



ARTICLE

DOI: 10.1038/s41467-018-05020-y

OPEN

Microbial biogeography of 925 geothermal springs in New Zealand

Jean F. Power ^{1,2}, Carlo R. Carere^{1,3}, Charles K. Lee², Georgia L.J. Wakerley², David W. Evans¹, Mathew Button⁴, Duncan White⁵, Melissa D. Climo^{5,6}, Annika M. Hinze⁴, Xochitl C. Morgan⁷, Ian R. McDonald², S. Craig Cary² & Matthew B. Stott ^{1,6}

Geothermal springs are model ecosystems to investigate microbial biogeography as they represent discrete, relatively homogenous habitats, are distributed across multiple geographical scales, span broad geochemical gradients, and have reduced metazoan interactions. Here, we report the largest known consolidated study of geothermal ecosystems to determine factors that influence biogeographical patterns. We measured bacterial and archaeal community composition, 46 physicochemical parameters, and metadata from 925 geothermal springs across New Zealand (13.9–100.6 °C and pH < 1–9.7). We determined that diversity is primarily influenced by pH at temperatures <70 °C; with temperature only having a significant effect for values >70 °C. Further, community dissimilarity increases with geographic distance, with niche selection driving assembly at a localised scale. Surprisingly, two genera (*Venenivibrio* and *Acidithiobacillus*) dominated in both average relative abundance (11.2% and 11.1%, respectively) and prevalence (74.2% and 62.9%, respectively). These findings provide an unprecedented insight into ecological behaviour in geothermal springs, and a foundation to improve the characterisation of microbial biogeographical processes.

¹Geomicrobiology Research Group, Department of Geothermal Sciences, GNS Science, Taupō 3384, New Zealand. ²Thermophile Research Unit, School of Science, University of Waikato, Hamilton 3240, New Zealand. ³Department of Chemical and Process Engineering, University of Canterbury, Christchurch 8140, New Zealand. ⁴Department of Computer Science, University of Waikato, Hamilton 3240, New Zealand. ⁵Wairakei Research Centre, GNS Science, Taupō 3384, New Zealand. ⁶School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand. ⁷Department of Microbiology and Immunology, University of Otago, Dunedin 9054, New Zealand. Correspondence and requests for materials should be addressed to S.C.C. (email: caryc@waikato.ac.nz) or to M.B.S. (email: matthew.stott@canterbury.ac.nz)

Biogeography identifies patterns of diversity across defined spatial or temporal scales in an attempt to describe the factors which influence these distributions. Recent studies have shown that microbial community diversity is shaped across time and space^{1,2} via a combination of environmental selection, stochastic drift, diversification, and dispersal limitation^{3,4}. The relative impact of these ecological drivers on diversity is the subject of ongoing debate, with differential findings reported across terrestrial, marine, and human ecosystems^{5–8}.

Geothermally-heated springs are ideal systems to investigate microbial biogeography, because, in comparison to terrestrial environments, they represent discrete, aquatic habitats with broad geochemical and physical gradients distributed across proximal and distal geographic distances. The relatively constrained microbial community structures, typical of geothermal springs compared to soils and sediments, also allow for the robust identification of diversity trends. Separate studies have each implicated temperature^{9–11}, pH¹², and seasonality¹³ as the primary drivers of bacterial and archaeal communities in these ecosystems; with niche specialisation observed within both local and regional populations^{14,15}. Other geochemical variables, particularly hydrogen^{16,17} and nitrogen^{18,19}, may also contribute to community structure. Concurrently, the stochastic action of

microbial dispersal is thought to be a significant driver behind the distribution of microorganisms²⁰, with endemism and allopatric speciation reported in intercontinental geothermal springs^{21,22}. It is important to note that significant differences have been found between aqueous and soil/sediment samples from the same springs^{10,12,23}; emphasising that the increased relative homogeneity of aqueous samples make geothermal water columns excellent candidate environments for investigating large-scale taxa–geochemical associations. Despite these findings, a lack of sampling quantity/density and physicochemical gradient scales, uniformity in sampling methodology, and a concurrent assessment of geographic distance, within a single study, has hindered a holistic view of microbial biogeography from developing.

The Taupō Volcanic Zone (TVZ) is a region rich in geothermal springs and broad physicochemical gradients spanning 8000 km² in New Zealand's North Island (Fig. 1), making it a tractable model system for studying microbial biogeography. The area is a rifting arc associated with subduction at the Pacific-Australian tectonic plate boundary, resulting in a locus of intense magmatism²⁴. The variable combination of thick, permeable volcanic deposits, high heat flux, and an active extensional (crustal thinning) setting favours the deep convection of groundwater and exsolved magmatic volatiles that are expressed as

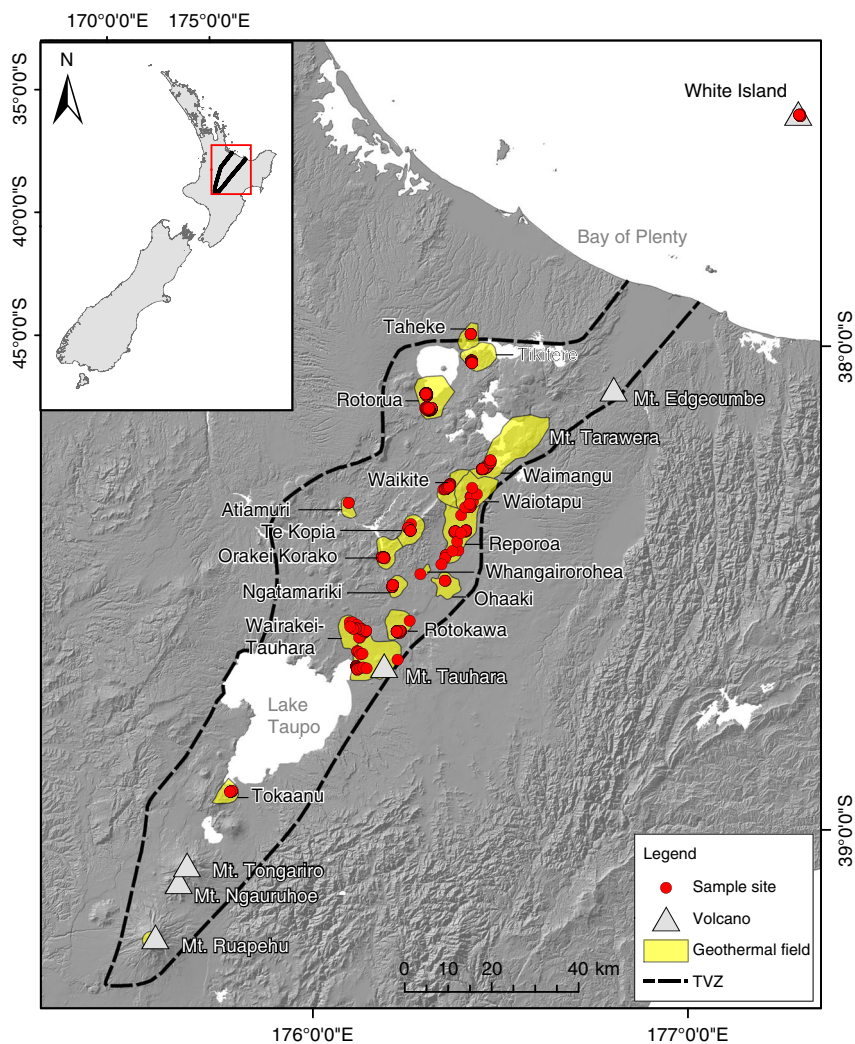


Fig. 1 Map of the Taupō Volcanic Zone (TVZ), New Zealand. The geothermal fields from which samples were collected are presented in yellow. All sampled geothermal springs ($n = 1019$) are marked by red circles. The panel insert highlights the location of the TVZ in the central north island of New Zealand. The topographic layers for this map were obtained from Land Information New Zealand (LINZ; CC-BY-4.0) and the TVZ boundary defined using data from Wilson et al.²⁴

physicochemically-heterogeneous surface features in geographically distinct geothermal fields²⁵. Previous microbiological studies across the region have hinted at novel diversity and function present within some of these features^{26–30}, however investigations into the biogeographical drivers within the TVZ are sparse and have focused predominantly on soil/sediments or individual hotspots^{9,11,31}.

Here, we report the diversity and biogeography of microbial communities found in geothermal springs, collected as part of the 1000 Springs Project. This project aimed to catalogue the microbial biodiversity and physicochemistry of New Zealand geothermal springs to serve as a conservation, scientific, and indigenous cultural knowledge repository; a publicly accessible database of all springs surveyed is available online (<http://1000Springs.org.nz>). In order to determine the influence of biogeographical processes on bacterial and archaeal diversity and community structure within geothermal springs, we collected water-columns and metadata from 1019 spring samples within the TVZ over a 21-month period using rigorously standardised methodologies. We then performed community analysis of the bacterial and archaeal population (16S rRNA gene amplicon sequencing) and quantified 46 physicochemical parameters for each sample. This work represents the largest known microbial ecology study on geothermal aquatic habitats at a regional scale and complements a parallel study on protist diversity in New Zealand geothermal springs³². Our results demonstrate both the relative influence of physicochemical parameters (e.g. pH) and the effect of geographic isolation on the assemblage of communities in these extreme ecosystems. Collectively, these findings expand our knowledge of the constraints that govern universal microbial biogeographical processes.

Results and Discussion

Geothermal feature sampling. Recent biogeography research has demonstrated that microbial diversity patterns are detectable and are influenced by both deterministic⁶ and stochastic processes⁷. A lack of consensus on the relative impact of these factors, however, has been exacerbated by the absence of data across broad physicochemical gradients, and sampling scales and density across both geographic distance and habitat type. The inherent heterogeneity of terrestrial soil microbial ecosystems³³ further confounds attempts to distinguish true taxa–geochemical associations. To provide greater resolution to the factors driving microbial biogeography processes, we collected 1019 geothermal water-column samples from across the TVZ and assessed physicochemical and microbial community composition (Fig. 1). Samples included representatives of both extreme pH (< 0–9.7) and temperature (13.9–100.6 °C) (Supplementary Fig. 1). The filtering of low-quality and temporal samples yielded a final data set of 925 individual geothermal springs for spatial-statistical analysis (details can be found in Supplementary Methods). From these 925 springs, a total of 28,381 operational taxonomic units (OTUs; 97% similarity) were generated for diversity studies.

Microbial diversity is principally driven by pH, not temperature.

Reduced microbial diversity in geothermal springs is often attributed to the extreme environmental conditions common to these areas. Temperature and pH are reported to be the predominant drivers of microbial diversity^{9,34}, but their influence relative to other parameters has not been investigated over large geographic and physicochemical scales with appropriate sample density. Our analysis of microbial richness and diversity showed significant variation spanning pH, temperature, and geographical gradients within the TVZ (richness: 49–2997 OTUs, diversity: 1.1–7.3 Shannon index; Supplementary Figs. 2 and 3). As

anticipated, average OTU richness (386 OTUs; Supplementary Fig. 4) was reduced in comparison to studies of non-geothermal temperate terrestrial³⁵ and aquatic¹ environments. Further, OTU richness was maximal at the geothermally-moderate temperature of 21.5 °C and at circumneutral pH 6.4. This is consistent with the hypothesis that polyextreme habitats prohibit the growth of most microbial taxa, a trend reported in both geothermal and non-geothermal environments alike^{5,9}. A comparison of 46 individual physicochemical parameters (Supplementary Table 1) confirmed pH as the most significant factor influencing diversity (16.4%, linear regression: $P < 0.001$; Supplementary Fig. 3), with diversity increasing from acidic to alkaline pH. Further, multiple regression analysis showed NO_3^- , turbidity (TURB), oxidation–reduction potential (ORP), dissolved oxygen (dO), NO_2^- , Si, and Cd also had meaningful contributions (Supplementary Table 2). Cumulatively, along with pH, these factors accounted for 26.6% of the observed variation in Shannon diversity. Correlation of pH with Shannon index (Pearson's coefficient: $|r| = 0.41$, $P < 0.001$) and significance testing between samples binned by pH increments (Kruskal–Wallis: $H = 179.4$, $P < 0.001$; Supplementary Fig. 1) further confirmed pH as a major driver of variation in alpha diversity. This finding is consistent with reports of pH as the primary environmental predictor of microbial diversity in several ecosystems, both in New Zealand and globally (e.g. soil^{5,36}, water^{32,37}, alpine^{38,39}).

It has been previously hypothesised that pH has significant influence on microbial community composition because changes in proton gradients will drastically alter nutrient availability, metal solubility, or organic carbon characteristics⁵. Similarly, acidic pH will also reduce the number of taxa observed due to the decreased number that can physiologically tolerate these conditions⁴⁰ compared to non-acidic habitats. Here, we demonstrate that pH had the most significant effect on diversity across all springs measured, but due to our high sampling frequency, we see this influence diminishes at temperatures > 70 °C (Fig. 2). Inversely, the effect of temperature on diversity was lessened in springs where pH was < 4 (Supplementary Fig. 5). There is some evidence that suggests thermophily predates acid tolerance^{40,41}, thus it is possible the added stress of an extreme proton gradient across cell membranes has constrained the diversification of the thermophilic chemolithoautotrophic organisms common to these areas⁴². Indeed, a recent investigation of thermoacidophily in Archaea suggests hyperacidophily (growth $< \text{pH } 3.0$) may have only arisen as little as ~ 0.8 Ga⁴⁰, thereby limiting the opportunity for microbial diversification; an observation highlighted by the paucity of these microorganisms in extremely acidic geothermal ecosystems^{11,40}. It is also important to note that salinity has previously been suggested as an important driver of microbial community diversity^{43,44}. The quantitative data in this study showed only minimal influence of salinity (proxy as conductivity (COND)) on diversity (linear regression: $R^2 = 0.001$, $P = 0.2720$; Supplementary Table 1), bearing in mind that the majority of the geothermal spring samples in this study had salinities substantially less than that of seawater.

The relationship between temperature and alpha diversity reported in this research starkly contrasts a previous intercontinental study comparing microbial community diversity in soil/sediments from 165 geothermal springs⁹, which showed that a strong relationship ($R^2 = 0.40$ – 0.44) existed. In contrast, our data across the entire suite of samples, revealed that temperature had no significant influence on observed community diversity ($R^2 = 0.002$, $P = 0.201$; Supplementary Fig. 3, Supplementary Table 1). This result increased marginally for archaeal-only diversity ($R^2 = 0.013$, $P = 0.0005$), suggesting that temperature has a more profound effect on this domain than it does on bacteria. However, the primers used in this study are known to be unfavourable

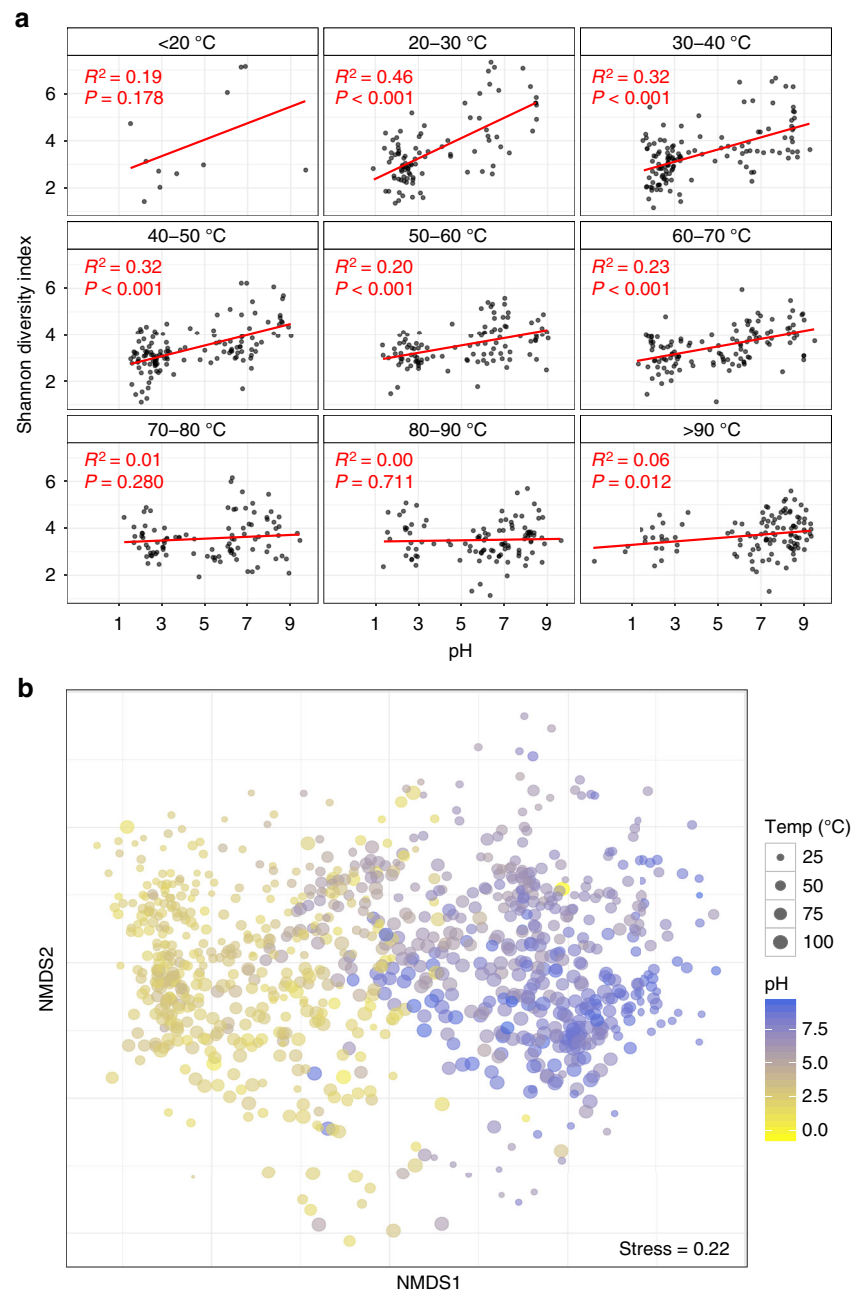


Fig. 2 Alpha and beta diversity as a function of pH and temperature. **a** pH against alpha diversity via Shannon index of all individual springs ($n = 925$) in 10 °C increments, with linear regression applied to each increment. **b** Non-metric multidimensional scaling (NMDS) plot of beta diversity (via Bray–Curtis dissimilarities) between all individual microbial community structures sampled ($n = 925$)

towards some archaeal clades⁴⁵, therefore it is likely that extensive archaeal diversity remains undetected in this study. The lack of influence of temperature on whole community diversity was further substantiated via multiple linear modelling (Supplementary Table 2), and significance and correlation testing (Kruskal–Wallis: $H = 16.2$, $P = 0.039$; Pearson's coefficient: $|r| = 0.04$, $P = 0.201$). When samples were split into pH increments, like Sharp et al.⁹, we observed increasing temperature only significantly constrained diversity above moderately acidic conditions ($\text{pH} > 4$; Supplementary Fig. 5). However, the magnitude of this effect was, in general, far less than previously reported and is likely a consequence of the sample type (e.g. soil/sediments versus aqueous) and density of samples processed¹². Many samples from geothermal environments are recalcitrant to traditional DNA extraction protocols and research in these areas

has therefore focused on those with greater biomass abundance^{9,34} (i.e. soils, sediments, streamers, or biomats). Whereas aqueous samples typically exhibit a more homogenous chemistry and community structure, the heterogeneity of terrestrial samples is known to affect microbial population (e.g. particle size, depth, nutrient composition)³³. Our deliberate use of aqueous samples extends the results of previous small-scale work^{10,31} and also permits the robust identification of genuine taxa–geochemical relationships in these environments.

Microbial communities are influenced by pH, temperature, and source fluid. Throughout the TVZ, beta diversity correlated more strongly with pH (Mantel: $\rho = 0.54$, $P < 0.001$) than with temperature (Mantel: $\rho = 0.19$, $P < 0.001$; Fig. 2, Supplementary

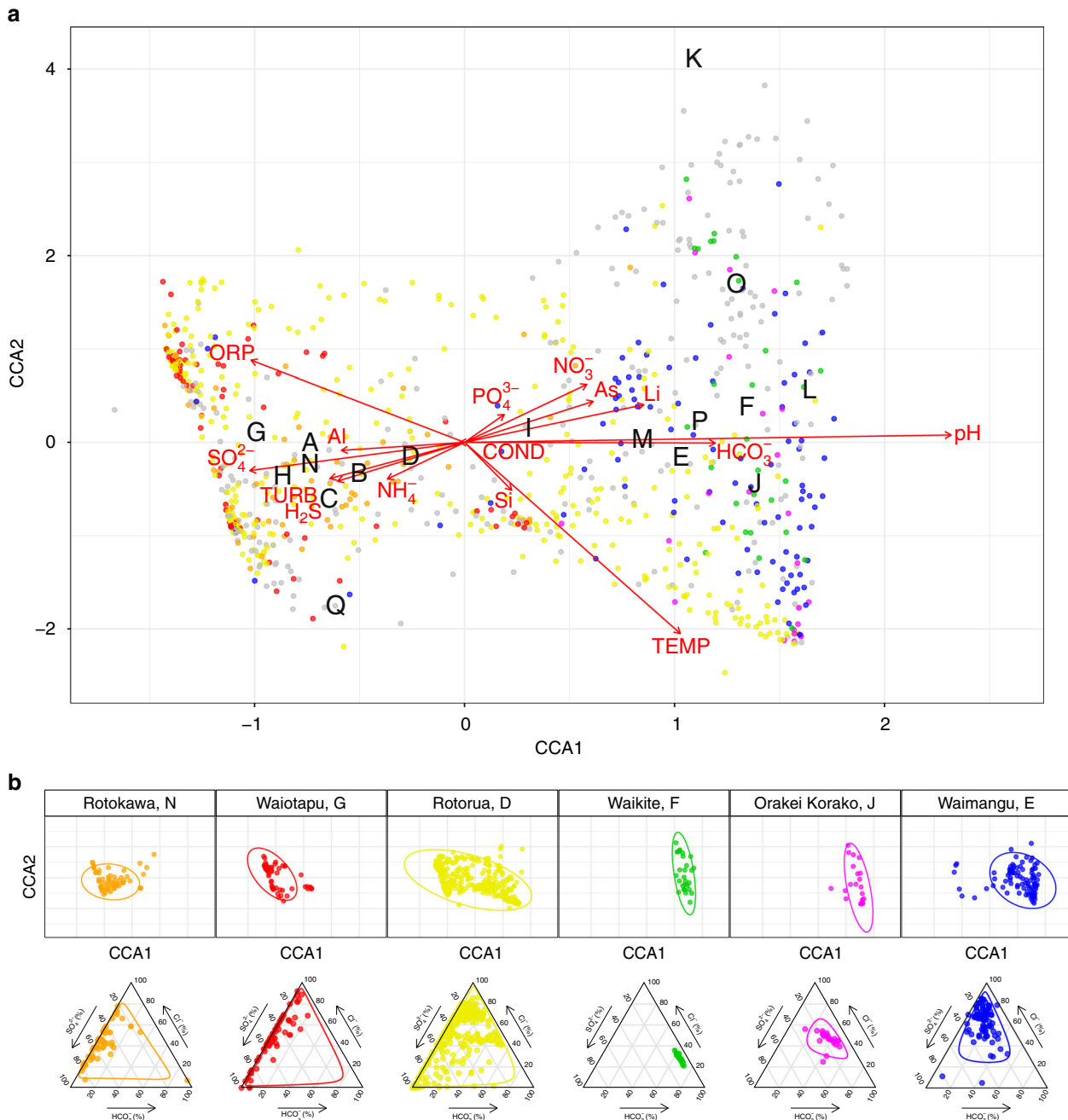


Fig. 3 Constrained correspondence analysis (CCA) of beta diversity with significant physicochemistry. **a** A scatter plot of spring community dissimilarities ($n = 923$), with letters corresponding to centroids from the model for geothermal fields (A–Q; White Island, Taheke, Tikitere, Rotorua, Waimangu, Waikite, Waitapu, Te Kopia, Reporoa, Orakei Korako, Whangairorohea, Ohaaki, Ngatamariki, Rotokawa, Wairakei-Tauhara, Tokaanu, Misc). Coloured communities are from fields represented in the subpanel. Constraining variables are plotted as arrows (COND: conductivity, TURB: turbidity), with length and direction indicating scale and area of influence each variable had on the model. **b** The top panel represents a subset of the full CCA model, with select geothermal fields shown in colour (including 95% confidence intervals). The bottom panel shows their respective geochemical signatures as a ratio of chloride (Cl^-), sulfate (SO_4^{2-}), and bicarbonate (HCO_3^-)

Table 3). This trend was consistent in pH- and temperature-binned samples (ANOSIM: $|R| = 0.46$ and 0.18 , respectively, $P < 0.001$; Supplementary Fig. 6); further confirming pH, more so than temperature, accounted for observed variations in beta diversity. Congruent with our finding that pH influences alpha diversity at lower temperatures ($< 70^\circ\text{C}$), the effect of temperature reducing beta diversity had greater significance above 80°C (Wilcox: $P < 0.001$; Supplementary Fig. 6). The extent of measured physicochemical properties across 925 individual habitats, however, allowed us to explore the environmental

impact on community structures beyond just pH and temperature. Permutational multivariate analysis of variance in spring community assemblages showed that pH (12.4%) and temperature (3.9%) had the greatest contribution towards beta diversity, followed by ORP (1.4%), SO_4^{2-} (0.8%), TURB (0.8%), and As (0.7%) ($P < 0.001$; Supplementary Table 4). Interestingly, constrained correspondence analysis of the 15 most significant, non-collinear, and variable parameters (Supplementary Tables 4 and 5; pH, temperature, TURB, ORP, SO_4^{2-} , NO_3^- , As, NH_4^+ , HCO_3^- , H_2S , COND, Li, Al, Si, and PO_4^{3-}),

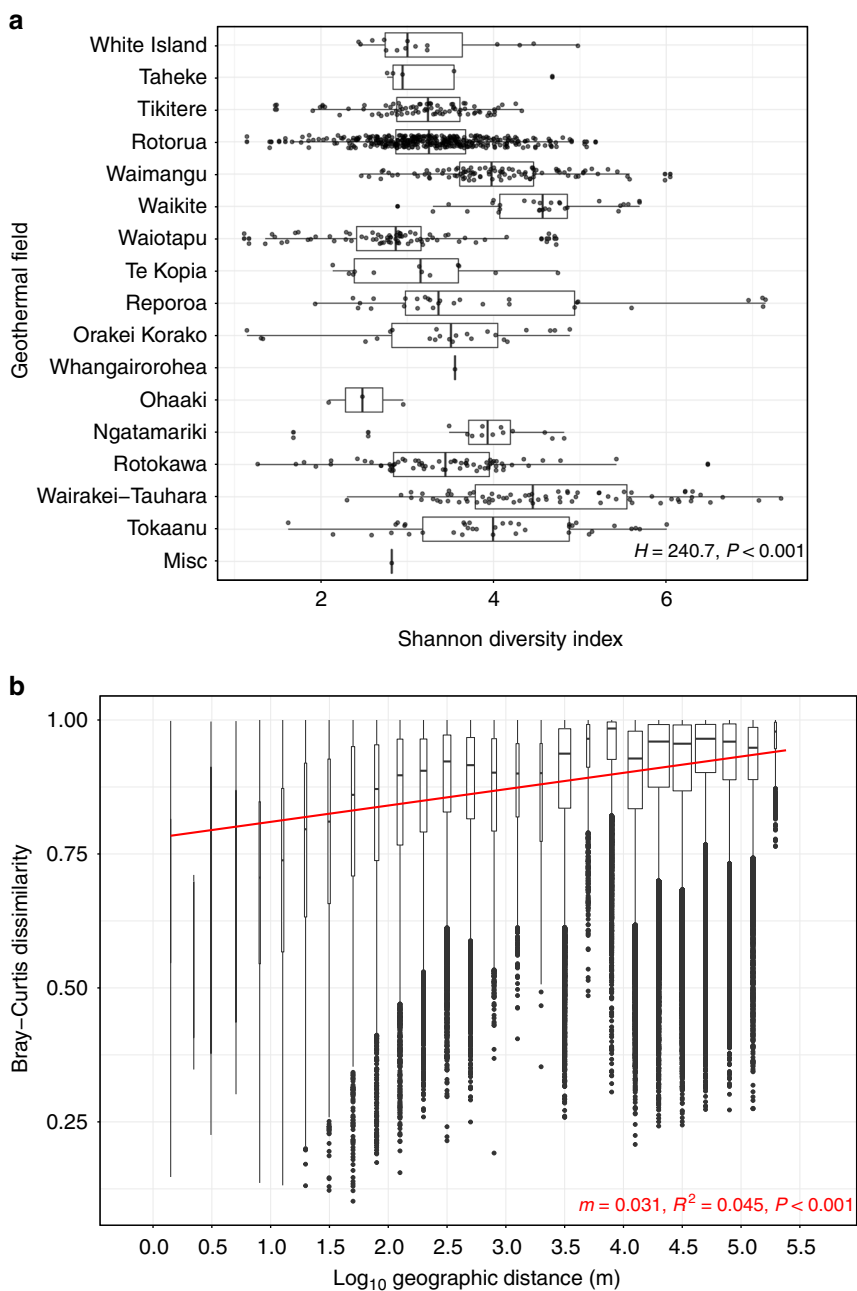


Fig. 4 Alpha and beta diversity as a function of geographic distance. **a** Alpha diversity scales (via Shannon index) across individual springs ($n = 925$), split by geothermal fields. Fields are ordered from north to south (H : Kruskal–Wallis test). **b** A distance-decay pattern of beta diversity (via Bray–Curtis dissimilarities of 925 springs) against pairwise geographic distance in metres, with linear regression applied ($m =$ slope). Geographic distance is split into bins to aid visualisation of the spread

along with geothermal field locations, only explained 10% of variation in beta diversity (Fig. 3), indicating physicochemistry, or at least the 46 parameters measured, were not the sole drivers of community composition.

Geothermal fields are known to express chemical signatures characteristic of their respective source fluids⁴⁶, implying autocorrelation could occur between location and geochemistry. We therefore investigated whether typical geochemical conditions exist for springs within the same geothermal field and whether specific microbial community assemblages could be predicted. Springs are usually classified according to these source fluids; alkaline-chloride or acid-sulfate. High-chloride features are typically sourced from magmatic waters and have little interaction with groundwater aquifers. At depth, water–rock

interactions can result in elevated bicarbonate concentrations and, consequently, neutral to alkaline pH in surface features. Acid-sulfate springs ($< \text{pH } 3$), in contrast, form as steam-heated groundwater couples with the eventual oxidation of hydrogen sulfide into sulfate (and protons). Rarely, a combination of the two processes can occur; leading to intermediate pH values⁴⁷. It is unknown, however, whether these source fluid characteristics are predictive of their associated microbial ecosystems. Bray–Curtis dissimilarities confirmed that, like alpha diversity (Kruskal–Wallis: $H = 240.7, P < 0.001$; Fig. 4), community structures were significantly different between geothermal fields (ANOSIM: $|R| = 0.26, P < 0.001$; Supplementary Fig. 7). Gradient analysis comparing significant geochemical variables and geography further identified meaningful intra-geothermal field

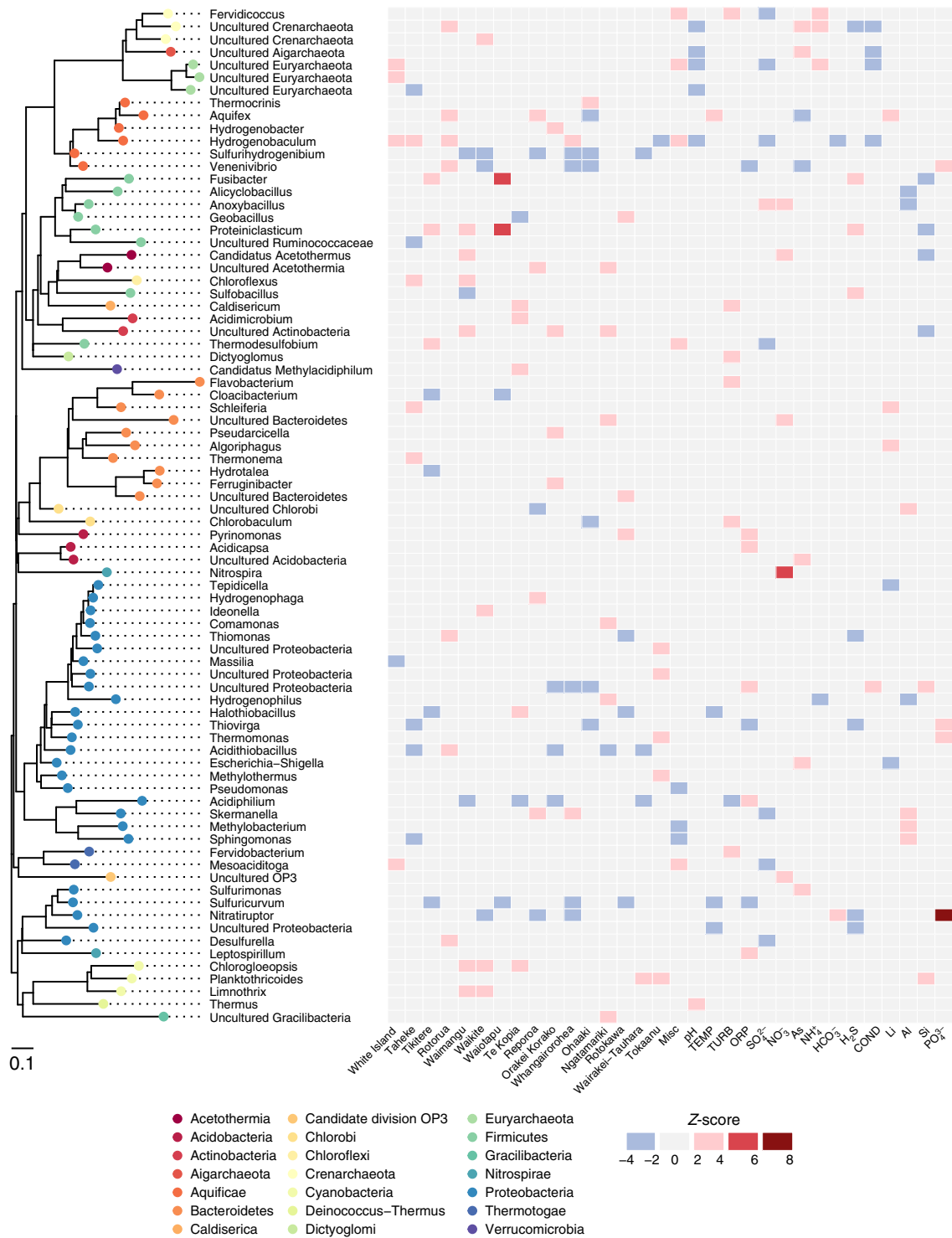


Fig. 5 Taxonomic association with location and physicochemistry. The heat map displays positive (red) and negative (blue) association of genus-level taxa (> 0.1% average relative abundance) with each geothermal field and significant environmental variables, based on Z-scores of abundance log ratios. Each taxon is colour-coded to corresponding phylum on the approximately maximum-likelihood phylogenetic tree

clustering of microbial communities (95% CI; Fig. 3 and Supplementary Fig. 8). Further, characteristic geochemical signatures from these fields were identified and analysis suggests they could be predictive of community composition. For example, the Rotokawa and Waikite geothermal fields (approx. 35 km apart) (Fig. 3 position N and F) display opposing ratios of HCO₃⁻, SO₄²⁻, and Cl⁻, with corresponding microbial

communities for these sites clustering independently in ordination space. Despite this association, intra-field variation in both alpha and beta diversity also occurred at other geothermal sites where geochemical signatures were not uniform across local springs (e.g. Rotokawa, Fig. 3 position D), demonstrating that correlation does not necessarily always occur between locational proximity and physicochemistry.

Aquificae and Proteobacteria are abundant and widespread. To determine whether individual microbial taxa favoured particular environmental conditions and locations, we first assessed the distribution of genera across all individual springs. Within 17 geothermal fields and 925 geothermal features, 21 phyla were detected with an average relative abundance > 0.1% (Fig. 5). We found that two phyla and associated genera, Proteobacteria (*Acidithiobacillus* spp.) and Aquificae (*Venenivibrio*, *Hydrogenobaculum*, *Aquifex* spp.), dominated these ecosystems (65.2% total average relative abundance across all springs), composing nine of the 15 most abundant genera > 1% average relative abundance (Table 1). Considering the broad spectrum of geothermal environmental conditions sampled in this study (we assessed microbial communities in springs across a pH gradient of nine orders of magnitude and a temperature range of ~87 °C) and the prevalence of these taxa across the region, this result was surprising. Proteobacteria was the most abundant phylum across all samples (34.2% of total average relative abundance; Table 1), found predominantly at temperatures less than 50 °C (Supplementary Fig. 9). Of the 19 most abundant proteobacterial genera (average relative abundance > 0.1%), the majority are characterised as aerobic chemolithoautotrophs, utilising either sulfur species and/or hydrogen for metabolism. Accordingly, the most abundant (11.1%) and prevalent (62.9%) proteobacterial genus identified was *Acidithiobacillus*⁴⁸, a mesophilic-moderately thermophilic, acidophilic autotroph that utilises reduced sulfur compounds, iron, and/or hydrogen as energy for growth.

Aquificae (order Aquificales) was the second most abundant phylum overall (31% average relative abundance across 925 springs) and included three of the four most abundant genera; *Venenivibrio*, *Hydrogenobaculum*, and *Aquifex* (11.2%, 10.0%, and 8.6%, respectively; Table 1). As Aquificae are thermophilic (T_{opt} 65–85 °C)⁴⁹, they were much more abundant in warmer springs (> 50 °C; Supplementary Fig. 9). The minimal growth temperature reported for characterised Aquificales species (*Sulfurihydrogenibium subterraneum* and *Sulfurihydrogenibium kristjanssonii*)⁴⁹ is 40 °C and may explain the low Aquificae abundance found in springs less than 50 °C. Terrestrial Aquificae are predominately microaerophilic chemolithoautotrophs that oxidise hydrogen or reduced sulfur compounds; heterotrophy is also observed in a few representatives⁴⁹. Of the 14 currently described genera within the Aquificae, six genera were relatively abundant in our dataset (average relative abundance > 0.1%; Fig. 5): *Aquifex*, *Hydrogenobacter*, *Hydrogenobaculum* and

Thermocrinis (family Aquificaceae), and *Sulfurihydrogenibium* and *Venenivibrio* (family Hydrogenothermaceae). No signatures of the Desulfurobacteriaceae were detected and is consistent with reports that all current representatives from this family are associated with deep-sea or coastal thermal vents⁴⁹. *Venenivibrio* (OTUs; $n = 111$) was also the most prevalent and abundant genus across all communities (Table 1). This taxon, found in 74.2% ($n = 686$) of individual springs sampled, has only one cultured representative, *Venenivibrio stagnispumantis* (CP.B2^T), which was isolated from the Waiotapu geothermal field in the TVZ²⁹. The broad distribution of this genus across such a large number of habitats was surprising, as growth of the type strain is only supported by a narrow set of conditions (pH 4.8–5.8, 45–75 °C). Considering this, and the number of *Venenivibrio* OTUs detected, we interpret this result as evidence there is substantial undiscovered phylogenetic and physiological diversity within the genus. The ubiquity of *Venenivibrio* suggests that either the metabolic capabilities of this genus extend substantially beyond those described for the type strain, and/or that many of the divergent taxa could be persisting and not growing under conditions detected in this study^{30,50}.

Geochemical and geographical associations exist at the genus level.

The two most abundant phyla, Proteobacteria and Aquificae, were found to occupy a characteristic ecological niche (< 50 and > 50 °C, respectively, Supplementary Fig. 9). To investigate specific taxa–geochemical associations beyond just temperature and pH, we applied a multivariate linear model to determine enrichment of taxa in association with geothermal fields and other environmental data (Fig. 5). The strongest associations between taxa and chemistry (Z -score > 4) were between *Nitrospira*–nitrate (NO_3^-) and *Nitratiruptor*–phosphate (PO_4^{3-}). *Nitrospira* oxidises nitrite to nitrate and therefore differential high abundance of this taxon in nitrate-rich environments is expected. Further, the positive *Nitratiruptor*– PO_4^{3-} relationship suggests phosphate is a preferred nutritional requirement for this chemolithoautotroph⁵¹ and informs future efforts to isolate members of this genus would benefit from additional phosphate or the presence of reduced phosphorous compounds in the culture medium^{52,53}. *Thermus* and *Hydrogenobaculum* were the only bacterial taxa to differentially associate (compared to other taxa) positively and negatively with pH respectively. This is consistent with the lack of acidophily phenotype (pH < 4) reported in *Thermus* spp.⁵⁴ and conversely, the preferred acidic ecological

Table 1 Average relative abundance and prevalence of phyla and genera

Phylum	Genus	Abundance	SD	Max	Prevalence
Aquificae	<i>Venenivibrio</i>	0.112	0.231	0.968	0.742
Proteobacteria	<i>Acidithiobacillus</i>	0.111	0.242	0.994	0.629
Aquificae	<i>Hydrogenobaculum</i>	0.100	0.235	0.999	0.608
Aquificae	<i>Aquifex</i>	0.086	0.212	0.971	0.497
Deinococcus-Thermus	<i>Thermus</i>	0.025	0.071	0.732	0.552
Proteobacteria	<i>Thiomonas</i>	0.024	0.101	0.941	0.396
Proteobacteria	<i>Desulfurella</i>	0.022	0.067	0.758	0.497
Crenarchaeota	Sulfolobaceae (f)	0.020	0.091	0.951	0.416
Euryarchaeota	Thermoplasmatales (o)	0.019	0.059	0.495	0.539
Proteobacteria	<i>Thiovirga</i>	0.015	0.077	0.816	0.374
Proteobacteria	Hydrogenophilaceae (f)	0.015	0.072	0.704	0.406
Thermodesulfobacteria	<i>Caldimicrobium</i>	0.015	0.052	0.651	0.519
Proteobacteria	<i>Hydrogenophilus</i>	0.013	0.045	0.432	0.484
Thermotogae	<i>Mesoaciditoga</i>	0.011	0.033	0.286	0.410
Parcubacteria	Parcubacteria (p)	0.010	0.024	0.193	0.608

Only taxa above a 1% average compositional threshold are shown. Maximum abundance of each taxon within individual features and standard deviation across all 925 springs. Where taxonomy assignment failed to classify to genus level, the closest assigned taxonomy is shown
SD = standard deviation, f = family, o = order, p = phylum

niche of *Hydrogenobaculum*⁵⁵. *Aquifex* was the only genus to display above average association with temperature, confirming abundance of this genera is significantly enhanced by hyperthermophily⁵⁶.

Similar to the chemical–taxa associations discussed above, differential abundance relationships were calculated with respect to individual geothermal fields (Fig. 5). A geothermal field, which contains springs across the pH scale (i.e. Rotorua), was closely associated with the highly abundant and prevalent *Acidithiobacillus* and *Venenivibrio*. On the other hand, a predominantly acidic geothermal system (i.e. Te Kopia), produced the only positive associations with “*Methylococcus*” (Verrucomicrobia), *Acidimicrobium* (Actinobacteria), *Terrimonas* (Bacteroidetes), and *Halothiobacillus* (Proteobacteria). These relationships are likely describing sub-community requirements that are otherwise not captured by conventional spatial–statistical analysis, therefore providing insight into previously unrecognised microbe–niche interactions.

Distance-decay patterns differ at local and regional scales.

Environmental selection, ecological drift, diversification, and dispersal limitation all contribute to distance-decay patterns⁴. While several studies have shown microbial dispersal limitations and distance-decay patterns exist in diverse geothermal and non-geothermal environments (e.g. hot springs²¹, freshwater streams¹, salt marshes), the point of inflection between dispersal limitation and selection, at regional and local geographic scales, remains under-studied. We identified a positive distance-decay trend with increasing geographic distance between 925 geothermal spring communities across the TVZ region (linear regression: $m = 0.031$, $P < 0.001$; Fig. 4). This finding strongly suggests that dispersal limitation exists between individual geothermal fields. Increasing the resolution to within individual fields, distance-decay patterns are negligible compared to the regional scale (Supplementary Table 6). Interestingly, the greatest pairwise difference ($y = 1$) between Bray–Curtis dissimilarities was also observed in springs classified as geographically-adjacent (< 1.4 m). In the 293 geothermal spring pairs separated by < 1.4 m, temperature had a greater correlation with beta diversity than pH (Spearman’s coefficient: $\rho = 0.44$ and 0.30 , respectively, $P < 0.001$). This result illustrates the stark spatial heterogeneity and selective processes that can exist within individual geothermal fields. Congruently, each OTU was detected in an average of only 13 springs (Supplementary Fig. 4). We propose that physical dispersal within geothermal fields is therefore not limiting, but the physicochemical diversity of geothermal springs acts as a barrier to the colonisation of immigrating taxa. However, even between some neighbouring springs with similar (95% CI) geochemical signatures, we did note some dissimilar communities were observed (for example, Waimangu geothermal field; Fig. 3 position E). These differing observations can be explained by either one of three ways: Firstly, the defining parameter driving community structure was not one of the 46 physicochemical variables measured in this study (e.g. dissolved organic carbon or hydrogen); secondly, through the process of dispersal, the differential viability of some extremophilic taxa restricts gene flow and contributes to population genetic drift within geothermal fields⁵⁷. We often consider “extremophilic” microorganisms living in these geothermal environments as the epitome of hardy and robust. In doing so, we overlook that their proximal surroundings (i.e. immediately outside the host spring) may not be conducive to growth and survival⁵⁸ and therefore the divergence of populations in neighbouring, chemically-homogenous spring ecosystems is plausible. Thirdly, aeolian immigration²⁰ from the non-geothermal surrounding environment could alter the perceived

composition of a community, even when immigrants are not competing to survive. Future work could be undertaken to understand individual population responses to community-wide selective pressures and the temporal nature of ecosystem functioning.

Conclusion

This study presents data on both niche and neutral drivers of microbial biogeography in 925 geothermal springs at a near-national scale. Our comprehensive data set, with sufficient sampling density and standardised methodology, is the first of its kind to enable a robust spatio-chemical statistical analysis of microbial communities at the regional level across broad physicochemical gradients. Unequivocally, pH drives diversity and community complexity structures within geothermal springs. This effect, however, was only significant at temperatures < 70 °C. We also identified specific taxa associations and finally demonstrated that geochemical signatures can be indicative of community composition. Although a distance-decay pattern across the entire geographic region indicated dispersal limitation, the finding that 293 adjacent community pairs exhibited up to 100% dissimilarity suggests niche selection drives microbial community composition at a localised scale (e.g. within geothermal fields).

This research provides a comprehensive dataset that should be used as a foundation for future studies (e.g. diversification and drift elucidation on targeted spring taxa). It complements the recently published Earth Microbiome Project⁴⁴ by expanding our knowledge of the biogeographical constraints on aquatic ecosystems using standardised quantification of broad physicochemical spectrums. There is also potential to use the two studies to compare geothermal ecosystems on a global scale. Finally, our research provides a springboard to assess the cultural, recreational, and resource development value of the microbial component of geothermal springs, both in New Zealand and globally. Many of the features included in this study occur on culturally-important and protected land for Māori, therefore this or follow-on future projects may provide an avenue for exploration of indigenous knowledge, while assisting in conservation efforts and/or development.

Methods

Field sampling and processing. Between July 2013 and April 2015, 1019 aqueous samples were collected from 974 geothermal features within 18 geothermal fields in the TVZ. A 3 L water column sample was taken (to 1 m depth where possible) from each geothermal spring, lake, stream, or the catchment pool of geysers for microbial and chemical analyses. Samples were collected either at the centre of the feature or at ~ 3 m from the edge to target well-mixed and/or more representative samples, depending on safety and size of the spring. Comprehensive physical and chemical measurements, and field observational metadata were recorded contemporaneously with a custom-built application and automatically uploaded to a database (Supplementary Table 7). All samples were filtered within 2 h of collection and stored accordingly. Total DNA was extracted using a modified CTAB method⁵⁹ with the PowerMag Microbial DNA Isolation Kit using SwiftMag technology (MoBio Laboratories, Carlsbad, CA, USA). The V4 region of the 16S rRNA gene was amplified in triplicate using universal Earth Microbiome Project⁴⁴ primers F515 (5′-GTGCCAGCMGCCGCGGTAA-3′) and R806 (5′-GGACTA CVSGGGTATCTAAT-3′), details of which can be found in Supplementary Methods. SPRIselect (Beckman Coulter, Brea, CA, USA) was used to purify DNA following amplification. Amplicon sequencing was performed using the Ion PGM System for Next-Generation Sequencing with the Ion 318v2 Chip and Ion PGM Sequencing 400 Kits (ThermoFisher Scientific, Waltham, MA, USA).

Forty-six separate physicochemical parameters were determined for each geothermal spring sample collected. Inductively coupled plasma–mass spectrometry (ICP-MS) was used to determine the concentrations of aqueous metals and non-metals (31 species; a full list is provided in Supplementary Table 7), and various UV–vis spectrometry methods were used to determine aqueous nitrogen species (NH_4^+ , NO_3^- , NO_2^-), Fe^{2+} , and PO_4^{3-} , with H_2S , HCO_3^- , and Cl^- determined via titration, and SO_4^{2-} concentration measured via ion chromatography (IC). COND, dO, ORP, pH, and TURB were determined using a Hanna Instruments (Woonsocket, RI, USA) multiparameter fieldmeter at room

temperature. Spring temperature (TEMP) was measured in situ immediately after sampling, using a Fluke 51-II thermocouple (Fluke, Everett, WA, USA).

Expanded details on sampling procedures, sample processing, and protocols for DNA extraction, DNA amplification, and chemical analyses can be found in Supplementary Methods and Supplementary Table 7.

DNA sequence processing. DNA sequences were processed through a custom pipeline utilising components of UPARSE⁶⁰ and QIIME⁶². An initial screening step was performed in mothur⁶³ to remove abnormally short (< 275 bp) and long (> 345 bp) sequences. Sequences with long homopolymers (> 6) were also removed. A total of 47,103,077 reads were quality filtered using USEARCH v7⁶¹ with a maximum expected error of 1% (fastq_maxee = 2.5) and truncated from the forward primer to 250 bp. Retained sequences (85.4% of initial reads) were dereplicated and non-unique sequences removed. Next, reads were clustered to 97% similarity and chimera checked using the cluster_otus command in USEARCH, and a de novo database was created of representative OTUs. 93.2% of the original pre-filtered sequences (truncated to 250 bp) mapped to these OTUs, and taxonomy was assigned using the Ribosomal Database Project Classifier⁶⁴ (with a minimum confidence score of 0.5) against the SILVA 16S rRNA database (123 release, July 2015)⁶⁵. The final read count was 43,202,089, with a mean of 43,905 reads per sample. Chloroplasts and mitochondrial reads were removed (1.0% and 0.5%, respectively of the final read count) and rarefaction was performed to 9500 reads per sample. As a consequence, 94 samples were then removed from the dataset (this included a set of samples collected temporally), leaving a final number of 925 individual samples (see Supplementary Methods). This QC screen also resulted in the removal of one geothermal field from the study ($n = 17$).

Statistical analyses. All statistical analyses and visualisation were performed in the R environment⁶⁶ using phyloseq⁶⁷, vegan⁶⁸, and ggplot2⁶⁹ packages. Alpha diversity was calculated using the estimate_richness function in phyloseq. A series of filtering criteria were applied to the 46 geochemical parameters measured in this study to identify metadata that significantly correlated with alpha diversity in these spring communities. First, collinear variables (Pearson correlation coefficient $|r| > 0.7$) were detected⁷⁰. The best-fit linear regression between alpha diversity (using Shannon's index) and each variable was used to pick a representative from each collinear group. This removed variables associated with the same effect in diversity. Multiple linear regression was then applied to remaining variables, before and after a stepwise Akaike information criterion (AIC) model selection was run⁷¹. Samples were also binned incrementally by pH (single pH units), temperature (10 °C increments) and geographic ranges (geothermal field) (Supplementary Fig. 1), with non-parametric Kruskal–Wallis (H) testing performed to identify any significant differences between groups. Finally, correlation of pH and temperature against Shannon diversity was calculated using Pearson's coefficient $|r|$.

Bray–Curtis dissimilarity was used for all beta diversity comparisons. For ordination visualisations, a square-root transformation was applied to OTU relative abundances prior to non-metric multidimensional scaling ($k = 2$) using the metaMDS function in the vegan package. ANOSIM ($|R|$) was used to compare beta diversity across the same pH, temperature, and geographic groups (i.e. geothermal fields) used for alpha diversity analyses, followed by pairwise Wilcoxon testing with Bonferroni correction to highlight significance between individual groups. Linear regression was applied to pairwise geographic distances against spring community dissimilarities to assess the significance of distance-decay patterns. These comparisons were similarly performed on spring communities constrained to each geothermal field. A second series of filtering criteria was applied to geochemical parameters to identify metadata that significantly correlated with beta diversity. Mantel tests were performed between beta diversity and all 46 physicochemical variables using Spearman's correlation coefficient (ρ) with 999 permutations. In decreasing order of correlation, metadata were added to a PERMANOVA analysis using the adonis function in vegan. Metadata significantly correlating with beta diversity ($P < 0.01$) was assessed for collinearity using Pearson's coefficient $|r|$ ⁷⁰. In each collinear group ($|r| > 0.7$), the variable with the highest mantel statistic was chosen as the representative. Low variant geochemical variables (SD < 0.25 ppm) were then removed to allow a tractable number of explanatory variables for subsequent modelling. Constrained correspondence analysis (using the cca function in vegan) was then applied to OTUs, geothermal field locations, and the reduced set of metadata. OTUs were first agglomerated to their respective genera (using the tax_glom function in phyloseq) and then low abundant taxa (< 0.7% of total mean taxon abundance) were removed. Typical geochemical signatures within each geothermal field were used to produce ternary diagrams of Cl^- , SO_4^{2-} and HCO_3^- ratios using the ggtern package⁷².

Finally, to detect significant associations between taxa, geochemistry, and other metadata (i.e. geothermal field observations), a multivariate linear model was applied to determine log enrichment of taxa using edgeR⁷³. To simplify the display of taxonomy in this model, we first agglomerated all OTUs to their respective genera or closest assigned taxonomy group (using the tax_glom function in phyloseq), and then only used taxa present in at least 5% of samples and > 0.1% average relative abundance. Log fold enrichments of taxa were transformed into Z-scores and retained if absolute values were > 1.96. Results were visualised using ggtree⁷⁴. A phylogenetic tree was generated in QIIME by confirming alignment of representative OTU sequences using PyNAST⁷⁵, filtering the alignment to remove

positions which were gaps in every sequence and then building an approximately maximum-likelihood tree using FastTree⁷⁶ with a midpoint root.

Code availability. All code used for statistics and figures is available through GitHub (<https://gitlab.com/morganlab/collaboration-1000Springs/1000Springs>).

Data availability. Raw sequences have been deposited into the European Nucleotide Archive (ENA) under study accession number PRJEB24353. A queryable website for the 1000 Springs Project is available at the URL: <http://1000springs.org.nz>. Other relevant data supporting the findings of the study are available in this article and its Supplementary Information files, or from the corresponding authors upon request.

Received: 12 February 2018 Accepted: 8 June 2018

Published online: 23 July 2018

References

- Lear, G. et al. Following Rapoport's rule: the geographic range and genome size of bacterial taxa decline at warmer latitudes. *Environ. Microbiol.* **19**, 3152–3162 (2017).
- Ward, C. S. et al. Annual community patterns are driven by seasonal switching between closely related marine bacteria. *ISME J.* **11**, 1412–1422 (2017).
- Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. H. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**, 497–506 (2012).
- Nemergut, D. R. et al. Patterns and processes of microbial community assembly. *Microbiol. Mol. Biol. Rev.* **77**, 342–356 (2013).
- Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75**, 5111–5120 (2009).
- Delgado-Baquerizo, M. et al. It is elemental: soil nutrient stoichiometry drives bacterial diversity. *Environ. Microbiol.* **19**, 1176–1188 (2017).
- Hernando-Morales, V., Ameneiro, J. & Teira, E. Water mass mixing shapes bacterial biogeography in a highly hydrodynamic region of the Southern Ocean. *Environ. Microbiol.* **19**, 1017–1029 (2017).
- Costello, E. K. et al. Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
- Sharp, C. E. et al. Humboldt's spa: microbial diversity is controlled by temperature in geothermal environments. *ISME J.* **8**, 1166–1174 (2014).
- Cole, J. K. et al. Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. *ISME J.* **7**, 718–729 (2013).
- Ward, L. et al. Microbial community dynamics in Inferno Crater Lake, a thermally fluctuating geothermal spring. *ISME J.* **11**, 1158–1167 (2017).
- Colman, D. R. et al. Ecological differentiation in planktonic and sediment-associated chemotrophic microbial populations in Yellowstone hot springs. *FEMS Microbiol. Ecol.* **92**, 137 (2016).
- Briggs, B. R. et al. Seasonal patterns in microbial communities inhabiting the hot springs of Tengchong, Yunnan Province, China. *Environ. Microbiol.* **16**, 1579–1591 (2014).
- Beam, J. P., Jay, Z. J., Kozubal, M. A. & Inskeep, W. P. Niche specialization of novel Thaumarchaeota to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *ISME J.* **8**, 938–951 (2014).
- Miller-Coleman, R. L. et al. Korarchaeota diversity, biogeography, and abundance in Yellowstone and Great Basin hot springs and ecological niche modeling based on machine learning. *PLoS ONE* **7**, e35964 (2012).
- Meyer-Dombard, D. R., Shock, E. L. & Amend, J. P. Archaeal and bacterial communities in geochemically diverse hot springs of Yellowstone National Park, USA. *Geobiology* **3**, 211–227 (2005).
- Spears, J. R., Walker, J. J., McCollom, T. M. & Pace, N. R. Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc. Natl. Acad. Sci. USA* **102**, 2555–2560 (2005).
- Loiaco, S. T. et al. Evidence for high-temperature in situ nifH transcription in an alkaline hot spring of Lower Geyser Basin, Yellowstone National Park. *Environ. Microbiol.* **14**, 1272–1283 (2012).
- Havig, J. R., Raymond, J., Meyer-Dombard, D. R., Zolotova, N. & Shock, E. L. Merging isotopes and community genomics in a siliceous sinter-depositing hot spring. *J. Geophys. Res. Biogeosci.* **116**, 1–16 (2011).
- Herbold, C. W., Lee, C. K., McDonald, I. R. & Cary, S. C. Evidence of global-scale aeolian dispersal and endemism in isolated geothermal microbial communities of Antarctica. *Nat. Commun.* **5**, 3875 (2014).
- Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**, 976–978 (2003).

22. Papke, R. T., Ramsing, N. B., Bateson, M. M. & Ward, D. M. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**, 650–659 (2003).
23. Hou, W. et al. A comprehensive census of microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. *PLoS ONE* **8**, e53350 (2013).
24. Wilson, C. J. N. et al. Volcanic and structural evolution of Taupo Volcanic Zone, New-Zealand: a review. *J. Volcanol. Geotherm. Res.* **68**, 1–28 (1995).
25. Chambeffort, I. & Bignall, G. Taupo Volcanic Zone geothermal systems, New Zealand: exploration, science and development. *Geothermics* **59**, 147–356 (2016).
26. Stott, M. B. et al. Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. *Environ. Microbiol.* **10**, 2030–2041 (2008).
27. Dunfield, P. F. et al. Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450**, 879–882 (2007).
28. Anders, H. et al. *Limisphaera ngatamarikiensis* gen. nov., sp. nov., a thermophilic, pink-pigmented coccus isolated from subaqueous mud of a geothermal hot spring. *Int. J. Syst. Evol. Microbiol.* **65**, 1114–1121 (2015).
29. Hetzer, A., McDonald, I. R. & Morgan, H. W. *Venenivibrio stagnispumantis* gen. nov., sp. nov., a thermophilic hydrogen-oxidizing bacterium isolated from Champagne Pool, Waiotapu, New Zealand. *Int. J. Syst. Evol. Microbiol.* **58**, 398–403 (2008).
30. Greening, C. et al. Persistence of the dominant soil phylum *Acidobacteria* by trace gas scavenging. *Proc. Natl Acad. Sci. USA* **112**, 10497–10502 (2015).
31. Childs, A. M., Mountain, B. W., O’Toole, R. & Stott, M. B. Relating microbial community and physicochemical parameters of a hot spring: Champagne Pool, Wai-o-tapu, New Zealand. *Geomicrobiol. J.* **25**, 441–453 (2008).
32. Oliverio, A. M. et al. The ecology and diversity of microbial eukaryotes in geothermal springs. *ISME J.* <https://doi.org/10.1038/s41396-018-0104-2> (2018).
33. Tecon, R. & Or, D. Biophysical processes supporting the diversity of microbial life in soil. *FEMS Microbiol. Rev.* **41**, 599–623 (2017).
34. Inskeep, W. P., Jay, Z. J., Tringe, S. G., Herrgård, M. J. & Rusch, D. B. The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone Geothermal Ecosystem. *Front. Microbiol.* **4**, 67 (2013).
35. Ramirez, K. S. et al. Biogeographic patterns in below-ground diversity in New York City’s Central Park are similar to those observed globally. *Proc. R. Soc. B, Biol. Sci.* **281**, 20141988 (2014).
36. Hermans, S. M., Buckley, H. L., Case, B. S., Curran-Courne, F. & Taylor, M. Bacteria as emerging indicators of soil condition. *Appl. Environ. Microbiol.* **83**, e02826-16 (2017).
37. Bååth, E. & Kritzberg, E. pH tolerance in freshwater bacterioplankton: trait variation of the community as measured by leucine incorporation. *Appl. Environ. Microbiol.* **81**, 7411–7419 (2015).
38. Wu, J., Anderson, B. J., Buckley, H. L., Lewis, G. & Lear, G. Aspect has a greater impact on alpine soil bacterial community structure than elevation. *FEMS Microbiol. Ecol.* **93**, 253 (2017).
39. Yashiro, E. et al. Local environmental factors drive divergent grassland soil bacterial communities in the Western Swiss Alps. *Appl. Environ. Microbiol.* **82**, 6303–6316 (2016).
40. Colman, D. R. et al. Geobiological feedbacks and the evolution of thermoacidophiles. *ISME J.* **12**, 225–236 (2018).
41. McCarthy, S. et al. Expanding the limits of thermoacidophily by adaptive evolution. *Appl. Environ. Microbiol.* **82**, 857–867 (2016).
42. Stetter, K. O. Extremophiles and their adaptation to hot environments. *FEBS Lett.* **452**, 22–25 (1999).
43. Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc. Natl Acad. Sci. USA* **104**, 11436–11440 (2007).
44. Thompson, L. R. et al. A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
45. Walters, W. et al. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* **1**, e0009–e0015 (2015).
46. Reysenbach, A.-L. & Shock, E. L. Merging genomes with geochemistry in hydrothermal ecosystems. *Science* **296**, 1077–1082 (2002).
47. Jiang, X. & Takacs-Vesbach, C. D. Microbial community analysis of pH 4 thermal springs in Yellowstone National Park. *Extremophiles* **21**, 135–152 (2017).
48. Kelly, D. P. & Wood, A. P. Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen. nov., *Halothiobacillus* gen. nov. and *Thermithiobacillus* gen. nov. *Int. J. Syst. Evol. Microbiol.* **50**, 511–516 (2000).
49. Gupta, R. S. The Phylum Aquificae. In *The Prokaryotes*. 417–445 (2014).
50. Ji, M. et al. Atmospheric trace gases support primary production in Antarctic desert ecosystems. *Nature* **552**, 400–403 (2017).
51. Nakagawa, S., Takai, K., Inagaki, F., Horikoshi, K. & Sako, Y. *Nitratiruptor tergaricus* gen. nov., sp. nov. and *Nitratifactor salsuginis* gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the Epsilon-Proteobacteria isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough. *Int. J. Syst. Evol. Microbiol.* **55**, 925–933 (2005).
52. Schink, B. & Friedrich, M. Bacterial metabolism: phosphite oxidation by sulphate reduction. *Nature* **406**, 37 (2000).
53. White, A. K. & Metcalf, W. W. Microbial metabolism of reduced phosphorus compounds. *Annu. Rev. Microbiol.* **61**, 379–400 (2007).
54. Yu, T. T. et al. *Thermus amyloliquefaciens* sp. nov., isolated from a hot spring sediment sample. *Int. J. Syst. Evol. Microbiol.* **65**, 2491–2495 (2015).
55. Shima, S. & Suzuki, K.-I. *Hydrogenobacter acidophilus* sp. nov., a thermoacidophilic, aerobic, hydrogen-oxidizing bacterium requiring elemental sulfur for growth. *Int. J. Syst. Evol. Microbiol.* **43**, 703–708 (1993).
56. Huber, R. et al. *Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Syst. Appl. Microbiol.* **15**, 340–351 (1992).
57. Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D. & Horner-Devine, M. C. Drivers of bacterial β -diversity depend on spatial scale. *Proc. Natl Acad. Sci. USA* **108**, 7850–7854 (2011).
58. Campbell, K. M. et al. *Sulfolobus islandicus* meta-populations in Yellowstone National Park hot springs. *Environ. Microbiol.* **19**, 2334–2347 (2017).
59. Hjort, K. & Bernander, R. Changes in cell size and DNA content in *Sulfolobus* cultures during dilution and temperature shift experiments. *J. Bacteriol.* **181**, 5669–5675 (1999).
60. Archer, S. D. J., McDonald, I. R., Herbold, C. W. & Cary, S. C. Characterisation of bacterioplankton communities in the meltwater ponds of Bratina Island, Victoria Land, Antarctica. *FEMS Microbiol. Ecol.* **89**, 451–464 (2014).
61. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
62. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
63. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
64. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
65. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).
66. R Development Core Team. R: A Language and environment for statistical computing. (2017).
67. McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of Microbiome Census data. *PLoS ONE* **8**, e61217 (2013).
68. Oksanen, J. et al. vegan: Community ecology package. R package version 2.4-3. (2017).
69. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, New York, 2009).
70. Dormann, C. F. et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).
71. Aho, K., Derryberry, D. & Peterson, T. Model selection for ecologists: the worldview of AIC and BIC. *Ecology* **95**, 631–636 (2014).
72. Hamilton, N. ggtern: An extension to ‘ggplot2’, for the creation of ternary diagrams. R package version 2.1.5. (2016).
73. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
74. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
75. Caporaso, J. G. et al. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267 (2010).
76. Price, M. N., Dehal, P. S. & Arkin, A. P. Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).

Acknowledgements

The authors wish to acknowledge all landowners and Māori collaborators for access and support of this research. *Mana whenua* (customary rights) is acknowledged for all data generated arising from geothermal features within *rohe* of *iwi* (tribal territories). The primary collection and processing of samples was funded by an MBIE Smart Idea grant (C05X1203—Microbial Bioinventory of Geothermal Ecosystems) awarded to M.B.S. and S.C.C., colloquially known as the 1000 Springs Project (<http://1000springs.org.nz>). J.F.P. was also supported by a GNS Science Postgraduate Scholarship (under the Geothermal Resources of New Zealand Research Programme), and the University of Waikato Hilary Jolly Memorial Scholarship. We thank Karen Houghton and Hanna-Annette Peach for assistance with field work, Kevin Lee for advice on bioinformatics, and Jayadev Payyakkal

Viswam for help with DNA extractions. We also thank Isabelle Chambefort, Ed Mroczek, and Nellie Olsen for valuable comments.

Author contributions

M.B.S., S.C.C., J.F.P., I.R.M., and M.D.C. designed the study. J.F.P., D.W.E., M.B.S., C.R.C., and G.L.J.W. undertook field work and processing of samples. M.B., D.W., M.B.S., J.F.P., and A.M.H. designed the field application, database, and website. G.L.J.W. performed DNA extractions and sequencing. J.F.P. and C.K.L. processed DNA sequences. J.F.P. and X.C.M. performed data analysis and statistics. J.F.P., C.R.C., and M.B.S. wrote the manuscript, with assistance from S.C.C., X.C.M., C.K.L., I.R.M., and G.L.J.W.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-05020-y>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018