

RESEARCH ARTICLE

Open Access



Prokaryotes of renowned Karlovy Vary (Carlsbad) thermal springs: phylogenetic and cultivation analysis

Tereza Smrhova¹ , Kunal Jani¹ , Petr Pajer², Gabriela Kapinusova¹ , Tomas Vylita³, Jachym Suman¹ , Michal Strejcek¹ and Ondrej Uhlík^{1*}

Abstract

Background: The extreme conditions of thermal springs constitute a unique aquatic habitat characterized by low nutrient contents and the absence of human impacts on the microbial community composition. Thus, these springs may host phylogenetically novel microorganisms with potential use in biotechnology. With this hypothesis in mind, we examined the microbial composition of four thermal springs of the world-renowned spa town of Karlovy Vary (Carlsbad), Czechia, which differ in their temperature and chemical composition.

Results: Microbial profiling using 16S rRNA gene sequencing revealed the presence of phylogenetically novel taxa at various taxonomic levels, spanning from genera to phyla. Many sequences belonged to novel classes within the phyla Hydrothermae, Altiarchaeota, Verrucomicrobia, and TA06. Cultivation-based methods employing oligotrophic media resulted in the isolation of 44 unique bacterial isolates. These include strains that withstand concentrations of up to 12% NaCl_{w/v} in cultivation media or survive a temperature of 100 °C, as well as hitherto uncultured bacterial species belonging to the genera *Thermomonas*, *Paenibacillus*, and *Cellulomonas*. These isolates harbored stress response genes that allow them to thrive in the extreme environment of thermal springs.

Conclusions: Our study is the first to analyze the overall microbial community composition of the renowned Karlovy Vary thermal springs. We provide insight into yet another level of uniqueness of these springs. In addition to their unique health benefits and cultural significance, we demonstrate that these springs harbor phylogenetically distinct microorganisms with unusual life strategies. Our findings open up avenues for future research with the promise of a deeper understanding of the metabolic potential of these microorganisms.

Keywords: Amplicon sequencing analysis, Cultivation analysis, Thermal water springs, Phylogenetic novelty

Background

Thermal springs occur all over the world, mostly in places with volcanic activity, and constitute a unique aquatic microbial habitat [1]. This uniqueness is characterized by their low nutrient content, broad geochemical gradients,

and the lack of human impact on both their chemical and microbiological composition [2]. Thus, as-yet-unexplored thermal springs represent model ecosystems for investigating subsurface microbial biogeography. Moreover, a certain analogy can be found between the conditions of such ecosystems and those postulated to have been present in early Earth [3]. Deep thermal springs can host hitherto unknown microorganisms that potentially harbor novel biologically active compounds that might be useful for a broad spectrum of biotechnological applications [4, 5]. In particular, they are great candidates

*Correspondence: ondrej.uhlik@vscht.cz

¹ Department of Biochemistry and Microbiology, Faculty of Food and Biochemical Technology, University of Chemistry and Technology, Prague, Technická 3, 166 28 Prague 6, Czech Republic
Full list of author information is available at the end of the article



for discovering compounds that are biologically active at high temperatures [6, 7]. Furthermore, exploration of these microorganisms could very well contribute to a deeper understanding of evolution and uncover some missing branches in the tree of life [3].

Culture-dependent and culture-independent methods have both been used to analyze the microbial composition of thermal springs. Culture-dependent methods have several limitations, the main one being that less than 10% of the total number of microorganisms in the environment are estimated to be routinely cultivable [8]. This limitation is primarily due to differences between laboratory and environmental conditions, the latter being often difficult to mimic [9]. However, cultivating microorganisms in pure cultures is crucial for describing their bioactive potential as well as for assigning function to genes previously identified by culture-independent methods, thus bridging the gap between culturing and culture-independent methods [10]. On the other hand, the large amount of sequencing data obtained by culture-independent methods enables the description of a much more comprehensive range of microorganisms than is possible using traditional culture-dependent approaches. Sequencing the ubiquitous prokaryotic 16S rRNA marker gene has been a cornerstone of culture-independent microbial community analyses for decades [11–13]. The major limitation of 16S rRNA gene sequence analysis is that it does not provide functional information that can be determined using metagenomic or pure culture analyses. Thus, a combination of marker gene sequence analyses and culture-dependent methods have been successfully used to analyze microbial diversity in many oligotrophic and extreme environments, including thermal springs [14–16].

Underground water systems are ubiquitous and can serve as an insight into the subsurface microscopic world [17]. Their investigation provides an opportunity to uncover unique prokaryotic taxa in many places on Earth [18–22]. Amongst the most renowned localities are the hot springs of Yellowstone National Park, where culture-independent analyses have discovered numerous hitherto uncultured microorganisms, mostly novel clades of Archaea [23]. Other widely investigated thermal springs in volcanic zones of New Zealand have been found to harbor unique microbial communities conditioned by temperature, pH, and geographical location [18, 24]. The warmest Czech thermal springs are located in the renowned spa town of Karlovy Vary (Carlsbad). This town is located where the Eger Rift meets the Karlovy Vary Thermal Spring Line, separating the Saxothuringian and Teplá-Barrandian tectonostratigraphic zones in the Bohemian Massif [25]. This intercontinental rift was created during the Neogene tectogenesis, with thermal

springs being estimated to have first emerged in the area around 230,000 years ago. Their genesis was allowed thanks to the combination of terraced bedrock and alkaline magmatism [26]. This combination led to volcanic activity which caused a release of CO₂ and results in bicarbonate-sulfate–chloride-enriched waters containing gaseous and dissolved CO₂ [27]. Due to their chemical composition, these waters are used for gastrointestinal tract therapies and other balneological procedures. The balneological attraction of this town began as early as the fourteenth century and has continued to grow since, with more than a million visitors to its spas recorded in 2019 (<https://www.karlovyvary.cz>). Despite their extensive use and recognition as a natural and cultural heritage site, only a few studies have been conducted on the microbial communities colonizing these thermal springs [28–30]. The first study by Pěčková et al. (1991) used a culture-dependent approach and was exclusively focused on one bacterial genus, *Thermus* [28]. The other two studies [29, 30] focused on the analysis of specific biomolecules from a few selected isolates. However, a complete picture of the microbial life inhabiting the Karlovy Vary thermal springs has remained missing thus far.

In this study, we provide insights into the composition of prokaryotic communities in four Karlovy Vary thermal springs. Phylogenetic analysis of 16S rRNA gene sequencing data highlights the unique composition of prokaryotic communities in each spring and reveals the presence of as-yet-undetected taxa that form a major part of the microbial communities therein. We also characterize a collection of bacterial isolates, including some novel species, which were obtained using different culturing media. To reveal the potential of the isolates for biotechnological applications, we further investigate their tolerance to NaCl and higher temperatures by observing their growth under these types of stress. Finally, a whole-genome analysis of the phylogenetically novel bacterial species is provided.

Methods

Sampling of thermal spring water

Four thermal springs differing in their chemical composition and temperature (Additional file 1: Table S1) were examined in this study, Vřídlo (V, 72.0 °C), Mlýnský (M, 59.3 °C), Sadový starý (S, 46.3 °C), and Štěpánka (P, 18.3 °C). Samples were collected at two time points, in autumn 2018 and spring 2019. A total volume of 25 L of water was collected directly from the constructed pipes into sterile 2 L and 1 L glass bottles (SIMAX, CZ). To control for the asepticity of the collection process, the fallout of possible air contamination was sampled using the same type of 2 L sampling bottle filled with sterile deionized water, which was left open during the

whole process of sampling. Collected samples and control bottles were immediately transferred to the laboratory (<4 h). Cells were filtered onto 0.22 µm membrane filters (VWR, USA), and membrane filters with retained cells from 20 L of water were used for the extraction of metagenomic DNA, whereas filters with retained cells from the 3 L of water were used for the cultivation of microorganisms. The remaining 2 L were used for the enriched cultivation approach.

Metagenomic DNA extraction

Metagenomic DNA was extracted from the filters using a DNeasy PowerWater Kit (QIAGEN, DE) according to the manufacturer's instructions. DNA concentrations were measured with a Qubit fluorometer using a dsDNA high-sensitivity assay (Invitrogen, USA) and extracted DNA was used for 16S rRNA gene sequencing.

Amplicon sequencing of 16S rRNA genes

Primers targeting the variable V4–V5 region of the 16S rRNA gene, 515F 5'-GTGYCAGCMGCNGCGG-3' and 926R 5'-CCGYCAATTYMTTTRAGTTT-3' (Fraraccio et al. [31] adapted from Parada et al. [32] and modified according to Klindworth et al. [33]), were used for the PCR. The final PCR volume was 15 µL and contained 1 × KAPA HiFi HotStart ReadyMix (Kapa Biosystems, USA), 0.3 µM of each primer, and extracted metagenomic DNA (<10 ng). The amplification was performed for 5 min of initial denaturation at 95 °C; followed by 25 cycles of: 20 s of denaturation at 98 °C, 15 s of annealing at 50 °C, and 40 s of elongation at 72 °C; and ended with final elongation for 5 min at 72 °C. Amplicons (0.5 µL) were used as templates for a second round of amplification using primers with sequencing adapters as described previously [31]. PCR was performed as described above except that the total volume was increased to 25 µL, primers were increased to 1 µM each, and the number of cycles was decreased to 10. The amplicons tagged with adapters were purified using AMPure XP Beads (Beckman Coulter, USA; bead:DNA ratio 0.8:1). Further library preparation and sequencing were performed in the DNA Core Lab of the University of Alaska Fairbanks, USA as follows: the concentration of amplicons was normalized to 1–2 ng using a SequalPrep Kit (Thermo Fisher Scientific, USA), the samples were pooled, subjected to 8-cycle PCR to add sequencing adapters, and sequenced in an Illumina MiSeq using the MiSeq Reagent Kit v3. In summary, 40 DNA samples were sequenced, including technical duplicates of: (1) each spring sample collected at two time points, (2) 4 samples of airborne contaminant control, (3) 5 blank samples, and (4) 3 mock community samples.

Sequencing data processing

Obtained sequencing data were processed in the R environment [34]. The steps of the DADA2 pipeline [35] were followed with the following exceptions: maximum expected error was set to 1 (maxEE=1) and truncation length set to 250 for the forward fastq files. Taxonomy was assigned to unique amplicon sequence variants (ASVs) using the SILVA SSU r138 database [36]. A mock community of 15 bacterial strains (Additional file 1: Table S2) was analyzed to ensure sequencing accuracy. The package phyloseq [37] was used for data handling and visualization. All MiSeq reads were deposited in the NCBI Short Read Archive under SRA study number PRJNA781448.

Five replicates of blank samples were included as controls for the entire process from DNA to amplicons and their processing. The R package Decontam [38] was used for detection and removal of contaminating ASVs, using both frequency-based and prevalence-based methods. Moreover, both eukaryotic ASVs and ASVs present in aseptic sampling controls were removed manually from the dataset. Sequences identified as contamination accounted for <0.05% of all reads. Sequencing coverage was calculated using the package iNEXT [39].

Phylogenetic novelty at the taxonomic rank class was confirmed using the script SSUnique with default settings [40]. Novel ASVs were aligned using SINA aligner [41] with the five closest references from the prokaryotic tree of life based on 16S rRNA gene. Constructed phylogenetic tree was visualized using the package ggtree [42] in R.

Microbial culturing

Two cultivation approaches were used for the isolation of indigenous microorganisms from the thermal spring waters. In the first approach, further referred to as cultivation with concentrated inoculum, cells retained from 2 L of spring water were resuspended overnight in 25 mL of filtered water from each respective spring. One hundred µL of the resulting cell suspensions was used to inoculate seven different oligotrophic media in 6 replicates (Additional file 1: Table S3). Distilled water in the media was replaced with filtrates from the corresponding thermal spring [43], and the plates were solidified using Noble Agar (1.8%). The plates were incubated at 58 °C (samples V, M, P), 44 °C (S, M, P), and 16 °C (P) for over 2 months. Cultivation temperatures of thermal spring P were selected to correspond to both those underground and at the outflow (Additional file 1: Table S1).

In the second approach, further referred to as cultivation with enriched inoculum, 2 L of unfiltered water from each thermal spring was incubated in sterile bottles for

over 2 months in the dark at 37 °C. After that, 200 µL of water was inoculated into different concentrations of Luria–Bertani liquid medium (1/1, 1/2, 1/5, 1/10, 1/20, and 1/50 medium LB). After 1 week of cultivation in liquid media at 37 °C and 130 rpm, the enriched cultures were plated on diluted solid media (Plate Count Agar) in 5 replicates for each concentration of LB medium, and incubation continued at 37 °C.

During both cultivation approaches, solid media were screened for the appearance of bacterial colonies, which were continuously re-streaked. The resulting isolates were subjected to mass spectrometric analysis as described earlier by Strejcek et al. [44].

Characterization of microbial isolates

The obtained collection of microbial cultures was identified using MALDI Biotyper[®] 3.1 (Bruker Daltonics, Germany) and dereplicated based on the similarity of protein mass spectra as described earlier [44]. Briefly, obtained spectra were clustered using a 0.9 cosine similarity cut-off; a representative isolate from each cluster was selected for further taxonomic identification either by mass spectrometry analysis [44] or 16S rRNA gene Sanger sequencing. Dereplicated isolates were tested for salt tolerance and lethal temperature. Salt tolerance in the range of NaCl 0–15_{w/v}% was monitored by a single serial dilution spotting method adapted from Thomas et al. [45]. In brief, two technical replicates of bacterial cultures, standardized to 0.5 McFarland, were serially diluted three times and 20 µL of each dilution was spotted on solid R2A plates containing varying amounts of salt. The number of colony-forming units (CFUs) was recorded after 72 h of incubation. The lethal temperature of isolates was determined by heating standardized inoculum for 11 min (1 min of temperation, 10 min of incubation) to the examined temperature. Subsequently, the inoculum was streaked on R2A solid media and the viability of isolates was confirmed by the growth of colonies after 72 h. The range of examined lethal temperatures was 50–100 °C with a 5 °C increment. Both characterizations were performed in two biological replicas. The R packages ggplot2 [46], DECIPHER [47], and ape [48] were used for the further visualization and processing of cultivation data.

Genome sequencing of phylogenetically novel isolates

To isolate genomic DNA of phylogenetically novel isolates, cultures were grown overnight in liquid LB media. Genomic DNA was isolated from cell pellets with a Pure-Link[™] Genomic DNA Mini Kit (Invitrogen[™], USA) following manufacturer's instructions. The resulting DNA was subjected to library preparation and sequencing using a Nanopore MinION instrument (FLO-MIN106

flow cell) by following the same protocols as described in Lopez-Echartea et al. [49]. Base-calling was performed by using guppy (3.2.10) [50], adapters were trimmed using Porechop (<https://github.com/rrwick/Porechop>) and reads were quality filtered using Filtlong (<https://github.com/rrwick/Filtlong>). Sequencing statistics for all novel genomes are provided in Additional file 1: Table S6. Genome assembly was carried out using Canu software (v1.7) [51]. The obtained genome assemblies were polished using Medaka (v1.4.3) (<https://github.com/nanoporetech/medaka>), which maps the Nanopore reads to the draft assembly. The quality of the genome assemblies was evaluated with CheckM [52], available on the Kbase server [53].

Genomes were annotated using Prokka [54] and the PATRIC online server [55]. Additionally, KEGG Decoder was used to evaluate the metabolic potential of the strains [56]. The 16S rRNA gene sequences were searched against the EZtaxon database. Taxonomic affiliations of potentially novel strains were deduced by retrieving genomes of the most closely related taxa and subjecting them to OrthoANI [57] and EzAAI analyses [58]. Phylogenetic relationship of each strain to its most closely related taxon was established by computing the core gene phylogenetic trees using UBCG with default settings [59]. Constructed trees were visualized using MEGAX [60]. Genomic dissimilarities between two strains belonging to the same genus were derived by employing BPGA [61]. The obtained unique genes from each strain were identified by using KofamKOALA [62] and were graphically interpreted using STAMP (v2.1.3) [63]. All genomic data generated in this study are available under the NCBI BioProject ID PRJNA772595.

Results

Amplicon sequencing of 16S rRNA genes

A total of 1,442,275 16S rRNA gene sequence reads were retrieved after DADA2 pipeline processing, averaging 180,284 sequences per sample, with a minimum of 66,149 sequences per sample. Of those, 2,430 unique ASVs were identified. While determining sequencing accuracy based on mock community analysis, we found that 98.33% ± 0.25 pp of reads were 100% identical to those expected.

The microbial composition significantly differed in the sampled springs (permutation test for CCA, $R^2=0.7484$, p value=0.007), with the proportion of archaeal ASVs increasing with increasing water temperature (Fig. 1a). In total, 68% of ASVs detected remained unclassified at the genus level. In the warmest spring, Vřídlo, unclassified ASVs comprised 97% of all reads (Additional file 1: Table S4). However, members of several well-described genera were observed amongst the 50 most abundant

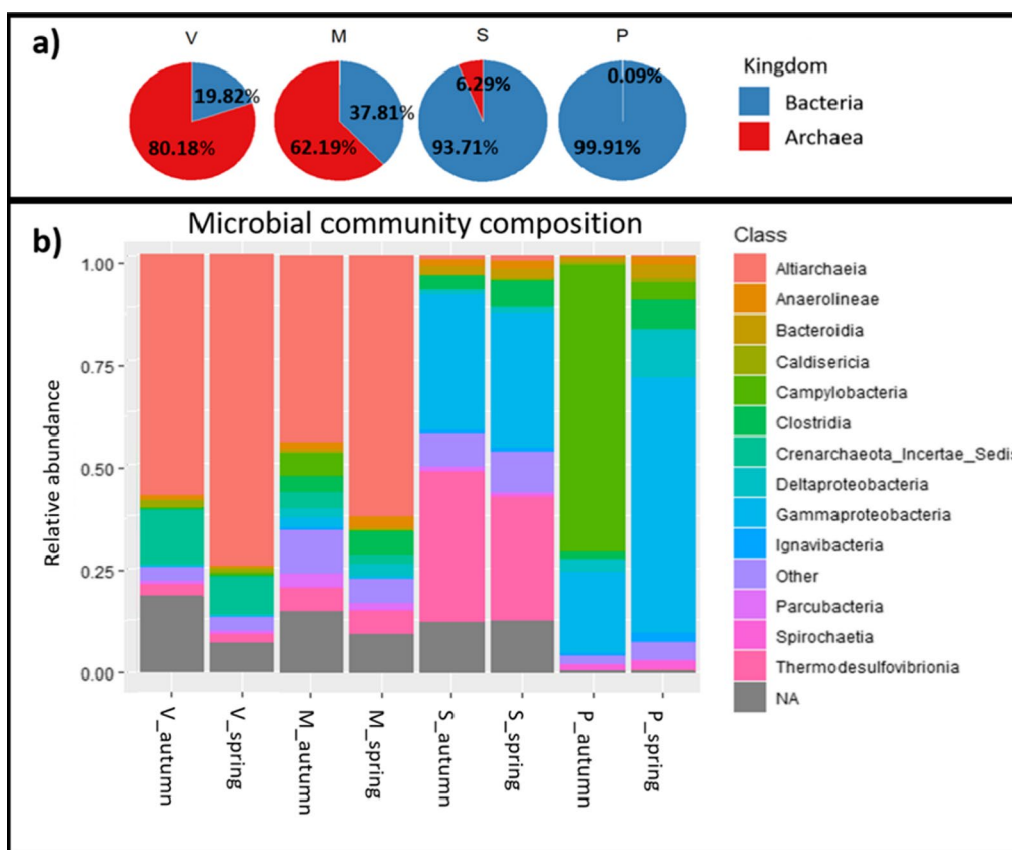


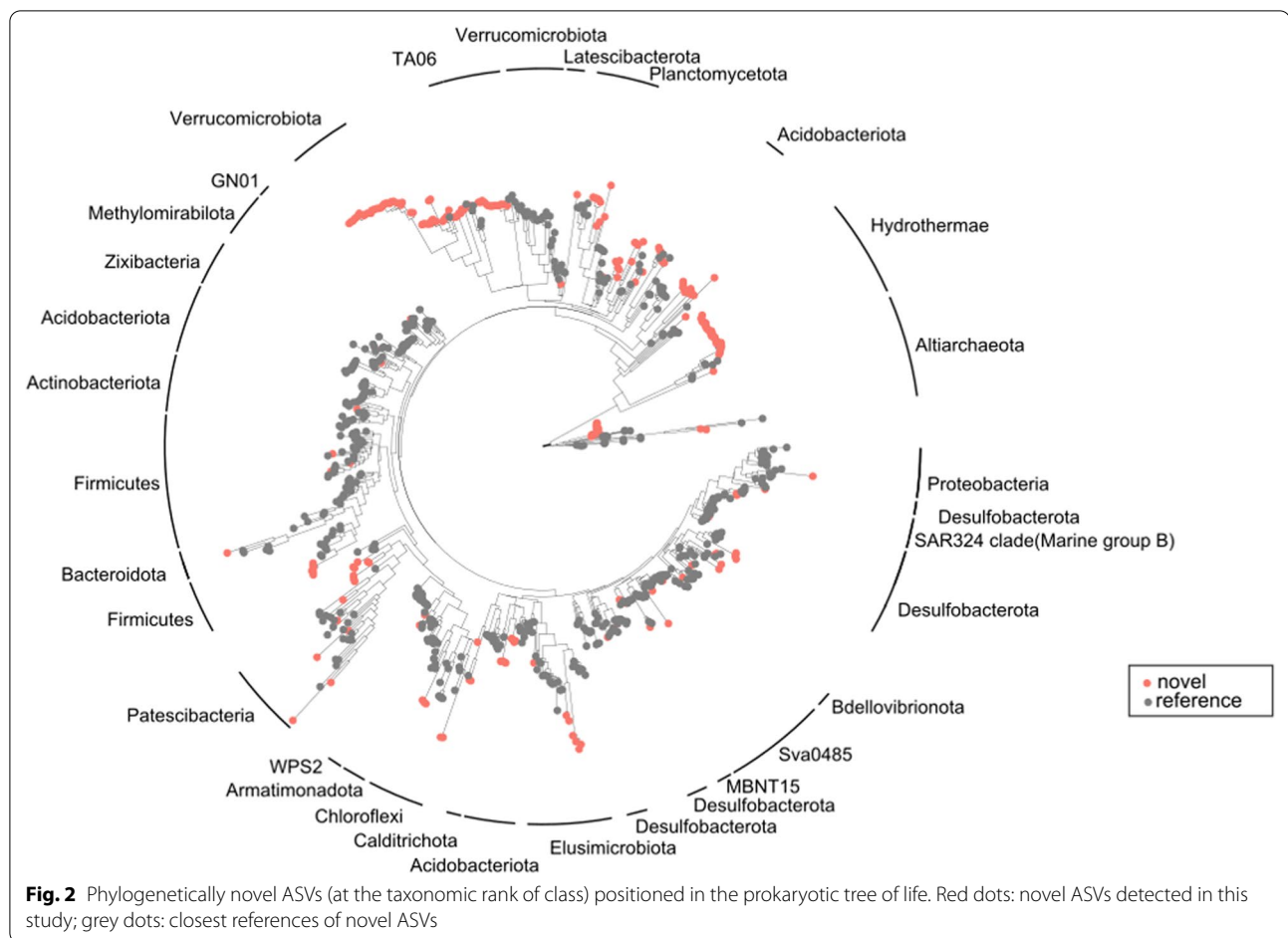
Fig. 1 Microbial community composition of the four examined Karlovy Vary thermal springs. Pie charts **a** show the fractions of ASVs belonging to Bacteria or Archaea (counted from combined spring and autumn sequencing data). Bar plots **b** show the community composition at the taxonomic level of class for each thermal spring at the two sampled time points (autumn and spring). Classes referred to as other represent all classes with a relative abundance < 0.05% per sample; NA refers to unclassified ASVs

unique ASVs from each spring. These genera included the chemoheterotrophic thermophilic genera *Thermoflexus* and *Hydrogenobacter*, the facultatively anaerobic genus *Anaerobacillus* in the warmest spring V. Genera involved in the metabolism of nitrogen, such as *Candidatus Nitrotoga* (S, P), as well as iron and sulfur, such as *Rhodofera* (P), *Sideroxydans* (S), *Desulfacinum* (P), *Thiomonas* (S), *Thermodesulfitimonas* (P), *Halothiobacillus* (P), *Sulfurimonas* (M, P), *Sulfuricurvum* (P), and *Thermoanaerobaculum* (M). The genera *Rhodofera* and *Sulfurimonas* represented over 60% of the total reads in the coolest spring (P), with *Sulfurimonas* being predominant in the autumn, and *Rhodofera* in the spring. Such inconsistency in microbial composition between the autumn and spring was only observed in thermal spring P. Despite this inconsistency, all examined thermal springs did not significantly differ in their seasonal microbial community composition (Pairwise multilevel comparison using adonis, $R^2 = 0.046$, p value = 0.125). The number of potentially novel classes of prokaryotes

that could not be classified using the SILVA_SSU_r138 database increased with increasing temperature of the thermal spring (Fig. 1b). While for the coolest spring P < 2% of reads were unclassified at the level of class, this number was 17% in springs S and M, and 19% in the warmest spring V. These reads represent potentially novel microorganisms and accounted for 226 novel bacterial ASVs and 60 archaeal ASVs (Fig. 2). In fact, several entirely novel clades of ASVs were detected across the phylogenetic tree, the largest of which were in the phyla Verrucomicrobiota, TA06, Hydrothermae (Bacteria), and Altiarchaeota (Archaea) (Fig. 2).

Cultivation with concentrated and enriched inocula

The concentrated inoculum cultivation approach with thermal water resulted in the isolation of 172 isolates before dereplication, with the distribution and yield differing based on the oligotrophic medium used (Fig. 3). The cultivation on 10× diluted R2A agar yielded the highest number of isolates. Other media with filtered

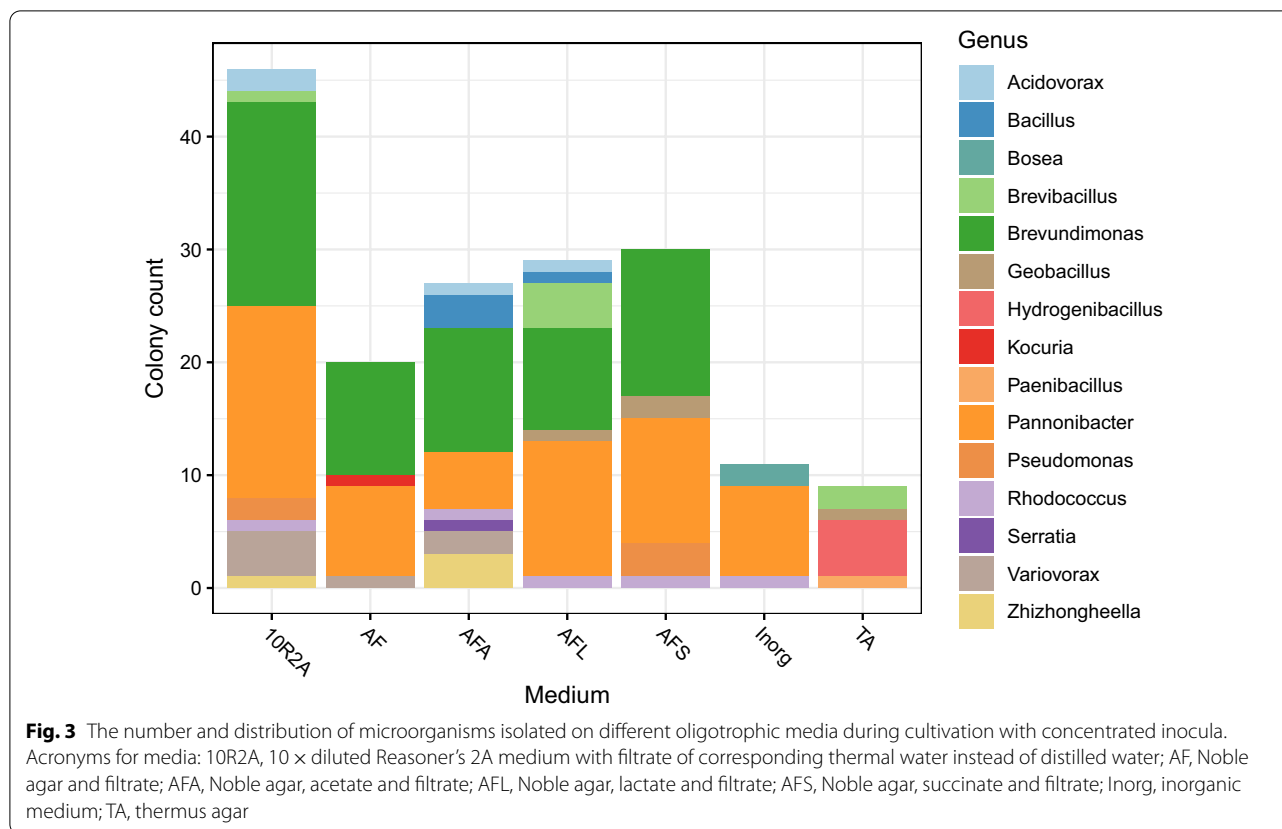


spring water yielded similar numbers of isolates, while the specialized media for autotrophs (inorganic medium) and the genus *Thermus* (thermus agar) yielded the lowest number of isolates, eleven and nine, respectively. Despite the lower number of isolates grown on these media, they provided unique cultures of the genus *Hydrogenibacillus* and *Bosea*. None of the isolates originated from thermal spring V, and members of only three genera, *Brevibacillus*, *Hydrogenibacillus* and *Paenibacillus* were isolated from thermal spring M, which was the second warmest spring. The two other thermal springs, S and P, provided more diverse isolates, although they were both dominated by one major genus, *Pannonibacter* and *Brevundimonas*, respectively (Fig. 4a). The enriched inoculum cultivation approach resulted in 76 isolates before dereplication. In contrast to the concentrated inoculum approach, seven isolates were isolated from the warmest sample V, which were identified as *Bacillus* and *Dietzia*. Furthermore, members of the genera *Rehaibacterium* (thermal spring M), *Thermomonas* (thermal spring S), *Paenibacillus* and *Cellulomonas* (thermal spring P) were isolated. Similarly to the cultivation with concentrated

inoculum, the genus *Pannonibacter* dominated the isolates from thermal spring S (Fig. 4b). All isolated strains were bacteria and were affiliated with the phyla Actinobacteria, Firmicutes, and Proteobacteria.

Characteristics of the cultures

Mass spectrometry-based dereplication of the collection of 248 isolates from both cultivation approaches resulted in the identification of 44 unique taxa (Fig. 5). The 16S rRNA gene sequence analysis of this dereplicated collection identified six potentially novel species. Isolates M2, M3, P25, and P26 were identified as members of the genus *Paenibacillus*. These four isolates had identical 16S rRNA gene sequences, but were not dereplicated due to their different origins (spring M and P) and morphologies (M2 and P25 form whitish colonies on solid media, whereas M3 and P26 form transparent colonies). The remaining two isolates, S9 and P24, were identified as members of the genera *Thermomonas* and *Cellulomonas*, respectively.



The 16S rRNA gene sequences of the dereplicated collection of microbial isolates were compared to ASVs in the amplicon data set in order to estimate the abundance of the culturable microorganisms in the springs. Thirteen sequences from the isolates were identical to sequences in the amplicon data. The remaining 30 isolate sequences did not match any ASVs despite the sequencing effort, which indicated that sample coverage was over 99% for each sample (Additional file 1: Table S5). These 30 isolate sequences mostly belonged to the order Bacillales, which is known for its ability to form endospores [64].

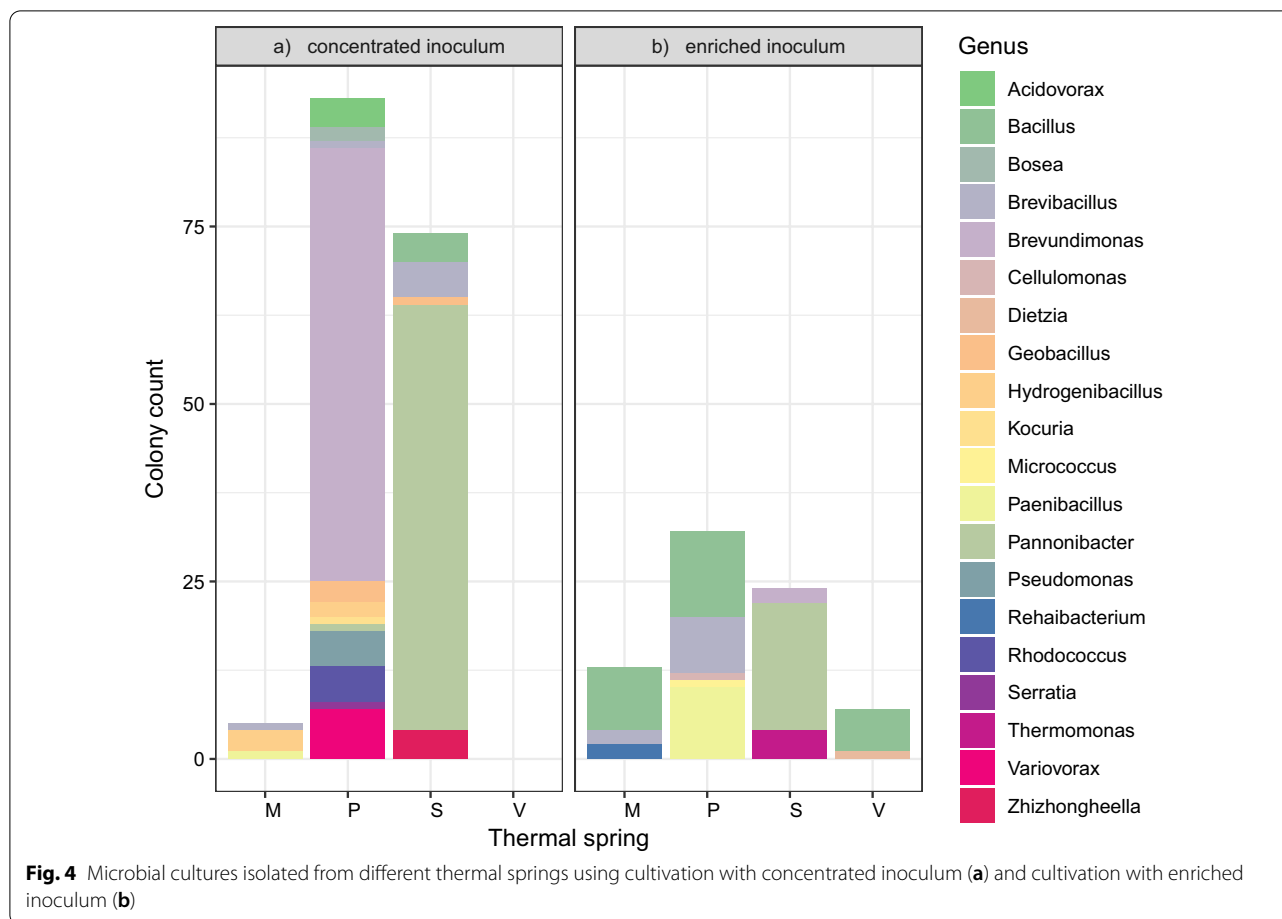
The collection of isolates was characterized in terms of salt tolerance and lethal temperature (Fig. 6). Over 60% of the isolates were able to grow at >3% NaCl, which is approximately the salt concentration of seawater [65] (Fig. 6a). Over 25% of the isolates were able to grow at ≥6% NaCl and can be considered halotolerant [66]. The most salt-tolerant isolate was *Bacillus zhangzhouensis* (isolate M6), which was able to grow at 12% NaCl. Overall, the majority of *Bacillus* isolates were able to grow at >5% NaCl. Other halotolerant genera included *Kocuria* (isolate P14, 10% NaCl) and *Dietzia* (isolate V3, 7% NaCl).

The lethal temperature of the majority of isolates was between 55 and 70 °C (Fig. 6b). However, the isolates of

Brevibacillus and some isolates of the *Bacillus* and *Paenibacillus* genera were able to survive temperatures >80 °C. Notably, the lethal temperature of isolates M6, M9, M10, M12, and S1 was 100 °C; moreover, isolates M1, M4, M8, S3, P7, P21, and P22 were able to survive exposure to boiling water, thus their lethal temperature is presented as >100 °C. Overall, trends in salt and temperature tolerance were related to bacterial genera rather than the isolation source. Spearman's rank correlation between salt and heat tolerance was 0.273 (p value = 0.009).

Genomic analysis of phylogenetically novel isolates

The four isolates *Thermomonas* sp. S9, *Paenibacillus* sp. P25, *Paenibacillus* sp. P26 and *Cellulomonas* sp. P24 were found to have 16S rRNA gene sequence identities of <98.65% to their closest cultured relatives, indicating their potential for representing phylogenetically novel species. Therefore, their full genomes were sequenced and analyzed. Genome statistics, including the number of scaffolds, total size, G + C content, number of coding DNA sequences (CDS) and protein-coding sequences, completeness and contamination of these isolates are given in Additional file 1: Table S6. The phylogenetic trees based on whole genome marker genes are provided in the Additional file 1: Figs. S8–S10.



(See figure on next page.)

Fig. 5 Phylogenetic tree based on 16S rRNA gene sequences of unique thermal spring isolates. The neighbor-joining phylogenetic tree was constructed using the distance model JC69. The outgroup sequence with accession number CP001792 represents *Fibrobacter succinogenes* subsp. *Succinogenes* S85. The green dots represent a match of the 16S rRNA sequence of the culture to ASVs. The blue stars represent potentially novel species

Thermomonas sp. S9

The strain S9 exhibited 98.5% 16S rRNA gene sequence similarity to *Thermomonas haemolytica* A50-7-3^T. OrthoANI values for the S9 genome and available genomes of the related members of the family *Lysobacteraceae* were in the range of 92.8–81.4% [67]. The results of core-gene-based phylogeny depicted the monophyletic clustering of strain S9, showing divergence from the type genera, such as *Thermomonas*, *Luteimonas*, and *Pseudoxanthomonas*. The distribution of functional categories revealed the predominance of genes belonging to metabolism, energy, protein processing, cellular processes, stress response, defense, virulence, and membrane transport (Additional file 1: Table S7). Strain S9 was found to

harbor a wide range of stress and heat response genes of which up to 50% belong to the heat shock *dnaK* gene cluster. Furthermore, the strain S9 was found to harbor genes for glutathione biosynthesis and gamma-glutamyl cycle, which were absent in all other strains (Fig. 7). Collective observations of the genomic characteristics provide evidence that strain S9 represents a novel species of the genus *Thermomonas* in the family *Lysobacteraceae*, phylum Proteobacteria.

Paenibacillus sp. P25 and Paenibacillus sp. P26

The strains P25 and P26 showed 96.8% 16S rRNA gene sequence identity to *Paenibacillus konkukensis* SK-3146^T. An OrthoANI comparison between the available

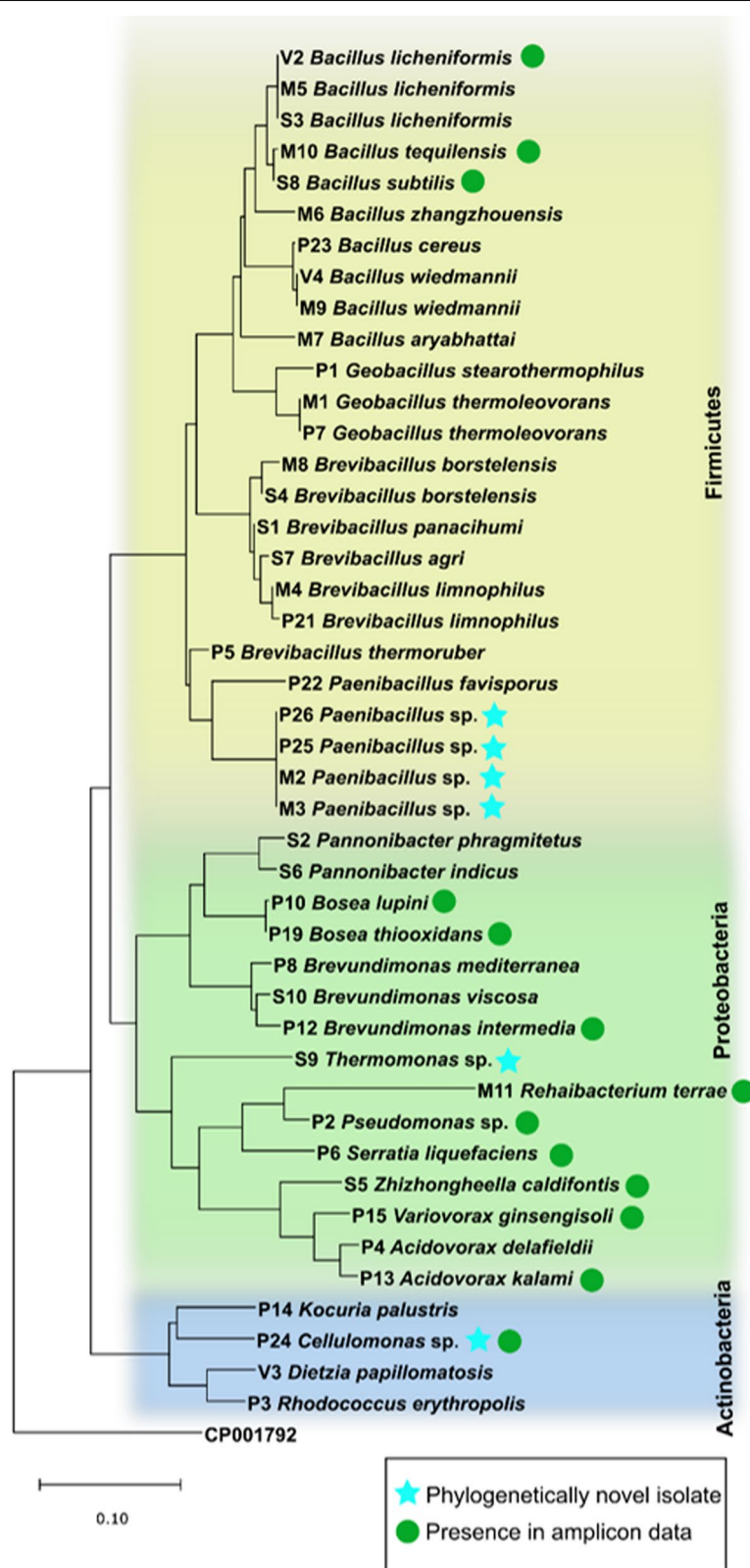
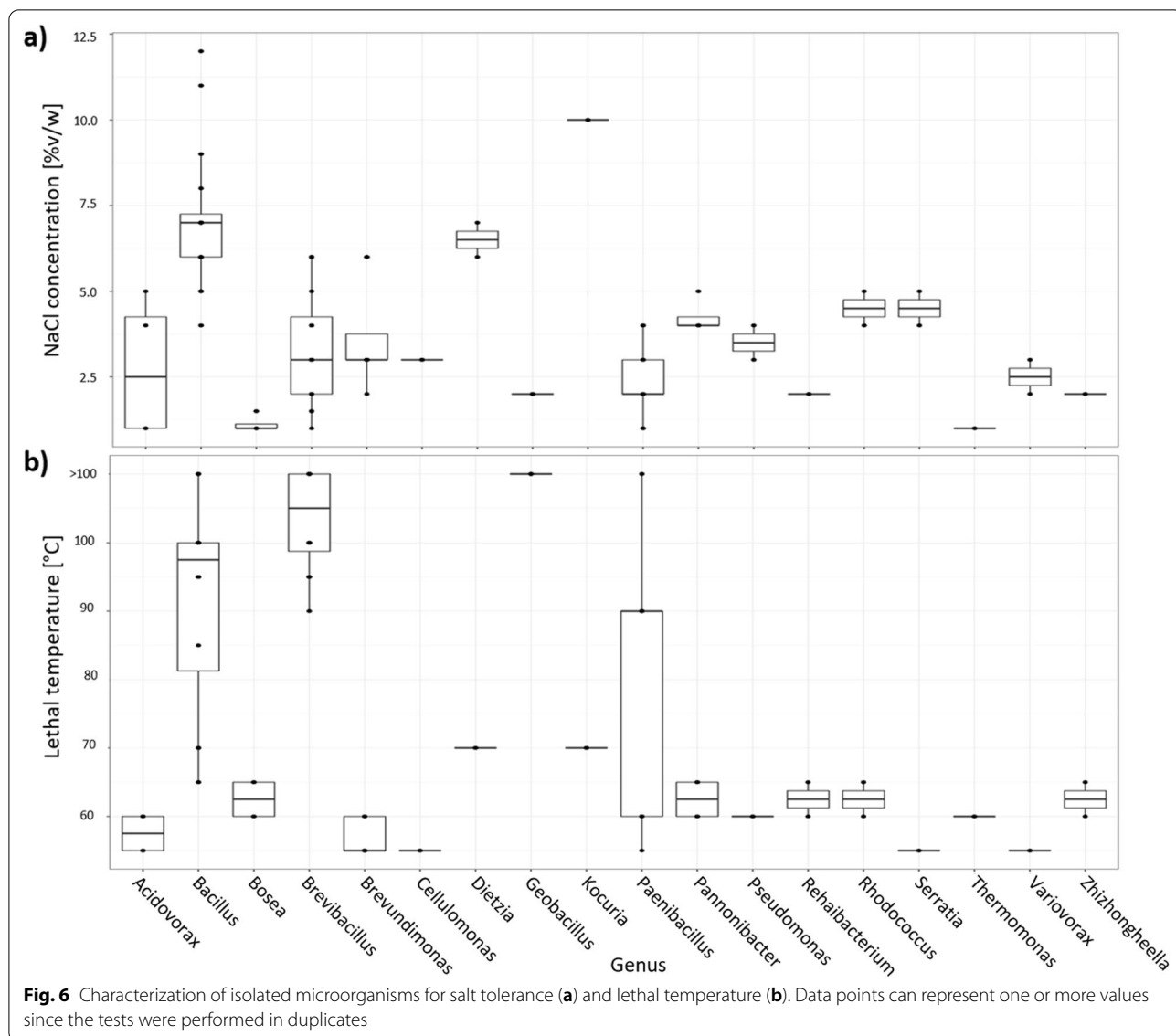
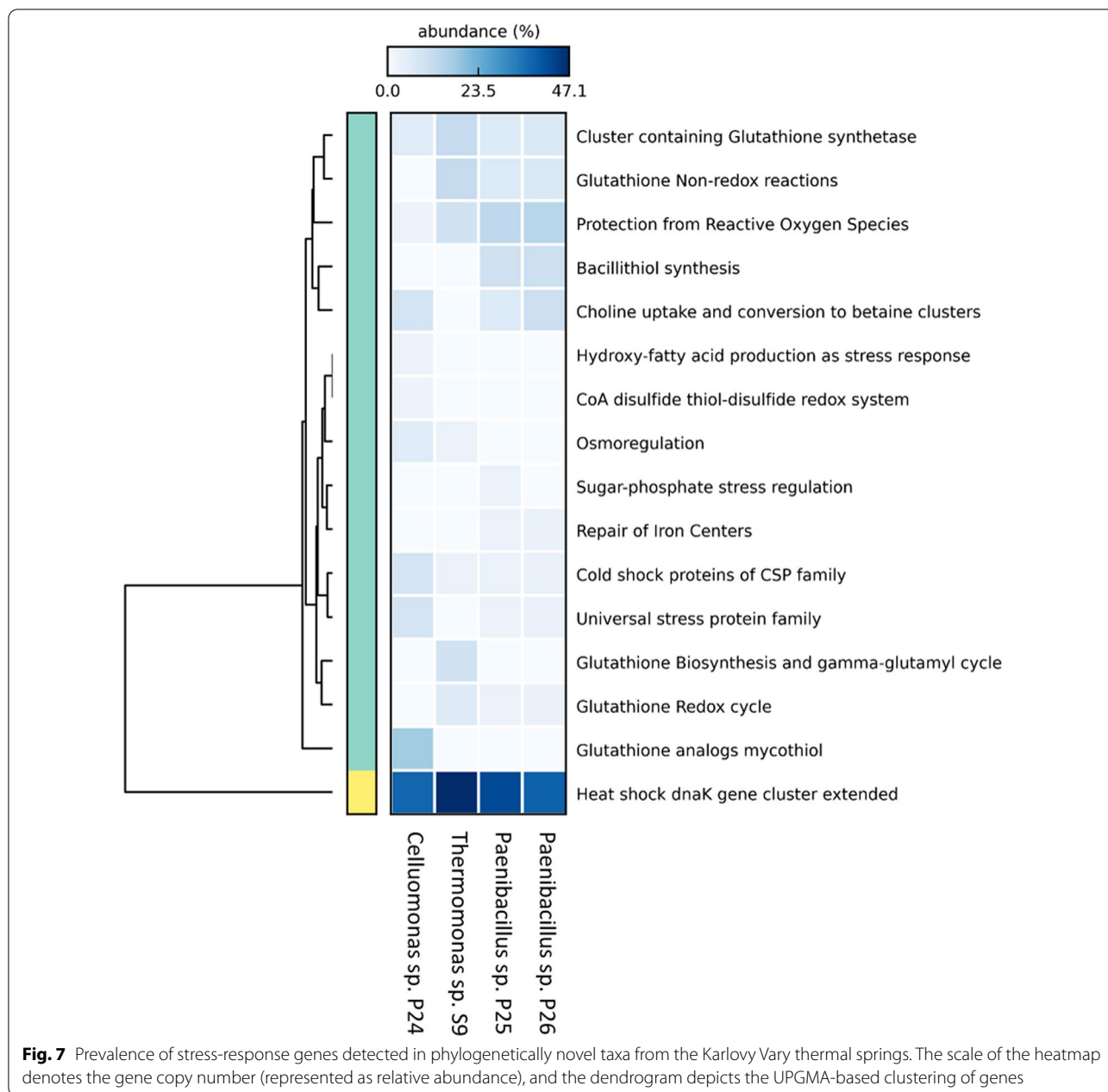


Fig. 5 (See legend on previous page.)



genomes of the genus *Paenibacillus* and the genomes of the P25 and P26 strains resulted in values of $\leq 79.5\%$. OrthoANI homology of strains P25 and P26 was 99.7%, and average amino acid identity was 99.1%. Additionally, core-gene-based phylogenetic analysis revealed a separate clustering of P25 and P26 with the *Paenibacillus* genus (Additional file 1: Fig. S9). Genome annotations of the two novel strains revealed subtle differences in genomic architecture. Both strains featured various gene clusters for stress response, most of which were shared, except for sugar-phosphate stress regulation, which was only featured in strain P25 (Fig. 7). Based on the genomic characterization, we propose strains P25 and P26 to represent members of a novel species of the genus *Paenibacillus* in the family Paenibacillaceae, phylum Firmicutes.

Although strains P25 and P26 exhibited a high degree of sequence similarity (Fig. 7), some phenotypic differences were detected. Whereas strain P25 can grow at 3% NaCl and its lethal temperature is 90 °C, strain P26 can grow at 2% NaCl and its lethal temperature is 60 °C (Fig. 6). These differences could potentially be explained by the observed subtle divergence in their genomes. A genome-genome comparison revealed the presence of 49 and 36 unique genes for strains P25 and P26, respectively, as well as differences in the abundances of KEGG categories (Additional file 1: Fig. S11). The strains P25 and P26 encoded for, respectively, 76 and 77 antibiotic resistance genes, including vancomycin resistance, as detected by PATRIC (Additional file 1: Table S6).



Celluomonas sp. P24

The strain P24 showed 97.9% 16S rRNA gene sequence similarity with *Celluomonas cellulasea* DSM 20118^T. OrthoANI comparisons between related members of the *Cellulomonadaceae* family and strain P24 was $\leq 79.8\%$. Additionally, the results of core-gene-based phylogeny depicted monophyletic clustering of strain P24, exhibiting divergence from the type species of the genus *Celluomonas*. The distribution of functional categories included metabolism, energy, protein

processing, stress response, defense, virulence, cellular processes, and finally membrane transport. The genome was shown to also feature a CRISPR array. The strain P24 was found to harbor the glutathione analog mycothiol, a CoA disulfide thiol-disulfide redox system, and hydroxy-fatty acid production genes as unique functional traits which were absent in the other genomes (Fig. 7). The results from genomic characterization and phylogenetic assessment indicate that strain P24 is a novel species of the genus *Celluomonas* in the family Cellulomonadaceae, phylum Actinobacteria.

Despite originating from the same thermal spring area, all novel strains harbored notably distinct gene complexes, which are presumably required to cope with stress in their extreme environmental niches (Fig. 7).

Discussion

Microbial community composition

In our study, we used a combination of 16S rRNA gene amplicon sequencing analysis and microbial cultivation approaches to analyze the microbial community composition of four Karlovy Vary thermal springs differing in their temperature and chemical composition. Figure 1 depicts the presence of up to 19% as-yet-unclassified, potentially novel microbial taxa at the class level. Despite this, the set of 50 most abundant unique ASVs from each spring contains genera that are commonly reported in literature on similar aquatic habitats, such as the genera *Desulfacinum* and *Thiomonas* [68], *Sulfuricurvum* [69], *Thermoflexus* [70], *Hydrogenobacter* [71], *Sideroxydans* [72], *Halothiobacillus*, or *Sulfurimonas* [73], some of which in similar relative abundances compared to our dataset. Genera known to be involved in sulfate reduction [74] and iron oxidation [75] were found mainly in thermal spring P, which is also the spring with the highest iron and sulfate cation concentrations of the examined thermal springs (Additional file 1: Table S1).

Microbial cultivation

The two cultivation approaches yielded a collection of 44 unique bacterial isolates. The lowest number of obtained colonies originated from thermal springs V and M (Fig. 4), which can be ascribed to the high abundance of archaea, whose cultivation is generally very difficult [76, 77], and previously undetected taxa (Fig. 1). The very low rate of culturability may have also been connected with the presence of VBNC microorganisms that use the uncultivable state as a response to environmental stress such as high temperature [78].

Although the concentrated inoculum approach yielded a higher number of isolates (Fig. 4a), the enriched inoculum approach led to the isolation of novel species belonging to rare taxa in our amplicon dataset (Fig. 4b), including novel species of genera *Thermomonas* (S), *Pae-nibacillus* and *Cellulomonas* (P). These findings confirm that such unconventional cultivation approaches are successful in isolating rare taxa [79] and that prolonged cultivation can enhance culturability of slow-growing and rare microbes [80]. At the same time, the enriched inoculum cultivation approach led to the enrichment of *Bacillus*-like genera, which are ubiquitous and often represent the cultivable part of microbes from thermal springs [15, 81]. Even though roughly a third of isolates were detected in the amplicon sequencing data, most of

these taxa formed a very rare fraction of the amplicon dataset (0.019% of sequencing reads). These results are in line with the findings that the dominant taxa in the environment are often not recovered by microbial cultivation [82]. Most of the genera which were isolated but not found in the amplicon sequencing data belonged to either the genus *Bacillus* and related genera whose cells might have been present as spores in the environment and could be missed by sequencing analyses [83], or as Kurm et al. described in their study, represented cultivable rare taxa [84].

Extreme environmental conditions are challenging to mimic in the laboratory, further complicating the cultivation of microbes. However, alternatives to classical cultivation on Petri dishes exist. For example, some rare taxa can be enriched using dilution to extinction [85], an Ichip platform [86], or diffusion chambers simulating natural environments [87, 88]. Continued microbial cultivation efforts are crucial for the study of phylogenetically novel strains, as well as for the discovery of potential biotechnologically attractive strains and enzymes that are active under extreme conditions [89], such as the widely used Taq polymerase [90]. According to Raddadi et al. [7], thermozyms produced by thermophilic microorganisms are often able to resist proteolysis under a combination of high salinity and high temperature [91]. Thus, we characterized the collection of microorganisms in terms of both salt tolerance and tolerance to high temperatures (Fig. 6). We could not confirm a significant correlation between salt and high temperature tolerance (Spearman's rank correlation = 0.273, p value = 0.009). However, isolate M6 of the species *Bacillus zhangzhouensis* was found to be halophilic and able to tolerate extremely high temperatures (Fig. 6). Moreover, a strain of the same species previously demonstrated biotechnological application potential thanks to the production of alkaliphilic enzymes [92]. Even though this coupled resistance is highly likely to be associated with the ability of bacilli to form spores, the genus *Bacillus* and other members of the family *Bacillaceae* are commonly found to be enzymatically active in hot springs [93, 94]. In addition to the family *Bacillaceae*, the genus *Kocuria* exhibited tolerance to higher salt concentrations (Fig. 6a). It is in line with the previous findings on the tolerance of this genus to different harsh conditions and its biotechnological potential [95]. Further analyses of the collection of isolates for potential biotechnological use is beyond the scope of this study and will be a subject of future research.

Phylogenetic novelty in thermal springs

Most of prokaryotic diversity likely remains unexplored [96], and so it is not surprising that an extreme environment, such as thermal springs, will host

as-yet-undescribed taxa which may represent the majority of its microbial community. Such observations were made by Bourrain et al. [22] and Fasesan et al. [97] in recent years, with both groups pointing out that as-yet-undescribed taxa constitute a stable part of the microbial community in heavily used thermal waters. Bourrain's group examined Avène Thermal Spring Water (21 °C) and obtained 12% unassigned reads at the genus level, a number similar to what we obtained in the coolest spring, P (15%). In contrast, the number of unassigned reads at the genus level in the hottest spring V in our study comprised 98% of the total reads (Additional file 1: Table S4), a fraction that is commonly observed in hot springs [98]. Such a high representation of unassigned reads is caused by the presence of a potentially novel order of the class Altiarchaeia that dominates the two warmest springs, V and M (Fig. 1). The presence of novel archaeal taxa in hot springs is in agreement with previous studies, such as the one by Inskeep et al. [99], who were able to discover and describe a novel archaeal phylum in Yellowstone National Park's geothermal springs by metagenomics.

Despite the high number of unassigned, thus potentially novel sequences in the amplicon data, there is often no confirmation or further analyses of these sequences. To avoid this, we used the tool SSUnique [40] to highlight phylogenetic novelty amongst the unassigned sequences in our amplicon dataset (Fig. 2). Four of the novel classes form extensive clades in the well-established phyla Hydrothermae, Altiarchaeota, Verrucomicrobia, and TA06. Phylum Hydrothermae was first identified by Jungbluth et al. in 2017 using high-quality metagenome-assembled genome analysis to describe the representative bacterium EM3 [100]. Since then, no members of this phylum have been cultivated, which is often the case in hot spring environments [101, 102]. In our dataset, putative members of this phylum were detected mostly in sample V, forming up to 14% of reads. Phylum Altiarchaeota was represented by 60 different ASVs that dominated samples V and M in our dataset (Fig. 2). Although this phylum is widespread in many aquatic habitats [103, 104], no members have been successfully cultivated yet. The first description was done by Probst et al., who introduced *Candidatus* Altiarchaeum hamiconexum as a member of the phylum Euryarchaeota [103], although it was later reclassified as the novel phylum Altiarchaeota, which is a member of the superphylum DPANN. The members of Altiarchaeota are known to be autotrophs and biofilm-forming archaea, and thus represent a carbon sink in thermal springs [105]. Novel clades of the phyla Verrucomicrobia (M, P) and TA06 (S) formed a small fraction of our amplicon data, but their

presence in various freshwater habitats such as acidic hot springs, thermal springs, or cold lakes has been repeatedly reported [106–110]. While no representatives of the phylum TA06 have been cultivated thus far, the phylum Verrucomicrobia has many characterized strains, although its ecological impact in aquatic habitats remains unclear [111]. To sum up, novel taxa, or entire clades of the aforementioned phyla, are great candidates for targeted genomics and could reveal important information about the ecology of thermal springs. Despite our cultivation efforts, we were unable to provide cultured representatives of novel clades of bacteria. To achieve this, live-FISH based on 16S rRNA genes combined with cell-sorting systems [112] could be used for single-cell genome sequencing and the genomic information can be later used for modeling the nutrient requirements [113] of sorted cells.

The modifications of common culture techniques used herein led to the isolation of several potentially novel species (Fig. 5). All the sequenced genomes of the four isolates had OrthoANI values < 95%, which is estimated as the cut-off for species demarcation [114], and are therefore presented herein as novel species (Fig. 5). Novel members of the genus *Paenibacillus* are introduced as two different strains of the same species. Based on the genomic information, both strains harbor antibiotic resistance genes, including resistance to vancomycin (Additional file 1: Table S6). Although this type of resistance was already found in the genomes of a *Paenibacillus* strain isolated from soil [115], it is not typically reported for members of this genus from environments with a lack of human impact, such as Karlovy Vary thermal springs. This specific resistance may play a role in the survival strategy for a stressful extreme environment with a low nutrient content [116, 117]. Due to the type of origin of the novel taxa, we focused our attention on the presence of stress response genes in their genomes. As expected, the most abundant stress response gene in our dataset was the heat shock *dnaK* gene cluster, associated with response to high temperatures (Fig. 7). *Thermomonas* sp. S9 harbored the highest abundance of these genes, corroborating with findings that this genus is often associated with warm to hot environments [118]. In fact, one of seven current members of the genus, the species *T. hydrothermalis*, has been commonly isolated from hot springs and used as a source of thermostable enzymes [119, 120]. In contrast, the genera *Cellulomonas* and *Paenibacillus* encompass many species and their appearance is almost ubiquitous, including thermal springs [121, 122]. Novel representatives of the genus *Paenibacillus* are very often found in thermal waters thanks to their ubiquity and ability to form spores [123–125].

Conclusions

In summary, our study is the first to thoroughly analyze the microbial community composition of Karlovy Vary thermal springs. The modified cultivation approaches enabled us to obtain a collection of 44 indigenous microorganisms, including members of 3 novel species. Due to their extremophilic origins, these microorganisms can be further investigated for potential biotechnological applications. The taxonomic novelty of microbes inhabiting the Karlovy Vary thermal springs was further confirmed by 16S rRNA gene amplicon sequencing, revealing numerous classes of previously undescribed taxa. Future metagenomic analyses will further unravel the community structure and metabolic potential of Karlovy Vary thermal spring microbial communities. Such information will aid in defining subsequent targeted cultivation experiments that could lead to the discovery of novel extremophilic enzymes with the potential of decreasing water and energy consumption in industrial biotechnology.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40793-022-00440-2>.

Additional file 1. Table S1: Chemical composition, temperature, and flow rates of the examined thermal springs. **Table S2:** Mock community composition. **Table S3:** List and composition of the used media. **Table S4:** Percentage of unclassified ASVs at different taxonomic levels. **Table S5:** Sequencing coverage computed using package iNEXT. **Table S6:** Genome statistics of the phylogenetically novel bacterial species and results from the annotation server PATRIC. **Table S7:** Functional categories distribution in the genomes of the phylogenetically novel bacterial species. **Figure S8:** UBCG-based core-genome phylogenetic analysis depicting the distinct positioning of strain S9 with members of the genus *Thermomonas*. Bootstrap values (expressed as percentages of 1000 replications) of above 70% are shown at the branch points. **Figure S9:** UBCG-based core-genome phylogenetic analysis depicting the distinct positioning of strains P25 and P26 with members of the genus *Paenibacillus*. Bootstrap values (expressed as percentages of 1000 replications) of above 70% are shown at the branch points. **Figure S10:** UBCG based core-genome phylogenetic analysis depicting the distinct positioning of strain P24 with members of the genus *Cellulomonas*. Bootstrap values (expressed as percentages of 1000 replications) of above 70% are shown at the branch points. **Figure S11:** Differences in KEGG categories abundances in the genomes of strains P25 and P26.

Acknowledgements

Not applicable.

Author contributions

TS, JS, MS, and OU designed the study. OU and MS supervised the project. TS, TV, and OU collected the samples. TS, GK, and PP carried out the experiments. TS, MS, and KJ analyzed the data. TS, KJ, MS, and OU contributed to the interpretation of the results. TS and KJ wrote the manuscript with support from MS, JS, TV, and OU. All authors approved the final version of the manuscript.

Funding

The research reported herein was supported by the Czech Science Foundation under Grant No. 18-000365. Further support from the ELIXIR CZ research infrastructure project (Ministry of Education, Youth and Sports of the Czech Republic Grant No. LM2018131), including access to computing and storage

facilities, is gratefully acknowledged. Kunal Jani gratefully acknowledges support from the European Structural and Investment Funds, OP RDE-funded project 'ChemJets2' (CZ.02.2.69/0.0/0.0/18_053/ 0016974). The experimental part at the Military Health Institute laboratories was supported by funding from the project of the Ministry of Defense of the Czech Republic through the Long-term Organization Development Plan No. 907930101413.

Availability of data and materials

All MiSeq reads generated and analyzed during the current study are available in the NCBI Short Read Archive under SRA study number PRJNA781448. The genome data generated in this study are available under the NCBI BioProject ID PRJNA772595.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Biochemistry and Microbiology, Faculty of Food and Biochemical Technology, University of Chemistry and Technology, Prague, Technická 3, 166 28 Prague 6, Czech Republic. ²Military Health Institute, Ministry of Defence of the Czech Republic, Prague, Czech Republic. ³Institute of Balneology and Spa Sciences, Karlovy Vary, Czech Republic.

Received: 22 November 2021 Accepted: 26 August 2022

Published online: 11 September 2022

References

- Alfaro C, Wallace M. Origin and classification of springs and historical review with current applications. *Environ Geol.* 1994;24(2):112–24.
- Des Marais DJ, Walter MR. Terrestrial hot spring systems: introduction. *Astrobiology.* 2019;19(12):1419–32.
- Djokic T, Van Kranendonk MJ, Campbell KA, Walter MR, Ward CR. Earliest signs of life on land preserved in ca. 3.5 Ga hot spring deposits. *Nat Commun.* 2017;8(1):15263.
- Dalmaso GZ, Ferreira D, Vermelho AB. Marine extremophiles: a source of hydrolases for biotechnological applications. *Mar Drugs.* 2015;13(4):1925–65.
- Lewis K, Epstein S, D'Onofrio A, Ling LL. Uncultured microorganisms as a source of secondary metabolites. *J Antibiot.* 2010;63(8):468–76.
- Brock TD, Freeze H. *Thermus aquaticus* gen. n. and sp. N., a nonsporulating extreme thermophile. *J Bacteriol.* 1969;98(1):289–97.
- Raddadi N, Cherif A, Daffonchio D, Neifar M, Fava F. Biotechnological applications of extremophiles, extremozymes and extremolytes. *Appl Microbiol Biotechnol.* 2015;99(19):7907–13.
- Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol.* 1985;39(1):321–46.
- Stewart EJ. Growing unculturable bacteria. *J Bacteriol.* 2012;194(16):4151–60.
- Gutleben J, Chaib De Mares M, van Elsas JD, Smidt H, Overmann J, Sipkema D. The multi-omics promise in context: from sequence to microbial isolate. *Crit Rev Microbiol.* 2018;44(2):212–29.
- Fox GE, Pechman KR, Woese CR. Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to prokaryotic systematics. *Int J Syst Evol.* 1977;27(1):44–57.
- Clarridge JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev.* 2004;17(4):840–62.
- Nygaard AB, Tunsjø HS, Meisal R, Charnock C. A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA

- gene sequencing to characterize building-dust microbiomes. *Sci Rep.* 2020;10(1):3209.
14. Strunecký O, Kopejtká K, Goecke F, Tomasch J, Lukavský J, Neori A, Kahl S, Pieper DH, Pilarski P, Kaftan D, et al. High diversity of thermophilic cyanobacteria in Rupite hot spring identified by microscopy, cultivation, single-cell PCR and amplicon sequencing. *Extremophiles.* 2019;23(1):35–48.
 15. Najar IN, Sherpa MT, Das S, Das S, Thakur N. Microbial ecology of two hot springs of Sikkim: predominate population and geochemistry. *Sci Total Environ.* 2018;637–638:730–45.
 16. Hobel CFV, Marteinsson VT, Hreggvidsson GÓ, Kristjánsson JK. Investigation of the microbial ecology of intertidal hot springs by using diversity analysis of 16S rRNA and chitinase genes. *Appl Environ Microbiol.* 2005;71(5):2771–6.
 17. Paces T, Smejkal V. Magmatic and fossil components of mineral waters in the Eager-river continental rift. In: Wanty R, Seal R, editors. *Water–rock interaction. II.* Rotterdam: AA Balkema Publishers; 2004. p. 167–72.
 18. Ward L, Taylor MW, Power JF, Scott BJ, McDonald IR, Stott MB. Microbial community dynamics in Inferno Crater Lake, a thermally fluctuating geothermal spring. *ISME J.* 2017;11(5):1158–67.
 19. Kato S, Itoh T, Yuki M, Nagamori M, Ohnishi M, Uematsu K, Suzuki K, Takashina T, Ohkuma M. Isolation and characterization of a thermophilic sulfur- and iron-reducing thaumarchaeote from a terrestrial acidic hot spring. *ISME J.* 2019;13(10):2465–74.
 20. Amin A, Ahmed I, Salam N, Kim B-Y, Singh D, Zhi X-Y, Xiao M, Li W-J. Diversity and distribution of thermophilic bacteria in hot springs of Pakistan. *Microb Ecol.* 2017;74(1):116–27.
 21. Knapik K, Becerra M, González-Siso M-I. Microbial diversity analysis and screening for novel xylanase enzymes from the sediment of the Lobios Hot Spring in Spain. *Sci Rep.* 2019;9:11195.
 22. Bourrain M, Suzuki MT, Calvez A, West NJ, Lions J, Lebaron P. In-depth prospection of Avène Thermal Spring Water reveals an uncommon and stable microbial community. *J Eur Acad Dermatol Venereol.* 2020;34(S5):8–14.
 23. Inskeep W, Jay Z, Tringe S, Herrgard M, Rusch D. The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Front Microbiol.* 2013;4:67.
 24. Power JF, Carere CR, Lee CK, Wakerley GLJ, Evans DW, Button M, White D, Climo MD, Hinze AM, Morgan XC, et al. Microbial biogeography of 925 geothermal springs in New Zealand. *Nat Commun.* 2018;9(1):2876.
 25. Vrba J. Thermal mineral water springs in Karlovy Vary. *Environ Geol.* 1996;27(2):120–5.
 26. Vrba J. Origin and occurrence of carbon dioxide and gaseous mineral waters, Variscian platform, central Europe. *Econ Geol.* 1964;59(5):874–82.
 27. Vylita T, Žák K, Čilek V, Hercman H, Mikšíková L. Evolution of hot-spring travertine accumulation in Karlovy Vary/Carlsbad (Czech Republic) and its significance for the evolution of Teplá Valley and Ohře/Eger Rift. *Zeitschrift für Geomorphol.* 2007;51(4):427–42.
 28. Pěčková M. Properties of a hyperthermophilic bacterium (*Thermus* sp.) isolated from a Carlsbad spring. *Folia Microbiol.* 1991;36(6):515–21.
 29. Kolouchová I, Timkina E, Matátková O, Kyselová L, Řezanka T. Analysis of bacteriohopanoids from thermophilic bacteria by liquid chromatography-mass spectrometry. *Microorganisms.* 2021;9(10):2062.
 30. Gharwalová L, Palyzová A, Marešová H, Kolouchová I, Kyselová L, Řezanka T. Identification of homologous polyprenols from thermophilic bacteria. *Microorganisms.* 2021;9(6):1168.
 31. Fraraccio S, Strejček M, Dolinova I, Macek T, Uhlik O. Secondary compound hypothesis revisited: selected plant secondary metabolites promote bacterial degradation of cis-1,2-dichloroethylene (cDCE). *Sci Rep.* 2017;7(1):8406.
 32. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol.* 2016;18(5):1403–14.
 33. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 2013;41(1):e1–e1.
 34. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
 35. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13(7):581–3.
 36. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res.* 2014;42(D1):D643–8.
 37. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One.* 2013;8:e61217.
 38. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome.* 2018;6(1):226.
 39. Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol Evol.* 2016;7(12):1451–6.
 40. Lynch MDJ, Neufeld JD. SSUnique: detecting sequence novelty in microbiome surveys. *mSystems.* 2016;1(6):e00133-00116.
 41. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics.* 2012;28(14):1823–9.
 42. Yu G. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinform.* 2020;69(1):e96.
 43. Hugenholtz P. Exploring prokaryotic diversity in the genomic era. *Genome Biol.* 2002;3(2):REVIEWS0003.
 44. Strejček M, Smrhova T, Junkova P, Uhlik O. Whole-cell MALDI-TOF MS versus 16S rRNA gene analysis for identification and dereplication of recurrent bacterial isolates. *Front Microbiol.* 2018;9:1294.
 45. Thomas P, Sekhar AC, Upreti R, Mujawar MM, Pasha SS. Optimization of single plate-serial dilution spotting (SP-SDS) with sample anchoring as an assured method for bacterial and yeast cfu enumeration and single colony isolation from diverse samples. *Biotechnol Rep.* 2015;8:45–55.
 46. Wickham H, Navarro D, Pedersen TL. *Ggplot2: elegant graphics for data analysis.* New York: Springer; 2016.
 47. Wright ES. RNAconTest: comparing tools for noncoding RNA multiple sequence alignment based on structural consistency. *RNA.* 2020;26(5):531–40.
 48. Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 2019;35(3):526–8.
 49. Lopez-Echartea E, Suman J, Smrhova T, Ridl J, Pajer P, Strejček M, Uhlik O. Genomic analysis of dibenzofuran-degrading *Pseudomonas veronii* strain Pvy reveals its biodegradative versatility. *G3 Genes.* 2020;11(2):jkaa030.
 50. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 2019;20(1):129.
 51. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27(5):722–36.
 52. 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(7):1043–55.
 53. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S. KBase: the United States department of energy systems biology knowledgebase. *Nat Biotechnol.* 2018;36(7):566–9.
 54. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068–9.
 55. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, et al. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res.* 2019;48(D1):D606–12.
 56. Graham E, Heidelberg J, Tully B. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J.* 2018;12(7):1861–6.
 57. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek.* 2017;110(10):1281–6.
 58. Kim D, Park S, Chun J. Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity. *J Microbiol.* 2021;59(5):476–80.

59. Na S-I, Kim YO, Yoon S-H, Sung-min H, Baek I, Chun J. UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J Microbiol.* 2018;56(4):281–5.
60. Kumar S, Stecher G, Li M, Nkryaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35(6):1547.
61. Chaudhari NM, Gupta VK, Dutta C. BPGA—an ultra-fast pan-genome analysis pipeline. *Sci Rep.* 2016;6(1):1–10.
62. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics.* 2020;36(7):2251–2.
63. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics.* 2014;30(21):3123–4.
64. Setlow P. Observations on research with spores of Bacillales and Clostridiales species. *J Appl Microbiol.* 2019;126(2):348–58.
65. Millero FJ, Feistel R, Wright DG, McDougall TJ. The composition of Standard Seawater and the definition of the Reference-Composition Salinity Scale. *Deep-Sea Res I Oceanogr Res Pap.* 2008;55(1):50–72.
66. Larsen H. Halophilic and halotolerant microorganisms—an overview and historical perspective. *FEMS Microbiol Rev.* 1986;39:3–7.
67. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, Da Costa MS, Rooney AP, Yi H, Xu X-W, De Meyer S, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol.* 2018;68(1):461–6.
68. Sharma N, Kumar J, Abedin MM, Sahoo D, Pandey A, Rai AK, Singh SP. Metagenomics revealing molecular profiling of community structure and metabolic pathways in natural hot springs of the Sikkim Himalaya. *BMC Microbiol.* 2020;20(1):246.
69. Chen Y, Wu L, Boden R, Hillebrand A, Kumaresan D, Moussard H, Baciú M, Lu Y, Colin MJ. Life without light: microbial diversity and evidence of sulfur- and ammonium-based chemolithotrophy in Movile Cave. *ISME J.* 2009;3(9):1093–104.
70. Rojas-Gätjens D, Arce-Rodríguez A, Puente-Sánchez F, Avendaño R, Libby E, Conejo-Barboza G, Mora-Amador R, Rojas K, Pieper DH, Chavarría M. Rapid shift in microbial community structure in a neutral hydrothermal hot spring from Costa Rica. *bioRxiv.* 2020.
71. DeCastro M-E, Doane MP, Dinsdale EA, Rodríguez-Belmonte E, González-Siso M-I. Exploring the taxonomical and functional profile of As Burgas hot spring focusing on thermostable β -galactosidases. *Sci Rep.* 2021;11(1):101.
72. Krauze P, Kämpf H, Horn F, Liu Q, Voropaev A, Wagner D, Alawi M. Microbiological and geochemical survey of CO₂-dominated mofette and mineral waters of the Cheb Basin, Czech Republic. *Front Microbiol.* 2017;8:2446.
73. Deja-Sikora E, Gołębiwski M, Kalwasińska A, Krawiec A, Kosobucki P, Walczak M. *Comamonadaceae* OTU as a remnant of an ancient microbial community in sulfidic waters. *Microb Ecol.* 2019;78(1):85–101.
74. Han Y, Perner M. The globally widespread genus *Sulfurimonas*: versatile energy metabolisms and adaptations to redox clines. *Front Microbiol.* 2015;6:989.
75. Emerson D, Moyer C. Isolation and characterization of novel iron-oxidizing bacteria that grow at circumneutral pH. *Appl Environ Microbiol.* 1997;63(12):4784–92.
76. Bornemann TLV, Adam PS, Turzynski V, Schreiber U, Figueroa-Gonzalez PA, Rahlff J, Köster D, Schmidt TC, Schunk R, Krauthausen B, et al. Genetic diversity in terrestrial subsurface ecosystems impacted by geological degassing. *Nat Commun.* 2022;13(1):284.
77. Sun Y, Liu Y, Pan J, Wang F, Li M. Perspectives on cultivation strategies of Archaea. *Microb Ecol.* 2020;79(3):770–84.
78. Oliver JD. The viable but nonculturable state for bacteria: status update. *Microbe.* 2016;11(4):159–64.
79. Davis KER, Joseph SJ, Janssen PH. Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. *Appl Environ Microbiol.* 2005;71(2):826–34.
80. Bender KE, Glover K, Archey A, Barton HA. The impact of sample processing and media chemistry on the culturable diversity of bacteria isolated from a cave. *Int J Speleol.* 2020;49(3):3.
81. Rahul K, Sharma RC. Microbial diversity and physico-chemical attributes of two hot water springs in the Garhwal Himalaya, India. *J Microbiol Biotechnol Food Sci.* 2019;8(6):1249–53.
82. Burns DG, Camakarís HM, Janssen PH, Dyll-Smith ML. Combined use of cultivation-dependent and cultivation-independent methods indicates that members of most Haloarchaeal groups in an Australian crystallizer pond are cultivable. *Appl Environ Microbiol.* 2004;70(9):5258–65.
83. Egan M, Dempsey E, Ryan CA, Ross RP, Stanton C. The sporobiota of the human gut. *Gut Microbes.* 2021;13(1):1–17.
84. Kurm V, Van Der Putten WH, Hol WHG. Cultivation-success of rare soil bacteria is not influenced by incubation time and growth medium. *PLoS ONE.* 2019;14(1):e0210073.
85. Kim S, Park MS, Song J, Kang I, Cho J-C. High-throughput cultivation based on dilution-to-extinction with catalase supplementation and a case study of cultivating acI bacteria from Lake Soyang. *J Microbiol.* 2020;58(11):893–905.
86. Nichols D, Cahoon N, Trakhtenberg EM, Pham L, Mehta A, Belanger A, Kanigan T, Lewis K, Epstein SS. Use of Ichip for high-throughput *in situ* cultivation of “uncultivable” microbial species. *Appl Environ Microbiol.* 2010;76(8):2445.
87. Kaeberlein T, Lewis K, Epstein SS. Isolating “uncultivable” microorganisms in pure culture in a simulated natural environment. *Science.* 2002;296(5570):1127–9.
88. Jung D, Seo E-Y, Owen JS, Aoi Y, Yong S, Lavrentyeva EV, Ahn TS. Application of the filter plate microbial trap (FPMT), for cultivating thermophilic bacteria from thermal springs in Barguzin area, eastern Baikal, Russia. *Biosci Biotechnol Biochem.* 2018;82(9):1624–32.
89. Coker JA. Extremophiles and biotechnology: current uses and prospects. *F1000Research.* 2016;5:396.
90. Tindall KR, Kunkel TA. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry.* 1988;27(16):6008–13.
91. Li X, Zhang S, Zhang Q, Gan L, Jiang G, Tian Y, Shi B. Characterization and application of a novel halotolerant protease with no collagenase activity for cleaner dehairing of goatskin. *Process Biochem.* 2022;113:203–15.
92. Moridshahi R, Bahreini M, Sharifmoghaddam M, Asoodeh A. Biochemical characterization of an alkaline surfactant-stable keratinase from a new keratinase producer, *Bacillus zhangzhouensis*. *Extremophiles.* 2020;24(5):693–704.
93. Soy S, Nigam VK, Sharma SR. Cellulolytic, amyolytic and xylanolytic potential of thermophilic isolates of Surajkund hot spring. *J Biosci.* 2019;44(5):124.
94. Panosyan H, Margaryan A, Birkeland N-K. Geothermal springs in Armenia and Nagorno-Karabakh: potential sources of hydrolase-producing thermophilic bacilli. *Extremophiles.* 2020;24(4):519–36.
95. Timkina E, Drábová L, Palyzová A, Řezanka T, Matátková O, Kolouchová I. *Kocuria* strains from unique radon spring water from Jachymov Spa. *Ferment.* 2022;8(1):35.
96. Jiao JY, Liu L, Hua ZS, Fang BZ, Zhou EM, Salam N, Hedlund BP, Li WJ. Microbial dark matter coming to light: challenges and opportunities. *Natl Sci Rev.* 2021;8(3):nwaa280.
97. Fasesan D, Dawkins K, Ramirez R, Rasheed-Jada H, Onilude A, Nash O, Esiobu N. Analysis of a tropical warm spring microbiota using 16S rRNA metabarcoding. *Adv Microbiol.* 2020;10:145–65.
98. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L, Neufeld JD. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *mSystems.* 2018;3(5):e00055-00018.
99. Inskeep W, Jay Z, Herrgard M, Kozubal M, Rusch D, Tringe S, Macur R, deM R, Jennings R, Boyd E, Spear J, et al. Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Front Microbiol.* 2013;4:95.
100. Jungbluth SP, Amend JP, Rappé MS. Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. *Sci Data.* 2017;4(1):170037.
101. López-López O, Cerdán ME, González-Siso MI. Hot spring metagenomics. *Life.* 2013;3(2):308–20.
102. Marín-Paredes R, Tapia-Torres Y, Martínez-Romero E, Quesada M, Servín-Garcidueñas LE, Newton ILG. Metagenome assembly and metagenome-assembled genome of “*Candidatus Aramenus sulfurataquae*” from thermal sediments from the Los Azufres volcanic complex. *Microbiol Resour Announc.* 2021;10(39):e00379-00321.

103. Probst AJ, Weinmaier T, Raymann K, Perras A, Emerson JB, Rattei T, Wanner G, Klingl A, Berg IA, Yoshinaga M, et al. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. *Nat Commun*. 2014;5(1):5497.
104. Bird JT, Baker BJ, Probst AJ, Podar M, Lloyd KG. Culture independent genomic comparisons reveal environmental adaptations for Altiarchaeales. *Front Microbiol*. 2016;7:1221.
105. Dombrowski N, Lee J-H, Williams TA, Offre P, Spang A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol Lett*. 2019;366(2):fnz008.
106. Bai Y, Wang J, Zhan Z, Guan L, Jin L, Zheng G, Huang Z. The variation of microbial communities in a depth profile of peat in the Gahai lake wetland natural conservation area. *Geomicrobiol J*. 2018;35(6):484–90.
107. Schneider D, Arp G, Reimer A, Reitner J, Daniel R. Phylogenetic analysis of a microbialite-forming microbial mat from a hypersaline lake of the Kiritimati Atoll, Central Pacific. *PLoS ONE*. 2013;8(6):e66662.
108. Chiang E, Schmidt ML, Berry MA, Biddanda BA, Burtner A, Johengen TH, Palladino D, Deneff VJ. Verrucomicrobia are prevalent in north-temperate freshwater lakes and display class-level preferences between lake habitats. *PLoS ONE*. 2018;13(3):e0195112.
109. Erikstad H-A, Ceballos RM, Smestad NB, Birkeland N-K. Global biogeographic distribution patterns of thermoacidophilic Verrucomicrobia methanotrophs suggest allopatric evolution. *Front Microbiol*. 2019;10:1129.
110. Magnabosco C, Tekere M, Lau MCY, Linage B, Kuloyo O, Erasmus M, Cason E, van Heerden E, Borgonie G, Kieft TL, et al. Comparisons of the composition and biogeographic distribution of the bacterial communities occupying South African thermal springs with those inhabiting deep subsurface fracture water. *Front Microbiol*. 2014;5:679.
111. He S, Stevens SLR, Chan L-K, Bertilsson S, del Rio TG, Tringe SG, Malmstrom RR, McMahon KD, Hallam SJ. Ecophysiology of freshwater Verrucomicrobia inferred from metagenome-assembled genomes. *mSphere*. 2017;2(5):e00277-00217.
112. Batani G, Bayer K, Böge J, Hentschel U, Thomas T. Fluorescence in situ hybridization (FISH) and cell sorting of living bacteria. *Sci Rep*. 2019;9(1):18618.
113. Kumar VS, Maranas CD. GrowMatch: an automated method for reconciling *in silico/in vivo* growth predictions. *PLoS Comput Biol*. 2009;5(3):e1000308.
114. Lee I, Kim YO, Park S-C, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol*. 2016;66(2):1100–3.
115. Guardabassi L, Perichon B, Heijenoort JV, Blanot D, Courvalin P. Glycopeptide resistance *vanA* operons in *Paenibacillus* strains isolated from soil. *Antimicrob Agents Chemother*. 2005;49(10):4227–33.
116. D'Costa VM, King CE, Kalan L, Morar M, Sung WWL, Schwarz C, Froese D, Zazula G, Calmels F, Debruyne R, et al. Antibiotic resistance is ancient. *Nature*. 2011;477(7365):457–61.
117. Waglechner N, Wright GD. Antibiotic resistance: it's bad, but why isn't it worse? *BMC Biol*. 2017;15(1):84.
118. Alves MP, Rainey FA, Nobre MF, da Costa MS. *Thermomonas hydrothermalis* sp. Nov., a new slightly thermophilic γ -proteobacterium isolated from a hot spring in Central Portugal. *Syst Appl Microbiol*. 2003;26(1):70–5.
119. Al-Daghistani HI, Mohammad BT, Kurniawan TA, Singh D, Rabadi AD, Xue W, Avtar R, Othman MHD, Shirazian S. Characterization and applications of *Thermomonas hydrothermalis* isolated from Jordan's hot springs for biotechnological and medical purposes. *Process Biochem*. 2021;104:171–81.
120. Abdollahi P, Ghane M, Babaekhou L. Isolation and characterization of thermophilic bacteria from Gavmesh Goli hot spring in Sabalan geothermal field, Iran: *Thermomonas hydrothermalis* and *Bacillus altitudinis* isolates as a potential source of thermostable protease. *Geomicrobiol J*. 2021;38(1):87–95.
121. Jessen JE, Sveinsson T, Scully S, Orlygsson J. Ethanol production by a *Paenibacillus* species isolated from an Icelandic hot spring: production yields from complex biomass. *Icel Agric Sci*. 2015;28:15–24.
122. Kumar M, Yadav AN, Tiwari R, Prasanna R, Saxena AK. Evaluating the diversity of culturable thermotolerant bacteria from four hot springs of India. *J Biodivers Biopros Dev*. 2014;1(3):1000127.
123. Mead DA, Lucas S, Copeland A, Lapidus A, Cheng J-F, Bruce DC, Goodwin LA, Pitluck S, Chertkov O, Zhang X, et al. Complete genome sequence of *Paenibacillus* strain Y4.12MC10, a novel *Paenibacillus lautus* strain isolated from Obsidian hot spring in Yellowstone National Park. *Stand Genom Sci*. 2012;6(3):366–85.
124. Bouraoui H, Rebib H, Aissa MB, Touzel JP, O'donohue M, Manai M. *Paenibacillus marinum* sp. Nov., a thermophilic xylanolytic bacterium isolated from a marine hot spring in Tunisia. *J Basic Microbiol*. 2013;53(11):877–83.
125. Fujinami S, Takeda-Yano K, Onodera T, Satoh K, Sano M, Takahashi Y, Narumi I, Ito M. Draft genome sequence of calcium-dependent *Paenibacillus* sp. strain TCA20, isolated from a hot spring containing a high concentration of calcium ions. *Genome Announc*. 2014;2(5):e00866-00814.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

