chlorobi, Deferribacteres, Fusobacteria, Gemmatimonadetes, Nitrospirae, Planctomycetes, Spirochaetes, Synergistetes, Tenericutes, Thermomicrobia, and Verrucomicrobia) were minor phyla with <500 CDSs each (Table S2; Fig 3A). Annotation of raw metagenomic reads corresponding to 16S rRNA genes detected 12 phyla in all – these included all the 10 major phyla mentioned above, plus one of the 13 minors phyla (Synergistetes), and a new phylum (Armatimonadetes) that was not represented in the CDS catalog (Table S3; Fig 3B). 16S amplicon-based OTU analysis identified 16 bacterial phyla, which included the 10 major ones identified from CDS annotation, plus two of the 13 minor phyla (Acidobacteria and Synergistetes), and Armatimonadetes, and three new phyla (Dictyoglomi, Rhodothermaeota and Spirochaetes) that were not represented among the CDSs or the raw reads encoding 16S rRNAs (Table S4).

Annotation of raw metagenomic reads corresponding to bacterial 16S rRNA genes revealed 92 bacterial genera, while OTU analysis of PCR amplified 16S rRNA gene sequences identified 121. Between these two sets of generic names obtained using two different approaches, 50 genera were common (in Tables S3 and S4), of which seven (*Fervidobacterium*, *Halomonas*, *Paracoccus*, *Sulfurihydrogenibium*, *Tepidimonas*, *Thermus* and *Vibrio*) were represented in the MAGs obtained from the assembled metagenomic data, and two (*Hydrogenobacter* and *Thiofaba*) were represented by very closely related MAGs (namely, LotusPond_MAG_Unclassified_Aquificaceae and LotusPond_MAG_Unclassified_Chromatiales; detailed phylogenomic relationships are given in the Supplementary Results). Of all the genera represented among the MAGs (Tables 1 and S5; Fig 4), *Thermosynechococcus* alone was not detected in any 16S rRNA gene sequence-based analysis; however, LotusPond_MAG_Thermosynechococcus_sp. did encompass a near-complete 16S rRNA gene, 99.6% similar to the *Thermosynechococcus elongatus* BP-1 homolog. This

indicated that the nonappearance of *Thermosynechococcus* in 16S rRNA gene sequence based analyses was a bioinformatic glitch.

Discrepancies are known to arise in the microbial diversities revealed by analyzing the sample sample via whole metagenome shotgun sequencing and PCR-amplified 16S rRNA gene sequencing. A number of studies have found 16S amplicon sequencing to underrepresent taxa at different hierarchical levels [60-63], whereas the opposite outcome has also been reported by quite a few studies [64, 65]. The present data demonstrated that the annotation of raw metagenomic reads corresponding to 16S rRNA genes reliably delineated most part of the diversity under investigation, at both the phylum and genus levels. Although this approch did not fare well in revealing the sarcely represented taxa, the diversity revealed by this approach was fairly reproducible in the corresponding data obtained via taxonomic classification of CDSs annotated within the assembled metagenome, and the OTUs clustered from PCR-amplified 16S rRNA gene sequences. The latter two approaches, in turn, appeared to be more adept to unearthing the less-abundant taxa, at the phylum as well as genus level. With all the three approaches having their own strengths and limitations, a simultaneous use of all of them seems to be the best practice in obtaining a comprehensive picture of microbial diversity in a given environmental sample.

Lotus Pond's vent-water microbiome remains stable over time

Over the last ten years, the Lotus Pond has been explored for its vent-water bacterial (but not archaeal) diversity, via 16S amplicon sequencing alone, on five previous time-points [10, 17]. A comparison of the present 16S amplicon-based OTU data with those from the past showed that all the 16 bacterial phyla detected in the current investigation (Table S4), except Rhodothermaeota and Synergistetes, had been identified in previous exploration(s). The five previous explorations too had collectively discovered 16 bacterial phyla, of which only Ignavibacteriae and Nitrospirae went undetected in the present 16S amplicon-based survey. Furthermore, 11 bacterial phyla, namely Actinobacteria, Aquificae, Bacteroidetes, Chloroflexi, Cyanobacteria, Deinococcus-Thermus, Firmicutes, Proteobacteria, Synergistetes, Thermodesulfobacteria and Thermotogae were consistently detected in the present annotation of CDSs, 16S rRNA-encoding metagenomic reads, and OTUs clustered from 16S amplicon sequences; all of these, except Synergistetes, were detectable in one or more previous studies of 16S amplicon sequencing and analysis.

At the genus level, 121 entities were detected in the current PCR-based investigation, out of which 59 (Table S4) were identified in one or more previous explorations, which in turn had collectively discovered 156 genera. Seven out of the 97 genera that were identified previously, but went undetected in the present PCR-based analysis (these were *Acidovorax*, Armatimonadetes_gp5, *Buttiauxella*, Cyanobacteria GpIV, *Klebsiella*, *Ornithinimicrobium* and *Sphingomonas*), got identified in the present annotation of raw metagenomic reads corresponding to 16S rRNA genes (Table S3). In this way, a total of 66 bacterial genera (Table 2) were found to be present consistently in Lotus Pond's vent-water over a period of approximately ten years. Notably, six out of these 66 genera, namely *Fervidobacterium*, *Halomonas*, *Paracoccus*,

Sulfurihydrogenibium, *Tepidimonas* and *Thermus*, were unequivocally represented in the population genome bins (MAGs) obtained from the current metagenomic data, while two more, namely *Hydrogenobacter* and *Thiofaba*, were represented by very closely related entities, if not direct member strains of the genera. In this way, there were only two such bacterial genera - namely *Thermosynechococcus* and *Vibrio* - that figured among the MAGs obtained in the current study but were not represented in any of the previous OTU data.

Table 2. Bacterial genera that were detected in the present exploration of the Lotus Pond vent-water as well as in one or more previous studies of the same habitat using 16S rRNA gene sequence analysis.

Phylum to which the genera belonged	**Names of the genera[1,2]**	**Maximum temperature for Laboratory-growth[3] (°C)**	**Sampling date of the previous study in which the genus was also identified**	
			23 July, 2013 [17]	**20 October, 2014 [10]**
Actinobacteria	***Brevibacterium***	42	+	+
	Corynebacterium	42	+	+
	Mycobacterium	43	-	+
	Nocardioides	37	-	+
	Ornithinimicrobium	50	+	-
	Rothia	38	+	-
Armatimonadetes	Gp5	-	-	+
	Gp7	-	-	+
Aquificae	***Hydrogenobacter***	85	+	+
	Sulfurihydrogenibium	80	+	+
Bacteroidetes	***Chryseobacterium***	37	+	+
	Cloacibacterium	40	+	+
	Pedobacter	35	-	+
	Porphyromonas	37	+	-
	Prevotella	45	+	-
Chloroflexi	***Chloroflexus***	59	+	+
Cyanobacteria	**GpXIII**	-	-	+
	GpIV	-	-	+
Deinococcus-Thermus	***Thermus***	80	+	+
	Truepera	50	+	+
Dictyoglomi	***Dictyoglomus***	80	-	+
Firmicutes	***Anoxybacillus***	70	+	+
	Bacillus	70	+	+
	Chryseomicrobium	45	+	-
	Enterococcus	45	+	-
	Exiguobacterium	49	+	+
	Gemella	37	-	+
	Lactococcus	40	+	-
	Lysinibacillus	45	-	+
	Planococcus	42	+	+
	Sporolactobacillaceae _incertae_sedis	-	-	+
	Staphylococcus	40	+	+
	Streptococcus	45	+	+
Alphaproteobacteria	***Azospirillum***	37	+	+
	Bradyrhizobium	37	+	+
	Brevundimonas	42	+	+
	Paracoccus	45	+	+

	Sphingobium	37	-	+
	Sphingomonas	45	+	+
	Sulfitobacter	40	+	+
Betaproteobacteria	***Achromobacter***	42	+	+
	Acidovorax	37	+	+
	Burkholderia	37	+	+
	Comamonas	44	-	+
	Delftia	40	-	+
	Herbaspirillum	45	+	-
	Pelomonas	37	+	+
	Ralstonia	41	+	+
	Tepidimonas	60	+	+
Gammaproteobacteria	***Aeromonas***	41	+	+
	Acinetobacter	37	+	+
	Buttiauxella	42	-	+
	Enhydrobacter	41	+	+
	Escherichia/Shigella	37	-	+
	Halomonas	50	+	+
	Klebsiella	37	-	+
	Marinobacter	50	+	+
	Pseudomonas	42	-	+
	Psychrobacter	38	+	+
	Rheinheimera	35	-	+
	Serratia	35	+	-
	Shewanella	30	+	-
	Stenotrophomonas	42	+	+
	Thiofaba	51	+	+
	Thiovirga	34	+	-
Thermotogae	***Fervidobacterium***	90	+	+

[1] Genera detected in the present PCR-amplified 16S rRNA gene sequence analysis are indicated by bold fonts.

[2] Names of the genera that were detected in the present annotation of metagenomic reads corresponding to 16S rRNA genes are underscored.

[3] Full references concerning the maximum temperatures for laboratory-growth of all the above genera have been cited in relation to Tables S3 and S4, within the Supplementary Information file.

Lotus Pond is not reserved for thermophiles, albeit Aquificales predominate numerically and nutritionally

Literature survey for the phenotypic characteristics of the cultured members of the 92 bacterial genera identifed via taxonomic classification of 16S rRNA encoding raw metagenomic reads showed that the constituent strains of only 29 genera had reports of growth in the laboratory above 45°C (see Table S3 and references therein). A similar exercise for the 121 genera detected via OTU-level clustering of PCR amplified 16S rRNA gene sequences showed that member strains of only 30 genera had reports of *in vitro* growth above 45°C (Table S4). Among the 66 bacterial genera regularly present in Lotus Pond vent-water, only 15 were found to have strains reported for laboratory growth above 45°C: these were *Fervidobacterium*, *Hydrogenobacter*, *Sulfurihydrogenibium*, *Thermus*, *Dictyoglomus*, *Anoxybacillus*, *Bacillus*, *Tepidimonas*,

Chloroflexus, *Thiofaba*, *Truepera*, *Halomonas*, *Marinobacter*, *Ornithinimicrobium* and *Exiguobacterium* (Table 2). Corroborative to this high diversity of bacterial mesophiles, 41.2% of all taxonomically classifiable CDSs identified in the assembled metagenome were ascribable to Proteobacteria, whereas only 24.5% were affiliated to Aquificae (Table S2; Fig 3A). Interestingly, at least 70 CDSs for different cold shock proteins (most of them were DNA-binding) could be detected in the Lotus Pond metagenome, out of which, again, 22 were ascribable to Aquificae, two each to Deinococcus-Thermus and Thermotogae, and one to Thermodesulfobacteria; the rest were affiliated to Proteobacteria (38), Firmicutes (5), Actinobacteria (1), and Synergistetes (1). These genomic features corroborate that the Lotus pond vent-water is neither a "Thermophiles only" habitat nor a closed, biophysically firewalled, ecosystem. Instead it hosts physiologically diverse microorganisms, including mesophiles and psychrophiles, which come into the system from the adjoining territories of the cold and arid Changthang Plateau. Furthermore, it highlights the potential fluxes of the thermophilic bacteria out of the high temperature ecosystem, into the cold desert biome, over time and space. All these suppositions are in agreement with the biogeodynamic model of the Puga hydrothermal system [10], as well as the "hot spring – river" ecological continuum that embodies the topography of the Lotus Pond site (Fig 1).

Although phylogenetic relatives of mesophilic bacteria overwhelmed the taxonomic, as well as the overall genetic, diversity of the Lotus Pond vent-water microbiome, thermophilic Aquificales dominated the habitat in terms of relative abundance. Only two MAGs affiliated to Aquificaceae/*Hydrogenobacter* (maximum temperature for laboratory-growth of Aquificaceae is 95°C [66]), and *Sulfurihydrogenibium* (maximum temperature for laboratory-growth: 80°C [67]), collectively made up 59.4% of the whole metagenome (Table S5). *Hydrogenobacter* and *Sulfurihydrogenibium* also accounted for ~80% of all 16S rRNA-encoding reads identified within the raw metagenomic sequence dataset (Table S3).

Among the other taxa identified as MAGs, *Fervidobacterium* (maximum temperature for laboratory-growth: 90°C [68]), *Thermus* (maximum temperature for laboratory-growth: 80°C [69]), the family Thermodesulfobacteriaceae of the phylum Thermodesulfobacteria (maximum temperature for laboratory-growth: 100°C [70]), and the order Desulfurococcales within the class Thermoprotei of the phylum Crenarchaeota (maximum temperature for laboratory-growth: 110°C [71]) have strains growing at temperatures above or within the ten-year range (78-85°C) recorded for Lotus Pond's vent-water [10, 16, 17]. These MAGs together accounted for ~3% of the metagenome.

Cyanobacteria have been found in 73-85°C springs [72, 73], but *Thermosynechococcus* are restricted to ≤70°C waters [74, 75] while its cultured strains are known to grow only at ≤60°C [76, 77]. *Halomonas* [78], *Paracoccus* [28, 79], *Tepidimonas* [80], Chromatiales [81] and *Vibrio* [82] have temperature maxima for laboratory-growth at 50°C, 45°C (or 50°C under special conditions), 60°C, 51°C, and 50°C respectively. Thus, eight out of the 15 Lotus Pond MAGs, collectively accounting for 5.6% of the metagenome, were affiliated to genera encompassing strains that grow only at temperatures below the range known for Lotus Pond's vent-water. That said, *Halomonas*, *Paracoccus*, *Tepidimonas*, *Thiofaba* and *Vibrio* have been isolated

from hydrothermal vents or vent-adjacent environments, and found to exhibit varying degrees of thermal endurance [28, 80, 81, 83-85].

The Aquificae not only predominated the Lotus Pond vent-water microbiome in terms of relative abundance, but also acted as the key primary producers of the ecosystem, accounting for most of the genes present in the metagenome for the most prevalent pathways of chemolithotrophy (sulfur oxidation) and autotrophy (carbon fixation), namely the Sox mechanism and the rTCA cycle respectively. While the Sox multienzyme complex system is thought to have originated in ancient thermophilic bacteria ancestral to *Aquificae* and *Epsilonproteobacteria* [86], the putatively primordial rTCA cycle has supposedly originated in sulfide-rich hydrothermal vent environments of the early Earth [87]. Although the constituent protein subunits of Sox have coevolved parallel to one another as orthologous lineages in many extant sulfur chemolithotrophs, the system has also spread across the taxonomic spectrum of *Bacteria* through extensive paralogy of the components [86, 88]. On the other hand, in the extant biota, the rTCA cycle is known to operate in members of Aquificales, and Chlorobi, Deltaproteobacteria, Epsilonproteobacteria, Nitrospirae, and Crenarchaeota [89-92], while the present exploration revealed the presence of rTCA cycle homologs in MAGs affiliated to Aquificaceae, and unclassified Bacteria, Desulfurococcales, *Fervidobacterium*, *Paracoccus*, *Tepidimonas*, *Thermus*, and *Vibrio*. The occurrence of rTCA cycle genes in the unexpected residents of this hydrothermal habitat (phylogenetic relatives of known mesophiles) thus shed new light on the ecological amplitude, selection, and evolution (including horizontal spread) of this ancient autotrophic pathway in the context of thermal stress.

The preponderance of Sox and rTCA in Lotus Pond conforms not only to the pathways' ecology and evolution, but also the habitat's physical chemistry, spatial abundance and isotopic ratios of redox sulfur species, and the copious presence of HCO_3^- [10, 16]. In contrast, the existence of a substantial cache of genes for the reductive pentose phosphate cycle, including in the MAGs belonging to *Tepidimonas*, *Thermosynechococcus* and unclassified Chromatiales, is idiosyncratic to the biophysics of the ecosystem, and necessitates a renewed attention to the feasibility of this pathway at temperatures above the upper limit (70-75°C) known thus far for its functionality [93]. Furthermore, in this context it is noteworthy that several homologs of (i) the phosphoenolpyruvate carboxylase gene *ppc* (K01595), (ii) the malate dehydrogenase *mdh* (K00024), (iii) the oxaloacetate decarboxylating and $NADP^+$ dependent malate dehydrogenase gene *maeB* (K00029), and (iv) the pyruvate, orthophosphate dikinase *ppdK* (K01006) were found to be parts of the Lotus Pond vent-water metagenome, plus a few MAGs such as those belonging to *Paracoccus* and *Tepidimonas*. The protein products of these genes are known to govern the carbon fixation pathway called crassulacean acid metabolism (CAM) in plants adapted to aridity and high temperature stress [94]. To avoid gaseous exchange during the hot day time, CAM plants store CO_2 in the form of malic acid at night, transport the malate to chloroplasts during the day, re-convert it CO_2 and concentrate the same in the proximity of ribulose bisphosphate carboxylase molecules to carry out efficient carbon fixation [94]. A new line of pure

culture based research of physiology and biochemistry needs to be initiated to explore the plausible roles of analogous mechanisms in bacteria adapted to thermal stress.

Genomic resources that can help mesophiles survive in the Lotus Pond habitat

Lotus Pond's assembled metagenome encompassed several such genes that could be potentially decisive in thermal adaptation, especially in the context of the vent-water's geochemistry and the spring's topography (Supplementary Results). A majority of these CDSs were ascribable to phylogenetic relatives of mesophilic bacteria, and their most conspicuous examples included copious genes encoding heat shock proteins, molecular chaperones and chaperonin complexes that render the quality control of DNA and proteins amid biophysical stress. The metagenome contained several genes encoding proteins that control/modulate/inhibit DNA gyrase activity. In true thermophiles, reverse gyrases (RGs) are indispensible as they render positive supercoils in the DNA to counter-balance the negative supercoils promoted by gyrases, and *in situ* heat [95]. RGs also help avert incorrect aggregation of denatured DNA, and facilitate correct annealing [96]. Expectedly, Lotus Pond MAGs affiliated to Aquificaceae, Desulfurococcales, *Fervidobacterium* and *Sulfurihydrogenibium* encoded RGs, but those belonging to phylogenetic relatives of mesophilic or moderately thermophilic bacteria did not. Nevertheless, MAGs affiliated to *Halomonas*, *Paracoccus*, *Tepidimonas*, and *Vibrio* did encode gyrase inhibitors/modulators such as YacG/PmbA/TldD which are known to nullify gyrase activity [97, 98]; notably, several homologs of these genes ascribable to other mesophilic taxa were also present in the assembled metagenome. Future studies of metatranscriptomics and metaproteomics are needed alongside multi-omic investigations of pure culture isolates to ascertain whether and how the above specialties help hot spring mesophiles endure high temperatures over time and space.

The metagenome encoded different toxin–antitoxin (TA) systems, including the Type II, Type IV, RelE / RelB, PrlF / YhaV, YdaS / YdaT, HicAB, and MraZ / MraW systems, which act as regulatory modules for growth and survival of bacterial cells under diverse stress conditions [99]. Other ecologically advantageous genes having large numbers of homologs in the assembled metagenome encoded the following: (a) universal stress proteins, homologs of which are also possessed by bacterial mesophiles occupying ecological niches adjacent to deep sea hydrothermal vents [84]; (b) methionine sulfoxide reductases (Msr), which repair oxidatively damaged/inactivated proteins by converting their methionine sulfoxides to methionines; (c) enzymes known to protect cells from the toxic effects of peroxides, which in turn are often abundant in hot spring waters [100]; (d) diverse proteins adapted structurally and/or functionally to high concentrations of heavy metals, which is a common feature of hydrothermal discharges [101-104] including those of Lotus Pond and other hot springs of Puga Valley [5, 10, 14-16]; (e) different Mnt proteins that render cellular transport and homeostasis of manganese, a key cofactor of different enzymes protecting against oxidative damage; (f) fatty acid desaturases that may help increase the rigidity of cell membranes (via unsaturation of fatty acids).

Numerous genes were also there in the Lotus Pond metagenome for flagellar structure/function, chemotaxis, cell adhesion/aggregation, biofilm formation, and quorum sensing, all of which have the potentials to assist *in situ* microorganisms respond to environmental conditions and position themselves dynamically at their biophysically/biochemically best-suited niches along the hydrothermal (Fig. 1A) and biogeochemical [10] gradients occurring in the vent (Lotus Pond) to river (Rulang) trajectory.

Why are virulence genes abundant in Lotus Pond?

The Lotus Pond metagenome encompassed several CDSs for the formation of biofilms, adhesion of cells, biosynthesis of hemagglutinin and hemolysin, and stationary phase metabolisms including persister cell formation (Table S1). While biofilm formation, cellular adhesion, and haemolysin synthesis may all have key roles in osmoregulation and extreme-temperature response [105, 106], formation of persister cells can help bacterial mesophiles attain ultra-slow metabolic states, which have recently been reported to be central to the endurance of high *in situ* temperatures [28]. Furthermore, the metagenome encompassed more than a thousand genes for constructing Type-I through Type-IX SSs, excepting the Type-V and Type-VII variants. While all these molecular machineries secreting biomacromolecules to the cells' exterior have clearly documented roles in bacterial survival within eukaryotic hosts (and consequent virulence), their potentially multifaceted ecological roles are also not unappreciated [84]. For instance, macromolecules moved across the cell membrane by Type-II SS often serve as local and long-range effectors/promoters of nutrient acquisition and biofilm formation under adverse environmental conditions [107, 108]. Type-III SS, besides translocating effector bacterial proteins into host cells, plays important roles in stress response including those involving high salts and peroxides in the chemical milieu of the bacterium [109]. Systems level network analyses have also revealed interrelations between Type-III and Type-IV SSs on one hand and stress management on the other [110]. Furthermore, Type-IV SS homologs are also sometimes involved in inter-bacterial killing, which can not only eliminate competition but also turn out to be beneficial for the killers in nutrient deprived habitats [111, 112]. Type-VI SS homologs, besides being central to survival within eukaryotic hosts, regulate quorum sensing with wide-ranging implications for environmental adaptation [113, 114]. Type-VIII SS is implicated in adhesion, aggregation, and biofilm formation [115], while the Type-IX SS, reported thus far from *Bacteroidetes* alone, confers gliding motility to the environmental strains of the phylum [116]. Thus, in the context of the mesophiles present in the Lotus Pond vent-water, both these SSs might turn out to be instrumental in placing the native bacteria to biophysically favorable sites within the hydrothermal gradients (Fig. 1A).

Antibiosis as an instrument of ecosystem domination

Abundance of genes for antibiotic synthesis and resistance within the microbiomes of hydrothermal environments is a topic of considerable interest in contemporary microbiological research. Many thermophilic

bacteria have been reported to produce antibiotics, which in turn generate and spread antibiotic resistance in the ecosystem [117, 118].

Lotus Pond's vent-water metagenome contained thousands of genes for the biosynthesis and resistance of a wide array of antibiotics (Fig 6 and Table S7). An approximate one third of the antibiotic-synthesizing CDSs (ASCs) detected were ascribable to the Aquificae, with most of them being concerned with the production of monobactams, streptomycin, prodigiosin, acarbose and validamycin, novobiocin, carbapenems, and vancomycin, while a few were responsible for the biosynthesis of phenazine, and neomycin/kanamycin/gentamicin. In contrast, Thermodesulfobacteria, Deinococcus-Thermus, Archaea as a whole, and Thermotogae accounted for only 11, 5, 2, and 1 percent of the ASCs respectively. Notably, the Lotus Pond thermophiles outside of Aquificae possessed either a lesser variety of ASCs or had lesser numbers of CDSs for the individual antibiotics putatively synthesized. Furthermore, the Aquificae were the only Lotus Pond thermophiles to have a rich repertoire of antibiotic-resisting CDSs (ARCs), and of all the ARCs detected in the assembled metagenome they accounted for an approximate one third. In terms of the antibiotic variety resisted, the Aquificae had CDSs for protection against beta-lactam antiibiotics, CAMPs, vancomycin, aminoglycosides, fluoroquinolones (in that order of CDS count), and a host of other drugs via efflux pumping. The Thermodesulfobacteria and Deinococcus-Thermus members present *in situ* also possessed these ARCs, albeit in far lesser numbers. Qualitatively, Thermotogae too had the same ARCs as Aquificae, Thermodesulfobacteria and Deinococcus-Thermus (except those meant for resistance against CAMPs), but their corresponding CDS counts were far less than those recorded for the other three phyla. The archaea native to the Lotus Pond habitat were virtually devoid of any ARC, except for a single aminoglycoside resistance CDS.

In contrast to the thermophiles of the habitat, Lotus Pond's mesophiles (these included some facultative, moderate thermophiles as well), led by Alphaproteobacteria and Gammaproteobacteria, possessed a wider variety of ASCs, and a more even distribution of CDSs across the individual antibiotics putatively synthesized (the two groups also collective accounted for one third of all ASCs detected). For instance, they produced penicillins and cephalosporins, and staurosporine, which the Aquificae, Thermodesulfobacteria and Thermotogae did not (Fig 6 and Table S7). Furthermore, the Lotus Pond mesophiles, especially the Proteobacteria, which collective accounted for almost a half of all ARCs detected in the assembled metagenome, possessed genes to resist a wider variety of antibiotics. For instance, they had CDSs for protection against imipenem, polymyxins, and tetracycline, over and above the antibiotics against which the Aquificae and Thermodesulfobacteria were found to possess resistance (Fig 6 and Table S7). Thermotogae, on the other hand, had ARCs only against aminoglycosides, beta-lactam antiibiotics, fluoroquinolones, vancomycin, and a host of other drugs via efflux pumping.

The potent antibiosis machinery possessed by the few Aquificae species present in Lotus Pond's vent-water microbiome seems to have helped them get rid of potential niche overlap with their ecological equivalents, especially the thermophilic/hyperthermophilic archaea. Taking into account the cellular and

molecular peculiarities of the archaea, and consequentially their innate immunity against most antibacterial substances, it seems that the rich genetic repertoire of the Lotus Pond Aquificae for producing aminoglycosides such as streptomycin and neomycin/kanamycin/gentamicin, which act via inhibition of protein synthesis [119], and novobiocin that inhibits DNA gyrase [120], could have been instrumental in virtually vanquishing most archaeal populations from this habitat. The thermophilic/hyperthermophilic archaea, despite being equally adept to managing thermal stress, were consigned to the status of microscopic minorities within the community apparently because they were almost unarmed with ARCs, so practically defense-less against these antibiotics. However, for the Aquificae, antibiosis as a stratagem for ecosystem dominance or ecological supremacy was just not as effective in decimating the high *in situ* diversities of the mesophilic to moderately thermophilic Proteobacteria, Firmicutes, Cyanobacteria, Chloroflexi and Bacteroidetes that infiltrate Lotus Pond, or for that matter the few native thermophilic groups such as Deinococcus-Thermus, Thermotogae and Thermodesulfobacteria that were not totally devoid of ASCs and ARCs. That said, Aquificae's potent contrivances of biomolecular warfare must have exacerbated the other groups' bioenergetic costs of thermal endurance. In that way, the few Aquificae species present in the Lotus Pond vent-water not only managed to critically restrict the proliferation of the few bacterial thermophiles present *in situ*, but also keep a tab on the flux of the mesophilic bacteria that infiltrate the ecosystem.

Data availability

The whole metagenome shotgun sequence dataset was submitted to the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI), USA, under the BioProject PRJNA296849, with the BioSample and Run accession numbers SAMN31749247 and SRR22364517 respectively. The bacterial and archaeal 16S amplicon sequence datasets were submitted to the SRA, under the BioProject PRJNA296849, with BioSample accession numbers SAMN32724325 and SAMN32724260, and Run accession numbers SRR23072972 and SRR23072970, respectively.

Acknowledgements

This research was financed by Bose Institute (via intramural faculty grants) as well as the Science and Engineering Research Board (SERB), Government of India (GoI) (SERB grant number EMR/2016/002703). N.M. received fellowships from SERB and Bose Institute. S.D. and J.S. obtained their fellowships from Council of Scientific and Industrial Research, GoI. S.C. and M.M. received a fellowship from the Department of Biotechnology, GoI. We thank Professor Shubhra Ghosh Dastidar and Sri Sanjib Kumar Gupta for helping with computational resources.

Author contributions

W.G. conceived the program, designed the experiments, interpreted the results, and wrote the paper. N.M. anchored and harnessed all the modules of the study, performed the experiments of microscopy, metagenomics and bioinformatics, and analyzed and curated the data. S.D., J.S. and C.R. performed bioinformatics studies, while S.C. and M.M. carried out microscopic studies. W.G., N.M., S.D. and S.C. performed the field works. R.C. and C.R. made intellectual contributions to the overall work as well as the manuscript. All the authors read and vetted the paper.

Competing interests

The authors declare no competing interests.

Supporting information

Supplementary Methods

- Enumeration of microbial cells
- Extraction and sequencing of metagenome
- Assembly and annotation of the metagenomic sequence
- Construction of population genome bins or metagenome-assembled genomes (MAGs)
- PCR-amplified 16S rRNA gene sequence analysis

Supplementary Results

- Taxonomic relationships of the population genome bins obtained from the assembled metagenome
- Metabolically specialized (environmentally advantageous) functions encoded by the Lotus Pond MAGs
- Metagenome-encoded functions that can help bacterial mesophiles survive in the Lotus Pond habitat

Supplementary Tables

Table S1. Complete annotation of the 220687 protein coding sequences identified within the assembled metagenome of Lotus Pond's vent-water. Since this Table is more than one page long it has been provided as an Excel sheet named Table S1, within the Excel Workbook named Supplementary_Dataset.

Table S2. Taxonomic affiliation of the archaeal and bacterial protein-coding sequences predicted within the assembled metagenome of Lotus Pond’s vent-water.

Table S3. Taxonomic classification of the metagenomic reads which corresponded to archaeal and bacterial 16S rRNA gene sequences.

Table S4. Taxonomic classification of the OTUs obtained via PCR-amplified 16S rRNA gene sequence analysis.

Table S5. Characteristic features of the population genome bins constructed from the assembled metagenome of Lotus Pond's vent-water. Since this Table is more than one page long it has been provided as an Excel sheet named Table S5, within the Excel Workbook named Supplementary_Dataset.

Table S6. Metabolic categories over which the different CDSs ascribable to clusters of orthologous genes were distributed. Since this Table is more than one page long it has been provided as an Excel sheet named Table S6, within the Excel Workbook named Supplementary_Dataset.

Table S7. Numbers of antibiotic biosynthesis and resistance related CDSs detected across microbial phyla within the assembled metagenome of Lotus Pond's vent-water.

Supplementary References

- References used in Supplementary Methods and Supplementary Results
- References used in Table S3, S4 and S5

Supplementary Figure

Fig. S1. Rarefaction curves showing the proportionality between OTU-level diversity revealed and the number of 16S rRNA gene sequence reads analyzed, for archaea or bacteria, in the vent-water sample.

References

1. Wächtershäuser G. From volcanic origins of chemoautotrophic life to Bacteria, Archaea and Eukarya. Philos Trans R Soc B. 2006;361:1787-1808.

2. Patel BH, Percivalle C, Ritson DJ, Duffy CD, Sutherland JD. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. Nat Chem. 2015;7:301-307.

3. Ranjan S, Todd ZR, Sutherland JD, Sasselov DD. Sulfidic anion concentrations on early earth for surficial origins-of-life chemistry. Astrobiology. 2018;18:1023-1040.

4. Fields PA. Protein function at thermal extremes: balancing stability and flexibility. Comp Biochem Physiol A Mol Integr Physiol. 2001;129:417-431.

5. Roy C, Rameez MJ, Haldar PK, Peketi A, Mondal N, Bakshi U, et al. Microbiome and ecology of a hot spring-microbialite system on the Trans-Himalayan Plateau. Sci Rep. 2020;10:5917.

6. Kristjansson JK, Stetter KO. Thermophilic bacteria. In: Kristjansson JK, Stetter KO (eds). Thermophilic bacteria. 1st ed. CRC press, 2021. pp 1-18.

7. Hall JR, Mitchell KR, Jackson-Weaver O, Kooser AS, Cron BR, Crossey LJ, et al. Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes. Appl Environ Microbiol. 2008;74:4910-4922.

8. Everroad RC, Otaki H, Matsuura K, Haruta S. Diversification of bacterial community composition along a temperature gradient at a thermal spring. Microbes Environ. 2012;27:374-381.

9. Tang J, Liang Y, Jiang D, Li L, Luo Y, Shah MM, et al. Temperature-controlled thermophilic bacterial communities in hot springs of western Sichuan, China. BMC Microbiol. 2018;18:134.

10. Roy C, Mondal N, Peketi A, Fernandes S, Mapder T, Volvoikar SP, et al. Geomicrobial dynamics of Trans-Himalayan sulfur–borax spring system reveals mesophilic bacteria's resilience to high heat. J Earth Syst Sci. 2020;129:157.

11. Chiriac CM, Szekeres E, Rudi K, Baricz A, Hegedus A, Dragoş N, et al. Differences in temperature and water chemistry shape distinct diversity patterns in thermophilic microbial communities. Appl Environ Microbiol. 2017;83:e01363-17.

12. Ruhl IA, Sheremet A, Smirnova AV, Sharp CE, Grasby SE, Strous M, et al. Microbial functional diversity correlates with species diversity along a temperature gradient. Msystems. 2022;7:e00991-21.

13. Gansser A. Geology of the Himalayas. London: Interscience Publishers, 1964.

14. Garrett DE. Borates: Handbook of deposits, processing, properties, and use. Academic Press, 1998.

15. Rai AP. Compilation of data on chemical analysis of water and gas samples from North West Himalaya and adjoining areas. Bulletin Series-C, No. 5, Geological Survey of India, 2001.

16. Ghosh W, Mallick S, Haldar PK, Pal B, Maikap SC, Gupta SK. Molecular and cellular fossils of a mat-like microbial community in geothermal boratic sinters. Geomicrobiol J. 2012;29:879-885.

17. Roy C, Alam M, Mandal S, Haldar PK, Bhattacharya S, Mukherjee T, et al. Global association between thermophilicity and vancomycin susceptibility in bacteria. Front Microbiol. 2016;7:412.

18. Mondal N, Peketi A, Mapder T, Roy C, Mazumdar A, Chakraborty R, et al. Indus and Nubra Valley hot springs affirm the geomicrobiological specialties of Trans-Himalayan hydrothermal systems. J Earth Syst Sci. 2022;131:12.

19. Wang S, Hou W, Dong H, Jiang H, Huang L, Wu G, et al. Control of temperature on microbial community structure in hot springs of the Tibetan Plateau. PLoS One. 2013;8:e62901.

20. Zhang Y, Wu G, Jiang H, Yang J, She W, Khan I, et al. Abundant and rare microbial biospheres respond differently to environmental and spatial factors in Tibetan hot springs. Front in Microbiol. 2018;9:2096.

21. Das S, Sherpa MT, Najar IN, Thakur N. Prevalence of methanogens in the uncultured Sikkim hot spring solfataric mud archaeal microbiome. Environ Sustain. 2020;3:453-469.

22. Ghosh W, Roy C, Roy R, Nilawe P, Mukherjee A, Haldar PK, et al. Resilience and receptivity worked in tandem to sustain a geothermal mat community amidst erratic environmental conditions. Sci Rep. 2015;5:12179.

23. Guo L, Wang G, Sheng Y, Sun X, Shi Z, Xu Q, et al. Temperature governs the distribution of hot spring microbial community in three hydrothermal fields, Eastern Tibetan Plateau Geothermal Belt, Western China. Sci Total Environ. 2020;720:137574.

24. Kumar V, Kumar S, Singh D. Metagenomic insights into Himalayan glacial and kettle lake sediments revealed microbial community structure, function, and stress adaptation strategies. Extremophiles. 2022;26:3.

25. Nagar S, Talwar C, Motelica-Heino M, Richnow HH, Shakarad M, Lal R, et al. Microbial Ecology of Sulfur Biogeochemical Cycling at a Mesothermal Hot Spring Atop Northern Himalayas, India. Front Microbiol. 2022;13:848010.

26. Zhang P, Xiong J, Qiao N, Luo S, Yang Q, Li X, et al. High Variation in Protist Diversity and Community Composition in Surface Sediment of Hot Springs in Himalayan Geothermal Belt, China. Microorganisms. 2023;11:674.

27. Roy C, Bakshi U, Rameez MJ, Mandal S, Haldar PK, Pyne P, et al. Phylogenomics of an uncultivated, aerobic and thermophilic, photoheterotrophic member of Chlorobia sheds light into the evolution of the phylum Chlorobi. Comput Biol Chem. 2019;80:206-216.

28. Mondal N, Roy C, Chatterjee S, Sarkar J, Dutta S, Bhattacharya S, et al. Thermal endurance by a hot-spring-dwelling phylogenetic relative of the mesophilic *Paracoccus*. Microbiol Spectr. 2022;10:e01606-22.

29. Steller LH, Nakamura E, Ota T, Sakaguchi C, Sharma M, Van Kranendonk MJ. Boron isotopes in the Puga geothermal system, India, and their implications for the habitat of early life. Astrobiology. 2019;19:1459-1473.

30. Sarkar S, Moitra H, Bhattacharya S, Dagar A, Ray D, Gupta S, et al. Spectroscopic studies on the Puga Hot Spring Deposits, Ladakh, an astrobiological Martian analog site in India. J Geophys Res Planets. 2022;127:e2022JE007299.

31. Harinarayana T, Azeez KA, Murthy DN, Veeraswamy K, Rao SE, Manoj C, et al. Exploration of geothermal structure in Puga geothermal field, Ladakh Himalayas, India by magnetotelluric studies. J Appl Geophys. 2006;58:280-295.

32. Quan X, Nitta K, Matoba O, Xia P, Awatsuji Y. Phase and fluorescence imaging by combination of digital holographic microscopy and fluorescence microscopy. Opt Rev. 2015;22:349-353.

33. Absher M. Hemocytometer counting. In Kruse PF, Patterson MK (eds). Tissue culture. Academic Press, 1973. pp 395-397.

34. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674-1676.

35. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.

36. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309-D314.

37. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38:5825-5829.

38. Stoddard SF, Smith BJ, Hein R, Roller BR, Schmidt TM. rrn DB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res. 2015;43:D593-D598.

39. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357-359.

40. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590-D596.

41. Seemann T. barrnap 0.9: rapid ribosomal RNA prediction. 2018. https://github.com/tseemann/barrnap.

42. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823-1829.

43. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165.

44. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605-607.

45. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11:1144-1146.

46. Sieber CM, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018;3:836-843.

47. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043-1055.

48. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020;36:1925-1927.

49. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genom. 2008;9:75.

50. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res. 2022;50:D801-D807.

51. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110:1281-1286.

52. Na SI, Kim YO, Yoon SH, Ha SM, Baek I, Chun J. UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. J Microbiology. 2018;56:280-285.

53. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol. 2020; *37*:291-294.

54. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30:1312-1313.

55. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47:W256-W259.

56. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35:W182-W185.

57. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013;10:996-998.

58. Evans RN, Blaha G, Bailey S, Steitz TA. The structure of LepA, the ribosomal back translocase. Proc Natl Acad Sci USA. 2008;105:4673-4678.

59. Heller JL, Kamalampeta R, Wieden HJ. Taking a step back from back-translocation: an integrative view of LepA/EF4's cellular function. Mol Cell Biol. 2017;37:e00653-16.

60. Campanaro S, Treu L, Kougias PG, Zhu X, Angelidaki I. Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies. Sci Rep. 2018;8:1926.

61. Laudadio I, Fulci V, Palone F, Stronati L, Cucchiara S, Carissimi C. Quantitative assessment of shotgun metagenomics and 16S rDNA amplicon sequencing in the study of human gut microbiome. OMICS J Integr Biol. 2018;22:248-254.

62. Brumfield KD, Huq A, Colwell RR, Olds JL, Leddy MB. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. PLoS One. 2020;15:e0228899.

63. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. Sci Rep. 2021;11:3030.

64. Tessler M, Neumann JS, Afshinnekoo E, Pineda M, Hersch R, Velho LF, et al. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. Sci Rep. 2017;7:6589.

65. Peterson D, Bonham KS, Rowland S, Pattanayak CW, Resonance Consortium, Klepac-Ceraj V. Comparative analysis of 16S rRNA gene and metagenome sequencing in pediatric gut microbiomes. Front Microbiol. 2021;12:670336.

66. Huber R, Wilharm T, Huber D, Trincone A, Burggraf S, König H, et al. *Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. Syst Appl Microbiol. 1992;15:340-351.

67. O'Neill AH, Liu Y, Ferrera I, Beveridge TJ, Reysenbach AL. *Sulfurihydrogenibium rodmanii* sp. nov., a sulfur-oxidizing chemolithoautotroph from the Uzon Caldera, Kamchatka Peninsula, Russia, and emended description of the genus *Sulfurihydrogenibium*. Int J Syst Evol Microbiol. 2008;58:1147-1152.

68. Cai J, Wang Y, Liu D, Zeng Y, Xue Y, Ma Y, et al. *Fervidobacterium changbaicum* sp. nov., a novel thermophilic anaerobic bacterium isolated from a hot spring of the Changbai Mountains, China. Int J Syst Evol Microbiol. 2007;57:2333-2336.

69. Chung AP, Rainey FA, Valente M, Nobre MF, da Costa MS. *Thermus igniterrae* sp. nov. and *Thermus antranikianii* sp. nov., two new species from Iceland. Int J Syst Evol Microbiol. 2000;50:209-217.

70. Kashefi K, Holmes DE, Reysenbach AL, Lovley DR. Use of Fe (III) as an electron acceptor to recover previously uncultured hyperthermophiles: isolation and characterization of *Geothermobacterium ferrireducens* gen. nov., sp. nov. Appl Environ Microbiol. 2002;68:1735-1742.

71. Stetter KO, König H, Stackebrandt E. *Pyrodictium* gen. nov., a new genus of submarine disc-shaped sulphur reducing archaebacteria growing optimally at 105°C. Syst Appl Microbiol. 1983;4:535-551.

72. Copeland JJ. Yellowstone thermal myxophyceae. Ann N Y Acad Sci. 1936;36:4-223.

73. Ward DM, Castenholz RW, Miller SR, Whitton B. Cyanobacteria in geothermal habitats. In: Whitton BA (ed). Ecology of cyanobacteria II: their diversity in space and time. Springer Science and Business media, 2012. pp 39-63.

74. Katoh H, Itoh S, Shen JR, Ikeuchi M. Functional analysis of psbV and a novel c-type cytochrome gene psbV2 of the thermophilic cyanobacterium *Thermosynechococcus elongatus* strain BP-1. Plant Cell Physiol. 2001;42:599-607.

75. Komárek J, Johansen JR, Šmarda J, Strunecký O. Phylogeny and taxonomy of *Synechococcus*-like cyanobacteria. Fottea. 2020;20:171-191.

76. Yamaoka T, Satoh K, Katoh S. Photosynthetic activities of a thermophilic blue-green alga. Plant Cell Physiol. 1978;19:943-954.

77. Liang Y, Tang J, Luo Y, Kaczmarek MB, Li X, Daroch M. *Thermosynechococcus* as a thermophilic photosynthetic microbial cell factory for CO_2 utilisation. Bioresour Technol. 2019;278:255-265.

78. Guan TW, Xiao J, Zhao K, Luo XX, Zhang XP, Zhang LL. *Halomonas xinjiangensis* sp. nov., a halotolerant bacterium isolated from a salt lake. Int J Syst Evol Microbiol. 2010;60:349-352.

79. Sun X, Luo P, Li M. *Paracoccus angustae* sp. nov., isolated from soil. Int J Syst Evol Microbiol. 2015;65:3469-3475.

80. Chen WM, Huang HW, Chang JS, Han YL, Guo TR, Sheu SY. *Tepidimonas fonticaldi* sp. nov., a slightly thermophilic betaproteobacterium isolated from a hot spring. Int J Syst Evol Microbiol. 2013;63:1810-1816.

81. Mori K, Suzuki KI. *Thiofaba tepidiphila* gen. nov., sp. nov., a novel obligately chemolithoautotrophic, sulfur-oxidizing bacterium of the Gammaproteobacteria isolated from a hot spring. Int J Syst Evol Microbiol. 2008;58:1885-1891.

82. Meng YC, Liu HC, Zhou YG, Cai M, Kang Y. *Vibrio gangliei* sp. nov., a novel member of Vibrionaceae isolated from sawdust in a pigpen. Int J Syst Evol Microbiol. 2018;68:1969-1974.

83. Kaye JZ, Marquez MC, Ventosa A, Baross JA. *Halomonas neptunia* sp. nov., *Halomonas sulfidaeris* sp. nov., *Halomonas axialensis* sp. nov. and *Halomonas hydrothermalis* sp. nov.: halophilic bacteria isolated from deep-sea hydrothermal-vent environments. Int J Syst Evol Microbiol. 2004;54:499-511.

84. Hasan NA, Grim CJ, Lipp EK, Rivera IN, Chun J, Haley BJ, et al. Deep-sea hydrothermal vent bacteria related to human pathogenic *Vibrio* species. Proc Natl Acad Sci USA. 2015;112:E2813-E2819.

85. Du R, Gao D, Wang Y, Liu L, Cheng J, Liu J, et al. Heterotrophic sulfur oxidation of *Halomonas titanicae* SOB56 and its habitat adaptation to the hydrothermal environment. Front Microbiol. 2022;13:888833.

86. Ghosh W, Mallick S, DasGupta SK. Origin of the Sox multienzyme complex system in ancient thermophilic bacteria and coevolution of its constituent proteins. Res Microbiol. 2009;160:409-420.

87. Kitadai N, Kameya M, Fujishima K. Origin of the reductive tricarboxylic acid (rTCA) cycle-type CO_2 fixation: A perspective. Life. 2017;7:39.

88. Ghosh W, Dam B. Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. FEMS Microbiol Rev. 2009;33:999-1043.

89. Schäfer S, Barkowski C, Fuchs G. Carbon assimilation by the autotrophic thermophilic archaebacterium *Thermoproteus neutrophilus*. Arch Microbiol. 1986;146:301-308.

90. Beh M, Strauss G, Huber R, Stetter KO, Fuchs G. Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaebacterium *Thermoproteus neutrophilus*. Arch Microbiol. 1993;160:306-311.

91. Campbell BJ, Cary SC. Abundance of reverse tricarboxylic acid cycle genes in free-living microorganisms at deep-sea hydrothermal vents. Appl Environ Microbiol. 2004;70:6282-6289.

92. Hügler M, Huber H, Stetter KO, Fuchs G. Autotrophic CO_2 fixation pathways in archaea (Crenarchaeota). Arch Microbiol. 2003;179:160-173.

93. Berg IA. Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. Appl Environ Microbiol. 2011;77:1925-1936.

94. Heyduk K. Evolution of Crassulacean acid metabolism in response to the environment: past, present, and future. Plant Physiol. 2022;190:19-30.

95. Lipscomb GL, Hahn EM, Crowley AT, Adams MW. Reverse gyrase is essential for microbial growth at 95°C. Extremophiles. 2017;21:603-608.

96. Kampmann M, Stock D. Reverse gyrase has heat-protective DNA chaperone activity independent of supercoiling. Nucleic Acids Res. 2004;32:3537-3545.

97. Murayama N, Shimizu H, Takiguchi S, Baba Y, Amino H, Horiuchi T, et al. Evidence for involvement of *Escherichia coli* genes pmbA, csrA and a previously unrecognized gene tldD, in the control of DNA gyrase by letD (ccdB) of sex factor F. J Mol Biol. 1996;256:483-502.

98. Sengupta S, Nagaraja V. YacG from *Escherichia coli* is a specific endogenous inhibitor of DNA gyrase. Nucleic Acids Res. 2008;36:4310-4316.

99. Page R, Peti W. Toxin-antitoxin systems in bacterial growth arrest and persistence. Nat Chem Biol. 2016;12:208-214.

100. Mesle MM, Beam JP, Jay ZJ, Bodle B, Bogenschutz E, Inskeep WP. Hydrogen peroxide cycling in high-temperature acidic geothermal springs and potential implications for oxidative stress response. Front Mar Sci. 2017;4:130.

101. Mulkidjanian AY, Bychkov AY, Dibrova DV, Galperin MY, Koonin EV. Origin of first cells at terrestrial, anoxic geothermal fields. Proc Natl Acad Sci USA. 2012;109:E821-E830.

102. Sherpa MT, Das S, Thakur N. Physicochemical analysis of hot water springs of Sikkim-Polok tatopani, borong tatopani and Reshi tatopani. Recent Res Sci Technol. 2013;5:63-67.

103. Van Kranendonk MJ, Baumgartner R, Djokic T, Ota T, Steller L, Garbe U, et al. Elements for the origin of life on land: a deep-time perspective from the Pilbara Craton of Western Australia. Astrobiology. 2021;21:39-59.

104. Jawadi HA, Malistani HA, Moheghy MA, Sagin J. Essential trace elements and arsenic in thermal springs, Afghanistan. Water. 2021;13:134.

105. Smith B, Oliver JD. In situ and in vitro gene expression by *Vibrio vulnificus* during entry into, persistence within, and resuscitation from the viable but nonculturable state. Appl Environ Microbiol. 2006;72:1445-1451.

106. Kent AG, Garcia CA, Martiny AC. Increased biofilm formation due to high-temperature adaptation in marine *Roseobacter*. Nat Microbiol. 2018;3:989-995.

107. Alvarez-Martinez CE, Sgro GG, Araujo GG, Paiva MR, Matsuyama BY, Guzzo CR, et al. Secrete or perish: the role of secretion systems in *Xanthomonas* biology. Comput Struct Biotechnol J. 2021 Jan 1;19:279-302.

108. Naskar S, Hohl M, Tassinari M, Low HH. The structure and mechanism of the bacterial type II secretion system. Mol Microbiol. 2021;115:412-424.

109. Liu X, Luo Y, Li Z, Wei G. Functional analysis of a putative type III secretion system in stress adaption by *Mesorhizobium alhagi* CCNWXJ12-2^{T}. Front Microbiol. 2018;9:263.

110. Banerjee G, Quan FS, Mondal AK, Sur S, Banerjee P, Chattopadhyay P. Interrelation between Stress Management and Secretion Systems of *Ralstonia solanacearum*: An In Silico Assessment. Pathogens. 2022;11:730.

111. Souza DP, Oka GU, Alvarez-Martinez CE, Bisson-Filho AW, Dunger G, Hobeika L, et al. Bacterial killing via a type IV secretion system. Nat Commun. 2015;6:6453.

112. Bayer-Santos E, Cenens W, Matsuyama BY, Oka GU, Di Sessa G, Mininel ID, et al. The opportunistic pathogen *Stenotrophomonas maltophilia* utilizes a type IV secretion system for interbacterial killing. PLoS Pathog. 2019;15:e1007651.

113. Weber B, Hasic M, Chen C, Wai SN, Milton DL. Type VI secretion modulates quorum sensing and stress response in *Vibrio anguillarum*. Environ Microbiol. 2009;11:3018-3028.

114. Yu KW, Xue P, Fu Y, Yang L. T6SS mediated stress responses for bacterial environmental survival and host adaptation. Int J Mol Sci. 2021;22:478.

115. Hufnagel DA, Evans ML, Greene SE, Pinkner JS, Hultgren SJ, Chapman MR. The catabolite repressor protein-cyclic AMP complex regulates csgD and biofilm formation in uropathogenic *Escherichia coli*. J Bacteriol. 2016;198:3329-3334.

116. Lasica AM, Ksiazek M, Madej M, Potempa J. The type IX secretion system (T9SS): highlights and recent insights into its structure and function. Front Cell Infect Microbiol. 2017;7:215.

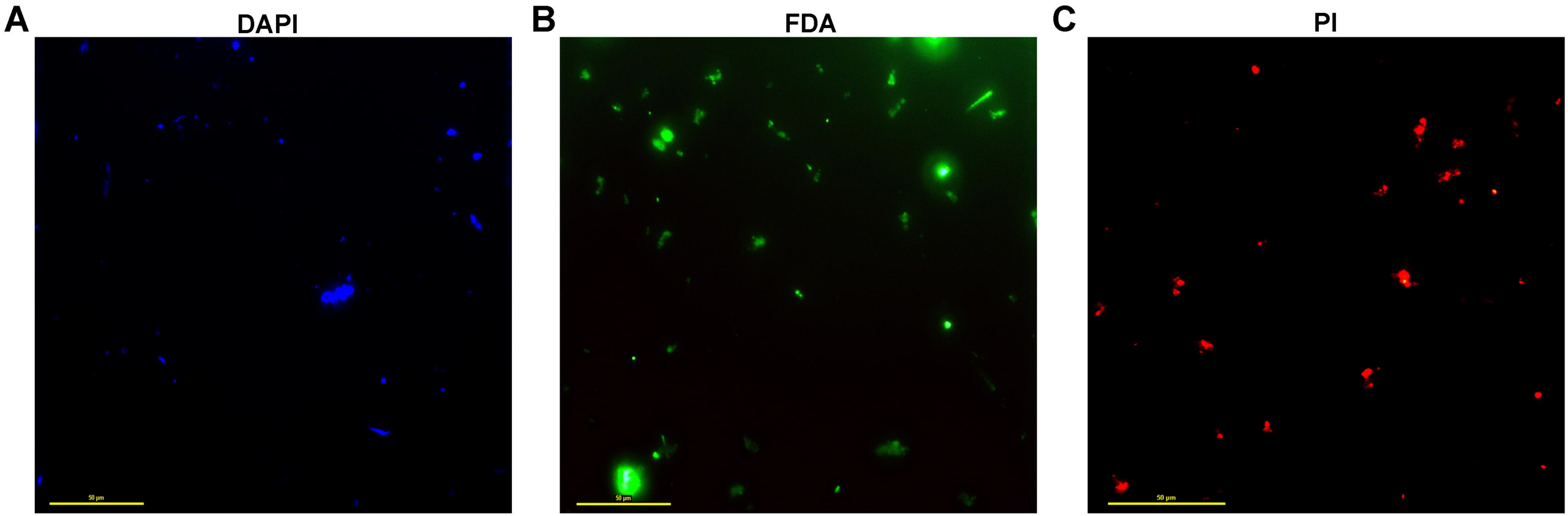
117. Pednekar P, Jain R, Mahajan G. Anti-infective potential of hot-spring bacteria. J Glob Infect Dis. 2011;3:241.

118. Mahajan GB, Balachandran L. Sources of antibiotics: Hot springs. Biochem Pharmacol. 2017;134:35-41.

119. Krause KM, Serio AW, Kane TR, Connolly LE. Aminoglycosides: an overview. Cold Spring Harbor perspectives in medicine. 2016;6:a027029.

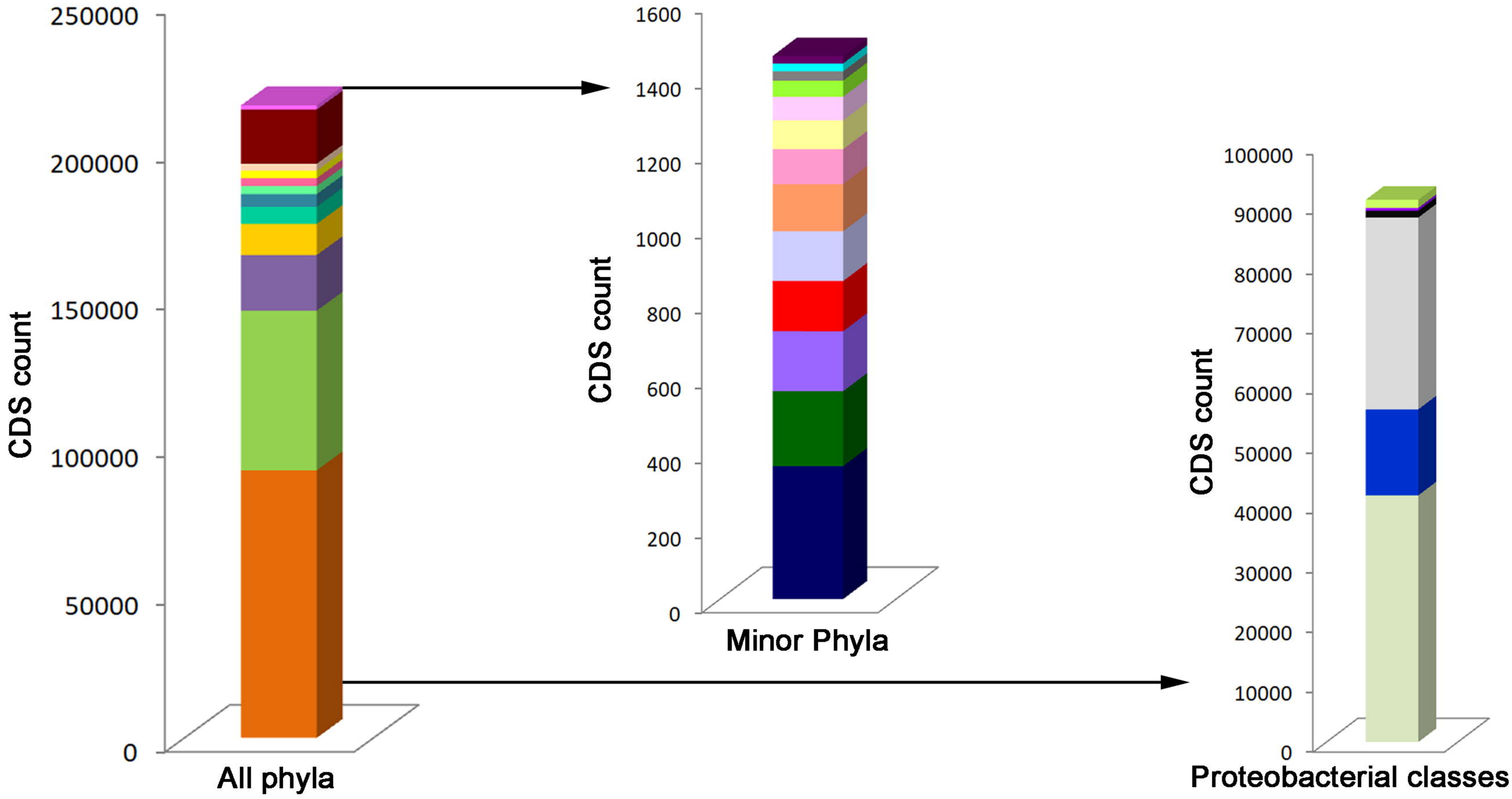
120. May JM, Owens TW, Mandler MD, Simpson BW, Lazarus MB, Sherman DJ, et al. The antibiotic novobiocin binds and activates the ATPase that powers lipopolysaccharide transport. J Am Chem Soc. 2017;139:17221-4.

A
B
C


A
DAPI
50 µm
B
FDA
50 µm
C
PI
50 µm

A

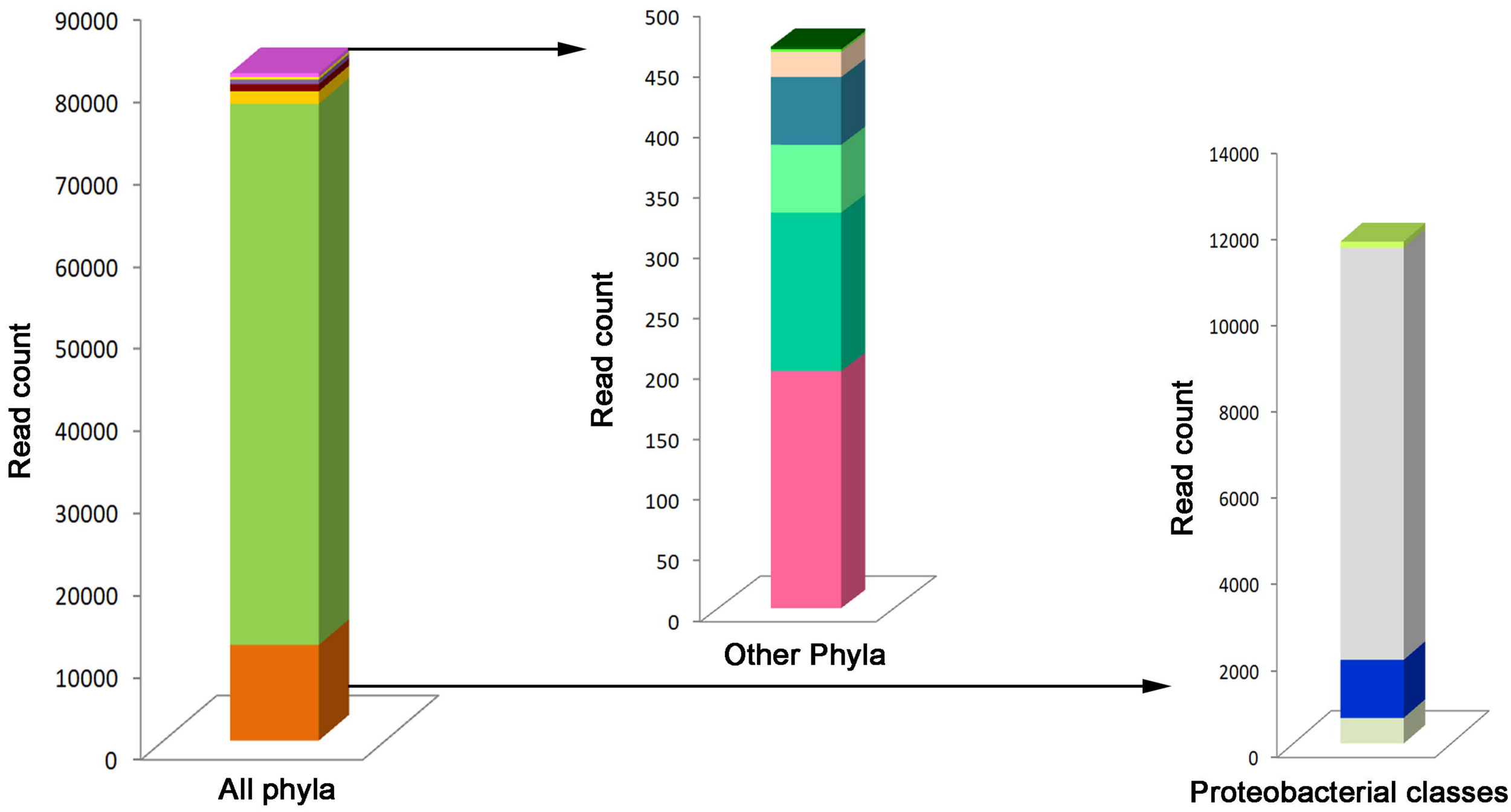
Taxonomic classification of bacterial CDSs identified in the assembled metagenome



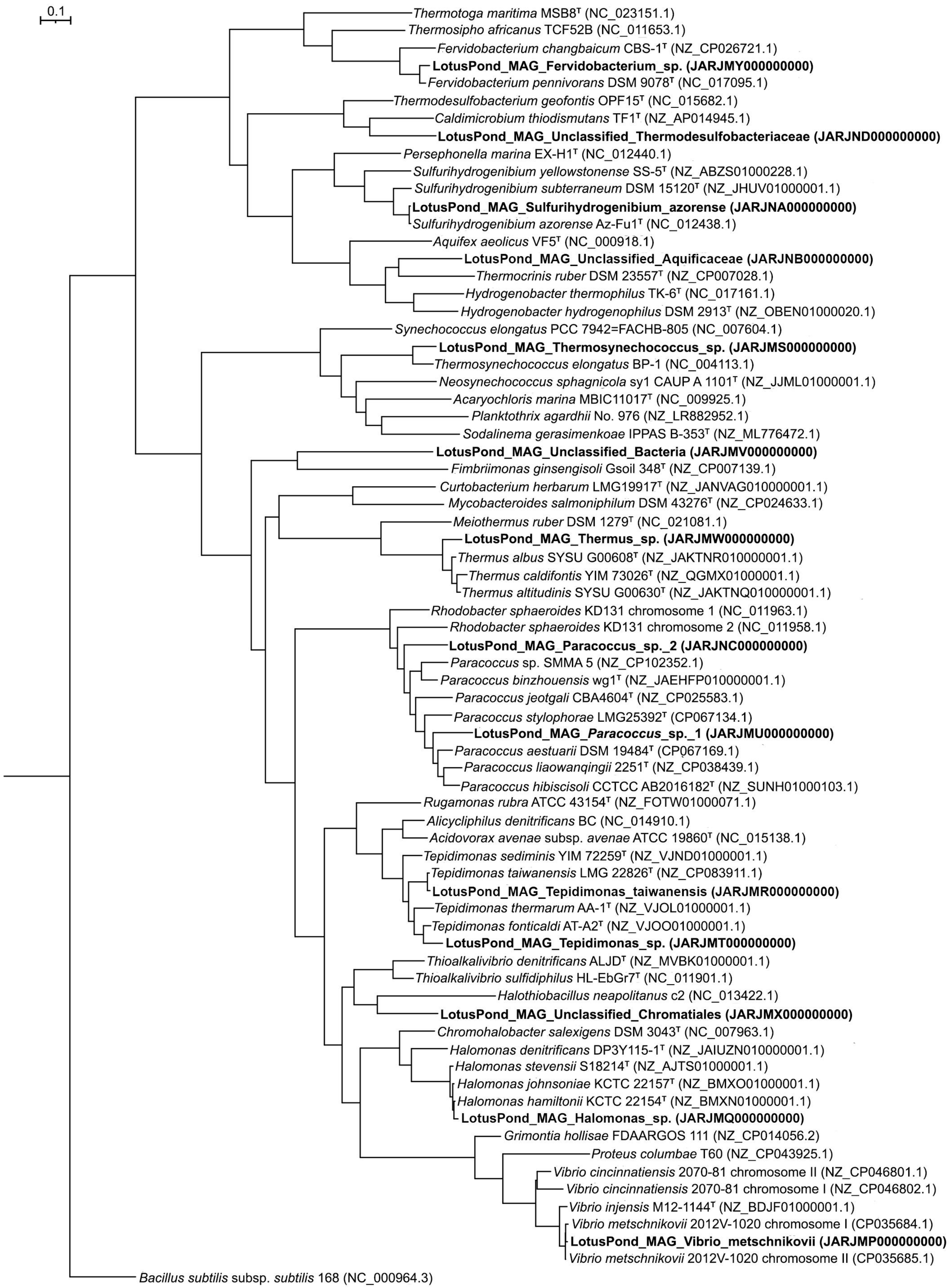
Proteobacteria (90793)
Aquificae (54151)
Firmicutes (18872)
Deinococcus-Thermus (10591)
Cyanobacteria (5827)
Chloroflexi (4254)
Bacteroidetes (2753)
Thermotogae (2671)
Thermodesulfobacteria (2547)
Actinobacteria (2357)
Unclassified Bacteria (18386)
Minor phyla (1450)
Planctomycetes (355)
Synergistetes (200)
Verrucomicrobia (160)
Acidobacteria (134)
Fusobacteria (133)
Nitrospirae (126)
Chlorobi (93)
Spirochaetes (77)
Thermomicrobia (63)
Deferribacteres (43)
Tenericutes (25)
Gemmatimonadetes (21)
Chlamydiae (20)
Alphaproteobacteria (41306)
Betaproteobacteria (14394)
Gammaproteobacteria (32111)
Deltaproteobacteria (1145)
Epsilonproteobacteria (466)
Other Proteobacteria (1371)

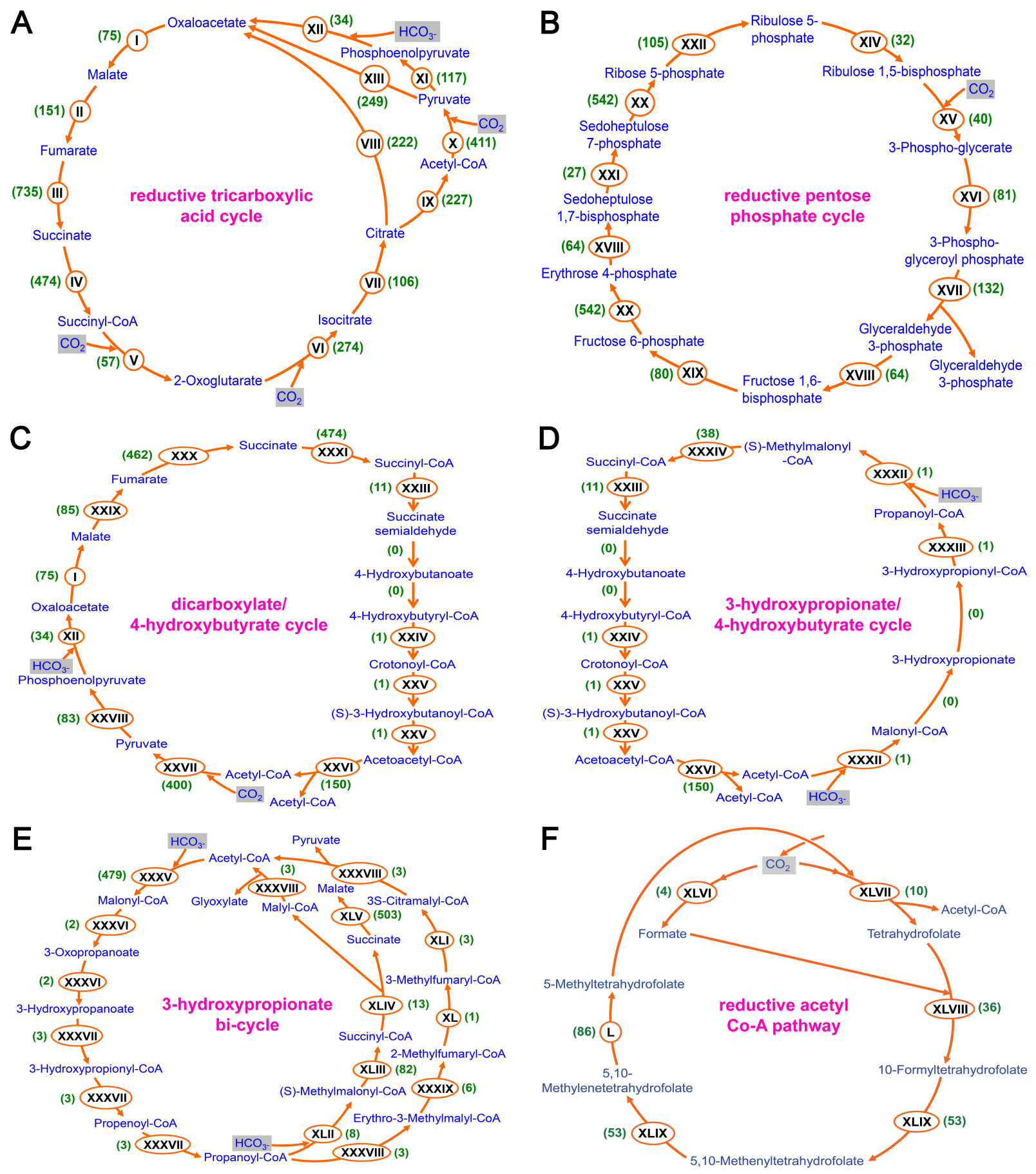
B

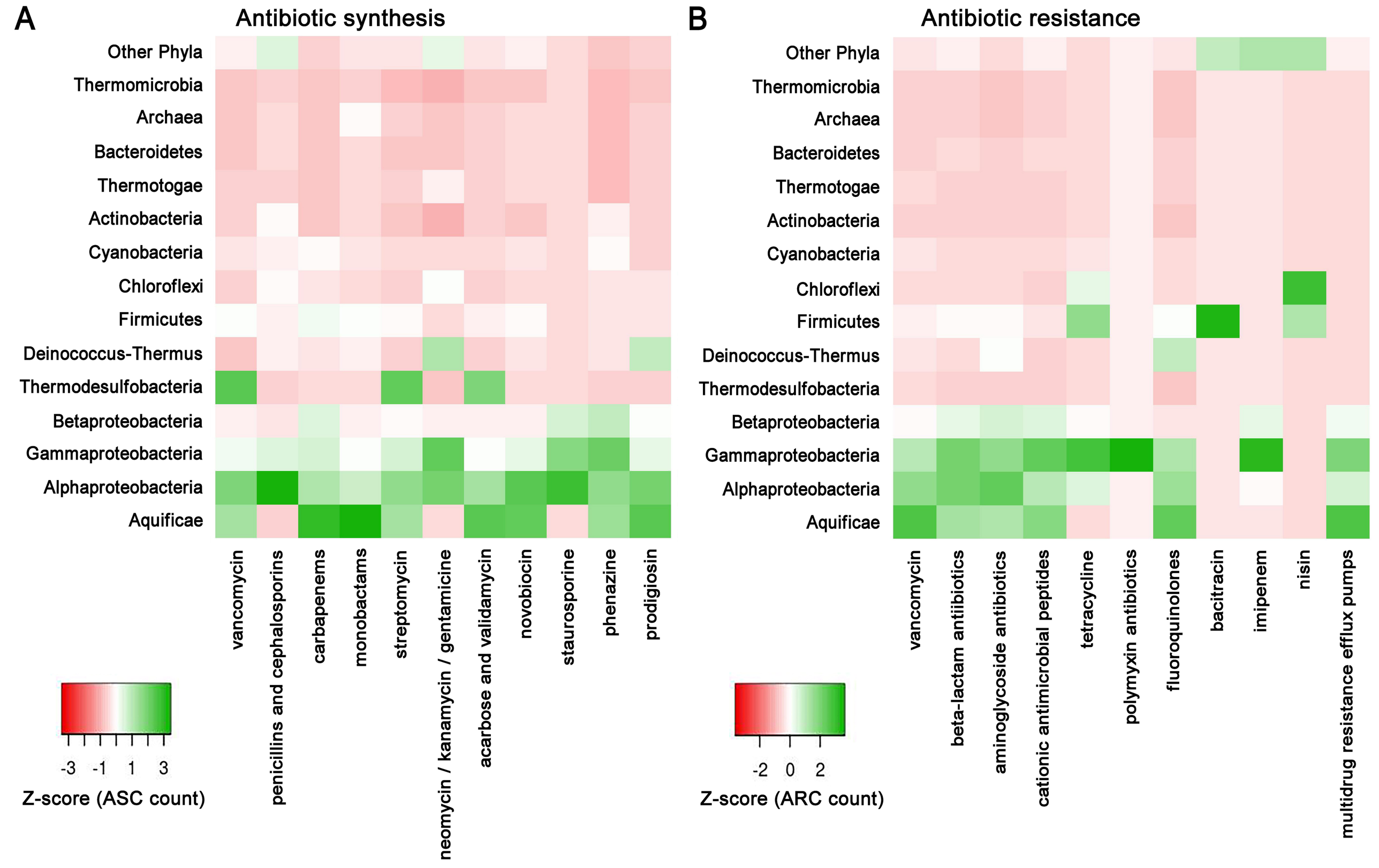
Taxonomic classification of metagenomic reads which corresponded to bacterial 16S rRNA genes



Proteobacteria (11639)
Aquificae (65853)
Deinococcus-Thermus (1517)
Unclassified Bacteria (875)
Firmicutes (567)
Thermodesulfobacteria (308)
Other Phyla (464)
Thermotogae (196)
Cyanobacteria (131)
Bacteroidetes (56)
Chloroflexi (56)
Actinobacteria (21)
Armatimonadetes (2)
Synergistetes (2)
Alphaproteobacteria (595)
Betaproteobacteria (1345)
Gammaproteobacteria (9553)
Other Proteobacteria (146)


0.1
Thermotoga maritima MSB8T (NC_023151.1)
Thermosipho africanus TCF52B (NC_011653.1)
Fervidobacterium changbaicum CBS-1T (NZ_CP026721.1)
LotusPond_MAG_Fervidobacterium_sp. (JARJMY000000000)
Fervidobacterium pennivorans DSM 9078T (NC_017095.1)
Thermodesulfobacterium geofontis OPF15T (NC_015682.1)
Caldimicrobium thiodismutans TF1T (NZ_AP014945.1)
LotusPond_MAG_Unclassified_Thermodesulfobacteriaceae (JARJND000000000)
Persephonella marina EX-H1T (NC_012440.1)
Sulfurihydrogenibium yellowstonense SS-5T (NZ_ABZS01000228.1)
Sulfurihydrogenibium subterraneum DSM 15120T (NZ_JHUV01000001.1)
LotusPond_MAG_Sulfurihydrogenibium_azorense (JARJNA000000000)
Sulfurihydrogenibium azorense Az-Fu1T (NC_012438.1)
Aquifex aeolicus VF5T (NC_000918.1)
LotusPond_MAG_Unclassified_Aquificaceae (JARJNB000000000)
Thermocrinis ruber DSM 23557T (NZ_CP007028.1)
Hydrogenobacter thermophilus TK-6T (NC_017161.1)
Hydrogenobacter hydrogenophilus DSM 2913T (NZ_OBEN01000020.1)
Synechococcus elongatus PCC 7942=FACHB-805 (NC_007604.1)
LotusPond_MAG_Thermosynechococcus_sp. (JARJMS000000000)
Thermosynechococcus elongatus BP-1 (NC_004113.1)
Neosynechococcus sphagnicola sy1 CAUP A 1101T (NZ_JJML01000001.1)
Acaryochloris marina MBIC11017T (NC_009925.1)
Planktothrix agardhii No. 976 (NZ_LR882952.1)
Sodalinema gerasimenkoae IPPAS B-353T (NZ_ML776472.1)
LotusPond_MAG_Unclassified_Bacteria (JARJMV000000000)
Fimbriimonas ginsengisoli Gsoil 348T (NZ_CP007139.1)
Curtobacterium herbarum LMG19917T (NZ_JANVAG010000001.1)
Mycobacteroides salmoniphilum DSM 43276T (NZ_CP024633.1)
Meiothermus ruber DSM 1279T (NC_021081.1)
LotusPond_MAG_Thermus_sp. (JARJMW000000000)
Thermus albus SYSU G00608T (NZ_JAKTNR010000001.1)
Thermus caldifontis YIM 73026T (NZ_QGMX01000001.1)
Thermus altitudinis SYSU G00630T (NZ_JAKTNQ010000001.1)
Rhodobacter sphaeroides KD131 chromosome 1 (NC_011963.1)
Rhodobacter sphaeroides KD131 chromosome 2 (NC_011958.1)
LotusPond_MAG_Paracoccus_sp._2 (JARJNC000000000)
Paracoccus sp. SMMA 5 (NZ_CP102352.1)
Paracoccus binzhouensis wg1T (NZ_JAEHFP010000001.1)
Paracoccus jeotgali CBA4604T (NZ_CP025583.1)
Paracoccus stylophorae LMG25392T (CP067134.1)
LotusPond_MAG_Paracoccus_sp._1 (JARJMU000000000)
Paracoccus aestuarii DSM 19484T (CP067169.1)
Paracoccus liaowanqingii 2251T (NZ_CP038439.1)
Paracoccus hibiscisoli CCTCC AB2016182T (NZ_SUNH01000103.1)
Rugamonas rubra ATCC 43154T (NZ_FOTW01000071.1)
Alicycliphilus denitrificans BC (NC_014910.1)
Acidovorax avenae subsp. avenae ATCC 19860T (NC_015138.1)
Tepidimonas sediminis YIM 72259T (NZ_VJND01000001.1)
Tepidimonas taiwanensis LMG 22826T (NZ_CP083911.1)
LotusPond_MAG_Tepidimonas_taiwanensis (JARJMR000000000)
Tepidimonas thermarum AA-1T (NZ_VJOL01000001.1)
Tepidimonas fonticaldi AT-A2T (NZ_VJOO01000001.1)
LotusPond_MAG_Tepidimonas_sp. (JARJMT000000000)
Thioalkalivibrio denitrificans ALJDT (NZ_MVBK01000001.1)
Thioalkalivibrio sulfidiphilus HL-EbGr7T (NC_011901.1)
Halothiobacillus neapolitanus c2 (NC_013422.1)
LotusPond_MAG_Unclassified_Chromatiales (JARJMX000000000)
Chromohalobacter salexigens DSM 3043T (NC_007963.1)
Halomonas denitrificans DP3Y115-1T (NZ_JAIUZN010000001.1)
Halomonas stevensii S18214T (NZ_AJTS01000001.1)
Halomonas johnsoniae KCTC 22157T (NZ_BMXO01000001.1)
Halomonas hamiltonii KCTC 22154T (NZ_BMXN01000001.1)
LotusPond_MAG_Halomonas_sp. (JARJMQ000000000)
Grimontia hollisae FDAARGOS 111 (NZ_CP014056.2)
Proteus columbae T60 (NZ_CP043925.1)
Vibrio cincinnatiensis 2070-81 chromosome II (NZ_CP046801.1)
Vibrio cincinnatiensis 2070-81 chromosome I (NZ_CP046802.1)
Vibrio injensis M12-1144T (NZ_BDJF01000001.1)
Vibrio metschnikovii 2012V-1020 chromosome I (CP035684.1)
LotusPond_MAG_Vibrio_metschnikovii (JARJMP000000000)
Vibrio metschnikovii 2012V-1020 chromosome II (CP035685.1)
Bacillus subtilis subsp. subtilis 168 (NC_000964.3)


A
reductive tricarboxylic acid cycle
Oxaloacetate
Malate
Fumarate
Succinate
Succinyl-CoA
2-Oxoglutarate
Isocitrate
Citrate
Acetyl-CoA
Pyruvate
Phosphoenolpyruvate
HCO3-
CO2
(75) I
(151) II
(735) III
(474) IV
(57) V
(274) VI
(106) VII
(222) VIII
(227) IX
(411) X
(117) XI
(34) XII
(249) XIII
B
reductive pentose phosphate cycle
Ribulose 5-phosphate
Ribulose 1,5-bisphosphate
3-Phospho-glycerate
3-Phospho-glyceroyl phosphate
Glyceraldehyde 3-phosphate
Glyceraldehyde 3-phosphate
Fructose 1,6-bisphosphate
Fructose 6-phosphate
Erythrose 4-phosphate
Sedoheptulose 1,7-bisphosphate
Sedoheptulose 7-phosphate
Ribose 5-phosphate
CO2
(105) XXII
(32) XIV
(40) XV
(81) XVI
(132) XVII
(64) XVIII
(80) XIX
(542) XX
(27) XXI
(542) XX
C
dicarboxylate/ 4-hydroxybutyrate cycle
Succinate
Succinyl-CoA
Succinate semialdehyde
4-Hydroxybutanoate
4-Hydroxybutyryl-CoA
Crotonoyl-CoA
(S)-3-Hydroxybutanoyl-CoA
Acetoacetyl-CoA
Acetyl-CoA
Acetyl-CoA
Pyruvate
Phosphoenolpyruvate
Oxaloacetate
Malate
Fumarate
HCO3-
CO2
(474) XXXI
(11) XXIII
(0)
(0)
(1) XXIV
(1) XXV
(1) XXV
(150) XXVI
(400) XXVII
(83) XXVIII
(34) XII
(75) I
(85) XXIX
(462) XXX
D
3-hydroxypropionate/ 4-hydroxybutyrate cycle
(S)-Methylmalonyl -CoA
Succinyl-CoA
Succinate semialdehyde
4-Hydroxybutanoate
4-Hydroxybutyryl-CoA
Crotonoyl-CoA
(S)-3-Hydroxybutanoyl-CoA
Acetoacetyl-CoA
Acetyl-CoA
Acetyl-CoA
Malonyl-CoA
3-Hydroxypropionate
3-Hydroxypropionyl-CoA
Propanoyl-CoA
HCO3-
HCO3-
(38) XXXIV
(11) XXIII
(0)
(0)
(1) XXIV
(1) XXV
(1) XXV
(150) XXVI
(1) XXXII
(0)
(0)
(1) XXXIII
(1) XXXII
E
3-hydroxypropionate bi-cycle
HCO3-
Acetyl-CoA
Pyruvate
Malonyl-CoA
Glyoxylate
Malyl-CoA
Malate
3S-Citramalyl-CoA
Succinate
3-Methylfumaryl-CoA
Succinyl-CoA
2-Methylfumaryl-CoA
(S)-Methylmalonyl-CoA
Erythro-3-Methylmalyl-CoA
3-Oxopropanoate
3-Hydroxypropanoate
3-Hydroxypropionyl-CoA
Propenoyl-CoA
Propanoyl-CoA
HCO3-
(479) XXXV
(3) XXXVIII
(3) XXXVIII
(503) XLV
(3) XLI
(13) XLIV
(1) XL
(82) XLIII
(6) XXXIX
(8) XLII
(3) XXXVIII
(2) XXXVI
(2) XXXVI
(3) XXXVII
(3) XXXVII
(3) XXXVII
F
CO2
Formate
Acetyl-CoA
Tetrahydrofolate
5-Methyltetrahydrofolate
10-Formyltetrahydrofolate
5,10-Methylenetetrahydrofolate
5,10-Methenyltetrahydrofolate
reductive acetyl Co-A pathway
(4) XLVI
(10) XLVII
(36) XLVIII
(86) L
(53) XLIX
(53) XLIX


A
Antibiotic synthesis
Other Phyla
Thermomicrobia
Archaea
Bacteroidetes
Thermotogae
Actinobacteria
Cyanobacteria
Chloroflexi
Firmicutes
Deinococcus-Thermus
Thermodesulfobacteria
Betaproteobacteria
Gammaproteobacteria
Alphaproteobacteria
Aquificae
vancomycin
penicillins and cephalosporins
carbapenems
monobactams
streptomycin
neomycin / kanamycin / gentamicine
acarbose and validamycin
novobiocin
staurosporine
phenazine
prodigiosin
-3 -1 1 3
Z-score (ASC count)
B
Antibiotic resistance
Other Phyla
Thermomicrobia
Archaea
Bacteroidetes
Thermotogae
Actinobacteria
Cyanobacteria
Chloroflexi
Firmicutes
Deinococcus-Thermus
Thermodesulfobacteria
Betaproteobacteria
Gammaproteobacteria
Alphaproteobacteria
Aquificae
vancomycin
beta-lactam antiibiotics
aminoglycoside antibiotics
cationic antimicrobial peptides
tetracycline
polymyxin antibiotics
fluoroquinolones
bacitracin
imipenem
nisin
multidrug resistance efflux pumps
-2 0 2
Z-score (ARC count)