



# Effects of Physicochemical Factors on Prokaryotic Biodiversity in Malaysian Circumneutral Hot Springs

Chia S. Chan<sup>1</sup>, Kok-Gan Chan<sup>2</sup>, Robson Ee<sup>2</sup>, Kar-Wai Hong<sup>2</sup>, Maria S. Urbietta<sup>3</sup>, Edgardo R. Donati<sup>3</sup>, Mohd S. Shamsir<sup>1</sup> and Kian M. Goh<sup>1\*</sup>

<sup>1</sup> Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia, <sup>2</sup> Division of Genetics and Molecular Biology, Faculty of Science, Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia, <sup>3</sup> CINDEFI (CCT, La Plata-CONICET, UNLP), Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina

## OPEN ACCESS

### Edited by:

Hongchen Jiang,  
Miami University, United States

### Reviewed by:

Song Zhaoqi,  
Shangqiu Normal University, China  
Weiguo Hou,  
China University of Geosciences,  
China  
Brandon Briggs,  
University of Alaska Anchorage,  
United States

### \*Correspondence:

Kian M. Goh  
gohkianmau@utm.my

### Specialty section:

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

**Received:** 26 April 2017

**Accepted:** 21 June 2017

**Published:** 06 July 2017

### Citation:

Chan CS, Chan K-G, Ee R,  
Hong K-W, Urbietta MS, Donati ER,  
Shamsir MS and Goh KM (2017)  
Effects of Physicochemical Factors on  
Prokaryotic Biodiversity in Malaysian  
Circumneutral Hot Springs.  
*Front. Microbiol.* 8:1252.  
doi: 10.3389/fmicb.2017.01252

Malaysia has a great number of hot springs, especially along the flank of the Banjaran Titiwangsa mountain range. Biological studies of the Malaysian hot springs are rare because of the lack of comprehensive information on their microbial communities. In this study, we report a cultivation-independent census to describe microbial communities in six hot springs. The Ulu Slim (US), Sungai Klah (SK), Dusun Tua (DT), Sungai Serai (SS), Semenyih (SE), and Ayer Hangat (AH) hot springs exhibit circumneutral pH with temperatures ranging from 43°C to 90°C. Genomic DNA was extracted from environmental samples and the V3–V4 hypervariable regions of 16S rRNA genes were amplified, sequenced, and analyzed. High-throughput sequencing analysis showed that microbial richness was high in all samples as indicated by the detection of 6,334–26,244 operational taxonomy units. In total, 59, 61, 72, 73, 65, and 52 bacterial phyla were identified in the US, SK, DT, SS, SE, and AH hot springs, respectively. Generally, Firmicutes and Proteobacteria dominated the bacterial communities in all hot springs. Archaeal communities mainly consisted of Crenarchaeota, Euryarchaeota, and Parvarchaeota. In beta diversity analysis, the hot spring microbial memberships were clustered primarily on the basis of temperature and salinity. Canonical correlation analysis to assess the relationship between the microbial communities and physicochemical variables revealed that diversity patterns were best explained by a combination of physicochemical variables, rather than by individual abiotic variables such as temperature and salinity.

**Keywords:** 16S rRNA amplicon sequencing, hot spring metagenome, saline pool, microbial community, thermophile diversity, microbial symbiosis, microbiome

**Abbreviations:** AH, Ayer Hangat; BOD, biochemical oxygen demand; CCA, canonical correspondence analysis; C:N, carbon-to-nitrogen ratio; COD, chemical oxygen demand; DT, Dusun Tua; OTU, operational taxonomic unit; PCA, principal component analysis; PCoA, principal coordinate analysis; QIIME, Quantitative Insights Into Microbial Ecology; SE, Semenyih; SK, Sungai Klah; SS, Sungai Serai; TN, total nitrogen; TOC, total organic carbon; UPGMA, unweighted pair group method with arithmetic mean; US, Ulu Slim.

## INTRODUCTION

The study of extremophiles provides insights into the origin and evolution of life. Biologists believe that extremophiles inhabiting extreme environments such as hot springs are the closest living descendants of the earliest life forms on earth (Woese et al., 1990; Olsen et al., 1994). Additionally, these extreme environments comprise relatively simple microbial ecosystems as compared with more complex environments such as soils (Xu et al., 2014), wastewater (Shanks et al., 2013), marine sediments (Zheng et al., 2014), and the human gastrointestinal tract (Trosvik and de Muinck, 2015). Therefore, studying the life in hot springs enables us to interrogate the interactions between organisms and the environment. Microbial communities in hot springs, especially those in Yellowstone National Park (USA) (Blank et al., 2002; Kan et al., 2011; Inskeep et al., 2013), Japan (Kubo et al., 2011; Nishiyama et al., 2013; Masaki et al., 2016), Iceland (Tobler and Benning, 2011; Menzel et al., 2015), China (Hou et al., 2013; Song et al., 2013), and India (Badhai et al., 2015; Sangwan et al., 2015), are extensively studied. Among these, the Octopus and Mushroom Springs at Yellowstone National Park have the longest history of research of nearly 50 years (Thiel et al., 2016). Recently, a research team in New Zealand initiated the 1,000 Springs Research Project to examine geothermal microbial biota in 1,000 hot-spring features in the Taupo Volcanic Zone<sup>1</sup>. Microbial populations in heated springs are influenced by physicochemical factors, such as temperature, pH, dissolved oxygen, and water chemistry (Mathur et al., 2007; Sharp et al., 2014; Chan et al., 2015). In hot springs with temperatures higher than 75°C, Aquificae, Deinococcus-Thermus, Thermodesulfobacteria, Thermotogae, and some thermophilic members of Proteobacteria and Firmicutes are the commonly found bacterial phyla (Blank et al., 2002; Hou et al., 2013; Song et al., 2013), in addition to archaeal phyla such as Crenarchaeota, Euryarchaeota, and Thaumarchaeota (Hou et al., 2013; Nishiyama et al., 2013; Chan et al., 2015). In hot springs with lower temperatures, thermophilic photosynthetic bacteria such as Cyanobacteria and Chloroflexi, together with phyla such as Proteobacteria may be the main population (Kubo et al., 2011; Nishiyama et al., 2013; Badhai et al., 2015). As the maximum temperature for photosynthesis is 75°C (Ferris and Ward, 1997), Cyanobacteria and Chloroflexi are therefore not predominantly present in hot springs with temperatures higher than this threshold.

The Banjaran Titiwangsa mountain range, with elevations of 900–2,100 m and a length of about 480 km, is the most prominent mountain cluster on Peninsular Malaysia. Most of the hot springs on the peninsula are located along the western flank of Banjaran Titiwangsa and are concentrated along major fault zones (Sum et al., 2010). The Malaysian hot springs differ from the geothermal systems in Yellowstone National Park, Japan, Kamchatka, and New Zealand in several aspects. Malaysia has a tropical climate and is located in a non-volcanic area. The water arising from these hot springs is heated geothermally; groundwater that percolates deeply into the Earth's crust comes

in contact with the rocks that get heated as a result of the geothermal gradient originating from the Earth's interior. Pressure is generated and forces the water to discharge through pores and fissures within the Earth's crust toward the surface to form hot springs (Baïoumy et al., 2015). In countries with volcanic activities in tectonically active zones, rain or melted snow comes in contact with near subsurface magma that may heat the water sufficiently to form superheated water bodies. Acidic hot springs (pH < 6.0) are often found in volcanic areas; however, hot springs in Malaysia are circumneutral or slightly alkaline. Hot-spring water usually has high concentrations of various elements owing to mineralization of dissolved solid elements from the adjacent areas. The composition of hot water is mainly determined by chemical interactions with reservoir rocks and rock-forming minerals along the ascent path, which may cause the spring water to be acidic or alkaline.

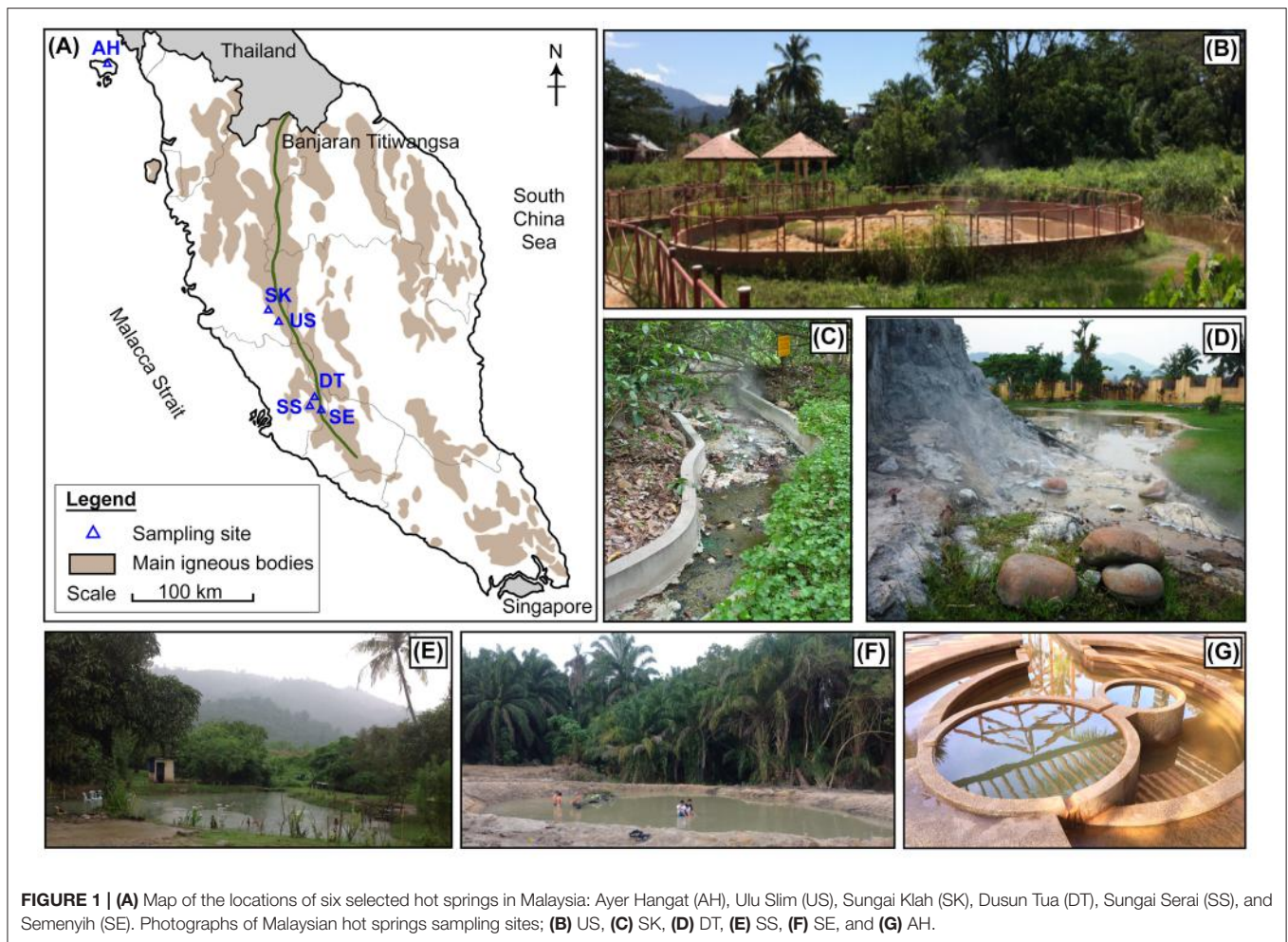
Earlier reports summarized the different physicochemical conditions in 46 known Malaysian hot springs (Samsudin et al., 1997; Sum et al., 2010). These hot springs have temperatures ranging from 36°C to 102°C, but predominantly lower than 80°C. Malaysian hot springs exhibit differences in physical appearance and chemical contents, which are expected to influence the microbial communities; nevertheless, limited studies have been conducted to understand the microbial diversity in Malaysian (Goh et al., 2011; Chan et al., 2015) and South-East Asian (Baker et al., 2001; Kanokratana et al., 2004; Huang et al., 2013) hot springs. In this report, a coordinated geochemical and molecular survey was conducted for six hot springs. Ulu Slim, Sungai Klah, and Dusun Tua were selected as these sites are the hottest geothermal springs in Malaysia. Ayer Hangat was chosen because it is the only saline hot spring in Malaysia. Semenyih and Sungai Serai hot springs were selected as they exhibit similar temperature (~45°C) and pool size and are located close to each other. This work utilized 16S rRNA gene markers to understand microbial diversity and community structure. We expected our findings to elucidate possible relationships with hot-spring physicochemical variables.

## MATERIALS AND METHODS

### Study Sites and Water Physicochemical Characteristics

Hot springs at Ulu Slim (US) (3°53'55.79"N, 101°29'52.44"E), Sungai Serai (SS) (3°5'27.71"N, 101°47'39.06"E), Dusun Tua (DT) (3°8'21.23"N, 101°50'10.33"E), and Semenyih (SE) (3°2'32.81"N, 101°52'19.87"E) are located along the Banjaran Titiwangsa main mountain range (**Figure 1**). Additionally, the Ayer Hangat (AH) (6°25'22.31"N, 99°48'48.97"E) hot spring, located outside the main mountain range, was included in this study because of its high saline water content (**Table 1**). At each hot spring, water temperature was measured using a portable thermometer. Water and sediment samples were collected in sterile bottles that were closed immediately after sampling. Samples were collected in at least three sites at each hot spring. The samples were kept at ambient temperature

<sup>1</sup><http://www.1000springs.org.nz>



and transferred immediately to the laboratory, where they were stored at 4°C prior to DNA extraction. Water samples of each site were sent to the Allied Chemists Laboratory Sdn. Bhd. (Malaysia) for physiochemical analyses in accordance with the American Public Health Association, Standard Methods for the Examination of Water and Wastewater (APHA) and United States Environmental Protection Agency (USEPA) guidelines (refer to **Table 2** for the list of analyses).

## DNA Extraction and 16S rRNA Gene Sequencing

DNA extraction was conducted in the laboratory. Equal volumes of water and sediment samples for each hot spring were collected and pooled to represent the overall microbiome of each site. DNA extraction and amplicon generation from environmental samples were conducted as previously described (Chan et al., 2015), with modification of the amount of starting material used for DNA extraction. In brief, genomic DNA extracted from the samples was amplified and dual-index barcoded for multiplex sequencing using primer pair (S-D-Bact-0341-b-S17 and S-D-Bact-0785-a-A-21) targeted to the V3–V4 regions

of the 16S rRNA gene (Klindworth et al., 2013). The forward (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-CCTACGGGNGGCWGCAG-3') and reverse (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-GACTACHVGGGTATCTAATCC-3') primers contained Illumina overhang adapter sequences (underlined regions). Amplicons were paired-end (2 × 300 bp, MiSeq v3 reagent) sequenced on an Illumina MiSeq sequencer (San Diego, CA, USA) at the High Impact Research Institute at the University of Malaya, Malaysia. The sequence data have been submitted to NCBI SRA under Bioproject accession number PRJNA378468. To compare microbial profiles obtained in this study with those reported earlier for the Malaysian SK hot spring (Chan et al., 2015), the SRA for the latter (Bioproject no. PRJEB7059) was retrieved and processed similarly to the reads generated in this work.

## Data Analysis

The raw sequence data were evaluated and filtered to ensure that >80% of the base calls in a sequence had a Phred quality score of 20 using the FASTQ Quality Filter ( $q = 20$ ,

**TABLE 1** | Descriptions of the hot springs studied.

Name	Sample type <sup>a</sup>	Descriptions
Ulu Slim (US)	Water and sediment (spring heads and pond)	<ul style="list-style-type: none"> <li>• The hottest spring in Malaysia.</li> <li>• A roughly cylindrical shallow pool (~4 m diameter, ~10 cm depth) with two moderate spring heads (110°C).</li> <li>• Clear water and fine clays at the bottom.</li> <li>• Biofilms (&lt;5 mm thickness) with colors of yellow, orange, light brown, and light green are formed on the spring.</li> </ul>
Sungai Klah <sup>b</sup> (SK)	Water and sediment (middle of stream)	<ul style="list-style-type: none"> <li>• Shallow water flow stream (~150 m length, ~1.5 m width, ~5 cm depth)</li> <li>• Water temperatures between 50–110°C.</li> <li>• Multiple spring heads with clear running water.</li> <li>• Clays, rocks, and sands at the bottom.</li> </ul>
Dusun Tua (DT)	Water and sediment (spring head and pond)	<ul style="list-style-type: none"> <li>• A shallow pool with small spring source (75°C) outflow continually from a 1.5-m man-made cement-fountain landscape.</li> <li>• Multiple thick biomats (1–5 cm) of various colors (dark brown, red, white, dark green, and orange) are formed on the fountain, while a green filamentous mat is present on the bank of the collecting pool.</li> <li>• Clear water with silicate sands and clays at the bottom.</li> </ul>
Sungai Serai (SS)	Water and sediment (center and edge of pool)	<ul style="list-style-type: none"> <li>• Roughly round-shaped pool with diameter ~10 m, and depth ~0.5 m.</li> <li>• Many bubbling sources, but non-visible spring head.</li> <li>• Murky water with green biomats on the bank of the pool.</li> <li>• Loam soil at the bottom.</li> </ul>
Semenyih (SE)	Water and sediment (center and edge of pool)	<ul style="list-style-type: none"> <li>• Roughly round-shaped pool with diameter ~10 m, and depth ~0.7 m.</li> <li>• Many bubbling sources and no visible outflow.</li> <li>• Non-visible spring head.</li> <li>• Murky water with green biomats on the bank of the spring.</li> <li>• Fine silicate sands at the bottom.</li> </ul>
Ayer Hangat (AH)	Water	<ul style="list-style-type: none"> <li>• The only saline hot spring in Malaysia.</li> <li>• Water is trapped in a deep, cylindrical man-made pond with ~1-m diameter.</li> <li>• Formation of small degassing bubbles with slow outflow.</li> <li>• Gray-yellowish water mixed with fine salt.</li> <li>• Thin light-brown and green biomats (&lt;1 cm) are formed on the surrounding wall.</li> </ul>

<sup>a</sup>Sample type used for genomic DNA extraction.

<sup>b</sup>Data collected from *Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing* (Chan et al., 2015).

$p = 80$ ) of the FASTX-Toolkit<sup>2</sup>. Sequences that passed the quality filtering, were free from ambiguous characters, and of  $\geq 200$  bp were merged using PEAR (Zhang et al., 2014). Chimeric sequences were identified and discarded using the UCHIME algorithm (Edgar et al., 2011) implemented in the USEARCH package (Edgar, 2010). The sequences were then analyzed using the Quantitative Insights Into Microbial Ecology (QIIME) pipeline version 1.9.1 (Caporaso et al., 2010b) with default parameters unless otherwise noted. Briefly, sequences were clustered into operational taxonomic units (OTUs) at 97% similarity with USEARCH-based (Edgar, 2010) open-reference OTUs picking protocols using the Greengenes 13\_8 reference sequences (McDonald et al., 2012). Taxonomy was assigned to sequences using UCLUST (Edgar, 2010), retrained on Greengenes 13\_8 with a minimum of one OTU size, in QIIME. Representative sequences, which were selected as the centroid sequence of each OTU, were aligned using PyNAST aligner (Caporaso et al., 2010a) with a minimum sequence

length of 26 bases, and those that failed to align were removed.

## Statistical Analysis

Multivariate principal component analysis (PCA) of 23 physicochemical variables including sampling-site water temperature, pH, alkalinity, acidity, color, turbidity, aluminum, arsenic, chloride, fluoride, CaCO<sub>3</sub>, iron, magnesium, phosphate, sodium, sulfate, sulfur, total nitrogen (TN), total organic carbon (TOC), biochemical oxygen demand under 5-day incubation (BOD; at 20, 60, and 80°C), and chemical oxygen demand (COD) was carried out to determine which physicochemical variables were related to the observed hot spring community patterns. All measured physicochemical variables were checked for normality and log-transformed before PCA. A PCA plot was generated using PAST software version 3.13 (Hammer et al., 2001).

To estimate species richness from the data sets, the original OTU table with singletons was used to conduct rarefaction analysis and estimate alpha diversity. Rarefaction curves based on observed OTUs were generated using QIIME. The

<sup>2</sup>[http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit)

**TABLE 2** | Physicochemical properties of water samples from the six Malaysian hot springs studied.

Test variable	Methods	Unit	US	SK <sup>a</sup>	DT	SS	SE	AH
Temperature range	Thermometer	°C	80–110	60–110	55–75	40–45	40–50	40–50
Sampling temperature	Thermometer	°C	90	75	70	43	43	45
pH <sup>b</sup>	APHA 4500 H <sup>+</sup> B		7.2	8.2	7.0	6.9	6.9	7.1
Alkalinity	APHA 2320 B	mg L <sup>-1</sup>	94	76	106	122	136	294
Acidity	APHA 2310 B	mg L <sup>-1</sup>	<1	<1	<1	<1	20	<1
Color	APHA 2120 B	TCU	<5	75	<5	<5	<5	5
Turbidity	APHA 2130 B	NTU	1.0	130	<0.05	1.0	0.5	1.1
C:N ratio (TOC/TN) <sup>c</sup>			1.5	1.6	0.1	0.3	0.1	5.0
Aluminum (Al)	APHA 3030 F/USEPA 6010 B	mg L <sup>-1</sup>	0.07	0.96	0.04	0.04	0.04	0.04
Arsenic (As)	APHA 3030 F/USEPA 6010 B	mg L <sup>-1</sup>	0.03	0.07	ND (<0.01)	ND (<0.01)	ND (<0.01)	ND (<0.01)
Chloride (Cl <sup>-</sup> )	APHA 4500-Cl <sup>-</sup> B	mg L <sup>-1</sup>	<1	2	1	3	4	13,832
Fluoride (F <sup>-</sup> )	APHA 4500-F <sup>-</sup> D	mg L <sup>-1</sup>	0.4	1.1	6.9	11	9.4	21
Hardness (CaCO <sub>3</sub> )	APHA 2340C	mg L <sup>-1</sup>	13	<1	5	27	27	5,020
Iron (Fe)	APHA 3030 F/USEPA 6010 B	mg L <sup>-1</sup>	ND (<0.02)	0.65	ND (<0.02)	0.03	ND (<0.02)	ND (<0.02)
Magnesium (Mg <sup>+</sup> )	APHA 3030 F/USEPA 6010 B	mg L <sup>-1</sup>	0.8	0.5	<0.1	0.2	0.2	394
Phosphate (PO <sub>4</sub> <sup>3-</sup> )	APHA 3030 G/USEPA 6010 B	mg L <sup>-1</sup>	0.1	0.2	0.1	0.3	0.7	0.4
Sodium (Na <sup>+</sup> )	APHA 3030 F/USEPA 6010 B	mg L <sup>-1</sup>	43	27	51	45	48	7,905
Sulfate (SO <sub>4</sub> <sup>2-</sup> )	APHA 4500-SO <sub>4</sub> E	mg L <sup>-1</sup>	3	8	6	1	1	947
Sulfur (S)	APHA 3030 F/USEPA 6010 B	mg L <sup>-1</sup>	5.2	3.9	2.7	0.5	1.1	477
TN	APHA 3030 F/USEPA 6010 B	mg L <sup>-1</sup>	<0.2	5.6	2.7	3.0	6.1	<0.2
TOC	APHA 5310 B	mg L <sup>-1</sup>	0.3	9.04	0.4	0.8	0.9	1.0
BOD 5 days at 20°C	APHA 5210 B	mg L <sup>-1</sup>	<5	5	<5	<5	<5	30
BOD 5 days at 60°C	APHA 5210 B	mg L <sup>-1</sup>	<5	10	<5	<5	<5	20
BOD 5 days at 80°C	APHA 5210 B	mg L <sup>-1</sup>	<5	5	<5	<5	<5	<5
COD	APHA 5220 B	mg L <sup>-1</sup>	7.4	35	<5	<5	<5	90

<sup>a</sup>Data collected from *Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing* (Chan et al., 2015).

<sup>b</sup>The pH values were measured at room temperature.

<sup>c</sup>The C:N ratios were calculated on mass basis.

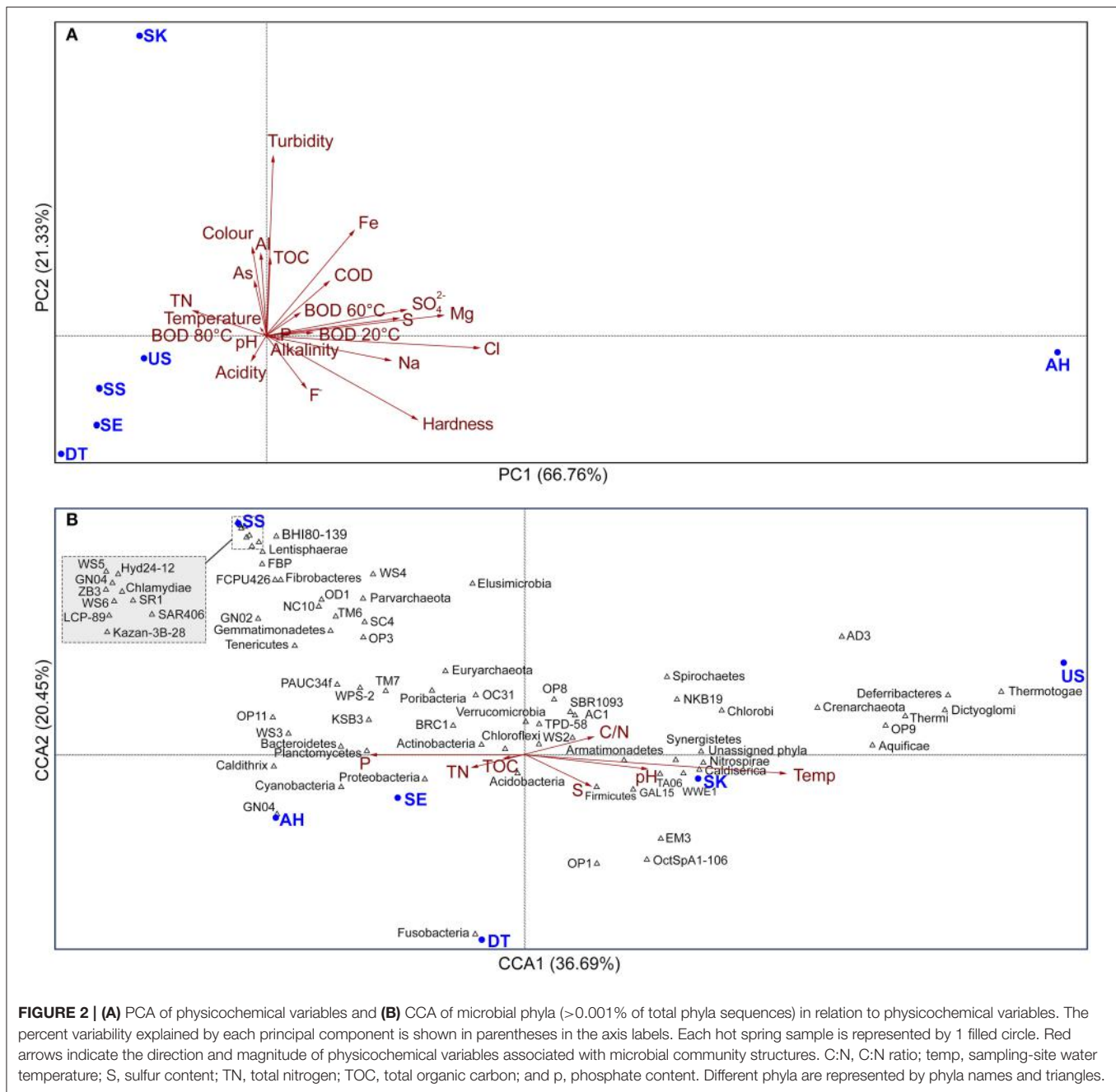
Abbreviations: APHA, in accordance with American Public Health Association, Standard Methods for the Examination of Water and Wastewater; USEPA, United States Environmental Protection Agency; TN, total nitrogen; TOC, total organic carbon; BOD, Biochemical oxygen demand; COD, Chemical oxygen demand; ND, not detected.

alpha diversity estimates, including Good's coverage, Shannon–Wiener's diversity index, and Simpson's index of diversity, were calculated. Sampling completeness was evaluated using the Good's average estimator, which calculates the probability that a randomly selected amplicon sequence from a sample has already been sequenced (Good, 1953). The Shannon–Wiener diversity index (Spellerberg and Fedor, 2003) was used to explain the entropy, taking into account the species richness and evenness of the community, which varied from 0 for communities with a single taxon, to high values for highly diverse communities. Simpson's index of diversity (1-D) (Simpson, 1949) was used to describe the diversity in a community, ranging from 0 to 1, with 1 indicating maximum diversity in a sample.

Singletons were discarded and the OTU table was rarefied to a depth of 313,337 sequences per sample (75% sequences to the lowest number of sequences found among the sample dataset) to minimize the effect of sampling effort. Further, diversity analyses including beta diversity analysis were conducted using QIIME, with the script `core_diversity_analyses.py`. Jackknifed unweighted pair group method with arithmetic mean (UPGMA) clustering was performed to compare microbial community

similarity among the hot spring samples based on UniFrac phylogenetic distances (weighted and unweighted) (Lozupone et al., 2007) and non-phylogenetic Bray–Curtis dissimilarity distances. A relative small UniFrac distance implies that two communities are compositionally similar, harboring lineages that share a common evolutionary history. Unweighted UniFrac only accounts for the community membership, while weighted UniFrac accounts for community structure (membership and relative abundance) (Lozupone et al., 2007). Biplots of principal coordinate analysis (PCoA) were generated in QIIME and visualized in 3D using Emperor (Vázquez-Baeza et al., 2013).

Canonical correspondence analysis (CCA) was performed using PAST software, to explore relationships of microbial community patterns at the phylum level with physicochemical variables. By considering that predominant species have greater influence within the communities, only 75 major OTUs with relative abundance of >0.001% across all sample data sets were used as a community matrix for CCA. The significance of the CCA models and the explanatory factors were tested using 999 permutations. The ordination on the x- and y-axis and the length of the corresponding arrows indicate the relative importance



of physicochemical variables explaining the taxon distribution across communities. Unless specified otherwise, all statistical analyses were done in QIIME and/or PAST (Hammer et al., 2001).

## RESULTS

### Site Descriptions and Physicochemical Characteristics of Hot Springs

In this study, six hot springs with different physical and chemical conditions were studied (Table 1). Basically, the selected sites could be categorized as: (1) pond with high NaCl content (AH hot

spring), (2) fast-flowing streamer (SK hot spring), and (3) non-saline pool with standing water (US, DT, SS, and SE hot springs). Data on microbial diversity in SK were previously reported by our group (Chan et al., 2015) and were compared with data for the other five hot springs studied in the current work (Figure 1, Table 1). The six hot springs exhibit different temperatures (40–110°C) and pH (6.9–8.2). The highest temperature was recorded at US (80–110°C) followed by SK (60–110°C) and DT (55–75°C). Hot springs SS, SE, and AH have lower temperatures (40–50°C). The water chemistry including alkalinity, acidity, color, turbidity, and major ions varied among the hot springs (Table 2). In comparison to other sites, SK water has the highest turbidity,

**TABLE 3** | Sequencing data profiles and alpha diversity indexes.

	US	SK <sup>a</sup>	DT	SS	SE	AH
Number of raw sequences	1,053,625	480,983	1,028,376	1,761,134	1,167,238	1,100,225
Average size read (bp)	35–301	35–301	35–301	45–301	35–301	35–301
Mean GC content (%)	58	56	54	53	54	55
High quality sequences <sup>b</sup>	1,013,171	429,677	952,470	1,650,101	1,084,632	1,057,048
Cleaned sequences <sup>c</sup>	990,519	424,188	914,072	1,627,045	987,849	984,566
PyNAST aligned sequences	987,974	421,383	912,827	1,608,109	985,204	982,466
Sequences without singletons	974,062	417,783	898,358	1,569,331	970,574	968,368
Total OTUs <sup>d</sup>	7,326	6,334	9,083	26,244	11,504	6,430
Taxonomy assigned sequences	801,105	376,334	861,537	1,264,280	912,975	909,979
Bacteria	785,038	373,507	857,088	1,243,916	908,541	909,890
Archaea	16,067	2,827	4,449	20,364	4,434	89
Unassigned sequences	172,957	41,449	36,821	305,051	57,599	58,389
<b>ALPHA DIVERSITY ANALYSIS</b>						
Good's coverage	0.985	0.987	0.983	0.974	0.982	0.985
Shannon-Wiener	6.505	9.020	6.123	9.929	7.128	7.513
Simpson	0.962	0.992	0.933	0.984	0.955	0.977

<sup>a</sup>Data collected from Bioproject accession number PRJEB7059.

<sup>b</sup>Sequences which passed the quality filtration and sequence read merging process.

<sup>c</sup>Sequences obtained after chimera removal.

<sup>d</sup>Amount of observed OTUs after singletons removal.

TOC, aluminum, arsenic, and iron, while AH had distinctly high concentrations of chloride, fluoride, CaCO<sub>3</sub> (hardness), magnesium, sodium, sulfate, and sulfur. The chemical contents in water samples from high-temperature springs (US and DT) are relatively similar, except for the concentrations of fluoride, CaCO<sub>3</sub>, magnesium, and TN. On the other hand, the water of two low-temperature hot springs (SE and SS) differs in the concentrations of iron, phosphate, TN, and acidity. Additionally, the sulfur content in AH, US, and SK is relatively high, probably associated with their location in the northern part of Peninsular Malaysia. The TN content in SK and SE is quite high, with 5.6 mg L<sup>-1</sup> and 6.1 mg L<sup>-1</sup>, respectively, while it is <0.2 mg L<sup>-1</sup> for US and AH. Water in DT, SS, and SE has a carbon-to-nitrogen ratio (C:N) of 0.1–0.3, while the C:N for US and SK is quite similar (1.5–1.6). Besides its high COD and BOD levels, AH water has a high C:N ratio of 5.

PCA of the physicochemical variables separated the hot springs into three physicochemically distinct habitats (Group-1: AH; Group-2: SK, and Group-3: US, DT, SS, and SE). **Figure 2A** shows the distribution of the physicochemical variables formed by the first two components of the analysis, which explained 88.09% of the total variance. PC1 explained 66.76% of the observed variation, and clearly separated the saline hot spring AH from the non-saline hot springs (US, SK, DT, SS, and SE). PC1 was correlated with chloride, magnesium, sodium, sulfate ions, hardness, and sulfur, and high values for these variables might be related to the presence of seawater, as the AH hot spring is located on an island (**Figure 1A**). PC2, accounting for 21.33% of the total variation, was related primarily to pH, color, turbidity, aluminum, iron, and TOC. The relatively higher values of these variables in SK separated this hot spring from the other five springs; the presence of plant litter in SK increases the TOC

(Chan et al., 2015). Group-3, a separate cluster of US, DT, SS, and SE, showed inverse correlations with most of the physicochemical variables, except temperature, TN, and acidity (**Figure 2A**).

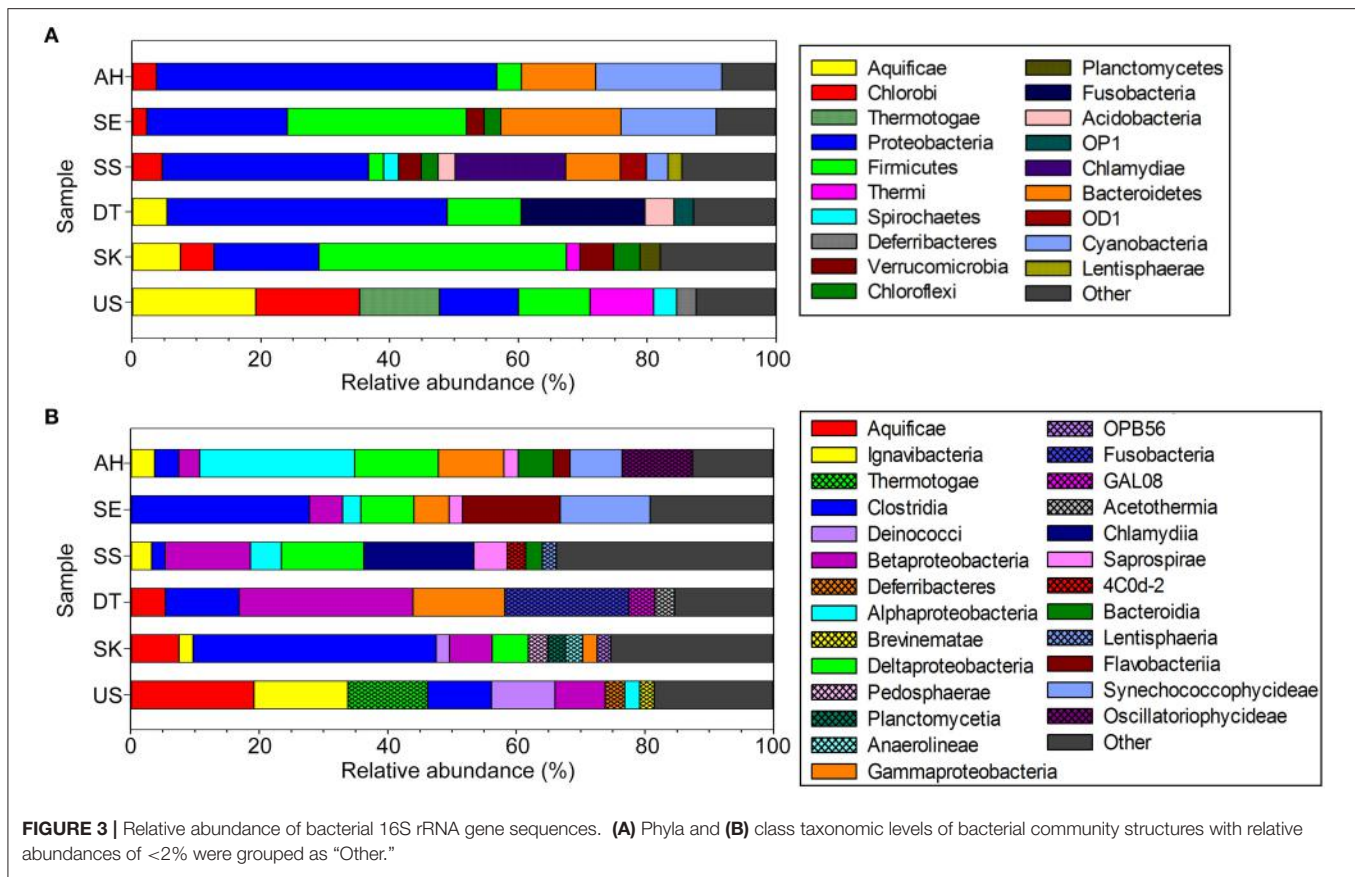
## 16S rRNA Gene Microbial Profiles

As summarized in **Table 3**, the sizes of the six 16S rRNA gene sequence datasets ranged from 0.60 to 2.32 Gb, with between 480,983 and 1,761,134 paired-end reads. After quality filtering, sequence merging, and chimera removal, more than 98.8% of the cleaned sequences were aligned against Greengenes database using the PyNAST (Caporaso et al., 2010a) alignment algorithm. After removal of the singletons (sequences that are present exactly once in a sample), 80.6–95.9% of the aligned sequences were taxonomically classified. About 4.1–19.4% of these sequences could not be assigned to known taxa, perhaps due to a lack of suitable reference sequences in the database. The majority of the assigned OTUs were classified as bacteria, i.e., 97.99% for US, 99.25% for SK, 99.48% for DT, 98.39% for SS, 99.51% for SE, and 99.99% for AH. Archaeal OTUs comprised relatively small percentages: 2.01% for US and 1.61% for SS, and <1% of the total populations for other sites (SK: 0.75%, DT: 0.52%, SE: 0.49%, and AH: 0.01%).

## Taxonomic Composition of the Prokaryotic Communities

### Bacterial Diversity

Within the domain Bacteria, more than 99.4% of the assigned sequences were classified at the phylum level. In total, 59, 61, 72, 73, 65, and 52 bacterial phyla were detected in US, SK, DT, SS, SE, and AH, respectively, yet less than six phyla are dominant in each hot spring with >10% relative abundance of total sequences at the phylum level. The bacterial communities within



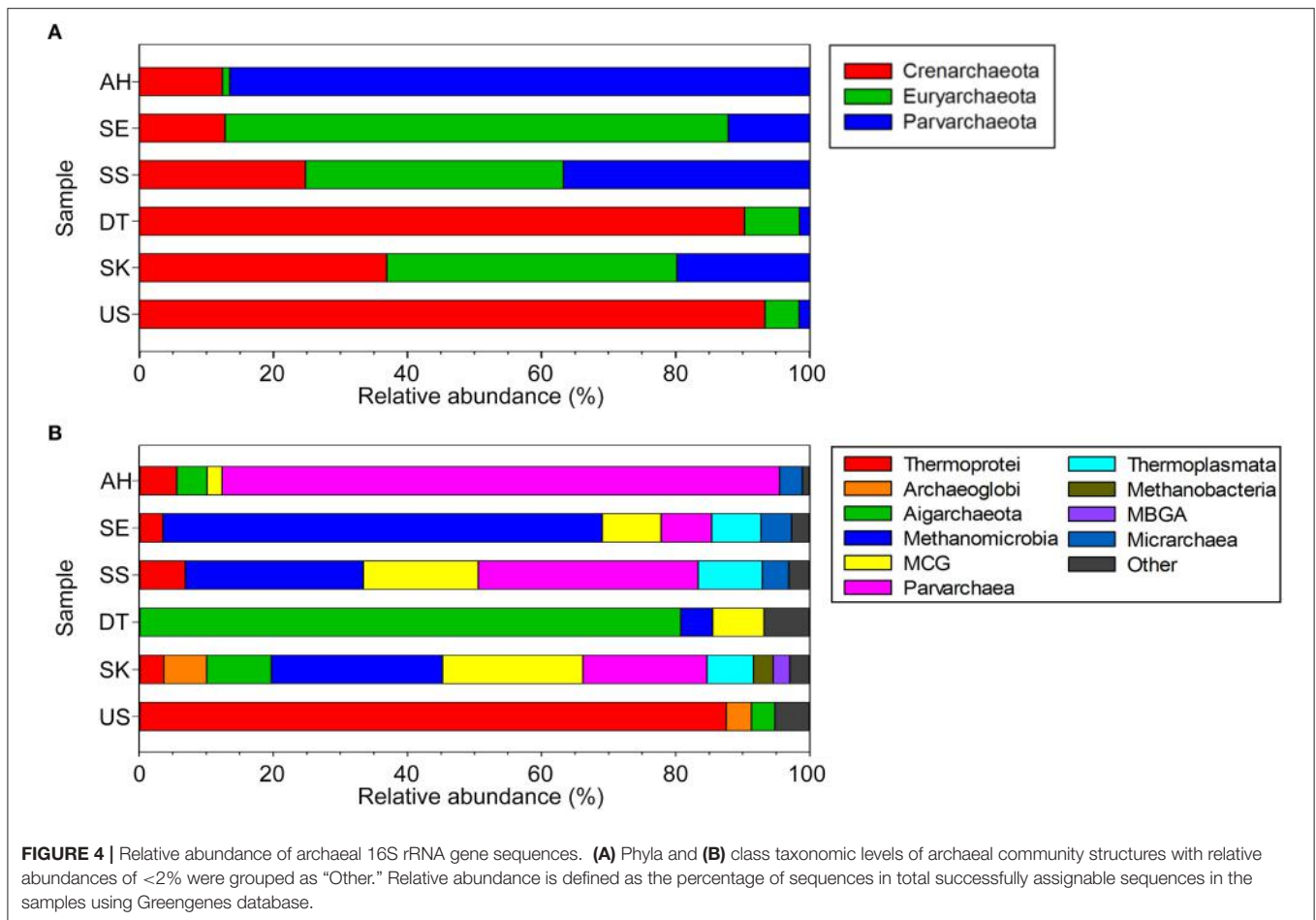
each site were unique. In general, the phyla Aquificae (relative abundance of 19.2%), Chlorobi (16.1%), Thermotogae (12.4%), Proteobacteria (12.2%), and Firmicutes (11.1%) dominated US. Phyla Firmicutes (38.5%) and Proteobacteria (16.3%) dominated SK, while Proteobacteria (43.5%), Fusobacteria (19.3%), and Firmicutes (11.5%) dominated DT. The SS hot spring was dominated by Proteobacteria (32.1%) and Chlamydiae (17.2%). SE water samples had great proportions of Firmicutes (27.8%), Proteobacteria (21.8%), Bacteroidetes (18.7%), and Cyanobacteria (14.8%). AH was dominated with Proteobacteria (52.8%), Cyanobacteria (19.6%), and Bacteroidetes (11.5%) (**Figure 3A**).

Firmicutes can be regarded as a signature phylum for circumneutral hot springs based on their abundance. Among the studied sites, SK and SE have the greatest dominance of Firmicutes (38.5 and 27.8% of total phyla, respectively), followed by US and DT (~11%), while this phylum is relatively less abundant in SS and AH hot springs (<4%). In general, at least 89.6% of total Firmicutes OTUs were assigned to the class Clostridia (mainly in the order Clostridiales, comprising obligate anaerobe and endospore-forming bacteria) (**Figure 3B**), while the remaining classes were Bacilli and Erysipelotrichi. In SK and SE, Firmicutes were mainly represented by the genera *Pelosinus* (15.4% of total Firmicutes) and *Acidaminobacter* (74.7%). Furthermore, Firmicutes in SE were dominated by the genus *Acidaminobacter*, with 37.8% of the total genera.

Proteobacteria form another signature phylum as the phylum was dominant in DT, SS, and AH (43.5, 32.1, and 52.8% of total phyla, respectively). High abundance of Proteobacteria was also detected in the US (12.2%), SK (16.3%), and SE (27.8%) hot springs. Betaproteobacteria form the most abundant Proteobacteria class in US, SK, DT, and SS, while AH was dominated by Alphaproteobacteria (**Figure 3B**). Moreover, Alpha-, Beta-, Gamma-, and Deltaproteobacteria were detected in all six Malaysian hot springs, but at different percentages. DT was dominated by *Vogesella* (43.3% of total genera), a genus within Betaproteobacteria. *Vogesella* is frequently found in freshwater bodies and currently, seven species have been described in this genus. *Vogesella lacus* (Chou et al., 2009) and *Vogesella perlucida* (Chou et al., 2008) are the only species that can grow at temperatures <40°C. The genus *Hahella* represented the main Gammaproteobacteria in AH (13.9%). *Hahella* are marine bacteria that require NaCl to grow. To date, only three species within the genus *Hahella* have been described, i.e., *H. chejuensis* (Lee et al., 2001), *H. ganghwensis* (Baik et al., 2005), and *H. antarctica* (Lee et al., 2008). *H. chejuensis* is able to grow at temperatures of up to 45°C (Lee et al., 2001).

Owing to its high water temperature (80–110°C), hyperthermophilic phyla including Aquificae and Thermotogae prevailed in US. Both phyla were also detected in SK (7.5% Aquificae and 1.7% Thermotogae) which has a considerable high temperature, but were insignificant in SS, SE, and AH (<0.1% of





total phyla). Aquificae in US and SK were mainly comprised of the genus *Hydrogenobacter*, with 98.9% in US and 88.6% in SK of total Aquificae, equivalent to 31.6 and 13.3% of total OTUs at the genus level, respectively. With an average temperature of 75°C, in DT, Aquificae (5.4% of total phyla) were more abundant than Thermotogae (0.2%).

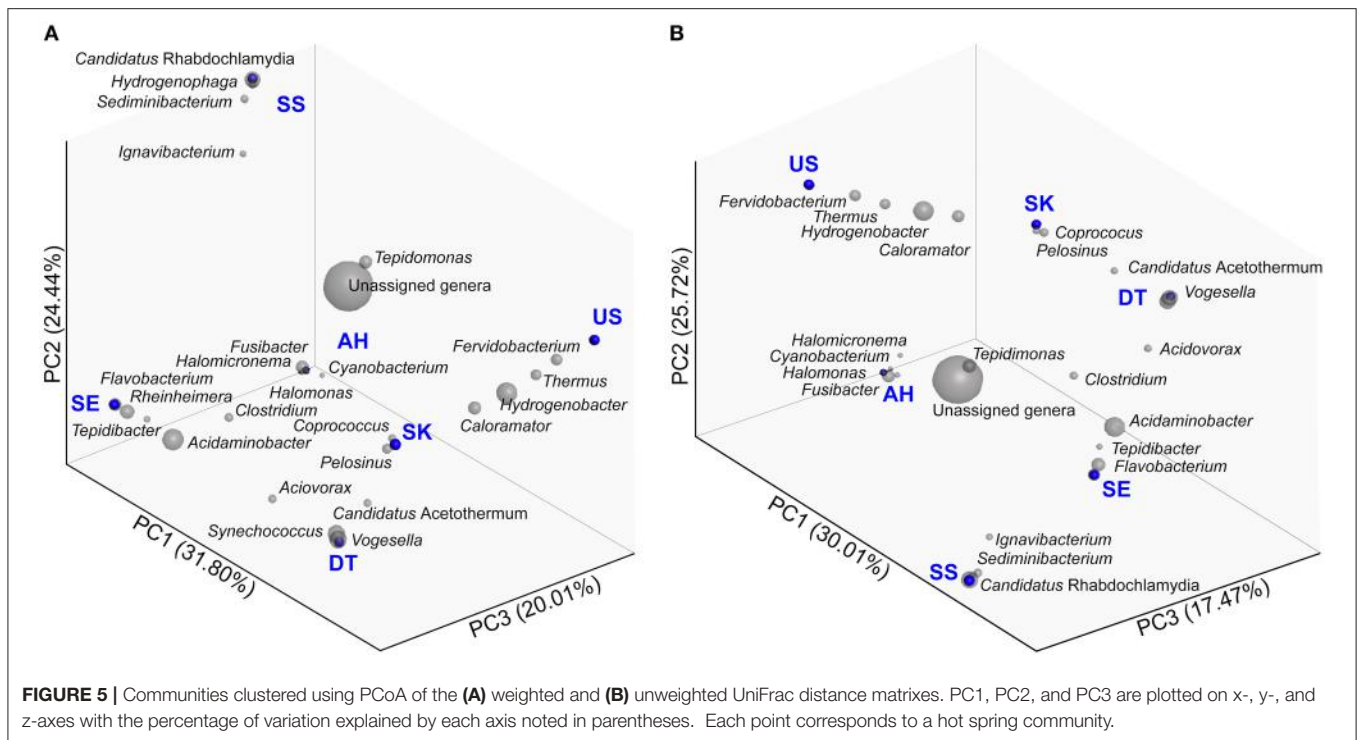
### Archaeal Diversity

Only 3 archaeal phyla, Euryarchaeota, Parvarchaeota, and Crenarchaeota, were detected in the studied sites (**Figure 4A**). Euryarchaeota were found to be the main archaea in SK (43.3% total archaea phylum), SS (38.5%), and SE (75.1%). Parvarchaeota prevailed in AH, with 86.5% of total archaeal phyla. An earlier metagenomics study elucidated that members of Parvarchaeota are very small in size (cells are <500 nm in diameter) and are from lineages without cultivated representatives that branch near the crenarchaeal/euryarchaeal divide (Baker et al., 2010). In SK, SS, and SE hot springs, the majority of the Euryarchaeota were methanogenic archaea of the class Methanomicrobia, genus, *Methanosaeta*. At present, *Methanosaeta* is the only genus affiliated to the Methanosaetaceae family. *Methanosaeta* spp. utilize acetate as a sole energy source to produce methane and carbon dioxide (Welte and Deppenmeier, 2011).

The phylum Crenarchaeota constituted the major archaeal member in the US (93.3%) and DT (90.3%) hot springs. In US, 93.7% of total Crenarchaeota were related to the class Thermoprotei. The hyperthermophilic genus *Aeropyrum* (90.4% of total genera) was identified in US; strains of this genus grow optimally at 85–95°C (Sako et al., 1996; Nakagawa et al., 2004). In DT (**Figure 4B**), 55% of the detected Aigarchaeota showed the closest similarity to the uncultivated archaeon “*Candidatus Caldiarchaeum*” (Nunoura et al., 2011). *Ca. C. subterraneum* was first discovered in a non-cultured metagenomic library from a microbial mat at a geothermal water stream of a sub-surface gold mine with a temperature of 70°C (Hirayama et al., 2005). It was proposed that *Ca. C. subterraneum* lives symbiotically with acetogenic “*Candidatus Acetothermus autotrophicus*” for organic carbon supply, since *Ca. C. subterraneum* has an extremely poor carbon fixation potential (Takami et al., 2015).

### Alpha Diversity Analysis

Since rarefaction curves did not reach a plateau (Supplementary Figure 1), total species richness was not estimated. Therefore, Good’s coverage estimator was used; all six hot spring samples had an estimated coverage of at least 97% of the entire sampled population (**Table 3**). Additionally, non-phylogeny-based metrics including Shannon–Wiener’s and Simpson’s



diversity indexes were computed (Table 3). A greater (Shannon–Wiener) diversity of OTUs was found in SS (9.9) and SK (9.0), followed by AH (7.5) and SE (7.1). Both US (6.5) and DT (6.1) samples have relatively lower diversity, based on both the Shannon–Wiener and Simpson indexes. It has been reported that high temperature has a negative effect on diversity (Sharp et al., 2014; Li et al., 2015). The microbial communities in high-temperature hot springs (US and DT, >70°C) were dominated by fewer genera than those in low-temperature sites (SS, AH, and SE, <50°C). Nevertheless, although the SK hot spring has a high water temperature of 50–110°C, the diversity index for SK is high.

## Beta Diversity Analysis

The above alpha diversity metrics provided an overview of the microbial diversity of each sampling site in this study. On the other hand, beta diversity is useful for documenting the structure of communities that may occur between samples categories or across environmental gradients (Lozupone and Knight, 2005). Here, we compared the six Malaysian hot spring samples using jackknifed UPGMA to cluster the community samples with weighted and unweighted UniFrac phylogenetic distances (Lozupone and Knight, 2005; Lozupone et al., 2007). Using PCoA with weighted UniFrac (Figure 5A), PC1 explains 32% of the variation, while PC2 and PC3 explain 24 and 20%, respectively. The AH hot spring presented as an outlier in PCoA owing to its high salinity. Weighted UniFrac PCoA and the corresponding dendrogram (Supplementary Figure 2A) were unable to show the effect of temperature on the microbial communities in the six hot springs. In comparison, PCoA with unweighted UniFrac (73% of total variation explained by PC1, PC2, and PC3) (Figure 5B) and the corresponding dendrogram (Supplementary Figure 2B) yielded a clearer clustering that sorted the hot springs

into three groups in accordance to variation in temperature and salinity. The three groups are: Group-1: high salinity with moderate temperature (AH); Group-2: low salinity with moderate temperature (SE and SS); and Group-3: low salinity with high temperature (US, SK, and DT). A similar grouping pattern was identified based on Bray–Curtis dissimilarity using the jackknifed UPGMA dendrogram (Supplementary Figure 2C).

## Relationships between Microbial Community and Physicochemical Variables

Multivariate analyses are widely used in ecology studies to elucidate the relationships between abundances of certain organisms and environmental parameters. In this study, the CCA approach (ter Braak and Verdonschot, 1995) was used to identify possible relationships between microbial communities in hot springs and local physicochemical variables. Only physicochemical variables with normally distributed values, including temperature, pH, C:N, TOC, TN, sulfur, and phosphate, were used in this study (Figure 2B). The placement of hot spring samples in the plot is influenced by their environmental characteristics. The eigenvalues for each axis generated by CCA indicate how much of the variation seen in the genera data can be explained by the canonical axis. In this analysis, 61% of the correlation between OTUs, hot spring samples, and physicochemical variables were explained by two axes.

Most variables, including temperature, pH, C:N, and sulfur content showed a positive correlation (Figure 2B), while inverse correlations were observed for TOC, TN, and phosphate content. Among the tested variables, temperature was the most influential variable affecting the distribution of OTUs. In Figure 2B, plots

for US and SK are situated in a right side, as these sites are characterized by high temperatures, while SS, AH, and SE hot springs fall in the opposite side of the plot. The right side was occupied by thermophilic phyla, including Thermotogae, Dictyoglomi, Thermi, Aquificae, Caldiserica, Crenarchaeota, Deferribacteres, candidate divisions OP9 (“Atribacter”) (Nobu et al., 2016), OP1, EM3, and OctSpA1-106 (Youssef et al., 2015), with a strong, positive correlation with temperature. Additionally, candidate divisions EM3, OP1, and OctSpA1-106 are also positively correlated with sulfur content, and these OTUs were frequently found in US, SK, and DT, which are rich in sulfur (Table 2). Furthermore, phosphate is positively correlated with Caldithrix, Bacteroidetes, Cyanobacteria, Proteobacteria, WS3, and OP11, which were found in the AH and SE hot springs with slightly higher phosphate contents. Proteobacteria, Cyanobacteria, and Bacteroidetes were also positively correlated with TOC and TN, while Chlorobi, candidate division NKB19, and Spirochaetes were positively correlated with C:N. The microbial community from SS was negatively correlated with most of the measured physicochemical variables.

## DISCUSSION

### Microbial Community Composition and Diversity

All reported Malaysian hot springs have circumneutral pH. In the present study, the distributions of microbiota in six hot springs with different physical and physicochemical characteristics were studied. The five non-saline hot springs can be differentiated based on their water temperature, i.e., moderate-temperature (SS and SE) and high-temperature (US, SK, and DT) thermal springs. The moderate-temperature hot spring AH is especially interesting, as it is located 2 km away from an open sea, on an island. The concentrations of sodium, magnesium, chloride, fluoride, and sulfate ions in the AH hot spring were nevertheless lower than those in a seawater sample collected from Malaysia (data not shown). Therefore, it is possible that the water in the AH pond is a mixture of sea- and groundwater. The compounds in seawater (including chloride, magnesium, sodium, and sulfate ions) clearly influenced the microbial community that distinguishes AH from the non-saline US, SK, DT, SS, and SE hot springs. The microbial community structures in the US and SK hot springs are quite similar; both sites cluster together in the biplots in Figures 5A,B and share similar OTUs such as *Thermus*, *Hydrogenobacter*, and *Caloramator*.

Generally, the dominant phyla in the studied hot springs are similar, but the sites differ with respect to the overall composition (Figures 3A, 4A). Firmicutes and Proteobacteria are the phyla consistently present in circumneutral hot springs. Site-specific taxa assigned at the genus level included *Fervidobacterium* in US; *Coprococcus* and *Pelosinus* in SK; *Vogesella* in DT; *Candidatus Rhabdochlamydia*, *Sediminibacterium*, and *Hydrogenophaga* in SS; *Rheinheimera* and *Flavobacterium* in SE; and *Halomicronema*, *Halomonas*, *Fusibacter*, and *Cyanobacterium* in AH (Figure 5A). Interestingly, *Candidatus Rhabdochlamydia* (Kostanjšek et al., 2004), an intracellular bacterium that was first found in the terrestrial isopod *Porcellio scaber* appeared to be one of the

dominant genera in SE. This finding suggests that endosymbiosis of thermophilic microbiota can occur in the hot springs and is not restricted to SE, but also possibly occurs in DT, where we discovered tiny reddish crustaceans (order, Isopoda; data not shown).

In the current study, archaea appeared to be a minority in the prokaryotic community. This result is consistent with our previous shotgun metagenome analyses for SK (Chan et al., 2015) and US (data not shown). High-temperature environments were previously generally believed to be the realm of archaea (Urbietta et al., 2014; Li et al., 2015). However, recent studies applying molecular methods have revealed that bacteria rather are the predominant prokaryotic communities in such environments (Badhai et al., 2015; López-López et al., 2015). The factors that allow bacteria to dominate in high-temperature habitats are not well understood. Our findings revealed that archaea are not dominant in circumneutral hot springs, which in agreement with several recent reports with similar pH ranges (Wang et al., 2013; Merkel et al., 2017). Though insightful, the above findings are preliminary, as they are based on molecular methods, which inherently assume that both bacteria and archaea are detected at the same level. Moreover, other technical factors including DNA extraction method (Zielińska et al., 2017), primer selection (Cai et al., 2013), primer combinations, library preparation protocols, and sequencing platforms should be considered.

Higher Shannon–Wiener or Simpson diversity index values indicate greater species richness. The moderate-temperature springs SS, SE, and AH were more diverse than the higher-temperature springs at US and DT (Table 3). This shows that increments in habitat temperature result in decreased taxonomic richness and diversity, which is in agreement with earlier findings (Miller et al., 2009; Tobler and Benning, 2011; Inskeep et al., 2013; Sharp et al., 2014). Yet, markedly high microbial richness and diversity exist in SK, which may be owing to the stream-like physical appearance of this hot spring. Along the heated stream, the pH gradient varies by 1.5 unit, and the temperature fluctuates (60–110°C) because of the presence of multiple spring heads. Moreover, the stream is shallow and the water flows rapidly, thus generating sufficient aeration for aerophiles. Finally, the presence of plant litter in SK is associated with high TOC (additional carbon source) (Hou et al., 2013), which favors the growth of microorganisms. Thus, fluctuations in physicochemical features in a microenvironment likely enable a wider range of microbial species to survive.

### Physicochemical Factors Regulating Microbial Community Structure

Quantitative weighted UniFrac analysis (Figure 5A) of our samples suggested that the five non-saline hot springs had similar microbial communities, while AH represented an outlier in the dendrogram (Supplementary Figure 2A). The high salinity of AH should be the main factor responsible for the unique microbial community based on the relative abundances of OTUs. Compared to quantitative weighted UniFrac, unweighted UniFrac is a qualitative distance metric that only considers the presence or absence of OTUs. Unweighted UniFrac further confirmed that microbial membership in the saline spring (AH) largely differs from that in the non-saline hot springs.

Besides salinity, unweighted UniFrac showed clustering of the microbial community structures by temperature for the non-saline hot springs (Figures 5B and Supplementary Figure 2B). This explained that temperature is a crucial factor in identifying the changes in community membership rather than community composition in this study. Taken together, our data indicate that salinity and temperature are the main factors in shaping the microbial community structures in these six Malaysian hot springs.

Besides temperature and salinity as the most influential factors, we were interested in identifying other factors potentially affecting the microbiota in circumneutral hot springs. Based on the CCA plot, it is possible that TOC and TN, or the C:N ratio affect growth efficiencies, and thus shape the microbial communities (Michaud et al., 2014; Wan et al., 2015). AH has the highest C:N among the six Malaysian hot springs, partly because seawater has a high C:N (Meyers, 1994). DT, SS, and SE hot springs are more species-rich, and these sites have low C:N (<0.5) (Touratier et al., 1999). Even though DT, US, and SK are high-temperature springs, the dominant genera in DT are different from those in the other two springs, probably owing to low C:N ratio. As SK is located in a forest, it has a relatively high TOC, possibly related to the presence of plant litter along the stream. The decomposition of plant litter is one of the processes involved in nutrient and carbon cycling in ecosystems, and results in the release of dissolved organic matter back to earth (Kindler et al., 2011). The high quantity of organic matter in SK might explain the high microbial diversity and richness of this hot spring in comparison to other springs.

CCA showed a positive correlation between phosphate content and phosphate-solubilizing bacteria from phyla Proteobacteria, Actinobacteria, and Bacteroidetes (Sharma et al., 2013). This suggests that Cyanobacteria, Planctomycetes, Caldithrix, candidate divisions OP11, and WS3 are likely involved in the process of releasing phosphorus from insoluble compounds to the environment. Additionally, the negative correlation between Cyanobacteria and water temperature was shown for high-temperature springs (US, SK, and DT hot springs). Such observation is explained by the restriction of photosynthesis when the temperature is higher than 75°C (Ferris and Ward, 1997). In another aspect, the positive correlation between sulfur content and candidate divisions OP1, EM3, and OctSpA1-106 suggests that sulfur may affect the growth of these uncultured prokaryotes. These candidate divisions may be important for sulfur cycling in hot springs. In comparison to aforementioned variables, other abiotic factors, including chloride, fluoride, sodium, sulfate, iron, magnesium, arsenic, and aluminum play less of a role in determining the predominant microbial members in the Malaysian hot springs.

## CONCLUSION

The six Malaysian hot springs in this study have different physical and physicochemical characteristics. Investigating hot spring microbiomes with simple microbial composition is important for understanding microbe-mediated biogeochemical cycles and ecosystem functioning. This is the first study to identify

the physicochemical factors that drive variations in microbial community structures in Malaysian hot springs. Firmicutes and Proteobacteria were the signature phyla in all six hot springs that along with the presence of site-specific taxa contributed to the uniqueness of each hot spring. Temperature was found to be the most influential factor shaping the microbiome of Malaysian hot springs, as was anticipated. Generally, overall microbial diversity and richness were negatively affected by temperature. As an exception, SK hosted high microbial richness and diversity despite its high temperature, probably due to its physical characters that enable a wider range of microbial species to survive. Variables such as salinity, C:N ratio, phosphate, and sulfur content are probably secondary factors that affect circumneutral hot spring microbial communities. Nevertheless, other variables should not be ignored in microbial ecology studies, as all abiotic factors collectively contribute to the dynamics of microbial populations. Understanding microbial community dynamics and genomic variability of community members in hot springs with different ecologies is important to elucidate community functions and their importance for the maintenance of hot spring ecosystems.

## AUTHOR CONTRIBUTIONS

CC, KC, and KG contributed to the conception and design of the study. CC, RE, and KH produced data. CC, KH, MS, and KG conducted the bioinformatics and statistical analyses. CC, KC, and KG wrote and reviewed the manuscript. MU, ED, and MS helped in interpretation of data and contributed to the discussion of the results followed by reviewing the manuscript. All authors read and approved the final manuscript, and are agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## FUNDING

This work was supported by the University of Malaya via High Impact Research Grants (UM.C/625/1/HIR/MOHE/CHAN/01 [Grant No. A-000001-50001] and UM.C/625/1/HIR/MOHE/CHAN/14/1 [Grant No. H-50001-A000027]) awarded to KC. KG is grateful for funding received from Universiti Teknologi Malaysia GUP (Grant 15H50). Mohd Shahir Shamsir appreciates funding provided to this study by Universiti Teknologi Malaysia GUP (Grant 15H16).

## ACKNOWLEDGMENTS

We are grateful to Teong Han Chew for assistance on part of the Python programming work.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01252/full#supplementary-material>

## REFERENCES

- Badhai, J., Ghosh, T. S., and Das, S. K. (2015). Taxonomic and functional characteristics of microbial communities and their correlation with physicochemical properties of four geothermal springs in Odisha, India. *Front. Microbiol.* 6:1166. doi: 10.3389/fmicb.2015.01166
- Baik, K. S., Seong, C. N., Kim, E. M., Yi, H., Bae, K. S., and Chun, J. (2005). *Hahella ganghwensis* sp. nov., isolated from tidal flat sediment. *Int. J. Syst. Evol. Microbiol.* 55, 681–684. doi: 10.1099/ijls.0.63411-0
- Baioumy, H., Nawawi, M., Wagner, K., and Arifin, M. H. (2015). Geochemistry and geothermometry of non-volcanic hot springs in West Malaysia. *J. Volcanol. Geotherm. Res.* 290, 12–22. doi: 10.1016/j.jvolgeores.2014.11.014
- Baker, B. J., Comolli, L. R., Dick, G. J., Hauser, L. J., Hyatt, D., Dill, B. D., et al. (2010). Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8806–8811. doi: 10.1073/pnas.0914470107
- Baker, G. C., Gaffar, S., Cowan, D. A., and Suharto, A. R. (2001). Bacterial community analysis of Indonesian hot springs. *FEMS Microbiol. Lett.* 200, 103–109. doi: 10.1111/j.1574-6968.2001.tb10700.x
- Blank, C. E., Cady, S. L., and Pace, N. R. (2002). Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park. *Appl. Environ. Microbiol.* 68, 5123–5135. doi: 10.1128/AEM.68.10.5123-5135.2002
- Cai, L., Ye, L., Tong, A. H. Y., Lok, S., and Zhang, T. (2013). Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PLoS ONE* 8:e53649. doi: 10.1371/journal.pone.0053649
- Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., and Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26, 266–267. doi: 10.1093/bioinformatics/btp636
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Chan, C. S., Chan, K.-G., Tay, Y.-L., Chua, Y.-H., and Goh, K. M. (2015). Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front. Microbiol.* 6:177. doi: 10.3389/fmicb.2015.00177
- Chou, J.-H., Chou, Y.-J., Arun, A. B., Young, C.-C., Chen, C. A., Wang, J.-T., et al. (2009). *Vogesella lacus* sp. nov., isolated from a soft-shell turtle culture pond. *Int. J. Syst. Evol. Microbiol.* 59, 2629–2632. doi: 10.1099/ijls.0.009266-0
- Chou, Y.-J., Chou, J.-H., Lin, M.-C., Arun, A. B., Young, C.-C., and Chen, W.-M. (2008). *Vogesella perlucida* sp. nov., a non-pigmented bacterium isolated from spring water. *Int. J. Syst. Evol. Microbiol.* 58, 2677–2681. doi: 10.1099/ijls.0.65766-0
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381
- Ferris, M. J., and Ward, D. M. (1997). Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* 63, 1375–1381.
- Goh, K. M., Chua, Y. S., Rahman, R. N. Z. R. A., Chan, R., and Illias, R. M. (2011). A comparison of conventional and miniprimer PCR to elucidate bacteria diversity in Malaysia Ulu Slim hot spring using 16S rDNA clone library. *Rom. Biotechnol. Lett.* 16, 6247–6255.
- Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* 40, 237–264. doi: 10.1093/biomet/40.3-4.237
- Hammer, Ø., Harper, D. A. T., and Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 4, 1–9.
- Hirayama, H., Takai, K., Inagaki, F., Yamato, Y., Suzuki, M., Nealson, K. H., et al. (2005). Bacterial community shift along a subsurface geothermal water stream in a Japanese gold mine. *Extremophiles* 9, 169–184. doi: 10.1007/s00792-005-0433-8
- Hou, W., Wang, S., Dong, H., Jiang, H., Briggs, B. R., Peacock, J. P., et al. (2013). A comprehensive census of microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. *PLoS ONE* 8:e53350. doi: 10.1371/journal.pone.0053350
- Huang, Q., Jiang, H., Briggs, B. R., Wang, S., Hou, W., Li, G., et al. (2013). Archaeal and bacterial diversity in acidic to circumneutral hot springs in the Philippines. *FEMS Microbiol. Ecol.* 85, 452–464. doi: 10.1111/1574-6941.12134
- Inskeep, W. P., Jay, Z. J., Tringe, S. G., Herrgard, M. J., and Rusch, D. B. (2013). The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Front. Microbiol.* 4:67. doi: 10.3389/fmicb.2013.00067
- Kan, J., Clingenpeel, S., Macur, R. E., Inskeep, W. P., Lovalvo, D., Varley, J., et al. (2011). Archaea in yellowstone lake. *ISME J.* 5, 1784–1795. doi: 10.1038/ismej.2011.56
- Kanokratana, P., Chanapan, S., Pootanakit, K., and Eurwilaichitr, L. (2004). Diversity and abundance of bacteria and archaea in the Bor Khlueng hot spring in Thailand. *J. Basic Microbiol.* 44, 430–444. doi: 10.1002/jobm.200410388
- Kindler, R., Siemens, J. A. N., Kaiser, K., Walmsley, D. C., Bernhofer, C., Buchmann, N., et al. (2011). Dissolved carbon leaching from soil is a crucial component of the net ecosystem carbon balance. *Glob. Change Biol.* 17, 1167–1185. doi: 10.1111/j.1365-2486.2010.02282.x
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1. doi: 10.1093/nar/gks808
- Kostanjšek, R., Štrus, J., Drobne, D., and Avguštin, G. (2004). Candidatus *Rhabdochlamydia porcellionis*, an intracellular bacterium from the hepatopancreas of the terrestrial isopod *Porcellio scaber* (Crustacea: Isopoda). *Int. J. Syst. Evol. Microbiol.* 54, 543–549. doi: 10.1099/ijls.0.02802-0
- Kubo, K., Knittel, K., Amann, R., Fukui, M., and Matsuura, K. (2011). Sulfur-metabolizing bacterial populations in microbial mats of the Nakabusa hot spring, Japan. *Syst. Appl. Microbiol.* 34, 293–302. doi: 10.1016/j.syapm.2010.12.002
- Lee, H. K., Chun, J., Moon, E. Y., Ko, S. H., Lee, D. S., Lee, H. S., et al. (2001). *Hahella chejuensis* gen. nov., sp. nov., an extracellular-polysaccharide-producing marine bacterium. *Int. J. Syst. Evol. Microbiol.* 51, 661–666. doi: 10.1099/00207713-51-2-661
- Lee, K., Lee, H. K., and Cho, J.-C. (2008). *Hahella antarctica* sp. nov., isolated from Antarctic seawater. *Int. J. Syst. Evol. Microbiol.* 58, 353–356. doi: 10.1099/ijls.0.65389-0
- Li, H., Yang, Q., Li, J., Gao, H., Li, P., and Zhou, H. (2015). The impact of temperature on microbial diversity and AOA activity in the Tengchong Geothermal Field, China. *Sci. Rep.* 5:17056. doi: 10.1038/srep17056
- López-López, O., Knapik, K., Cerdán, M.-E., and González-Siso, M.-I. (2015). Metagenomics of an alkaline hot spring in Galicia (Spain): microbial diversity analysis and screening for novel lipolytic enzymes. *Front. Microbiol.* 6:1291. doi: 10.3389/fmicb.2015.01291
- Lozupone, C. A., Hamady, M., Kelley, S. T., and Knight, R. (2007). Quantitative and qualitative  $\beta$  diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* 73, 1576–1585. doi: 10.1128/AEM.01996-06
- Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005
- Masaki, Y., Tsutsumi, K., Hirano, S. I., and Okibe, N. (2016). Microbial community profiling of the Chinoike Jigoku (“Blood Pond Hell”) hot spring in Beppu, Japan: isolation and characterization of Fe(III)-reducing *Sulfolobus* sp. strain GA1. *Res. Microbiol.* 167, 595–603. doi: 10.1016/j.resmic.2016.04.011
- Mathur, J., Bizzoco, R. W., Ellis, D. G., Lipson, D. A., Poole, A. W., Levine, R., et al. (2007). Effects of abiotic factors on the phylogenetic diversity of bacterial communities in acidic thermal springs. *Appl. Environ. Microbiol.* 73, 2612–2623. doi: 10.1128/AEM.02567-06
- McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618. doi: 10.1038/ismej.2011.139
- Menzel, P., Gudbergdóttir, S. R., Rike, A. G., Lin, L., Zhang, Q., Contursi, P., et al. (2015). Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* 70, 411–424. doi: 10.1007/s00248-015-0576-9

- Merkel, A. Y., Pimenov, N. V., Rusanov, I. I., Slobodkin, A. I., Slobodkina, G. B., Tarnovetckii, I. Y., et al. (2017). Microbial diversity and autotrophic activity in Kamchatka hot springs. *Extremophiles* 21, 307–317. doi: 10.1007/s00792-016-0903-1
- Meyers, P. A. (1994). Preservation of elemental and isotopic source identification of sedimentary organic matter. *Chem. Geol.* 114, 289–302. doi: 10.1016/0009-2541(94)90059-0
- Michaud, L., Lo Giudice, A., Interdonato, F., Triplet, S., Ying, L., and Blancheton, J. P. (2014). C/N ratio-induced structural shift of bacterial communities inside lab-scale aquaculture biofilters. *Aqua. Eng.* 58, 77–87. doi: 10.1016/j.aquaeng.2013.11.002
- Miller, S. R., Strong, A. L., Jones, K. L., and Ungerer, M. C. (2009). Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Appl. Environ. Microbiol.* 75, 4565–4572. doi: 10.1128/AEM.02792-08
- Nakagawa, S., Takai, K., Horikoshi, K., and Sako, Y. (2004). *Aeropyrum camini* sp. nov., a strictly aerobic, hyperthermophilic archaeon from a deep-sea hydrothermal vent chimney. *Int. J. Syst. Evol. Microbiol.* 54, 329–335. doi: 10.1099/ijs.0.02826-0
- Nishiyama, M., Yamamoto, S., and Kurosawa, N. (2013). Microbial community analysis of a coastal hot spring in Kagoshima, Japan, using molecular- and culture-based approaches. *J. Microbiol.* 51, 413–422. doi: 10.1007/s12275-013-2419-z
- Nobu, M. K., Dodsworth, J. A., Murugapiran, S. K., Rinke, C., Gies, E. A., Webster, G., et al. (2016). Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. *ISME J.* 10, 273–286. doi: 10.1038/ismej.2015.97
- Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., et al. (2011). Insights into the evolution of archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* 39, 3204–3223. doi: 10.1093/nar/gkq1228
- Olsen, G. J., Woese, C. R., and Overbeek, R. (1994). The winds of (evolutionary) change: breathing new life into microbiology. *J. Bacteriol.* 176, 1–6. doi: 10.1128/jb.176.1.1-6.1994
- Sako, Y., Nomura, N., Uchida, A., Ishida, Y., Morii, H., Koga, Y., et al. (1996). *Aeropyrum pernix* gen. nov., sp. nov., a novel aerobic hyperthermophilic archaeon growing at temperatures up to 100°C. *Int. J. Syst. Bacteriol.* 46, 1070–1077. doi: 10.1099/00207713-46-4-1070
- Samsudin, A. R., Hamzah, U., Rahman, R. A., Siwar, C., Mohd, M. F., and Othman, R. (1997). Thermal springs of Malaysia and their potential development. *J. Asian Earth Sci.* 15, 275–284. doi: 10.1016/S1367-9120(97)00012-6
- Sangwan, N., Lambert, C., Sharma, A., Gupta, V., Khurana, P., Khurana, J. P., et al. (2015). Arsenic rich Himalayan hot spring metagenomics reveal genetically novel predator–prey genotypes. *Environ. Microbiol. Rep.* 7, 812–823. doi: 10.1111/1758-2229.12297
- Shanks, O. C., Newton, R. J., Kelty, C. A., Huse, S. M., Sogin, M. L., and McLellan, S. L. (2013). Comparison of the microbial community structures of untreated wastewaters from different geographic locales. *Appl. Environ. Microbiol.* 79, 2906–2913. doi: 10.1128/AEM.03448-12
- Sharma, S. B., Sayyed, R. Z., Trivedi, M. H., and Gobi, T. A. (2013). Phosphate solubilizing microbes: sustainable approach for managing phosphorus deficiency in agricultural soils. *Springerplus* 2:587. doi: 10.1186/2193-1801-2-587
- Sharp, C. E., Brady, A. L., Sharp, G. H., Grasby, S. E., Stott, M. B., and Dunfield, P. F. (2014). Humboldt's spa: microbial diversity is controlled by temperature in geothermal environments. *ISME J.* 8, 1166–1174. doi: 10.1038/ismej.2013.237
- Simpson, E. H. (1949). Measurement of diversity. *Nature* 163:688. doi: 10.1038/163688a0
- Song, Z.-Q., Wang, F.-P., Zhi, X.-Y., Chen, J.-Q., Zhou, E.-M., Liang, F., et al. (2013). Bacterial and archaeal diversities in Yunnan and Tibetan hot springs, China. *Environ. Microbiol.* 15, 1160–1175. doi: 10.1111/1462-2920.12025
- Spellerberg, I. F., and Fedor, P. J. (2003). A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' index. *Glob. Ecol. Biogeogr.* 12, 177–179. doi: 10.1046/j.1466-822X.2003.00015.x
- Sum, C. W., Irawan, S., and Fathaddin, M. T. (2010). "Hot springs in the Malay Peninsula," in *Proceedings World Geothermal Congress 2010* (Bali), 1–5.
- Takami, H., Arai, W., Takemoto, K., Uchiyama, I., and Taniguchi, T. (2015). Functional classification of uncultured "*Candidatus* Caldiarchaeum subterraneum" using the MAPLE system. *PLoS ONE* 10:e0132994. doi: 10.1371/journal.pone.0132994
- ter Braak, C. J. F., and Verdonschot, P. F. M. (1995). Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquat. Sci.* 57, 255–289. doi: 10.1007/BF00877430
- Thiel, V., Wood, J. M., Olsen, M. T., Tank, M., Klatt, C. G., Ward, D. M., et al. (2016). The dark side of the mushroom spring microbial mat: life in the shadow of chlorophototrophs. I. Microbial diversity based on 16S rRNA gene amplicons and metagenomic sequencing. *Front. Microbiol.* 7:919. doi: 10.3389/fmicb.2016.00919
- Tobler, D. J., and Benning, L. G. (2011). Bacterial diversity in five Icelandic geothermal waters: temperature and sinter growth rate effects. *Extremophiles* 15, 473–485. doi: 10.1007/s00792-011-0378-z
- Touratiere, F., Legendre, L., and Vézina, A. (1999). Model of bacterial growth influenced by substrate C:N ratio and concentration. *Aquat. Microb. Ecol.* 19, 105–118. doi: 10.3354/ame019105
- Trosvik, P., and de Muinck, E. J. (2015). Ecology of bacteria in the human gastrointestinal tract—identification of keystone and foundation taxa. *Microbiome* 3:44. doi: 10.1186/s40168-015-0107-4
- Urbietta, M. S., Toril, E. G., Giaveno, M. A., Bazán, Á. A., and Donati, E. R. (2014). Archaeal and bacterial diversity in five different hydrothermal ponds in the Copahue region in Argentina. *Syst. Appl. Microbiol.* 37, 429–441. doi: 10.1016/j.syapm.2014.05.012
- Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., and Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. *Gigascience* 2:16. doi: 10.1186/2047-217X-2-16
- Wan, X., Huang, Z., He, Z., Yu, Z., Wang, M., Davis, M. R., et al. (2015). Soil C:N ratio is the major determinant of soil microbial community structure in subtropical coniferous and broadleaf forest plantations. *Plant Soil* 387, 103–116. doi: 10.1007/s11104-014-2277-4
- Wang, S., Hou, W., Dong, H., Jiang, H., Huang, L., Wu, G., et al. (2013). Control of temperature on microbial community structure in hot springs of the Tibetan Plateau. *PLoS ONE* 8:e62901. doi: 10.1371/journal.pone.0062901
- Welte, C., and Deppenmeier, U. (2011). Membrane-bound electron transport in *Methanoseta thermophila*. *J. Bacteriol.* 193, 2868–2870. doi: 10.1128/JB.00162-11
- Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576
- Xu, Z., Hansen, M. A., Hansen, L. H., Jacquiod, S., and Sørensen, S. J. (2014). Bioinformatic approaches reveal metagenomic characterization of soil microbial community. *PLoS ONE* 9:e93445. doi: 10.1371/journal.pone.0093445
- Youssef, N. H., Couger, M. B., McCully, A. L., Criado, A. E. G., and Elshahed, M. S. (2015). Assessing the global phylum level diversity within the bacterial domain: a review. *J. Adv. Res.* 6, 269–282. doi: 10.1016/j.jare.2014.10.005
- Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620. doi: 10.1093/bioinformatics/btt593
- Zheng, B., Wang, L., and Liu, L. (2014). Bacterial community structure and its regulating factors in the intertidal sediment along the Liaodong Bay of Bohai Sea, China. *Microbiol. Res.* 169, 585–592. doi: 10.1016/j.micres.2013.09.019
- Zielińska, S., Radkowski, P., Blendowska, A., Ludwig-Gałęzowska, A., Loś, J. M., and Loś, M. (2017). The choice of the DNA extraction method may influence the outcome of the soil microbial community structure analysis. *Microbiologypopen*. doi: 10.1002/mbo3.453. [Epub ahead of print].

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

Copyright © 2017 Chan, Chan, Ee, Hong, Urbietta, Donati, Shamsir and Goh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.