Microbial biogeography of 1,000 geothermal springs:

spatial, temporal, and allopatric dynamics of extremophiles in the

Taupō Volcanic Zone, Aotearoa-New Zealand

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Biological Sciences

at

The University of Waikato

by

JEAN FLORENCE POWER

THE UNIVERSITY OF

WAIKATO

Te Whare Wānanga o Waikato

2023

ABSTRACT

Geothermal springs are model ecosystems to investigate microbial biogeography as they represent discrete, relatively homogeneous habitats, are distributed across multiple geographical scales, and span broad geochemical gradients. This thesis reports the largest consolidated study of geothermal ecosystems to determine causal factors that influence both spatial and temporal biogeographical patterns of extremophiles. Bacterial and archaeal community composition, 46 physicochemical parameters, and metadata from 925 geothermal spring water columns across the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand were measured. Standardised methodologies were employed to define microbial diversity via 16S rRNA gene amplicon sequencing, and associated physicochemistry, ensuring confidence in sample comparison. Over an 8,000 km^2 spatial scale, multivariate statistical analyses determined diversity was primarily influenced by pH at temperatures <70 °C; with temperature only having a significant effect at >70 °C. Further, community dissimilarity increased with increasing geographic distance across the region, highlighting niche selection driving assembly at a local scale.

This research also describes the first comprehensive temporal study of geothermal microbial communities in Aotearoa-New Zealand. One hundred and fifteen water column samples from 31 geothermal features were taken over a 34-month period to ascertain microbial community stability, community response to both natural and anthropogenic disturbances in the local environment, and temporal variation in spring diversity across the pH range found in the TVZ. Results indicated temperature and associated groundwater physicochemistry were the most likely parameters to vary stochastically in these geothermal features, with community abundances rather than composition more readily affected by a changing environment. However, variation in pH (pH ±1) had a more significant effect on community structure than temperature (±20 °C), with alpha diversity failing to be an adequate measure of temporal microbial disparity in geothermal features outside of circumneutral conditions. While a substantial physicochemical disturbance was required to shift community structures at the phylum level, geothermal ecosystems were resilient at this broad taxonomic rank and returned to a pre-disturbed state if environmental conditions re-established.

The discovery that genus *Venenivibrio* (phylum Aquificota) exhibited the greatest average abundance (11.2 %) and distribution (74.2 %) of all taxa across 925 geothermal ecosystems

in the TVZ, and an apparent absence of this taxon in global geothermal systems, led to the hypothesis that allopatric speciation enabled the evolution of an endemic bacterial genus to Aotearoa-New Zealand. Maximal read abundance of *Venenivibrio* occurred in geothermal features with pH 4-6, 50-70 °C, and low oxidation-reduction potentials, highlighting a specific environmental niche that could enhance habitat isolation. Genomic analysis of the only characterised species for the genus, *Venenivibrio stagnispumantis* CP.B2[T], confirmed a chemolithoautotrophic metabolism dependent on hydrogen oxidation. While similarity between *Venenivibrio* populations illustrated dispersal was not limited across the TVZ, extensive amplicon, metagenomic, and phylogenomic analyses of local and global microbial communities from DNA sequence databases indicated *Venenivibrio* is geographically restricted to the Aotearoa-New Zealand archipelago. It was concluded that combined geographical and physicochemical constraints prevent this taxon from distributing on a broader scale, resulting in the establishment of an endemic bacterial genus.

Collectively, the findings from this thesis provide an insight into ecological behaviour in geothermal springs, highlighting the diverse controls between different taxa within the same habitat-type, and expand understanding of microbial biogeography in extreme ecosystems.

THESIS PREFACE

The research outlined in this thesis was made possible by the collection of microbial, physicochemical, and locational data across 1,088 samples from 974 geothermal features in the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand, as part of a Smart Ideas research grant (C05X1203), colloquially called the 1,000 Springs Project. The project's goal was to produce a '*Microbial Bioinventory of Geothermal Ecosystems*', and detailed information on each geothermal feature can be found at https://1000springs.org.nz/. The sampling of data from this number of geothermal features was globally unprecedented and involved months of logistical organisation. Robust methodologies ensured all samples had identical collection, processing, and archiving, which took almost three years to complete. This permitted use of this unique dataset to confidently investigate if extremophilic microorganisms in the TVZ displayed patterns of biogeography, and what underlying processes were driving the distribution of microbial diversity across the region. To this end, spatial patterns of microbial diversity were examined from 925 geothermal springs, with associated physicochemical and geographical attributes enabling niche identification. Temporal diversity patterns were scrutinised in 31 features over a 34-month period. The most abundant taxon found, genus *Venenivibrio*, was examined to understand causal mechanisms behind the distribution of this unique population in the TVZ. Overall, three data chapters in this thesis elucidate the novel biodiversity that exists in geothermal habitats in Aotearoa-New Zealand, and crucially identify the native characteristics that locally select for these indigenous microorganisms.

To briefly outline the thesis structure, Chapter 1 introduces microbiology, the concept of microbial biogeography, and microbial ecology in geothermal ecosystems through a thorough review of published literature. Research aims for the thesis are outlined, along with rationale for hypotheses and how investigation of these was planned. Chapter 2 is the first of the data chapters and focuses on spatial patterns of microbial biogeography in TVZ geothermal springs. Chapter 3 continues with temporal dynamics of geothermal microbial communities, followed by an emphasis on the most abundant taxon in these extreme ecosystems and its apparent endemism to Aotearoa-New Zealand in Chapter 4. The thesis is concluded by a synthesis of major findings and future work that can continue this novel research in Chapter 5. References for the entire thesis are consolidated into Appendix A. Finally, as the three data chapters were prepared as manuscripts for peer-reviewed publication, extensive supplementary material is included in Appendix B-F.

PUBLICATIONS RESULTING FROM THIS THESIS

- Chapter 2 | Published in *Nature Communications*, 2018

Jean F. Power, Carlo R. Carere, Charles K. Lee, Georgia L.J. Wakerley, David W. Evans, Mathew Button, Duncan White, Melissa D. Climo, Annika M. Hinze, Xochitl C. Morgan, Ian R. McDonald, S. Craig Cary, & Matthew B. Stott. Microbial biogeography of 925 geothermal springs in New Zealand. *Nat. Commun.* **9**, 2876 (2018).

- Chapter 3 | Published in *Frontiers in Microbiology*, 2023

Jean F. Power, Caitlin L. Lowe, Carlo R. Carere, Ian R. McDonald, S. Craig Cary, & Matthew B. Stott. Temporal dynamics of geothermal ecosystems in Aotearoa-New Zealand. *Front. Microbiol.* **14**, 1094311 (2023).

- Chapter 4 | Under review in *Nature Communications*, 2023

Jean F. Power, Carlo R. Carere, Holly E. Welford, Daniel T. Hudson, Kevin C. Lee, John W. Moreau, Thijs J. G. Ettema, Anna-Louise Reysenbach, Charles K. Lee, Daniel R. Colman, Eric S. Boyd, Xochitl C. Morgan, Ian R. McDonald, S. Craig Cary, & Matthew B. Stott. Allopatric speciation in the bacterial phylum Aquificota enables genus-level endemism in Aotearoa-New Zealand. *Nat. Commum.* In Review (2023).

- Appendix E | Published in *Microbiology Resource Announcements*, 2023

Jean F. Power, Holly E. Welford, Carlo R. Carere, McDonald, Ian R. McDonald, S. Craig Cary, & Matthew B. Stott. Draft genome sequence of *Venenivibrio stagnispumantis* CP.B2^T, isolated from Champagne Pool, Waiotapu, Aotearoa-New Zealand. *Microbiol. Resour. Announc.* **12**, e01074-22 (2023).

ACKNOWLEDGEMENT OF MANA WHENUA

All landowners and Māori collaborators in the Taupō Volcanic Zone, Aotearoa-New Zealand, are acknowledged as having *mana whenua* (customary rights) for data generated from ecosystems within *rohe* of *iwi* (tribal territories).

In particular, Ngāti Tahu-Ngāti Whaoa are recognised as having mana whenua of the novel bacterium *Venenivibrio stagnispumantis* type strain CP.B2, and all associated data (genomic, environmental, and phenotypic) linked with this strain, as it was isolated from Champagne Pool in the Waiotapu geothermal field, Taupō Volcanic Zone.

Sincere gratitude is expressed to all landowners and Māori collaborators for access to geothermal features and permission to collect samples, and for their continued support and enthusiasm for this, and related, research projects.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES & TABLES

LIST OF ABBREVIATIONS

16S rRNA	16 Svedberg ribosomal ribonucleic acid
ABC	ATP-binding cassette
AIC	Akaike information criterion
ANI	Average nucleotide identity
ANOSIM	Analysis of similarities
ASV	Amplicon sequence variant
ATP	Adenosine triphosphate
BAM	Binary sequence alignment map
BED	Browser extensible data
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
CCA	Constrained correspondence analysis
COG	Clusters of orthologous genes
COND	Conductivity
CP	Champagne Pool
CP.B2^T	Type strain for *Venenivibrio stagnispumantis*
dNTP	Deoxynucleoside triphosphate
dO	Dissolved oxygen
dsDNA	Double-stranded DNA
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
EC	Enzyme Commission
EDTA	Ethylenediaminetetraacetic acid
EMP	Earth Microbiome Project
EMP	Embden-Meyerhof-Parnas glycolytic pathway
ENA	European Nucleotide Archive
FDR	False discovery rate
GenBank	National Genetic Sequence Database
GOLD	Genomes Online Database
GTDB	Genome Taxonomy Database
HydDB	Hydrogenase Database
ICP-MS	Inductively coupled plasma-mass spectrometry
IMG/M	Integrated Microbial Genomes & Microbiomes
IMNGS	Integrated Microbial Next-Generation Sequencing
INSDC	International Nucleotide Sequence Database Collaboration
IQR	Interquartile range
JGI	Joint Genome Institute
KEGG/K	Kyoto Encyclopedia of Genes and Genomes
LINZ	Toitū Te Whenua	Land Information New Zealand
MAG	Metagenome-assembled genome
MBIE	Ministry of Business, Innovation, & Employment, Aotearoa-New Zealand Government
MG-RAST	Metagenomic Rapid Annotations using Subsystems Technology

mya	Million years ago
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing
NMDS	Non-metric multidimensional scaling
NZ	Aotearoa-New Zealand
OP	Obsidian Pool
ORP	Oxidation reduction potential
OTU	Operational taxonomic unit
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PES	Polyethersulfone
Pfam	Protein families
PP	Polypropylene
ppm	Parts per million
p-value	Probability value
QIIME	Quantitative Insights into Microbial Ecology
R^2	Coefficient of determination for linear regression
RCF	Relative centrifugal force
RDP	Ribosomal Database Project
RED	Relative evolutionary divergence
RefSeq	Reference Sequence Database
RPM	Revolutions per minute
rTCA	Reductive tricarboxylic acid
SAG	Single amplified genome
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SOX	Sulfur oxidation
SRA	Sequence Read Archive
SSU	Small subunit
TE	Tris-EDTA
Tris	Tris(hydroxymethyl)aminomethane
TRU	Thermophile Research Unit
TVZ	Taupō Volcanic Zone, Aotearoa-New Zealand
UV-vis	Ultraviolet-visible
WGS84	World Geodetic System 1984
YNP	Yellowstone National Park, USA
ρ	Greek small letter rho for correlation coefficients

CHAPTER 1

LITERATURE REVIEW

1.1 DIVERSITY AND ECOLOGY OF MICROORGANISMS

Microorganisms play a direct role in almost all geochemical pathways on the planet. They provide the building blocks for nearly every ecosystem and help maintain the health and prosperity of humans and animals alike. Indeed, it has been described how a world without microorganisms would result in a catastrophic failure of all geochemical cycles, food production, and waste treatment which would lead to starvation, disease, social, and economic unrest and ultimately, the end of the world as we know it[1]. Yet so much remains to be understood on the dynamics of these organisms. Their diversity is large, their function vital, and their distribution across many environments is widespread[2–4]. The fields of microbial diversity and ecology attempt to define these disciplines; therefore, it is imperative that we understand the driving forces behind them.

Microbial diversity studies focus on either a single group of microorganisms (i.e., population) or every member present in a sample ecosystem (i.e., community). The distinction between populations and communities is important, as it has been suggested individual populations are governed by different influences[5]. Examples of population studies include: *Sulfolobus*[6,7], Cyanobacteria[8], and *Acidithiobacillus*[9] lineages in hot springs; Nitrososphaerota (formerly Thaumarchaeota)[10], sulfate-reducing bacteria[11], and heterotrophic protists[12] in marine environments; *Curtobacterium* in leaf litter[13]; and 'Candidatus Udaeobacter copiosus' in soil[14]. While targeted population studies can infer activity and function of individual taxa, community dynamics are often required to elucidate ecosystem functioning and advances in high-throughput DNA sequencing technologies have allowed both increased sample size and geographical scale in comparing microbial community structures. Examples of large-scale microbial community studies include: spatial and seasonal gradients in a coastal marine environment using 596 samples[15]; bacterioplankton in the marine water column over a 10 year timeframe[16]; temporal dynamics in a treated wastewater habitat over five years[17]; bacterial connectivity between over 500 soil, sediment, and stream water samples from within a forest catchment area[18]; global diversity of soil fungi in nearly 15,000 samples from 365

individual sites[19]; and a massive 27,751 microbial communities from 21 biomes across the globe[20]. Both population- and community-level studies complement each other and are necessary to disentangle complex ecosystem behaviour in any microbial habitat.

When measuring microbial diversity, there are several attributes that need to be considered. There is a lack of consensus on a universal method for defining a microbial species[21,22], as traditional taxonomic classification based on phenotype is difficult to apply to microorganisms[21]. Individual taxa identified through environmental high-throughput gene surveys are often defined as operational taxonomic units (OTUs), with some reports suggesting this resolution is too coarse and important lineages go undetected[23]. Others propose alternate characterisations, such as ecotypes based on functionality[24,25], microbial diversity units[26], or the more recent amplicon sequence variants (ASVs)[27]. Whatever definition is used, it is imperative that the same unit of measurement is used to assess the relative abundance of taxa when comparing microbial communities across multiple ecosystems. The standardisation of sample collection and processing from environmental sites is also necessary to ensure robust comparison between samples and studies, and allow repeatability if required[28,29], with an increased sampling effort essential to incorporate both dominant and rare taxa in any microbial diversity study[30].

Microbial ecology takes the diversity of microorganisms and broadens the picture by looking at environmental and inter-species interactions. This gives a more holistic understanding of ecosystem functionality, while also providing insight into the evolutionary path of microbial taxa[31]. Many studies have linked ecology to evolution and report that evolutionary changes arise in response to local selection pressures and population dynamics[32]. This also applies to macroorganisms[33] but can be more meticulously investigated in microorganisms due to their size, shorter generation times, and more ubiquitous nature[7,34]. Measuring diversity is the first step in describing microbial community dynamics but understanding both taxa distributions and interactions with the local physicochemical environment is pertinent to expanding our knowledge on microbial evolutionary processes.

1.2   BIOGEOGRAPHY

Biogeography is a discipline of ecology which identifies patterns of diversity across a defined scale and attempts to describe the underlying processes which govern these distributions. Biogeography of macroorganisms has been a fundamental aspect of ecology studies since Darwinian times. Whittaker first penned the concept of comparative community diversity[35], while MacArthur discussed patterns of species and attempted to understand them[36]. Indeed, he and Wilson authored a book on the theory of island biogeography which has become a staple in any ecologist's book collection, describing the balance between immigration and extinction in island populations[37]. There have been many models suggested to explain biogeographical patterns but no consensus has been reached[38,39]. In 1994, Palmer reviewed 120 published theories and discussed the importance of moving towards a unification of hypotheses[40]. Common diversity patterns observed include relationships with area, time, latitude and elevation[41–44]. However, understanding the processes behind these observed patterns is important to elucidate the driving mechanisms behind the diversification of life[45]. A review in 2010 classified all processes into four distinctions (i.e., selection, dispersal, speciation and drift), which encompass all theoretical and conceptual models in community ecology[46]. Processes of biogeography have also been defined as either deterministic or stochastic[47], and/or neutral or niche[48], with Hubbell's neutral theory stating stochastic processes (e.g., immigration and extinction) are responsible for controlling species abundance in an ecosystem[49]. Rather than reinventing the wheel, microbiologists looked to these established models to investigate if similar processes affected microbial behaviour.

Empirical evidence of research involving microbial biogeography began in 1913[50,51], which included the popular Baas-Becking theory that '*everything is everywhere; but, the environment selects*'. In recent years, microbial ecologists have questioned this theory and investigated if the same biogeographical patterns and processes of animals and plants can be applied to microorganisms[52–54]. Studies have reported alternate patterns for microbial biogeography[55,56], suggesting attributes of microorganisms (e.g., size, metabolism, generation time, evolutionary scale) result in different controls for diversity, abundance and distribution. Also, scale of sample size is vastly different between the two groups. Soil aggregates may contain multiple micro-habitats[57,58], whereas a much larger area is needed for research into macroorganisms. The consensus is that microorganisms do display biogeographical patterns and, while debate continues on the contributing processes, this is still the case for plants and

animals and is not likely to be resolved soon[59–61]. Microbial ecologists have agreed on characterising microbial biogeographical processes into four distinctions, similar to Vellend's work on macroorganisms[46]. Reviewers describe these as selection, dispersal, drift and diversification[45,62], which will be discussed individually in Section 1.4 of this chapter. The first step, however, in investigating biogeography lies in mapping patterns of diversity.

1.3   PATTERNS OF BIOGEOGRAPHY

Patterns of microbial diversity can be mapped via spatial and temporal dynamics[63,64]. Many studies focus on spatial distribution which is necessary if comparative community diversity is the main objective, providing a snapshot of the sample environment[65]. True local diversity of an ecosystem needs longitudinal analysis to obtain a complete picture of community dynamics[66], with evidence of time also affecting spatial difference in aqueous environments being reported[67]. Examples of microbial spatial studies include bacterial diversity gradients across elevation scales[56,68], global distribution of dominant archaeal populations in soil[69], bacterial community structure in large volume flowing water[70], and spatial patterns of soil microorganisms across an urban park setting[71]. Temporal patterns of microbial communities have been reported in: geothermal springs[72]; soils representing different land-use types[66,73]; the marine water column over 10 years[16]; stream biofilm communities annually[74]; and a range of environments that include air, soil, aquatic, wastewater, human-, and plant-associated microbial biomes[75]. Temporal variation is also often described following a disturbance to microbial habitats, such as a wildfire[76], hurricane[77], climate change[78], or a geological event[79].

Diversity patterns can also be subdivided into three categories, based on the scale or definition of the study area. Alpha diversity examines community structure through the number and/or evenness of taxa within a single habitat[80], beta diversity is a direct comparison between two or more community assemblages[81], and gamma diversity steps up the comparison to a regional level[26]. Many suggestions have been made on how to map biogeographical patterns, with network analysis diagrams[5,82], geographic information systems (GIS)[83] or even specifically designed software programs for microbial communities[84,85]. There have been reports of disparate microbial patterns occurring between spatial and temporal analysis of microbial community assembly[86], indicating the ideal study investigates microbial biogeography through both space and time to comprehensively interpret the driving forces behind ecosystem assembly.

1.4   PROCESSES OF BIOGEOGRAPHY

As already discussed, biogeographical processes have been grouped into four categories for plant, animal, and microbial ecology; selection, dispersal, drift, and diversification[45,46,53,62]. The consensus agrees all processes play a part in producing observed population and community patterns across all ecosystems.

1.4.1   Selection

The majority of research into microbial biogeography has focused on selection due to the ability to measure local environmental conditions[87–89], with several physicochemical parameters reported as the primary driver of microbial diversity across multiple habitats. An investigation into 111 datasets revealed salinity was the main driver of bacterial communities in diverse physical environments[90]. Sediments were shown to be more phylogenetically diverse than soils in this study, whereas soils had increased species-level diversity but below-average phylogenetic diversity[90]. In contrast, pH was found to be the main driver of soil community assembly in two separate studies[55,91], while temperature directly influenced diversity in geothermal microbial communities[92–94]. Species-sorting (i.e., environmental niche partitioning for microbial populations) was found in lakes, urban stormwater runoff, ocean water, and wastewater treatment plants[17,95–97]. Other defining parameters reported as influencing microbial structure include seasonality[16,98], latitude[64,99], and elevation[56]. Additionally, interactions between different taxonomic groups were shown to affect overall diversity of a community[100–103], with variation in genomes thought to be associated with social and ecological interplay[32]. These ecological interactions can be classified as a selective process, given the influence certain microorganisms have on their local physicochemical environment. Clearly, there is no consensus on a single environmental parameter driving microbial diversity, suggesting selection pressures are specific for individual populations and/or communities based on habitat-type. These studies indicate that measuring a full suite of physicochemical variables would aid interpretation of microbial biogeography in any ecosystem.

1.4.2   Dispersal

Dispersal can be a passive or stochastic process, facilitated by the small size of microorganisms as highlighted, for example, by the near global ubiquity of *Pseudomonas syringae* via rainclouds[83]. Endospores of thermophilic *Firmicutes* were reported across the

seabed floor, indicating microbial dispersal can also occur by ocean currents[104]. Other mechanisms enabling microbial dispersal include rivers and streams, with water retention time and immigration major players in shaping aquatic microbial communities[105,106]. Aeolian dispersal was responsible for cosmopolitan taxa discovered in isolated microbial communities in both Antarctica and Africa[107,108], which was reinforced by a recent study suggesting that most microbial strains (average nucleotide identity ≥99.9 %) are globally distributed[109]. A study of conspecific populations in hot springs revealed geographic barriers can limit microbial dispersal to create endemic taxa[6], proposing '*dispersal limitation*' as a more accurate description for this biogeographical process[45]. Measuring dispersal, however, is not as straightforward as selection and recent advances are attempting to calculate this empirical data[109–111].

1.4.3   Diversification

Diversification refers to mutation, speciation, horizontal gene transfer, evolutionary drift, or any significant change to the genotypes of microbial communities[45,62]. Evolutionary change in response to ecological conditions or geographical barriers is a fundamental concept in microbial ecology and has been reported in species- and/or strain-level lineages[8,13,86,112]. The suggested rate of divergence for the 16S rRNA gene (e.g., ~1 % per 50 million years) could limit the detection of more recent diversification within microorganisms[113], with other factors like dormancy also affecting diversification rates[114]. The use of sole markers, like partial 16S rRNA gene sequences, to capture genetic variability is being superseded by multiple loci, whole genome analysis, or even full-length 16S rRNA gene sequences that increase ability to detect gene flow and conspecific lineages with populations[7,13,115–117].

1.4.4   Drift

Ecological drift in microbial communities refers to stochastic changes that occur to population abundance levels due to births and deaths[62]. Measurement of this process has been difficult but recent advances are attempting to quantify stochasticity to incorporate all mechanisms driving microbial diversity[118–120]. The process of drift is thought to affect low-abundance microorganisms more readily than dominant members of a community[62], although a recent study found that dominant taxa in soil habitats were more likely to change in abundance after a disturbance than rare taxa[121]. Quantitative analysis of controlled bacterial communities found drift contributes to the stochasticity observed in environmental surveys, with rates of drift increasing when dispersal decreased[122]. Some argue that biogeographical

processes should not be split into ecological and evolutionary distinctions, and so, include evolutionary drift into this category[45]. It should be noted here again that while a study might show a main driver of microbial diversity in a given sample environment, all processes contribute to community assembly and dynamics. Taxa are likely to be differentially affected by each process, and not react in the same way to individual selection pressures or interferences. A combined approach, investigating all defining processes of microbial diversity within an ecosystem, is desired to understand whole community dynamics and ecological role in the local environment[5].

1.5   MICROBIAL ECOLOGY OF GEOTHERMAL FEATURES

Geothermal features act as model systems to investigate microbial biogeography due to their extreme environmental conditions and distinct separation from neighbouring habitats. Microbial life in geothermal features was brought into prominence by Thomas Brock in the 1960s[123,124], which led to expansive research isolating and characterising novel microorganisms from these ecosystems on a global scale[125–130]. A breakthrough study of endemic *Sulfolobus* populations found in geographically separated geothermal fields highlighted using these microbial '*islands*' to study biogeography[6]. Since then, patterns of diversity (and causal processes) have been proposed for microbial populations and communities in terrestrial geothermal ecosystems from Antarctica[131,132], Canada[92], Chile[133], China[94], Iceland[134], Japan[135–137], Aotearoa-New Zealand[138,139], Pakistan[140], Russia[141,142], Thailand[143], and the USA[93,144] (in particular Yellowstone National Park[145–147]), with some studies including geothermal systems across multiple continents[6,92,148,149]. Research of whole microbial community assemblages has been prevalent[92–94,132,144], but there is also focus on the biogeography of individual populations in these habitats, including *Acidiothiobacillus*[9], Actinobacteria[149], Aquificales[150], Korarchaeota[151], Nanoarchaeota[152], *Sulfolobus*[6,86], Thaumarchaeota[147], and Verrucomicrobia[153]. More recently, biogeography research of geothermal environments has expanded to include thermophilic viruses/phages[154,155], and microbial eukaryotes[148,156]. Using geothermal springs to study microbial biogeography provides a unique opportunity to investigate controlling influences behind the diversity and evolution of life on Earth[157,158].

Selection is the predominantly studied process of microbial biogeography in geothermal ecosystems[92,94,140,147,150,159]. Dispersal, drift, and diversification can be difficult to quantity,

whereas the ability to measure environmental parameters from individual sites, particularly water columns, allows the relationship between physicochemistry and community structure to be examined. The extreme nature of geothermal features must also conceivably influence microbial diversity and community assembly, promoting the effect of environmental selection on microbial biogeography. In multiple studies, temperature was found to be the main driver of microbial community structure in geothermal sediments, mats and/or streamers[92,93,143], particularly where there was a temperature gradient[131,136,145,160], with some studies demonstrating other physicochemical conditions, like pH[94,131,149,161] and/or sulfur[141,143,162], also drive microbial community assembly in geothermal features. Allopatric speciation, the process of geographic isolation enabling the evolution of new species, has been proposed as an additional mechanism for increasing microbial diversity[6,8]. This proposition was corroborated for other thermophilic populations[112,148,149,151], with analysis at finer geographic scale demonstrating allopatry at species/strain level also occurs locally (i.e., within the same region)[150,163]. Geothermal communities in remote locations (e.g., Antarctica) exhibited both endemic and globally cosmopolitan taxa[107], while dispersal of thermophilic endospores in marine sediments is controlled by connectivity of local water masses to ocean currents[104]. Additionally, historical geological events can also shape the distribution and diversity of extremophiles in geothermal systems[79,164,165], highlighting contemporary physicochemical conditions do not solely drive diverging populations in geothermal ecosystems.

Other units of measurement beyond the 16S rRNA gene (e.g., whole genome sequencing, functionality) can also enhance our understanding of how microbial diversity is shaped in geothermal features. Whole genome sequencing found diversification in a single population, suggesting novel lineages are becoming evolutionarily independent of each other within the same environment[7], and that gene sweeps can occur specific to niche differentiation[166]. Functional analysis reported decreased temperature limits to transcripts or inferred function when compared to the presence of corresponding genes[167,168]. Recently, microbial diversity and ecology of geothermal habitats has been expanded via the use of metagenomes[150,161,168], metagenome-assembled genomes (MAGs)[142,155,169] & single-amplified genomes (SAGs)[152,170], which aim to preclude the occurrence of traditional biases in ecology studies. The majority of geothermal microbial research has also focused on spring sediments rather than water columns, due to increased biomass (and subsequent DNA recovery) that is usually associated with solid samples[171]. However, ecological differentiation has been reported

between planktonic and sediment microbial communities from within the same geothermal feature[93,94,172], to which oxygen availability was suggested as the distinguishing factor explaining this dichotomy[173]. Archaeal taxa usually have greater abundances than bacterial taxa in geothermal sediments[146,165,174], although some bacterial populations, for example *Sulfurihydrogenibium*, have been reported to dominate biofilms in geothermal areas[155]. Greater temporal changes also occurred in sediments rather than the aqueous component of a hot spring[175], with seasonal fluid recharge brought on by large precipitation events destabilising microbial populations in these ecosystems[72,86,169]. Therefore, it is prudent to be aware of biases that can occur through both sample medium and unit of measurement in observing ecological processes in any ecosystem to describe community assembly. Focusing on the aqueous component of geothermal springs, with sufficient sample size, could provide a more direct link with selection pressure exerted by physicochemical conditions in these habitats.

1.6   RESEARCH AIMS

A review of literature investigating microbial biogeography has emphasised the need for a statistically robust dataset, collected and analysed with consistent methodologies, to confidently identify controlling mechanisms behind diversity in microbial ecosystems. Geothermal features are model systems to investigate these controls due to their extreme and isolated nature, with multiple volcanic regions around the globe still unexplored in detail. The focus on geothermal sediment samples has also introduced a significant knowledge gap on the diversity of spring water columns, with alternating controls reported for community and population structures. Disparate environmental parameters have been found as primary drivers of geothermal community assembly, indicating that an extensive suite of physicochemical variables must be measured to ensure credibility in defining environmental niches. Additionally, evidence highlights the many limitations associated with sole use of the 16S rRNA gene as a marker for defining both taxonomic diversity and evolutionary changes, proposing whole genome sequencing, either through isolates or metagenomic samples, as an avenue for designating genotypic variation.

The Taupō Volcanic Zone (TVZ) in Aotearoa-New Zealand presents an ideal opportunity to address these shortcomings that currently exist in the literature. The 8,000 km^2 region in Aotearoa-New Zealand's North Island is both compact and accessible, extending from

Whakaari-White Island near the Bay of Plenty coastline to Mount Ruapehu in the Central Plateau. It is an active volcanic and geothermal zone sitting near the Pacific-Australasian convergent plate boundary, and is home to an estimated 10,000 geothermal features[176]. Extremophilic microorganisms have been isolated from individual features across the region, showing novel diversity and function in these ecosystems[127,177–184]. Ecology studies are less frequent, however, and have focused on soil, sinter and/or sediment samples[9,92,185,186], or individual springs[138,139,187–189].

Therefore, to embrace this opportunity and develop a greater understanding of the driving influences behind microbial biogeography, I propose to sample and investigate the spatial and temporal biogeography of 1,000 geothermal spring water columns in the TVZ, both at community and population level, from the 1,000 Springs Project (MBIE title: *Microbial Bioinventory of Geothermal Ecosystems*). This collaborative project, between the Te Whare Wānanga o Waikato-University of Waikato and Te Pū Ao-GNS Science, was a first-of-its-kind for any geothermal system in the world, with a primary goal to collate physicochemical and microbial diversity information of 1,000 geothermal springs in the TVZ. All data collected is publicly presented via a user-friendly website, https://1000springs.org.nz/, to increase public knowledge of geothermal microbial diversity, complement mātauranga (i.e., indigenous knowledge), and assist kaitiakitanga (i.e., guardianship) of these novel ecosystems in Aotearoa-New Zealand. Evidence has shown that big data sets are needed to accurately identify patterns of biodiversity[20,71,190], and large-scale investigations into geothermal ecology in Aotearoa-New Zealand have so far focused on soil, sediment, sinter and/or biomat samples[9,92]. Indeed, large-scale microbial studies directly sampled from Yellowstone National Park geothermal features have included only ≤100 sediments and/or biofilms[151,159,191], or 15 geothermal water columns (which were also compared to associated sediments)[172]. The attributes of soil are known to affect the microbial populations (e.g., particle size, depth, nutrient composition)[192,193], whereas aqueous samples exhibit a more homogeneous chemical and community structure, providing greater credibility in linking microbial communities with the physicochemical environment. A large-scale microbial ecology study concentrating entirely on geothermal planktonic communities across an entire region is unprecedented and will expand our knowledge on the driving forces behind microbial diversity and evolution.

1.7   HYPOTHESES & OBJECTIVES

The three main hypotheses proposed by this study are:

- Hypothesis 1 (H1): Extremophiles exhibit spatial biogeographical patterns and are predominantly governed by contemporary environmental conditions and dispersal limitation
- Hypothesis 2 (H2): Temporal patterns of microbial biogeography remain consistent in geothermal springs if physicochemical conditions are stable
- Hypothesis 3 (H3): Individual population diversity and distribution can be controlled by different processes to microbial communities

To test these hypotheses, the following objectives are proposed:

1. Measure microbial and physicochemical components of 1,000 geothermal spring water columns in the TVZ.
2. Define spatial patterns and underlying drivers of microbial community assembly across the region through taxonomic classification, diversity metrics, and dissimilarity statistics.
3. Sample a subset of the spatial dataset over time to investigate if temporal patterns of microbial biogeography exist for geothermal ecosystems in the TVZ.
4. Highlight the biogeography of a single geothermal population within the region and determine factors that shape distribution.

1.7.1   Measure microbial and physicochemical components of 1,000 geothermal spring water columns in the TVZ

In this objective, I will collect >1,000 geothermal spring samples from the TVZ. Consistent sampling and processing techniques will be established to ensure confidence in sample comparison, with the 16S rRNA gene used to taxonomically define geothermal bacterial and archaeal communities. Forty-six physicochemical parameters from the local environment will be measured to complement microbial data, with whole genome and metagenomic analyses also performed on a single population from these habitats to compare if a biogeographical dichotomy exists between community and population structures. Due to the significant sample size planned for this study, the hypervariable V4 region of the 16S rRNA gene will be used to measure microbial diversity, using the original primers from the Earth Microbiome Project[20]. At the time of methodology development in 2013, this was the best available

primer set to capture both bacterial and archaeal signatures from environmental samples[194]. Consistent analyses are also planned to measure the 46 physicochemical parameters from each spring environment using either a multiparameter field meter *in situ*, inductively coupled plasma-mass spectrometry (ICP-MS), UV-vis spectrometry, titration, or ion chromatography. Field work for a study of this scale will be meticulously planned and executed to ensure consistency across sampling and sample processing, with appropriate access permissions acquired and health and safety practices completed.

1.7.2 Define spatial patterns and underlying drivers of microbial community assembly

Sequencing reads of the 16S rRNA gene will be clustered into operational taxonomic units to classify community structure using a curated, taxonomic database. Bacterial and archaeal patterns will be additionally analysed by both alpha (i.e., OTU richness and Shannon indices) and beta (i.e., Bray-Curtis dissimilarity) diversity metrics to enable an accurate representation of microbial biogeography to be reported. From the four processes of biogeography discussed in this literature review, selection and dispersal will be investigated in detail. It is proposed that environmental selection is the main driver of diversity in geothermal environments[144,147], and with features in close proximity to each other having substantial physicochemical dissimilarities, unique microbial islands are created only metres apart. The scale of their heterogeneous physicochemical makeup would suggest that microbial diversity is selected based on these characteristics, and this hypothesis will be investigated using the broad scale of physicochemical attributes measured in conjunction with linear regression modelling, correlation statistics, and multivariate analysis. Dispersal is also thought to be a significant driver of community assembly, mainly due to microorganisms' size and ability to spread easily[83,107]. This process will be examined by confirming the presence or absence of a distance-decay pattern (i.e., microbial communities decrease in similarity to each other with increased geographic distance) across the region.

1.7.3 Temporal patterns of microbial biogeography in TVZ geothermal ecosystems

Using identical methodologies to those discussed in Section 1.7.2, longitudinal diversity patterns of bacterial and archaeal communities in the TVZ will be studied over a multiyear period to ascertain microbial stability potential. I will do this by investigating three different categories of samples: control sites to determine stochasticity in geothermal microbial structures; disturbed sites to investigate the effect of both short- and long-term, and natural and anthropogenic disturbances to resident geothermal communities; and pH sites to assess

the state of community structure across the wide pH range found in TVZ geothermal springs (pH 3-9) while temperature remains relatively constant (60-70 °C). Taxonomic, alpha, and beta diversity measures will be employed to assess the level of ecosystem stability in these extreme habitats.

1.7.4   Biogeography of a single geothermal population within the TVZ

The most abundant population in TVZ geothermal spring water columns, as identified from Sections 1.7.1 and 1.7.2, will be examined in more detail to assess factors that shape taxon diversity and distribution. In order to define diversity, I will assess in depth the 16S rRNA gene diversity of this taxon at different taxonomic hierarchies (using data generated from Section 1.7.1), with both OTU number and read abundance used as a proxy for defining diversity. Patterns of distribution and abundance, and statistically defined associations with physicochemistry will be investigated. Whole genome sequencing will be performed on any local isolates that have been characterised to complement this diversity analysis. Additionally, I will examine metagenomic samples from available TVZ geothermal springs to assess the level of agreement with amplicon sequencing, and metagenome-assembled genomes of the taxon will be produced if sufficient sequencing reads are present. Finally, phylogenomics will rigorously define taxonomy and phylogeny of the population.

1.8   THESIS OVERVIEW

1.8.1   Chapter 2 | Microbial biogeography of 925 geothermal springs in Aotearoa-New Zealand

The initial part of the study involved the collection of microbial and physicochemical data from geothermal springs in the TVZ. A total of 974 individual springs and 1,019 samples were collected, with all data collated into a custom-made, publicly available database via https://1000springs.org.nz/. Prokaryotic (for the purposes of this thesis refers to domain Bacteria and Archaea) cells from the water column, collected onto a sterile column filter, were sent to the Thermophile Research Unit at the Te Whare Wānanga o Waikato-University of Waikato for DNA extraction and subsequent 16S rRNA gene sequencing, using the Ion PGM™ System for Next-Generation Sequencing. The physicochemical samples were either measured *in situ*, analysed at the Wairakei Research Centre, Te Pū Ao-GNS Science or at the Te Aka Mātuatua-School of Science, Te Whare Wānanga o Waikato-University of Waikato.

A custom bioinformatics pipeline was designed to process amplicon sequences and cluster the reads into operational taxonomic units (OTUs) for taxonomic classification and determination of associated read abundance. Both microbial and physicochemical data were then analysed to investigate both local (i.e., alpha) and regional (i.e., beta) diversity, and the underlying processes driving any observed patterns. Statistical analyses implemented included alpha diversity metrics, beta dissimilarity indices, non-metric multi-dimensional scaling, and canonical correspondence analysis. This chapter formed the basis of the manuscript Power *et al*. (2018)[195], which was published in *Nature Communications*. The data generated by this chapter also contributed to an additional manuscript on protistan diversity published in *The ISME Journal* (Oliverio *et al*. [2018])[156], for which I was a second author.

1.8.2 Chapter 3 | Temporal dynamics of geothermal microbial communities in Aotearoa-New Zealand

Three categories of geothermal features analysed in the spatial study were selected to investigate temporal behaviour of microbial communities in the TVZ. These three categories were: control sites, which included seasonality as a potential influence on community structure; the effect of a short-term, natural disturbance on resident microbial communities in a geothermally sourced stream, and a long-term, anthropogenic disturbance to microbial communities of a geothermal wetland undergoing restoration; and stochastic variation across time of the microbial communities in geothermal features from across the pH range found in TVZ systems (which included the collection of an additional 69 samples). The taxonomic, alpha, and beta diversity of 31 geothermal features were examined over a 34-month timeframe, accompanied by 44 physicochemical variables measured for each sample. This chapter was developed into the manuscript Power *et al*. (2023)[196] which was published in the journal *Frontiers in Microbiology*.

1.8.3 Chapter 4 | Allopatric speciation in the bacterial phylum Aquificota enables genus-level endemism in Aotearoa-New Zealand

The most abundant genus found in TVZ geothermal springs was *Venenivibrio* (family Hydrogenothermaceae) from the phylum Aquificota. Extensive investigations into the ecology of this phylum and *Venenivibrio* populations in the TVZ were performed, which involved traditional alpha diversity measures, total read abundances of the genus across 467 geothermal springs, and statistical analyses with spring physicochemistry. Beta diversity between *Venenivibrio* populations was examined to investigate dispersal capability across the

region, with evidence for the taxon outside of the Aotearoa-New Zealand archipelago explored. The genome of the type strain for the genus, *Venenivibrio stagnispumantis* CP.B2^T, was sequenced and annotated to corroborate a chemolithoautotrophic lifestyle, with predicted metabolism compared to other Hydrogenothermaceae. Both local (i.e., from within Aotearoa-New Zealand; *n*=16) and global (i.e., from outside Aotearoa; *n*=188) metagenomes were taxonomically classified and aligned to known Hydrogenothermaceae genomes, with metagenome-assembled genomes (MAGs; *n*=4) and phylogenomics used to confirm exclusivity to Aotearoa-New Zealand. This chapter was developed into a manuscript that has been submitted to *Nature Communications* for peer-review.

1.8.4   Chapter 5 | Summary, conclusions, & future work

The final chapter of this thesis summarises major findings from spatial, temporal, and population analyses of 925 geothermal spring ecosystems in Aotearoa-New Zealand. A study of this scale for geothermal systems is unprecedented, and the volume of data has allowed for statistically robust conclusions to be drawn on the patterns and underlying processes of microbial diversity in these extreme habitats. As with most research, the findings have unearthed many questions, and avenues for future work, using this study as a sound basis, are explored.

1.8.5   Appendices

All references for the thesis are consolidated into Appendix A. The supplementary information for Chapter 2, including supplementary methods, figures, and tables, is in Appendix B. Appendix C is the supplementary information that complements Chapter 3, and Appendix D comprises extensive information to support findings reported in Chapter 4. The genome sequencing of *Venenivibrio stagnispumantis* CP.B2^T was published in *Microbiology Resource Announcements*[197], which is outlined in Appendix E. Finally, Appendix F contains the signed authorship forms for all chapters prepared as manuscripts for peer-review.

CHAPTER 2

MICROBIAL BIOGEOGRAPHY OF 925 GEOTHERMAL SPRINGS IN AOTEAROA-NEW ZEALAND

Jean F. Power, Carlo R. Carere, Charles K. Lee, Georgia L.J. Wakerley, David W. Evans, Mathew Button, Duncan White, Melissa D. Climo, Annika M. Hinze, Xochitl C. Morgan, Ian R. McDonald, S. Craig Cary, & Matthew B. Stott

2.1 PREFACE

This chapter introduces the first study on spatial microbial biogeography using geothermal spring water columns from the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand, which was published as a manuscript in the journal *Nature Communications*[195]. A literature review of microbial biogeography, in particular focusing on geothermal ecosystems, highlighted the need for extensive sampling scales and density across both geographic distance and physicochemical gradients to statistically define causal processes of microbial diversity in these extreme habitats. Importantly, previous microbial biogeography research in geothermal systems had focused on geothermal sediment samples over hot spring water columns. This chapter was conceived to close these gaps in the literature by contextualising the microbial diversity, along with associated physicochemical and geographical data, from 1,000 geothermal springs in the TVZ. In doing so, the first hypothesis that extremophiles exhibit spatial biogeographical patterns could be tested, with environmental selection and dispersal proposed as mechanisms constraining patterns of diversity across the region.

While I wrote this chapter and performed all analyses presented here, the collection of 1,019 samples from 974 individual geothermal springs was a massive collaborative effort between Te Pū Ao-GNS Science and the Te Whare Wānanga o Waikato-University of Waikato to create a bioinventory of microbial diversity and ecology in TVZ geothermal springs, colloquially named the 1,000 Springs Project. The majority of co-authors for this chapter were part of this project, which was directed by principal investigators Matthew Stott and

Craig Cary. Melissa Climo was involved in the initial planning stage and significantly contributed to the successful MBIE grant application.

As the primary author of this research chapter, I was charged with field sampling for the 1,000 Springs Project. I planned, coordinated and executed all field work, and developed standardised methodologies for collection, processing, recording of metadata, and storage of samples for both microbial and physicochemical analyses, with assistance from David Evans, Carlo Carere, Matthew Stott, and various staff at Te Pū Ao-GNS Science, Taupō. I filtered 2 L of each spring sample to collect planktonic communities from the water column, and measured spring temperature, pH, oxidation-reduction potential, conductivity, dissolved oxygen, turbidity, and ferrous iron from each sample. I prepared samples for all remaining physicochemical analyses, which were carried out by the New Zealand Geothermal Analytical Laboratory at Te Pū Ao-GNS Science and the Te Aka Mātuatua-School of Science, Te Whare Wānanga o Waikato-University of Waikato. I contributed via end-user feedback to the custom-made field application, database, and website design, which were produced by Mathew Button, Duncan White, and Annika Hinze for comparably collecting, collating, and presenting data. DNA extractions and sequencing were done by the Thermophile Research Unit at the Te Whare Wānanga o Waikato-University of Waikato, primarily by Georgia Wakerley. While Charles Lee processed sequences that appear on the 1,000 Springs Project website, I developed my own custom bioinformatical pipeline for processing of raw sequences into operational taxonomic units (OTUs) and associated read abundances that were used in this chapter. I performed all statistical and ecology-based analyses for the chapter, with direction from Matthew Stott and Carlo Carere. Xochitl Morgan provided statistical support, in particular for the multivariate linear model to define differential enrichment of taxa abundances against physicochemistry and location. Finally, I wrote the first draft of this chapter, which was then revised by Matthew Stott, Carlo Carere, Craig Cary, and Ian McDonald, with final edits made by all co-authors. A signed co-authorship form confirming these contributions can be found in Appendix F.

The remainder of this chapter and associated appendices are a reproduction of the manuscript text. Supplementary material accompanying this chapter is presented in Appendix B. All scripts developed for statistics and figures are available through GitLab (https://gitlab.com/morganlab/collaboration-1000Springs/1000Springs), with raw sequences deposited in the European Nucleotide Archive (ENA) under study accession PRJEB24353.

2.2 ABSTRACT

Geothermal springs are model ecosystems to investigate microbial biogeography as they represent discrete, relatively homogeneous 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 Aotearoa-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 at >70 °C. Further, community dissimilarity increases with increasing 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 %) 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.

*This chapter was written before recent nomenclature changes to bacterial and archaeal phyla names. For example, the bacterial phyla Aquificae and Proteobacteria, which are discussed in this chapter, are now known as Aquificota and Pseudomonadota, respectively. For additional valid phyla names, please refer to Oren & Garrity, 2021[198].

2.3   INTRODUCTION

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[199,200] via a combination of environmental selection, stochastic drift, diversification and dispersal limitation[45,62]. 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[91,201–203].

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[92,93,139], pH[172], and, seasonality[72] as the primary drivers of bacterial and archaeal communities in these ecosystems; with niche specialisation observed within both local and regional populations[147,151]. Other geochemical variables, particularly hydrogen[204,205] and nitrogen[167,206] may also contribute to community structure. Concurrently, the stochastic action of microbial dispersal is thought to be a significant driver behind the distribution of microorganisms[107], with endemism and allopatric speciation reported in intercontinental geothermal springs[6,8]. It is important to note that significant differences have been found between aqueous and soil/sediment samples from the same springs[93,94,172], 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 gradients 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 8,000 km^2 in Aotearoa-New Zealand's North Island (Figure 2.1), 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 magmatism[207]. 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 physicochemically heterogeneous surface features in geographically distinct geothermal fields[208]. Previous microbiological studies across the region have hinted at novel diversity and function present within some of these features[127,177,179,182,209], however investigations into the biogeographical drivers within the TVZ are sparse and have focused predominantly on soil/sediments or individual hot springs[92,139,187].

Here we report the diversity and biogeography of microbial communities found in geothermal springs, collected as part of the 1,000 Springs Project. This project aimed to catalogue the microbial biodiversity and physicochemistry of Aotearoa-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 (https://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 1,019 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 Aotearoa-New Zealand geothermal springs[156]. 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.
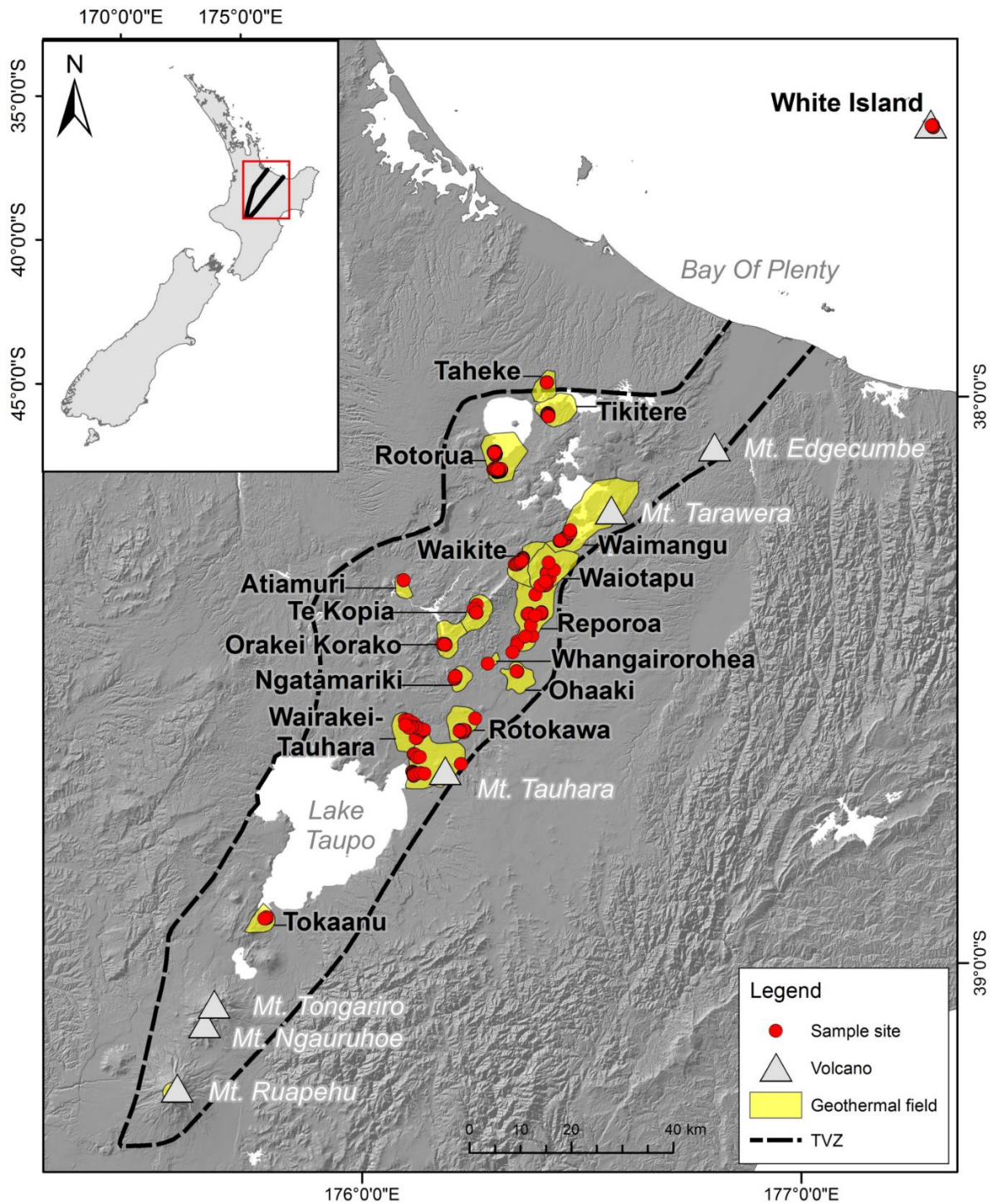
Figure 2.1 - Map of the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand. The geothermal fields from which samples were collected are presented in yellow. All sampled geothermal springs (*n*=1,019) are marked by red circles. The panel insert highlights the location of the TVZ in the central North Island of Aotearoa-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.* (1995)[207].

2.4 METHODS

2.4.1 Field sampling & processing

Between July 2013 and April 2015, 1,019 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 (Table B.1). All samples were filtered within two hours of collection and stored accordingly. Total DNA was extracted using a modified CTAB method[210] 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 Project[20] primers F515 (5'-GTGCCAGCMGCCGCGGTAA-3') and R806 (5'-GGACTACVSGGGTATCTAAT-3'), details of which can be found in the 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 Table B.1), 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). Conductivity (COND), dissolved oxygen (dO), oxidation-reduction potential (ORP), pH, and turbidity (TURB) were determined using a Hanna Instruments (Woonsocket, RI, USA) multiparameter field meter 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 the Supplementary methods and Table B.1.

2.4.2 DNA sequence processing

DNA sequences were processed through a custom pipeline utilising components of UPARSE[211] and QIIME[212]. An initial screening step was performed in mothur[213] 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 v7[211] 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 operational taxonomic units (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 Classifier[214] (with a minimum confidence score of 0.5) against the SILVA 16S rRNA database (123 release, July 2015)[215]. 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 9,500 reads per sample. Consequently, 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).

2.4.3 Statistical analyses

All statistical analyses and visualisation were performed in the R environment[216] using phyloseq[217], vegan[218] and ggplot2[219] 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 ($|r|$>0.7) were detected[220]. 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 associating 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 run[221]. Samples were also binned incrementally by pH (single pH units), temperature (10 °C increments), and geographic ranges (geothermal field; Figure B.1), with non-parametric Kruskal-Wallis (H) testing coefficient 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 Wilcox 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 ($P<0.01$) were assessed for collinearity using Pearson's coefficient $|r|$[221]. 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 package (v2.1.5)[222].

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 edgeR[223]. 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 visualized using ggtree[224]. A phylogenetic tree was generated in QIIME by confirming alignment of representative OTU sequences using PyNAST[225], filtering the alignment to remove positions which were gaps in every sequence and then building an approximately maximum-likelihood tree using FastTree[226] with a midpoint root.
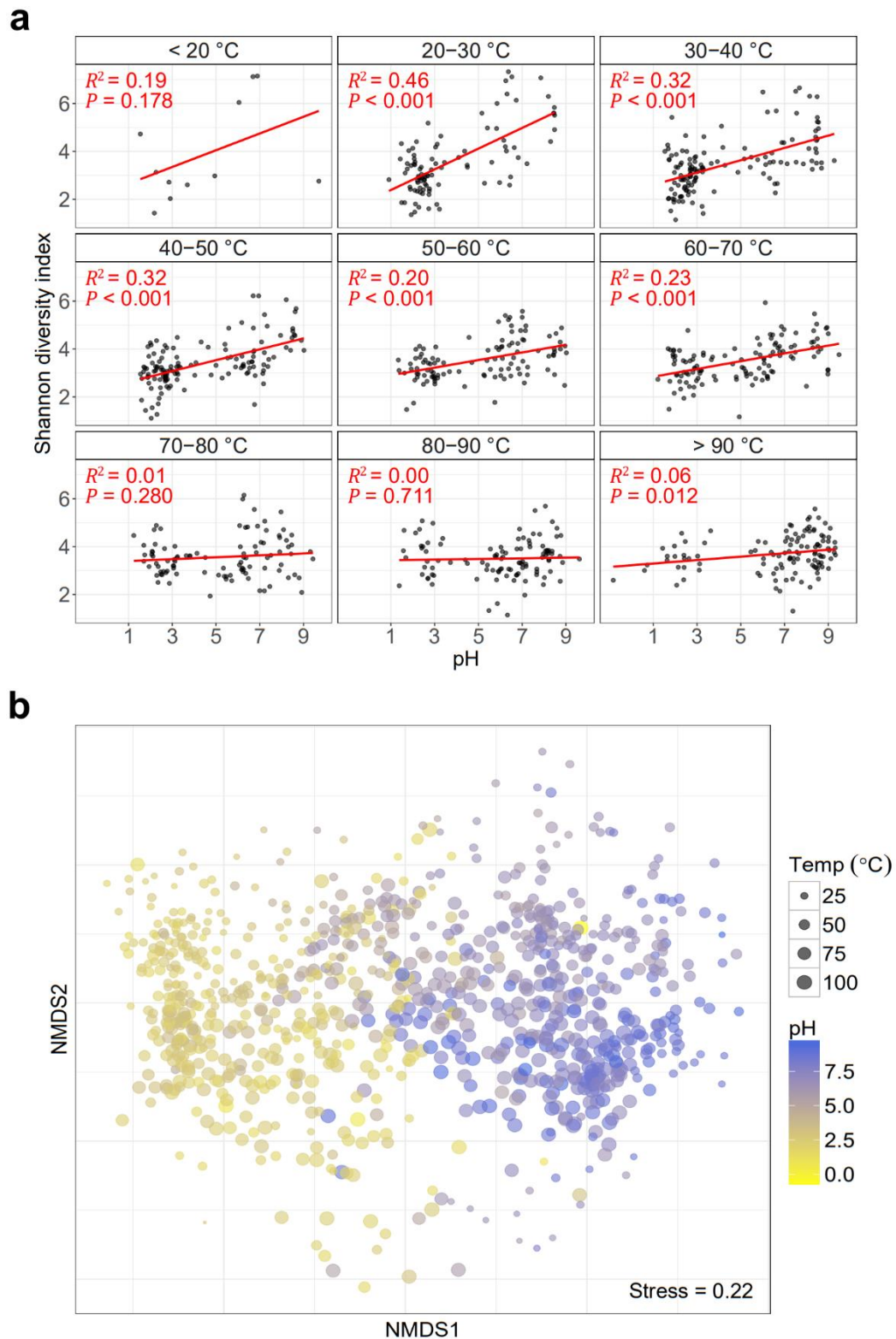
Figure 2.2 - Alpha and beta diversity as a function of pH and temperature. **(a)** pH against alpha diversity via Shannon index of all individual springs (n=925) in 10 °C increments, with linear regression applied to each increment. **(b)** Non-metric multidimensional scaling (NMDS) plot of beta diversity (via Bray-Curtis dissimilarities) between all individual microbial community structures sampled (n=925).

2.5 RESULTS & DISCUSSION

2.5.1 Geothermal feature sampling

Recent biogeography research has demonstrated microbial diversity patterns are detectable and are influenced by both deterministic[201] and stochastic processes[202]. A lack of consensus on the relative impact of these factors, however, has been exacerbated by an 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 ecosystems[57] further confounds attempts to distinguish true taxa-geochemical associations. To provide greater resolution to the factors driving microbial biogeography processes, we collected 1,019 geothermal water-column samples from across the TVZ and assessed physicochemical and microbial community composition (Figure 2.1). Samples included representatives of both extreme pH (<0–9.7) and temperature (13.9–100.6 °C) (Figure B.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 Section B.1.5). From these 925 springs, a total of 28,381 operational taxonomic units (OTUs; 97 % similarity) were generated for diversity studies.

2.5.2 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[92,159], 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; Figure B.2 & Figure B.3). As anticipated, average OTU richness (386 OTUs; Figure B.4) was reduced in comparison to studies of non-geothermal temperate terrestrial[71] and aquatic[199] 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[91,92]. A comparison of 46 individual physicochemical parameters (Table B.2) confirmed pH as the most significant factor influencing diversity

(16.4 %, linear regression: *P*<0.001; Figure B.3), with diversity increasing from acidic to alkaline pH. Further multiple regression analysis showed NO_3^-, turbidity, ORP, dissolved oxygen, NO_2^-, Si, and Cd also had meaningful contributions (Table B.3). 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; Figure B.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 Aotearoa-New Zealand and globally (e.g. soil[91,227], water[156,228], alpine[68,229]).

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 characteristics[91]. Similarly, acidic pH will also reduce the number of taxa observed due to the decreased number that can physiologically tolerate these conditions[191] 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 (Figure 2.2). Inversely, the effect of temperature on diversity was lessened in springs where pH was <4 (Figure B.5). There is some evidence that suggests thermophily predates acid tolerance[191,230], 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 areas[231]. Indeed, a recent investigation of thermoacidophily in Archaea suggests hyperacidophily (growth pH <3.0) may have only arisen as little as ~0.8 Ga[191], thereby limiting the opportunity for microbial diversification; an observation highlighted by the paucity of these microorganisms in extremely acidic geothermal ecosystems[139,191]. It is also important to note that salinity has previously been suggested as an important driver of microbial community diversity[20,90]. The quantitative data in this study showed only minimal influence of salinity (proxy as conductivity) on diversity (linear regression: R^2=0.001, *P*=0.272; Table B.2), bearing in mind that the majority of the geothermal spring samples in this study had salinities substantially less than that of seawater.
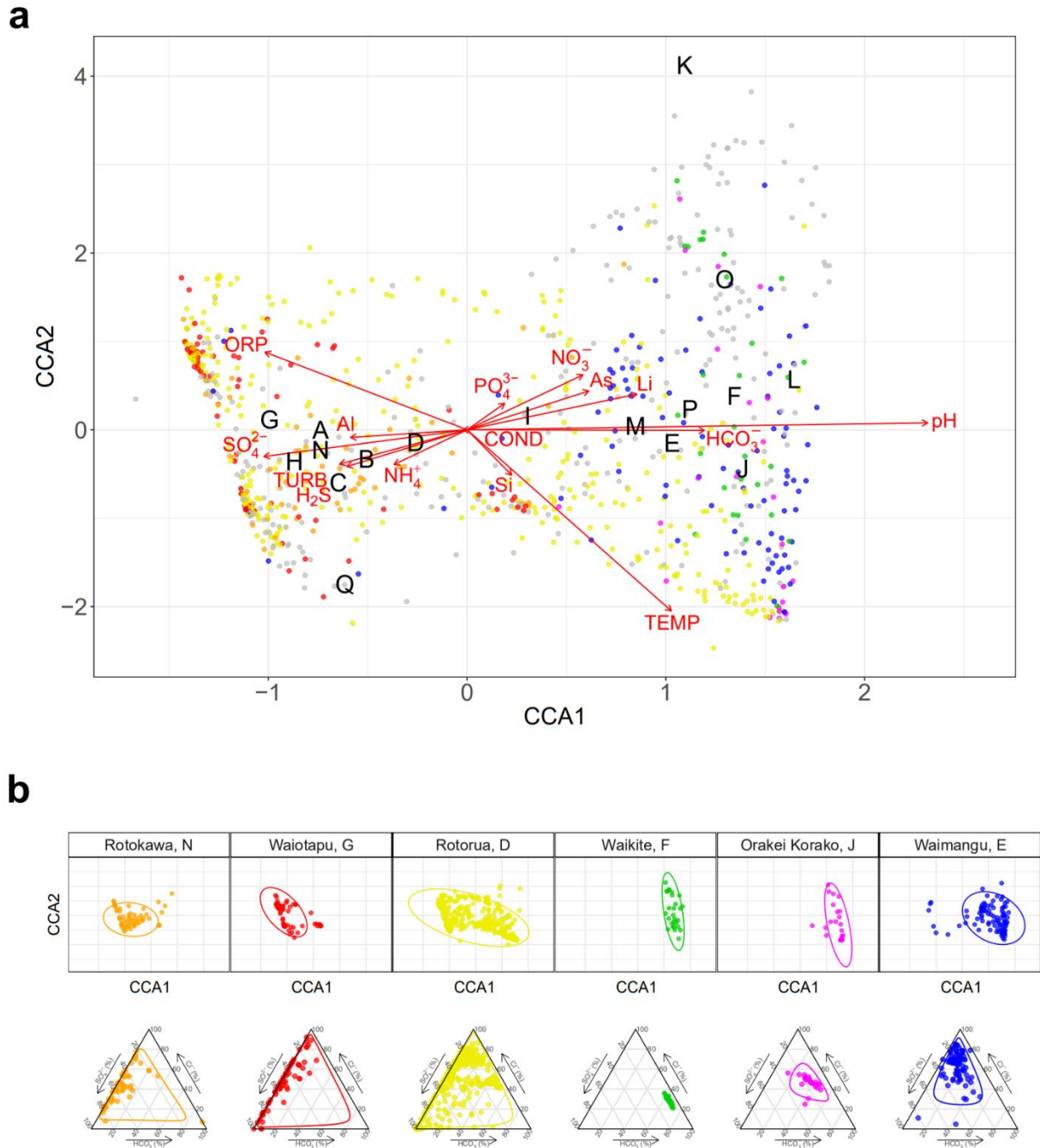
Figure 2.3 - Constrained correspondence analysis (CCA) of beta diversity with significant physicochemistry. **(a)** A scatter plot of spring community dissimilarities (n=923), with letters corresponding to centroids from the model for geothermal fields (A-Q; White Island, Taheke, Tikitere, Rotorua, Waimangu, Waikite, Waiotapu, 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^-).

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 springs[92], which showed a strong relationship (R^2=0.40–0.44) existed. In contrast, our data across the entire suite of samples, revealed temperature had no significant influence on observed community diversity (R^2=0.002, P=0.201; Figure B.3, Table B.2). This result increased marginally for archaeal-only diversity (R^2=0.013, P=0.0005), suggesting temperature has a more profound effect on this domain than it does Bacteria. However, the primers used in this study are known to be unfavourable towards some archaeal clades[232], therefore it is likely 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 (Table B.3), 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.* (2014)[92], we observed increasing temperature only significantly constrained diversity above moderately acidic conditions (pH >4; Figure B.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 processed[172]. 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[92,159] (i.e., soils, sediments, streamers or biomats). Whereas aqueous samples typically exhibit a more homogeneous chemistry and community structure, the heterogeneity of terrestrial samples is known to affect microbial populations (e.g., particle size, depth, nutrient composition)[57]. Our deliberate use of aqueous samples extends the results of previous small-scale work[93,187] and also permits the robust identification of genuine taxa-geochemical relationships in these environments.

2.5.3 Microbial communities are influenced by pH, temperature and source fluid

Throughout the TVZ, beta diversity correlated more strongly with pH (Mantel: ρ=0.54, P<0.001) than with temperature (Mantel: ρ=0.19, P<0.001; Figure 2.2, Table B.4). This trend was consistent in pH- and temperature-binned samples (Figure B.6; ANOSIM: $|R|$=0.46 and 0.18 respectively, P<0.001); 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 °C), the effect of temperature reducing beta diversity had greater significance above 80 °C (Wilcox: P<0.001; Figure B.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 %), turbidity (0.8 %), and As (0.7 %, $P<0.001$; Table B.5). Interestingly, constrained correspondence analysis of the 15 most significant, non-collinear, and variable parameters (Table B.5 & Table B.6; pH, temperature, turbidity, ORP, SO_4^{2-}, NO_3^-, As, NH_4^+, HCO_3^-, H_2S, conductivity, Li, Al, Si, and PO_4^{3-}), along with geothermal field locations, only explained 10 % of variation in beta diversity (Figure 2.3), 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 fluids[233], 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 (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 values[234]. 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: $H=240.7$, $P<0.001$; Figure 2.4), community structures were significantly different between geothermal fields (ANOSIM: $|R|=0.26$, $P<0.001$; Figure B.7). Gradient analysis comparing significant geochemical variables and geography further identified meaningful intra-geothermal field clustering of microbial communities (95 % CI; Figure 2.3 & Figure B.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; Figure 2.3 position N & F) display opposing ratios of HCO_3^-, SO_4^{2-}, 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., Rotorua, Figure 2.3 position D), demonstrating that correlation does not necessarily always occur between locational proximity and physicochemistry.

Table 2.1 - Average relative abundance and prevalence of phyla and genera. Only taxa above a 1 % average compositional threshold are shown. Maximum abundance of each taxon within individual features and standard deviation (SD) across all 925 springs. Where taxonomy assignment failed to classify to genus level, the closest assigned taxonomy is shown (f=family, o=order, p=phylum).

Phylum*	Genus	Abundance	SD*	Max	Prevalence
Aquificae	*Venenivibrio*	0.112	0.231	0.968	0.742
Proteobacteria	*Acidithiobacillus*	0.111	0.242	0.994	0.629
Aquificae	*Hydrogenobaculum*	0.100	0.235	0.999	0.608
Aquificae	*Aquifex*	0.086	0.212	0.971	0.497
Deinococcus-Thermus	*Thermus*	0.025	0.071	0.732	0.552
Proteobacteria	*Thiomonas*	0.024	0.101	0.941	0.396
Proteobacteria	*Desulfurella*	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	*Thiovirga*	0.015	0.077	0.816	0.374
Proteobacteria	Hydrogenophilaceae (f)	0.015	0.072	0.704	0.406
Thermodesulfobacteria	*Caldimicrobium*	0.015	0.052	0.651	0.519
Proteobacteria	*Hydrogenophilus*	0.013	0.045	0.432	0.484
Thermotogae	*Mesoaciditoga*	0.011	0.033	0.286	0.410
Parcubacteria	Parcubacteria (p)	0.010	0.024	0.193	0.608

*This table was generated before recent nomenclature changes to bacterial and archaeal phyla. For current valid names, please refer to Oren & Garrity, 2021[198].

2.5.4 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 % (Figure 2.5). 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 2.1). 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 2.1), found predominantly at temperatures less than 50 °C (Figure B.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*[235], 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 2.1). As Aquificae are thermophilic (T_{opt} 65–85 °C)[236], they were much more abundant in warmer springs (>50 °C ; Figure B.9). The minimal growth temperature reported for characterised Aquificales species (*Sulfurihydrogenibium subterraneum* and *S. kristjanssonii*)[236] 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 representatives[236]. Of the 14 currently described genera within the Aquificae, six genera were relatively abundant in our dataset (average relative abundance >0.1 %; Figure 2.5): *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 vents[236]. *Venenivibrio* (OTUs; *n*=111) was also the most prevalent and abundant genus across all communities (Table 2.1). This taxon, found in 74.2 % (*n*=686) of individual springs sampled, has only one cultured representative, *V. stagnispumantis* (CP.B2[T]), which was isolated from the Waiotapu geothermal field in the TVZ[179]. 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[209,237].

2.5.5 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 °C and >50 °C, respectively; Figure B.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 (Figure 2.5). 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 chemolithoautotroph[238] 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[239,240]. *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.[241] and conversely, the preferred acidic ecological niche of *Hydrogenobaculum*[242]. *Aquifex* was the only genus to display above average association with temperature, confirming abundance of this taxon is significantly enhanced by hyperthermophily[243].

Similar to the chemical-taxa associations discussed above, differential abundance relationships were calculated with respect to individual geothermal fields (Figure 2.5). 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 '*Methylacidiphilum*' (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.

a

$H = 240.7$, $P < 0.001$

b

$m = 0.031$, $R^2 = 0.045$, $P < 0.001$
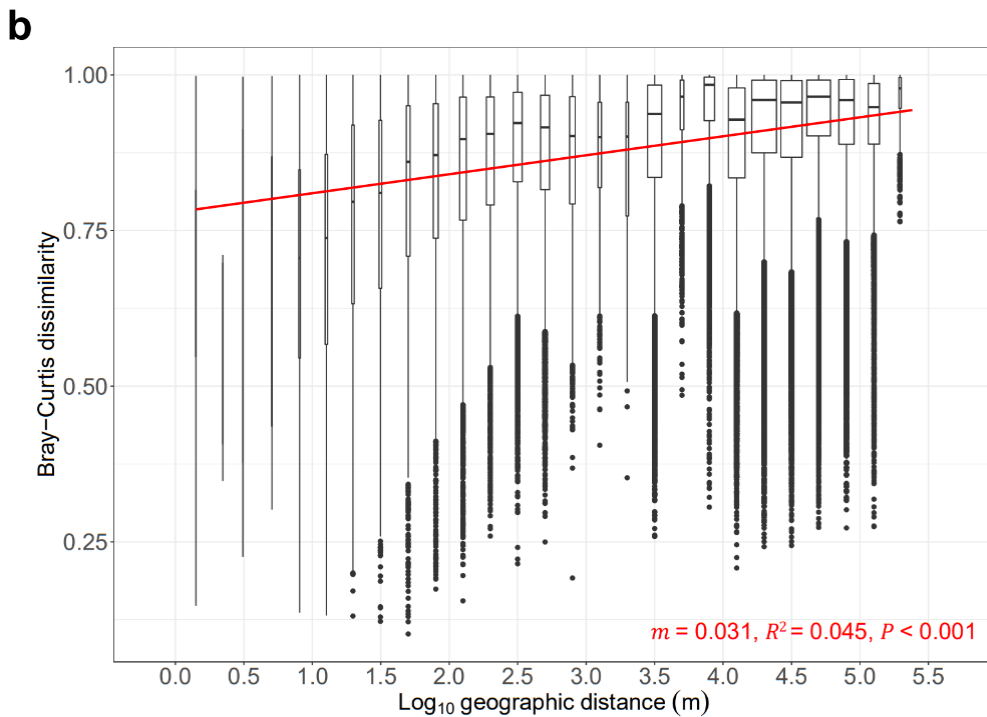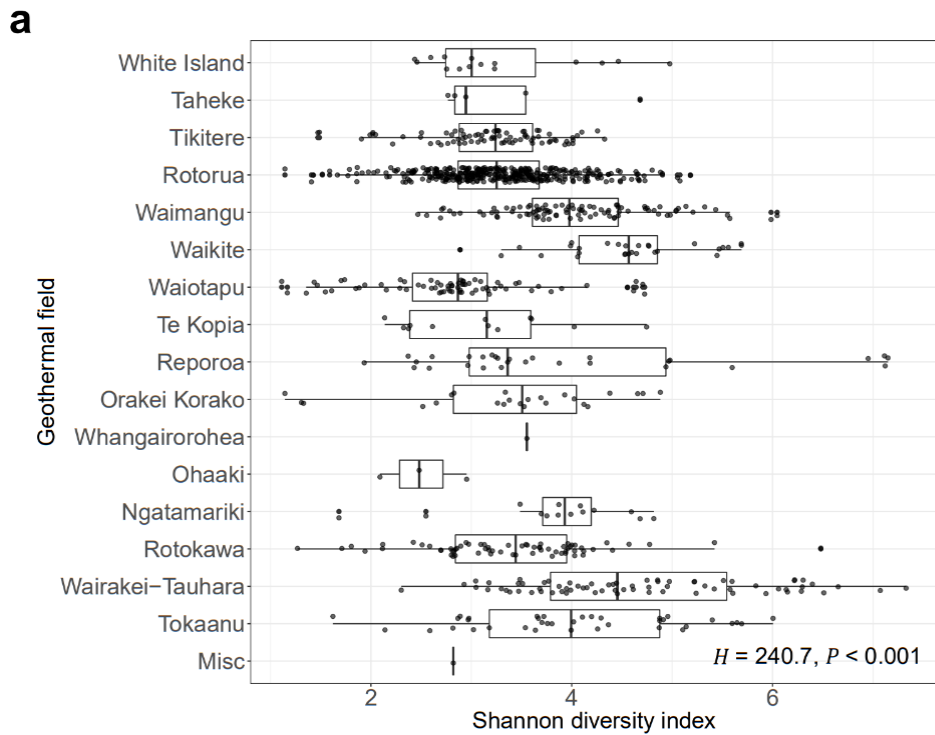
Figure 2.4 - Alpha and beta diversity as a function of geographic distance. **(a)** Alpha diversity scales (via Shannon index) across individual springs (*n*=925), split by geothermal fields. Fields are ordered from north to south (*H*: Kruskal-Wallis test). **(b)** A distance-decay pattern of beta diversity (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.
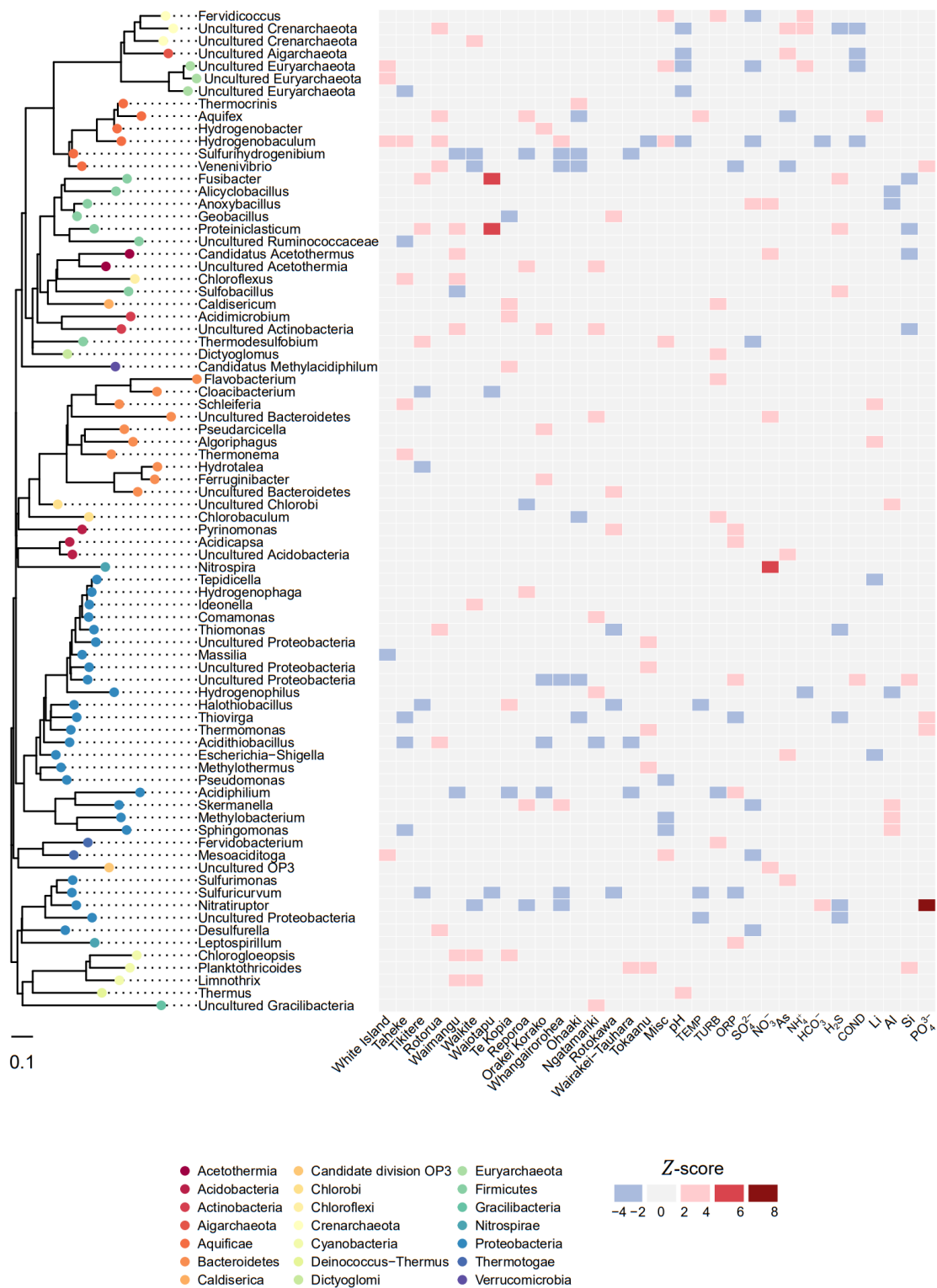
Figure 2.5 - Taxonomic association with location and physicochemistry. The heat map displays positive (red) and negative (blue) association of genus-level taxa (>0.1 % average relative abundance) with each geothermal field and significant environmental variables, based on Z-scores of abundance log ratios. Each taxon is colour-coded to corresponding phylum* on the approximately maximum-likelihood phylogenetic tree.

*This figure was generated before recent nomenclature changes to bacterial and archaeal phyla. For current valid names, please refer to Oren & Garrity, 2021[198].

2.5.6 Distance-decay patterns differ at local and regional scales

Environmental selection, ecological drift, diversification, and dispersal limitation all contribute to distance-decay patterns[62]. While several studies have shown microbial dispersal limitations and distance-decay patterns exist in diverse geothermal and non-geothermal environments (e.g., hot springs[6], freshwater streams[199], salt marshes[244]), 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 (slope=0.031, $P<0.001$; Figure 2.4). This finding strongly suggests dispersal limitation exists between individual geothermal fields. Increasing the resolution to within individual fields, distance-decay patterns are negligible compared to the regional scale (Table B.7). 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 (Figure B.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; Figure 2.3 position E). These differing observations can be explained 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 fields[86]. 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 survival[245] and therefore the divergence of populations in neighbouring, chemically-homogeneous spring ecosystems is plausible. Thirdly, aeolian immigration[107] 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.

2.6   SUMMARY

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 Project[20] 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 Aotearoa-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.

2.7 SUPPLEMENTARY INFORMATION

Additional methodology (Section B.1), Figures B.1-B.9 (Section B.2) and Tables B.1-B.7 (Section B.3) can be found via Supplementary Information in Appendix B accompanying this thesis.

2.8 DATA AVAILABILITY

Raw sequences have been deposited into the European Nucleotide Archive (ENA) under study accession number PRJEB24353. A query-able website for information on each individual sample is available at the URL: https://1000springs.org.nz. All code used for statistics and figures is available through the GitLab URL: https://gitlab.com/morganlab/collaboration-1000Springs/1000Springs.

CHAPTER 3

TEMPORAL DYNAMICS OF GEOTHERMAL MICROBIAL COMMUNITIES IN AOTEAROA-NEW ZEALAND

Jean F. Power, Caitlin L. Lowe, Carlo R. Carere, Ian R. McDonald, S. Craig Cary, & Matthew B. Stott

3.1   PREFACE

This chapter focuses on patterns of microbial biogeography in TVZ geothermal ecosystems across time, rather than space. It complements Chapter 2 of this thesis, which investigated bacterial and archaeal diversity patterns in the water columns of 925 individual geothermal features across a spatial range of ~8,000 km^2. Abundant taxa, including *Venenivibrio* and *Acidithiobacillus,* were found to be prevalent across the region, with pH contributing the greatest control over extremophilic diversity from the 46 physicochemical parameters measured from each spring. A weak distance-decay pattern of microbial communities was observed across all 925 springs, indicating there was some limitation to dispersal at the community level. However, stark heterogeneity in both community structure and physicochemistry was detected in some springs immediately adjacent to each other, which proposed that while dispersal within geothermal fields may not be restricted, environmental selection at the local scale appears to define community assembly in these habitats. The samples from Chapter 2, however, should be seen as snapshots of diversity, taken at a single point in time, with microbial communities in other non-extreme environments often reported as temporally dynamic. Further, we know microorganisms can respond to disturbances in the local environment, and indeed, evidence has suggested a certain degree of stochasticity exists in population abundances through time. However, due to the controlling influence of extreme physicochemistry found in geothermal springs, and the finding of Chapter 2 that pH was the main driver of microbial diversity across the broad range of physicochemistry found in TVZ ecosystems, I wanted to test whether the same controls existed for extremophiles across time, and if a substantial change in physicochemistry was required to induce community turnover.

To this end, a subset of the 925 geothermal springs from Chapter 2 were selected to sample temporally. One hundred and fifteen samples were collected from 31 geothermal features that ranged in size, source fluid physicochemistry, and initial community structure (based on the diversity snapshot taken for Chapter 2). Control sites with varying types of source fluid were chosen to monitor any stochasticity in community abundances over time (Category A). Geothermal stream sites were monitored before and after resident microbial communities were disturbed either by a natural, short-term, or an anthropogenic, long-term disruption (Category B). Finally, as research from Chapter 2 suggested pH was the principal, underlying constraint of microbial diversity in these habitats, springs were studied from across the pH range found in TVZ geothermal systems (pH 3, 5, 7, and 9) over the timeframe of one year (pH sites). Taxonomic, alpha, and beta diversity assessments, in conjunction with correlation statistics, were deployed to investigate these questions.

Samples for the control and disturbed sites (Categories A and B; sample prefix P1) were collected concurrently during field work for Chapter 2. This ensured standardised collection and processing methodologies were used to robustly compare all samples. I performed this field work and sample processing, as outlined in Chapter 2, Section 2.1, with assistance from Matthew Stott, Carlo Carere, and various staff at Te Pū Ao-GNS Science, Taupō. The pH sites (Category C; sample prefix P2), were collected and processed by both myself and Caitlin Lowe. DNA extractions and sequencing were performed by Caitlin Lowe, again at the Thermophile Research Unit, Te Whare Wānanga o Waikato-University of Waikato, using the same procedures at outlined in Chapter 2 and Appendix B. Ms. Lowe used the data from Category C samples to write her MSc thesis in 2017[246]. For this chapter, I re-analysed the raw sequences from all Category C sites, along with Categories A and B, using the custom bioinformatical pipeline I developed for Chapter 2. This ensured identical sample processing and analyses were performed for all categories to confidently compare and interpret underlying processes driving any temporal variation in TVZ geothermal springs. Additionally, I performed all statistical and ecology-based analyses presented in this chapter, and I wrote the chapter as a manuscript that has been published in the journal *Frontiers in Microbiology*[196]. Matthew Stott, Craig Cary, Ian McDonald, and Carlo Carere contributed to experimental design of this chapter and edited the text. A signed co-authorship form confirming these contributions can be found in Appendix F.

The remainder of this chapter and associated appendices are a reproduction of the manuscript text. Supplementary material to accompany this chapter is outlined in Appendix C. All scripts developed for statistics and figures are available through GitLab (https://gitlab.com/morganlab/collaboration-1000Springs/1000Springs), with raw sequences used in this chapter deposited in the European Nucleotide Archive (ENA) under study accessions PRJEB24353 and PRJEB55115.

3.2 ABSTRACT

Microbial biogeography studies, in particular for geothermal-associated habitats, have focused on spatial patterns and/or individual sites, which have limited ability to describe the dynamics of ecosystem behaviour. Here, we report the first comprehensive temporal study of bacterial and archaeal communities from an extensive range of geothermal features in Aotearoa-New Zealand. One hundred and fifteen water column samples from 31 geothermal ecosystems were taken over a 34-month period to ascertain microbial community stability (control sites), community response to both natural and anthropogenic disturbances in the local environment (disturbed sites), and temporal variation in spring diversity across different pH values (pH 3, 5, 7, and 9) all at a similar temperature of 60-70 °C (pH sites). Identical methodologies were employed to measure microbial diversity via 16S rRNA gene amplicon sequencing, along with 44 physicochemical parameters from each feature, to ensure confidence in comparing samples across timeframes. Our results indicated temperature and associated groundwater physicochemistry were the most likely parameters to vary stochastically in these geothermal features, with community abundances rather than composition more readily affected by a changing environment. However, variation in pH (pH ±1) had a more significant effect on community structure than temperature (±20 °C), with alpha diversity failing to adequately measure temporal microbial disparity in geothermal features outside of circumneutral conditions. While a substantial physicochemical disturbance was required to shift community structures at the phylum level, geothermal ecosystems were resilient at this broad taxonomic rank and returned to a pre-disturbed state if environmental conditions re-established. These findings highlight the diverse controls between different microbial communities within the same habitat-type, expanding our understanding of temporal dynamics in extreme ecosystems.

3.3 INTRODUCTION

Microbial biogeography identifies patterns and processes of diversity across space and time. While spatial studies focus on contemporary community structures and allow direct comparison between target ecosystems, this type of analysis only captures a snapshot of community diversity and fails to consider the dynamic nature of microbial assemblages[5,17]. A true description of local diversity requires both spatial and temporal examination to obtain a complete picture of habitat structure[66]. Neutral processes, such as immigration/extinction of species and population drift, have been shown to affect community abundances over time[62,106]. Physicochemical flux can also alter microbial communities and/or populations in an ecosystem[247], with some taxonomic groups acting as indicators for overall community change[15,16]. Interestingly, sampling time can affect spatial difference between microbial communities as sufficient time can be needed to build felicitous composition from migration[67,109], which reinforces the use of both spatial and temporal scales for microbial ecology studies[63,64]. This includes understanding the distinction between stochastic and deterministic change to obtain a more accurate representation of ecosystem behaviour.

Previous research into temporal dynamics in geothermal ecosystems, including microbial mats, sediments, and water columns, has presented conflicting results on the processes driving microbial community variation. Little to no disparity in microbial community structure was observed in geothermal mats across multiple temporal scales from three separate studies in Yellowstone National Park (YNP)[248–250]. A five-year study across the water column of Boiling Springs Lake in the USA again revealed a stable community despite a significant seasonal temperature cycle[174]. However, other studies have shown a seasonality effect in microbial mats, water and sediment samples from geothermal features, where seasons coincided with dichotomous weather patterns (e.g., monsoon rainfall or snow melt)[72,169,251]. Time-dependent differences to alpha diversity (i.e., richness and relative abundances), but not community dissimilarity have been reported in two separate studies[175,251]. Geothermal eukaryotic algae (order Cyanidiales) have been shown to increase in relative abundance and photosynthetic activity with sunlight during summer[252], while temperature induced community variation across seasons in microbial mats from the Patagonian Andes[133]. Wang *et al.* (2014) discussed limited temporal change across entire communities of geothermal springs in China but demonstrated certain populations were susceptible to changes in pH, temperature and dissolved organic carbon[175]. Within Aotearoa-

New Zealand, temporal studies of geothermal microbial communities are limited to just two features: Champagne Pool (estimated volume 50,000 m^3), where no changes to community structure, pH or temperature were reported in the water column over a two-year period[138]; and Inferno Crater Lake (estimated volume 18,500-65,200 m^3), where community turnover was associated with recurrent geological reservoir cycling evident by extreme changes in water level (~8 m) and temperature (~30-70 °C)[139]. There are multiple plausible reasons for the disparity in these studies; variation in community structure not recorded due to low resolution of taxa, small sample sizes, insufficient temporal scales, environmental differentiation, and/or the absence of control sites. In order to generate a consensus in the drivers of geothermal microbial variation over time, further studies are needed, preferably with consistent methodologies across multiple hot spring types, to develop a more holistic view of ecosystem dynamics.

The 1,000 Springs Project (https://1000springs.org.nz)[195] is a study investigating microbial biogeography across the Taupō Volcanic Zone (TVZ), a geothermal region in the North Island of Aotearoa-New Zealand (Figure C.1)[207]. From July 2013 to April 2015, 1,019 spring samples were collected from 18 geothermal fields and 974 individual features across the TVZ. An initial biogeography study, conducted on 925 of these springs, reported pH to be the most significant physicochemical parameter driving spatial community composition[195]. To help address the knowledge gap on temporal biogeography in TVZ geothermal features, and to determine whether the spatial study represented a simplistic characterisation of these microbial communities, 19 sites were sampled over a 34-month period to ascertain both long-term microbial community stability (i.e., control sites) and reaction to physicochemical disturbances in the local environment (i.e., disturbed sites). In addition, we sampled 12 geothermal features displaying moderate temperatures (60-70 °C) across a range of pH values (pH 3, 5, 7, or 9) every two months for a one-year period to assess drivers of community structure across the pH scale (i.e., pH sites). Our results indicate that while pH remains a significant constraining parameter of microbial diversity in geothermal features, temperature and groundwater physicochemical conditions are more likely to vary stochastically with time; resulting in a dynamic system where abundance levels of resident populations can be in a constant state of flux.

3.4   METHODS

3.4.1   Sample description & collection

All samples were collected from the water columns of geothermal features within the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand (Figure C.1), as per detailed methodologies outlined in Power *et al.* (2018)[195]. We gathered three different categories of temporal samples to define whether, and how, geothermal microbial communities changed over time, and what the primary drivers of these changes, if any, could be. The three categories were defined as follows: category A - natural microbial variation in geothermal features with both known and unknown physicochemical source fluid (control sites); category B - both natural and anthropogenic disturbances to geothermal features (disturbed sites); and category C – microbial variation over time in geothermal features with different pH values (pH sites). Category A and B samples were collected between July 2013 and October 2016 and have the sample prefix P1 (Table C.1, Table C.2, and Table C.3). Category C samples were collected between December 2015 and October 2016 and have the prefix P2 (Table C.4).

Category A (control sites) represents 13 samples from five distinct geothermal features. Three of these sites (Wairakei Features 5, 13 and 14; Wairakei-Tauhara geothermal field) were sampled twice within a five-month period on the basis of deeply-sourced hydrothermal fluid inputs presenting minimal interaction with groundwater[253]. These features are sourced from the Wairakei steam field injection pipe network[254], where surface fluid originates from a depth of at least 1,500 m[255]. A fourth spring (Radiata Pool, Ngatamariki geothermal field) was sampled twice 10 months apart, with elevated concentrations of bicarbonate (519 ppm) suggesting a longer subsurface residence time for the water column than previous control sites[256]. The fifth control site, Champagne Pool from the Waiotapu geothermal field, was selected to investigate the influence of seasonality on microbial communities within a large (~65 m diameter, ~50,000 m^3 volume), well established (ca. 900 years ago), and deeply-sourced geothermal spring[138,257]. This feature was sampled five times within a 16-month period.
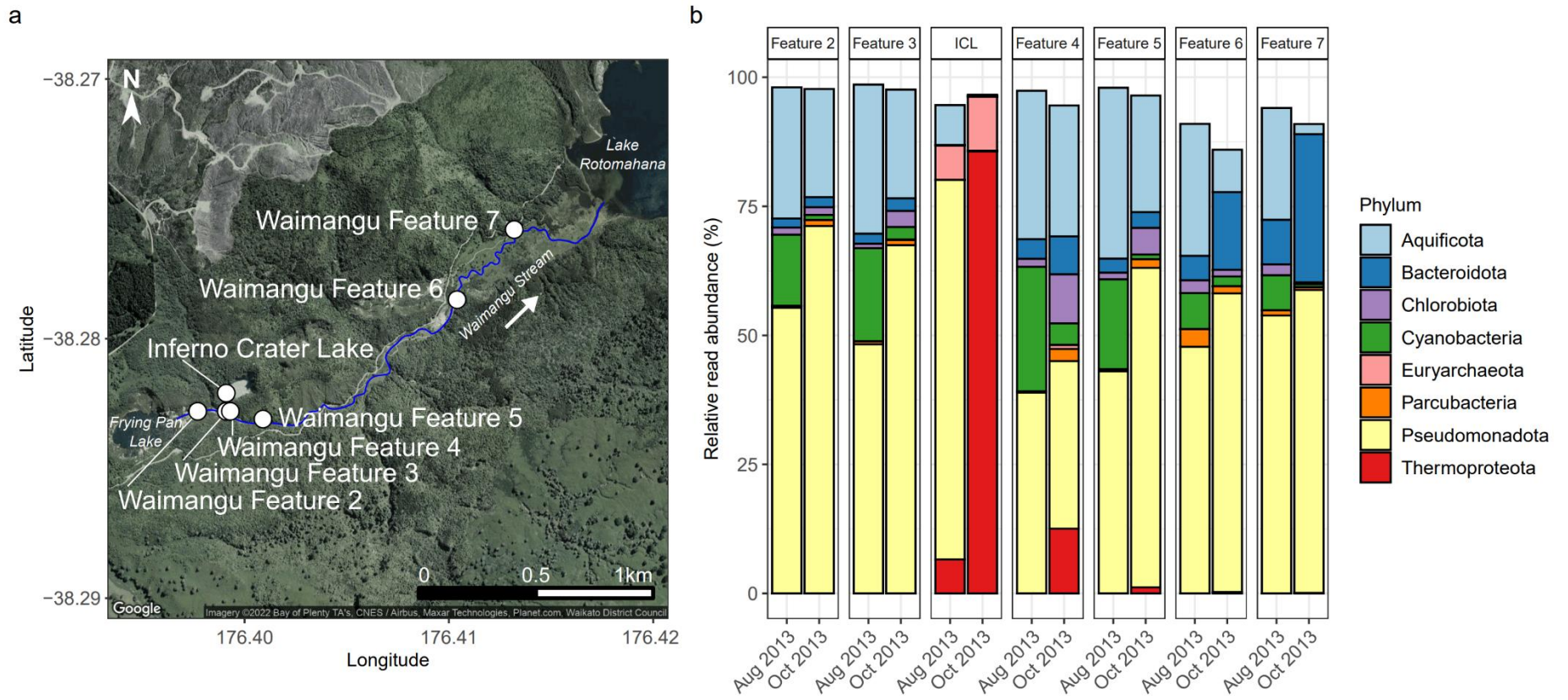
Figure 3.1 - Short-term, natural disturbance of Waimangu Stream by Inferno Crater Lake. **(a)** Waimangu Stream (outlined in blue) originates from Frying Pan Lake and flows down a rift valley to Lake Rotomahana. Sampling locations along the stream (Waimangu Features 2-7) are depicted by white circles, along with Inferno Crater Lake which overflows into Waimangu Stream every ~30-40 days. Waimangu Features 2 and 3 are upstream of the Inferno Crater Lake overflow (by ~150 and 10 m respectively), while Waimangu Features 4-7 are downstream (by ~35, 200, 1400, and 1950 m respectively). **(b)** Microbial community composition and relative abundance of Waimangu Stream features and Inferno Crater Lake (ICL) before (August 2013) and during (October 2013) the disturbance, measured by amplicon sequencing of the 16S rRNA gene. Only phyla >1 % average relative abundance across all samples in this category are shown   .

Category B samples (disturbed sites) were collected from two geothermal fields before and after a natural, short-term (i.e., pulse), or an anthropogenic, long-term (i.e., press) disturbance[258] to the resident microbial communities. The first set of sites were sampled from the Waimangu geothermal field (Figure 3.1), where Waimangu Stream (a.k.a. Hot Water Creek) is naturally disturbed every ~30-40 days by Inferno Crater Lake, a geothermal spring that cycles in both temperature (~30-70 °C) and water level (~8 m; Figure C.2)[139]. The stream is continuously sourced (~104 l s^{-1}) from Frying Pan Lake (~45 °C, pH 7, and 200,000 m^3 in volume), and experiences additional discontinuous sourcing (~79 l s^{-1}) from Inferno Crater Lake (~70 °C, pH 2, and 65,200 m^3 in volume) for 1-2 days when the spring overflows at the end of a geological cycle[259]. Six sites were sampled along the stream (two sites upstream of the Inferno Crater Lake overflow by 10 and 150 m, and four sites downstream of the overflow by 35, 200, 1400 and 1950 m), along with Inferno Crater Lake, both before (August 2013) and during overflow (October 2013), resulting in 14 samples for analysis. The second set of category B sites was collected from the Waikite geothermal field (Figure 3.2) which was undergoing a long-term wetland restoration[260]. Historically, the area was altered to create farming pasture by diverting Otamakokore Stream and draining the wetland, and rehabilitation to reverse these processes began in 2009. This included the installation of a weir on the stream in 2014, which raised the water level of the wetland by ~1.17m[260]. Seven sites across the geothermal field were sampled twice before (December 2013 and January 2014), and twice after this disturbance (April 2015 and October 2016); one site upstream from the wetland (Waikite Feature 3), the inlet to the wetland (Waikite Feature 4), two sites within the wetland (Waikite Features 5 and 6), the outlet to the wetland (Waikite Feature 8), and two geothermal springs independently situated near the wetland, colloquially known as 'Pig Scorcher' and 'Big Spring' (Figure 3.2). These two sites were sampled to check if alterations to the wetland affected other features within the geothermal field, with reduced timepoints taken for health and safety reasons. We sampled five- and 23-months post weir installation to allow microbial communities sufficient time to stabilise after the disturbance, with a total of 25 samples from Waikite used for subsequent analysis.

49
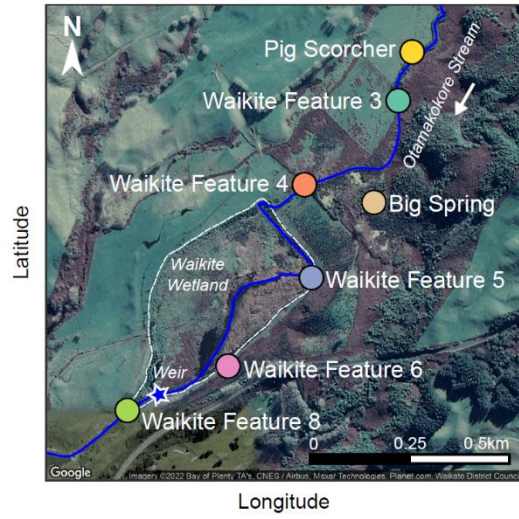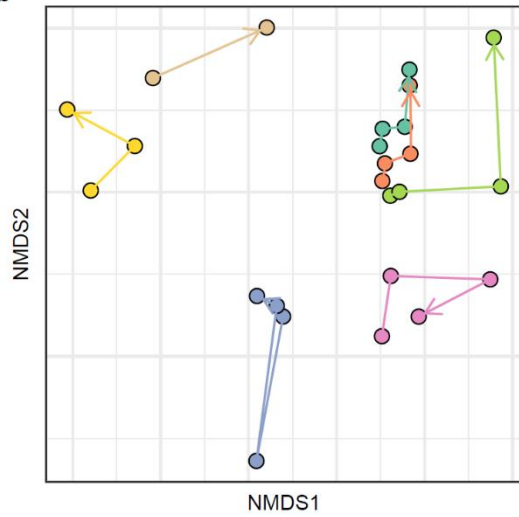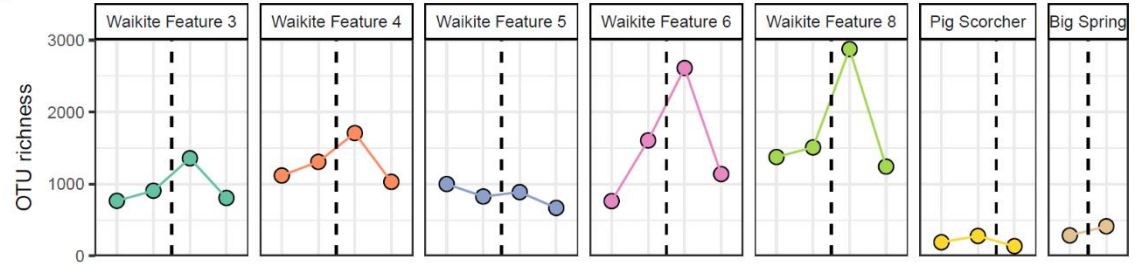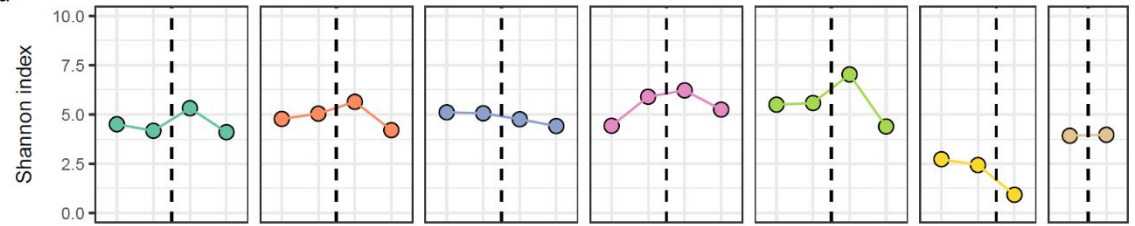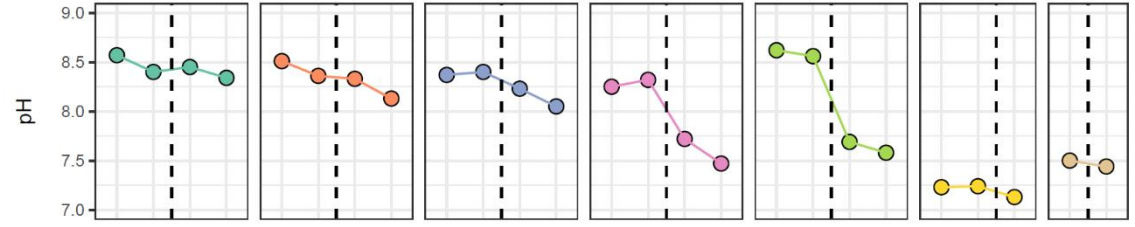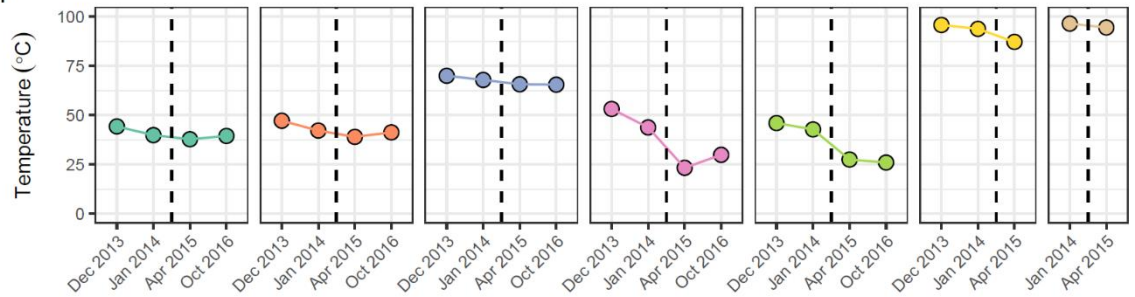
Figure 3.2 - Temporal microbial diversity and physicochemistry of the Waikite wetland during restoration. **(a)** Aerial map of the Waikite geothermal field, including the wetland (dashed white line) and Otamakokore Stream (solid blue line). The location of the weir is highlighted by the star, with stream sites and geothermal features sampled distinguished by colour. **(b)** Non-metric multidimensional scaling (NMDS; n=25, stress=0.11) of beta diversity, calculated using Bray-Curtis dissimilarities between 16S rRNA gene sequences of microbial communities in these features. Samples are coloured by feature name, with time indicated by arrows. **(c)** Variation in alpha diversity of samples measured via OTU richness (i.e., the number of OTUs per community). **(d)** Alpha diversity was also measured by the Shannon diversity index to indicate evenness of the communities over time. **(e)** The pH of features at the time of sampling. (f) Temperature variation of all sites during the wetland restoration. The black dashed line indicates the construction of a weir in the southwest corner of the wetland.

Finally, category C samples (pH sites) targeted 12 geothermal features within the same temperature range (60-70 °C) but at different pH levels (pH 3, 5, 7 and 9). Our spatial study on the microbial communities of 925 geothermal springs across the TVZ demonstrated pH as the major driver of diversity in these ecosystems[195]. Thus, to investigate whether pH had a similar effect on geothermal communities temporally, three features for each pH group were sampled from the Rotorua, Waiotapu, Atiamuri, and Ohaaki geothermal fields (Figure C.1). To minimise the influence of temperature on community structure, features were selected between 60-70 °C. This also allowed for decreased taxa numbers common to hot temperature springs to robustly identify microbial community change. The 12 sites were sampled every two months from December 2015 to October 2016, with three sites inaccessible in December 2015.

3.4.2   Microbial and physicochemical analyses

Microbial DNA extraction, sequencing, processing and the measurement of 44 physicochemical parameters from each sample were performed as described by Power *et al.* (2018)[195], unless otherwise stated. Briefly, the 16S rRNA gene was analysed using the original Earth Microbiome Project primers[20] and the Ion PGM system for next-generation sequencing. To remain consistent with the initial spatial biogeography study on these ecosystems[195], raw sequences were processed using USEARCH[211] and QIIME[212] to generate operational taxonomic units (OTUs) and respective read abundances, with taxonomy assigned using the RDP classifier[214] and the SILVA database v123[215]. Rarefaction to the lowest sequence read count that did not compromise OTU diversity (n=11,131 and 5,000 sequencing reads for categories A and B, and category C, respectively) was applied to all samples. Conductivity (COND), oxidation-reduction potential

(ORP), and pH were determined using a Hanna Instruments multiparameter field meter (Woonsocket, RI, USA), with spring temperature (TEMP) measured by a Fluke 51-II thermocouple (Everett, WA, USA). Inductively coupled plasma-mass spectrometry (ICP-MS), UV-vis spectrometry, titration, and ion chromatography were used to measure the following aqueous metals and non-metals: aluminum (Al), ammonium (NH_4^+), arsenic (As), barium (Ba), bicarbonate (HCO_3^-), boron (B), bromine (Br), cadmium (Cd), caesium (Cs), calcium (Ca), chloride (Cl^-), chromium (Cr), cobalt (Co), copper (Cu), ferrous iron (Fe^{2+}), hydrogen sulfide (H_2S), iron (Fe), lead (Pb), lithium (Li), magnesium (Mg), manganese (Mn), mercury (Hg), molybdate (Mo), nickel (Ni), nitrate (NO_3^-), nitrite (NO_2^-), phosphate (PO_4^{3-}), potassium (K), rubidium (Rb), selenium (Se), silicon (Si), silver (Ag), sodium (Na), strontium (Sr), sulfate (SO_4^{2-}), sulfur (S), thallium (Tl), uranium (U), vanadium (V), and zinc (Zn). In addition, six variables were analysed for category C samples. These included rainfall data (December 2015 to October 2016) obtained from the CliFlo National Climate Database, which is maintained by the National Institute of Water and Atmospheric Research (NIWA) in Aotearoa-New Zealand[261]. Total daily depth (mm) was recorded from 00:00 to 23:59 New Zealand Standard Time (NZST) of each sampling date at the Whakarewarewa, Rotorua weather station (EK577135). The average daily water level depth (m) of five geothermal monitoring bores (M24-28) from the Rotorua geothermal field was also retrieved through the Environmental Data Portal managed by the Bay of Plenty regional council[262].

3.4.3   Statistical analyses

All statistical analyses and figures were produced using R v4.0.3[216] and the phyloseq package v1.32.0[217], unless otherwise stated. Operational taxonomic units (OTUs) generated from processing of the 16S rRNA gene sequences, and associated sample metadata, were imported into R using the import_biom function. The dataset was then divided into either category A, B, or C using the subset_samples function. To assess variation in the read abundances of dominant taxa, OTUs in each sample were agglomerated to corresponding phylum (or genus) using the tax_glom function (with the argument NArm=FALSE), with reads transformed to percentage relative abundance using the transform_sample_counts function. Only microbial taxa with an average relative abundance of >1 % were retained to allow robust identification of fluctuations to dominant community members (using the filter_taxa function), with plot_bar implemented to

visualise these compositions over time. Using original OTUs (prior to phylum agglomeration), alpha diversity via OTU richness (i.e., number of OTUs) and evenness (i.e., Shannon diversity index) was calculated using the estimate_richness function. The package ggplot2 v3.3.2[219] was used to plot alpha diversity and physicochemical parameters (e.g., pH and temperature) over time. Beta diversity between spring communities was generated via Bray-Curtis dissimilarities using the vegdist function (with the 'bray' method) from the vegan package v2.5-6[218]. Non-metric multidimensional scaling (NMDS) was implemented to visualise these dissimilarities via the metaMDS function (k=2), also from the vegan package, in conjunction with ggplot2. Correlation testing between alpha diversity and spring physicochemistry was performed using the dist function with the 'Euclidean' dissimilarity index, and the cor.test function with Spearman's correlation coefficient (ρ) from the base R stats package. For beta diversity, Bray-Curtis dissimilarities were checked for correlation with spring physicochemistry using the mantel function from vegan, again with Spearman's correlation coefficient (ρ). Correlation heatmaps were plotted using the geom_tile function in ggplot2. Analysis of similarities (ANOSIM; vegan package) was performed to compare beta diversity between pH groups for category C samples. Satellite maps of both Waimangu and Waikite geothermal fields (category B sites) were generated from Google using the ggmap package v3.0.0[263], with latitude and longitude coordinates (WGS84) of geothermal features added via ggplot2. The topographic layers for the TVZ map were obtained from Toitū Te Whenua-Land Information New Zealand (LINZ; CC-BY-4.0), and the TVZ boundary was defined using data from Wilson *et al*. (1995)[207].
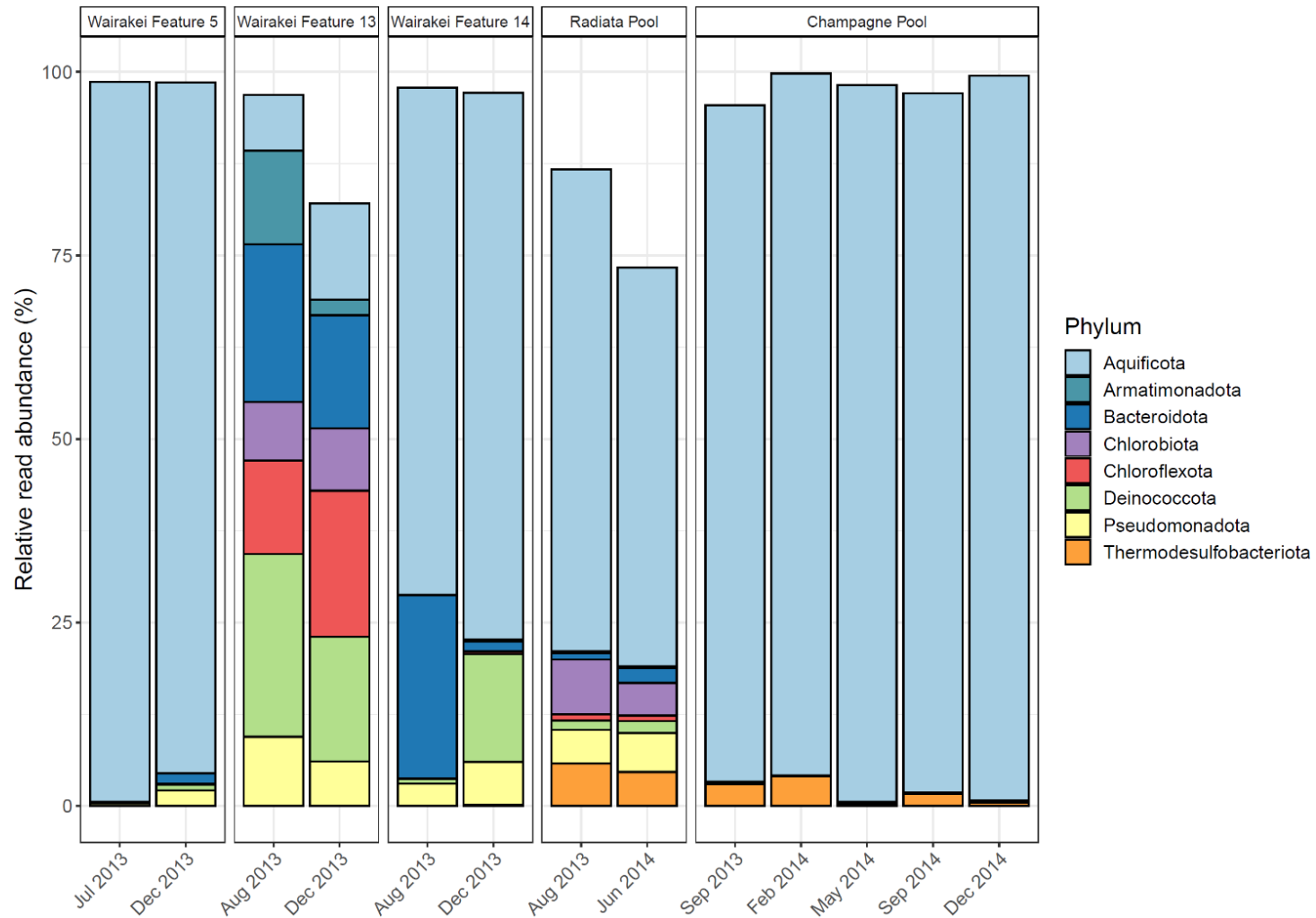
Figure 3.3 - Temporal microbial community composition and relative abundance of control features in the Taupō Volcanic Zone (TVZ). Amplicon sequencing of the 16S rRNA gene was used to measure read abundance of taxa in spring communities, with only phyla >1 % average relative abundance across all samples in this category shown. Control features were sampled from the Wairakei-Tauhara (Wairakei Features 5, 13, and 14), Ngatamariki (Radiata Pool), and Waiotapu (Champagne Pool) geothermal fields .

3.5   RESULTS

3.5.1   Relative read abundances of taxa fluctuated in some control geothermal features over time

Microbial community composition remained broadly consistent in control features throughout the duration of this study, but relative read abundances of these dominant taxa were variable in some sites (Figure 3.3; Table C.5). Notable abundance changes in geothermal features from the Wairakei field included increased Chloroflexota coupled with decreased Armatimonadota and Deinococcota in Feature 13, and increased Deinococcota with decreased Bacteroidota in Feature 14. These changes corresponded with temporal variation in community richness (i.e., the number of OTUs) and evenness (Shannon diversity index), while both pH and temperature remained stable for all features (Figure C.3; Table C.5). Similarly, no significant change was observed in microbial community taxa and physicochemistry of Radiata Pool, Ngatamariki, with Aquificota displaying the greatest variation to relative abundance (SD=8.0 %). Community structure in Champagne Pool, Waiotapu, was dominated by Aquificota in all samples, with OTU richness, evenness, pH, and temperature remaining relatively constant over all seasons throughout the 15-month sampling timeframe (Figure C.3; Table C.5). Additionally, no significant correlations (p<0.05, Spearman's correlation coefficient) were observed between either alpha diversity metric with pH and temperature. Beta diversity analysis for all control features grouped samples by their respective geothermal spring (Figure C.3), indicating decreased temporal variation at a community level for these sites.

3.5.2   Alpha and beta diversity remained consistent despite substantial short-term physicochemical disturbance

Evidence of a pulse disturbance in the Waimangu geothermal field was observed in microbial communities of Waimangu Stream downstream from Inferno Crater Lake overflow, with variation in relative abundances of resident taxa also occurring irrespective to this disturbance (Figure 3.1; Table C.6). The microbial community of Inferno Crater Lake drastically changed between timepoints, indicated by an almost complete community turnover from Pseudomonadota (72.4 % of the total spring community assigned to *Acidithiobacillus* sp.) to Thermoproteota (85.6 % of the total spring community assigned to Sulfolobaceae), which corresponded with an

associated temperature change of 41.0 to 69.1 °C for this feature during overflow (Figure C.4; Table C.6). These Thermoproteota signatures were subsequently evidenced in downstream sites of the overflow, decreasing in relative abundance from 12.5 % of the total microbial community in Waimangu Feature 4 to 0.1 % in Waimangu Feature 7. Other variable phyla across Waimangu Stream included increases in Pseudomonadota (not *Acidithiobacillus* sp.) in all sites except Feature 4, and decreases in Cyanobacteria in all sites. Features 6 and 7, which were near the stream outflow into Lake Rotomahana, also had increases in Bacteroidota, which corresponded with decreases in obligately thermophilic Aquificota. The effect of Inferno Crater Lake overflow was apparent in the physicochemistry of features immediately downstream of the overflow entry point (Figure C.4; Table C.6), where pH decreased from 6.7 to 3.0 in Feature 4, and 7.4 to 3.0 in Feature 5. Temperature fluctuation in downstream sites was less apparent, with the greatest temperature increase due to overflow (+6.7 °C) occurring in Waimangu Feature 4. Interestingly, alpha diversity analyses did not reflect the physicochemical results, with the greatest changes observed in OTU richness of Waimangu Feature 4 and Feature 6 (Figure C.4; Table C.6). Despite a complete shift in microbial community structure in Inferno Crater Lake between the two timepoints, both alpha diversity measures for this ecosystem had negligible variation, with an associated temperature difference of 28.1 °C (Figure C.4; Table C.6). The greatest dissimilarity in beta diversity between the temporal stream samples was found in Features 6 and 7 (Figure C.4). Even with the pH and temperature changes in Features 4 and 5 post-disturbance, beta diversity did not show variance between the microbial communities of these features over time.

3.5.3 Diversity metrics showed initial response to long-term anthropogenic disturbance

In contrast to the short-term, natural disturbance observed at Waimangu, variations in alpha and beta diversity of microbial communities did correspond with a long-term, anthropogenic disturbance at the Waikite wetland, in particular for sites located immediately before and after the installation of a weir (Features 6 and 8 respectively; Figure 3.2; Table C.7). In all wetland sites, OTU richness and evenness then returned to pre-disturbance levels for the final sampling timepoint (23 months post-weir installation). Beta diversity showed an initial response to the disturbance for all sites in or near Otamakokore Stream, which was then amplified for the final timepoint of Features 3, 4 and 8. Features with an additional geothermal input (Feature 5 and

'Pig Scorcher') demonstrated variability in beta diversity irrespective of the disturbance, which was also apparent in the relative abundances of community structure in these sites (Figure 3.4; Table C.7). For example, the 'Pig Scorcher' geothermal spring had a notable decrease in Deinococcota before the disturbance was introduced to the wetland, which coincided with an increase in Aquificota. Additionally, the microbial community of the 'Big Spring' exhibited fluctuations in relative phyla abundances before and after the weir installation. However, as only two timepoints were sampled for this site, the question of whether these changes were associated with the disturbance remain unresolved. While the resident taxa of microbial communities upstream to the wetland (Features 3 and 4) remained constant over time, increases in the relative abundances of Pseudomonadota, coupled with decreases in Cyanobacteria and Bacteroidota, did correspond with the weir installation (Figure 3.4). Similarly, the relative abundances of Pseudomonadota increased, while Bacteroidota and Verrucomicrobiota decreased in Waikite Feature 6, a site within the wetland, although compositional changes not related to the disturbance were also noted in the microbial community of this feature. The outlet to the wetland (Feature 8) showed increases in Parcubacteria and Chlamydiota at the first timepoint post-disturbance, but these reduced to similar levels as the first timepoint by the end of the experiment. Overall, Pseudomonadota increased and Bacteroidota decreased at the wetland outlet across the three years analysed. Regarding physicochemistry, all features in and around the wetland decreased in both pH and temperature from December 2013 to October 2016 (Figure 3.2; Table C.7). The majority of these changes were minor (pH: $SD \leq 0.2$; temperature: $SD \leq 4.5$ °C), except for Waikite Features 6 and 8 which had standard deviations of pH 0.4 and 13.5 °C, and pH 0.6 and 10.3 °C respectively. Despite these variations, no significant correlations ($p<0.05$, Spearman's coefficients) were observed between diversity and temperature for all features, with only beta diversity of Waikite Features 4 and 8 producing meaningful correlations with pH (Figure C.5).
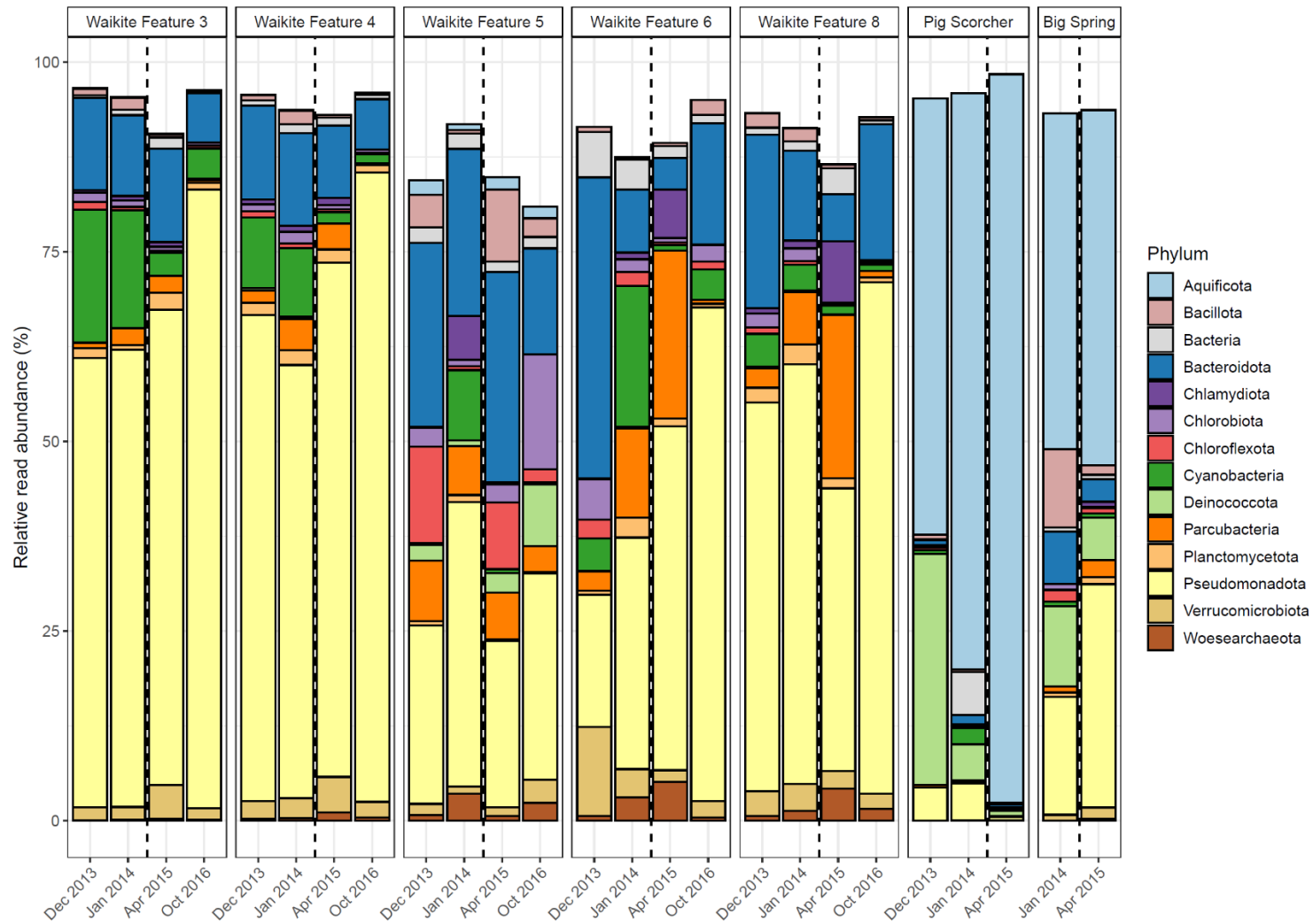
Figure 3.4 - Temporal microbial community composition and relative abundance of Waikite geothermal features. Amplicon sequencing of the 16S rRNA gene was used to measure read abundance of taxa in spring communities, with only phyla >1 % average relative abundance across all samples in this category shown. The vertical dashed line indicates the construction of a weir near the outlet of the wetland during restoration.

3.5.4 Temperature is more variable than pH over time in geothermal features

Similar to control and disturbed samples, temporal variations were observed in the relative abundances of geothermal microbial communities across a range of pH values (pH 3 to 9; Figure 3.5; Table C.8). Six of the original 69 samples did not yield sufficient sequence reads (<5,000) after rarefaction, resulting in 63 samples for final community analysis. The greatest changes to relative abundances occurred in two features from the pH 3 group (Whakarewarewa Features 51 and 53), and two features from the pH 9 group (Whakarewarewa Feature 1 and Ohaaki Feature 2). Aquificota taxa were involved in all these changes, with relative abundance of the phylum increasing in Whakarewarewa Feature 51, but decreasing in Whakarewarewa Feature 53 over the timeline of the experiment. These variations were coupled with decreasing Thermoproteota (family Sulfolobaceae; Figure C.6) and increasing Pseudomonadota respectively, while the balance of Aquificota in the two pH 9 features were inversely proportional to Deinococcota taxa. Aquificota was also the most abundant phylum in five of six features from the pH 5 and 7 groups (as genus *Venenivibrio*; Figure C.6), with notable variation to this phylum only occurring in Kuirau Park Feature 60. Whangapoa Feature 1 (pH 7 group) was the only feature across all 12 sites in this category to present change independent of Aquificota; here, relative abundances of Armatimonadota, Bacteroidota, and Deinococcota fluctuated over the 10-months analysed.

Alpha and beta diversity metrics for all 12 features in this category produced contradictory results (Figure 3.6; Table C.8), with the greatest variation in alpha diversity presented in the pH 7 group features. Conversely, beta diversity suggested geothermal features from pH 3, 5 and 9 groups were more likely to have varied community structure over time, and was more representative of the variation in relative abundances than alpha diversity. pH remained relatively stable over time for the majority of features, with only Whakarewarewa Feature 51 (pH 3 group) having a standard deviation >0.3 pH units (Figure 3.6; Table C.8). Analysis of similarities confirmed that variation in beta diversity was greater between pH groups than within pH groups (p=0.001, ANOSIM; Figure C.7), indicating the significant influence of pH on driving diversity in these ecosystems on a spatial scale. However, temperature exhibited a greater temporal variability than pH, with only four features having a standard deviation of ≤ 2 °C (Figure 3.6; Table C.8). Whakarewarewa Feature 1 (pH 9 group) had the greatest fluctuation in temperature overall, changing from 63.4 to 80.6 °C, over the sampling period. The effect of

temperature on these ecosystems was highlighted by six features producing significant positive correlations between temperature and beta diversity ($p<0.05$, Spearman's coefficient; Figure C.8). The physicochemical relationship with alpha diversity was not as conclusive, with only four features demonstrating significant correlation between temperature and either OTU richness or evenness ($p\leq0.03$, Spearman's coefficient; Figure C.8). Corresponding with the pH stability we observed in these geothermal features, Whakarewarewa Feature 51 (group pH 3) was the only feature to produce a positive correlation between pH with both alpha and beta diversity metrics (OTU richness, $\rho=0.70$, $p=0.03$; Bray-Curtis dissimilarities, $\rho=0.84$, $p=0.01$; Spearman's correlation coefficient).
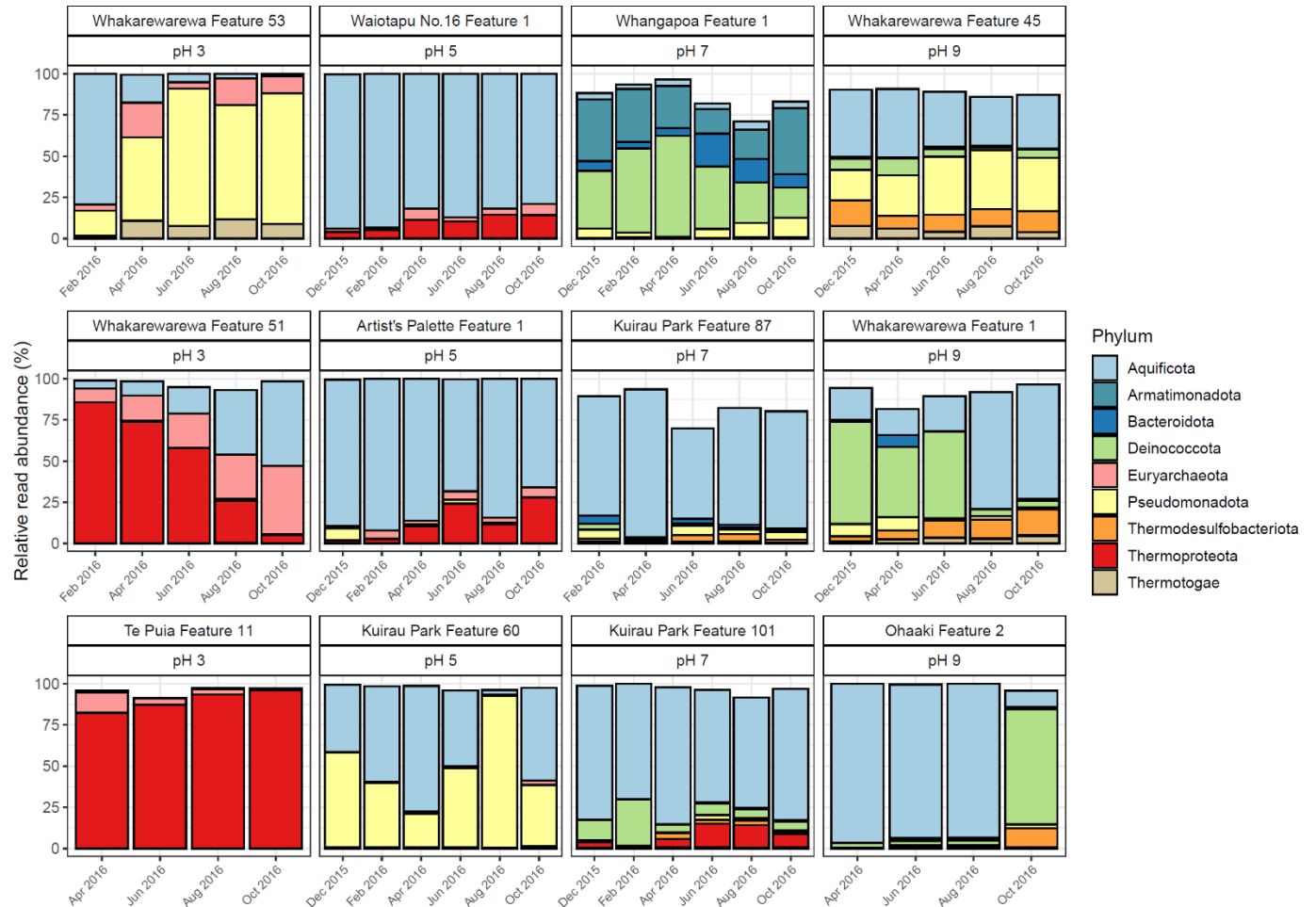
Figure 3.5 - Temporal microbial community composition and relative abundance of geothermal features across the pH scale. Amplicon sequencing of the 16S rRNA gene was used to measure read abundance of taxa in spring communities, with only phyla >1 % average relative abundance across all samples in this category shown. Geothermal features are grouped according to their respective pH group (pH 3, 5, 7, and 9).
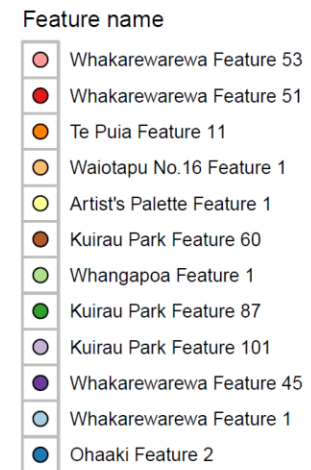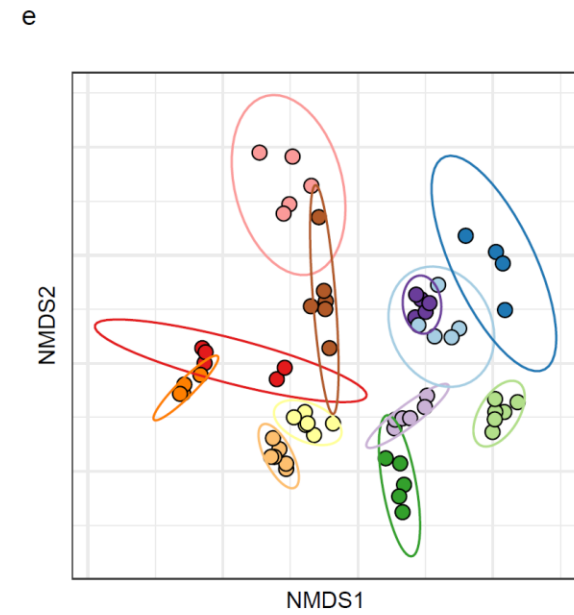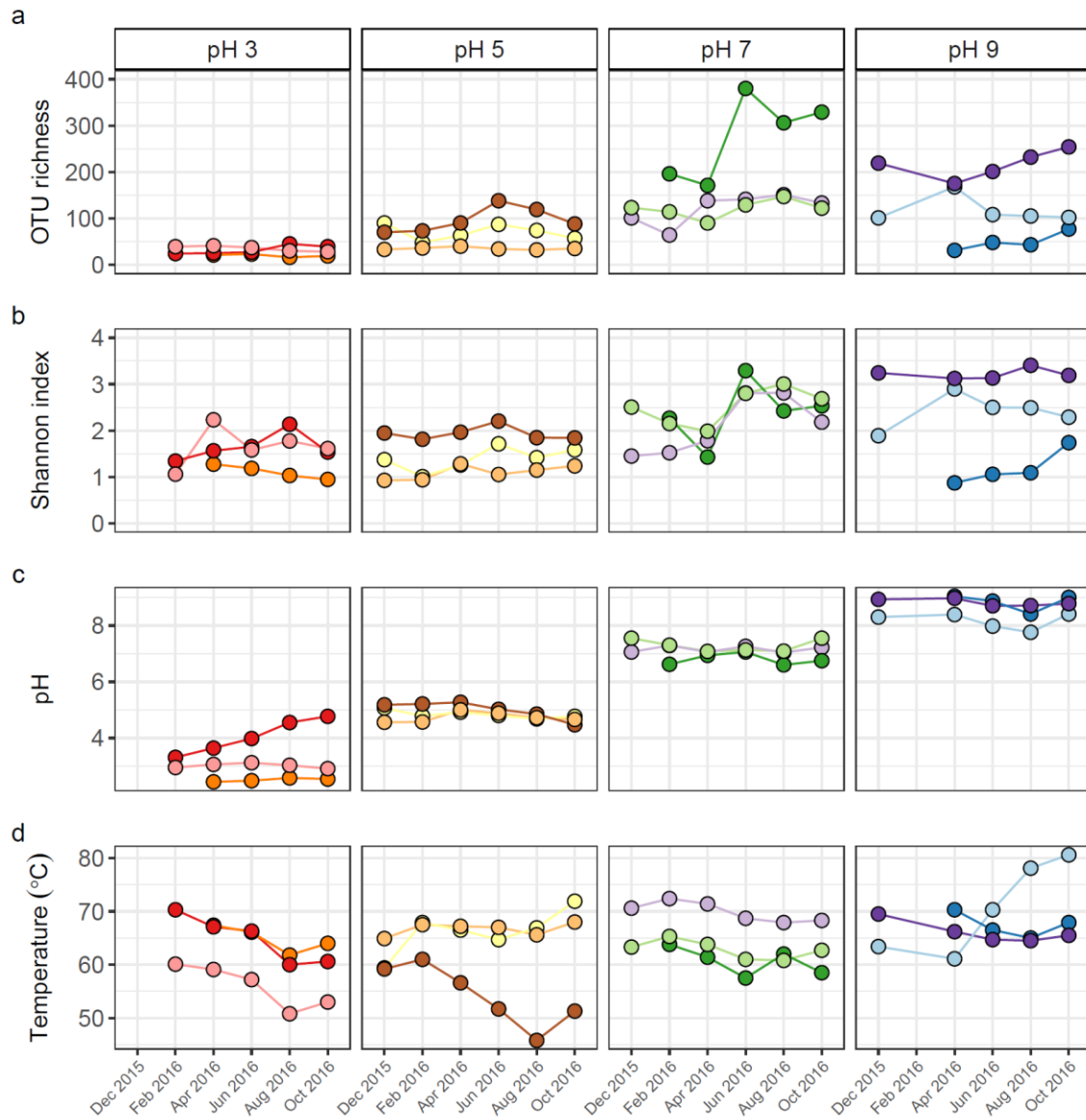
Figure 3.6 - Temporal diversity and physicochemistry of geothermal features across the pH scale. **(a)** The microbial communities of 12 geothermal springs were measured six times over one year by amplicon sequencing of the 16S rRNA gene, with variation in alpha diversity between temporal samples indicated by OTU richness (i.e., the number of OTUs per community). **(b)** Alpha diversity was also measured by the Shannon diversity index to indicate evenness of the communities over time. **(c)** Spring pH of the features at the time of sampling. **(d)** Spring temperature of the geothermal features. **(e)** Beta diversity of the 12 features is shown by a non-metric multidimensional scaling (NMDS; n=62, stress=0.15) of Bray-Curtis dissimilarities between all temporal samples, with data ellipses generated from multivariate t-distribution with a 95% confidence level. All samples are coloured according to feature name, with alpha diversity and physicochemistry further grouped by pH group.

3.6 DISCUSSION

Microbial community analysis of 31 geothermal features from the Taupō Volcanic Zone, Aotearoa-New Zealand, suggests abundances of taxa can vary stochastically over time due to a complex confluence of both apparent and cryptic parameters, with a sustained physicochemical change necessary to provoke extensive community turnover. Conversely, the composition of microbial taxa did not change in the five features sampled temporally as control sites (Figure 3.3), with physicochemistry also remaining stable (Figure C.3). Of these control sites, Wairakei Feature 5 and Champagne Pool displayed the least variation in relative abundances. These sites are both replenished via deeply-sourced hydrothermal fluids that exhibit limited interaction with groundwater[255,257]. The rate of physicochemical change for these fluids would be much slower than other inputs to shallower (i.e., acidic) geothermal springs, such as meteoric and/or groundwater, which are more readily affected by the level of the water table and atmospheric events[86,169]. Thus, the microbial communities of geothermal features with deeply sourced hydrothermal fluids are sustained by a more constant environmental niche, which consequently promotes a stable ecosystem structure. The physical structure of a geothermal feature is an important attribute to note, as it has been proposed that smaller springs are more susceptible to environmental change[86]. Even though diel variation to the physicochemistry of Champagne Pool was suggested to be a consequence of microbial activity[264,265], our findings indicate that the microbial communities of this large geothermal spring are unaffected by seasonality and weather events. These results are consistent with the inferred dogma for non-extreme environments; espousing a stable environment encourages low turnover of dominant taxa in resident microbial communities[247], with a disturbance that alters major physicochemical conditions required to shift community structures at the phylum level.

While analysis of microbial community response to both natural and anthropogenic disturbances corroborated fluctuation in abundances of taxa irrespective of measured environmental change, variation to both taxa composition and relative abundance were noted in geothermal features in the immediate vicinity of perturbations. The influence of Inferno Crater Lake, as evidenced by the dominant population of Thermoproteota (family Sulfolobaceae) and physicochemical conditions during overflow (Figure 3.1), was clearly apparent in Waimangu Stream sites downstream of the pulse disturbance. The pH of the stream was more readily affected than temperature during this coalescence (pH 6.7 to 3.0, and 49.9 to 56.6 °C; Waimangu Feature 4; Figure C.4), indicating minimal buffering capacity of the stream water; with the volumetric rate of Inferno Crater Lake overflow (~79 l s^{-1}) only marginally increasing the temperature of the stream. However, these physicochemical changes to Waimangu Stream are not conserved as the Inferno Crater Lake overflow is short-lived (2-3 days)[139]. With microbial communities in the stream experiencing this disturbance every 30-40 days, we now know resident populations quickly revert back to pre-disturbed states; the previous overflow event to this study ended eight days prior to sampling (Figure C.2). Waimangu Stream (~104 l s^{-1}) provides a constant reservoir of microbial populations from Frying Pan Lake that can quickly re-colonise downstream sites once the disturbance has ended, thereby providing a short residence time for the disturbed communities[266]. Frying Pan Lake also undergoes reservoir cycling, albeit less extreme than Inferno Crater Lake[259], which could explain the alternating abundances of both Cyanobacteria and Pseudomonadota observed in all stream sites. Similar to reports on freshwater microbial assemblages[267,268], these findings suggest geothermal stream communities are resilient and can return to a pre-disturbance state if physicochemical conditions stabilise, and sufficient re-colonisation is attainable. However, microbial variation in Waimangu Stream and Inferno Crater Lake were not clear in alpha and beta diversity measures, with the greatest temporal dissimilarity in beta diversity occurring in the two sites furthest away from the disturbance (Figure C.4). These features could be influenced by other geothermal springs not associated with upstream samples, and/or proximity to the stream outflow into Lake Rotomahana likely facilitates lacustrine water inputs to these ecosystems, supported by an observed increase in Bacteroidota which are commonly found in freshwater lake epilimnia[269]. Nevertheless, even though phylum-level changes occurred in stream communities immediately downstream of the coalescence disturbance point, these short-term variances were not reflected in diversity analyses.

Conversely, the long-term disturbance at the Waikite geothermal field was most evident in the beta diversity of temporal microbial communities at the outlet to the wetland, which corresponded with the greatest variation in physicochemistry of all sites studied in this area (Figure 3.2). Even though there was a temperature decrease of 20 °C at the outlet, diversity only correlated with a decrease in pH from 8.6 to 7.6, indicating that a larger magnitude of change in temperature may be required than pH to influence microbial populations in geothermal ecosystems. Interestingly, alpha diversity of all stream sites around the wetland showed increases in diversity in the first timepoint post-disturbance, but these had returned to pre-disturbance levels by the final timepoint (Figure 3.2). These observations advocate the importance of using multiple diversity metrics to investigate microbial communities over time so that a cohesive picture of ecosystem behaviour is obtained[270]. Both relative abundances and beta diversity of microbial communities within (Features 5 and 6) and outside the wetland ('Pig Scorcher' and 'Big Spring') fluctuated irrespectively of the perturbation (Figure 3.4), with no correlation to pH or temperature (Figure C.5). Features 5 and 6 have additional geothermal inputs to Otamakokore Stream, and are relatively low volume features, suggesting they are more susceptible to fluctuation in environmental conditions[86]. A notable temperature difference (93.8 to 87.2 °C) was observed for the 'Pig Scorcher' spring after the weir installation, however, this physicochemical change was not attributed to wetland dynamics[260]. It should be noted that both sampling timepoints after the disturbance occurred outside of the austral summer, which could partially explain the decrease observed in Cyanobacteria at all wetland sites. Even though limited Cyanobacterial change was reported with or without ultraviolet radiation in geothermal microbial mats[249], seasonal variation was detected in other geothermally-associated Cyanobacteria populations between dry and rainy seasons[251]. Greater temporal variation can occur in the microbial communities of geothermal sediments than associated water columns[175], and with shorter residence times for taxa in stream sites than other geothermal features, physical processes enabling community assembly through time would conceivably alternate across sample types. Overall, these findings suggest the abundances of planktonic microbial communities acclimatise to sustained physicochemical change in the local environment, but there is a general instability or hysteresis in geothermal habitats, even when physicochemistry appears relatively stable.

Similar to the previous categories of geothermal features, relative abundances of microbial communities from all pH groups varied over the one-year sampling timeframe (Figure 3.5),

with temperature fluctuating more than pH in these habitats (Figure 3.6). The pH 7 group had the most change in alpha diversity, particularly for Shannon diversity index (Figure 3.6), which could be a result of the increased number of taxa that prefer circumneutral conditions[153]. In contrast, the greatest changes in beta diversity were detected in pH 3, 5 and 9 groups (Figure 3.6), which corresponded with the variation observed in relative abundances of taxa in these features. Like microbial community analysis of Inferno Crater Lake, these results suggest alpha diversity fails to adequately measure temporal variation outside of circumneutral pH. This could be due in part to reduced diversity potential to fill any voids created by changing communities at either ends of the pH spectrum for geothermal springs, with the magnitude change that occurs with varying pH outside of circumneutral being more impactful. Previous longitudinal research has suggested alpha diversity can fail to present change despite highly divergent microbial communities[270], again reinforcing the use of multiple diversity metrics when investigating temporal scales. Whakarewarewa Feature 51 (pH 3 group) was the only feature that significantly correlated with pH for both alpha and beta diversity metrics (Figure C.8), with a corresponding pH increase of 3.3 to 4.8 observed. This feature also had a positive correlation between diversity and temperature, along with six other features from across the pH range (Figure C.8), suggesting temperature is more likely to fluctuate over time than pH in geothermal features. At least eight of the features in this category are shallow (<2 m in depth), and so are more susceptible to meteoric and groundwater fluctuations that occur seasonally in Aotearoa-New Zealand[271].

Comparable to the microbial community of Inferno Crater Lake, Thermoproteota populations (family Sulfolobaceae) decreased in Whakarewarewa Feature 51 as the temperature decreased and pH increased over the study period (Figure 3.6 & Figure C.6). This population decrease, characteristic to Sulfolobaceae when pH increases[272], was counteracted with increased Aquificota, and indeed, Aquificota featured prominently in geothermal community fluctuations from across the pH scale. This could be a result of the widespread abundance of this phylum found throughout the TVZ[195], but temporal changes to Aquificota populations have also been observed in other geothermal systems worldwide[72,139,169], suggesting hypersensitivity in this phylum to physicochemical change. Interestingly, Ohaaki Feature 2 (pH 9 group) had an almost complete community shift from Aquificota to Deinococcota (Figure 3.5), which correlated with both groundwater and hydrogen sulfide levels. Hydrogen sulfide gas frequently percolates through geothermally-heated soil in the TVZ[273], and consequently is common to geothermal springs that are influenced by groundwater.

Therefore, this observed shift indicates water table levels directly affect the microbiology of this spring. We also know that historic modification, stemming from local geothermal power generation, influenced both temperature and water levels of the spring[274], so anthropogenic impacts could explain the community variation observed in this study. Regardless, geothermal features from a range of source fluids across the pH scale exhibited both stochastic and deterministic fluctuation in the relative abundances of microbial communities, proposing that all geothermal features do not adhere to the same processes of community assembly through time.

3.7 SUMMARY

Our study of 31 individual features indicates that taxa abundances fluctuate temporally in geothermal microbial communities: with stochastic or unexplained variation still occurring across a wide pH range. Temperature is more variable than pH over time in these habitats, however, community assemblages appear more sensitive to changes in pH than temperature. Geothermal features with a deeply-sourced fluid input are less susceptible to physicochemical change and consequently have a more stable microbial community; suggesting source fluid (either hydrothermal, meteoric or a mixture of both) as a main driver of temporal variation. Our unique dataset also infers geothermal ecosystems can recover from short-term physicochemical disturbances, but how we define change is pertinent as multiple diversity metrics produced disparate results over the breadth of feature categories analysed, implying these measures are not always representative of change.

There are limitations to this study that must be considered, such as focusing on large shifts in the relative abundance of phyla. Even though a recent report indicated dominant taxa are more susceptible to a changing environment than rare taxa[121], it is likely that environmental changes actively select for finer-scale phenotypic or metabolic traits not detectable via the taxonomic analyses undertaken in this study[275]. Focusing on genome-wide analysis could also infer other factors beyond physicochemistry than induce temporal change in thermophilic taxa[86]. Additionally, the reliance on diversity metrics[80] and the use of relative over absolute abundances[276] as a proxy for designating change should be taken into account. While the effect of traditional biases, such as primer design, DNA extraction, and sequencing[232,277–279], were minimised by the use of identical methodologies for processing of

all samples, our study highlights the importance of using both replicates and increased sampling frequency in order to obtain a true consensus on core geothermal microbial communities.

3.8 SUPPLEMENTARY INFORMATION

Figures C.1-C.8 (Section C.1) and Tables C.1-C.8 (Section C.2) can be found via the Supplementary Information in Appendix C accompanying this thesis.

3.9 DATA AVAILABILITY

Raw amplicon sequences have been deposited into the European Nucleotide Archive (ENA) under study accession numbers PRJEB55115 and PRJEB24353, with a full list of individual sample accessions found in Tables C.1-C.4 (Appendix C.2). All code used for statistics and figures is available through the 1,000 Springs Project GitLab URL: https://gitlab.com/morganlab/collaboration-1000Springs/1000Springs.

CHAPTER 4

ALLOPATRIC SPECIATION IN THE BACTERIAL PHYLUM AQUIFICOTA ENABLES GENUS-LEVEL ENDEMISM IN AOTEAROA-NEW ZEALAND

Jean F. Power, Carlo R. Carere, Holly E. Welford, Daniel T. Hudson, Kevin C. Lee, John W. Moreau, Thijs J. G. Ettema, Anna-Louise Reysenbach, Charles K. Lee, Daniel R. Colman, Eric S. Boyd, Xochitl C. Morgan, Ian R. McDonald, S. Craig Cary, & Matthew B. Stott

4.1   PREFACE

The final research chapter of this thesis establishes the concept of endemism for a bacterial genus in the Aotearoa-New Zealand archipelago. This outcome was not anticipated when hypotheses and objectives were being developed for the study and was first indicated through Chapter 2 analyses involving the microbial biogeography of 925 geothermal spring ecosystems in the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand. The most abundant and widespread genus across all geothermal features sampled, with physicochemical ranges of 13.9-100.6 °C and pH <1-9.7, was *Venenivibrio*, from the family Hydrogenothermaceae within the phylum Aquificota. This taxon, designated *Venenivibrio stagnispumantis* CP.B2[T], was first isolated in 2007 from Champagne Pool in the Waiotapu geothermal field, TVZ and was validly characterised as a novel genus within the Hydrogenothermaceae alongside sister genera *Sulfurihydrogenibium*, *Persephonella*, and *Hydrogenothermus*. Surprisingly, no trace of *Venenivibrio* could be found outside of Aotearoa-New Zealand, either through characterisation of additional species or an initial search of DNA sequence databases. This apparent restriction to Aotearoa-New Zealand, in conjunction with widespread abundance in TVZ geothermal features, identified *Venenivibrio* as the ideal population to explore the third hypothesis that individual population diversity and distribution can be controlled by different processes to whole microbial communities.

Subsets of both Aquificota- and *Venenivibrio*-assigned 16S rRNA gene sequencing reads, and associated metadata, were extracted from the dataset produced in Chapter 2. Biogeographical patterns and underlying processes were investigated for both subsets to

assess taxa controls at phylum and genus level, but also to ascertain the degree of conformity between drivers of community and population diversity in TVZ geothermal ecosystems. This was achieved using similar analytical tools as previous chapters, such as linear regression modelling and correlation statistics, to allow direct comparison between chapters. Both alpha and beta diversity were assessed, with a concentrated focus on the effect of pH and temperature driving *Venenivibrio* abundance and distribution across the locality. The genome of CP.B2^T was sequenced, assembled, and annotated to confirm a chemolithoautotrophic lifestyle and whether any genotypic evidence could infer the taxon's apparent endemic behaviour. Protein-coding genes, intrinsic to core metabolism, were compared to other Hydrogenothermaceae genomes to search for nonconformity between putative physiologies. Nine amplicon and metagenomic DNA sequence databases were extensively searched for *Venenivibrio* on a global scale, complemented by a literature search for any mention of the genus. The microbial communities of 188 global hot spring metagenomes were taxonomically classified to thoroughly search for *Venenivibrio* traces, along with 16 metagenomes sourced from TVZ geothermal springs. To confirm that congruence between 16S rRNA gene sequences of the Hydrogenothermaceae was not masking the presence of *Venenivibrio* in these classifications, or indeed in any of the 16S rRNA gene databases, a selection of local and global metagenomes (with putative *Venenivibrio* and/or closely aligning Hydrogenothermaceae) were aligned to the genomes of CP.B2^T, *Sulfurihydrogenibium yellowstonense* SS-5^T, *Sulfurihydrogenibium* sp. Y03AOP1, and *Persephonella hydrogeniphila* 29W^T. This also validated the presence of *Venenivibrio* in Aotearoa-New Zealand geothermal features, which was corroborated by testing the classification and alignment accuracy of these analyses using mock communities of known taxon concentrations. Additionally, metagenome-assembled genomes (MAGs) were created, again from both local and global metagenomic samples, to further substantiate the sole ubiquity of *Venenivibrio* in the TVZ. Finally, phylogenomics of all publicly available genomes and MAGs classified to the Hydrogenothermaceae confirmed phylogeny of these taxa.

This work was consolidated into a manuscript that is under peer-review in the journal *Nature Communications*. In parallel to the work stated above, CP.B2^T culture experiments were performed by Holly Welford as part of her MSc thesis[280], with direction from Matthew Stott, and results were added to the manuscript to complement findings from this chapter. Specific growth characteristics of CP.B2^T were reanalysed to interrogate the observed widespread

prevalence of *Venenivibrio* in the TVZ, along with testing of *Venenivibrio* growth in hot spring water from outside of Aotearoa-New Zealand.

As the primary author for this manuscript, I led the study experimental design with direction from Matthew Stott and Carlo Carere. I coordinated and performed all field work for the 16S rRNA gene samples (as described in Chapter 2). I designed and implemented all biogeographical analyses of Aquificota and *Venenivibrio* in the TVZ. I assembled and annotated the CP.B2^T genome, and compared putative function to other Hydrogenothermaceae. I led the experimental design of the metagenome section, with input from Xochitl Morgan, Matthew Stott, and Carlo Carere. Because of access issues to high-performance computing, Daniel Hudson and Xochitl Morgan performed the metagenomic classification and alignment, testing of mock communities, and MAG generation. Kevin Lee, John Moreau, Thijs Ettema, Anna-Louise Reysenbach, Charles Lee, Daniel Colman, and Eric Boyd supplied either metagenomic or geothermal spring samples. I sourced all publicly available Hydrogenothermaceae genomes and MAGs for phylogenomic classification. I wrote the manuscript, with assistance from Matthew Stott, Carlo Carere, Craig Cary, and Ian McDonald, and all co-authors edited the text. A signed co-authorship form confirming these contributions can be found in Appendix F.

The remainder of this chapter and associated appendices are a reproduction of the manuscript text submitted to *Nature Communications*. Supplementary information accompanying this chapter is outlined in Appendix D. All scripts developed for statistics and figures are available at https://gitlab.com/morganlab/collaboration-1000Springs/1000Springs (under the venenivibrio_manuscript directory), with raw amplicon sequences used in this chapter deposited in the European Nucleotide Archive (ENA) under study accessions PRJEB24353. Shotgun sequencing of the *V. stagnispumantis* CP.B2^T genome was deposited in ENA under the accession ERR10117592, with the assembled genome deposited in the Genomes Online Database (GOLD Analysis ID Ga0311387) for associated annotation with the Integrated Microbial Genomes system (IMG Taxon ID 2799112217). Aotearoa-New Zealand MAGs are available under Project ID TAONGA-AGDR00025 in the Genomics Aotearoa Data Repository (https://doi.org/10.57748/vpk8-zp44). A separate manuscript was generated for the sequencing, assembly, and annotation of the *Venenivibrio stagnispumantis* CP.B2^T genome, which has been published in *Microbiology Resource Announcements*[197]. I am first author on this announcement and a draft version of the text can be found in Appendix E.

4.2 ABSTRACT

Allopatric speciation has been difficult to examine among microorganisms, with prior reports of endemism restricted to sub-genus level taxa. Here, we present evidence of an endemic bacterial genus, *Venenivibrio* (phylum Aquificota), from the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand. 16S rRNA gene sequencing revealed widespread distribution of *Venenivibrio* in TVZ geothermal springs (74 %, *n*=686), with maximal read abundance occurring at pH 4-6, 50-70 °C, and low oxidation-reduction potentials. Genomic analysis and culture-based experiments of the only characterised species for the genus, *Venenivibrio stagnispumantis* CP.B2[T], confirmed a chemolithoautotrophic metabolism dependent on hydrogen oxidation, further highlighting a specific environmental niche that could enhance habitat isolation. While similarity between *Venenivibrio* populations illustrated dispersal is not limited across the TVZ, extensive amplicon, metagenomic, and phylogenomic analyses of global microbial communities from DNA sequence databases indicates *Venenivibrio* is geographically restricted to the Aotearoa-New Zealand archipelago. We conclude that combined geographical and physicochemical constraints have resulted in the establishment of an endemic bacterial genus.

4.3 INTRODUCTION

A central tenet of biology is that geographic isolation enables the evolution of new species (i.e., allopatric speciation)[281]. MacArthur and Wilson used island ecosystems to demonstrate that the physical barrier of the ocean can also limit migration to provoke divergence of distinct species that are geographically separated[37], which can ultimately result in the formation of endemic taxa. While these models have been validated for macroorganisms across multiple biomes[41,282], evidence for microbial allopatry is largely constrained to strain-level hot spring and hydrothermal vent lineages[6,8,112,148,163]. Microbial taxa in other non-extreme habitats present no limits to dispersal[115,283,284], reinforcing the supposition that not all microorganisms diversify by the same mechanisms[285] and habitat fragmentation can acutely influence dispersal capability[286]. Gene flow, as a measure of genetic variability between conspecific taxa, can be interrupted by environmental and geographic isolation[285,287], with ecological niche boundaries contributing to speciation[288]. While there are some reports of endemism occurring in microbial communities[107,108,289], evaluating diversification within populations has focused on strains[6,13,163], with no indication of genus-level endemism being observed[109].

The bacterial phylum Aquificota (formerly Aquificae)[198] was first reported from hot marine sediments in 1992[243], and is prevalent in geographically patchy, high temperature environments[236]. Growth temperatures for Aquificota isolates range from ~45-90 °C across both marine (i.e., deep-sea thermal vents) and terrestrial (i.e., geothermal springs) habitats, with most, but not all, strains dependent exclusively on hydrogen and/or sulfur species as reductant for metabolism[236]. Aquificota are also predominantly microaerophilic[236], further restricting the range of environmental conditions amenable to growth and successful dispersal between 'island-like' habitats. Thus, taxa from this phylum are ideal candidates to investigate endemism, with ecosystems conducive to their growth facilitating speciation by allopatry, environmental selection, and competition for available niches.

We recently published a comprehensive study on the microbial diversity and biogeography across 925 geothermal springs in the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand[195]. While pH was the principal driver of microbial diversity in geothermal spring communities spanning the entire 8,000 km^2 region, Aquificota were the dominant taxa in springs >50 °C. Furthermore, three of the four most abundant genera in the study also belonged to Aquificota,

with the genus *Venenivibrio* (family Hydrogenothermaceae) identified in 74.2 % (*n*=686) of geothermal springs analysed (Table D.1). Only one species of this genus, *Venenivibrio stagnispumantis* strain CP.B2[T], has been formally described[179], with this strain isolated in 2007 from the Waiotapu geothermal field (*mana whenua*: Ngāti Tahu – Ngāti Whaoa), Aotearoa-New Zealand[138]. Interestingly, while sister genera of *Venenivibrio* (*Hydrogenothermus*, *Persephonella,* and *Sulfurihydrogenibium*) are globally distributed[112,135,141,290–294], we were unable to find any verifiable record of *Venenivibrio* taxa (cultivated species or via molecular signatures) outside of Aotearoa-New Zealand. The ubiquity of *Venenivibrio* within Aotearoa-New Zealand, but apparent absence of this taxon globally, led us to explore the hypothesis that the genus *Venenivibrio* is endemic to the Aotearoa-New Zealand archipelago.

To investigate the possibility of *Venenivibrio* endemism, we undertook a detailed ecological study of Aquificota in Aotearoa-New Zealand using both 16S rRNA gene amplicon and shotgun metagenomic data, as well as geothermal spring physicochemical measurements, while also searching for evidence of the *Venenivibrio* genus in publicly available DNA sequence databases. Genome analysis identified distinguishing characteristics that may explain the exclusivity of this taxon, with the type strain CP.B2[T] also cultured in spring waters *in vitro* from various locales to validate reported growth ranges and determine tolerances to physicochemical regimes of other geothermal provinces. Finally, phylogenomics confirmed placement of *Venenivibrio* as an independent genus, unequivocally distinguished from all other publicly available genomes and metagenome-assembled genomes (MAGs) within Hydrogenothermaceae. Our findings indicate *Venenivibrio* is only found in Aotearoa-New Zealand, with geography, associated dispersal limitation, and niche differentiation as possible mechanisms driving the apparent endemism of this bacterial genus.

4.4   METHODS

4.4.1   Aquificota & *Venenivibrio* ecology in Aotearoa-New Zealand

The data used to evaluate the distribution of Aquificota and *Venenivibrio* in Aotearoa-New Zealand were generated as part of the 1,000 Springs Project[195], a study investigating the biogeography of bacterial and archaeal communities in geothermal springs from the Taupō Volcanic Zone (TVZ). Briefly, microbial communities from the water columns of 925 geothermal springs were determined via 16S rRNA gene amplicon sequencing of the V4 region using the original 515F-806R Earth Microbiome Project primers[20], and the Ion PGM system for next-generation sequencing. Raw sequences were processed using USEARCH[211] and QIIME[212], rarefied to 9,500 reads per sample, and taxonomy was assigned to operational taxonomic units (OTUs) using the RDP classifier[214] (with a minimum confidence score of 0.5) and the SILVA SSU v123 database[215]. Forty-six physicochemical parameters were also measured from each spring; a full list is described in Supplementary methods. A detailed description of field sampling, sample processing, DNA extraction, DNA sequencing, OTU clustering and classification, and physicochemical analyses, can be found in Power *et al* (2018)[195].

To specifically investigate the ecology of Aquificota and *Venenivibrio* in Aotearoa-New Zealand geothermal springs, statistical analyses were performed using R v4.0.3[216] and the phyloseq package v1.32[217] (unless otherwise stated), with all figures generated using ggplot2 v3.3.2[219]. The 1,000 Springs Project dataset was first pruned to only samples containing reads assigned to the Aquificota using the subset_taxa function. Each Aquificota-assigned OTU was agglomerated to its respective genera using the tax_glom function to calculate taxon prevalence over the region, average relative read abundance of each genus across the original 925 springs sampled, and the number of OTUs found per genus. To conservatively analyse Aquificota diversity with respect to environmental variables, OTUs (not agglomerated to genus) with <20 reads across all samples and samples with <20 Aquificota-assigned reads were filtered from the dataset using the prune_taxa and prune_samples functions, respectively. The number of Aquificota OTUs per spring (i.e., OTU number) and the total abundance of Aquificota-assigned reads per spring (i.e., read abundance) were then calculated. Individual linear regression modelling was applied using Aquificota OTU number and read abundance separately as the response variable, against each physicochemical

parameter (*n*=46) as the predictor variable. This was followed by multiple linear regression using all physicochemical parameters, combined with a backward stepwise Akaike information criterion (AIC), to identify main predictor variables. Correlations of all 46 physicochemical variables against both Aquificota OTU number and read abundance were tested using the dist function to create Euclidean distance matrices, and the cor.test function to calculate Pearson's and Spearman's correlation coefficients. All coefficient *p*-values for both regressions and correlations were corrected using the p.adjust function with the false discovery rate (FDR) Benjamini-Hochberg procedure.

The data were then further pruned to contain *Venenivibrio*-assigned reads only. Again, OTUs with <20 total reads across all springs were removed, as were springs with <20 reads assigned to *Venenivibrio*. This resulted in 467 springs and 99 OTUs for all subsequent analyses. The number of *Venenivibrio* OTUs per spring (i.e., OTU number) and total read abundance of *Venenivibrio* per spring (i.e., read abundance) were calculated. As detailed for the Aquificota dataset, linear regression modelling and correlations were performed for both *Venenivibrio* diversity metrics against the 46 physicochemical variables measured for each spring.

4.4.2   Temperature, pH, and distribution of *Venenivibrio* in Aotearoa-New Zealand

To further scrutinise the effect of pH and temperature on these ecosystems, springs were binned by either one pH unit (i.e., springs with pH 4.01-5.00 binned as pH 4-5 etc.) or 10 °C (i.e., springs with temperatures of 60.1-70.0 °C binned as 60-70 °C etc.) increments and plotted against read abundance of *Venenivibrio* in each spring. To calculate a local polynomial regression of read abundance per spring, the loess function with default settings was applied to the data with both pH and temperature as the predictor variables. This model identified maximal abundance of the taxon as ≥4,000 reads and so, springs were then partitioned into 'high' (≥4000 reads) and 'low' (<4000 reads) *Venenivibrio* abundance to assess environmental differences between the two subsets (with 9,500 being the total number of reads per spring community). The get_googlemap function from the package ggmap v3.0.0[263] generated maps of the TVZ and individual geothermal fields, with either average relative read abundance in each geothermal field or read abundance in each spring added respectively, using WGS84 latitude and longitude coordinates. The proportional relative abundance of *Venenivibrio* OTUs per field and spring was also calculated. *Venenivibrio* hotspots were defined as springs with ≥85 % of sequencing reads from the microbial

community assigned to the genus, and were plotted on an outline of North Island, Aotearoa-New Zealand using the map_data and geom_polygon functions from ggplot2. The vegdist function from vegan v2.5-6[218] was used to generate Bray-Curtis dissimilarities between *Venenivibrio* populations in all springs, which was then coupled with pairwise geographic distance to investigate a distance-decay pattern across the region. Characteristics of each OTU that assigned to the genus *Venenivibrio* were determined, including distribution across the region (i.e., prevalence), average relative read abundance across all 925 springs originally sampled, and associated environmental variables including pH and temperature.

4.4.3   Genome annotation of *V. stagnispuman*tis CP.B2[T]

Extended annotation was performed on the draft genome sequence of CP.B2[T] [197] through the Integrated Microbial Genomes annotation pipeline v4.16.4[295] (IMG Genome ID 2799112217). Hydrogenase annotation was completed using HydDB[296]. Genes of interest from the CP.B2[T] genome were also manually compared to all available Hydrogenothermaceae genomes in IMG using the Function Search tool to identity distinguishing characteristics that might explain the putative exclusivity of *Venenivibrio*. A description of these genomes can be found in Supplementary methods.

4.4.4   Growth range reassessment of *V. stagnispumantis* CP.B2[T]

To interrogate the observed widespread distribution of *Venenivibrio* in Aotearoa-New Zealand, specific growth characteristics of the type strain CP.B2[T] were reanalysed including temperature, pH, salinity, and O_2 tolerances. Growth media were prepared as per Hetzer *et al.* (2008)[179], with amendments where necessary as described in Supplementary methods. Additionally, geothermal spring water was collected from both Champagne Pool (CP), Waiotapu, Aotearoa-New Zealand, the origin of CP.B2[T], and Obsidian Pool (OP), Yellowstone National Park, USA, to test *in vitro* growth capability of CP.B2[T] in a chemically similar, global spring. Both spring water samples were filtered twice, immediately in the field and then in the laboratory, using a Sterivex-GP 0.22 µm PES column filter to ensure sterility, and stored at 4 °C until use. Three variations of each sample (natal spring water, spring water with supplemented S^0, and spring water with supplemented S^0, NH_4Cl and KH_2PO_4) were inoculated with CP.B2[T] in triplicate and incubated at 70 °C for seven days.

4.4.5 Global search for 16S rRNA genes reported as, or closely related to, *Venenivibrio*

The following nine databases were screened for evidence of the 16S rRNA gene of *V. stagnispumantis* CP.B2[T]: NCBI's Nucleotide collection (which includes GenBank, EMBL, DDBJ, and RefSeq)[297], the Sequence Read Archive (SRA)[297], SILVA[215], the Ribosomal Database Project (RDP)[298], Greengenes[299], the Integrated Microbial Next Generation Sequencing (IMNGS) platform (which includes amplicon sequencing from SRA, DDBJ, and ENA)[300], JGI's Integrated Microbial Genomes and Microbiomes (IMG/M) system[295], the Earth Microbiome Project (EMP)[20], and the Qiita platform[301]. A 16S rRNA gene phylogenetic tree was also constructed of entries in the SILVA database (SSU r138.1) assigned to the genus *Venenivibrio*. Additionally, all NCBI databases (including PubMed Central) and Google Scholar were searched for any literature and/or samples containing the word '*Venenivibrio*'. Any 16S rRNA gene sequences putatively identified as *Venenivibrio* were manually checked for ≥95.0 % sequence similarity to *V. stagnispumantis* CP.B2[T] using BLASTN v2.13.0 with the megablast setting. Detailed search criteria for each database are outlined in Supplementary methods, along with the 16S rRNA gene phylogenetic tree members and construction parameters.

4.4.6 Screening metagenomes for *Venenivibrio*

We screened all available TVZ geothermal metagenomic samples (*n*=16) from 10 previously sampled geothermal springs (i.e., local metagenomes; Table D.2) for the presence of *Venenivibrio* using the metagenomic classifier Kraken2 v2.0.8[302]. The default Kraken2 database failed to classify Hydrogenothermaceae reads to genus or species level, so a custom database was built using the NCBI Taxonomy database[303] and RefSeq complete bacterial reference library[304] (accessed 4/Sep/2020). The *V. stagnispumantis* genome is not included in RefSeq, so the type strain genome CP.BP[T] was downloaded from IMG and added manually (GOLD Analysis ID Ga0170441; IMG Genome ID 2724679818). The output from Kraken2 was then searched for '*Venenivibrio*' using grep v2.21. TerrestrialMetagenomeDB[305] release 2.0 (January 2021), a curated and standardised metadata repository for terrestrial (non-marine) metagenomes in the databases SRA and MG-RAST, was used to search for *Venenivibrio* signatures in hot springs outside of Aotearoa-New Zealand (i.e., global metagenomes). The following terms were used in the search (accessed 31/Jan/2021); 'hot spring', 'hotspring', 'hydrothermal', and 'geothermal'. Resultant metagenomes were downloaded from either SRA using SRA Tools v2.9.3 or MG-RAST manually. These were

classified in the same way as local metagenomes using Kraken2, with the custom database described above.

Four local metagenomes with increased *V. stagnispumantis* read abundance, either by metagenomic or 16S rRNA gene community profiling, were aligned to both the type strain CP.B2^T (Ga0170441) and three of the most closely related family members using bowtie2 v2.3.5.1[306] (with default settings). These were identified as *Sulfurihydrogenibium yellowstonense* SS-5^T (RefSeq assembly accession GCF_000173615), *Persephonella hydrogeniphila* 29W^T (GCF_900215515), and *Sulfurihydrogenibium* sp. Y03AOP1 (GCF_000020325). In addition, six global metagenomes returning putative *V. stagnispumantis* signatures were also selected to map to the Hydrogenothermaceae genomes. Detailed selection criteria for the ten metagenomes used in this analysis can be found in Supplementary methods. Output files were first converted to BAM format using samtools v1.10[307] and then to BED format using the genomecov function in bedtools v2.29.2[308] to import into R. For each metagenomic sample (*n*=10), average coverage depth per contig was calculated and plotted against relative genome position to assess evenness across the breadth of the four genomes. The fraction of each contig covered (i.e., coverage breadth) was also determined. *De novo* assembly was then performed on four local and two global metagenomes using the ATLAS v2.6a3 workflow[309]. This included quality control of sequencing reads and binning into metagenome-assembled genomes (MAGs). Where the sequencing platform differed from Illumina, MEGAHIT v1.2.9[310] and MetaBAT 2 v2.10[311] were used for assembly and binning, respectively. All MAGs were analysed with FastANI v1.32[312] to calculate average nucleotide identity (ANI) to *V. stagnispumantis* CP.B2^T and CheckM v1.0.5[313] was used to determine genome completeness and contamination of matching bins.

A sequencing simulator (InSilicoSeq v1.5.1)[314] created synthetic metagenomes to test the classification and alignment accuracy of *Venenivibrio*, *Sulfurihydrogenibium*, and *Persephonella* sequencing reads. The iss function generated reads modelled on Illumina MiSeq sequencing technology from each of the following genomes: *V. stagnispumantis* CP.B2^T (Ga0170441), *Sulfurihydrogenibium* sp. Y03AOP1 (GCA_000020325.1), and *P. hydrogeniphila* 29W^T (GCA_900215515.1), along with 10 random genomes in NCBI (accessed 21/Dec/2020; Supplementary methods). Seqtk-1.3 (r106)[315] was then used to subsample reads to build eight mock communities. The output FASTQ files were classified

using Kraken2 with the custom database and aligned to Hydrogenothermaceae genomes using bowtie2 as previously described.

4.4.7 Phylogenomics of Hydrogenothermaceae

An approximate maximum-likelihood phylogenomic tree was built using SpeciesTree v2.2.0 in KBase[316] to confirm *Venenivibrio* as a distinct genus within the Hydrogenothermaceae. All publicly available genomes and MAGs from the family were imported from NCBI RefSeq[304], GTDB[317], and IMG/M[295] databases (07/Feb/2022), and the tree was generated using multiple sequence alignment with 49 house-keeping COG gene families as defined by SpeciesTree[316]. This also ensured robust classification of all assemblies to aid our search for *Venenivibrio* globally. Four MAGs generated from geothermal springs in Aotearoa-New Zealand (i.e., local) and two MAGs from other locales (i.e., global) were included. Additionally, the GTDB relative evolutionary divergence (RED) score of CP.B2^T was calculated[317] and average nucleotide identity (ANI) scores were generated of the Hydrogenothermaceae in IMG using the Compare Genomes function. Molecular clock analysis using full length 16S rRNA gene sequences of *V. stagnispumantis* CP.B2^T and *S. yellowstonense* SS-5^T was also completed using TimeTree[318] to estimate a pairwise divergence time of the most recent common ancestor to the two genera.
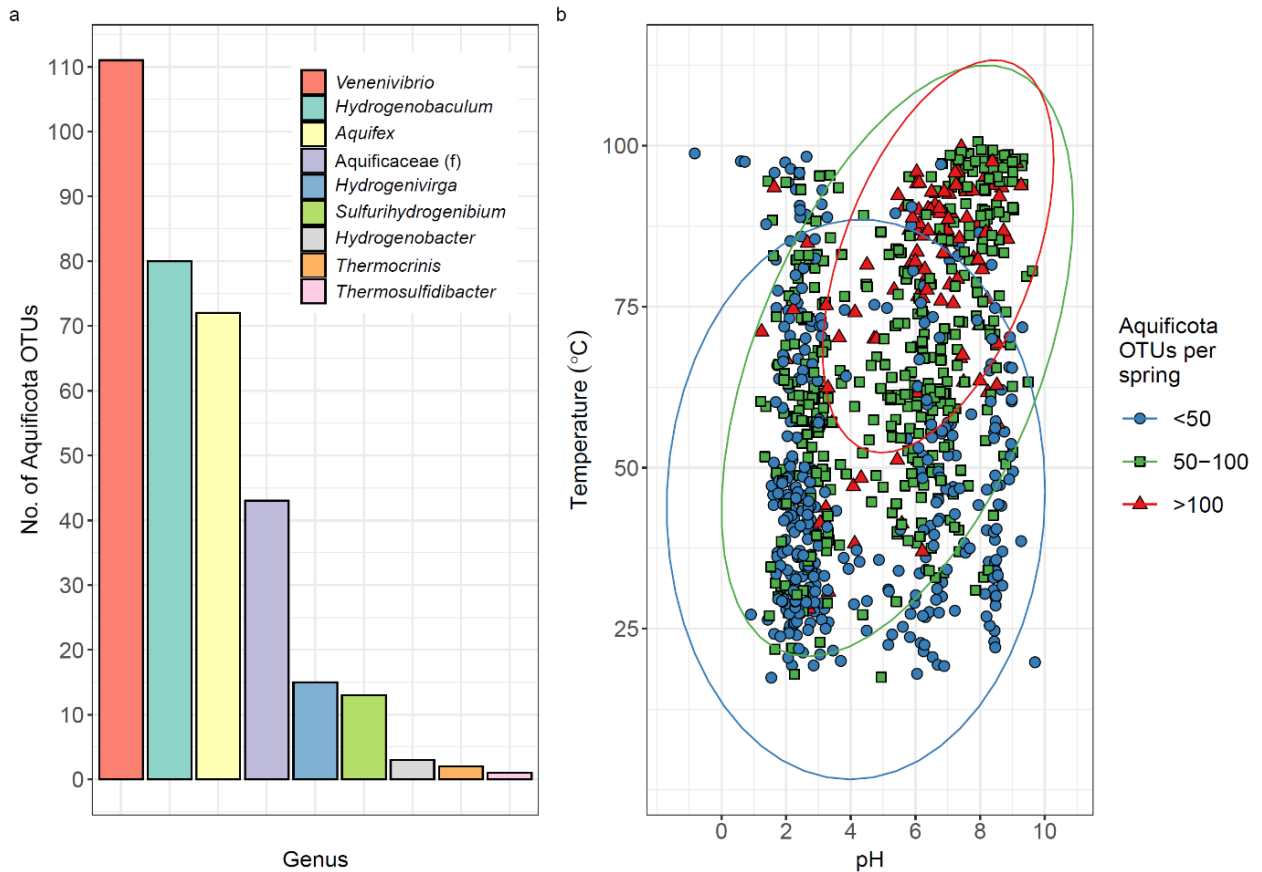
Figure 4.1 - Diversity of Aquificota 16S rRNA genes in the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand. **(a)** The division of Aquificota-assigned 16S rRNA gene operational taxonomic units (OTUs) to their respective genus from amplicon sequencing of the microbial communities found in 925 geothermal springs prior to filtering. Where taxonomy was unable to be assigned to an OTU at genus level, the corresponding family is shown (f). **(b)** All springs that contained Aquificota (*n*=891) in the TVZ are plotted according to environmental pH and temperature. The number of Aquificota-assigned OTUs in each spring is represented by blue (<50 OTUs), green (50-100 OTUs), or red (>100 OTUs), with data ellipses assuming multivariate t-distribution and a 95 % confidence level.

4.5 RESULTS

4.5.1 Aquificota & *Venenivibrio* ecology in Aotearoa-New Zealand

Of the 28,381 OTUs generated from the 1,000 Springs Project[195], 340 were assigned to the phylum Aquificota and were detected in 891 of the 925 individual geothermal springs analysed. Aquificota taxa comprised the Aquificaceae and Hydrogenothermaceae families (216 and 124 OTUs, respectively), with 111 *Venenivibrio*-assigned OTUs (Hydrogenothermaceae) found in 686 springs (Figure 4.1; Table D.1). The observed number of Aquificota OTUs increased with increasing spring temperature and pH (Figure 4.1). After conservatively removing exiguous Aquificota-assigned OTUs, 305 OTUs and their combined

read abundance per 783 springs were then analysed to assess the effect of 46 physicochemical variables on diversity in the TVZ. Temperature, oxidation-reduction potential (ORP), pH, and conductivity had the greatest linear relationship to changes in both Aquificota OTU number and read abundance per spring (Figure 4.2; Table D.3). Multiple linear modelling suggested 38.9 % of the variation in Aquificota OTU number was mostly attributable to temperature, ORP, and Ag ($p<0.001$, Table D.4), with 35.8 % of variation in read abundance predominantly related to temperature, ORP, pH, SO_4^{2-}, Mn, and Fe^{2+} ($p<0.001$; Table D.5). Temperature, pH, ORP, and HCO_3^- most strongly correlated to Aquificota OTU number per spring, with comparable results produced for read abundance ($p<0.001$, Pearson's and Spearman's correlation coefficients; Table D.6).

A reduced number of geothermal springs ($n=467$), containing 99 *Venenivibrio*-assigned OTUs, was used to investigate the ecology of this genus in Aotearoa-New Zealand after again conservatively removing exiguous OTUs (Figure D.1; Table D.7). Overall, the linear relationship between spring physicochemistry and the diversity of *Venenivibrio* appeared weaker than that of the entire phylum. Limited or no effect was observed for *Venenivibrio* OTU number and read abundance per spring with increased temperature and pH, with ORP and turbidity having the most association (Figure 4.2; Table D.8). Multiple linear modelling indicated Na, turbidity, and ORP principally accounted for 19.2 % of variation in *Venenivibrio* OTU number across all springs ($p\leq0.002$; Table D.9), with 18.3 % of variation in read abundance mostly influenced by ORP, turbidity, Li, and Na ($p\leq0.001$; Table D.10). Turbidity had the strongest correlation between both *Venenivibrio* diversity metrics, with ORP and H_2S also having significant effects ($p<0.001$, Pearson's and Spearman's correlation coefficients; Table D.11).
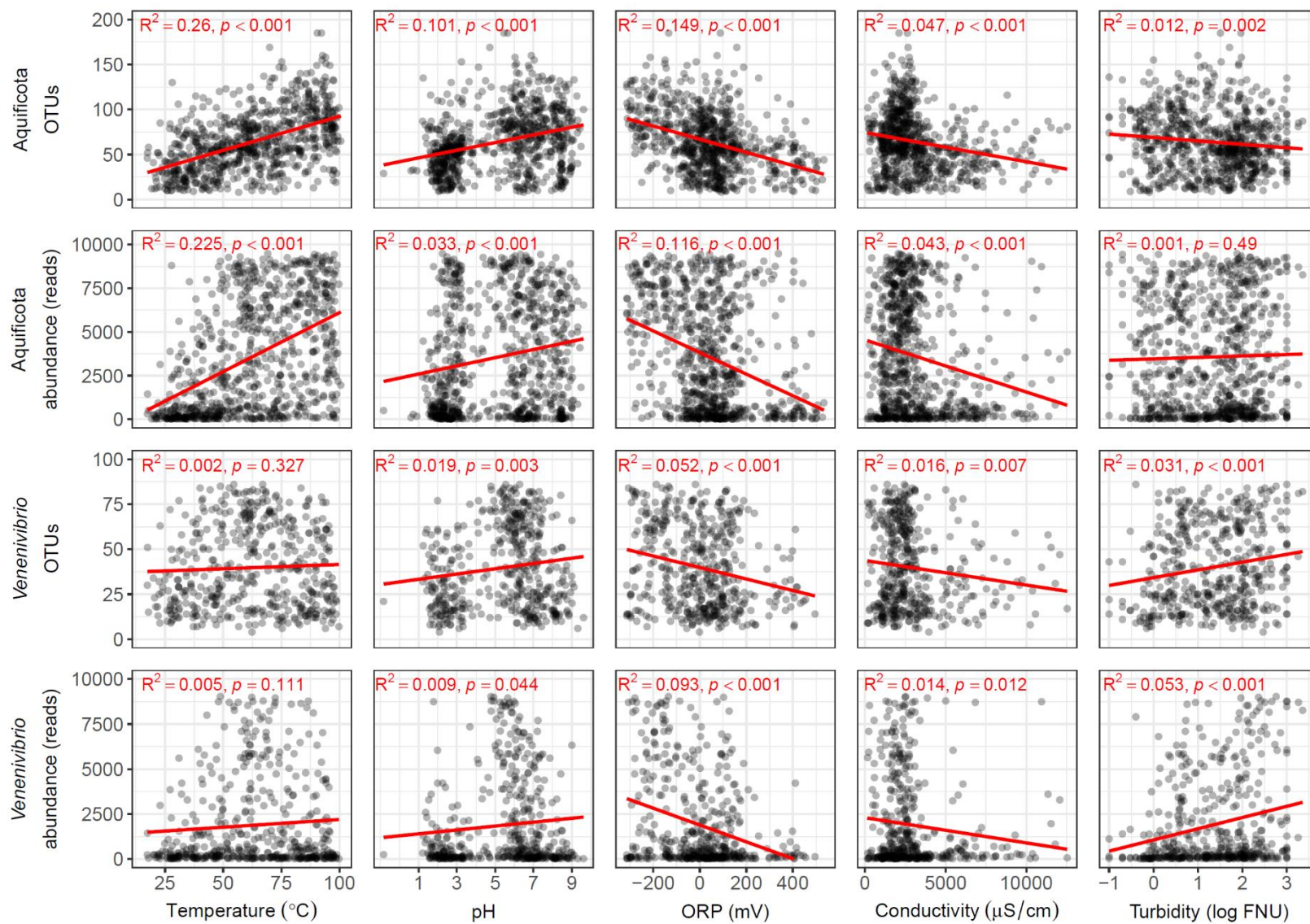
Figure 4.2 - Aquificota and *Venenivibrio* 16S rRNA gene diversity as a function of select environmental measurements in Aotearoa-New Zealand geothermal springs. Diversity is represented as the number of operational taxonomic units (OTUs) assigned to either the phylum or genus in each spring, and the total read abundance of Aquificota- or *Venenivibrio*-assigned reads per spring community (reads). Maximum sequence reads per spring community was 9,500. Linear regression was applied to each diversity metric against spring temperature, pH, oxidation-reduction potential (ORP), conductivity, and turbidity. OTU number and read abundance strongly correlated with each other (Pearson's and Spearman's coefficients: 0.674 and 0.714, respectively, *p*<0.001; Table D.11).

4.5.2   Temperature, pH, and distribution of *Venenivibrio* in Aotearoa-New Zealand

Polynomial regression highlighted *Venenivibrio* read abundance was greatest within a limited range of pH and temperature conditions (Figure 4.3). From the reduced dataset of 467 springs, microbial communities with ≥4,000 reads assigned to *Venenivibrio* (42 % of the spring community; *n*=91) had a median pH and temperature of 5.9 (IQR 1.2) and 64 °C (IQR 16.3), respectively (Figure 4.3). In comparison, springs with <4,000 reads (*n*=376) had a median pH and temperature of 6.1 (IQR 4.3) and 62.3 °C (IQR 41.3), respectively. The narrow pH and temperature range for *Venenivibrio* abundant springs was again evident when springs were subdivided into environmental increments, with the greatest read abundance observed at pH 4-6 and 50-70 °C (Figure 4.3). The ten most abundant *Venenivibrio* OTUs were also found in more acidic and hotter spring environments (≥1,000 reads per spring; pH 5.8 and 67.6 °C; Figure 4.4) than their less abundant counterparts (<1,000 reads; pH 6.0 and 62.9 °C; Figure D.2).

Considering the distribution of *Venenivibrio* across the TVZ, average relative read abundance in springs containing the taxon per geothermal field ranged from 0.8 to 35.7 %, with Waiotapu having the greatest abundance (Figure 4.5). Twenty individual springs located in six geothermal fields were found to have ≥85 % of the total microbial community assigned to *Venenivibrio* (Figure D.3), with a median pH and temperature of 5.5 and 66.0 °C, respectively. In contrast to a weak distance-decay pattern observed in microbial communities across the TVZ[195], *Venenivibrio* populations did not show this trend (R^2=0, *p*=0.007; Figure D.4), with 19 of the most abundant OTUs found in all 14 geothermal fields analysed (Figure 4.5). In particular, Whakaari/White Island, situated ~48 km offshore from the mainland, showed increased *Venenivibrio*-level evenness in spring communities. Conversely, most of the *Venenivibrio* read abundance in two springs from the Waiotapu geothermal field was assigned to a single OTU (OTU_5; 65 and 77 % of the total microbial community; Figure 4.4 & Figure D.5). While OTU_5 was the most prevalent *Venenivibrio*-OTU in springs across the

entire TVZ (n=456; Table D.7), its next greatest read abundance in an individual spring was 22 % of the microbial community. The third greatest read abundance of a single *Venenivibrio* OTU was OTU_26661, which comprised 62 % of a Rotorua spring community (Figure 4.4 & Figure D.5). All remaining OTUs had <42 % relative read abundance in any of the 467 microbial communities analysed.

4.5.3   Genome annotation of *V. stagnispumantis* CP.B2[T]

The CP.B2[T] genome encoded the capacity for a chemolithoautotrophic mode of energy generation and carbon fixation (Table D.12). This included genes for the Type I reductive tricarboxylic acid (rTCA) cycle[319], cytochrome bd (*cydAB*) as the sole respiratory terminal oxidase, and two [NiFe]-hydrogenases from groups 1b (*hynAB*) and 2d (*huaSL*)[320]. Despite obligately requiring a reduced source of sulfur for growth[179], predicted sulfur-metabolising genes were limited to two copies of a membrane-bound sulfide:quinone oxidoreductase (*sqr*). There was evidence of an extensive sulfur-trafficking network inside the cell, including three copies of a TusA-related sulfurtransferase, two rhodanese-type transferases (*sseA* and *pspE*), and the *dsrEFH* complex[321,322]. There was no evidence of capacity for either assimilatory sulfate reduction or the full SOX pathway for sulfur/thiosulfate oxidation. Genes for arsenic resistance (*arsRBC*) were detected, which presumably transports reduced arsenite [As^{3+}] or stibnite [Sb^{3+}] out of the cytosol via a membrane-bound efflux pump[323,324]; both As^{3+} and Sb^{3+} are present in elevated concentration in the CP.B2[T]-host environment[265]. Consistent with the characterisation of CP.B2[T] and an investigation into microbial-induced arsenic cycling of Champagne Pool[179,188], no evidence of described genes encoding dissimilatory arsenic metabolism was noted in the genome. Analysis of ten Hydrogenothermaceae genomes (including CP.B2[T]) to identify any distinguishing characteristics contributing to the apparent endemism of *Venenivibrio* revealed all genomes encoded ATP-citrate lyase (ACL; gene *aclAB*), used in the reductive TCA cycle. Interestingly, citrate synthase genes (*gltA*), required in the oxidative TCA cycle, were exclusive to *P. marina* EX-H1[T], *P. hydrogeniphila* 29W[T], and all *Sulfurihydrogenibium* spp. analysed. Most notably, CP.B2[T] was the only isolate which did not encode the full SOX pathway.
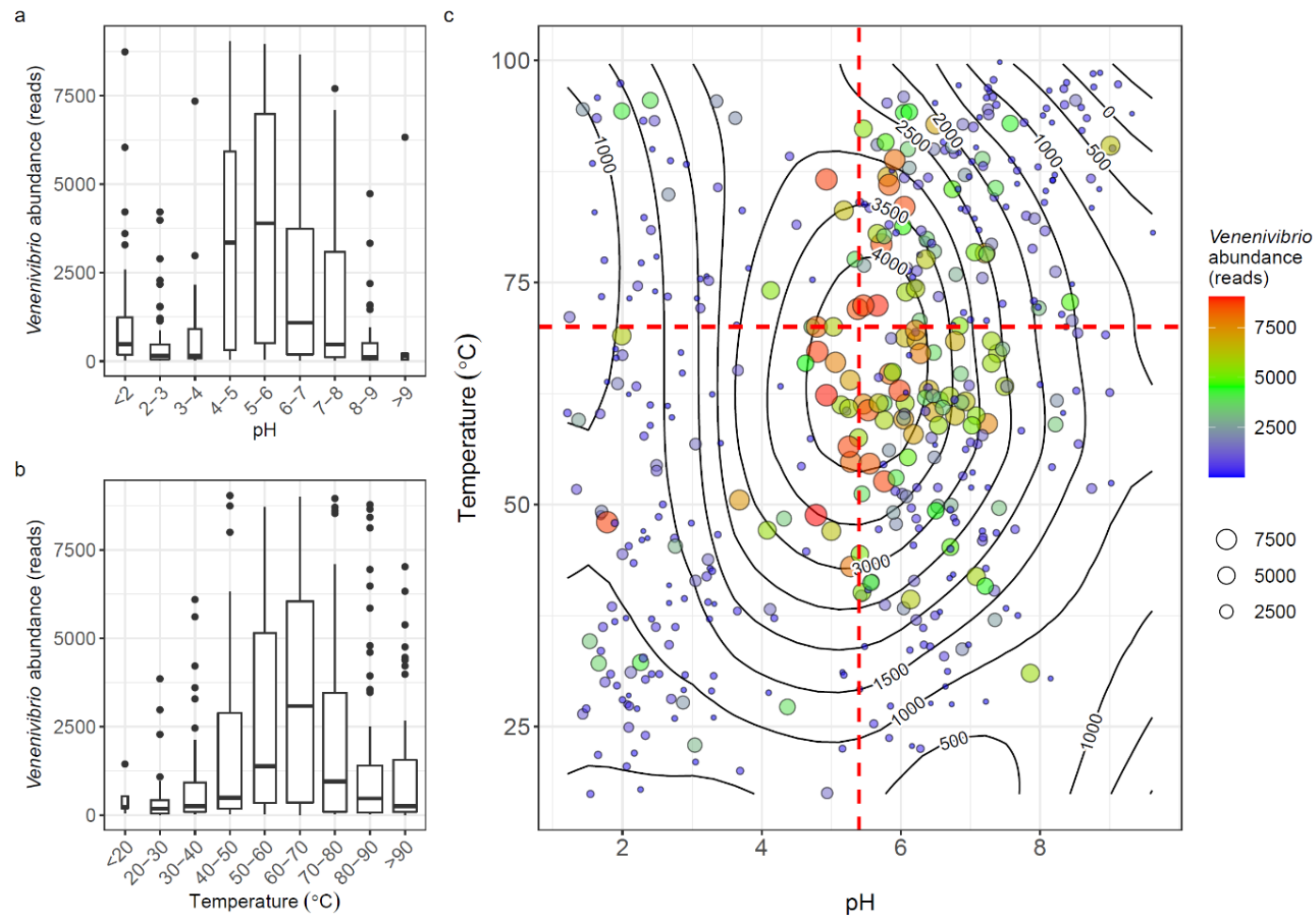
Figure 4.3 - Read abundance of *Venenivibrio* 16S rRNA genes as a function of geothermal spring pH and temperature in Aotearoa-New Zealand. **(a)** The read abundance of *Venenivibrio* 16S rRNA genes in geothermal springs (post filtering; *n*=467) split by increments of spring pH. **(b)** The read abundance of *Venenivibrio* 16S rRNA genes split by increments of spring temperature. **(c)** A local polynomial regression of *Venenivibrio* 16S rRNA gene read abundance (contours) was applied to examine the relationship to both spring pH and temperature (*n*=466), with abundance also described by colour and size of the points. Maximum sequence reads per spring community was 9,500. The dashed red lines represent the reported optimum pH and temperature of the type strain, *V. stagnispumantis* CP.B2^T, as per Hetzer *et al.* (2008).
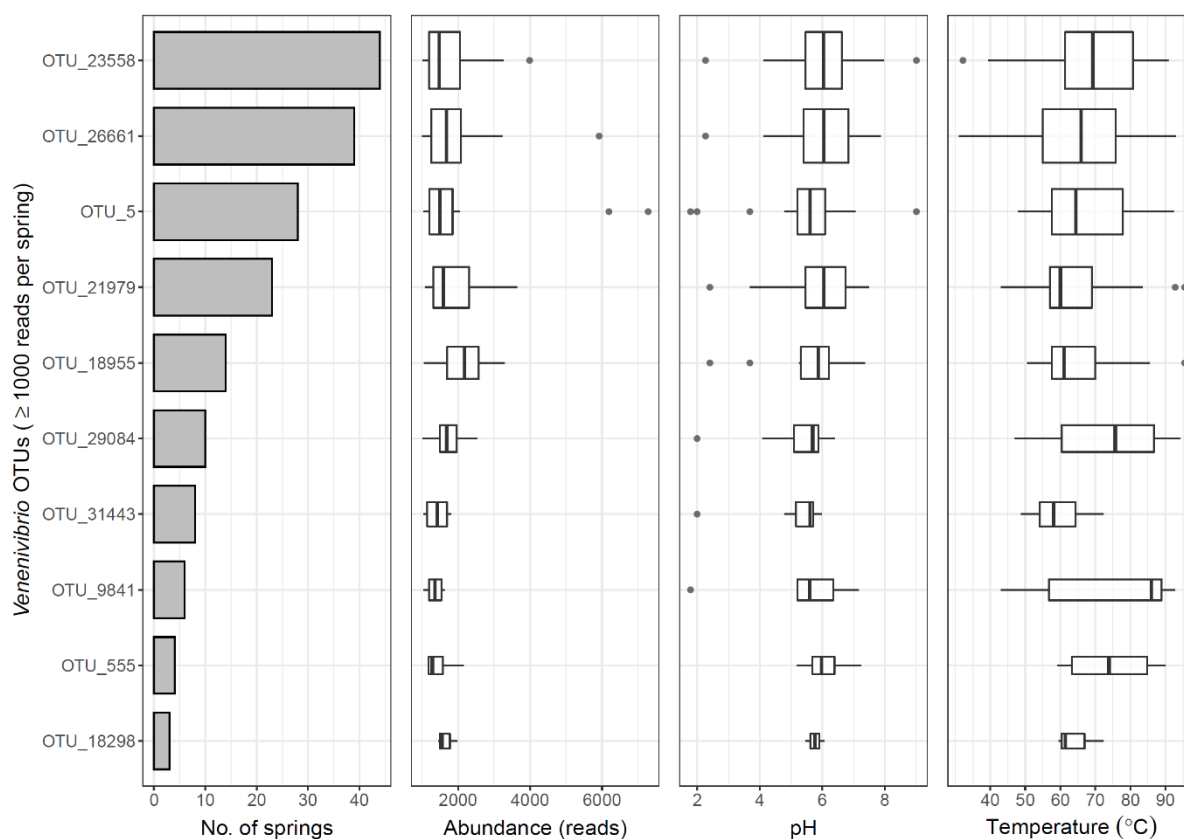
Figure 4.4 - Prevalence and read abundance of select *Venenivibrio* operational taxonomic units (OTUs) as a function of spring pH and temperature. Only OTUs with read abundances of ≥1,000 sequence reads per spring (10.5 % of the total community; *n*=10) that assigned to the genus *Venenivibrio* are shown. Maximum sequence reads per spring community was 9,500. Abundant OTUs are ordered via prevalence (i.e., no. of springs) across all springs where *Venenivibrio* was found (post-filtering; *n*=467). Median pH and temperature for abundant OTUs were 5.8 (IQR 0.9) and 67.6 °C (IQR 18.2), respectively.

4.5.4   Growth range reassessment of *V. stagnispumantis* CP.B2[T]

In light of the broad distribution of *Venenivibrio* across a range of temperature and pH conditions, we reinterrogated *in vitro* growth conditions of the type strain *V. stagnispumantis* CP.B2[T]. Results from these experiments differed considerably from those previously reported (Table D.13)[179]. Notably, viable growth was observed (via floc formation) up to 8 % w/v NaCl (previously reported as 0.8 % w/v NaCl) across a pH range of 3.0-8.5 (previously reported as pH 4.8-5.8). However, growth optima for temperature, pH, salinity, and O_2 tolerance did not change substantially, reinforcing the preferred conditions for increased abundance of the taxon (Figure 4.3). Further, to test whether there were some cryptic physicochemical conditions that restricted *Venenivibrio* growth outside Aotearoa-New Zealand, we tested growth in filtered spring water from Obsidian Pool, Yellowstone National

Park (YNP), USA, and Champagne Pool (TVZ) as a control. Obsidian Pool displays similar physicochemistry to Champagne Pool and is known to support resident populations of *Sulfurihydrogenibum*[292]. CP.B2^T grew in all conditions (natal spring water, spring water with supplemented S^0, and spring water with supplemented S^0, NH$_4$Cl, and KH$_2$PO$_4$) for both geothermal features.

4.5.5 Global search for 16S rRNA genes reported as, or closely related to, *Venenivibrio*

No significant evidence was found of the 16S rRNA gene of *V. stagnispumantis* CP.B2^T outside of Aotearoa-New Zealand (Figure D.6, Table D.14, Table D.15, & Table D.16). The nine microbial databases exhaustively searched included 971,117 samples, 26.7 billion 16S rRNA gene sequences, and 12.2 petabytes of sequence data. Putatively identified *Venenivibrio* 16S rRNA gene sequences were easily invalidated via manual interrogation of sequence similarity (<95 %) to *V. stagnispumantis* CP.B2^T.

4.5.6 Screening metagenomes for *Venenivibrio*

All 16 metagenomic samples from TVZ geothermal springs (i.e., local metagenomes) contained genomic signatures of *V. stagnispumantis*, with a maximum community read abundance of 45.1 % (Table D.2). Champagne Pool metagenomes (*n*=4), the locale where CP.B2^T was originally isolated, had read abundances from 8.6 to 27.7 % across the outflow channel, adjacent terrace, rim of the spring, and water column. A total of 188 metagenomes were associated with the term 'hot spring' (i.e., global metagenomes) from a database of 20,206 terrestrial samples[305]. The terms 'hotspring' and 'geothermal' yielded no results, while 'hydrothermal' resulted in 16 non-surface samples (i.e., deep-sea vent or subsurface habitats) that were excluded from classification analysis. Twenty-one of the 188 hot spring metagenomes returned putative traces of *Venenivibrio* from community composition analysis (Table D.17). Eight of these exhibited 0.01-1.08 % of the total community being assigned to *V. stagnispumantis*, with the remaining 13 metagenomes having negligible *Venenivibrio* read abundance (0 %) compared to each respective community.

To test the validity of these putative traces, four local and six global metagenomes with substantive *Venenivibrio*-assigned reads were aligned to both the *V. stagnispumantis* CP.B2^T genome and the most closely related Hydrogenothermaceae (Figure 4.6). All metagenomes from Aotearoa-New Zealand mapped to 91.8-96.8 % of the CP.B2^T genome, with an average coverage depth of 35-1006 x (Figure 4.6; Table D.18). In contrast, the three other

Hydrogenothermaceae genomes had relatively minimal coverage breadth (≤0.81 %) and depth (≤0.17 ×) by the local metagenomes (Figure 4.6; Table D.18). The global metagenomes, sourced from geothermal springs in Japan (*n*=3), USA (*n*=2), and Canada (*n*=1), aligned to 0.1-10.2 % of the CP.B2^T genome, with an average coverage depth of 0.0-16.2 x (Figure 4.6; Table D.18). These samples had slightly more coverage breadth across the other Hydrogenothermaceae genomes, mainly *Sulfurihydrogenibium* sp. Y03AOP1 (0.2-13.1 %) and *S. yellowstonense* SS-5^T (0.4-12.9 %). Additionally, metagenome-assembled genomes (MAGs) were created from four local and two global metagenomes (Table D.19). Each of the local metagenomes produced a MAG that had an average nucleotide identity (ANI) of 94.8-97.9 % to the *V. stagnispumantis* CP.B2^T genome. Conversely, MAGs from the two global samples with the closest congruity to CP.B2^T were ANI <80 % similar. Completeness for all MAGs ranged from 91.9-99.2 % (Table D.19). Mock communities designed to test the robustness and credibility of metagenomic analysis substantiated the classification and alignment of *Venenivibrio* in these metagenomic samples (Figure D.7; Table D.20, & Table D.21).

4.5.7   Phylogenomics of Hydrogenothermaceae

Thirty-eight publicly available genomes and MAGs (*n*=11 and 27 respectively) from the Hydrogenothermaceae were used to build an approximate maximum-likelihood phylogenomic tree, placing *V. stagnispumantis* strain CP.B2^T as a separate clade to all other genera (Figure 4.6). GTDB also positioned CP.B2^T as a novel genus within the Hydrogenothermaceae (RED value of 0.82). The four MAGs from Aotearoa-New Zealand geothermal springs positioned within this clade, with the two global MAGs generated by this study phylogenomically classifying as *Sulfurihydrogenibium* spp. Pairwise ANIs of Hydrogenothermaceae isolates to CP.B2^T were <80 %, with *Sulfurihydrogenibium yellowstonense* SS-5^T being the mostly closely related non-Aotearoa-New Zealand taxon. Molecular clock analysis between CP.B2^T and SS-5^T suggested divergence of the most recent common ancestor to the two genera occurred ~67 mya.

More detail on genome annotation (D.2.1) and growth tolerances (D.2.2) of CP.B2^T, individual database results for the CP.B2^T 16S rRNA gene search (D.2.3), and the synthetic metagenome analysis (D.2.4) can be found in Supplementary results.
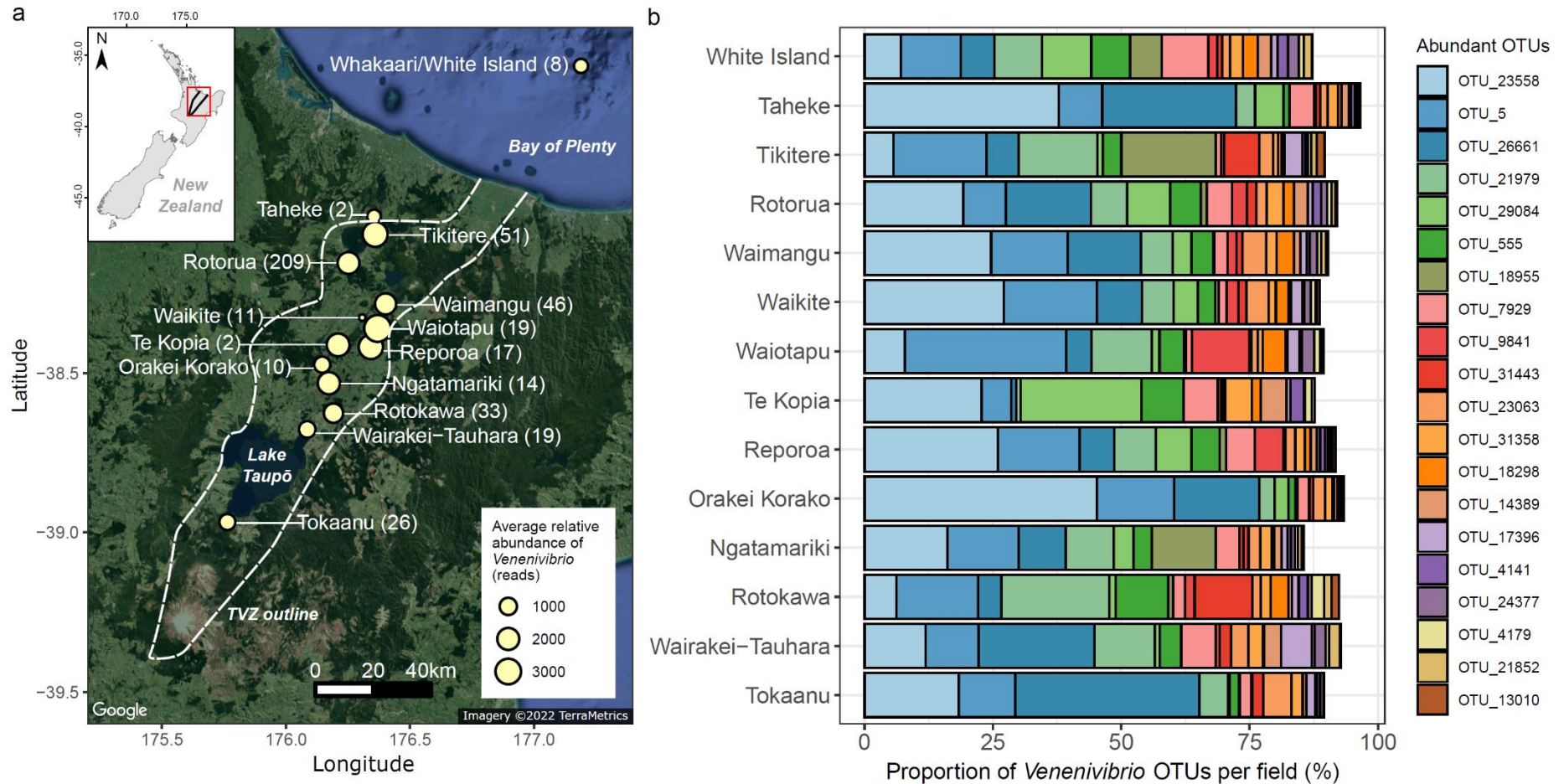
Figure 4.5 - Distribution of *Venenivibrio* 16S rRNA genes in Aotearoa-New Zealand. **(a)** A map of the Taupō Volcanic Zone (TVZ), with geothermal fields where *Venenivibrio* 16S rRNA genes were detected highlighted in yellow. Geothermal fields are scaled according to average abundance of *Venenivibrio*-assigned sequencing reads across springs in each field, with the number of springs containing *Venenivibrio* displayed in brackets. Maximum sequence reads per spring community was 9,500. **(b)** The proportion of the 20 most abundant *Venenivibrio* 16S rRNA gene operational taxonomic units (OTUs) per geothermal field, where 100 % equals the total relative abundance of *Venenivibrio* reads within each field. Geothermal fields are ordered north to south. A breakdown of this relative abundance in reads per spring can be found in Figure D.5.

a
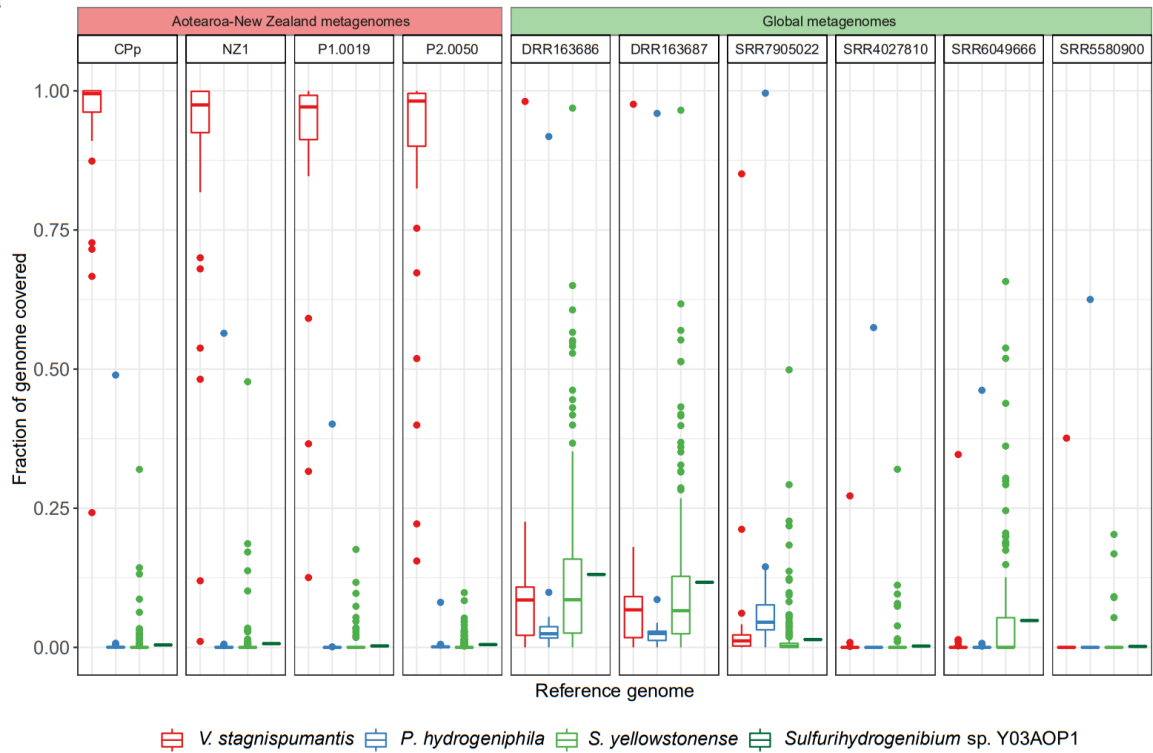
b
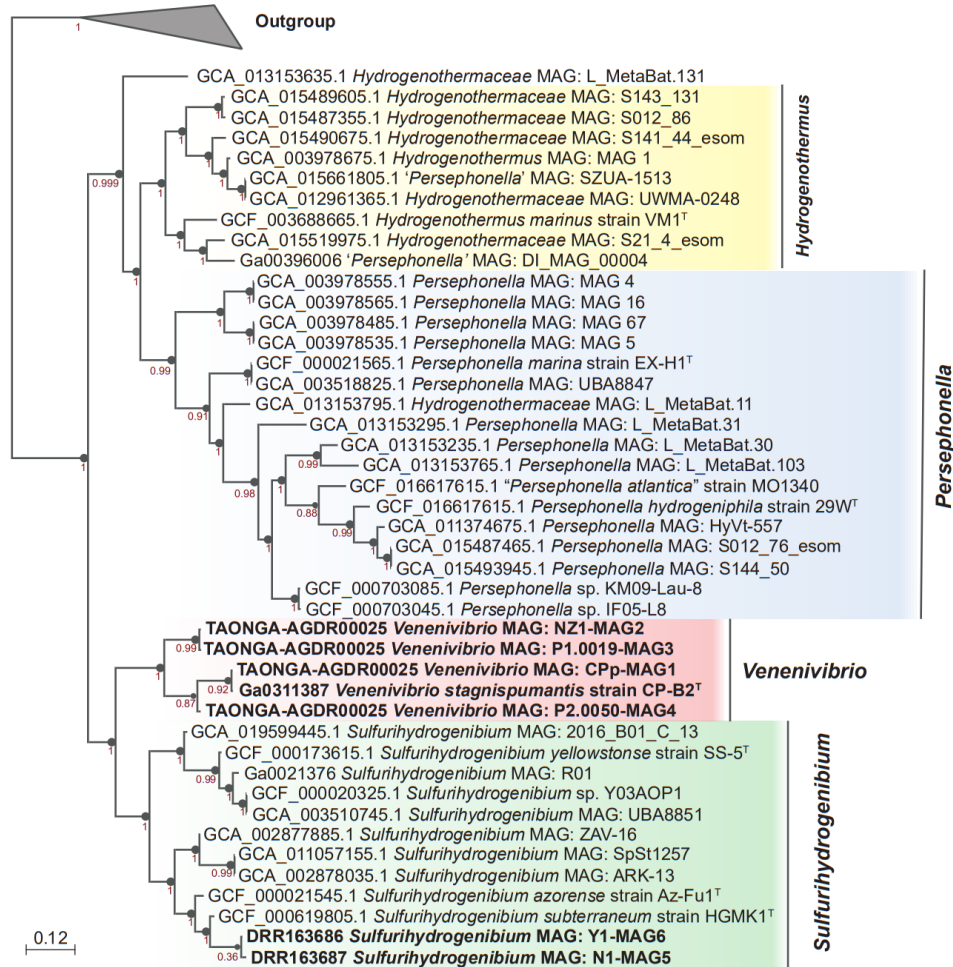
Figure 4.6 - Metagenomic evidence for *Venenivibrio* in Aotearoa-New Zealand. **(a)** Metagenome reads derived from four Aotearoa-New Zealand (left) and six global (right) geothermal spring microbial communities were mapped to the *V. stagnispumantis* CP.B2^T genome and three genomes of the most closely related strains from family Hydrogenothermaceae. Box plots represent average mapped coverage across each of the reference genome contigs (i.e., fraction of the reference genome covered) by the metagenomic samples. Samples are ordered by decreasing sequence similarity to CP.B2^T. **(b)** An approximate maximum-likelihood phylogenomic tree of all publicly available genome assemblies, including isolates and metagenome-assembled genomes (MAGs), from Hydrogenothermaceae. Four Aotearoa-New Zealand and two global MAGs generated by this study are highlighted in bold. The tree was built using 49 house-keeping COG gene families, with bootstrap confidence values indicated at nodes. Selected members of Aquificaceae (*Aquifex aeolicus* VF5^T [GCF 000008625.1], *Hydrogenobacter thermophilus* TK-6^T [GCF 000010785.1], and *Thermocrinis ruber* OC 1/4^T [GCF 000512735.1]) were included as an outgroup.

4.6 DISCUSSION

4.6.1 *Venenivibrio* is ubiquitously dispersed in Aotearoa-New Zealand, but not globally

The phylum Aquificota dominates geothermal spring water in the TVZ, a region spanning 8,000 km^2, with *Venenivibrio* identified as the most abundant genus across the 925 springs analysed (10 % average relative read abundance). Illustrating that local dispersal within the TVZ is not limited, *Venenivibrio* was found in 74.2 % of springs across a wide range of physicochemical conditions. While the reassessed growth ranges for the type strain, *V. stagnispumantis* CP.B2^T, expand the environmental conditions under which we could conceivably expect growth, this does not necessarily account for the physicochemical ranges where we detect *Venenivibrio* (pH 1.2-9.6, 17.4-99.8 °C; Figure 4.3). Amplicon sequence detection does not infer cell viability in springs where *Venenivibrio* densities are low[325], however, increasing the pool of cultivated *Venenivibrio* isolates will expand understanding of the dispersal potential and range of phenotypes possessed by the taxon. Additionally, the lack of a distance-decay pattern for *Venenivibrio* populations within the TVZ, and the distribution of *Venenivibrio*-assigned OTUs across all geothermal fields in the region, substantiates the pervasiveness of the genus within Aotearoa-New Zealand.

Twenty 'hotspots' of increased *Venenivibrio* read abundance were identified in the TVZ that may serve as reservoirs to facilitate local dispersal. The planktonic lifestyle of *Venenivibrio*, corroborated by our exclusive use of water column samples[195], the source material for the isolation of CP.B2$^{T\ 138}$, and annotation of a full suite of flagellar and chemotaxis genes in the

genome, likely contributes to this dispersal via subsurface geothermal aquifers and/or water vapour, enhanced by increased read abundance of the taxon in springs hot enough to generate steam (50-70 °C). Interestingly, *Venenivibrio* was not reported as an abundant community member of geothermal spring sediments and sinters (*n*=138) from the TVZ[9], which supports the planktonic lifestyle model promoting gene flow at a local scale.

An intensive search for *Venenivibrio* in both amplicon and metagenomic datasets did not yield any noteworthy results from outside Aotearoa-New Zealand. While a small number of global metagenomes initially classified trace levels of sequences as *Venenivibrio*, comprehensive interrogation refuted their classification within this genus. Alignment of global metagenomes to Hydrogenothermaceae genomes corroborated their lack of *Venenivibrio*, with MAGs created from these samples having only ≤80 % similarity to *V. stagnispumantis* CP.B2[T]. In general, reduced microbial biomass in geothermal spring water columns, compared to associated sediments[172], has fostered a sampling bias of these ecosystems. While this bias could partially explain the absence of reported *Venenivibrio* in datasets outside Aotearoa-New Zealand, dominant planktonic taxa are usually detectable (in lower abundance) within sediments from the same geothermal feature[93,172,173]. Therefore, despite prior evidence supporting the aeolian transport of some cosmopolitan taxa to reach geographically isolated habitats over geological time[107,108], our findings show *Venenivibrio* has not adhered to this process of community assembly.

We propose that dispersal of *Venenivibrio* outside of Aotearoa-New Zealand is limited. Local-global overlap in microbial diversity depends on multiple factors including: community size, global diversity, inter-patch environmental heterogeneity, and patch connectivity[326]. Aotearoa-New Zealand exists as an isolated archipelago within the South Pacific Ocean; near the convergent boundary of the Pacific-Australian tectonic plates. The nearest substantial landmasses are New Caledonia to the northwest (~1,400 km) and Australia to the west (~1,900 km). Contrary to the comparatively small TVZ, Aotearoa-New Zealand's isolated position within the Pacific Ocean disadvantages *Venenivibrio* from quickly reaching optimal habitats offshore via atmospheric transport. Although there is ample evidence of global microbial aeolian dispersal[83,107], strong selection occurs during long-range atmospheric transport[327], with stress tolerance being an important trait to aid dissemination[284]. Neither *Venenivibrio* nor any known members of Aquificota produce spores or cysts that would increase stress tolerance and survival over prolonged periods of travel. Other

constraining factors include the prevailing westerly winds in Aotearoa-New Zealand that would drive aerosol dispersal east, where the nearest substantial land mass is the South American continent. Ocean salinity is likely to further restrict *Venenivibrio* distribution, with optimal growth of the type strain CP.B2^T observed at 0.0-0.2 % w/v NaCl. Sodium prominently featured in modelling of all 46 environmental parameters from geothermal springs that contained *Venenivibrio*, indicating diversity was affected by concentrations of this solute. Additionally, no traces of *Venenivibrio* (or indeed any Aquificota) were found in a global aerosol dataset of 596 near ground air, soil, and high-altitude microbial samples[328], indicating that substantial atmospheric transport is likely not conducive to this phylum. Intriguingly, growth of CP.B2^T in Obsidian Pool spring water (YNP) indicates that if viable cells do reach geothermal spring ecosystems outside Aotearoa-New Zealand, contemporary physicochemical conditions could allow *Venenivibrio* establishment, so long as the environmental niche does not substantively overlap with those of resident organisms. Nevertheless, the inability for the taxon to spread is exacerbated by a lack of suitable niches proximal to the TVZ. This prohibits the growth of sufficient cell numbers to create the 'stepping-stones' required to facilitate migration globally.

4.6.2 What are the specific growth requirements that facilitate *Venenivibrio* to be locally abundant, but globally stranded?

Maximal diversity of *Venenivibrio* occurs in reducing, hypoxic geothermal springs, with elevated concentrations of arsenic, antimony, and hydrogen sulfide, and with a limited pH and temperature range. The apparent specificity of *Venenivibrio*'s preferred growth environment was confirmed by the identification of ORP and turbidity as the main parameters associated with increased read abundance of the taxon, and reinforces dispersal limitation beyond geothermal fields through strict selection of appropriate environmental niches. The association with turbidity is slightly confounding as the most parsimonious explanation is that *Venenivibrio* associates with particles, an observation not supported by a recent study of TVZ sediments[9]. However, widespread distribution of *Venenivibrio* in the TVZ was also observed in springs far outside previously reported physicochemical optima[179] and was corroborated by our growth range re-evaluation of the type strain, *V. stagnispumantis* CP.B2^T. Either initial characterisation did not capture the full growth capabilities of the type strain, and/or there is much more diversity and physicochemical preferences at species/strain level represented by the 111 *Venenivibrio*-assigned OTUs found in our analysis. Even if sequencing error may have inflated this OTU number[329], we recently

isolated six divergent strains of *Venenivibrio* spp. from Aotearoa-New Zealand geothermal springs with disparate phenotypic traits and morphology (unpublished data), confirming diversity within the genus is currently underrepresented. Nevertheless, springs abundant with *Venenivibrio*-assigned reads still confirmed a preference for limited physicochemical regimes, in particular those with the infrequently encountered pH 4-6[253], further supporting that the genus requires a specialised habitat to thrive.

These specific growth characteristics were substantiated by genome annotation of *V. stagnispumantis* CP.B2[T], which demonstrated a streamlined metabolism capable of both autotrophy and microaerophily. The obligatory reliance on hydrogen as an electron donor (via Type 1b and 2d NiFe-hydrogenases) suggests a lack of metabolic versatility to utilise other energy sources that are often present in hot springs, with no fully characterised pathway for sulfur oxidation found. Comparative genomics within the Hydrogenothermaceae validated this metabolically constrained lifestyle, with *Sulfurihydrogenibium* spp., in particular, encoding three enzymes/pathways (citrate synthase as part of the oxidative TCA cycle, a full SOX system for thiosulfate/sulfur oxidation, and a cytochrome-cbb3 terminal oxidase) conspicuously absent from the CP.B2[T] genome. Contrasting the prominence of planktonic *Venenivibrio* in TVZ spring water columns over sediments[9,195], *Sulfurihydrogenibium* spp. also seem to dominate within filamentous mat communities in geothermal areas[141,150,155,294,330]. This dichotomy in growth media between the genera, accompanied by the additional metabolic flexibility found in *Sulfurihydrogenibium* spp., could have formed alternate environmental niches, a mechanism that can aid speciation[288]. The reduced number of hydrogenases in *Sulfurihydrogenibium* spp. isolated from Yellowstone National Park[292], for example, highlights that these taxa appear to have dissimilar metabolic strategies. These differing traits suggest environmental selection and niche differentiation may have induced ancestral divergence, followed by speciation, within the Hydrogenothermaceae family, which was either driven or reinforced by geographic isolation of these habitats.

4.6.3   Is *Venenivibrio* an endemic genus to Aotearoa-New Zealand?

Identification of *Venenivibrio* in 686 amplicon and 16 metagenomic samples from Aotearoa-New Zealand geothermal springs categorically confirmed the presence of the genus in the country; with four of the derived MAGs having >95 % similarity to *V. stagnispumantis* CP.B2[T]. The type strain, CP.B2[T], was isolated from Champagne Pool, Waiotapu, in the TVZ[138], and while a recent campaign targeting isolation of new *Venenivibrio* strains from

TVZ springs yielded six isolates (unpublished data), no other strain from either genus or species has been validly characterised. The first evidence of this genus was a clone reported in 2003 from a geothermal spring in Kuirau Park, Aotearoa-New Zealand[331] and since then, all 16S rRNA gene sequences identified with ≥95 % nucleotide similarity have originated from Aotearoa-New Zealand[187,188,195]. A phylogenomic analysis of Aquificota genomes further confirmed CP.B2[T] as a distinct genus within the Hydrogenothermaceae, with only Aotearoa-New Zealand-generated MAGs included within this clade. All other global genomes and MAGs from the family placed within either the *Sulfurihydrogenibium*, *Persephonella,* or *Hydrogenothermus* genera, strengthening the domestic exclusivity of *Venenivibrio* to Aotearoa-New Zealand.

A recent study analysing 36,795 bacterial and archaeal genomes from ~7,000 locations around the world found most species (and even strains; average nucleotide identity ≥99.9 %) are globally distributed, but suggested continental-scale endemism does occur at a sub-strain taxonomic level[109]. Hot spring and subsurface lineages, however, displayed the slowest rates of dissemination in this analysis[109], reinforcing the concept that these ecosystems can act as 'isolated islands' of microbial evolution[86]. Endemism has long been suggested for thermophiles in geothermal springs, where physical constraints and extreme physicochemistry create discrete, isolated microbial islands[6,8], but no evidence of genus-level endemism has been reported. Given the specific growth requirements of *Venenivibrio*, multiple dispersal limitations, and the geographical setting of Aotearoa-New Zealand, it is plausible for endemism to occur within this genus. It is also important to highlight that time has been proposed as a main influence on building microbial community diversity in any ecosystem[109,332]. Indeed, a molecular clock analysis suggests *Venenivibrio* diverged from its sister genus *Sulfurihydrogenibium* ~67 mya, approximating a similar timeline to when the Zealandia landmass split from Gondwana via seafloor spreading[333]. Perhaps the geological age of Aotearoa-New Zealand has not allowed sufficient time for this specialised microorganism to disperse onto the global stage; owing to the multitude of hurdles it must overcome to successfully colonise a new habitat. This scenario, however, does not account for the distribution of *Sulfurihydrogenibium* globally (including Aotearoa-New Zealand, albeit at lower abundances) and suggests other factors such as niche contraction have contributed to this apparent endemism of *Venenivibrio*. Although it is possible ecosystems harbouring *Venenivibrio* populations outside of Aotearoa-New Zealand remain to be sampled, or that some congruence between 16S rRNA genes of Hydrogenothermaceae genera

may hide distribution of this taxon when amplicon sequencing is used as a proxy for designating species, our study clearly identifies *Venenivibrio* as a tangible endemic genus within Aotearoa-New Zealand. Given the considerable research on global geothermal springs, particularly from YNP, and the fact we have shown the type strain can grow in global spring water, it seems reasonable to expect reports of *Venenivibrio* if the genus was already inhabited elsewhere.

4.7 SUMMARY

Venenivibrio is the dominant bacterial genus found in the microbial communities of Aotearoa-New Zealand geothermal spring water. While the taxon occupies an extreme and isolated niche, local reservoirs are able to facilitate widespread dispersal across the TVZ and tolerance to environmental conditions outside of optimal ranges. Despite globally sourced geothermal springs supporting *Venenivibrio* growth, the taxon has failed to contemporarily distribute beyond Aotearoa-New Zealand, resulting in the establishment of an apparent endemic genus via niche differentiation and allopatric speciation.

Microbial communities are dynamic and environmental fluctuations over time also affect microbial behaviour, ecology and evolution[247,283]. Future studies investigating the longitudinal response of *Venenivibrio* to conditions encountered during dispersal (e.g., environmental stressors and other competitors) would provide a more complete picture of population dynamics. It was previously suggested *V. stagnispumantis* CP.B2[T] was adapted to a moderately acidic environment, hinting that a wider range of pH tolerance precluded the taxon's colonisation of Champagne Pool[179]. Certainly, our results show that while narrow pH and temperature optima increase abundance of the taxon in specific spring environments, tolerance of broader environmental conditions do exist within the genus, signifying the importance of future investigation at species- and strain-level. Questions remain over the timeline of diversification within the Hydrogenothermaceae, with preliminary molecular dating suggesting divergence between *Venenivibrio* and *Sulfurihydrogenibium* occurred ~67 mya, and how other family members have successfully managed intercontinental dispersal. Biogeographical patterns of *Sulfurihydrogenibium* have been linked to major geological events in the past[79], so elucidating the role of historic events (i.e., the separation of Zealandia from the Gondwana supercontinent and the Oruanui supereruption of the Taupō

volcano)[333,334], tracing gene flow among populations, and determining recombination efficiency will not only further establish the concept of endemism for *Venenivibrio*, but also enhance our understanding of microbial evolutionary processes.

4.8   SUPPLEMENTARY INFORMATION

Supplementary methods (Section D.1) and results (Section D.2), including Figures D.1-D.7 (Section D.3) and Tables D.1-D.21 (Section D.4), can be found via the Supplementary Information in Appendix D supporting this manuscript.

4.9   DATA AVAILABILITY

The 925 raw amplicon sequences analysed in this manuscript from the 1,000 Springs Project can be found under the study accession PRJEB24353. Shotgun sequencing of the *V. stagnispumantis* CP.B2^T genome was deposited in the European Nucleotide Archive under the accession ERR10117592, with the assembled genome deposited in the Genomes Online Database (GOLD Analysis ID Ga0311387) for associated annotation with the Integrated Microbial Genomes system (IMG Taxon ID 2799112217). Accessions for the 16S rRNA gene sequence and metagenomic search results (Table D.14, Table D.15, and Table D.16), and local and global metagenomes analysed (Table D.2 & Table D.17) are available in the Supplementary Information. Aotearoa-New Zealand MAGs (Table D.19) are available under Project ID TAONGA-AGDR00025 in the Genomics Aotearoa Data Repository (https://doi.org/10.57748/vpk8-zp44). All custom R scripts used for analyses and figures are available at https://gitlab.com/morganlab/collaboration-1000Springs/1000Springs, under the venenivibrio_manuscript directory.

CHAPTER 5

SUMMARY, CONCLUSIONS, & FUTURE WORK

5.1 THESIS SUMMARY & CONCLUSIONS

This thesis reports stochastic and deterministic drivers of microbial community and population assembly through both space and time, using a total of 1,088 geothermal spring water columns from within the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand. The comprehensive dataset used for these analyses, with sufficient sampling density and standardised methodology, is the first-of-its-kind that enabled a robust, spatiotemporal, and statistical association of planktonic geothermal taxa with broad physicochemical gradients at a regional level. The data was collected as part of the 1,000 Springs Project, an MBIE-funded multidisciplinary research project that aimed to measure and characterise the indigenous microbial diversity of geothermal spring ecosystems in the TVZ. All information gathered (i.e., bacterial, archaeal, physicochemical, and locational) is stored in a custom-made Amazon Relational Database Service (RDS) that is publicly presented through an Amazon Elastic Compute Cloud (EC2) web server at https://1000springs.org.nz/, and provides a unique resource for assessing ecosystem health and function, biotechnology potential, and conservation needs of these novel ecosystems.

5.1.1 Hypotheses & objectives

The literature review for this thesis identified gaps in the study of microbial biogeography, particularly in using geothermal ecosystems to ascertain mechanisms driving extremophilic diversity. Dissimilar environmental variables have been reported as controlling microbial diversity in extreme habitats[88,92,137], reinforcing the need to collect both adequate sample number and physicochemical parameters, with standardised methodologies, to statistically describe community assembly processes. It was particularly evident from previous research[9,92,159] that most studies have relied on geothermal solid samples (i.e., soils, biomats, sinters, and/or streamers), rather than water columns of geothermal features, when describing microbial biogeography, likely due to the ease of sampling and rate of DNA recovery. However, the heterogeneity of terrestrial soil microbial ecosystems confounds attempts to distinguish true taxa-geochemical associations[57,58]. Many global geothermal systems,

including the TVZ in Aotearoa-New Zealand, remained extensively unexplored, providing unique biomes to further chronicle novel diversity and ecosystem behaviour on the planet. Limitations associated with sole use of the 16S rRNA gene as a marker for designating taxonomy were also outlined in the literature review, with complementary analysis using whole genome sequencing recently advancing traditional ecosystem assessments.

To address these shortcomings, three hypotheses were proposed. The first hypothesis (H1), that *extremophiles exhibit spatial biogeographical patterns and are predominantly governed by contemporary environmental conditions and dispersal limitation*, was tested using microbial, physicochemical, and locational data generated by the 1,000 Springs Project. This dataset provided a unique opportunity to assess if patterns of extremophilic diversity existed in the TVZ using a sizable 1,019 geothermal spring samples. A study of this scale was unprecedented for geothermal springs globally and allowed statistical determination of selection and dispersal limitation as potential drivers contributing to community structure. The second hypothesis (H2), that *temporal patterns of microbial biogeography remain consistent in geothermal springs if physicochemical conditions are stable*, was based on previous global and TVZ research reporting microbial community assemblages only changed in both composition and abundance due to a significant change in the local physicochemical environment. As temporal studies from the TVZ were restricted to single springs, a substantial subset of the geothermal springs studied for the first hypothesis (n=31) was further analysed over time to assess if indeed a disturbance was needed to induce temporal disparity in resident microbial communities. A range of potential disturbances were investigated (i.e., short-term, long-term, natural, anthropogenic, and seasonal), along with springs that had no known physicochemical interruptions. Additionally, to tease apart the effect of pH on microbial communities through time, springs with comparable temperature across the pH range found in TVZ systems were scrutinised. Finally, the third hypothesis (H3), that *individual population diversity and distribution can be controlled by different processes to microbial communities*, was assessed by focusing on the most abundant and prevalent taxon found in geothermal features from across the TVZ. Individual diversity patterns, distribution, and controlling mechanisms were explored for this population, with whole genome sequencing (through a monoculture of an isolate and environmental metagenomic samples) also analysed to expand on putative function and phylogeny.

5.1.2 Chapter 2 Summary | Microbial biogeography of 925 geothermal springs in Aotearoa-New Zealand

The goal of this chapter was to explore potential biogeographical patterns through space using TVZ geothermal habitats as model systems, and establish if contemporary environmental conditions drove microbial community assembly in these ecosystems, with main findings published in *Nature Communications*[195]. Amplicon sequencing of the 16S rRNA gene from 925 individual geothermal spring water columns generated 28,381 OTUs from 43,202,089 quality-controlled reads (with a mean of 43,905 reads per sample). Average OTU number per spring community was 386, with a range of 49-2,997 OTUs. Evenness of the microbial communities also varied, with Shannon diversity index calculated as 1.1-7.3 across all samples. Prokaryotic community composition was dominated by the phylum Aquificota (genera *Venenivibrio*, *Hydrogenobaculum*, and *Aquifex*), in both abundance and prevalence, followed by Pseudomonadota (genus *Acidithiobacillus*). Individual taxon-defined environmental niches were identified through differential abundance statistics, with exemplar geochemical signatures within some geothermal fields indicative of community composition. Unequivocally, pH was the primary driver of microbial diversity and community complexity within geothermal habitats on a spatial scale, demonstrated through multiple statistical testing (including linear modelling, multiple regressions, non-parametric analysis of variance, correlation analyses, analysis of similarities, and constrained correspondence analysis). The effect of pH on these communities, however, was only significant at temperatures <70 °C, highlighting those multiple parameters work in synergy to constrain microbial diversity in geothermal environments. Although a distance-decay pattern indicated dispersal limitation for microbial communities across the entire region, the finding that 293 adjacent community pairs exhibited up to 100 % dissimilarity is illustrative of the stark spatial heterogeneity that exists within individual geothermal fields, and suggests environmental selection drives community composition at a localised scale.

While comprehensive, the data presented in this chapter should be perceived as a snapshot of the potential novelty that exists in these understudied, extreme ecosystems on a spatial scale. There was no analysis of longitudinal diversity patterns that would provide a more holistic representation of ecosystem behaviour. Due to the considerable number of samples collected, discussion focused around dominant taxa that were both abundant and widespread across the TVZ, with no identification of rare biosphere members (<0.1% average relative abundance) and associated ecological niches. The findings presented here also assumed putative function

through taxonomic and phylogenetic classification, and focused on community-level diversity patterns and controls rather than individual populations. Measurement of functionality in providing a real-time assessment of habitat performance was also excluded, with reports of geothermal spring function (via transcriptomics) displaying contrasting influences on community diversity against gene-based analysis[167]. This reinforces the dissimilarity that can occur between units of measurement for assessing community structures, highlighting that assumptions must be made when comparing microbial ecosystems presented with disparate measurements. Furthermore, the addition of extraneous taxa via sample contamination is also a consideration, in particular for some low abundance microorganisms. A prime example of this was presence of the acne-inducing[335] *Cutibacterium* (formerly *Propionibacterium*)[336] in the dataset, which was found in 0.04 % of all microbial sequences produced from the 925 springs. As screening dominant strains indicated that the vast majority of taxa were geothermal in origin, the impact of these contaminants was likely minimal to overall analyses for this chapter. Thus, it was decided against filtering human commensal taxa from the dataset, as many of the springs (particularly those <45 °C) are used for bathing and so, these taxa could not be discriminated as sample contaminants.

It is very clear the 16S rRNA gene can only tell us so much about the microbial community inhabiting a ecosystem[337,338]. Surprisingly, Archaea were not measured as dominant and widespread across these extreme environments. While it has been suggested that Bacteria can dominate hot spring microbial communities[158], the universal primer set used in this study is known to be less favourable to Archaea[232], even though it had the greatest 16S rRNA gene coverage at the time of study design. Investigations into other biogeographical processes, such as diversification and ecological drift, are also limited by sole use of a marker gene. Ecology is directly linked to evolution, and evolutionary changes can arise in response to local selection pressures and population dynamics[32]. Accordingly, changes in genotype should not be overlooked when examining ecosystem performance, which can be masked by slow evolution rates of the 16S rRNA gene[113], or dormancy, particularly in hot spring taxa[104,114]. Whole genome sequencing or flow cytometry can overcome some of these obstacles and can demonstrate novel population lineages becoming evolutionarily independent of each other within the same environment[7], or directly quantify stochasticity in populations[122]. While it is noteworthy to acknowledge the universal limitations inherent to 16S rRNA gene analysis, such as relic DNA and variable copy number inherently inflating sequence read presence and/or abundance[337,338], this chapter highlights the potential diversity

and habitat heterogeneity these springs accommodate on a large but accessible geographical scale. Large datasets, while initially required to develop a baseline of putative diversity and function in novel microbial ecosystems, need to be recognised as a starting point for more targeted research. There is a danger of over-reaching, so use of this chapter as a springboard for more targeted and detailed investigation using temporal, genomic, and/or population approaches is advised.

5.1.3   Chapter 3 Summary | Temporal dynamics of geothermal microbial ecosystems in Aotearoa-New Zealand

In this chapter, I sought to address the second hypothesis of the thesis that stated temporal patterns of geothermal microbial biogeography remain consistent if physicochemical conditions are stable. To test this hypothesis, 115 samples from 31 individual features were analysed over a 34-month period to determine longitudinal microbial variation in three categories of geothermal ecosystems. These categories were: control sites, with no known physicochemical variation; disturbed sites, where resident geothermal communities were either naturally or anthropogenically interrupted over short or long-term phases; and, pH sites, where microbial variation was monitored in features from across the pH scale found in the TVZ (with temperature differential reduced by all features in this category having ~60-70 °C). Amplicon sequencing of the 16S rRNA gene generated 919 OTUs across 13 samples for the control sites, 2,461 OTUs across 14 samples (natural, short-term disturbance) and 7,569 OTUs across 25 samples (anthropogenic, long-term disturbance) for the disturbed sites, and 1,170 OTUs across 63 samples for the pH sites.

Overall, findings indicated that taxa abundances fluctuated temporally across all categories of geothermal features; with stochastic or unexplained variation still occurring across a wide pH range and irrespective of disturbed environmental conditions by either natural or anthropogenic influences. This leaves the second hypothesis that biogeographical patterns through time in geothermal habitats need a physicochemical change to induce microbial variation unproven. Stochastic fluctuation, however, was reduced with increased stability in the source fluid, even across seasons, as demonstrated by the control geothermal features. This proposes that microbial communities of geothermal features with deeply sourced hydrothermal fluids are sustained by a more constant environmental niche, which consequently promotes a stable ecosystem structure. Additionally, significant variation in community composition, rather than abundance levels, did need a substantial fluctuation in

physicochemistry of the immediate local environment of the microbial community. This was clearly evident in both Inferno Crater Lake and Waimangu Stream sites at the Waimangu geothermal field, where not only did a significant temperature change in the lake (41.0 to 69.1 °C) induce almost complete community turnover from *Acidithiobacillus* to Sulfolobaceae, but these Sulfolobaceae taxa were introduced to stream communities ≤2,000 m downstream of the overflow confluence point into the stream (with corresponding changes of pH ≤4.4). Despite these extensive physicochemical changes, and corresponding introduction of foreign taxa at the phylum level to resident communities, the effect of the disturbance was short-lived, with sufficient and continuous re-colonisation from the stream source returning microbial assemblages to pre-disturbed states. Indeed, we know Inferno Crater Lake also returns to an *Acidithiobacillus*-dominated community once temperature cools at the end of a cycle every ~30-40 days[139], further highlighting the speed and resilience of microbial communities to re-stabilise after a major physicochemical disturbance to the local environment ends.

Interestingly, alpha and beta diversity of Inferno Crater Lake and Waimangu Stream sites did not reflect the variation evident in taxonomic and physicochemical analyses. This contrasted the second disturbed geothermal field studied at Waikite, where diversity analyses demonstrated a notable change to some microbial communities in the first timepoint measured after a weir was constructed at the outlet to the wetland. Even though temperature also cooled at this site by 20 °C, beta diversity only correlated with a decrease in pH from 8.6 to 7.6, indicating that a greater magnitude of change in temperature may be required than pH to impact geothermal microbial communities. Two sites within the wetland have direct geothermal fluid sources in addition to Otamakokore Stream (Waikite Features 5 and 6); while these showed an initial response to the disturbance, both alpha and beta diversity had returned to pre-disturbance levels by the final timepoint. Conversely, beta diversity analyses for stream sites (Waikite Features 3, 4, and 8) suggested microbial community structure continued to change at the end of the experiment. These results highlight two probabilities that should be considered when measuring microbial change; multiple diversity metrics (in this case taxonomic, alpha, and beta diversity) can provide contrasting results on the dynamics of microbial communities through time, and some geothermal features, like Waikite Features 5 and 6 with additional geothermal source fluid, are resilient to change and can return to pre-disturbed conditions, as corroborated by the findings from the Waimangu geothermal field.

The third category of geothermal features analysed showed that while pH was again presented as the most defining parameter of microbial community structure in these extreme habitats (as determined in Chapter 2), temperature was more temporally variable than pH in these ecosystems. This was visible in measuring both the variation in temperature and correlation with diversity metrics. Eight of the springs analysed in this category were shallow (<2 m depth), and so, are susceptible to meteoric and groundwater changes that seasonally occur in Aotearoa-New Zealand that could drive these temperature changes. Interestingly, the greatest change in relative abundances of taxa occurred in features at either end of the pH scale (pH 3 and 9), while pH 7 features demonstrated the greatest variation in alpha diversity. This is most likely a consequence of the increased number of taxa found in circumneutral conditions, which advocates conservatively using alpha diversity to identify microbial diversity variation at extreme acidity or alkalinity.

In summary, this chapter identified that geothermal features with a deeply sourced fluid input are less susceptible to physicochemical change and consequently have a more stable microbial community; this suggests source fluid (either hydrothermal, meteoric, or a mixture of both) as a main driver of temporal variation. Even though pH was again presented as the most defining parameter of microbial community structure in these extreme habitats, temperature was more variable than pH over time, but communities were more sensitive to changes in pH than temperature. Evidence also inferred geothermal ecosystems can recover from short-term physicochemical disturbances, but how we define this change is pertinent as multiple diversity metrics produced disparate results over the breadth of features analysed, implying these measures are not always representative of change. The limitations to this chapter are similar to Chapter 2: the inherent biases of solely using the 16S rRNA gene as a marker for signifying microbial change; the archaeal insensitivity of the specific primer pair used for DNA amplification; the focus of variation in community abundances at the phylum level which could fail to recognise finer taxonomic change in response to both stochastic or deterministic processes through time; the lack of discussion around rare taxa and putative function of dominant species, which could lead to discovery of novel microorganisms and subsequent metabolisms; the reliance on diversity metrics and the use of relative over absolute abundances as a proxy for designating change; and, focusing on community-wide variation through time, rather than individual populations. Even though up to 46 physicochemical parameters were analysed for each sample, it is most likely that other

influences, both biotic and abiotic, not measured by this study also effect microbial community fluctuations. To obtain a more holistic representation of ecosystem dynamics, more samples over longer timeframes and with greater sampling frequency need to be taken, with whole genome analysis essential to vigorously interrogate the role of stochasticity and evolutionary drift in temporal microbial fluctuations. Standardised collection and analyses of all samples in this chapter, however, have allowed confidence in comparing longitudinal responses in these extreme habitats to physicochemical change in the local environment, and create a comprehensive argument for increasing both replicates and sampling frequency in future temporal analysis of microbial communities in geothermal environments.

5.1.4   Chapter 4 Summary | Allopatric speciation enables genus-level endemism in the bacterial phylum Aquificota in Aotearoa-New Zealand

Chapter 4 continued with the theme of investigating microbial biogeography in geothermal features across the TVZ but the focus shifted to population-level analysis, rather than whole communities. From Chapter 2 investigation of spatial diversity patterns in 925 geothermal spring ecosystems, the most abundant and widespread taxon across the TVZ was found to be genus *Venenivibrio*, from the phylum Aquificota. The third hypothesis of this thesis stated that individual population diversity and distribution can be controlled by different processes to microbial communities, and so, I wanted to determine if pH and source fluid were also responsible for driving the apparent dominance of this taxon in the region. Data generated from Chapter 2 was divided into both Aquificota- and *Venenivibrio*-assigned sequencing reads, with OTU number and read abundance of each taxonomic group scrutinised against spring physicochemistry and geography to assess diversity patterns and controlling mechanisms of their distribution in Aotearoa-New Zealand.

Only one reported species of this genus, *Venenivibrio stagnispumantis* CP.B2[T], has been validly characterised[179] and this was isolated in 2007 from a geothermal spring known as 'Champagne Pool', situated in the Waiotapu geothermal field, TVZ, Aotearoa-New Zealand[138]. The species was reported as a thermophilic, microaerophilic, chemolithoautotroph that obligatorily used hydrogen as an electron donor, with a narrow pH range of 4.8-5.8 (optimum pH 5.4). Elemental sulfur or thiosulfate were necessary in media for growth, but hydrogen could not be replaced by either sulfur species under both anaerobic or microaerobic conditions[179]. Along with investigating 16S rRNA gene diversity patterns of this taxon, it was decided to sequence the genome of CP.B2[T] to elucidate if putative genotypic evidence could

resolve the specific ecology of this microorganism. Additionally, any available metagenomic samples from TVZ geothermal springs would be studied, to confirm the presence of *Venenivibrio* in Aotearoa-New Zealand, generate metagenome-assembled genomes (MAGs) of putative *Venenivibrio* strains (if possible), and also discover if any congruence existed between classification of Hydrogenothermaceae, in particular using the 16S rRNA gene, that could be masking presence of the genus elsewhere.

V. stagnispumantis CP.B2[T] phylogenetically classified as a member of the Hydrogenothermaceae family, along with sister genera *Hydrogenothermus*, *Persephonella*, and *Sulfurihydrogenibium*[179]. The former two genera are marine-associated and are generally isolated from deep-sea, hydrothermal vents[112,290,291,339]. *Sulfurihydrogenibium*, however, is commonly found in terrestrial geothermal systems throughout the world[141,150,155,293,294], including Aotearoa-New Zealand, albeit in much lesser abundance than *Venenivibrio*[195]. An initial search for *Venenivibrio* in published literature and DNA sequence databases failed to present any evidence of the genus outside of Aotearoa-New Zealand, indicating that the country's geographic isolation could be restricting dispersal of this genus onto the global stage; with allopatric speciation within the Hydrogenothermaceae enabling genus-level endemism. This chapter also extensively searched both amplicon and metagenomic sequence databases, along with published literature, to further strengthen the assertion of *Venenivibrio* endemism to the Aotearoa-New Zealand archipelago.

To summarise the main findings, widespread distribution and abundance of *Venenivibrio* was detected in 74 % (*n*=686) of sampled geothermal spring water columns (via 16S rRNA gene sequencing) in the TVZ. Analysis of 46 physicochemical parameters from each spring indicated oxidation-reduction potential and turbidity as the primary drivers of *Venenivibrio* diversity, with increased abundance occurring at pH 4-6 and 50-70 °C, highlighting the specific environmental niche that further induces ecosystem isolation. Sodium also prominently featured in modelling of all 46 physicochemical parameters against *Venenivibrio* diversity, suggesting that this solute might be involved in niche separation of marine-based *Hydrogenothermus* and *Persephonella*, and the terrestrial-based *Sulfurihydrogenibium* and *Venenivibrio*. Oxidation-reduction potential and turbidity are also common to environments conducive to other Hydrogenothermaceae growth, so perhaps some cryptic environmental parameter remains to be identified that, in conjunction with sodium, induced and/or reinforced niche differentiation within the family. Interestingly, decreased abundances of

Venenivibrio were found in physicochemical conditions far outside reported optima (e.g., for pH and salinity), suggesting environmental tolerance of the genus exceeds that of the type strain and enables widespread distribution locally. Congruently, limited dissimilarity was observed between *Venenivibrio* populations from across the region, reinforcing the concept that dispersal is not restricted at a local scale and is facilitated by reservoirs or 'hotspots' of increased abundance of the taxon. This lack of dispersal limitation locally is in contrast with the distance-decay pattern observed for whole microbial communities across the TVZ, and demonstrated that individual populations can be controlled by different biogeographical processes than whole communities in these extreme habitats. Genomic annotation of *V. stagnispumantis* CP.B2[T] confirmed a chemolithoautotrophic metabolism typical for oxygen-limiting environments, with a surprising absence of known genes encoding sulfur oxidation. Metagenomic and phylogenomic analyses of local and global geothermal spring microbial communities established *Venenivibrio* is geographically restricted to Aotearoa-New Zealand and distinct from *Hydrogenothermus*, *Sulfurihydrogenibium,* and *Persephonella*, with only CP.B2[T] and Aotearoa-New Zealand-derived MAGs forming a monophyletic cluster within the Hydrogenothermaceae. This was corroborated by an intensive search of amplicon and metagenomic DNA sequence databases (which included at least 970,000 samples, 26.7 billion 16S rRNA gene sequences, and 12.2 petabytes of sequence data). It was concluded that while dispersal of *Venenivibrio* is not limited within the Aotearoa-New Zealand archipelago, physicochemical and geographical constraints prevent this taxon from distributing on a broader scale, resulting in the establishment of an endemic bacterial genus.

Limitations to previous chapters around the sole use of the 16S rRNA gene were managed here by use of both genomic and metagenomic analysis to complement the initial analysis of diversity patterns reported. Genome annotation of *V. stagnispumantis* CP.B2[T] confirmed the extreme and isolated niche that was proposed for optimal growth of the taxon using 16S rRNA gene analysis. Metagenomic shotgun sequencing removed biases associated with primer design and DNA amplification, and still placed *Venenivibrio* as an abundant genus within Aotearoa-New Zealand geothermal springs. Whether *Venenivibrio* is exclusively abundant to geothermal spring water columns, and not associated sediments, remains to be seen. A recent, separate study of 138 hot spring silica sinters and sediments from the TVZ did not mention any *Venenivibrio* in these samples, and indeed, community composition reported decreased numbers of Aquificota across all sites[9]. The most abundant taxa were Thermoplasmatales and Sulfolobales, reinforcing previous reports that Archaea dominate

geothermal sediments over water columns[172]. Sriaporn *et al*. (2021) did describe *Acidithiobacillus* as the third most abundant taxon (11.4 % average relative abundance) across these 138 samples, which corroborated the similar average relative abundance of *Acidithiobacillus* (11.1 %) found in water columns from this study. Clearly, studying both planktonic and sediment-associated taxa for this thesis would have been preferred and to this end, sediments were collected alongside water columns and cryogenically preserved to allow future comparison of these sample types.

Another limitation of this chapter (and previous ones) was the use of operational taxonomic units (OTUs) to define 16S rRNA gene diversity, rather than the more recent amplicon sequence variants (ASVs). OTUs were still the preferred choice for processing amplicon sequencing at the time of method development for Chapter 2, and this methodology of *de novo* clustering was continued for subsequent chapters to allow confidence in comparing the results of each section. The use of ASVs over OTUs has been strongly advocated[340,341], due to the conservative approach of OTU clustering (especially at 97% sequence similarity) potentially eliminating novel taxa and the lack of clustering in ASVs allowing for more robust, multi-study comparison. It has been suggested, however, that ASVs could artificially inflate diversity in an ecosystem, while also incorrectly identifying sequencing error as new taxa[22], highlighting that limitations are inherent for any type of sample processing. Indeed, the next-generation DNA sequencing technology engaged by this study was the Ion Torrent PGM system, in which an initial, relatively high rate of sequencing error allowed the alternate Illumina platform to increase in popularity[342,343]. Error rates have since decreased for Ion Torrent sequencing[344,345], but the widespread use of Illumina use in ecology studies warrants a reservation when comparing this study to others based on different sequencing platforms.

Collectively, these findings from all chapters reinforce the supposition that multiple processes combine to shape diversity patterns of microorganisms through both space and time. Microbial behaviour, ecology, and evolution are dependent on deterministic and stochastic controls that can alternate depending on the scope of study, including: scale (i.e., spatial or temporal); taxonomic definition (i.e., community or population); sample type (i.e., solid or aqueous); geography (i.e., local, regional, or global); and habitat type. While the effect of traditional biases for microbial biogeography studies, such as primer design, DNA extraction, and sequencing, were minimised by the use of identical methodologies for collecting and processing of all samples, this thesis highlights the importance of using replicates, increased

sampling frequency, and multiple diversity measures to obtain a cohesive representation of core geothermal microbial communities.

5.2 FUTURE WORK

This research provides a comprehensive dataset that should be used as a foundation for future, more directed studies, focusing on some of the limitations already discussed in this chapter. It is likely that environmental changes actively select for finer-scale phenotypic or metabolic traits not detectable via the taxonomic analyses (i.e., 16S rRNA gene sequencing) undertaken in this study[275]. Additionally expanding focus on genome-wide analysis could infer other factors beyond physicochemistry and dispersal limitation that induce both spatial and temporal change in thermophilic taxa, for example, diversification and ecological drift, through tracing gene flow and determining recombination efficiency[13,122,346,347].

In all chapters of this thesis, putative metabolic function was assumed from gene-based analysis. Further work is needed to accurately define microbial metabolism and function in these ecosystems as an extended description of population and community dynamics, either *in situ* in geothermal features through transcriptomics[348], in a controlled, chemostat-based approach in the laboratory[184], or using 16S rRNA genes or metagenomics to infer predicted function from curated databases[349,350]. It is important to note that these functional databases are not yet at the same scale as those for taxonomic markers[351]. Some ecosystem capability through vital and/or novel taxa may go unnoticed if functionality alone is used to define microbial behaviour in any habitat. Determining actual function and cellular state, rather than genetic potential, of geothermal ecosystems, could lead to increased biotechnological application for novel enzymes from extremophiles, which have previously been used in pharmaceuticals, molecular biology, food products, industrial applications, greenhouse gas capture, and bioremediation[352,353]. Additionally, for *Venenivibrio*, it would be interesting to ascertain what genes were upregulated in response to certain environmental parameters that might explain niche preferentiality for the taxon, for example, by using transcriptomics to measure phenotypic response during arsenic stress[354] or resolve the uptake of sulfur into the cell[355]. Also, the full high affinity molybdate uptake system (modABCD) was not annotated in the *V. stagnispumantis* CP.B2[T] genome, whereas the operon was present in all other Hydrogenothermaceae genomes analysed. This could expand on the apparent lack of

mixotrophy within *Venenivibrio*, or give *Sulfurihydrogenibium* a competitive advantage in colonising a habitat amenable to both taxa, as molybdenum is an essential co-factor in molybdoenzymes used for bacterial metabolism (e.g., formate and sulfite dehydrogenases)[356]. CP.B2^T also had fewer copper exporting ATPases (copB), again compared to other family members, so using a function-based assessment like transcriptomics could narrow defining characteristics that have enabled the apparent endemism of the genus to Aotearoa-New Zealand.

Further, this thesis focused only on prokaryotic (i.e., archaeal and bacterial) communities and populations. Geothermal microbial ecosystems also include eukaryotic microorganisms, such as microalgae[148] and fungi[357], as well as viruses[358], so targeted research on these taxa would complement findings presented through this study. Indeed, the data generated by Chapter 2 has already been utilised to describe protistan diversity in TVZ geothermal springs[156], adding to a vital, but under-researched, component of ecosystem services in these habitats. As already stated in this chapter, the original Earth Microbiome Project primers (515f-806r) used by this study are known to contain biases towards Bacteria, and so, could have potentially eliminated detection of many Archaea in these ecosystems, in particular Thermoproteota and Nitrososphaerota[232]. It would be prudent for future study, in particular for geothermal sediments which are known to harbour increased abundance of Archaea[172], to use the modified 515f-806rB primer pair that are highly concordant with the original primer set to allow cross-comparison between studies[232]. An additional archaeal-specific primer set could also be used, if the focus was on identifying rare taxa (and associated abundances) underrepresented by the current study[359], which could also be mixed into one reaction mix with universal bacterial primers to ensure no partiality for either domain. Particularly, these primers could be used in conjunction with digital PCR to obtain absolute taxa abundances[360]. The collection and cryogenic storage of associated sediments from each geothermal spring water column collected for this study, along with the water column DNA already extracted and archived, will provide a unique opportunity to assess these shortcomings[361]. Ideally, using both universal and archaeal-specific primers would give the best comparison to the current dataset and representation of realistic archaeal diversity in TVZ geothermal springs. Metagenomic sequencing would also remove the biases of any primer set, but as sequencing and analysing metagenomes from >1,000 sites is not likely attainable for the immediate study of Aotearoa-New Zealand geothermal systems and would reduce biodiversity resolution of target assemblages, complementing 16S rRNA gene data with a reduced number of sample

metagenomes would aid confidence in defining microbial ecosystem behaviour in these environments.

While the focus of Chapter 4 was on the most abundant genus found in TVZ geothermal water columns, assessment of other populations, preferably both abundant and rare, remains to be examined. We know that some extremophiles were missed by the study design, for example, the novel genus and species from the Thermoproteota phylum, *Zestosphaera tikiterensis*, and its symbiotic nanoarchaeote, *Candidatus* Nanoclepta minutus. Both were recently isolated from anaerobic enrichment cultures taken from the Tikitere geothermal field in the TVZ[362]. Even though some taxa have either evaded detection or have not appeared in the broad-range analysis from this thesis, the dataset can be bioprospected to find specific geothermal niches with the appropriate physicochemical environment conducive to growth of a target microorganism. Enrichment of samples from geothermal springs with increased 16S rRNA gene abundance of the desired taxon can also aid in isolating pure cultures, which has already been utilised in the recent isolation of six divergent strains of *Venenivibrio* spp. from the TVZ (unpublished data). Obtaining and characterising novel isolates, or isolating co-cultures, is still imperative to determine microbial phenotypes and physiology, improve gene-based annotations, discover innovative biotechnological applications, decipher complex biotic interactions, and eliminate relic DNA as a contributor to community structures[337,363–365]. Additionally, elucidating drivers of the abundant acidophile *Acidithiobacillus* in an alternate physicochemical niche to *Venenivibrio*, in conjunction with the genome-wide analysis of the taxon in TVZ sinters and sediments[9], would enhance understanding of extremophilic biogeography in the region. Taxa from the phylum Armatimonadota were also found to be widespread in TVZ geothermal springs[195], but this novel phylum, first identified in 1998[158] and characterised in 2011[366], remains under-explored and -described[367]. The diversity and biogeography of Armatimonadota could be analysed in detail using this dataset, which could then inform future research on the metabolism, physiology, and ecosystem function of this novel phylum. With the exception of *Venenivibrio*, this thesis primarily focused on large-scale community patterns of diversity to gain initial insight into these novel ecosystems, providing a fundamental infrastructure for future, more targeted research on individual populations and taxa.

Indeed, population analysis via the genus *Venenivibrio* has introduced many questions. The apparent endemism of a bacterial genus is unprecedented. *Venenivibrio* presents an easily

accessible and tractable genus to validate this finding, as the overwhelming presence and abundance in Aotearoa-New Zealand geothermal springs preclude under-sampling and/or rarity as rebuttals for endemic classification. While allopatry has clearly played a part in the evolution of the genus within the Hydrogenothermaceae family, it is unlikely to be the sole cause of contemporary endemism. Other constraining factors, like physical dispersal mechanisms and environmental stress responses, need to be explored in more detail to assess key metabolic and physiological traits that either drive and/or maintain the apparent endemism. Multiple modes of transport could be investigated (i.e., aerosols, spring drainage streams, and/or geothermal aquifers), with the absence of *Venenivibrio* internationally likely reflecting a strong negative selection pressure (via pH, salt, and starvation extremes) in long-range atmospheric transport[327]. These barriers to migration are further enhanced by a lack of suitable environmental stepping stones immediately beyond Aotearoa-New Zealand, and this aspect could be more conclusively examined by including the nearest geothermal springs outside of the archipelago (e.g., in New Caledonia or Tonga)[368,369] for future research on the genus. Indeed, traces of the taxon were found in geothermal features on Raoul Island, as highlighted in Table D.2, Appendix D[370]. This island, situated ~1,100 km northeast of the North Island, is the northernmost outpost of Aotearoa-New Zealand and reinforces the concept that local dispersal of *Venenivibrio* is not restricted. Results from this chapter also show that while narrow pH and temperature optima increase abundance of *Venenivibrio* in specific spring environments, tolerance of broader environmental conditions do exist within the genus, signifying the importance of future investigation at species- and strain-level.

Additionally, the timeline of diversification within the Hydrogenothermaceae needs to be validly elucidated, with preliminary molecular dating suggesting divergence between *Venenivibrio* and *Sulfurihydrogenibium* occurred ~67 mya[318]. Most significantly, deciphering how other family members have successfully managed intercontinental dispersal, along with the dispersal-limiting mechanisms of *Venenivibrio* preventing dissemination onto the global stage, will be essential to explaining the isolated ubiquity of the genus in the Aotearoa-New Zealand archipelago. Investigating the diffusivity (i.e., a measure of dispersal rate over evolutionary time scales)[109] of the taxon would complement this avenue of exploration, as would more extensive comparative genomics of the Hydrogenothermaceae to differentiate metabolic capability between the genera. Elucidating the role of local historic events (i.e., the separation of Zealandia from the Gondwana supercontinent via seafloor spreading and the Oruanui supereruption of the Taupō volcano)[333,334], combined with the legacy of other

Hydrogenothermaceae[79], will not only further establish the concept of endemism for *Venenivibrio*, but also enhance our understanding of microbial evolutionary processes.

Finally, the data and findings produced by this thesis are available to researchers and bioinformaticians to support model development of microbial community establishment and structure, ecosystem health, and uniqueness[119,371–373]. The study provides sufficient data to robustly approximate community assembly using typical taxon-geochemical associations and geographical influence. The research also complements microbial ecology studies of non-extreme biomes in Aotearoa-New Zealand[374–377], and increases knowledge on the value of indigenous biodiversity in the country. Assessing ecosystem health and the effect of environmental change in geothermal springs is complex, in comparison to non-extreme environments[227,371], as the range of physicochemistry and microbial taxa found in these habitats make establishing a baseline difficult. This study provides a starting point, however, to determine typical microbial assemblages and corresponding physicochemical types[378], in particular to specific geothermal fields, that could proxy for environmental health and resilience in these novel habitats.

This study clearly provides a springboard to assess the cultural, recreational, and resource development value of the microbial component of geothermal springs, both in Aotearoa-New Zealand and globally. Not only are DNA sequences and associated metadata available to the scientific community in INSDC databases, the microbiological, physicochemical and locational information of each geothermal feature is presented in a user-friendly website (https://1000springs.org.nz/), increasing the scope of the study to several stakeholders. These include the general public in expanding science outreach, landowners and tourism operators of geothermal sites, commercial partners, geothermal power industry, government agencies, and Māori groups (under mana whenua). The findings of each chapter reiterate that microorganisms are taonga (i.e., valuable natural resources), with many of the features included in this research occurring on culturally important and protected land for Māori. Therefore, this or follow-on future projects provide an avenue for exploration of indigenous knowledge, while assisting in management, conservation, and/or protection effects of national government agencies to maintain these ecosystems as part of Aotearoa-New Zealand's treasured endemic species.

APPENDIX A

REFERENCES

1. Gilbert, J. A. & Neufeld, J. D. Life in a world without microbes. *PLoS Biol.* **12**, e1002020 (2014).

2. Nemergut, D. R. *et al.* Global patterns in the biogeography of bacterial taxa. *Environ. Microbiol.* **13**, 135–144 (2011).

3. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).

4. Wilson, E. *The Diversity of Life*. (Harvard University Press, 1992).

5. Fuhrman, J. A., Cram, J. A. & Needham, D. M. Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* **13**, 133–146 (2015).

6. Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Geographic barriers isolate endemic populations of hyperthermophilic Archaea. *Science* **301**, 976–978 (2003).

7. Cadillo-Quiroz, H. *et al.* Patterns of gene flow define species of thermophilic Archaea. *PLoS Biol.* **10**, e1001265 (2012).

8. Papke, R. T., Ramsing, N. B., Bateson, M. M. & Ward, D. M. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**, 650–659 (2003).

9. Sriaporn, C., Campbell, K. A., Van Kranendonk, M. J. & Handley, K. M. Genomic adaptations enabling *Acidithiobacillus* distribution across wide-ranging hot spring temperatures and pHs. *Microbiome* **9**, 135 (2021).

10. Hollibaugh, J. T. *et al.* Seasonal variation in the metatranscriptomes of a Thaumarchaeota population from SE USA coastal waters. *ISME J.* **8**, 685–698 (2014).

11. Green-Saxena, A., Dekas, A. E., Dalleska, N. F. & Orphan, V. J. Nitrate-based niche differentiation by distinct sulfate-reducing bacteria involved in the anaerobic oxidation of methane. *ISME J.* **8**, 150–163 (2013).

12. Pernice, M. C. *et al.* Global abundance of planktonic heterotrophic protists in the deep ocean. *ISME J.* **9**, 782–792 (2014).

13. Chase, A. B. *et al.* Maintenance of sympatric and allopatric populations in free-living terrestrial bacteria. *MBio* **10**, e02361-19 (2019).

14. Brewer, T. E., Handley, K. M., Carini, P., Gibert, J. A. & Fierer, N. Genome reduction in an abundant and ubiquitous soil bacterial lineage. *Nat. Microbiol.* **2**, 16198 (2016).

15. Fortunato, C. S. *et al.* Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. *ISME J.* **7**, 1899–1911 (2013).

16. Cram, J. A. *et al.* Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J.* **9**, 563–580 (2015).

17. Ju, F. & Zhang, T. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME J.* **9**, 683–695 (2014).

18. Hermans, S. M., Buckley, H. L., Case, B. S. & Lear, G. Connecting through space and time: catchment-

scale distributions of bacteria in soil, stream water and sediment. *Environ. Microbiol.* **22**, 1000–1010 (2020).

19. Tedersoo, L. *et al.* Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).

20. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).

21. Achtman, M. & Wagner, M. Microbial diversity and the genetic nature of microbial species. *Nat. Rev. Microbiol.* **6**, 431–440 (2008).

22. Schloss, P. D. Amplicon sequence variants artificially split bacterial genomes into separate clusters. *mSphere* **6**, e00191-21 (2021).

23. Tikhonov, M., Leach, R. W. & Wingreen, N. S. Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. *ISME J.* **9**, 68–80 (2014).

24. Cohan, F. M. What are bacterial species? *Annu. Rev. Microbiol.* **56**, 457–487 (2002).

25. Koeppel, A. F. & Wu, M. Species matter: the role of competition in the assembly of congeneric bacteria. *ISME J.* **8**, 531–540 (2014).

26. Escalas, A. *et al.* A unifying quantitative framework for exploring the multiple facets of microbial biodiversity across diverse scales. *Environ. Microbiol.* **15**, 2642–2657 (2013).

27. Eren, A. M. *et al.* Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* **4**, 1111–1119 (2013).

28. Dickie, I. A. *et al.* Towards robust and repeatable sampling methods in eDNA-based studies. *Mol. Ecol. Resour.* **18**, 940–952 (2018).

29. Lear, G. *et al.* Methods for the extraction, storage, amplification and sequencing of dna from environmental samples. *N. Z. J. Ecol.* **42**, 10-50A (2018).

30. Hermans, S. M., Buckley, H. L. & Lear, G. Perspectives on the impact of sampling design and intensity on soil microbial diversity estimates. *Front. Microbiol.* **10**, 1820 (2019).

31. O'Malley, M. A. The nineteenth century roots of 'everything is everywhere'. *Nat. Rev. Microbiol.* **5**, 647–652 (2007).

32. Cordero, O. X. & Polz, M. F. Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* **12**, 263–273 (2014).

33. Schluter, D. Ecology and the origin of species. *Trends Ecol. Evol.* **16**, 372–380 (2001).

34. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME J.* **8**, 1553–1565 (2014).

35. Whittaker, R. H. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* **30**, 279–338 (1960).

36. MacArthur, R. Patterns of species diversity. *Biol. Rev.* **40**, 510–533 (1965).

37. MacArthur, R. & Wilson, E. *The theory of island biogeography*. (Princeton University Press, 1967).

38. Leibold, M. A. *et al.* The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* **7**, 601–613 (2004).

39. Jones, K. E., Blackburn, T. M. & Isaac, N. J. B. Can unified theories of biodiversity explain mammalian macroecological patterns? *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2554–2563 (2011).

40. Palmer, M. W. Variation in species richness: towards a unification of hypotheses. *Folia Geobot.*

Phytotaxon. **29**, 511–530 (1994).

41. Qian, H. & Ricklefs, R. E. Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature* **407**, 180–182 (2000).

42. Payne, N. L. & Smith, J. A. An alternative explanation for global trends in thermal tolerance. *Ecol. Lett.* **20**, 70–77 (2017).

43. Kardol, P., Fanin, N. & Wardle, D. A. Long-term effects of species loss on community properties across contrasting ecosystems. *Nature* **557**, 710–713 (2018).

44. Peters, M. K. *et al.* Climate–land-use interactions shape tropical mountain biodiversity and ecosystem functions. *Nature* **568**, 88–92 (2019).

45. 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).

46. Vellend, M. Conceptual synthesis in community ecology. *Q. Rev. Biol.* **85**, 183–206 (2010).

47. Wang, J. *et al.* Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J.* **7**, 1310–1321 (2013).

48. Chase, J. M. Towards a really unified theory for metacommunities. *Funct. Ecol.* **19**, 182–186 (2005).

49. Hubbell, S. P. *The unified neutral theory of biodiversity and biogeography*. (Princeton University Press, 2001).

50. Beijerinck, M. W. *De infusies en de ontdekking der backteriën. Jaarboek van de Koninklijke Akademie van Wetenschappen* (Müller, Amsterdam, 1913).

51. O'Malley, M. A. 'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography. *Stud. Hist. Philos. Biol. Biomed. Sci.* **39**, 314–325 (2008).

52. Curtis, T. P., Sloan, W. T. & Scannell, J. W. Estimating prokaryotic diversity and its limits. *PNAS* **99**, 10494–10499 (2002).

53. Martiny, J. B. H. *et al.* Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).

54. Prosser, J. I. *et al.* The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* **5**, 384–392 (2007).

55. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *PNAS* **103**, 626–631 (2006).

56. Bryant, J. A. *et al.* Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *PNAS* **105**, 11505–11511 (2008).

57. Tecon, R. & Or, D. Biophysical processes supporting the diversity of microbial life in soil. *FEMS Microbiol. Rev.* **41**, 599–623 (2017).

58. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* **15**, 579–590 (2017).

59. Mendenhall, C. D., Karp, D. S., Meyer, C. F. J., Hadly, E. A. & Daily, G. C. Predicting biodiversity change and averting collapse in agricultural landscapes. *Nature* **509**, 213–217 (2014).

60. Burrows, M. T. *et al.* Geographical limits to species-range shifts are suggested by climate velocity. *Nature* **507**, 492–495 (2014).

61. Helmus, M. R., Mahler, D. L. & Losos, J. B. Island biogeography of the Anthropocene. *Nature* **513**, 543–546 (2014).

62. Nemergut, D. R. *et al.* Patterns and processes of microbial community assembly. *Microbiol. Mol. Biol. Rev.* **77**, 342–356 (2013).

63. Roeselers, G. *et al.* Microbial biogeography of drinking water: patterns in phylogenetic diversity across space and time. *Environ. Microbiol.* **17**, 2505–2514 (2015).

64. Ladau, J. *et al.* Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **7**, 1669–1677 (2013).

65. Trosvik, P., de Muinck, E. J. & Stenseth, N. C. Biotic interactions and temporal dynamics of the human gastrointestinal microbiota. *ISME J.* **9**, 533–541 (2015).

66. Lauber, C. L., Ramirez, K. S., Aanderud, Z., Lennon, J. & Fierer, N. Temporal variability in soil microbial communities across land-use types. *ISME J.* **7**, 1641–1650 (2013).

67. Winter, C., Matthews, B. & Suttle, C. A. Effects of environmental variation and spatial distance on Bacteria, Archaea and viruses in sub-polar and arctic waters. *ISME J.* **7**, 1507–1518 (2013).

68. 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**, fiw253 (2017).

69. Bates, S. T. *et al.* Examining the global distribution of dominant archaeal populations in soil. *ISME J.* **5**, 908–917 (2011).

70. Lear, G., Bellamy, J., Case, B. S., Lee, J. E. & Buckley, H. L. Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. *ISME J.* **8**, 1715–1726 (2014).

71. 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).

72. 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).

73. Hermans, S. M., Buckley, H. L., Curran-Cournane, F., Taylor, M. & Lear, G. Temporal variation in soil bacterial communities can be confounded with spatial variation. *FEMS Microbiol. Ecol.* **96**, fiaa192 (2020).

74. Gautam, A., Lear, G. & Lewis, G. D. Time after time: detecting annual patterns in stream bacterial biofilm communities. *Environ. Microbiol.* **24**, 2502–2515 (2022).

75. Shade, A., Gregory Caporaso, J., Handelsman, J., Knight, R. & Fierer, N. A meta-analysis of changes in bacterial and archaeal communities with time. *ISME J.* **7**, 1493–1506 (2013).

76. Ferrenberg, S. *et al.* Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *ISME J.* **7**, 1102–1111 (2013).

77. Afshinnekoo, E. *et al.* Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Syst.* **1**, 72–87 (2015).

78. Wilhelm, L., Singer, G. A., Fasching, C., Battin, T. J. & Besemer, K. Microbial biodiversity in glacier-fed streams. *ISME J.* **7**, 1651–1660 (2013).

79. Takacs-Vesbach, C., Mitchell, K., Jackson-Weaver, O. & Reysenbach, A.-L. Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. *Environ. Microbiol.* **10**, 1681–1689 (2008).

80.   Shade, A. Diversity is the question, not the answer. *ISME J.* **11**, 1–6 (2017).

81.   Whittaker, R. H. Evolution and measurement of species diversity. *Taxon* **21**, 213–251 (1972).

82.   Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351 (2012).

83.   Monteil, C. L., Bardin, M. & Morris, C. E. Features of air masses associated with the deposition of *Pseudomonas syringae* and *Botrytis cinerea* by rain and snowfall. *ISME J.* **8**, 2290–2304 (2014).

84.   Gonzalez, A. *et al.* Characterizing microbial communities through space and time. *Curr. Opin. Biotechnol.* **23**, 431–436 (2012).

85.   Gonzalez, A., Stombaugh, J., Lauber, C. L., Fierer, N. & Knight, R. SitePainter: a tool for exploring biogeographical patterns. *Bioinformatics* **28**, 436–438 (2012).

86.   Campbell, K. M. *et al. Sulfolobus islandicus* meta-populations in Yellowstone National Park hot springs. *Environ. Microbiol.* **19**, 2334–2347 (2017).

87.   Aguilar, M., Fiore-Donno, A.-M., Lado, C. & Cavalier-Smith, T. Using environmental niche models to test the 'everything is everywhere' hypothesis for *Badhamia. ISME J.* **8**, 737–745 (2014).

88.   Xiong, J. *et al.* Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environ. Microbiol.* **14**, 2457–2466 (2012).

89.   Lin, W., Wang, Y., Gorby, Y., Nealson, K. & Pan, Y. Integrating niche-based process and spatial process in biogeography of magnetotactic bacteria. *Sci. Rep.* **3**, 1643 (2013).

90.   Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *PNAS* **104**, 11436–11440 (2007).

91.   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).

92.   Sharp, C. E. *et al.* Humboldt's spa: microbial diversity is controlled by temperature in geothermal environments. *ISME J.* **8**, 1166–1174 (2014).

93.   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).

94.   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).

95.   Logue, J. B. & Lindström, E. S. Species sorting affects bacterioplankton community composition as determined by 16S rDNA and 16S rRNA fingerprints. *ISME J.* **4**, 729–738 (2010).

96.   Perryman, S. E., Rees, G. N., Walsh, C. J. & Grace, M. R. Urban stormwater runoff drives denitrifying community composition through changes in sediment texture and carbon content. *Microb. Ecol.* **61**, 932–940 (2011).

97.   Teeling, H. *et al.* Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**, 608–611 (2012).

98.   Podell, S. *et al.* Seasonal fluctuations in ionic concentrations drive microbial succession in a hypersaline lake community. *ISME J.* **8**, 979–990 (2013).

99.   Jun, W., Oliver, T. A., Ducklow, H. W., Amaral-zettler, L. A. & Sogin, M. L. Marine bacteria exhibit a bipolar distribution. *PNAS* **110**, 2342–2347 (2012).

100.   Ho, A. *et al.* The more, the merrier: heterotroph richness stimulates methanotrophic activity. *ISME J.* **8**,

1945–1948 (2014).

101.   Andrade-Domínguez, A., Salazar, E., Del Carmen Vargas-Lagunas, M., Kolter, R. & Encarnación, S. Eco-evolutionary feedbacks drive species interactions. *ISME J.* **8**, 1041–1054 (2014).

102.   Saleem, M., Fetzer, I., Harms, H. & Chatzinotas, A. Diversity of protists and bacteria determines predation performance and stability. *ISME J.* **7**, 1912–1921 (2013).

103.   Sztajer, H. *et al.* Cross-feeding and interkingdom communication in dual-species biofilms of *Streptococcus mutans* and *Candida albicans*. *ISME J.* **8**, 2256–2271 (2014).

104.   Müller, A. L. *et al.* Endospores of thermophilic bacteria as tracers of microbial dispersal by ocean currents. *ISME J.* **8**, 1153–1165 (2014).

105.   Read, D. S. *et al.* Catchment-scale biogeography of riverine bacterioplankton. *ISME J.* **9**, 516–526 (2015).

106.   Lee, J. E., Buckley, H. L., Etienne, R. S. & Lear, G. Both species sorting and neutral processes drive assembly of bacterial communities in aquatic microcosms. *FEMS Microbiol. Ecol.* **86**, 288–302 (2013).

107.   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).

108.   Vimercati, L., Darcy, J. L. & Schmidt, S. K. The disappearing periglacial ecosystem atop Mt. Kilimanjaro supports both cosmopolitan and endemic microbial communities. *Sci. Rep.* **9**, 10676 (2019).

109.   Louca, S. The rates of global bacterial and archaeal dispersal. *ISME J.* **16**, 159–167 (2022).

110.   Stegen, J. C. *et al.* Quantifying community assembly processes and identifying features that impose them. *ISME J.* **7**, 2069–2079 (2013).

111.   Albright, M. B. N. & Martiny, J. B. H. Dispersal alters bacterial diversity and composition in a natural community. *ISME J.* **12**, 296–299 (2018).

112.   Mino, S. *et al.* Biogeography of *Persephonella* in deep-sea hydrothermal vents of the Western Pacific. *Front. Microbiol.* **4**, 107 (2013).

113.   Ochman, H., Elwyn, S. & Moran, N. A. Calibrating bacterial evolution. *PNAS* **96**, 12638–12643 (1999).

114.   Locey, K. J. Synthesizing traditional biogeography with microbial ecology: the importance of dormancy. *J. Biogeogr.* **37**, 1835–1841 (2010).

115.   Hoetzinger, M., Pitt, A., Huemer, A. & Hahn, M. W. Continental-scale gene flow prevents allopatric divergence of pelagic freshwater bacteria. *Genome Biol. Evol.* **13**, evab019 (2021).

116.   Speth, D. R. *et al.* Microbial communities of Auka hydrothermal sediments shed light on vent biogeography and the evolutionary history of thermophily. *ISME J.* **16**, 1750–1764 (2022).

117.   Johnson, J. S. *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **10**, 5029 (2019).

118.   Zhou, J. *et al.* Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *PNAS* **111**, E836–E845 (2014).

119.   Evans, S., Martiny, J. B. H. & Allison, S. D. Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME J.* **11**, 176–185 (2017).

120.   Ning, D., Deng, Y., Tiedje, J. M. & Zhou, J. A general framework for quantitatively assessing ecological stochasticity. *PNAS* **116**, 16892–16898 (2019).

121. Kurm, V., Geisen, S. & Gera Hol, W. H. A low proportion of rare bacterial taxa responds to abiotic changes compared with dominant taxa. *Environ. Microbiol.* **21**, 750–758 (2019).

122. Fodelianakis, S., Valenzuela-Cuevas, A., Barozzi, A. & Daffonchio, D. Direct quantification of ecological drift at the population level in synthetic bacterial communities. *ISME J.* **15**, 55–66 (2021).

123. Brock, T. D. Life at high temperatures. *Science* **158**, 1012–1019 (1967).

124. Brock, T. D. & Freeze, H. *Thermus aquaticus* gen. n. and sp. n., a non-sporulating extreme thermophile. *J. Bacteriol.* **98**, 289–297 (1969).

125. Nakagawa, S. *et al. Sulfurihydrogenibium yellowstonense* sp. nov., an extremely thermophilic, facultatively heterotrophic, sulfur-oxidizing bacterium from Yellowstone National Park, and emended descriptions of the genus *Sulfurihydrogenibium. Int. J. Syst. Evol. Microbiol.* **55**, 2263–2268 (2005).

126. Itoh, T., Yoshikawa, N. & Takashina, T. *Thermogymnomonas acidicola* gen. nov., sp. nov., a novel thermoacidophilic, cell wall-less archaeon in order Thermoplasmatales, isolated from a solfataric soil in Hakone, Japan. *Int. J. Syst. Evol. Microbiol.* **57**, 2557–2561 (2007).

127. 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).

128. O'Neill, A. H., Liu, Y., Ferrera, I., Beveridge, T. J. & Reysenbach, A. L. *Sulfurihydrogenibium rodmanii* sp. nov., a sulfur-oxidizing chemolithoautotroph from the Uzon Caldera, Kamchatka Peninsula, Russia, and emended description of the genus *Sulfurihydrogenibium. Int. J. Syst. Evol. Microbiol.* **58**, 1147–1152 (2008).

129. Vésteinsdóttir, H., Reynisdóttir, D. B. & Örlygsson, J. *Thiomonas islandica* sp. nov., a moderately thermophilic, hydrogen- and sulfur-oxidizing betaproteobacterium isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* **61**, 132–137 (2011).

130. Dodsworth, J. A., Ong, J. C., Williams, A. J., Dohnalkova, A. C. & Hedlund, B. P. *Thermocrinis jamiesonii* sp. nov a thiosulfate-oxidizing, autotropic thermophile isolated from a geothermal spring. *Int. J. Syst. Evol. Microbiol.* **65**, 4769–4775 (2015).

131. Bendia, A. G. *et al.* A mosaic of geothermal and marine features shapes microbial community structure on Deception Island volcano, Antarctica. *Front. Microbiol.* **9**, 899 (2018).

132. Noell, S. E. *et al.* Unique Geothermal Chemistry Shapes Microbial Communities on Mt. Erebus, Antarctica. *Front. Microbiol.* **13**, 836943 (2022).

133. Mackenzie, R., Pedrós-Alió, C. & Díez, B. Bacterial composition of microbial mats in hot springs in Northern Patagonia: variations with seasons and temperature. *Extremophiles* **17**, 123–136 (2013).

134. Tobler, D. J. & Benning, L. G. Bacterial diversity in five Icelandic geothermal waters: temperature and sinter growth rate effects. *Extremophiles* **15**, 473–485 (2011).

135. Kubo, K., Knittel, K., Amann, R., Fukui, M. & Matsuura, K. Sulfur-metabolizing bacterial populations in microbial mats of the Nakabusa hot spring, Japan. *Syst. Appl. Microbiol.* **34**, 293–302 (2011).

136. Everroad, R. C., Otaki, H., Matsuura, K. & Haruta, S. Diversification of bacterial community composition along a temperature gradient at a thermal spring. *Microbes Environ.* **27**, 374–381 (2012).

137. Nishiyama, E. *et al.* The relationship between microbial community structures and environmental parameters revealed by metagenomic analysis of hot spring water in the Kirishima area, Japan. *Front. Bioeng. Biotechnol.* **6**, 202 (2018).

138. Hetzer, A., Morgan, H. W., McDonald, I. R. & Daughney, C. J. Microbial life in Champagne Pool, a geothermal spring in Waiotapu, New Zealand. *Extremophiles* **11**, 605–614 (2007).

139. Ward, L. *et al.* Microbial community dynamics in Inferno Crater Lake, a thermally fluctuating geothermal spring. *ISME J.* **11**, 1158–1167 (2017).

140. Amin, A. *et al.* Diversity and distribution of thermophilic bacteria in hot springs of Pakistan. *Microb. Ecol.* **74**, 116–127 (2017).

141. Merkel, A. Y. *et al.* Microbial diversity and autotrophic activity in Kamchatka hot springs. *Extremophiles* **21**, 307–317 (2017).

142. Wilkins, L. G. E., Ettinger, C. L., Jospin, G. & Eisen, J. A. Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Sci. Rep.* **9**, 3059 (2019).

143. Purcell, D. *et al.* The effects of temperature, pH and sulphide on the community structure of hyperthermophilic streamers in hot springs of northern Thailand. *FEMS Microbiol. Ecol.* **60**, 456–466 (2007).

144. Colman, D. R. *et al.* An analysis of geothermal and carbonic springs in the western United States sustained by deep fluid inputs. *Geobiology* **12**, 83–98 (2014).

145. Miller, S. R., Strong, A. L., Jones, K. L. & Ungerer, M. C. 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 (2009).

146. Inskeep, W. P. *et al.* Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Front. Microbiol.* **4**, 95 (2013).

147. 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).

148. Toplin, J. A., Norris, T. B., Lehr, C. R., McDermott, T. R. & Castenholz, R. W. Biogeographic and phylogenetic diversity of thermoacidophilic cyanidiales in Yellowstone National Park, Japan, and New Zealand. *Appl. Environ. Microbiol.* **74**, 2822–2833 (2008).

149. Valverde, A., Tuffin, M. & Cowan, D. A. Biogeography of bacterial communities in hot springs: a focus on the actinobacteria. *Extremophiles* **16**, 669–679 (2012).

150. Takacs-Vesbach, C. *et al.* Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three Aquificales lineages. *Front. Microbiol.* **4**, 84 (2013).

151. 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).

152. Munson-McGee, J. H. *et al.* Nanoarchaeota, their Sulfolobales host, and Nanoarchaeota virus distribution across Yellowstone National Park hot springs. *Appl. Environ. Microbiol.* **81**, 7860–7868 (2015).

153. Sharp, C. E. *et al.* Distribution and diversity of Verrucomicrobia methanotrophs in geothermal and

acidic environments. *Environ. Microbiol.* **16**, 1867–1878 (2014).

154. Zablocki, O., van Zyl, L. & Trindade, M. Biogeography and taxonomic overview of terrestrial hot spring thermophilic phages. *Extremophiles* **22**, 827–837 (2018).

155. McKay, L. J. *et al.* Sulfur cycling and host-virus interactions in Aquificales-dominated biofilms from Yellowstone's hottest ecosystems. *ISME J.* **16**, 842–855 (2022).

156. Oliverio, A. M. *et al.* The ecology and diversity of microbial eukaryotes in geothermal springs. *ISME J.* **12**, 1918–1928 (2018).

157. Nisbet, E. G. & Sleep, N. H. The habitat and nature of early life. *Nature* **409**, 1083–1091 (2001).

158. Hugenholtz, P., Pitulle, C., Hershberger, K. L. & Pace, N. R. Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* **180**, 366–376 (1998).

159. 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).

160. Cuecas, A., Portillo, M. C., Kanoksilapatham, W. & Gonzalez, J. M. Bacterial distribution along a 50 °C temperature gradient reveals a parceled out hot spring environment. *Microb. Ecol.* **68**, 729–739 (2014).

161. Inskeep, W. P. *et al.* Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS One* **5**, e9773 (2010).

162. Li, J., Peng, X., Zhang, L., Jiang, L. & Chen, S. Linking microbial community structure to S, N and Fe biogeochemical cycling in the hot springs at the Tengchong geothermal fields, southwest China. *Geomicrobiol. J.* **33**, 135–150 (2016).

163. Becraft, E. D., Wood, J. M., Cohan, F. M. & Ward, D. M. Biogeography of American northwest hot spring A/B′-lineage *Synechococcus* populations. *Front. Microbiol.* **11**, 77 (2020).

164. Elser, J. J. *et al.* Community structure and biogeochemical impacts of microbial life on floating pumice. *Appl. Environ. Microbiol.* **81**, 1542–1549 (2015).

165. Payne, D. *et al.* Geologic legacy spanning >90 years explains unique Yellowstone hot spring geochemistry and biodiversity. *Environ. Microbiol.* **21**, 4180–4195 (2019).

166. Shapiro, B. J. *et al.* Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**, 48–51 (2012).

167. Loiacono, 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).

168. Ruhl, I. A. *et al.* Microbial functional diversity correlates with species diversity along a temperature gradient. *mSystems* **7**, e00991-21 (2022).

169. Colman, D. R. *et al.* Seasonal hydrologic and geologic forcing drive hot spring geochemistry and microbial biodiversity. *Environ. Microbiol.* **23**, 4034–4053 (2021).

170. Eloe-Fadrosh, E. A. *et al.* Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* **7**, 10476 (2016).

171. Ishii, S. *et al.* Microbial population and functional dynamics associated with surface potential and carbon metabolism. *ISME J.* **8**, 963–978 (2013).

172. Colman, D. R. *et al.* Ecological differentiation in planktonic and sediment-associated chemotrophic microbial populations in Yellowstone hot springs. *FEMS Microbiol. Ecol.* **92**, fiw137 (2016).

173. Fernandes-Martins, M. C. *et al.* Ecological dichotomies arise in microbial communities due to mixing of deep hydrothermal waters and atmospheric gas in a circumneutral hot spring. *Appl. Environ. Microbiol.* **87**, e01598-21 (2021).

174. Siering, P. L. *et al.* Microbial biogeochemistry of Boiling Springs Lake: a physically dynamic, oligotrophic, low-pH geothermal ecosystem. *Geobiology* **11**, 356–376 (2013).

175. Wang, S. *et al.* Greater temporal changes of sediment microbial community than its waterborne counterpart in Tengchong hot springs, Yunnan Province, China. *Sci. Rep.* **4**, 7479 (2014).

176. Mongillo, M. A. & Clelland, L. *Concise listing of information on the thermal areas and thermal springs of New Zealand. New Zealand Department of Scientific and Industrial Research* **9**, (1984).

177. Dunfield, P. F. *et al.* Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450**, 879–882 (2007).

178. Saw, J. H. *et al.* Encapsulated in silica: genome, proteome and physiology of the thermophilic bacterium *Anoxybacillus flavithermus* WK1. *Genome Biol.* **9**, R161 (2008).

179. 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).

180. Crowe, M. A. *et al. Pyrinomonas methylaliphatogenes* gen. nov., sp. nov., a novel group 4 thermophilic member of the phylum Acidobacteria from geothermal soils. *Int. J. Syst. Evol. Microbiol.* **64**, 220–227 (2014).

181. Anders, H. *et al. Thermoflavifilum aggregans* gen. nov., sp. nov., a thermophilic and slightly halophilic filamentous bacterium from the phylum Bacteroidetes. *Int. J. Syst. Evol. Microbiol.* **64**, 1264–1270 (2014).

182. Anders, H. *et al. Limisphaera ngatamarikiensis* gen. nov., sp. nov., a thermophilic, pink-pigmented coccus isolated from subaqueous mud of a geothermal hotspring. *Int. J. Syst. Evol. Microbiol.* **65**, 1114–1121 (2015).

183. Houghton, K. M. *et al. Thermorudis pharmacophila* sp. nov., a novel member of the class Thermomicrobia isolated from geothermal soil, and emended descriptions of *Thermomicrobium roseum*, *Thermomicrobium carboxidum*, and *Thermorudis peleae*. *Int. J. Syst. Evol. Microbiol.* **65**, 4479–4487 (2015).

184. Carere, C. R. *et al.* Hydrogen oxidation influences glycogen accumulation in a verrucomicrobial methanotroph. *Front. Microbiol.* **10**, 1873 (2019).

185. Kaur, G., Mountain, B. W., Stott, M. B., Hopmans, E. C. & Pancost, R. D. Temperature and pH control on lipid composition of silica sinters from diverse hot springs in the Taupo Volcanic Zone, New Zealand. *Extremophiles* **19**, 327–344 (2015).

186. Carere, C. R. *et al.* Mixotrophy drives niche expansion of verrucomicrobial methanotrophs. *ISME J.* **11**, 2599–2610 (2017).

187. Childs, A. M., Mountain, B. W., O'Toole, R. & Stott, M. B. Relating microbial community and physicochemical parameters of a hot spring: Champagne Pool, Waiotapu, New Zealand. *Geomicrobiol. J.* **25**, 441–453 (2008).

188. Hug, K. *et al.* Microbial contributions to coupled arsenic and sulfur cycling in the acid-sulfide hot spring

Champagne Pool, New Zealand. *Front. Microbiol.* **5**, 569 (2014).

189. Stewart, L. C. *et al.* Interaction between ferruginous clay sediment and an iron-reducing hyperthermophilic *Pyrobaculum* sp. in a terrestrial hot spring. *FEMS Microbiol. Ecol.* **94**, fiy160 (2018).

190. Gilbert, J., O'Dor, R., King, N. & Vogel, T. M. The importance of metagenomic surveys to microbial ecology: or why Darwin would have been a metagenomic scientist. *Microb. Inform. Exp.* **1**, 5 (2011).

191. Colman, D. R. *et al.* Geobiological feedbacks and the evolution of thermoacidophiles. *ISME J.* **Advance on**, 1–12 (2017).

192. Sessitsch, A. *et al.* Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. *Appl. Environ. Microbiol.* **67**, 4215–4224 (2001).

193. Eilers, K. G., Debenport, S., Anderson, S. & Fierer, N. Digging deeper to find unique microbial communities: the strong effect of depth on the structure of bacterial and archaeal communities in soil. *Soil Biol. Biochem.* **50**, 58–65 (2012).

194. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).

195. Power, J. F. *et al.* Microbial biogeography of 925 geothermal springs in New Zealand. *Nat. Commun.* **9**, 2876 (2018).

196. Power, J. F. *et al.* Temporal dynamics of geothermal microbial communities in Aotearoa-New Zealand. *Front. Microbiol.* **14**, 1094311 (2023).

197. Power, J. F. *et al.* Draft genome sequence of *Venenivibrio stagnispumantis* CP.B2^T, isolated from Champagne Pool, Waiotapu, Aotearoa-New Zealand. *Microbiol. Resour. Announc.* **12**, e01074-22 (2023).

198. Oren, A. & Garrity, G. M. Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **71**, 005056 (2021).

199. 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).

200. Ward, C. S. *et al.* Annual community patterns are driven by seasonal switching between closely related marine bacteria. *ISME J.* **11**, 1412–1422 (2017).

201. Delgado-Baquerizo, M. *et al.* It is elemental: soil nutrient stoichiometry drives bacterial diversity. *Environ. Microbiol.* **19**, 1176–1188 (2017).

202. 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).

203. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).

204. 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).

205. Spear, J. R., Walker, J. J., McCollom, T. M. & Pace, N. R. Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *PNAS* **102**, 2555–2560 (2005).

206. 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. Biogeosciences* **116**, G01005 (2011).

207.   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).

208.   Chambefort, I. & Bignall, G. Taupo Volcanic Zone geothermal systems, New Zealand: exploration, science and development. *Geothermics* **59**, 147–356 (2016).

209.   Greening, C. *et al.* Persistence of the dominant soil phylum Acidobacteria by trace gas scavenging. *PNAS* **112**, 10497–10502 (2015).

210.   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).

211.   Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).

212.   Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).

213.   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).

214.   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).

215.   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).

216.   R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2020). Available at: http://www.r-project.org.

217.   McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).

218.   Oksanen, J. *et al.* vegan: community ecology package (v2.5-6). (2019). Available at: https://cran.r-project.org/package=vegan.

219.   Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer-Verlag New York, 2016).

220.   Dormann, C. F. *et al.* Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop.).* **36**, 27–46 (2013).

221.   Aho, K., Derryberry, D. & Peterson, T. Model selection for ecologists: the worldview of AIC and BIC. *Ecology* **95**, 631–636 (2014).

222.   Hamilton, N. ggtern: an extension to 'ggplot2', for the creation of ternary diagrams. (v2.1.5). (2016). Available at: https://cran.r-project.org/package=ggtern.

223.   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).

224.   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).

225.   Caporaso, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267 (2010).

226. 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).

227. Hermans, S. M. *et al.* Bacteria as emerging indicators of soil condition. *Appl. Environ. Microbiol.* **83**, e02826-16 (2017).

228. 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).

229. 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).

230. McCarthy, S. *et al.* Expanding the limits of thermoacidophily by adaptive evolution. *Appl. Environ. Microbiol.* **82**, 857–867 (2016).

231. Stetter, K. O. Extremophiles and their adaptation to hot environments. *FEBS Lett.* **452**, 22–25 (1999).

232. 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**, e0009-15 (2015).

233. Reysenbach, A.-L. & Shock, E. L. Merging genomes with geochemistry in hydrothermal ecosystems. *Science* **296**, 1077–1082 (2002).

234. Jiang, X. & Takacs-Vesbach, C. D. Microbial community analysis of pH 4 thermal springs in Yellowstone National Park. *Extremophiles* **21**, 135–152 (2017).

235. 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).

236. Gupta, R. S. The Phylum Aquificae. in *The Prokaryotes - Other Major Lineages of Bacteria and The Archaea* (eds. Rosenberg, E., Delong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 417–445 (Springer-Verlag Berlin Heidelberg, 2014).

237. Ji, M. *et al.* Atmospheric trace gases support primary production in Antarctic desert ecosystems. *Nature* **552**, 400–403 (2017).

238. Nakagawa, S., Takai, K., Inagaki, F., Horikoshi, K. & Sako, Y. *Nitratiruptor tergarcus* gen. nov., sp. nov. and *Nitratifractor salsuginis* gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the E-Proteobacteria isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough. *Int. J. Syst. Evol. Microbiol.* **55**, 925–933 (2005).

239. Schink, B. & Friedrich, M. Bacterial metabolism: phosphite oxidation by sulphate reduction. *Nature* **406**, 37 (2000).

240. White, A. K. & Metcalf, W. W. Microbial metabolism of reduced phosphorus compounds. *Annu Rev Microbiol* **61**, 379–400 (2007).

241. 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).

242. 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).

243. 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).

244.   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. *PNAS* **108**, 7850–7854 (2011).

245.   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).

246.   Lowe, C. L. Temporal dynamics of microbial communities in geothermal hotsprings of the Taupō Volcanic Zone. Master of Science (Research). The University of Waikato, Hamilton, New Zealand. (2017). Available at: https://hdl.handle.net/10289/11348.

247.   Nguyen, J., Lara-Gutiérrez, J. & Stocker, R. Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS Microbiol. Rev.* **45**, fuaa068 (2021).

248.   Ferris, M. J. & Ward, D. M. 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 (1997).

249.   Norris, T. B., McDermott, T. R. & Castenholz, R. W. The long-term effects of UV exclusion on the microbial composition and photosynthetic competence of bacteria in hot-spring microbial mats. *FEMS Microbiol. Ecol.* **39**, 193–209 (2002).

250.   Bowen De León, K., Gerlach, R., Peyton, B. M. & Fields, M. W. Archaeal and bacterial communities in three alkaline hot springs in Heart Lake Geyser Basin, Yellowstone National Park. *Front. Microbiol.* **4**, 330 (2013).

251.   Lacap, D. C., Barraquio, W. & Pointing, S. B. Thermophilic microbial mats in a tropical geothermal location display pronounced seasonal changes but appear resilient to stochastic disturbance. *Environ. Microbiol.* **9**, 3065–3076 (2007).

252.   Lehr, C. R. *et al.* Cyanidia (cyanidiales) population diversity and dynamics in an acid-sulfate-chloride spring in Yellowstone National Park. *J. Phycol.* **43**, 3–14 (2007).

253.   Colman, D. R., Lindsay, M. R. & Boyd, E. S. Mixing of meteoric and geothermal fluids supports hyperdiverse chemosynthetic hydrothermal communities. *Nat. Commun.* **10**, 681 (2019).

254.   Thain, I. A. & Carey, B. Fifty years of geothermal power generation at Wairakei. *Geothermics* **38**, 48–63 (2009).

255.   Bignall, G. *et al.* Geology of the Wairakei-Tauhara geothermal system, New Zealand. in *Proceedings World Geothermal Congress 2010* 1–8 (2010).

256.   Reyes, A. G., Christenson, B. W. & Faure, K. Sources of solutes and heat in low-enthalpy mineral waters and their relation to tectonic setting, New Zealand. *J. Volcanol. Geotherm. Res.* **192**, 117–141 (2010).

257.   Gallagher, A. *et al.* Hydrothermal eruption dynamics reflecting vertical variations in host rock geology and geothermal alteration, Champagne Pool, Waiotapu, New Zealand. *Bull. Volcanol.* **82**, 77 (2020).

258.   Shade, A. *et al.* Fundamentals of microbial community resistance and resilience. *Front. Microbiol.* **3**, 417 (2012).

259.   Scott, B. J. Cyclic activity in the crater lakes of Waimangu hydrothermal system, New Zealand. *Geothermics* **23**, 555–572 (1994).

260.   Reeves, R. R., Wilke, M., Cashmore, P., Macdonald, N. & Thompson, K. Physical and ecological effects of rehabilitating the geothermally influenced Waikite Wetland, New Zealand. *J. Environ.*

Manage. **228**, 279–291 (2018).

261. National Institute of Water and Atmospheric Research (NIWA). New Zealand's National Climate Database. (2016). Available at: https://cliflo.niwa.co.nz/.

262. Bay of Plenty Regional Council (BOPRC). New Zealand's Environmental Data Portal. (2016). Available at: https://envdata.boprc.govt.nz/.

263. Kahle, D. & Wickham, H. ggmap: spatial visualization with ggplot2. *R J.* **5**, 144–161 (2013).

264. Pope, J. G., McConchie, D. M., Clark, M. D. & Brown, K. L. Diurnal variations in the chemistry of geothermal fluids after discharge, Champagne Pool, Waiotapu, New Zealand. *Chem. Geol.* **203**, 253–272 (2004).

265. Ullrich, M. K., Pope, J. G., Seward, T. M., Wilson, N. & Planer-Friedrich, B. Sulfur redox chemistry governs diurnal antimony and arsenic cycles at Champagne Pool, Waiotapu, New Zealand. *J. Volcanol. Geotherm. Res.* **262**, 164–177 (2013).

266. Jones, E. F. *et al.* Stream microbial community structured by trace elements, headwater dispersal, and large reservoirs in sub-alpine and urban ecosystems. *Front. Microbiol.* **11**, 491425 (2020).

267. Shade, A. *et al.* Lake microbial communities are resilient after a whole-ecosystem disturbance. *ISME J.* **6**, 2153–2167 (2012).

268. Simon, M. *et al.* Resilience of freshwater communities of small microbial eukaryotes undergoing severe drought events. *Front. Microbiol.* **7**, 812 (2016).

269. Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D. & Bertilsson, S. A guide to the natural history of freshwater lake bacteria. *Microbiol. Mol. Biol. Rev.* **75**, 14–49 (2011).

270. Wagner, B. D. *et al.* On the use of diversity measures in longitudinal sequencing studies of microbial communities. *Front. Microbiol.* **9**, 1037 (2018).

271. Zheng, X. & Frederiksen, C. S. A study of predictable patterns for seasonal forecasting of New Zealand rainfall. *J. Clim.* **19**, 3320–3333 (2006).

272. Albers, S.-V. & Siebers, B. The family Sulfolobaceae. in *The Prokaryotes - Other Major Lineages of Bacteria and The Archaea* (eds. Rosenberg, E., De Long, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 323–346 (Springer-Verlag Berlin Heidelberg, 2014).

273. Chambefort, I. Sulfur in New Zealand geothermal systems: insights from stable isotope and trace element analyses of anhydrite from Rotokawa and Ngatamariki geothermal fields, Taupō Volcanic Zone. *New Zeal. J. Geol. Geophys.* **64**, 372–388 (2021).

274. Glover, R. B., Hunt, T. M. & Severne, C. M. Ohaaki Ngawha; Ohaaki Pool. *Proc. 18th New Zeal. Geotherm. Work.* 77–84 (1996).

275. Poretsky, R., Rodriguez-R, L. M., Luo, C., Tsementzi, D. & Konstantinidis, K. T. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One* **9**, e93827 (2014).

276. Props, R. *et al.* Absolute quantification of microbial taxon abundances. *ISME J.* **11**, 584–587 (2017).

277. Wagner Mackenzie, B., Waite, D. W. & Taylor, M. W. Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Front. Microbiol.* **6**, 130 (2015).

278. Tremblay, J. *et al.* Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* **6**, 771 (2015).

279. Wear, E. K., Wilbanks, E. G., Nelson, C. E. & Carlson, C. A. Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton. *Environ. Microbiol.* **20**, 2709–2726 (2018).

280. Welford, H. E. Revisiting the thermophilic bacterial genus *Venenivibrio*; physiological, phenotypic, and genotypic (re)characterisation of the type strain and novel *Venenivibrio* isolates. Master of Science. The University of Canterbury, Christchurch, NZ. (2022). Available at: https://hdl.handle.net/10092/104029.

281. Mayr, E. *Systematics and the origin of species*. (Columbia University Press, 1942).

282. Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K. & Triantis, K. A. Island biogeography: taking the long view of natures laboratories. *Science* **357**, aam8326 (2017).

283. Whittaker, K. A. & Rynearson, T. A. Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. *PNAS* **114**, 2651–2656 (2017).

284. Davison, J. *et al.* Microbial island biogeography: isolation shapes the life history characteristics but not diversity of root-symbiotic fungal communities. *ISME J.* **12**, 2211–2224 (2018).

285. Whitaker, R. J. Allopatric origins of microbial species. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 1975–1984 (2006).

286. Choudoir, M. J., Barberán, A., Menninger, H. L., Dunn, R. R. & Fierer, N. Variation in range size and dispersal capabilities of microbial taxa. *Ecology* **99**, 322–334 (2018).

287. Shapiro, B. J. & Polz, M. F. Microbial speciation. *Cold Spring Harb. Perspect. Biol.* **7**, a018143 (2015).

288. Baquero, F., Coque, T. M., Galán, J. C. & Martinez, J. L. The origin of niches and species in the bacterial world. *Front. Microbiol.* **12**, 657986 (2021).

289. Talbot, J. M. *et al.* Endemism and functional convergence across the North American soil mycobiome. *PNAS* **111**, 6341–6346 (2014).

290. Stöhr, R., Waberski, A., Völker, H., Tindall, B. J. & Thomm, M. *Hydrogenothermus marinus* gen. nov., sp. nov., a novel thermophilic hydrogen-oxidizing bacterium, and recognition of *Calderobacterium hydrogenophilum* as a member of the genus *Hydrogenobacter*. *Int. J. Syst. Evol. Microbiol.* **51**, 1853–1862 (2001).

291. Trembath-Reichert, E., Butterfield, D. A. & Huber, J. A. Active subseafloor microbial communities from Mariana back-arc venting fluids share metabolic strategies across different thermal niches and taxa. *ISME J.* **13**, 2264–2279 (2019).

292. Reysenbach, A. L. *et al.* Complete and draft genome sequences of six members of the Aquificales. *J. Bacteriol.* **191**, 1992–1993 (2009).

293. Hedlund, B. P. *et al.* Isolation of diverse members of the Aquificales from geothermal springs in Tengchong, China. *Front. Microbiol.* **6**, 157 (2015).

294. Hamamura, N., Meneghin, J. & Reysenbach, A. L. Comparative community gene expression analysis of Aquificales-dominated geothermal springs. *Environ. Microbiol.* **15**, 1226–1237 (2013).

295. Chen, I. M. A. *et al.* The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* **49**, D751–D763 (2021).

296. Søndergaard, D., Pedersen, C. N. S. & Greening, C. HydDB: a web tool for hydrogenase classification and analysis. *Sci. Rep.* **6**, 34212 (2016).

297. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic*

Acids Res. **49**, D10–D17 (2021).

298. Cole, J. R. *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**, D633–D642 (2014).

299. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).

300. Lagkouvardos, I. *et al.* IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci. Rep.* **6**, 33721 (2016).

301. Gonzalez, A. *et al.* Qiita: rapid, web-enabled microbiome meta-analysis. *Nat. Methods* **15**, 796–798 (2018).

302. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).

303. Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020**, baaa062 (2020).

304. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

305. Corrêa, F. B., Saraiva, J. P., Stadler, P. F. & Da Rocha, U. N. TerrestrialMetagenomeDB: a public repository of curated and standardized metadata for terrestrial metagenomes. *Nucleic Acids Res.* **48**, D626–D632 (2020).

306. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

307. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

308. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

309. Kieser, S., Brown, J., Zdobnov, E. M., Trajkovski, M. & McCue, L. A. ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinformatics* **21**, 257 (2020).

310. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics* **31**, 1674–1676 (2015).

311. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

312. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).

313. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

314. Gourlé, H., Karlsson-Lindsjö, O., Hayer, J. & Bongcam-Rudloff, E. Simulating Illumina metagenomic data with InSilicoSeq. *Bioinformatics* **35**, 521–522 (2019).

315. Li, H. Seqtk. (2021). Available at: https://github.com/lh3/seqtk.

316. Arkin, A. P. *et al.* KBase: the United States Department of Energy systems biology knowledgebase. *Nat.*

Biotechnol. **36**, 566–569 (2018).

317. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).

318. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

319. Campbell, B. J. & Cary, S. C. Abundance of reverse tricarboxylic acid cycle genes in free-living microorganisms at deep-sea hydrothermal vents. *Appl. Environ. Microbiol.* **70**, 6282–6289 (2004).

320. Greening, C. *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H_2 is a widely utilised energy source for microbial growth and survival. *ISME J.* **10**, 761–777 (2016).

321. Dahl, C. Cytoplasmic sulfur trafficking in sulfur-oxidizing prokaryotes. *IUBMB Life* **67**, 268–274 (2015).

322. Wang, R. *et al.* Sulfur oxidation in the acidophilic autotrophic *Acidithiobacillus* spp. *Front. Microbiol.* **9**, 2390 (2019).

323. Meng, Y. L., Liu, Z. & Rosen, B. P. As(III) and Sb(III) uptake by GlpF and efflux by ArsB in *Escherichia coli*. *J. Biol. Chem.* **279**, 18334–18341 (2004).

324. Fekih, I. Ben *et al.* Distribution of arsenic resistance genes in prokaryotes. *Front. Microbiol.* **9**, 2473 (2018).

325. Song, S. & Wood, T. K. 'Viable but non-culturable cells' are dead. *Environ. Microbiol.* **23**, 2335–2338 (2021).

326. Livermore, J. A. & Jones, S. E. Local–global overlap in diversity informs mechanisms of bacterial biogeography. *ISME J.* **9**, 2413–2422 (2015).

327. Archer, S. D. J. *et al.* Airborne microbial transport limitation to isolated Antarctic soil habitats. *Nat. Microbiol.* **4**, 925–932 (2019).

328. Archer, S. *et al.* Global biogeography of atmospheric microorganisms reflects diverse recruitment and environmental filtering. *Research Square* (2022). Available at: https://doi.org/10.21203/rs.3.rs-244923/v4.

329. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *PNAS* **103**, 12115–12120 (2006).

330. Dong, Y. *et al.* Physiology, metabolism, and fossilization of hot-spring filamentous microbial mats. *Astrobiology* **19**, 1442–1458 (2019).

331. Sunna, A. & Bergquist, P. L. A gene encoding a novel extremely thermostable 1,4-*β*-xylanase isolated directly from an environmental DNA sample. *Extremophiles* **7**, 63–70 (2003).

332. Rivett, D. W., Mombrikotb, S. B., Gweon, H. S., Bell, T. & van der Gast, C. Bacterial communities in larger islands have reduced temporal turnover. *ISME J.* **15**, 2947–2955 (2021).

333. Mortimer, N. *et al.* Zealandia: Earth's hidden continent. *GSA Today* **27**, 27–35 (2017).

334. Barker, S. J. *et al.* Taupō: an overview of New Zealand's youngest supervolcano. *New Zeal. J. Geol. Geophys.* **64**, 320–346 (2021).

335. Mollerup, S. *et al.* *Propionibacterium acnes*: Disease-causing agent or common contaminant? detection in diverse patient samples by next- generation sequencing. *J. Clin. Microbiol.* **54**, 980–987 (2016).

336. Scholz, C. F. P. & Kilian, M. The natural history of cutaneous propionibacteria, and reclassification of

selected species within the genus *Propionibacterium* to the proposed novel genera *Acidipropionibacterium*, *Cutibacterium* and *Pseudopropionibacterium*. *Int. J. Syst. Evol. Microbiol.* **66**, 4422–4432 (2016).

337. Carini, P. *et al.* Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat. Microbiol.* **2**, 16242 (2016).

338. Větrovský, T. & Baldrian, P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* **8**, e57923 (2013).

339. Ferrera, I., Banta, A. B. & Reysenbach, A. L. Spatial patterns of Aquificales in deep-sea vents along the Eastern Lau Spreading Center (SW Pacific). *Syst. Appl. Microbiol.* **37**, 442–448 (2014).

340. Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11**, 2639–2643 (2017).

341. Caruso, V., Song, X., Asquith, M. & Karstens, L. Performance of microbiome sequence inference methods in environments with varying biomass. *mSystems* **4**, e00163-18 (2019).

342. Quail, M. A. *et al.* A tale of three NGS sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341 (2012).

343. Salipante, S. J. *et al.* Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl. Environ. Microbiol.* **80**, 7583–7591 (2014).

344. Lahens, N. F. *et al.* A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. *BMC Genomics* **18**, 602 (2017).

345. Park, C., Kim, S. B., Choi, S. H. & Kim, S. Comparison of 16S rRNA gene based microbial profiling using five next-generation sequencers and various primers. *Front. Microbiol.* **12**, 715500 (2021).

346. Denef, V. J. *et al.* Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *PNAS* **107**, 2383–2390 (2010).

347. Zhou, Z., Tran, P. Q., Kieft, K. & Anantharaman, K. Genome diversification in globally distributed novel marine Proteobacteria is linked to environmental adaptation. *ISME J.* **14**, 2060–2077 (2020).

348. Jay, Z. J. *et al.* Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nat. Microbiol.* **3**, 732–740 (2018).

349. Langille, M. G. I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013).

350. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**, 1272–1277 (2016).

351. Xu, Z., Malmer, D., Langille, M. G. I., Way, S. F. & Knight, R. Which is more important for classifying microbial communities: who's there or what they can do? *ISME J.* **8**, 2357–2359 (2014).

352. van den Burg, B. Extremophiles as a source for novel enzymes. *Curr. Opin. Microbiol.* **6**, 213–218 (2003).

353. Littlechild, J. A. Enzymes from extreme environments and their industrial applications. *Front. Bioeng. Biotechnol.* **3**, 161 (2015).

354. Cleiss-Arnold, J. *et al.* Temporal transcriptomic response during arsenic stress in *Herminiimonas arsenicoxydans*. *BMC Genomics* **11**, 709 (2010).

355. Christel, S. *et al.* RNA transcript sequencing reveals inorganic sulfur compound oxidation pathways in the acidophile *Acidithiobacillus ferrivorans*. *FEMS Microbiol. Lett.* **363**, fnw057 (2016).

356. Peng, T., Xu, Y. & Zhang, Y. Comparative genomics of molybdenum utilization in prokaryotes and eukaryotes. *BMC Genomics* **19**, 691 (2018).

357. Radujković, D. *et al.* Prolonged exposure does not increase soil microbial community response to warming along geothermal gradients. *FEMS Microbiol. Ecol.* **94**, fix174 (2018).

358. Gudbergsdóttir, S. R., Menzel, P., Krogh, A., Young, M. & Peng, X. Novel viral genomes identified from six metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot springs. *Environ. Microbiol.* **18**, 863–874 (2016).

359. Bahram, M., Anslan, S., Hildebrand, F., Bork, P. & Tedersoo, L. Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment. *Environ. Microbiol. Rep.* **11**, 487–494 (2019).

360. Taylor, S. C., Laperriere, G. & Germain, H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci. Rep.* **7**, 2409 (2017).

361. Cary, S. C. & Fierer, N. The importance of sample archiving in microbial ecology. *Nat. Rev. Microbiol.* **12**, 789–790 (2014).

362. St. John, E. *et al.* A new symbiotic nanoarchaeote (*Candidatus* Nanoclepta minutus) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. *Syst. Appl. Microbiol.* **42**, 94–106 (2019).

363. Gutleben, J. *et al.* The multi-omics promise in context: from sequence to microbial isolate. *Crit. Rev. Microbiol.* **44**, 212–229 (2018).

364. Palatinszky, M. *et al.* Cyanate as an energy source for nitrifiers. *Nature* **524**, 105–108 (2015).

365. Kauffman, K. M. *et al.* Resolving the structure of phage–bacteria interactions in the context of natural diversity. *Nat. Commun.* **13**, 372 (2022).

366. Tamaki, H. *et al. Armatimonas rosea* gen. nov., sp. nov., of a novel bacterial phylum, Armatimonadetes phyl. nov., formally called the candidate phylum OP10. *Int. J. Syst. Evol. Microbiol.* **61**, 1442–1447 (2011).

367. Lee, K. C. Y., Dunfield, P. F. & Stott, M. B. The Phylum Armatimonadetes. in *The Prokaryotes - Other Major Lineages of Bacteria and The Archaea* (ed. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F.) 447–458 (Springer-Verlag Berlin Heidelberg, 2014).

368. Quéméneur, M. *et al.* Procaryotic diversity and hydrogenotrophic methanogenesis in an alkaline spring (La Crouen, New Caledonia). *Microorganisms* **9**, 1360 (2021).

369. Dragone, N. B. *et al.* The early microbial colonizers of a short-lived volcanic island in the Kingdom of Tonga. *MBio* **14**, e03313-22 (2023).

370. Stewart, L. C., Stucker, V. K., Stott, M. B. & Ronde, C. E. J. De. Marine-influenced microbial communities inhabit terrestrial hot springs on a remote island volcano. *Extremophiles* **22**, 687–698 (2018).

371. Lau, K. E. M. *et al.* A novel bacterial community index to assess stream ecological health. *Freshw. Biol.* **60**, 1988–2002 (2015).

372. Astudillo-García, C., Hermans, S. M., Stevenson, B., Buckley, H. L. & Lear, G. Microbial assemblages and bioindicators as proxies for ecosystem health status: potential and limitations. *Appl. Microbiol. Biotechnol.* **103**, 6407–6421 (2019).

373. Ning, D. *et al.* A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* **11**, 4747 (2020).

374. Lear, G. *et al.* The biogeography of stream bacteria. *Glob. Ecol. Biogeogr.* **22**, 544–554 (2013).

375. Cárdenas, C. A., Bell, J. J., Davy, S. K., Hoggard, M. & Taylor, M. W. Influence of environmental variation on symbiotic bacterial communities of two temperate sponges. *FEMS Microbiol. Ecol.* **88**, 516–527 (2014).

376. Boey, J. S., Mortimer, R., Couturier, A., Worrallo, K. & Handley, K. M. Estuarine microbial diversity and nitrogen cycling increase along sand–mud gradients independent of salinity and distance. *Environ. Microbiol.* **24**, 50–65 (2022).

377. Stott, M. B. & Taylor, M. W. Microbial ecology research in New Zealand. *N. Z. J. Ecol.* **40**, 12–28 (2016).

378. Hermans, S. M. *et al.* Using soil bacterial communities to predict physico-chemical variables and soil quality. *Microbiome* **8**, 79 (2020).

379. Wilson, A. D. The micro-determination of ferrous iron in silicate minerals by a volumetric and a colorimetric method. *Analyst* **85**, 823–827 (1960).

380. Rueckert, A. & Morgan, H. W. Removal of contaminating DNA from polymerase chain reaction using ethidium monoazide. *J. Microbiol. Methods* **68**, 596–600 (2007).

381. Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**, 635–645 (2014).

382. Katz, K. S. *et al.* STAT: a fast, scalable, MinHash-based *k*-mer tool to assess Sequence Read Archive next-generation sequence submissions. *Genome Biol.* **22**, 270 (2021).

383. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).

384. Hügler, M., Huber, H., Molyneaux, S. J., Vetriani, C. & Sievert, S. M. Autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum Aquificae: evidence for two ways of citrate cleavage. *Environ. Microbiol.* **9**, 81–92 (2007).

385. Aoshima, M. & Igarashi, Y. A novel oxalosuccinate-forming enzyme involved in the reductive carboxylation of 2-oxoglutarate in *Hydrogenobacter thermophilus* TK-6. *Mol. Microbiol.* **62**, 748–759 (2006).

386. Steffens, L. *et al.* High CO_2 levels drive the TCA cycle backwards towards autotrophy. *Nature* **592**, 784–788 (2021).

387. Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W. & Milo, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *PNAS* **110**, 10039–10044 (2013).

388. Spaans, S. K., Weusthuis, R. A., van der Oost, J. & Kengen, S. W. M. NADPH-generating systems in bacteria and archaea. *Front. Microbiol.* **6**, 742 (2015).

389. Borisov, V. B., Gennis, R. B., Hemp, J. & Verkhovsky, M. I. The cytochrome *bd* respiratory oxygen reductases. *Biochim. Biophys. Acta* **1807**, 1398–1413 (2012).

390. Trumpower, B. L. The protonmotive Q cycle: energy transduction by coupling of proton translocation to electron transfer by the cytochrome *bc*1 complex. *J. Biol. Chem.* **265**, 11409–11412 (1990).

391. Boughanemi, S., Infossi, P., Giudici-Orticoni, M. T., Schoepp-Cothenet, B. & Guiral, M. Sulfite oxidation by the quinone-reducing molybdenum sulfite dehydrogenase SoeABC from the bacterium *Aquifex aeolicus*. *Biochim. Biophys. Acta - Bioenerg.* **1861**, 148279 (2020).

392. Koch, T. & Dahl, C. A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds. *ISME J.* **12**, 2479–2491 (2018).

393. Cozen, A. E. *et al.* Transcriptional map of respiratory versatility in the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *J. Bacteriol.* **191**, 782–794 (2009).

394. Giloteaux, L. *et al.* Characterization and transcription of arsenic respiration and resistance genes during *in situ* uranium bioremediation. *ISME J.* **7**, 370–383 (2013).

395. Hu, M. *et al.* New arsenite oxidase gene (*aioA*) PCR primers for assessing arsenite-oxidizer diversity in the environment using high-throughput sequencing. *Front. Microbiol.* **12**, 691913 (2021).

396. Hamamura, N. *et al.* Linking microbial oxidation of arsenic with detection and phylogenetic analysis of arsenite oxidase genes in diverse geothermal environments. *Environ. Microbiol.* **11**, 421–431 (2009).

397. Lee, K. C. *et al.* The *Chthonomonas calidirosea* genome is highly conserved across geography and distinct chemical and microbial environments in the Taupō Volcanic Zone, New Zealand. *Appl. Environ. Microbiol.* **82**, 3572–3581 (2016).

398. Tamazawa, S., Takasaki, K., Tamaki, H., Kamagata, Y. & Hanada, S. Metagenomic and biochemical characterizations of sulfur oxidation metabolism in uncultured large sausage-shaped bacterium in hot spring microbial mats. *PLoS One* **7**, e49793 (2012).

399. Yeoh, Y. K. *et al.* The core root microbiome of sugarcanes cultivated under varying nitrogen fertilizer application. *Environ. Microbiol.* **18**, 1338–1351 (2016).

400. Grasby, S. E. *et al.* The Paint Pots, Kootenay National Park, Canada - a natural acid spring analogue for mars. *Can. J. Earth Sci.* **50**, 94–108 (2013).

401. Martijn, J. *et al.* Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon. *Environ. Microbiol.* **21**, 2485–2498 (2019).

402. Massello, F. L. *et al.* Meta-analysis of microbial communities in hot springs: recurrent taxa and complex shaping factors beyond pH and temperature. *Microorganisms* **8**, 906 (2020).

403. Giovannelli, D. *et al.* Insight into the evolution of microbial metabolism from the deep-branching bacterium, *Thermovibrio ammonificans*. *Elife* **6**, e18990 (2017).

404. Dixit, S. *et al.* Bacterial diversity and CAZyme potential revealed in Pandanus rich thermal spring cluster of India: a non-cultivable 16S rRNA sequencing approach. *Front. Microbiol.* **12**, 760573 (2021).

405. Yabe, S., Kato, A., Hazaka, M. & Yokota, A. *Thermaerobacter composti* sp. nov., a novel extremely thermophilic bacterium isolated from compost. *J. Gen. Appl. Microbiol.* **55**, 323–328 (2009).

406. Isobe, K. *et al.* A simple and rapid GC/MS method for the simultaneous determination of gaseous metabolites. *J. Microbiol. Methods* **84**, 46–51 (2011).

407. Peng, X. & Jones, B. Rapid precipitation of silica (opal-A) disguises evidence of biogenicity in high-temperature geothermal deposits: case study from Dagunguo hot spring, China. *Sediment. Geol.* **257–260**, 45–62 (2012).

408. Gupta, R. S. & Lali, R. Molecular signatures for the phylum Aquificae and its different clades: proposal for division of the phylum Aquificae into the emended order Aquificales, containing the families Aquificaceae and Hydrogenothermaceae, and a new order Desulfurobacteriales. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **104**, 349–368 (2013).

409. Härtig, C. *et al.* Chemolithotrophic growth of the aerobic hyperthermophilic bacterium *Thermocrinis ruber* OC 14/7/2 on monothioarsenate and arsenite. *FEMS Microbiol. Ecol.* **90**, 747–760 (2014).

410. François, D. X. *et al. Persephonella atlantica* sp. nov.: how to adapt to physico-chemical gradients in high temperature hydrothermal habitats. *Syst. Appl. Microbiol.* **44**, 126176 (2021).

411. Guo, L. *et al.* Temperature governs the distribution of hot spring microbial community in three hydrothermal fields, eastern Tibetan Plateau geothermal belt, western China. *Sci. Total Environ.* **720**, 137574 (2020).

412. Xiao, E. *et al.* The physiological response of *Arundo donax* and characteristics of anodic bacterial community in BE-CW systems: effects of the applied voltage. *Chem. Eng. J.* **380**, 122604 (2020).

413. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).

414. Jiang, L., Hengchao, X. U. & Qiao, H. Biomediated precipitation of calcium carbonate in a slightly acidic hot spring, Yunnan Province. *Acta Geol. Sin.* **91**, 145–155 (2017).

415. Pagaling, E. *et al.* Bacterial and archaeal diversity in two hot spring microbial mats from the geothermal region of Tengchong, China. *Extremophiles* **16**, 607–618 (2012).

416. Peng, X., Xu, H., Jones, B., Chen, S. & Zhou, H. Silicified virus-like nanoparticles in an extreme thermal environment: implications for the preservation of viruses in the geological record. *Geobiology* **11**, 511–526 (2013).

417. Pagaling, E. *et al.* Microbial biogeography of six salt lakes in Inner Mongolia, China, and a salt lake in Argentina. *Appl. Environ. Microbiol.* **75**, 5750–5760 (2009).

418. Sugihara, C. *et al.* Transition of microbiological and sedimentological features associated with the geochemical gradient in a travertine mound in northern Sumatra, Indonesia. *Sediment. Geol.* **343**, 85–98 (2016).

419. Adjeroud, M., Escuder-Rodríguez, J. J., González-Siso, M. I. & Kecha, M. Metagenomic investigation of bacterial and archaeal diversity of Hammam Essalihine hot spring from Khenchela, Algeria. *Geomicrobiol. J.* **37**, 804–817 (2020).

420. Schirmer, M. *et al.* Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43**, e37 (2015).

421. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes *de novo* assembler. *Curr. Protoc. Bioinforma.* **70**, e102 (2020).

422. Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D. & Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142–i150 (2018).

423. Mukherjee, S. *et al.* Genomes OnLine Database (GOLD) v.8: overview and updates. *Nucleic Acids Res.* **49**, D723–D733 (2021).

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

B.1 SUPPLEMENTARY METHODS

B.1.1 Field sampling

Three litres of spring water were collected for each sample taken. A custom-made stainless-steel device was used to capture the water column of each feature to a depth of 1 m where possible, either at the centre of the spring or at ~3 m from the edge for large features to target well-mixed and/or more representative samples, depending on safety and size of the spring (detailed sketches were drawn of each sampling location to facilitate replication). If the feature was on a slope or the safest sampling position was too far from the spring for the water sampler to operate correctly, a 500 mL PP Nalgene bottle (ThermoFisher Scientific, Waltham, MA, USA) was used to collect the water from the same part of the water column. Either this vessel or the water sampler was then used to aseptically fill a 2000 mL PP Nalgene bottle (ThermoFisher Scientific, Waltham, MA, USA) with spring water for subsequent filtering and DNA extraction. In addition, a 330 mL rubber-sealed glass bottle was collected for geochemical analyses and the 500 mL PP bottle was used to retain water for geophysical parameters. All vessels that contained a microbiological sample were subjected to the same stringent washing procedures, namely with detergent (Extran MA03, EMD Millipore, Billerica, MA, USA) and bleach (20 % v/v bleach:distilled water solution using 5 % sodium hypochlorite), followed by a final autoclave step (122 °C for 20 mins). All metadata were recorded on a custom-made application suitable for Android tablets. Metadata recorded *in situ* at the time of sampling included: sample number, sample date, feature name, feature type, location name, geothermal field, district, latitude and longitude coordinates, detailed description, ebullition, size, colour, spring temperature (same location as sample), and photographs/diagrams of the site. Spring temperature (TEMP) was measured *in situ* immediately after sampling. Parameters measured within two hours of sampling were pH, oxidation-reduction potential (ORP), conductivity (COND), dissolved oxygen (dO), turbidity (TURB), ferrous iron (Fe^{2+}) concentration and filtered volume (more details are given in Sample Processing). Entries were digitally linked to the corresponding sample ID and automatically uploaded to an Amazon Relational Database Service (RDS) and E3 Bucket

with structured query language (SQL). These results are visible on an Amazon Elastic Compute Cloud (EC2) web server, accessed through https://1000springs.org.nz/.

B.1.2 Sample processing

Within 2 hours of sampling, the contents of the 2000 mL Nalgene bottle were filtered through a Sterivex-GP 0.22 µm PES column filter (EMD Millipore, Billerica, MA, USA), using the Masterflex E/S Portable Sampler with a peristaltic L/S pump head and platinum-cured silicone L/S tubing (Cole-Parmer, Vernon Hills, IL, USA). All tubing was bleached and rinsed (first with reverse-osmosis water followed by approximately 150 mL sample water) between samples. Each sample was filtered until all 2,000 mL water was pushed through or the filter membrane became clogged. The filters were immediately cooled to 4 °C and then stored at -20 °C until DNA extraction. Filtrate from the column filter was used to fill three 50 mL tubes and two 15 mL tubes with spring water for varying geochemical analyses (0). 80 µL of this filtrate was also added to 4 mL ferrous iron reagent (which includes 0.63 mM 2,2-bipyridyl, 0.80 M ammonium acetate and 3.7 % v/v glacial acetic acid)[379] for colorimetric spectroscopy. A multiparameter field meter (Hanna Instruments, Woonsocket, RI, USA) was used to measure pH, oxidation-reduction potential (ORP), conductivity (COND), turbidity (TURB), dissolved oxygen (dO) and sample temperature from the air-tight 500 mL vessel (after samples had cooled to below room temperature). Where pH measured less than 1, a benchtop pH meter (Hanna Instruments, Woonsocket, RI, USA) calibrated to pH 0 was used. All geochemical sample vessels were then stored at either 4 °C or -20 °C until analyses were performed. Aqueous metals and non-metals measured by ICP-MS were Ag, Al, As, B, Ba, Br, Ca, Cd, Co, Cr, Cs, Cu, Fe, Hg, K, Li, Mg, Mn, Mo, Na, Ni, Pb, Rb, S, Se, Si, Sr, Tl, U, V, and Zn. Chemical analyses were performed at the Geomicrobiology Research Group (GRG) and the New Zealand Geothermal Analytical Laboratory (NZGAL), both at GNS Science in Wairakei, New Zealand and at the School of Science, University of Waikato, Hamilton, New Zealand. All samples and analyses are summarised in Table B.1.

B.1.3 DNA extraction

DNA extraction, amplification and sequencing were performed at the Thermophile Research Unit and DNA Sequencing Facility (University of Waikato, Hamilton, New Zealand) from a modified cetyl trimethylammonium bromide (CTAB) method[210]. Column filters were thawed, 500 µL of a 0.8 % w/v skim milk powder solution (local consumer brand, freshly prepared for each batch of extractions and treated with UV light for 20 min) added and mixed at 150

RPM on a Ratek orbital mixer at 65 °C for 15 mins. A buffer containing CTAB was used in the extraction lysis buffer which consisted of 2 % (v/v) cetyl trimethylammonium bromide, 1 % (v/v) polyvinyl pyrrolidone, 100 mM Tris-HCl, 1.4 M NaCl and 20 mM EDTA. The final extraction buffer contained 400 µL CTAB buffer, 200 µL PBS (100 mM) and 100 µL SDS (10 %). This was added to the filters which were then mixed at 150 RPM at 65 °C for 45 mins. The extraction buffer was pushed through the filter and collected. A further 0.7 mL CTAB buffer was added to the filters and mixed at 150 RPM for 15 mins. This filtrate was added to a separate tube, resulting in two extraction duplicates. Chloroform:isoamyl alcohol (24:1), in equal volumes (1:1) to the filtrate, was added to each duplicate and vortexed. These were centrifuged at 10,000 RCF for 12 mins at 4 °C. The aqueous top layers were transferred to new tubes, and 300 µL of chloroform:isoamyl (24:1) added. Again, these were vortexed and centrifuged at 10,000 RCF for 12 mins at 4 °C. The aqueous top layers were removed to fresh tubes. The subsequent steps were modified from the PowerMag Microbial DNA Isolation Kit using SwiftMag technology (MoBio Laboratories, Carlsbad, CA, USA). Equal volumes of 100 % molecular grade ethanol and SwiftMag beads (22:22 µL) were added to the aqueous phases and placed on a magnetic stand for 2 mins. The supernatants were removed and the beads washed with 1 mL of 100 % ethanol. The beads were then resuspended in 22 µL 1X TE (10 mM Tris-HCL containing 1 mM EDTA, pH 8.0). Using a magnet to retain the beads, resuspended DNA were collected, pooled and quantified using the Qubit dsDNA HS assay (ThermoFisher Scientific, Waltham, MA, USA). DNA was then either stored at 4 °C for subsequent PCR that day or at -20 °C for extended periods of time.

B.1.4 DNA amplification

PCR reactions were done in triplicate and each final concentration contained: 0.2 µM of forward and reverse primers, 0.016 µg/µL BSA, 0.24 mM of each dNTP, 1.2 X PCR buffer, 6 mM $MgCl_2$, 0.6 U TAQ polymerase and 0.5 ng of DNA to a final volume of 25 µL. Prior to the addition of primers, TAQ polymerase and DNA, the PCR master mix was treated with ethidium monoazide bromide (1 mg/mL stock) to remove exogenous DNA in the PCR reagents[380]. The amount of ethidium monoazide bromide added varied for each batch of the reagent made. This was determined by serial dilution and the resultant highest concentration that did not inhibit PCR (a 1/100 dilution was typical). The master mix was then incubated on ice for 1 min in darkness, followed by 1 min photoactivation using a halogen lamp. All PCR reagents were supplied by Life Technologies (ThermoFisher Scientific, Waltham, MA, USA), except for ethidium monazide bromide (Mediray, Auckland, New Zealand). The

following PCR thermocycling parameters were performed: initial 3 min at 97 °C denaturation, 30 cycles of 45 sec at 94 °C, 60 sec at 50 °C and 90 sec at 72 °C, followed by a final 10 min incubation at 72 °C.

The triplicate amplicons were pooled and purified using SPRIselect (Beckman Coulter, Brea, CA, USA) as per the manufacturer's instructions (recommended ratio of 0.8X SPRI to amplicon volume). Quality and final concentration of the libraries were verified and adjusted to 12 pM via HS Qubit 2.0 (ThermoFisher Scientific, Waltham, MA, USA) and 9100 BioAnalyser (Agilent Technologies, Santa Clara, CA, USA). Amplicon sequencing was then performed using the Ion PGM System for Next-Generation Sequencing (ThermoFisher Scientific, Waltham, MA, USA) with the Ion 318 Chip Kit v2 and 400-base read length chemistry.

B.1.5 Sample filtering for spatial analyses

A total of 1,019 samples were taken from 18 different geothermal fields across the entire TVZ for this study from July 2013 to April 2015 (Figure 1). Twenty-eight of these produced insufficient DNA yields for sequencing. Twenty-two failed to generate adequate sequence reads for downstream processing (<9,500 reads). Twenty-one geothermal springs were also sampled over time for future investigation of temporal variation (n=66). Forty-four temporal repeats (excluding one already removed due to low sequence reads) were therefore removed from the dataset, leaving a final 925 individual geothermal springs for spatial statistical analyses. This also removed all springs sampled from the geothermal field Atiamuri, which left a final number of 17 individual fields analysed.

B.1.6 Geochemical filtering

To build the constrained correspondence analysis (CCA) model, the 46 physicochemical variables measured physicochemistry had to be reduced to a tractable number. Mantel testing of all variables against Bray-Curtis similarities calculated showed significant correlations (Table B.4). In order of highest to lowest mantel statistic, the variables were added to a permutational multivariate analysis of variance (the adonis function in the vegan package in R). pH and temperature had the highest contributions to this model (12.4 and 3.9 % respectively, P<0.001, Table B.5). Manganese, caesium, cadmium, selenium, cobalt, iron, zinc, barium, chromium, calcium and nickel had a P-value greater than 0.01 and were removed. Collinear variables were identified in the remaining set (Pearson's coefficient:

|r|>0.7) and group representatives with the highest mantel statistic with beta diversity were chosen. This removed rubidium, potassium, vanadium, mercury, sodium, boron, bromine, sulfur, chloride and ferrous iron. The model was re-run on remaining variables and subsequently, magnesium was removed (P=0.062). Of the 24 variables that remained, those with low variation (standard deviation <0.25 ppm) were removed (Table B.6) – this included strontium, copper, lead, nitrite, molybdenum, thallium, silver and uranium. The remaining 15 variables were added to the CCA model, with geothermal fields and spring communities (Figure 2.3). Two springs also had to be removed from this model due to insufficient chemical analyses (n=923).

a
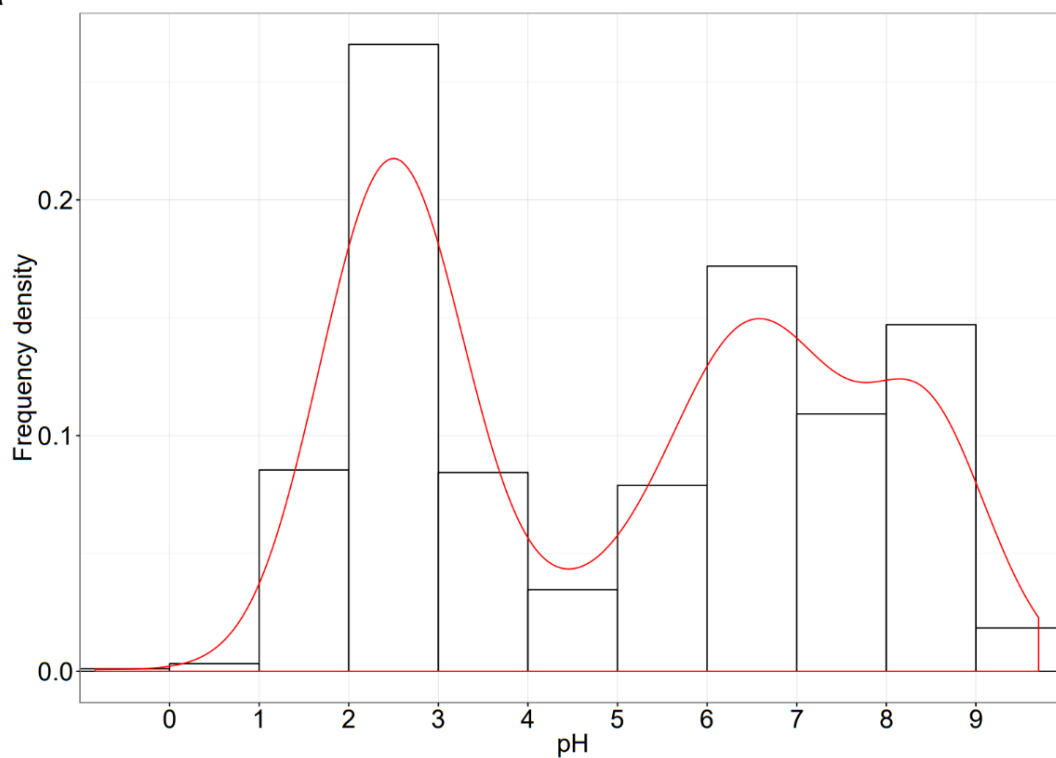
b
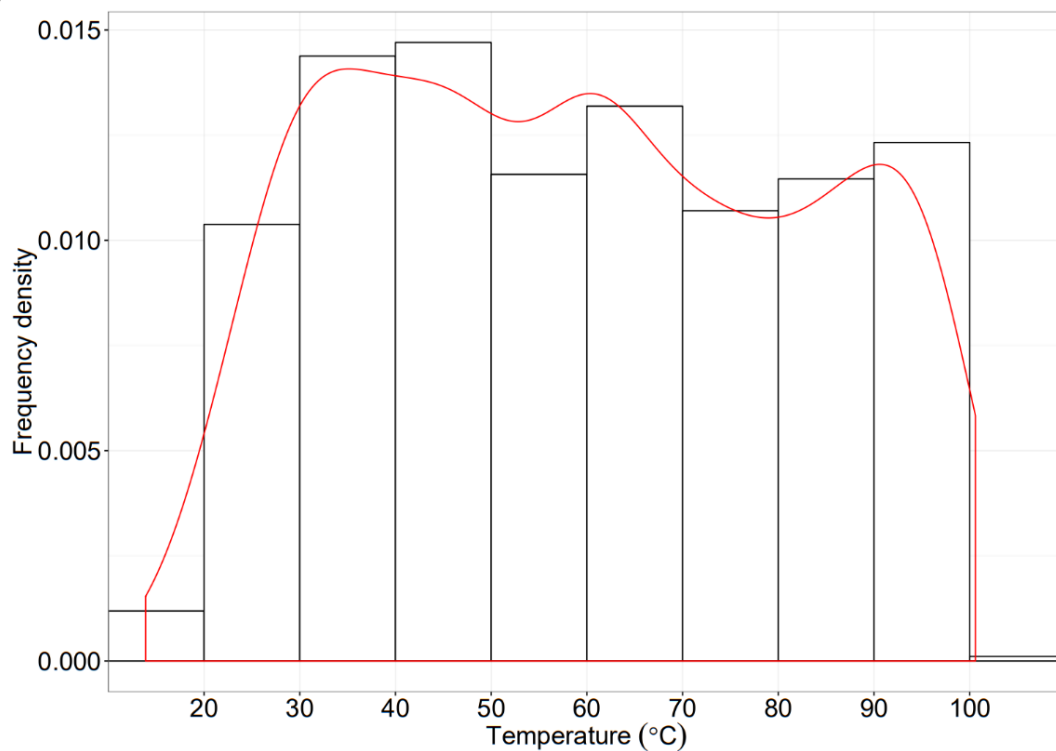
Figure B.1 - pH **(a)** and temperature **(b)** frequencies from all spring communities sampled. Red trendlines are a function of frequency density (*n*=925).
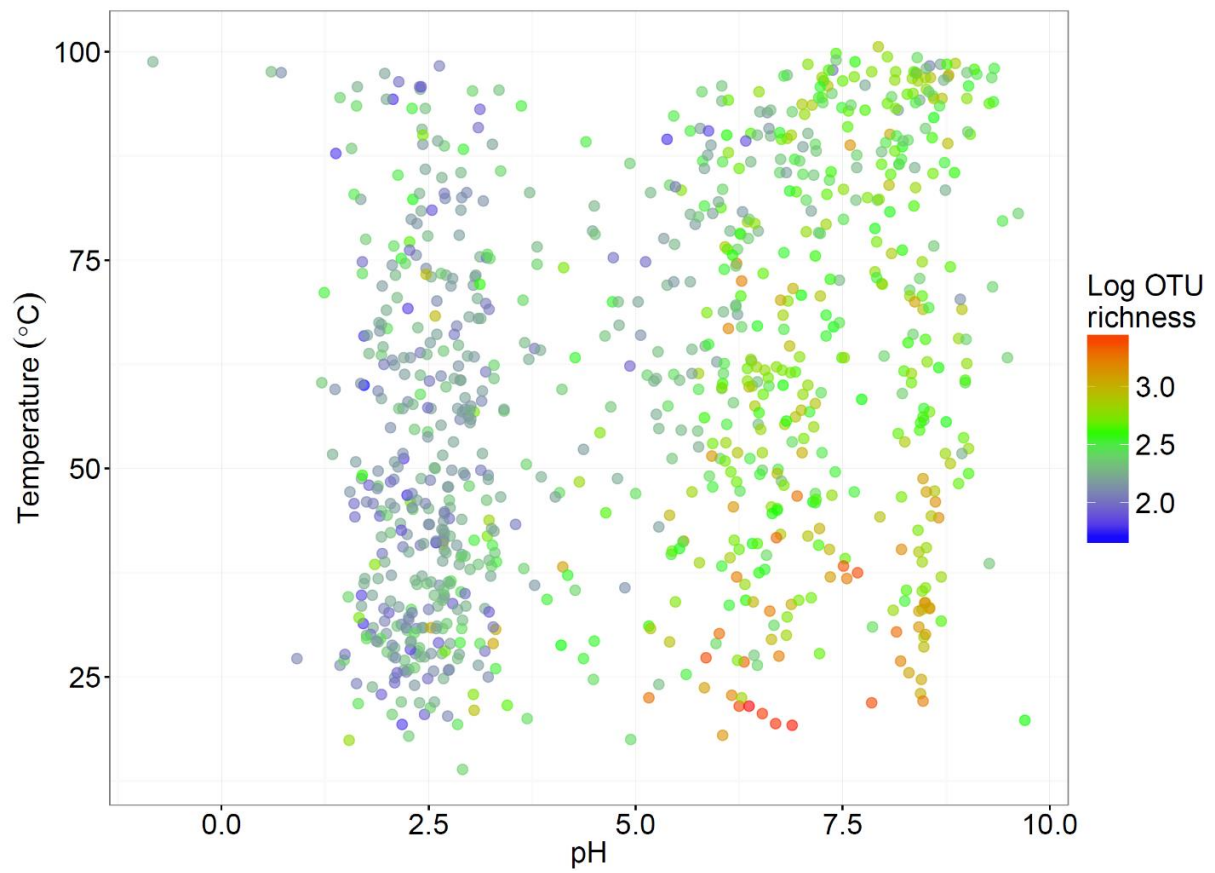
Figure B.2 - pH, temperature, and alpha diversity scales. A scatter plot of pH and temperature gradients for all springs sampled (*n*=925) is shown. The number of OTUs or richness is displayed in colour (range: 49-2997, mean: 386, median: 247).

a

$m = 0.157$, $R^2 = 0.164$, $P < 0.001$

b

$m = 0.002$, $R^2 = 0.002$, $P = 0.201$
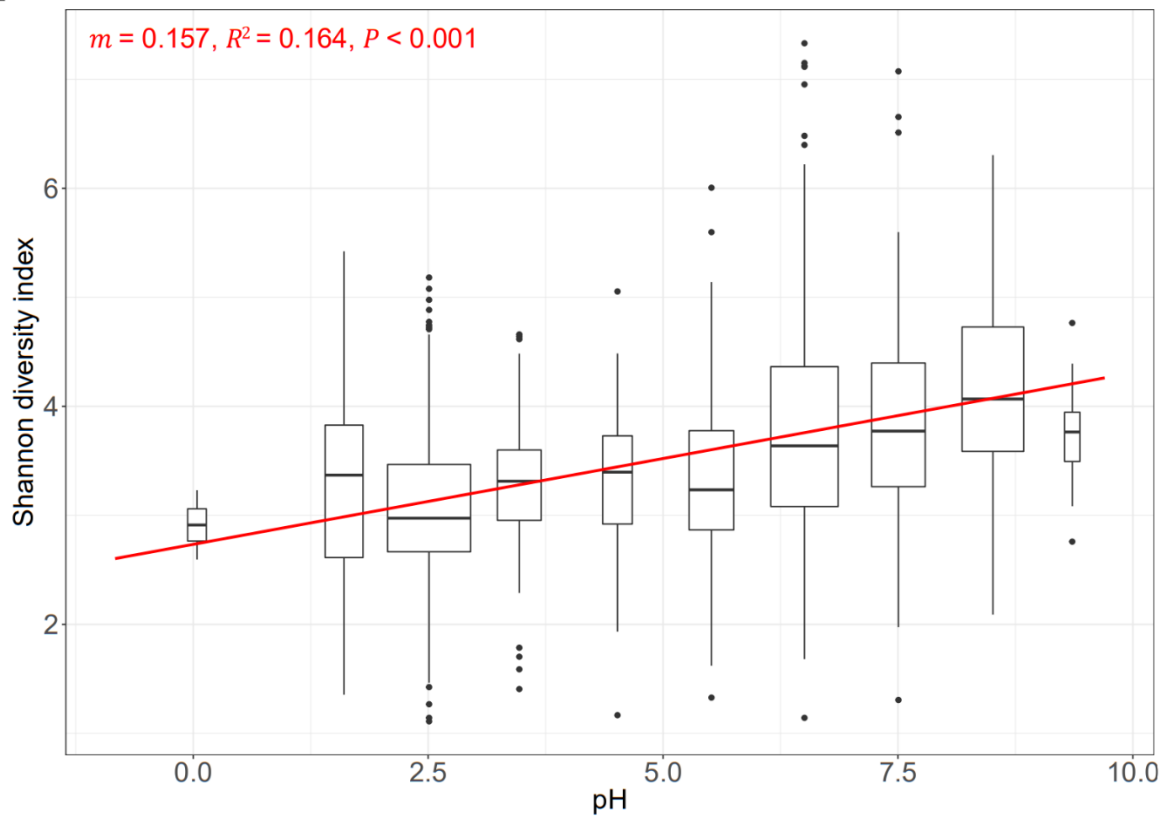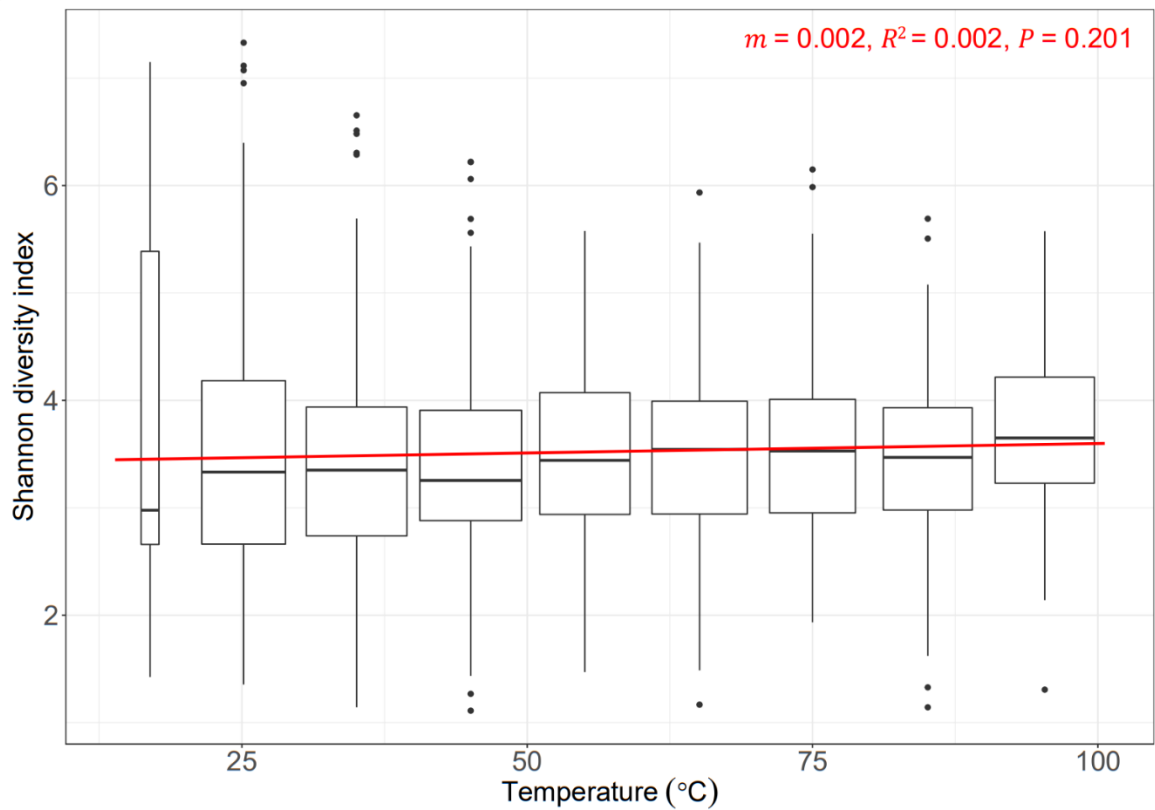
Figure B.3 - Alpha diversity against pH **(a)** and temperature **(b)**. Linear regression of Shannon diversity index against each variable is shown in red (*n*=925).

143

a
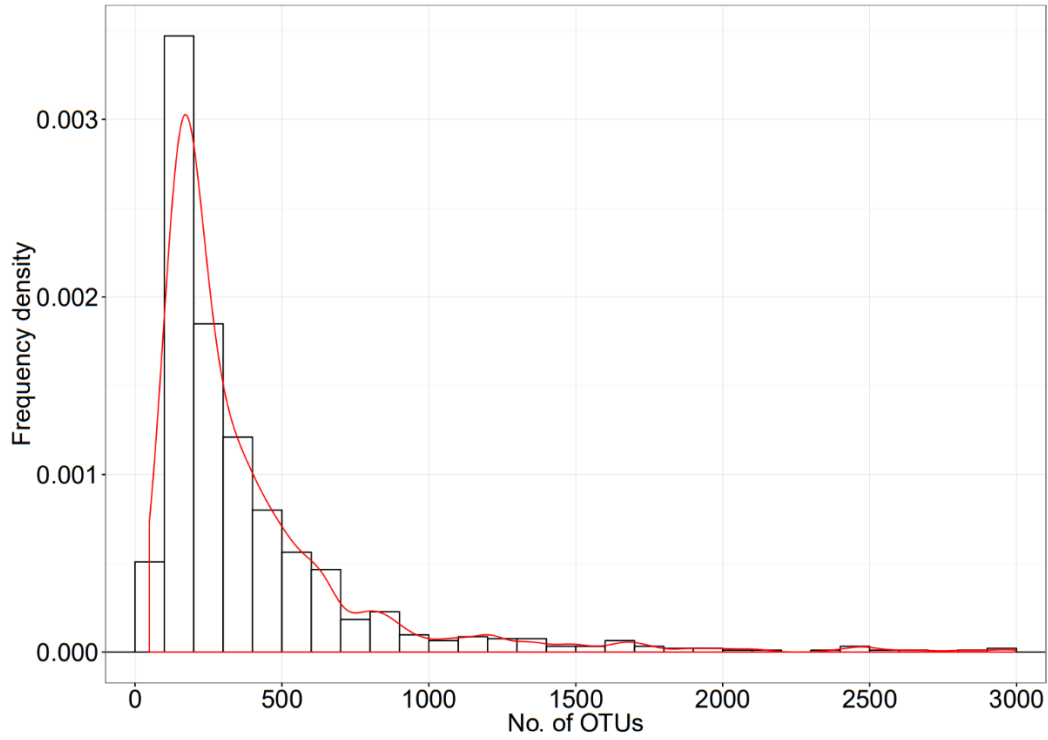
b
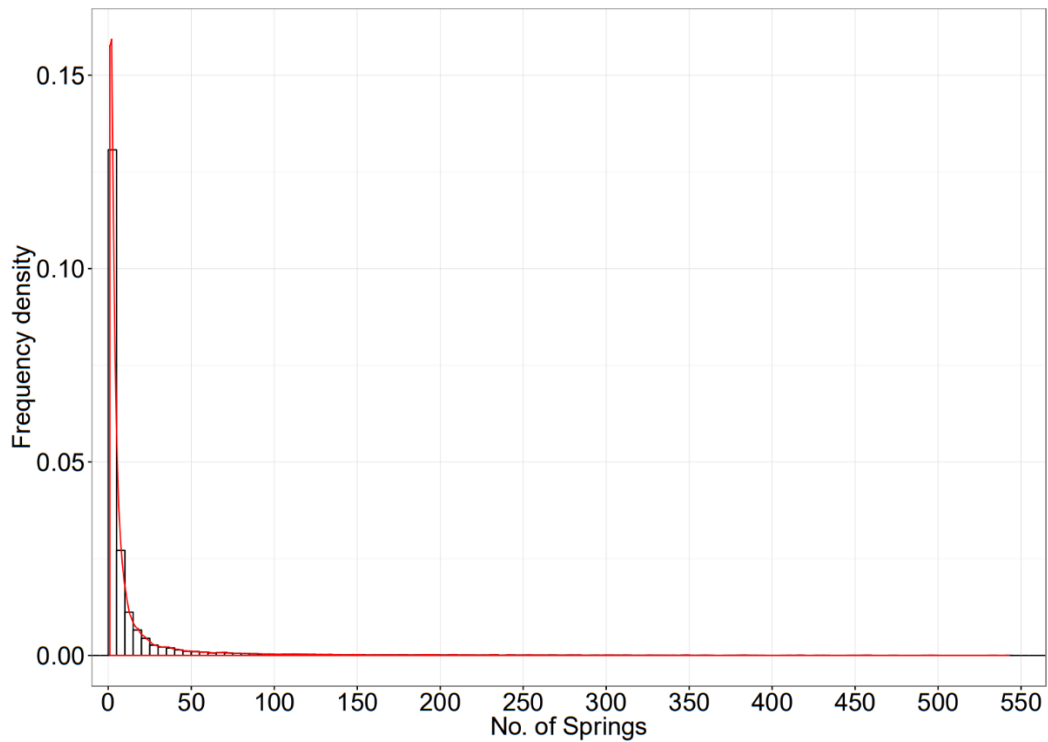
Figure B.4 - Quantity and distribution of OTUs across geothermal springs. **(a)** Histogram shows the number of OTUs found in each spring (n=28,381, range: 49-2,997, mean: 386, median: 247). **(b)** The distribution of each OTU across all springs (n=925, range: 1-547, mean: 13, median: 3).
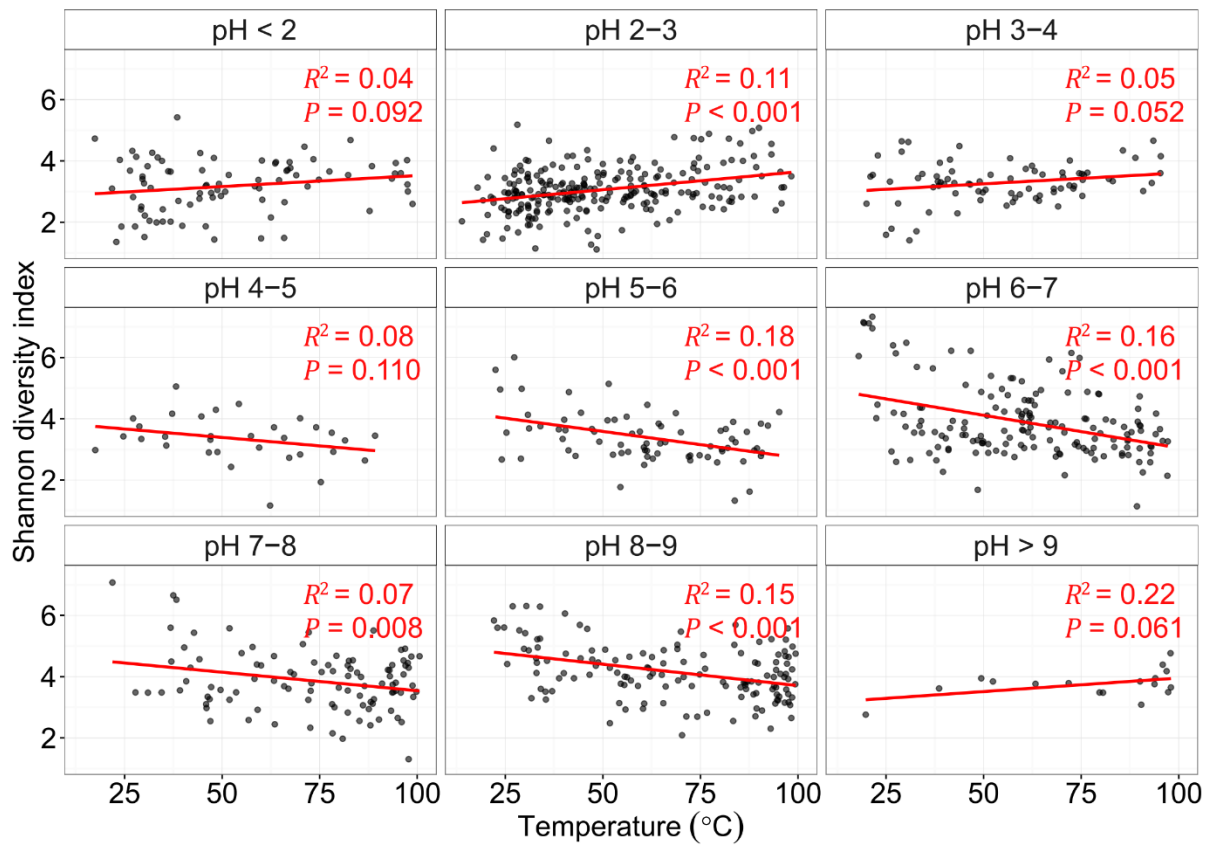
Figure B.5 - The relationship between alpha diversity and temperature. Samples (*n*=925) are split into pH increments. Linear regression of each increment against Shannon diversity index is shown in red.
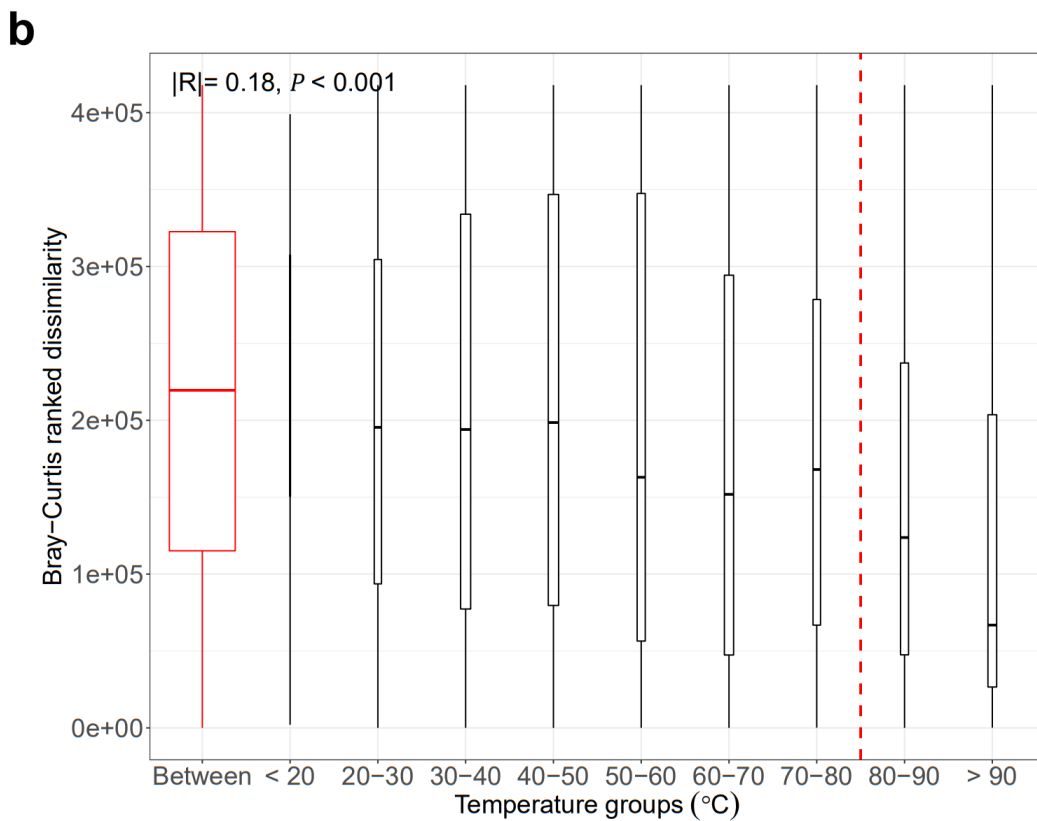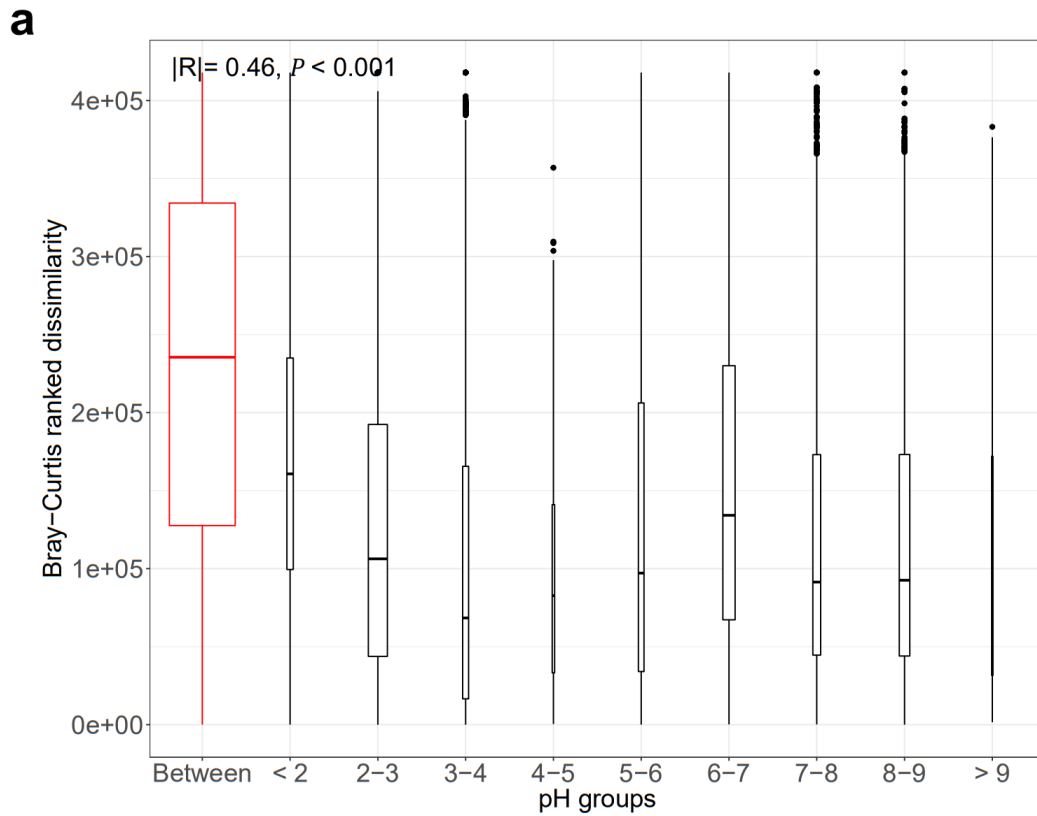
Figure B.6 - Analysis of similarities (ANOSIM: |R|) of pH and temperature binned samples (*n*=925) against beta diversity. The variation between pH **(a)** and temperature **(b)** groups is shown in red, with the variation within individual groups in black. The red dashed line in (b) refers to the greatest significant difference between individual temperature groups (Wilcoxon test: $P<2\text{x}10^{-16}$; >80 °C).
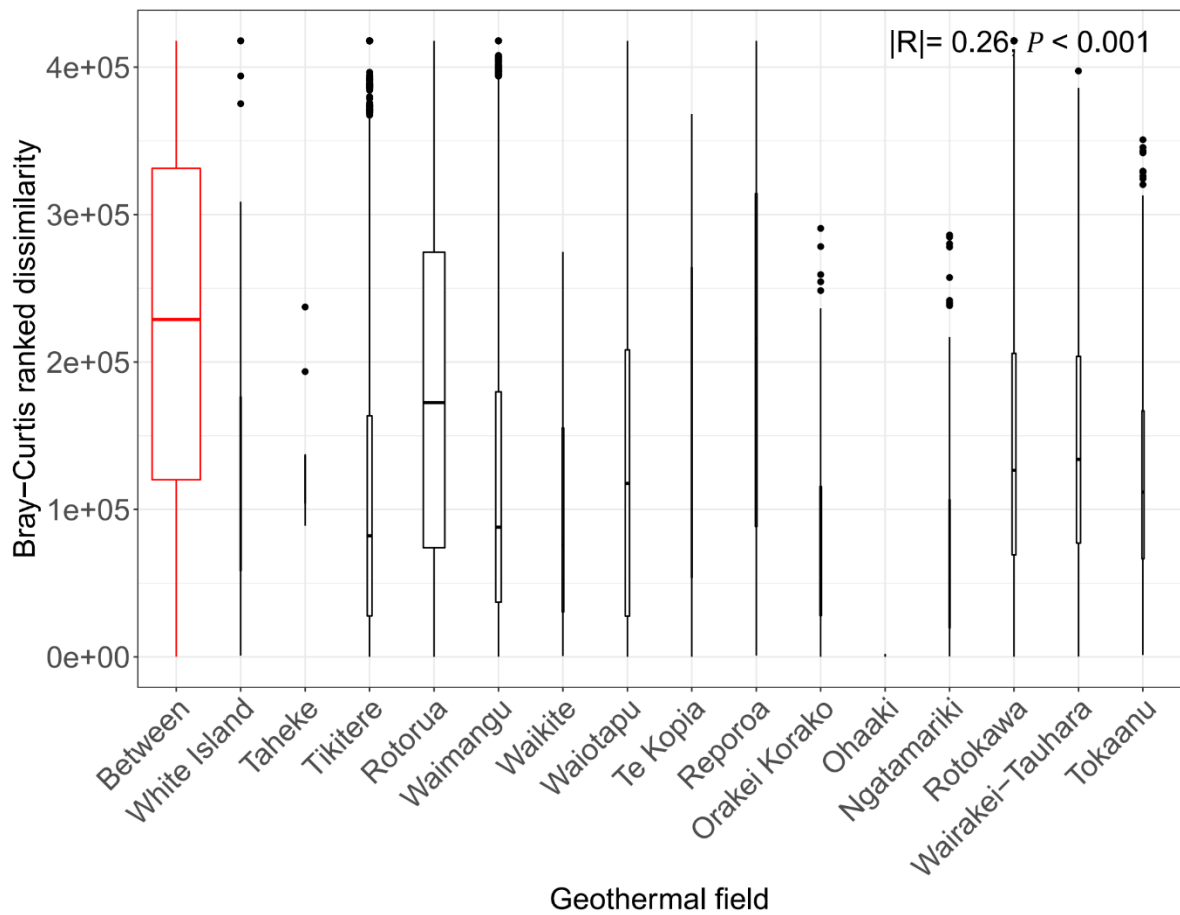
Figure B.7 - Analysis of similarities (ANOSIM: |R|) between geothermal fields and beta diversity. The overall variation between fields is plotted in red, with the variation within each individual field in black. The width of each bar indicates the number of springs per site (*n*=925 in total). Two geothermal fields (Whangairorohea and Misc) had only one spring sampled and therefore are not shown in this analysis.
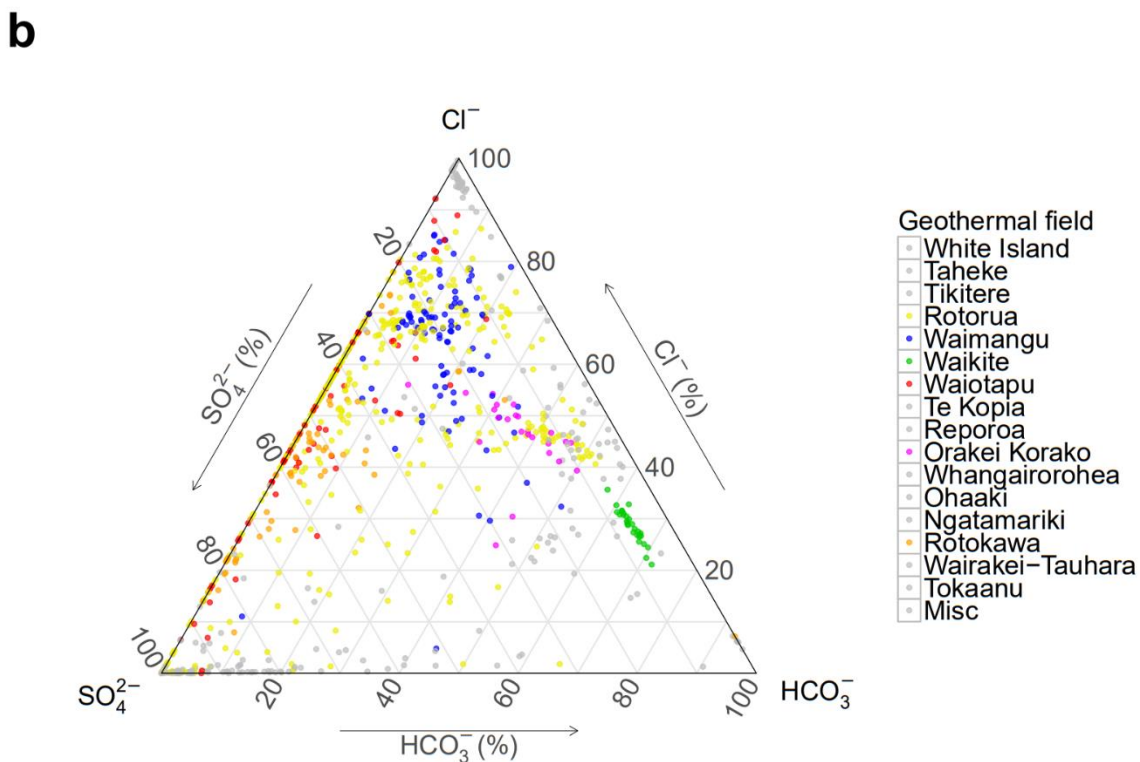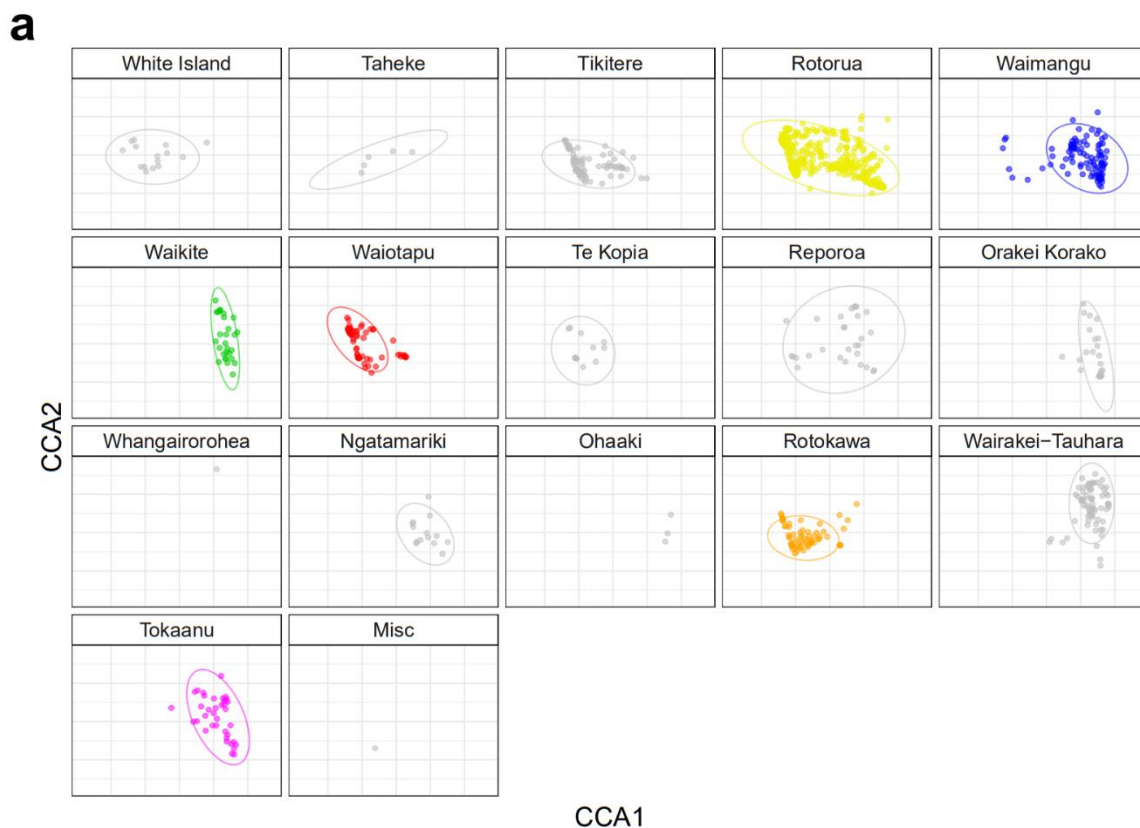
Figure B.8 - Constrained correspondence analysis (CCA) of geothermal fields and geochemical signatures. **(a)** CCA ordination for each geothermal field – full model is shown in Figure 2.3 (*n*=923). Data ellipses are plotted as 95 % confidence intervals with normal distribution. **(b)** Ternary diagram of geochemical signatures for springs sampled (*n*=923). Points are colour-coded to geothermal fields shown individually in Figure 2.3.

a

b
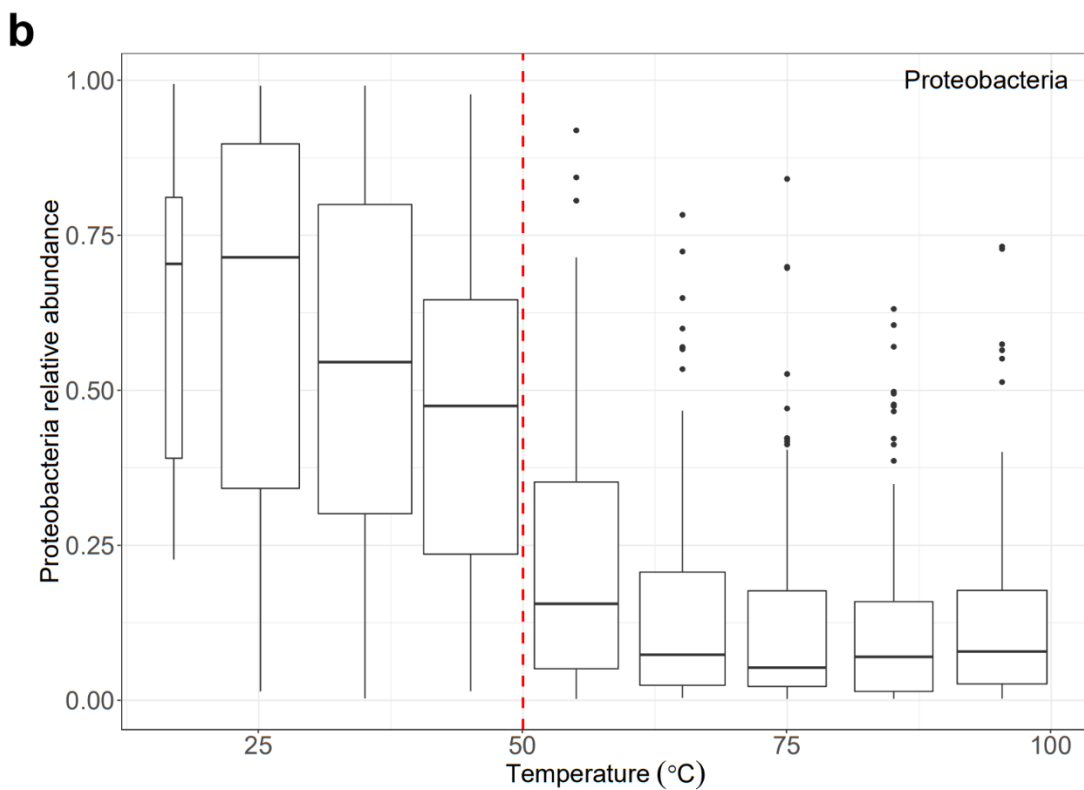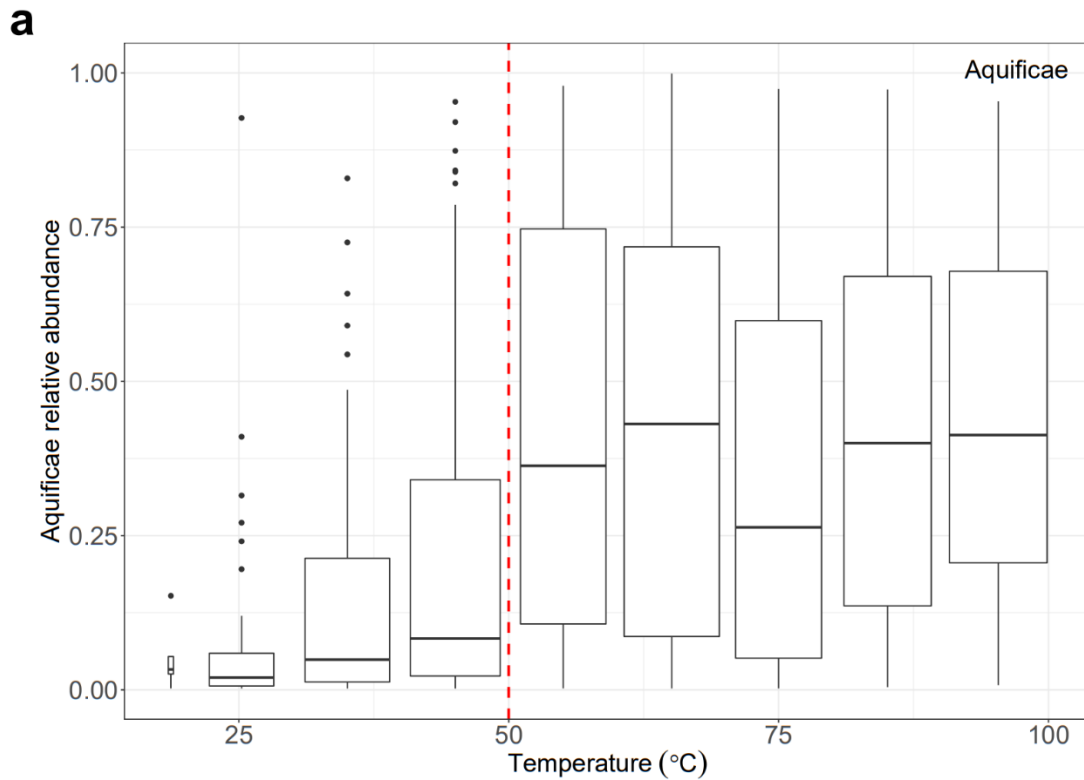
Figure B.9 - Relative abundance of **(a)** Aquificae* and **(b)** Proteobacteria* in each spring against temperature. The boxplots are binned by 10 °C increments. The red dotted line marks a substantial change in abundance levels at 50 °C (*n*=925 for both plots).

*The phyla Aquificae and Proteobacteria are now known as Aquificota and Pseudomonadota, respectively, due to nomenclature changes in bacterial taxonomy in 2021[198].

B.3 Supplementary tables

Table B.1 - A list of all DNA and physicochemical samples taken from each spring, and the subsequent processing and analyses performed for each individual parameter.

Parameter	Volume (mL)	Processing	Storage	Analytical Method
Microbial diversity	2,000	Filtered	-20 °C	DNA extraction, sequencing
Temperature	n.a.	Raw	NA (*in situ*)	Digital thermocouple
*Physical properties	500	Raw	RT	Multiparameter field meter
Fe^{2+}	0.08	Filtered	RT	Colorimetric spectroscopy
H_2S	330	Raw	4 °C	Iodometric titration
HCO_3^-	330	Raw	4 °C	HCO_3 titration
Cl^-	330	Raw	4 °C	Potentiometric titration/IC
SO_4^{2-}	50	Filtered	4 °C	IC
NH_4^+, PO_4^{3-}, NO_2^-, NO_3^-	50	Filtered	-20 °C	FIA
Back up sample	50	Filtered	4 °C	NA
**Elemental analysis	15	Filtered, acidified	4 °C	ICP-MS

n.a.: not applicable, **RT:** room temperature, **IC:** ion chromatography, **FIA:** flow injection analysis, **ICP-MS:** inductively coupled plasma-mass spectroscopy**.**
*Physical properties measured by the field meter were pH, oxidation-reduction potential (ORP), conductivity (COND), turbidity (TURB) and dissolved oxygen (dO)
**Elements measured using ICP-MS were Al, Ag, As, B, Ba, Br, Ca, Cd, Co, Cr, Cu, Cs, Fe, Hg, K, Li, Mg, Mo, Mn, Na, Ni, Pb, Rb, S, Se, Si, Sr, Tl, U, V and Zn.

Table B.2 - Linear regression of physicochemistry individually against alpha diversity using the Shannon diversity index.

	Slope	Std. Error	t-value	*P*-value	Signif. Codes	R^2
pH	0.1574	0.0117	13.480	<0.0001	***	0.164
NO_3^-	0.8205	0.0834	9.838	<0.0001	***	0.095
SO_4^{2-}	-0.0003	0.0000	-6.620	<0.0001	***	0.045
TURB	-0.0009	0.0001	-6.382	<0.0001	***	0.042
H_2S	-0.0140	0.0026	-5.304	<0.0001	***	0.030
Al	-0.0049	0.0010	-5.100	<0.0001	***	0.027
dO	0.0727	0.0148	4.899	<0.0001	***	0.025
S	-0.0006	0.0001	-4.675	<0.0001	***	0.023
U	-350.7932	76.2842	-4.599	<0.0001	***	0.022
HCO_3^-	0.0011	0.0003	4.229	<0.0001	***	0.019
Li	0.0321	0.0080	4.009	<0.0001	***	0.017
Na	0.0004	0.0001	3.887	<0.0001	***	0.016
As	0.0627	0.0195	3.216	0.0013	**	0.011
Cs	0.1131	0.0373	3.030	0.0025	**	0.010
K	0.0020	0.0007	2.831	0.0047	**	0.009
Br	0.0355	0.0133	2.664	0.0079	**	0.008
Se	14.3009	5.4569	2.621	0.0089	**	0.007

ORP	-0.0005	0.0002	-2.596	0.0096	**	0.007
NH_4^+	-0.0014	0.0005	-2.594	0.0096	**	0.007
Rb	0.1298	0.0534	2.429	0.0153	*	0.006
Fe	-0.0027	0.0012	-2.319	0.0206	*	0.006
Fe^{2+}	-0.0023	0.0010	-2.317	0.0207	*	0.006
NO_2^-	0.9100	0.4052	2.246	0.0250	*	0.005
B	0.0051	0.0023	2.188	0.0289	*	0.005
Mn	-0.0555	0.0297	-1.870	0.0619	.	0.004
V	-0.6407	0.3791	-1.690	0.0913	.	0.003
Si	-0.0010	0.0006	-1.656	0.0980	.	0.003
Cd	36.1309	23.0894	1.565	0.1180		0.003
Co	-9.3257	6.3612	-1.466	0.1430		0.002
Cr	-0.2915	0.2145	-1.359	0.1740		0.002
TEMP	0.0018	0.0014	1.279	0.2010		0.002
Ca	-0.0005	0.0004	-1.175	0.2400		0.001
Ni	-1.0608	0.9284	-1.143	0.2540		0.001
Mo	2.7628	2.4559	1.125	0.2610		0.001
COND	0.0000	0.0000	-1.098	0.2720		0.001
Hg	-2.3407	2.1961	-1.066	0.2870		0.001
Mg	-0.0009	0.0008	-1.022	0.3070		0.001
Ba	-0.0648	0.0639	-1.015	0.3100		0.001
Ag	-43.2878	47.1283	-0.919	0.3590		0.001
Zn	-0.1021	0.1115	-0.916	0.3600		0.001
Cl^-	0.0000	0.0000	-0.700	0.4840		0.001
Tl	5.0194	7.6863	0.653	0.5140		0.000
PO_4^{3-}	0.0269	0.1135	0.237	0.8130		0.000
Pb	-0.0390	0.3551	-0.110	0.9130		0.000
Sr	0.0100	0.1433	0.070	0.9440		0.000
Cu	0.0149	0.2460	0.060	0.9520		0.000

Signif. codes: 0-0.001 '***'; 0.001-0.01 '**'; 0.01-0.05 '*'; 0.05-0.1 '.'; 0.1-1 ' '
TURB: turbidity, dO: dissolved oxygen, ORP: oxidation-reduction potential, COND: conductivity

Table B.3 - Multiple linear regression model of significant physicochemical parameters against alpha diversity (Shannon Index), after collinear variables were removed and an Akaike information criterion (AIC) was applied. Variables were added to the model in order of highest to lowest best fit from singular linear regression.

	Estimate	Std. Error	t-value	*P*-value	Signif. codes
(Intercept)	2.7116	0.1107	24.502	<0.0001	***
pH	0.1568	0.0140	11.201	<0.0001	***
NO_3^-	0.7690	0.0905	8.502	<0.0001	***
TURB	-0.0002	0.0001	-1.768	0.0774	.
dO	0.0498	0.0131	3.807	0.0002	***
ORP	0.0004	0.0002	2.219	0.0268	*
NO_2^-	-1.6099	0.4104	-3.923	0.0001	***
Si	-0.0014	0.0005	-2.689	0.0073	**
Cd	48.7715	19.8181	2.461	0.0140	*

Multiple R-squared: 0.273, adjusted R-squared: 0.2666
F-statistic: 42.99 on 8 and 916 degrees of freedom, p-value: < 2.2e-16
Signif. codes: 0-0.001 '***'; 0.001-0.01 '**'; 0.01-0.05 '*'; 0.05-0.1 '.'; 0.1-1 ' '
TURB: turbidity, dO: dissolved oxygen

Table B.4 - Mantel tests using Spearman's correlation (permutations=999) of Bray-Curtis dissimilarities between all communities sampled and each individual geochemical parameter and geographic distance (km).

	Rho (ρ)	*P*-value	Signif. codes
pH	0.544	<0.001	***
SO_4^{2-}	0.248	<0.001	***
Al	0.224	<0.001	***
TEMP	0.193	<0.001	***
HCO_3^-	0.153	<0.001	***
TURB	0.145	<0.001	***
S	0.131	<0.001	***
km	0.122	<0.001	***
NH_4^+	0.103	<0.001	***
As	0.102	<0.001	***
COND	0.098	<0.001	***
Mo	0.086	<0.001	***
Fe	0.082	<0.001	***
Li	0.082	<0.001	***
Ag	0.079	<0.001	***
Tl	0.074	<0.001	***
NO_2^-	0.073	<0.001	***
V	0.072	<0.001	***
U	0.069	<0.001	***
Rb	0.068	<0.001	***
Na	0.068	<0.001	***
Cs	0.064	<0.001	***
Mn	0.063	<0.001	***
B	0.059	<0.001	***
K	0.058	<0.001	***
Mg	0.057	<0.001	***

ORP	0.057	<0.001	***
Cl$^-$	0.056	<0.001	***
Fe^{2+}	0.056	<0.001	***
Hg	0.055	<0.001	***
NO$_3^-$	0.055	<0.001	***
Si	0.046	<0.001	***
Sr	0.044	<0.001	***
Co	0.040	0.002	**
Br	0.038	0.002	**
Zn	0.044	<0.001	***
dO	0.041	<0.001	***
Se	0.031	0.003	**
Ca	0.029	0.007	**
Ba	0.020	0.017	*
Cd	0.020	0.068	.
H$_2$S	0.014	0.142	
Cr	0.007	0.276	
Pb	0.006	0.297	
PO$_4^{3-}$	0.005	0.344	
Cu	-0.008	0.736	
Ni	-0.020	0.953	

Signif. codes: 0-0.001 '***'; 0.001-0.01 '**'; 0.01-0.05 '*'; 0.05-0.1 '.'; 0.1-1 ' '
TEMP: temperature, TURB: turbidity, km: kilometres, COND: conductivity, dO: dissolved oxygen, ORP: oxidation reduction potential

Table B.5 - Permutational multivariate analysis of variance (using continuous variables only) of beta diversity using Bray-Curtis dissimilarities.

	Df	SumsOfSqs	MeanSqs	F.Model	R^2	P-value	Signif. codes
pH	1	44.13	44.125	154.945	0.124	0.001	***
TEMP	1	14.03	14.026	49.253	0.039	0.001	***
ORP	1	5.06	5.056	17.755	0.014	0.001	***
SO_4^{2-}	1	2.88	2.883	10.123	0.008	0.001	***
TURB	1	2.71	2.708	9.508	0.008	0.001	***
As	1	2.51	2.512	8.820	0.007	0.001	***
NO_3^-	1	2.25	2.249	7.896	0.006	0.001	***
NH_4^+	1	1.99	1.989	6.986	0.006	0.001	***
HCO_3^-	1	1.84	1.839	6.457	0.005	0.001	***
Rb	1	1.76	1.757	6.169	0.005	0.001	***
Ag	1	1.39	1.389	4.877	0.004	0.001	***
NO_2^-	1	1.25	1.254	4.405	0.004	0.001	***
K	1	1.22	1.223	4.294	0.003	0.001	***
COND	1	1.14	1.143	4.014	0.003	0.001	***
dO	1	1.10	1.096	3.848	0.003	0.001	***
H_2S	1	1.00	1.001	3.514	0.003	0.001	***
Tl	1	1.00	0.998	3.506	0.003	0.001	***
V	1	0.96	0.959	3.369	0.003	0.001	***
Mg	1	0.88	0.880	3.090	0.002	0.001	***
Al	1	0.84	0.838	2.942	0.002	0.001	***
Si	1	0.83	0.830	2.916	0.002	0.001	***
Sr	1	0.75	0.753	2.643	0.002	0.001	***
Hg	1	0.72	0.718	2.522	0.002	0.001	***
Mo	1	0.64	0.640	2.247	0.002	0.001	***
Pb	1	0.56	0.565	1.984	0.002	0.001	***
Na	1	1.02	1.020	3.582	0.003	0.002	**
B	1	0.95	0.945	3.320	0.003	0.002	**
Br	1	0.88	0.880	3.089	0.002	0.002	**
Li	1	0.87	0.866	3.041	0.002	0.002	**
PO_4^{3-}	1	0.65	0.653	2.292	0.002	0.002	**
Cu	1	0.61	0.612	2.149	0.002	0.002	**
U	1	0.74	0.744	2.613	0.002	0.003	**
S	1	0.83	0.827	2.903	0.002	0.004	**
Cl^-	1	0.75	0.748	2.628	0.002	0.005	**
Fe^{2+}	1	0.57	0.568	1.995	0.002	0.005	**
Mn	1	0.58	0.581	2.040	0.002	0.011	*
Cs	1	0.59	0.592	2.078	0.002	0.014	*
Cd	1	0.39	0.394	1.384	0.001	0.019	*
Se	1	0.47	0.473	1.660	0.001	0.055	.
Co	1	0.38	0.379	1.329	0.001	0.142	
Fe	1	0.36	0.357	1.255	0.001	0.145	
Zn	1	0.37	0.367	1.289	0.001	0.146	
Ba	1	0.37	0.369	1.296	0.001	0.172	
Cr	1	0.35	0.355	1.247	0.001	0.187	
Ca	1	0.34	0.344	1.209	0.001	0.225	
Ni	1	0.23	0.227	0.797	0.001	0.758	

Signif. codes: 0-0.001 '***'; 0.001-0.01 '**'; 0.01-0.05 '*'; 0.05-0.1 '.'; 0.1-1 ' '
TURB: turbidity, dO: dissolved oxygen, ORP: oxidation-reduction potential, COND: conductivity

Table B.6 - Variance and standard deviation (SD) of significant chemistry (ppm) correlating with beta diversity. Parameters with SD<0.25 ppm were removed for building the constrained correspondence analysis (CCA) model with a tractable number of variables.

	Variance(σ^2)	SD (ppm)
SO_4^{2-}	404969.24	636.37
HCO_3^-	15489.52	124.46
NH_4^+	3417.50	58.46
Si	2945.29	54.27
Al	452.90	21.28
H_2S	141.45	11.89
Li	15.49	3.94
As	2.58	1.61
NO_3^-	0.13	0.36
PO_4^{3+}	0.08	0.28
Sr	0.04	0.20
Cu	0.02	0.13
Pb	0.01	0.09
NO_2^-	0.01	0.08
Mo	0.00	0.01
Tl	0.00	0.00
Ag	0.00	0.00
U	0.00	0.00

Table B.7 - Linear regression of spring community dissimilarity in each geothermal field against geographic distance. Geothermal fields Whangairorohea, Ohaaki, and Misc were removed from this analysis due to low spring numbers present ($n<3$). Fields are ordered north to south.

	Slope	Std. Error	t-value	P-value	Signif. Codes	R^2
White Island	0.0003	0.0001	2.925	0.0042	**	0.077
Taheke	0.0003	0.0003	0.995	0.349		0.11
Tikitere	0	0	2.57	0.0102	*	0.002
Rotorua	0	0	25.85	<0.0001	***	0.01
Waimangu	0	0	1.858	0.0632	.	0.001
Waikite	0.0001	0	4.389	0	***	0.046
Waiotapu	0	0	10.11	<0.0001	***	0.04
Te Kopia	0	0.0001	-0.117	0.907		0
Reporoa	0	0	6.541	0	***	0.096
Orakei Korako	0.0001	0.0001	1.686	0.0929	.	0.01
Ngatamariki	0.0001	0.0001	2.566	0.0111	*	0.035
Rotokawa	0.0001	0	9.565	<0.0001	***	0.041
Wairakei-Tauhara	0	0	21.2	<0.0001	***	0.153
Tokaanu	0.0005	0	12.28	<0.0001	***	0.185

Signif. codes: 0-0.001 '***'; 0.001-0.01 '**'; 0.01-0.05 '*'; 0.05-0.1 '.'; 0.1-1 ' '

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

C.1 SUPPLEMENTARY FIGURES
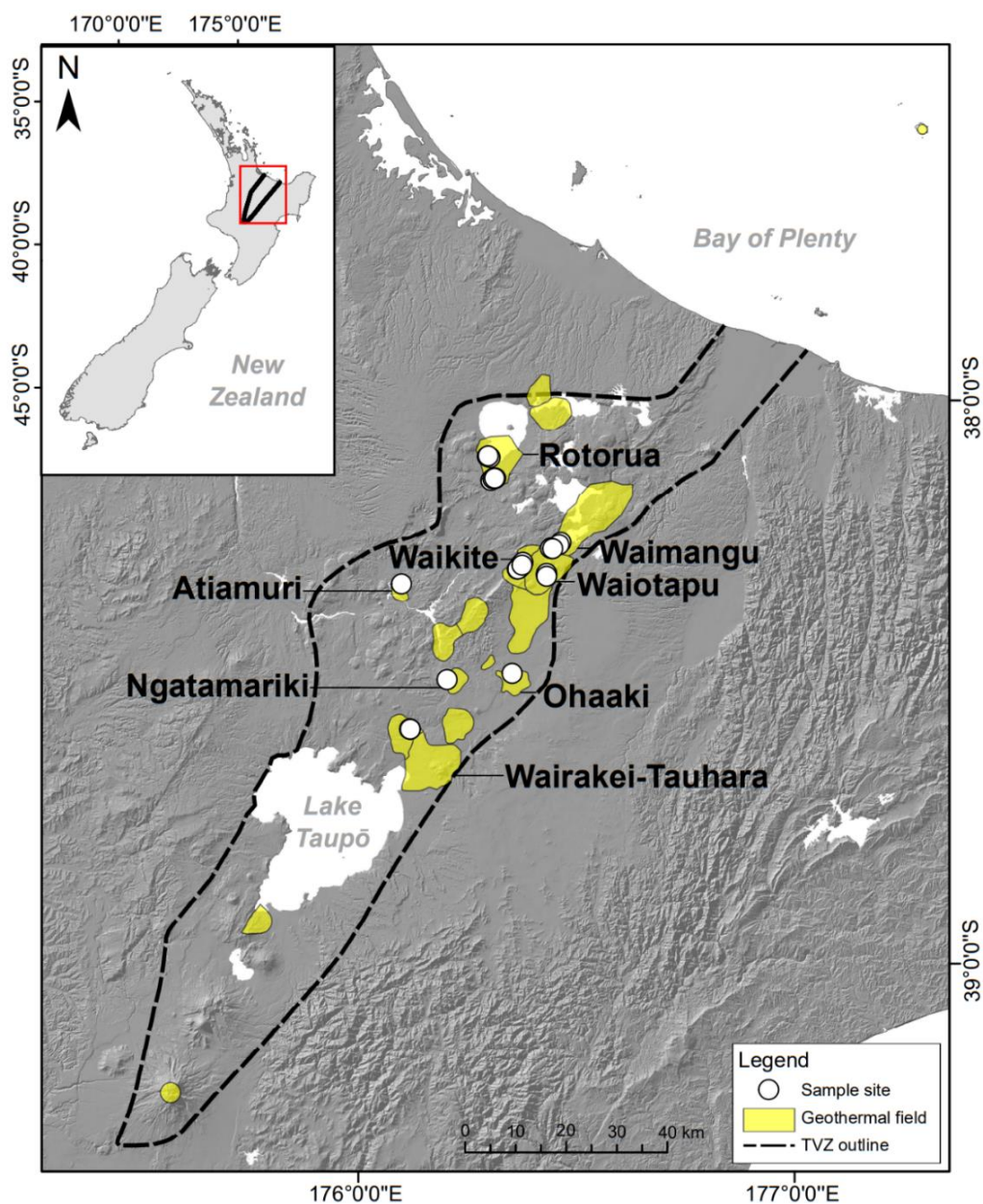
Figure C.1 - Map of the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand. Geothermal fields are presented in yellow, with fields used in this study labelled. Geothermal features sampled temporally are highlighted by white circles (*n*=31). The panel insert displays the location of the TVZ in the central North Island of Aotearoa-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*. (1995).
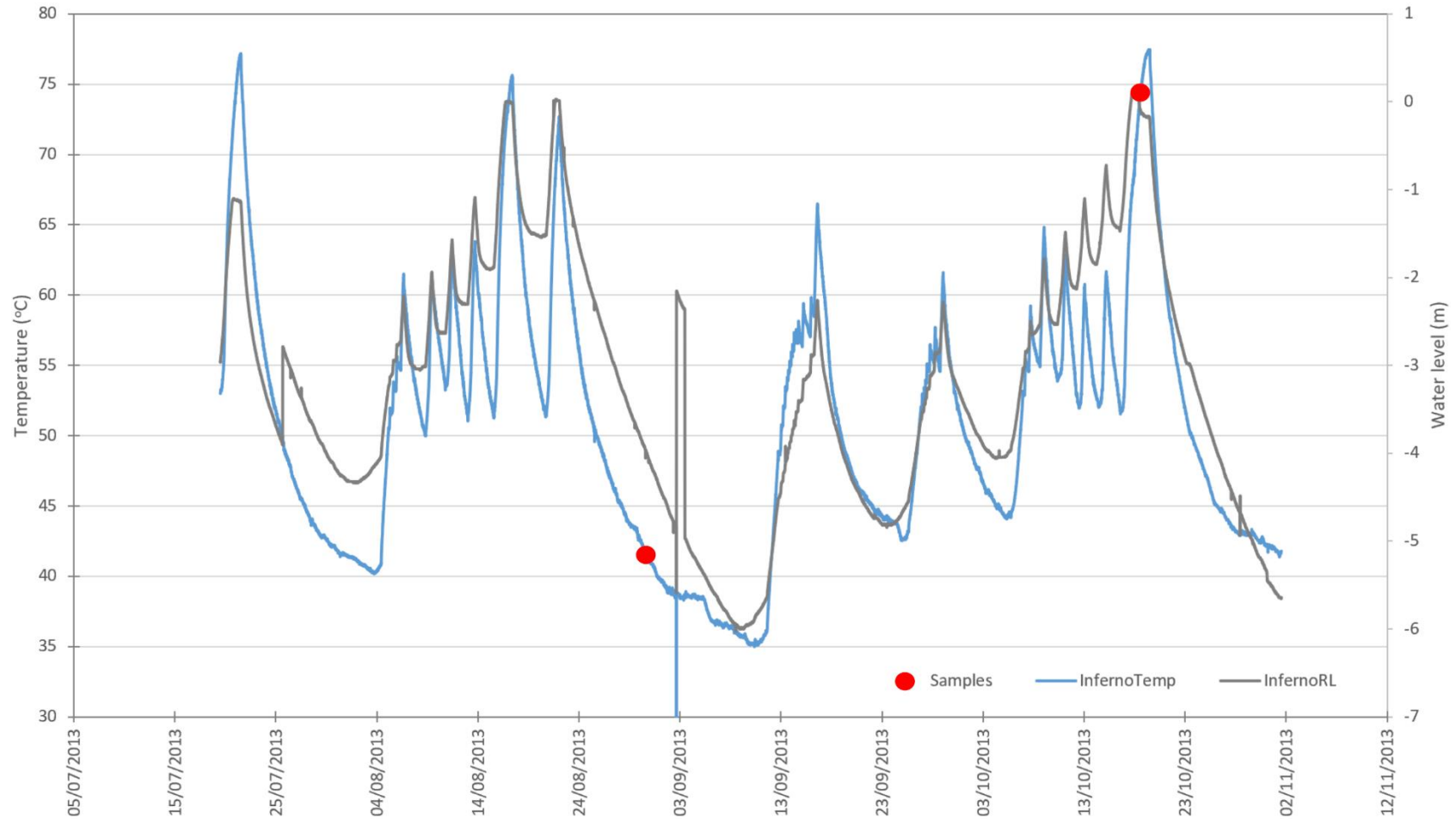
Figure C.2 - Temperature and water level profile of Inferno Crater Lake. The ~30-40 day cycle of Inferno Crater Lake in the Waimangu geothermal field is indicated by the rise and fall of temperature on the y-axis, with overflow events into Waimangu Stream occuring when lake temperatures reached >70 °C. The corresponding reduced water level (RL) of the lake is also shown in grey. Timepoints of samples taken for this study are represented in red**.**
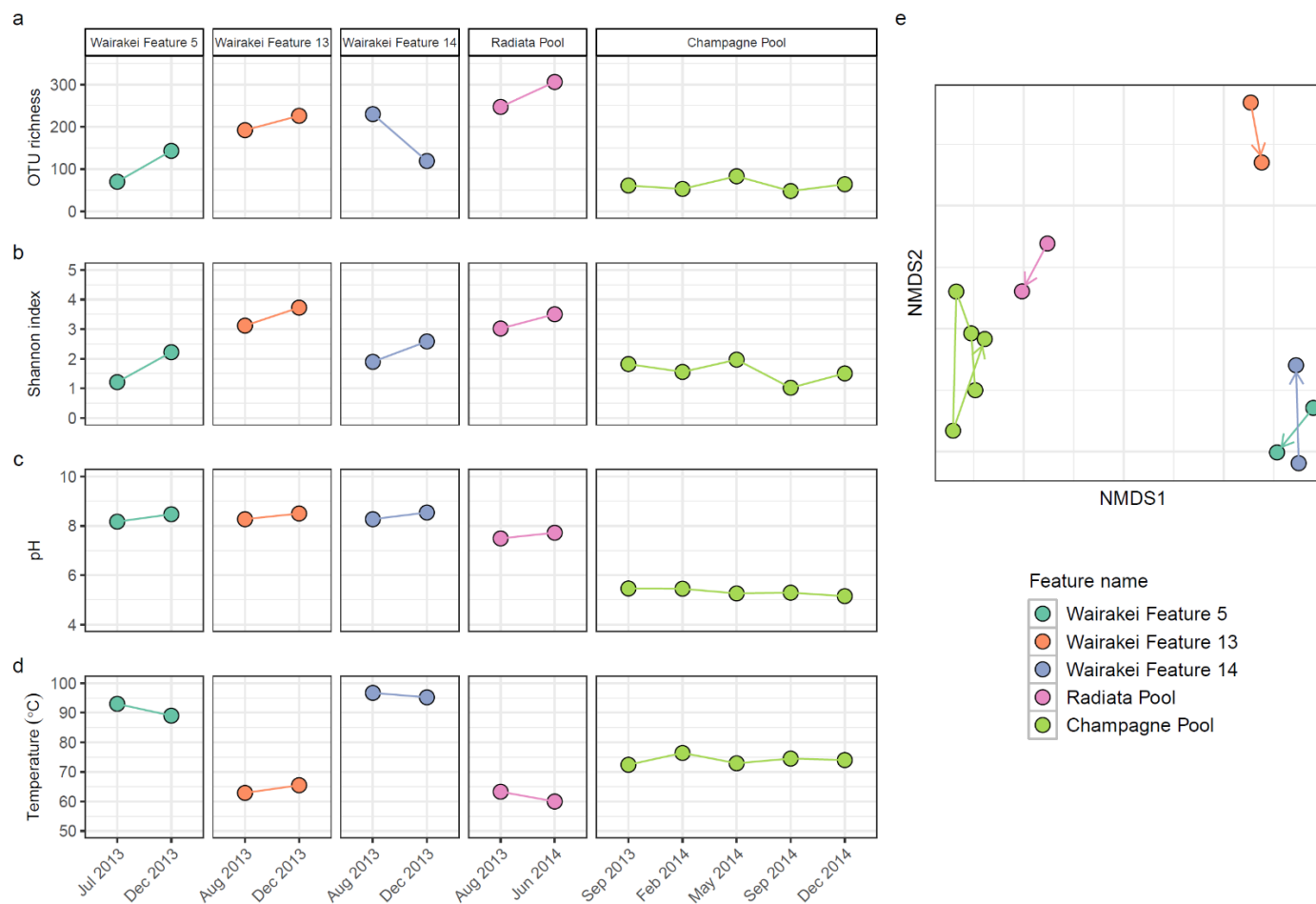
Figure C.3 - Temporal diversity and physicochemistry of control features in the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand. **(a)** The microbial communities of five control features were measured by amplicon sequencing of the 16S rRNA gene, with variation in alpha diversity between temporal samples indicated by OTU richness (i.e., the number of OTUs per community). **(b)** Alpha diversity was also measured by the Shannon diversity index to indicate evenness of the communities over time. **(c)** Spring pH of the features at the time of sampling. **(d)** Spring temperature of the control features. **(e)** Beta diversity of all five control features, distinguished by colour, is shown by a non-metric multidimensional scaling (NMDS; n=13, stress=0.1) of Bray-Curtis dissimilarities between samples. The timeline of sampling is indicated by the direction of the lines and arrows.
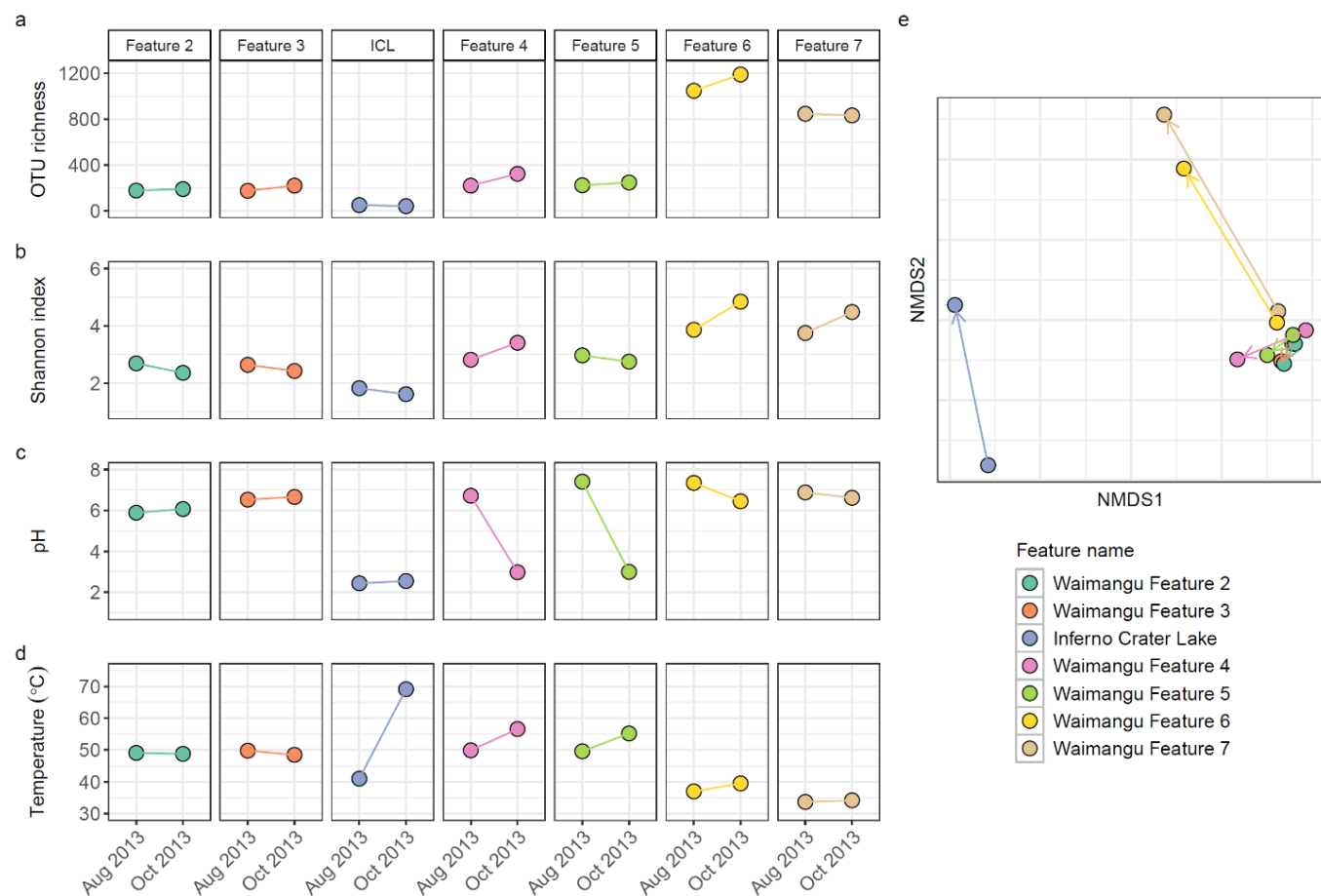
Figure C.4 - Temporal diversity and physicochemistry of disturbed features at Waimangu geothermal field. **(a)** The microbial communities of Waimangu Stream sites and Inferno Crater Lake (ICL) were measured by amplicon sequencing of the 16S rRNA gene, with variation in alpha diversity between samples indicated by OTU richness (i.e., the number of OTUs per community). **(b)** Alpha diversity was also measured by the Shannon diversity index to indicate evenness of the communities over time. **(c)** The pH of features at the time of sampling. **(d)** Temperature of both stream sites and Inferno Crater Lake. **(e)** Bray-Curtis dissimilarity was calculated for beta diversity, with nonmetric-multidimensional scaling (NMDS; n=14, stress=0.09) used to ordinate the samples in 2-dimensional space. Samples were taken before (August 2013) and during (October 2013) the overflow of Inferno Crater Lake into Waimangu Stream, with time indicated by the direction of the arrows.
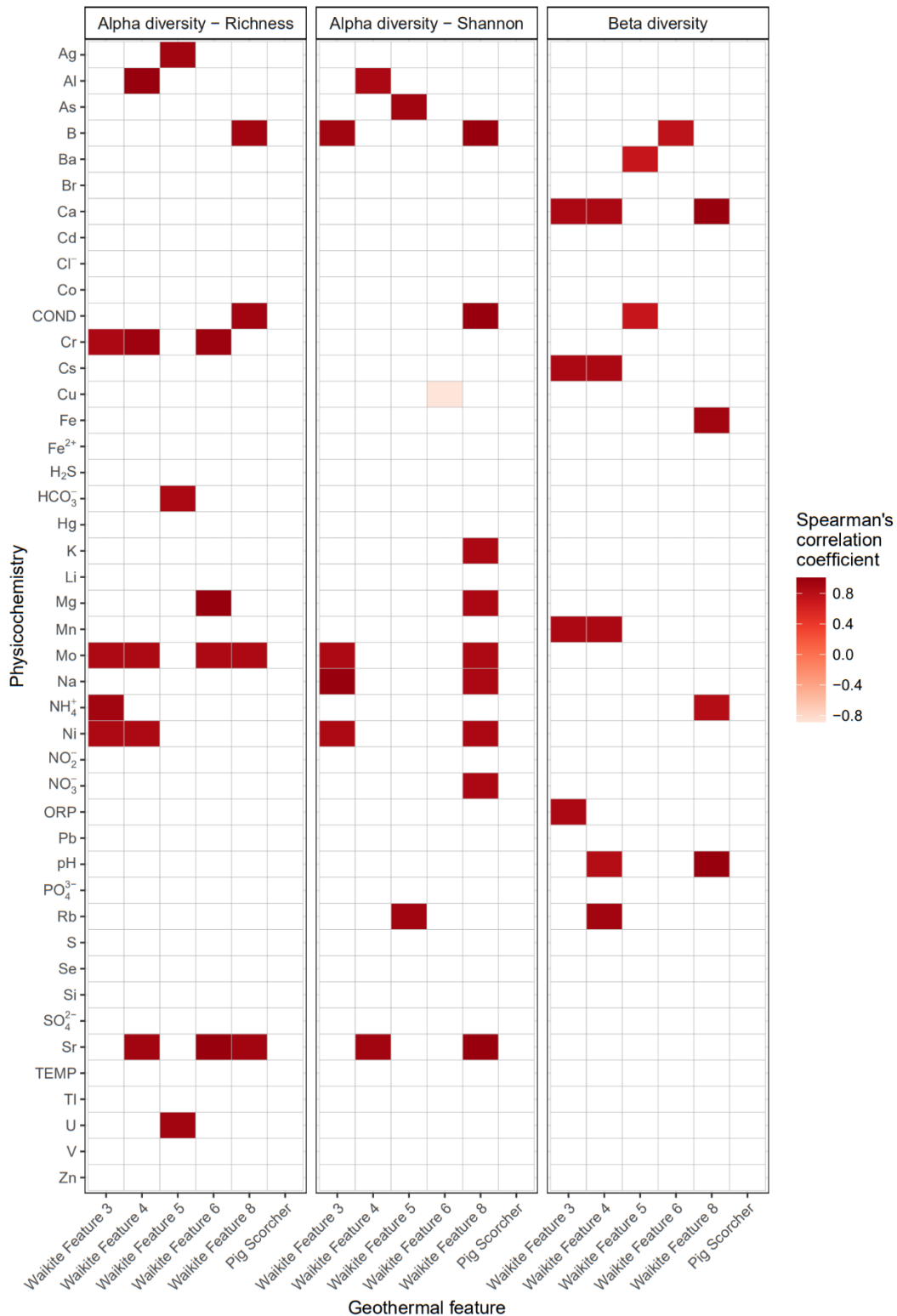
Figure C.5 - Relationship between diversity and physicochemistry of Waikite geothermal features. Diversity measures were calculated from amplicon sequencing of the 16S rRNA gene found in the microbial communities of these features over time. Alpha diversity was measured by both OTU richness (i.e., the number of OTUs in a community), and Shannon index (i.e., evenness of a community), with beta diversity determined by Bray-Curtis dissimilarities between temporal sites. Significant correlations ($p<0.05$) via Spearman's coefficient with all 44 physicochemical variables are displayed in colour.
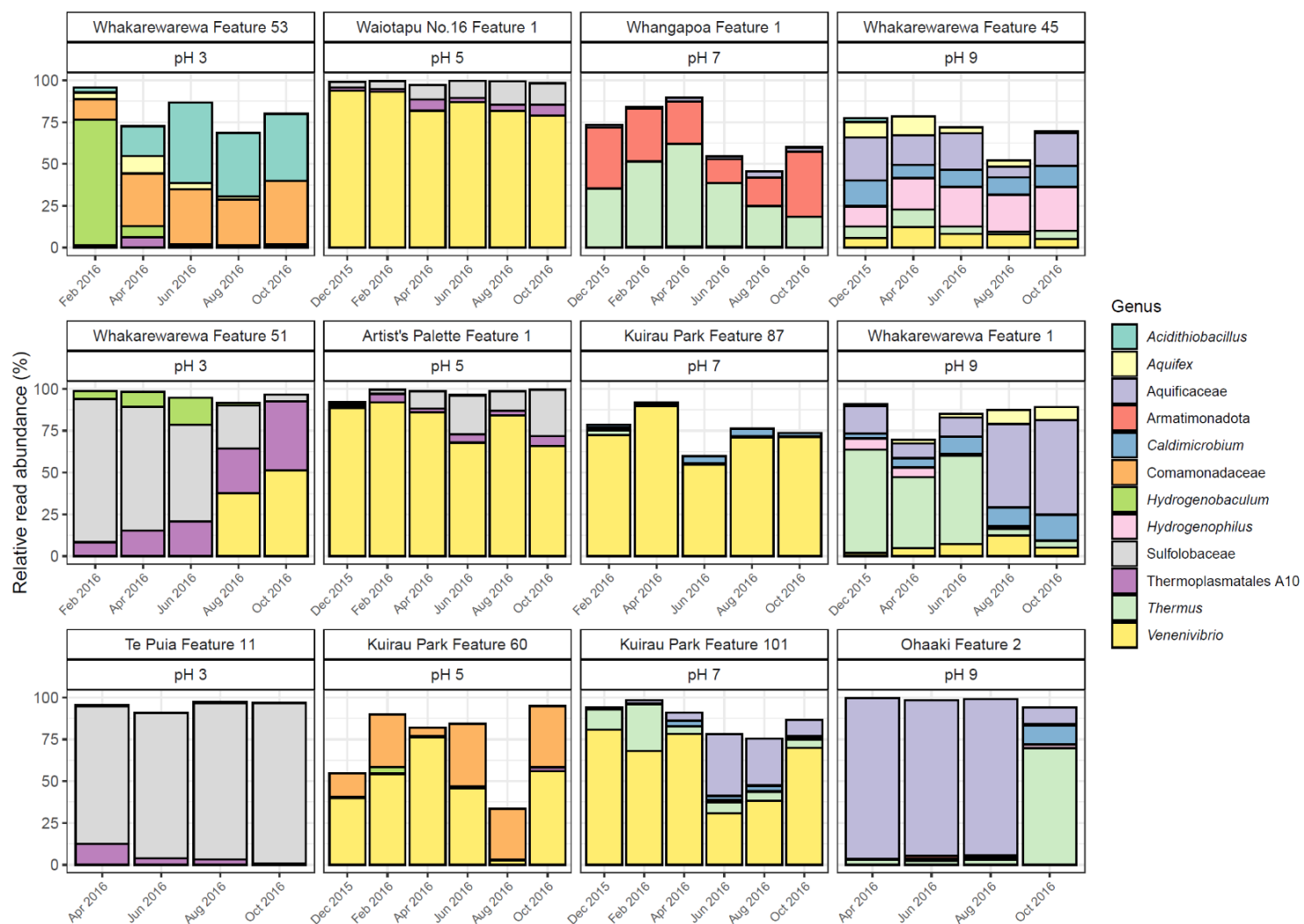
Figure C.6 - Temporal community composition and relative abundance of genera in geothermal features across the pH scale. Amplicon sequencing of the 16S rRNA gene was used to measure read abundance of taxa in spring communities, with only genera >1 % average relative abundance across all samples shown. Geothermal features are grouped according to their respective pH group (pH 3, 5, 7 and 9). Where taxonomy failed to classify to genus, the next corresponding taxonomic rank is displayed.
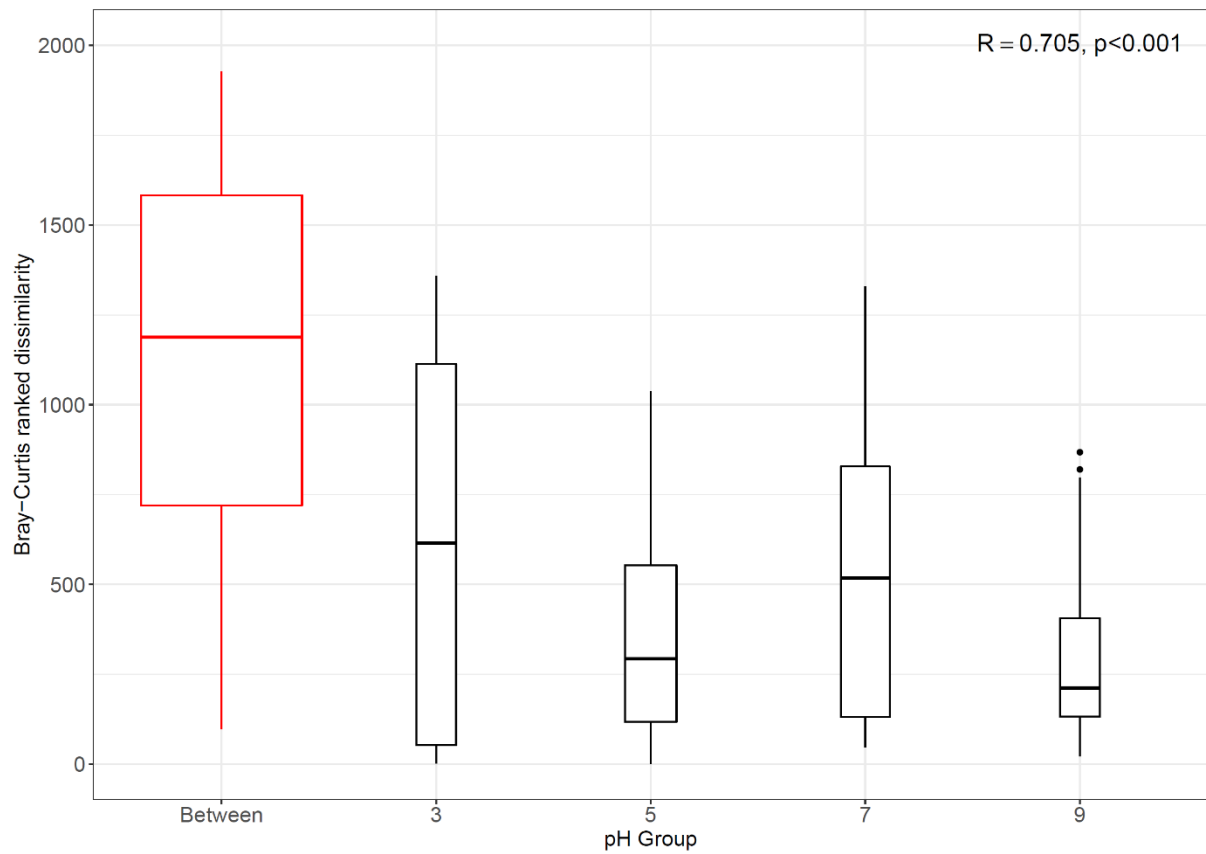
Figure C.7 - Beta diversity comparison of temporal samples grouped to pH. Bray-Curtis dissimilarity was used to calculate beta diversity from 16S rRNA gene sequencing of samples (*n*=63) from across the pH range found in Aotearoa-New Zealand geothermal features. Samples were binned according to pH group (3, 5, 7, and 9), with analysis of similarities (ANOSIM) demonstrating that microbial diversity between pH groups (red) was greater that the diversity within individual groups (R=0.705, *p*<0.001).
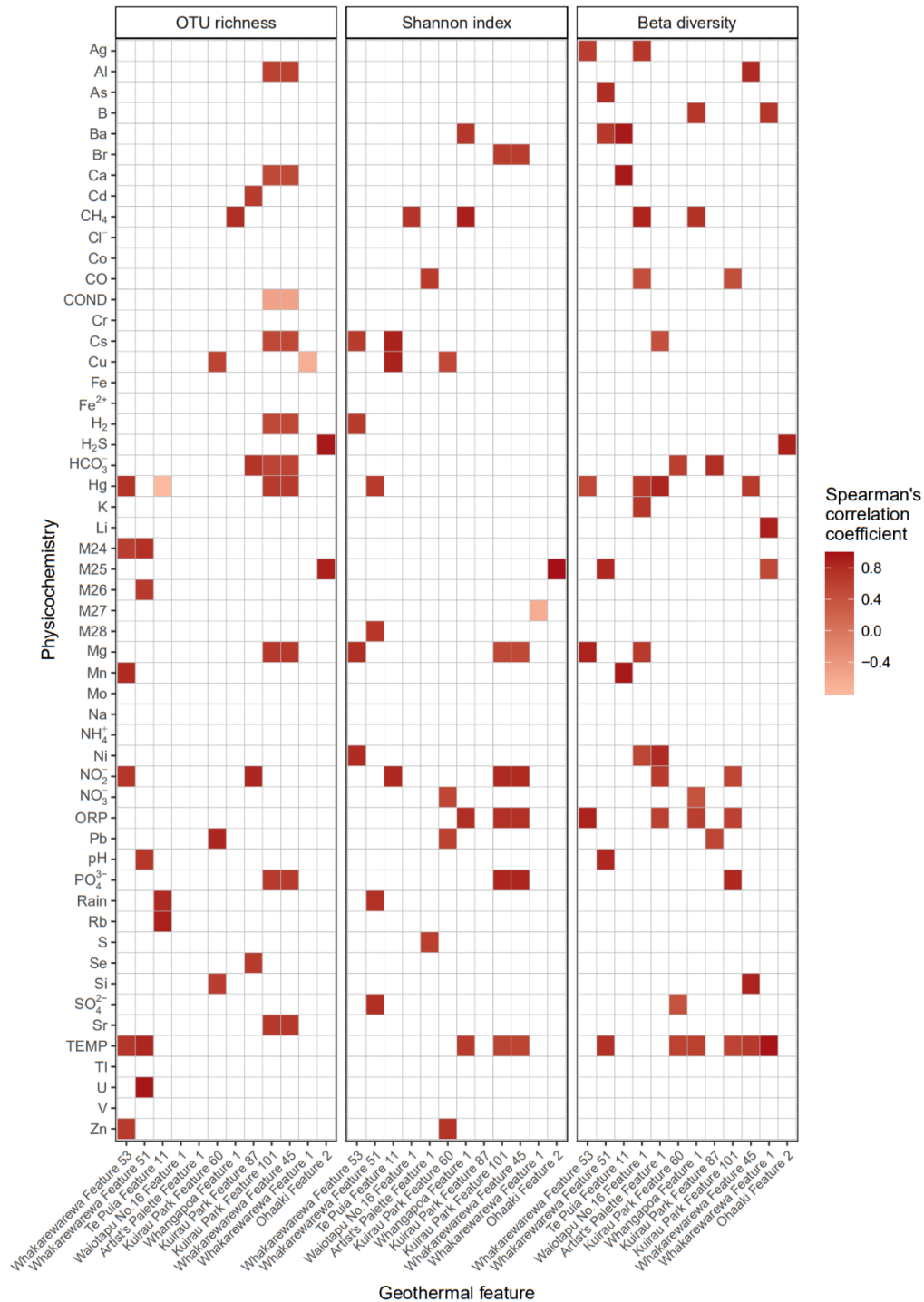
Figure C.8 - Relationship between diversity and physicochemistry of geothermal features across the pH range. Diversity was calculated from amplicon sequencing of the 16S rRNA gene found in the microbial communities of these features over time. Alpha diversity was measured by both OTU richness (i.e., the number of OTUs in a community), and Shannon index (i.e., evenness of a community), with beta diversity determined by Bray-Curtis dissimilarities between temporal sites. Significant correlations ($p<0.05$) with all 49 physicochemical variables are displayed in colour. Four features had positive correlations between alpha diversity and temperature ($\rho \geq 0.57$, $p \leq 0.03$, Spearman's coefficient), while the six features that had significant positive correlations between beta diversity and temperature ($\rho = 0.56$-0.98, $p < 0.05$) were Whakarewarewa Feature 51, Kuirau Park Feature 60, Whangapoa Feature 1, Kuirau Park Feature 101, Whakarewarewa Feature 45, and Whakarewarewa Feature 1.

C.2 SUPPLEMENTARY TABLES

Table C.1 - Control geothermal features and associated metadata. Alpha diversity is represented by OTU richness (i.e., number of OTUs) and evenness (i.e., Shannon index) measured in these ecosystems via 16S rRNA gene amplicon sequencing. Accessions for sample sequences are also included.

Sample ID	Study accession	Run accession	Feature name	Geothermal field	Date	pH	Temp (°C)	OTU Richness	Shannon
P1.0004	PRJEB24353	ERR2240132	Wairakei Terraces Feature 5	Wairakei-Tauhara	Jul 2013	8.2	93.0	70	1.2
P1.0140	PRJEB55115	ERR10020314	Wairakei Terraces Feature 5	Wairakei-Tauhara	Dec 2013	8.5	89.0	143	2.2
P1.0011	PRJEB55115	ERR10020306	Wairakei Terraces Feature 13	Wairakei-Tauhara	Aug 2013	8.3	62.9	192	3.1
P1.0148	PRJEB24353	ERR2240256	Wairakei Terraces Feature 13	Wairakei-Tauhara	Dec 2013	8.5	65.5	226	3.7
P1.0010	PRJEB55115	ERR10020305	Wairakei Terraces Feature 14	Wairakei-Tauhara	Aug 2013	8.3	96.7	230	1.9
P1.0149	PRJEB24353	ERR2240257	Wairakei Terraces Feature 14	Wairakei-Tauhara	Dec 2013	8.5	95.2	119	2.6
P1.0019	PRJEB24353	ERR2240145	Radiata Pool	Ngatamariki	Aug 2013	7.5	63.3	247	3.0
P1.0482	PRJEB55115	ERR10020323	Radiata Pool	Ngatamariki	Jun 2014	7.7	60.0	306	3.5
P1.0054	PRJEB24353	ERR2240180	Champagne Pool	Waiotapu	Sep 2013	5.5	72.4	61	1.8
P1.0236	PRJEB55115	ERR10020321	Champagne Pool	Waiotapu	Feb 2014	5.5	76.4	53	1.6
P1.0472	PRJEB55115	ERR10020322	Champagne Pool	Waiotapu	May 2014	5.3	72.9	83	2.0
P1.0675	PRJEB55115	ERR10020324	Champagne Pool	Waiotapu	Sep 2014	5.3	74.5	48	1.0
P1.0923	PRJEB55115	ERR10020325	Champagne Pool	Waiotapu	Dec 2014	5.2	74.0	64	1.5

Table C.2 - Disturbed features and associated metadata at Waimangu geothermal field. This subset of samples were taken from Waimangu Stream and Inferno Crater Lake, before (August 2013) and during (October 2013) an overflow of hydrothermal fluid from Inferno Crater Lake into the stream. Accessions for sample sequences are also included.

Sample ID	Study accession	Run accession	Feature name	Geothermal field	Date	pH	Temp (°C)	OTU Richness	Shannon
P1.0027	PRJEB24353	ERR2240153	Waimangu Stream Feature 2	Waimangu	Aug 2013	5.9	49.1	178	2.7
P1.0110	PRJEB55115	ERR10020313	Waimangu Stream Feature 2	Waimangu	Oct 2013	6.1	48.8	191	2.4
P1.0028	PRJEB24353	ERR2240154	Waimangu Stream Feature 3	Waimangu	Aug 2013	6.5	49.8	176	2.6
P1.0109	PRJEB55115	ERR10020312	Waimangu Stream Feature 3	Waimangu	Oct 2013	6.7	48.5	221	2.4
P1.0037	PRJEB24353	ERR2240163	Inferno Crater Lake	Waimangu	Aug 2013	2.4	41.0	51	1.8
P1.0108	PRJEB55115	ERR10020311	Inferno Crater Lake	Waimangu	Oct 2013	2.6	69.1	42	1.6
P1.0029	PRJEB24353	ERR2240155	Waimangu Stream Feature 4	Waimangu	Aug 2013	6.7	49.9	221	2.8
P1.0107	PRJEB55115	ERR10020310	Waimangu Stream Feature 4	Waimangu	Oct 2013	3.0	56.6	324	3.4
P1.0030	PRJEB24353	ERR2240156	Waimangu Stream Feature 5	Waimangu	Aug 2013	7.4	49.6	224	3.0
P1.0106	PRJEB55115	ERR10020309	Waimangu Stream Feature 5	Waimangu	Oct 2013	3.0	55.2	248	2.8
P1.0032	PRJEB24353	ERR2240158	Waimangu Stream Feature 6	Waimangu	Aug 2013	7.4	37.0	1049	3.9
P1.0105	PRJEB55115	ERR10020308	Waimangu Stream Feature 6	Waimangu	Oct 2013	6.5	39.5	1191	4.8
P1.0031	PRJEB24353	ERR2240157	Waimangu Stream Feature 7	Waimangu	Aug 2013	6.9	33.7	848	3.8
P1.0104	PRJEB55115	ERR10020307	Waimangu Stream Feature 7	Waimangu	Oct 2013	6.6	34.2	834	4.5

Table C.3 - Disturbed features and associated metadata at Waikite geothermal field. These features were analysed before and after a long-term, anthropogenic disturbance was introduced to a geothermal wetland undergoing rehabilitation. Accessions for sample sequences are included.

Sample ID	Study accession	Run accession	Feature name	Description	Geothermal field	Date	pH	Temp (°C)	OTU richness	Shannon
P1.0130	PRJEB24353	ERR2240240	Waikite Restoration Feature 3	Stream upgradient	Waikite	Dec 2013	8.6	44.3	766	4.5
P1.0171	PRJEB55115	ERR10020316	Waikite Restoration Feature 3	Stream upgradient	Waikite	Jan 2014	8.4	39.9	906	4.2
P1.1004	PRJEB55115	ERR10020329	Waikite Restoration Feature 3	Stream upgradient	Waikite	Apr 2015	8.5	37.8	1358	5.3
P1.1017	PRJEB55115	ERR10020337	Waikite Restoration Feature 3	Stream upgradient	Waikite	Oct 2016	8.3	39.5	805	4.1
P1.0131	PRJEB24353	ERR2240241	Waikite Restoration Feature 4	Wetland - inlet	Waikite	Dec 2013	8.5	47.2	1118	4.8
P1.0172	PRJEB55115	ERR10020317	Waikite Restoration Feature 4	Wetland - inlet	Waikite	Jan 2014	8.4	42.2	1308	5.0
P1.1005	PRJEB55115	ERR10020330	Waikite Restoration Feature 4	Wetland - inlet	Waikite	Apr 2015	8.3	39.0	1709	5.6
P1.1016	PRJEB55115	ERR10020336	Waikite Restoration Feature 4	Wetland - inlet	Waikite	Oct 2016	8.1	41.3	1030	4.2
P1.0132	PRJEB24353	ERR2240242	Waikite Restoration Feature 5	Wetland - hot	Waikite	Dec 2013	8.4	70.0	999	5.1
P1.0173	PRJEB55115	ERR10020318	Waikite Restoration Feature 5	Wetland - hot	Waikite	Jan 2014	8.4	67.9	826	5.1
P1.1007	PRJEB55115	ERR10020332	Waikite Restoration Feature 5	Wetland - hot	Waikite	Apr 2015	8.2	65.7	887	4.8
P1.1015	PRJEB55115	ERR10020335	Waikite Restoration Feature 5	Wetland - hot	Waikite	Oct 2016	8.1	65.5	667	4.4
P1.0133	PRJEB24353	ERR2240243	Waikite Restoration Feature 6	Wetland - warm	Waikite	Dec 2013	8.3	53.2	762	4.4
P1.0174	PRJEB55115	ERR10020319	Waikite Restoration Feature 6	Wetland - warm	Waikite	Jan 2014	8.3	43.8	1606	5.9
P1.1002	PRJEB55115	ERR10020327	Waikite Restoration Feature 6	Wetland - warm	Waikite	Apr 2015	7.7	23.3	2613	6.2
P1.1014	PRJEB55115	ERR10020334	Waikite Restoration Feature 6	Wetland - warm	Waikite	Oct 2016	7.5	29.9	1138	5.2
P1.0135	PRJEB24353	ERR2240245	Waikite Restoration Feature 8	Wetland - outlet	Waikite	Dec 2013	8.6	46.0	1375	5.5
P1.0176	PRJEB55115	ERR10020320	Waikite Restoration Feature 8	Wetland - outlet	Waikite	Jan 2014	8.6	42.8	1509	5.6
P1.1001	PRJEB55115	ERR10020326	Waikite Restoration Feature 8	Wetland - outlet	Waikite	Apr 2015	7.7	27.5	2877	7.0
P1.1013	PRJEB55115	ERR10020333	Waikite Restoration Feature 8	Wetland - outlet	Waikite	Oct 2016	7.6	26.0	1242	4.4
P1.0129	PRJEB24353	ERR2240239	Waikite Restoration Feature 2	Pig Scorcher	Waikite	Dec 2013	7.2	95.8	189	2.7
P1.0170	PRJEB55115	ERR10020315	Waikite Restoration Feature 2	Pig Scorcher	Waikite	Jan 2014	7.2	93.8	277	2.4
P1.1003	PRJEB55115	ERR10020328	Waikite Restoration Feature 2	Pig Scorcher	Waikite	Apr 2015	7.1	87.2	135	0.9
P1.0181	PRJEB24353	ERR2240282	Waikite Restoration Feature 13	Big Spring	Waikite	Jan 2014	7.5	96.5	284	3.9
P1.1006	PRJEB55115	ERR10020331	Waikite Restoration Feature 13	Big Spring	Waikite	Apr 2015	7.4	94.5	411	4.0

Table C.4 - pH features and associated metadata. Samples were taken from 12 geothermal features across the pH range commonly found in features from the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand (pH groups 3, 5, 7, and 9). All features had a similar temperature range of ~60-70 °C. Accessions for sample sequences are also included.

Sample ID	Study accession	Run accession	Feature name	pH group	Geothermal field	Date	pH	Temp (°C)	Richness	Shannon
P2.0019	PRJEB55115	ERR10020352	Whakarewarewa Feature 53	3	Rotorua	Feb 2016	3.0	60.1	39	1.1
P2.0024	PRJEB55115	ERR10020355	Whakarewarewa Feature 53	3	Rotorua	Apr 2016	3.1	59.1	41	2.2
P2.0036	PRJEB55115	ERR10020367	Whakarewarewa Feature 53	3	Rotorua	Jun 2016	3.1	57.2	37	1.6
P2.0048	PRJEB55115	ERR10020379	Whakarewarewa Feature 53	3	Rotorua	Aug 2016	3.0	50.8	30	1.8
P2.0060	PRJEB55115	ERR10020391	Whakarewarewa Feature 53	3	Rotorua	Oct 2016	2.9	53.0	28	1.6
P2.0017	PRJEB55115	ERR10020351	Whakarewarewa Feature 51	3	Rotorua	Feb 2016	3.3	70.3	24	1.3
P2.0026	PRJEB55115	ERR10020357	Whakarewarewa Feature 51	3	Rotorua	Apr 2016	3.6	67.1	25	1.6
P2.0038	PRJEB55115	ERR10020369	Whakarewarewa Feature 51	3	Rotorua	Jun 2016	4.0	66.3	27	1.7
P2.0050	PRJEB55115	ERR10020381	Whakarewarewa Feature 51	3	Rotorua	Aug 2016	4.6	60.0	45	2.1
P2.0062	PRJEB55115	ERR10020393	Whakarewarewa Feature 51	3	Rotorua	Oct 2016	4.8	60.6	39	1.5
P2.0023	PRJEB55115	ERR10020354	Te Puia Feature 11	3	Rotorua	Apr 2016	2.4	67.4	21	1.3
P2.0035	PRJEB55115	ERR10020366	Te Puia Feature 11	3	Rotorua	Jun 2016	2.5	66.1	23	1.2
P2.0047	PRJEB55115	ERR10020378	Te Puia Feature 11	3	Rotorua	Aug 2016	2.6	61.8	16	1.0
P2.0059	PRJEB55115	ERR10020390	Te Puia Feature 11	3	Rotorua	Oct 2016	2.5	64.0	19	0.9
P2.0002	PRJEB55115	ERR10020339	Waiotapu No.16 Feature 1	5	Waiotapu	Dec 2015	4.6	64.9	33	0.9
P2.0012	PRJEB55115	ERR10020347	Waiotapu No.16 Feature 1	5	Waiotapu	Feb 2016	4.6	67.5	36	0.9
P2.0032	PRJEB55115	ERR10020363	Waiotapu No.16 Feature 1	5	Waiotapu	Apr 2016	5.0	67.2	40	1.3
P2.0044	PRJEB55115	ERR10020375	Waiotapu No.16 Feature 1	5	Waiotapu	Jun 2016	4.9	67.0	34	1.1
P2.0056	PRJEB55115	ERR10020387	Waiotapu No.16 Feature 1	5	Waiotapu	Aug 2016	4.7	65.6	32	1.1
P2.0068	PRJEB55115	ERR10020399	Waiotapu No.16 Feature 1	5	Waiotapu	Oct 2016	4.7	68.0	35	1.2
P2.0003	PRJEB55115	ERR10020340	Artist's Palette Feature 1	5	Waiotapu	Dec 2015	5.1	59.4	90	1.4
P2.0011	PRJEB55115	ERR10020346	Artist's Palette Feature 1	5	Waiotapu	Feb 2016	4.8	67.9	48	1.0
P2.0033	PRJEB55115	ERR10020364	Artist's Palette Feature 1	5	Waiotapu	Apr 2016	4.9	66.5	63	1.3
P2.0045	PRJEB55115	ERR10020376	Artist's Palette Feature 1	5	Waiotapu	Jun 2016	4.8	64.7	87	1.7
P2.0057	PRJEB55115	ERR10020388	Artist's Palette Feature 1	5	Waiotapu	Aug 2016	4.7	66.9	74	1.4

Sample ID	Study accession	Run accession	Feature name	pH group	Geothermal field	Date	pH	Temp (°C)	Richness	Shannon
P2.0069	PRJEB55115	ERR10020400	Artist's Palette Feature 1	5	Waiotapu	Oct 2016	4.8	71.9	57	1.6
P2.0006	PRJEB55115	ERR10020342	Kuirau Park Feature 60	5	Rotorua	Dec 2015	5.2	59.2	70	1.9
P2.0015	PRJEB55115	ERR10020350	Kuirau Park Feature 60	5	Rotorua	Feb 2016	5.2	61.0	73	1.8
P2.0031	PRJEB55115	ERR10020362	Kuirau Park Feature 60	5	Rotorua	Apr 2016	5.3	56.6	90	2.0
P2.0043	PRJEB55115	ERR10020374	Kuirau Park Feature 60	5	Rotorua	Jun 2016	5.0	51.7	138	2.2
P2.0055	PRJEB55115	ERR10020386	Kuirau Park Feature 60	5	Rotorua	Aug 2016	4.9	45.8	119	1.8
P2.0067	PRJEB55115	ERR10020398	Kuirau Park Feature 60	5	Rotorua	Oct 2016	4.5	51.3	88	1.8
P2.0001	PRJEB55115	ERR10020338	Whangapoa Feature 1	7	Atiamuri	Dec 2015	7.6	63.3	123	2.5
P2.0010	PRJEB55115	ERR10020345	Whangapoa Feature 1	7	Atiamuri	Feb 2016	7.3	65.3	114	2.2
P2.0028	PRJEB55115	ERR10020359	Whangapoa Feature 1	7	Atiamuri	Apr 2016	7.1	63.8	90	2.0
P2.0040	PRJEB55115	ERR10020371	Whangapoa Feature 1	7	Atiamuri	Jun 2016	7.1	61.0	129	2.8
P2.0052	PRJEB55115	ERR10020383	Whangapoa Feature 1	7	Atiamuri	Aug 2016	7.1	60.8	147	3.0
P2.0064	PRJEB55115	ERR10020395	Whangapoa Feature 1	7	Atiamuri	Oct 2016	7.6	62.7	122	2.7
P2.0013	PRJEB55115	ERR10020348	Kuirau Park Feature 87	7	Rotorua	Feb 2016	6.6	63.8	196	2.3
P2.0029	PRJEB55115	ERR10020360	Kuirau Park Feature 87	7	Rotorua	Apr 2016	6.9	61.4	171	1.4
P2.0041	PRJEB55115	ERR10020372	Kuirau Park Feature 87	7	Rotorua	Jun 2016	7.1	57.5	380	3.3
P2.0053	PRJEB55115	ERR10020384	Kuirau Park Feature 87	7	Rotorua	Aug 2016	6.6	62.0	306	2.4
P2.0065	PRJEB55115	ERR10020396	Kuirau Park Feature 87	7	Rotorua	Oct 2016	6.8	58.5	329	2.5
P2.0005	PRJEB55115	ERR10020341	Kuirau Park Feature 101	7	Rotorua	Dec 2015	7.1	70.6	101	1.4
P2.0014	PRJEB55115	ERR10020349	Kuirau Park Feature 101	7	Rotorua	Feb 2016	7.3	72.4	64	1.5
P2.0030	PRJEB55115	ERR10020361	Kuirau Park Feature 101	7	Rotorua	Apr 2016	7.1	71.4	138	1.8
P2.0042	PRJEB55115	ERR10020373	Kuirau Park Feature 101	7	Rotorua	Jun 2016	7.3	68.7	141	2.8
P2.0054	PRJEB55115	ERR10020385	Kuirau Park Feature 101	7	Rotorua	Aug 2016	7.0	67.9	151	2.8
P2.0066	PRJEB55115	ERR10020397	Kuirau Park Feature 101	7	Rotorua	Oct 2016	7.2	68.3	133	2.2
P2.0007	PRJEB55115	ERR10020343	Whakarewarewa Feature 45	9	Rotorua	Dec 2015	8.9	69.5	219	3.2
P2.0025	PRJEB55115	ERR10020356	Whakarewarewa Feature 45	9	Rotorua	Apr 2016	9.0	66.2	175	3.1
P2.0037	PRJEB55115	ERR10020368	Whakarewarewa Feature 45	9	Rotorua	Jun 2016	8.7	64.7	201	3.1
P2.0049	PRJEB55115	ERR10020380	Whakarewarewa Feature 45	9	Rotorua	Aug 2016	8.7	64.5	232	3.4
P2.0061	PRJEB55115	ERR10020392	Whakarewarewa Feature 45	9	Rotorua	Oct 2016	8.8	65.5	254	3.2

Sample ID	Study accession	Run accession	Feature name	pH group	Geothermal field	Date	pH	Temp (°C)	Richness	Shannon
P2.0009	PRJEB55115	ERR10020344	Whakarewarewa Feature 1	9	Rotorua	Dec 2015	8.3	63.4	101	1.9
P2.0027	PRJEB55115	ERR10020358	Whakarewarewa Feature 1	9	Rotorua	Apr 2016	8.4	61.1	168	2.9
P2.0039	PRJEB55115	ERR10020370	Whakarewarewa Feature 1	9	Rotorua	Jun 2016	8.0	70.3	108	2.5
P2.0051	PRJEB55115	ERR10020382	Whakarewarewa Feature 1	9	Rotorua	Aug 2016	7.8	78.1	105	2.5
P2.0063	PRJEB55115	ERR10020394	Whakarewarewa Feature 1	9	Rotorua	Oct 2016	8.4	80.6	102	2.3
P2.0022	PRJEB55115	ERR10020353	Ohaaki Feature 2	9	Ohaaki	Apr 2016	9.0	70.3	31	0.9
P2.0034	PRJEB55115	ERR10020365	Ohaaki Feature 2	9	Ohaaki	Jun 2016	8.9	66.5	48	1.1
P2.0046	PRJEB55115	ERR10020377	Ohaaki Feature 2	9	Ohaaki	Aug 2016	8.4	65.0	43	1.1
P2.0058	PRJEB55115	ERR10020389	Ohaaki Feature 2	9	Ohaaki	Oct 2016	9.0	67.9	77	1.7

Table C.5 - Variation in physicochemistry, alpha diversity and taxon relative abundance of control geothermal features. Variation is presented as standard deviation (SD) from the mean of values measured for pH, temperature (°C), OTU richness, Shannon diversity index, and dominant taxa (>1 % average relative abundance across all samples in this category). Alpha diversity and relative abundances of taxa were calculated via amplicon sequencing of the 16S rRNA gene from DNA extracted from these ecosystems.

Geothermal feature	Wairakei Feature 5		Wairakei Feature 13		Wairakei Feature 14		Radiata Pool		Champagne Pool	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
pH	8.3	0.2	8.4	0.2	8.4	0.2	7.6	0.2	5.3	0.1
Temperature (°C)	91.0	2.8	64.2	1.8	96.0	1.1	61.7	2.3	74.0	1.6
OTU richness	106.5	51.6	209.0	24.0	174.5	78.5	276.5	41.7	61.8	13.4
Shannon index	1.7	0.7	3.4	0.4	2.2	0.5	3.3	0.3	1.6	0.4
Aquificota (%)	96.1	2.8	10.3	3.9	71.8	3.8	60.0	8.0	95.9	2.5
Armatimonadota (%)	0.0	0.0	7.5	7.5	0.2	0.1	0.3	0.0	0.0	0.0
Bacteroidota (%)	0.7	1.0	18.5	4.3	13.1	16.7	1.4	0.8	0.0	0.0
Chlorobiota (%)	0.0	0.0	8.2	0.4	0.0	0.0	5.9	2.1	0.0	0.0
Chloroflexota (%)	0.0	0.1	16.3	5.1	0.2	0.2	0.8	0.1	0.0	0.0
Deinococcota (%)	0.5	0.5	20.9	5.6	7.7	9.9	1.4	0.2	0.0	0.0
Pseudomonadota (%)	1.3	1.2	7.8	2.4	4.4	2.0	5.0	0.5	0.2	0.1
Thermodesulfobacteriota (%)	0.0	0.0	0.0	0.0	0.1	0.1	5.2	0.8	1.9	1.6

Table C.6 - Variation in physicochemistry, alpha diversity and taxon relative abundance of Waimangu geothermal features. Variation is presented as standard deviation (SD) from the mean of values measured for pH, temperature (°C), OTU richness, Shannon diversity index, and dominant taxa (>1 % average relative abundance across all samples in this category). Alpha diversity and relative abundances of taxa were calculated via amplicon sequencing of the 16S rRNA gene from DNA extracted from these ecosystems.

Geothermal feature	Waimangu Feature 2		Waimangu Feature 3		Inferno Crater Lake		Waimangu Feature 4		Waimangu Feature 5		Waimangu Feature 6		Waimangu Feature 7	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
pH	6.0	0.1	6.6	0.1	2.5	0.1	4.9	2.6	5.2	3.1	6.9	0.6	6.8	0.2
Temperature (°C)	49.0	0.2	49.2	0.9	55.1	19.9	53.3	4.7	52.4	4.0	38.3	1.8	34.0	0.4
OTU richness	184.5	9.2	198.5	31.8	46.5	6.4	272.5	72.8	236.0	17.0	1120.0	100.4	841.0	9.9
Shannon index	2.5	0.2	2.5	0.2	1.7	0.1	3.1	0.4	2.9	0.2	4.4	0.7	4.1	0.5
Aquificota (%)	23.2	3.1	24.9	5.5	4.0	5.3	27.0	2.4	27.8	7.5	16.8	12.3	11.8	13.9
Bacteroidota (%)	1.9	0.1	2.2	0.4	0.1	0.1	5.6	2.5	2.9	0.2	9.9	7.2	18.6	14.2
Chlorobiota (%)	1.4	0.1	2.0	1.6	0.0	0.0	5.5	5.6	3.2	2.8	1.9	0.8	1.2	1.3
Cyanobacteria (%)	7.4	9.1	10.2	11.0	0.0	0.0	14.2	14.1	9.2	11.7	4.4	3.6	3.7	4.4
Euryarchaeota (%)	0.0	0.0	0.0	0.0	8.6	2.7	0.4	0.5	0.0	0.0	0.0	0.0	0.0	0.0
Parcubacteria (%)	0.7	0.6	0.8	0.3	0.0	0.0	1.3	1.5	1.0	0.9	2.4	1.4	0.8	0.3
Pseudomonadota (%)	63.2	11.1	57.8	13.6	36.9	51.9	35.7	4.6	52.4	13.3	52.6	7.0	56.1	3.4
Thermoproteota (%)	0.0	0.0	0.0	0.0	46.1	55.9	6.3	8.8	0.6	0.8	0.1	0.2	0.0	0.1

Table C.7 - Variation in physicochemistry, alpha diversity and taxon relative abundance of Waikite geothermal features. Variation is presented as standard deviation (SD) from the mean of values measured for pH, temperature (°C), OTU richness, Shannon diversity index, and dominant taxa (>1 % average relative abundance across all samples in this category). Alpha diversity and relative abundances of taxa were calculated via amplicon sequencing of the 16S rRNA gene from DNA extracted from these ecosystems.

Geothermal feature	Waikite Feature 3		Waikite Feature 4		Waikite Feature 5		Waikite Feature 6		Waikite Feature 8		Pig Scorcher		Big Spring	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
pH	8.4	0.1	8.3	0.2	8.3	0.2	7.9	0.4	8.1	0.6	7.2	0.1	7.5	0.0
Temperature (°C)	40.4	2.8	42.4	3.5	67.3	2.1	37.6	13.5	35.6	10.3	92.3	4.5	95.5	1.4
OTU richness	958.8	272.6	1291.3	301.7	844.8	138.5	1529.8	800.4	1750.8	758.7	200.3	71.7	347.5	89.8
Shannon index	4.5	0.6	4.9	0.6	4.8	0.3	5.5	0.8	5.6	1.1	2.0	1.0	3.9	0.0
Aquificota (%)	0.1	0.0	0.0	0.0	1.4	0.5	0.0	0.0	0.0	0.0	76.5	19.3	45.6	1.8
Bacillota (%)	0.7	0.7	0.8	0.7	4.2	3.9	0.8	0.8	1.1	0.8	0.4	0.3	5.8	6.4
Bacteria (%)	0.7	0.5	0.9	0.3	1.7	0.3	3.2	2.3	1.5	1.3	2.0	3.2	0.6	0.0
Bacteroidota (%)	10.4	2.7	10.2	2.7	22.0	5.8	17.0	15.9	14.7	7.3	0.7	0.5	4.9	2.8
Chlamydiota (%)	0.4	0.3	0.6	0.4	1.5	2.8	1.9	3.0	2.5	3.8	0.0	0.0	0.3	0.4
Chlorobiota (%)	0.7	0.4	0.8	0.5	5.2	6.7	2.4	2.1	1.0	0.9	0.2	0.0	0.5	0.4
Chloroflexota (%)	0.5	0.3	0.5	0.3	5.9	5.8	1.4	0.9	0.4	0.3	0.3	0.1	1.1	0.6
Cyanobacteria (%)	10.0	7.6	5.3	4.5	2.6	4.5	6.9	7.9	2.5	1.7	0.9	1.1	0.6	0.1
Deinococcota (%)	0.1	0.1	0.2	0.1	3.4	3.3	0.1	0.1	0.1	0.1	12.0	16.1	8.1	3.5
Parcubacteria (%)	1.4	1.0	2.3	1.8	6.0	1.9	9.2	9.9	8.0	9.4	0.2	0.1	1.5	1.0
Planctomycetota (%)	1.3	0.7	1.6	0.4	0.5	0.3	1.2	1.0	1.6	0.8	0.0	0.0	0.7	0.2
Pseudomonadota (%)	65.9	10.5	68.0	10.9	27.6	7.0	39.6	20.4	52.8	12.4	3.2	2.4	22.5	9.8
Verrucomicrobiota (%)	2.3	1.4	2.9	1.2	1.6	0.9	4.8	4.7	2.8	0.7	0.0	0.0	1.1	0.5
Woesearchaeota (%)	0.1	0.1	0.5	0.4	1.8	1.4	2.3	2.2	1.9	1.6	0.0	0.0	0.1	0.2

Table C.8 - Variation in physicochemistry, alpha diversity and taxon relative abundance of geothermal features across the pH range. Variation is presented as standard deviation (SD) from the mean of values measured for pH, temperature (°C), OTU richness, Shannon diversity index, and dominant taxa (>1 % average relative abundance across all samples in this category). Alpha diversity and relative abundances of taxa were calculated via amplicon sequencing of the 16S rRNA gene from DNA extracted from these ecosystems.

pH group	3		3		3		5		5		5	
Geothermal feature	Whakarewarewa Feature 53		Whakarewarewa Feature 51		Te Puia Feature 11		Waiotapu No.16 Feature 1		Artist's Palette Feature 1		Kuirau Park Feature 60	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
pH	3.0	0.1	4.1	0.6	2.5	0.1	4.7	0.2	4.8	0.1	5.0	0.3
Temperature (°C)	56.0	4.0	64.9	4.4	64.8	2.5	66.7	1.2	66.2	4.1	54.3	5.7
OTU richness	35.0	5.7	32.0	9.4	19.8	3.0	35.0	2.8	69.8	16.8	96.3	26.8
Shannon index	1.7	0.4	1.6	0.3	1.1	0.1	1.1	0.2	1.4	0.2	1.9	0.1
Aquificota (%)	21.1	33.1	24.1	20.2	0.6	0.5	86.1	6.3	80.8	11.1	46.6	24.7
Armatimonadota (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Bacteroidota (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1
Deinococcota (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Euryarchaeota (%)	11.0	7.5	22.5	12.6	5.1	5.0	3.8	2.3	3.7	1.8	1.1	0.8
Pseudomonadota (%)	59.5	27.9	0.4	0.4	0.0	0.0	0.0	0.1	2.0	2.7	49.0	24.4
Thermodesulfobacteriota (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.2
Thermoproteota (%)	0.1	0.2	49.7	33.6	89.7	6.2	10.0	4.4	13.0	10.9	0.3	0.1
Thermotogota (%)	8.0	4.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.3	0.3

pH group	7		7		7		9		9		9	
Geothermal feature	Whangapoa Feature 1		Kuirau Park Feature 87		Kuirau Park Feature 101		Whakarewarewa Feature 45		Whakarewarewa Feature 1		Ohaaki Feature 2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
pH	7.3	0.2	6.8	0.2	7.2	0.1	8.8	0.1	8.2	0.3	8.8	0.3
Temperature (°C)	62.8	1.7	60.6	2.6	69.9	1.8	66.1	2.0	70.7	8.6	67.4	2.3
OTU richness	120.8	18.7	276.4	89.4	121.3	32.8	216.2	30.1	116.8	28.8	49.8	19.5
Shannon index	2.5	0.4	2.4	0.7	2.1	0.6	3.2	0.1	2.4	0.4	1.2	0.4
Aquificota (%)	3.8	0.7	71.7	12.4	74.8	7.3	35.6	5.2	39.3	28.2	73.2	42.3
Armatimonadota (%)	28.0	10.4	0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.0	1.0	0.6
Bacteroidota (%)	9.5	6.4	2.2	1.8	0.4	0.2	0.9	0.3	2.0	2.9	0.3	0.2
Deinococcota (%)	38.0	16.1	1.0	1.3	10.4	9.1	5.6	3.3	33.1	27.3	19.5	33.5
Euryarchaeota (%)	0.0	0.0	0.5	0.5	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0
Pseudomonadota (%)	6.0	4.0	4.0	1.9	1.0	0.8	29.3	7.6	3.9	3.4	1.4	0.9
Thermodesulfobacteriota (%)	0.1	0.0	2.6	1.6	1.8	1.2	11.4	2.9	9.2	5.0	3.0	5.5
Thermoproteota (%)	0.0	0.0	0.4	0.3	7.7	5.4	0.1	0.0	0.3	0.1	0.0	0.0
Thermotogota (%)	0.4	0.1	0.4	0.2	0.5	0.3	5.8	1.8	2.8	1.3	0.3	0.4

APPENDIX D

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

D.1 SUPPLEMENTARY METHODS

D.1.1 Aquificota & *Venenivibrio* ecology in Aotearoa-New Zealand

The 46 physicochemical parameters measured from each of the 925 geothermal springs were aluminium (Al), ammonium (NH_4^+), arsenic (As), barium (Ba), bicarbonate (HCO_3^-), boron (B), bromine (Br), cadmium (Cd), caesium (Cs), calcium (Ca), chloride (Cl^-), chromium (Cr), cobalt (Co), conductivity (COND), copper (Cu), dissolved oxygen (dO), ferrous iron (Fe^{2+}), hydrogen sulfide (H_2S), iron (Fe), lead (Pb), lithium (Li), magnesium (Mg), manganese (Mn), mercury (Hg), molybdate (Mo), nickel (Ni), nitrate (NO_3^-), nitrite (NO_2^-), oxidation-reduction potential (ORP), pH, phosphate (PO_4^{3-}), potassium (K), rubidium (Rb), selenium (Se), silicon (Si), silver (Ag), sodium (Na), strontium (Sr), sulfate (SO_4^{2-}), sulfur (S), temperature (TEMP), thallium (Tl), turbidity (TURB), uranium (U), vanadium (V) and zinc (Zn). Conductivity, ORP, dO, TURB and pH were determined using a Hanna Instruments multiparameter field meter (Woonsocket, RI, USA), with spring temperature measured by a Fluke 51-II thermocouple (Everett, WA, USA). Inductively coupled plasma-mass spectrometry (ICP-MS), UV-vis spectrometry, titration and ion chromatography were used to measure aqueous metals and non-metals. More details on this methodology can be found in Power *et al.* (2018)[195].

D.1.2 Genome annotation of *V. stagnispumantis* CP.B2[T]

The following nine publicly-available Hydrogenothermaceae genomes were used to compare metabolisms to *V. stagnispumantis* CP.B2[T], using IMG/MER: *Hydrogenothermus marinus* VM1[T] (IMG Taxon ID: 2734482257), *Persephonella hydrogeniphila* 29W[T] (2728369220), *P. marina* EX-H1[T] (643692030), *Persephonella* sp. IF05_L8 (2562617012), *Persephonella* sp. KM09_Lau8 (2561511220), *Sulfurihydrogenibium azorense* Az-Fu1[T] (643692050), *S. subterraneum* HGMK1[T] (2558860995), *Sulfurihydrogenibium* sp. Y03AOP1 (642555165), and *S. yellowstonense* SS-5[T] (645058708).

D.1.3   Growth range reassessment of *V. stagnispumantis* CP.B2[T]

To confirm growth conditions of type strain *V. stagnispumantis* CP.B2[T], medium was prepared in 9 mL aliquots (in triplicate) as per Hetzer *et al.* (2008)[179] (DSMZ medium 1146) with the addition of elemental sulfur (S^0), inoculated with 1 mL of strain CP.B2[T], and incubated at 70 °C statically for seven days, unless otherwise stated. Both positive and negative controls for each experiment were also prepared accordingly. Phase contrast microscopy was used to determine growth due to the tendency for CP.B2[T] to flocculate. The observed upper and lower growth limits were confirmed by subculturing the culture into the described optimal growth medium and conditions (DSMZ medium 1146; with S^0 and 2.5 % v/v O_2, at 70 °C and pH 5.5) to confirm viability. The growth temperature range for CP.B2[T] was tested between 38.5-79.7 °C using a custom-made temperature gradient oscillator (10 oscillations/min). To test a pH range of 3.0-8.5, DSMZ medium 1146 was prepared without the addition of MES except for where pH 5.5 – 6.5 was required. To buffer the medium outside this range, either 10 mM sodium citrate/citric acid (pH 3.0-5.0), or 2.38 g/L HEPES (pH 7.0-8.5) buffer were added. The medium pH was adjusted using either 1 M NaOH or 1 M HCl as required. Each enrichment was tested after incubation with both universal and range specific pH paper to ensure consistency of pH throughout the experiment. Salinity growth ranges were tested by the addition of 0.0-10.0 % (w/v) NaCl to growth medium. To test the O_2 tolerance of CP.B2[T], a headspace of 50:40:10 N_2:H_2:CO_2 was prepared to growth medium, with oxygen then added volumetrically on top of this gas mixture to final concentrations of 1.25, 2.5, 5.0, 7.5, 10, 12.5, 15, and 25 % (v/v).

In order to ascertain if *Venenivibrio* could grow in hot spring conditions outside Aotearoa-New Zealand, water column samples collected from Champagne Pool (CP), Waiotapu, Aotearoa-New Zealand and Obsidian Pool (OP), Yellowstone National Park, USA, were used as the basal growth medium for subsequent inoculations with CP.B2[T]. Both samples were filtered immediately in the field and then in the laboratory using a Sterivex-GP 0.22 µm PES column filter to ensure sterility, and stored at 4 °C until use. The two spring samples were prepared in triplicate as follows: original spring water with no additions; spring water plus NH_4Cl, KH_2PO_4, and S^0 (as per the concentrations specified in DSMZ medium 1146); and spring water with S^0 as the only supplement. In all cases, no trace metals were added to the hybrid spring media. Medium pH did not vary with the addition of amendments. Headspace composition for all experiments was 50:40:7.5:2.5 N_2:H_2:CO_2:O_2. An exponential phase

CP.B2^T was then inoculated into all test samples and incubated at 70 °C as previously described.

D.1.4 Global search for 16S rRNA genes reported as, or closely related to, *Venenivibrio*

Global databases (Nucleotide Collection, SRA, SILVA, RDP, Greengenes, IMNGS, IMG/M, EMP and Qiita) were searched for 16S rRNA gene sequences either reported as, or returning ≥95.0 % sequence similarity[381] to *Venenivibrio stagnispumantis* CP.B2^T. Criteria for each individual search are reported below, with the phylogenetic analyses and tree construction detailed at the end of the section.

D.1.4.1 Nucleotide Collection (NCBI)

A pairwise analysis of the full length 16S rRNA gene of *V. stagnispumantis* CP.BP^T (1506 bp; GenBank accession DQ989208.1 or RefSeq accession NR_044029) was undertaken by querying the sequence against the Nucleotide Collection (nr/nt) in NCBI using BLASTN v2.13.0 with the megablast setting (accessed 23/Jul/2021). Returned hits with ≥95.0 % sequence similarity and an e-value of zero were investigated.

D.1.4.2 Sequence Read Archive (NCBI)

All samples (both amplicon and metagenomic) from the Sequence Read Archive (SRA) were screened for reads that assigned to the genus *Venenivibrio* (tax_id=407997) using the Sequence Taxonomic Analysis Tool (STAT)[382] through the Google Cloud BigQuery platform (accessed 06/Dec/2021). As the genome of *V. stagnispumantis* CP.B2^T is not yet in the RefSeq database, only the 16S rRNA gene sequence (RefSeq accession NR_044029) was used by the STAT program to assign taxonomy to k-mers. BLASTN v2.13.0 (megablast setting) was used to confirm any putative similarity of samples to *V. stagnispuma*ntis CP.B2^T as required.

D.1.4.3 SILVA database

The Browser function in the 16S rRNA gene database SILVA was used to search the SSU r138.1 component (accessed 25/Mar/2022) for entries classified as the genus *Venenivibrio*. The SILVA Incremental Aligner (SINA v1.2.11) was then run with default settings to align any entries found. The search and classify function of SINA was enabled (with default

settings) and closest neighbours to *V. stagnispumantis* CP.B2^T (≥95 % sequence similarity) in the Ref NR database were identified.

D.1.4.4 Ribosomal Database Project (RDP)

RDP (release 11_6; accessed 25/May/2021) was screened by using Seqmatch to query the 16S rRNA gene sequence of *V. stagnispumantis* CP.B2^T. Search settings included type and non-type strains, uncultured and isolate sequences, all sizes (near full length and partial), good quality and KNN matches of 20.

D.1.4.5 Greengenes database

The 16S rRNA gene database Greengenes (v13_8; August 2013) was downloaded through the FTP site (ftp://greengenes.microbio.me/greengenes_release; accessed 02/Jul/2021) and both 97 % (97_otu_taxonomy.txt) and 99 % (99_otu_taxonomy.txt) OTU representative sets were screened for the presence of Hydrogenothermaceae and *Venenivibrio*. Corresponding sequences were extracted from the accompanying FASTA files, and sequence identity to the 16S rRNA gene of *V. stagnispumantis* CP.B2^T was checked using BLASTN v2.13.0 (megablast setting). Any additional sequences that mapped to these OTUs were also found in 97_otu_map.txt and 99_otu_map.txt. Finally, Greengenes OTU IDs were cross checked against accession numbers from GenBank to confirm identity (gg_13_5_accessions.txt; https://greengenes.secondgenome.com/; accessed 02/Jul/2021).

D.1.4.6 Integrated Microbial NGS (IMNGS) platform

The full length 16S rRNA gene of *V. stagnispumantis* CP.B2^T was queried using the Integrated Microbial Next Generation Sequencing (IMNGS) platform v1.0 build 2105 (accessed 21/May/2021). This database takes all available raw 16S rRNA gene amplicon samples from the International Nucleotide Sequence Database Collaboration (SRA, DDBJ and ENA), and runs them through a standardised pipeline to produce quality controlled OTUs and corresponding read abundance data. The gene was queried against the full database, using the discrete similarity function at a threshold of 95 % sequence similarity (minimum sample size 200 bp). The database was additionally screened for all OTUs assigned to the phylum Aquificae/Aquificota (accessed 25/May/2021). The IMNGS database also included 472 samples from the Tara Oceans project.

D.1.4.7 Integrated Microbial Genomes and Microbiomes (IMG/M) database

All genomes in IMG/M v.6.0 (accessed 21/Jul/2021) were searched for Hydrogenothermaceae using the Genomes by Taxonomy function. The 16S rRNA gene of *V. stagnispumantis* CP.B2^T was also searched against all genomes in the database using the BLAST option. Metagenomic samples were screened for Hydrogenothermaceae-assigned bins using the Bins by Taxonomy function under Find Genomes/Metagenome Bins.

D.1.4.8 Earth Microbiome Project (EMP)

The first release of the Earth Microbiome Project (EMP) was searched in multiple ways for the presence of *Venenivibrio*. Firstly, all samples classified using SILVA taxonomy (SSU r123) were downloaded through the EMP FTP site (ftp://ftp.microbio.me/emp/release1; accessed 22/Jun/2021) in the dataset emp_cr_silva_16S_123.release1.biom. This was converted to text and the grep command was used to find OTUs assigned to *Venenivibrio*. The reference OTUs used to assign this taxonomy (silva_123.97_otus_16S.consensus_taxonomy_all_levels.txt) were also checked to confirm presence of *Venenivibrio* in this taxonomy. Secondly, the software redbiom (v0.3.5; accessed 22/Jun/2021) was used to search for *Venenivibrio* in the EMP category (qiita_empo_3) for all contexts (i.e., analyses). As this software is still under active development, the taxon search function is limited to closed reference data using Greengenes taxonomy. The Greengenes database only assigns *Venenivibrio* to the genus level in the 99 % OTU representative set (g__Venenivibrio), therefore the corresponding OTU ID from the 97 % rep set (OTU ID 32720) was also searched using the feature option. Thirdly, the 16S rRNA gene sequence of *V. stagnispumantis* CP.B2^T was used to search for exact matches in EMP datasets analysed by deblur, again using the feature selection in redbiom. Finally, the EMP subset used to create trading cards (otu_summary.emp_deblur_90bp.subset_2k.rare_5000.tsv) was checked for *Venenivibrio*. This analysis created 90bp tag sequences using deblur with Greengenes taxonomy.

D.1.4.9 Qiita database

All publicly available datasets in the Qiita database (https://qiita.ucsd.edu/redbiom/; accessed 05/Jul/2021; version 2021.05 9799e8f) were searched using redbiom by taxon (g__Venenivibrio), feature (OTU IDs 32720 and 1142935), and sequence (16S rRNA gene sequence of *V. stagnispumantis* CP.B2^T). The software is still in active development so

currently only OTU IDs/taxa names for closed reference data using Greengenes taxonomy or exact sequences for deblur can be searched.

D.1.4.10 NCBI & Google Scholar word search

A word search for the term '*Venenivibrio*' was conducted for all NCBI databases, including PubMed Central (accessed 28/Apr/2022), and results were compared against the 16S rRNA gene of *V. stagnispumantis* CP.BP[T], again using BLASTN v2.13.0 (megablast setting). A similar word search was also conducted in Google Scholar (accessed 28/Apr/2022).

D.1.4.11 16S rRNA gene sequence phylogeny

A FASTA file, containing all near-full length trimmed and quality controlled 16S rRNA gene sequences assigned to the genus *Venenivibrio* in the SILVA database (SSU r138.1; accessed 25-Mar-2022), was imported into the ARB software ecosystem (LTP_09_2021)[383]. Only sequences >1,000 bp in length were included in the analyses to avoid long branch attraction artefacts. Six new *Venenivibrio* spp. isolates (CPO1, KUI1, KUI2, LRO1, LRO2, and OKO1; unpublished results) were also included. In addition, reference sequences (RefSeq) from sister genera, *Sulfurihydrogenibium* and *Persephonella*, plus a selection of Aquificaceae [*Hydrogenobaculum acidophilum* 3H-1[T] (D16296), 'Aquifex pyrophilus Kol5a' (M83548), *Hydrogenivirga caldilitoris* IBSK3[T] (AB120294), *Hydrogenivirga okinawensis* LS12-2[T] (AB235314), *Hydrogenobacter hydrogenophilus* Z-829[T] (HE616187), *Hydrogenobacter subterraneus* HGP1[T] (AB026268), *Thermothrix azorensis* TM[T] (GU233444), *Thermocrinis minervae* CR11[T] (LT670846), *Thermocrinis albus* HI 11/12[T] (CP001931), *Thermocrinis jamiesonii* GBS1[T] (KC526152), and *Thermocrinis ruber* OC 1/4[T] (CP007028)] were added for more distantly related *Venenivibrio* sequences.

All sequences were aligned to the original, quality controlled, *V. stagnispumantis* CP.B2[T] (DQ989208) available in ARB, or as appropriate, *Persephonella* or *Sulfurihydrogenibium* reference sequences. A phylogenetic tree (Figure D.6) was constructed using TREE-PUZZLE, a quartet-puzzling maximum-likelihood algorithm with 10,000 puzzling steps (substitution model used: HKY).

D.1.5 Screening metagenomes for *Venenivibrio*

D.1.5.1 Metagenome selection for Hydrogenothermaceae alignment

Four local metagenomic samples from Aotearoa-New Zealand were aligned to selected Hydrogenothermaceae genomes to calculate the fraction of the genomes covered by the mapped reads (i.e., coverage breadth). Three domestic springs with the greatest read abundance of *V. stagnispumantis* CP.B2^T from metagenomic community profiling were selected (Waiotapu No.15 Feature 1, Whakarewarewa Feature 51 and Champagne Pool; Table D.2), along with a fourth spring that had increased read abundance from 16S rRNA gene community analysis (Radiata Pool). A total of six global metagenomes were also aligned. These included two springs from the Kirishima region, Kyushu Island, Japan (SRA Accession DRR163686, DRR163687), Jinata Spring, Shikinejima Island, Japan (SRR7905022), Obsidian Pool, Yellowstone National Park, USA (SRR6049666), Dewar Creek, British Columbia, Canada (SRR5580900), and Great Boiling Spring, Nevada, USA (SRR4027810; Table D.17). The three samples from Japan had the greatest number of *V. stagnispumantis* reads (26,312-594,730) from community profiling of all global spring metagenomes analysed (*n*=188), and while the three other global samples from the USA and Canada had putative traces of *Venenivibrio*, they also represented springs with synonymous physicochemistry to the source location for the type strain CP.B2^T (Champagne Pool, Waiotapu geothermal field, Aotearoa-New Zealand).

D.1.5.2 Generation of synthetic metagenomes

One million reads, modelled on illumina MiSeq sequencing technology, were generated using the iss function in InSilicoSeq v1.5.1 from each of the following genomes; *V. stagnispumantis* CP.B2^T (GOLD Analysis ID Ga0170441), *Sulfurihydrogenibium* sp. Y03AOP1 (GenBank Assembly Accession GCA_000020325.1), and *P. hydrogeniphila* 29W^T (GCA_900215515.1), along with one million reads from 10 random genomes in NCBI (accessed 21/Dec/2020). The random selection of NCBI genomes used to generate the synthetic metagenomes were 4.2 % *Streptococcus agalactiae* FDAARGOS_670 (CP044090.1), 12.7 % *Myroides phaeus* 18QD1AZ29W (CP047050.1), 12.9 % *Lactobacillus johnsonii* 3DG (CP047409.1), 12.1 % *Enterococcus faecium* V1836 (CP044264.1), 3.2 % *Pseudomonas* sp. NP-1 (CP056030.1), 3.0 % Candidatus *Sulcia muelleri* PSPU (AP013293.1), 37.8 % *Klebsiella variicola* KP2757 (CP060807.1), 3.1 % *Serratia*

surfactantfaciens YD25 (CP016948.1), 1.4 % *Acinetobacter baumannii* VB2107 (CP051474.1), and 9.5 % *Stenotrophomonas* sp. WZN-1 (CP021768.1).

D.2 SUPPLEMENTARY RESULTS

D.2.1 Genome annotation of *V. stagnispumantis* CP.B2[T]

Forty-seven scaffolds from the draft genome sequence of CP.B2[T] were analysed with an estimated total size of 1.6 Mbp and 29.6 % mol G+C[197]. The number of protein coding genes was 1,707, with 1,409 of these having predicted function. Two 16S rRNA genes were detected. A second copy of the CP.B2[T] genome exists in GOLD (Analysis ID Ga0170441) and IMG (Taxon ID 2724679818), sequenced directly from the culture collection DSMZ (DSM 18763) which was used to corroborate annotation. Detailed annotation on carbon assimilation, electron transport, sulfur, nitrogen and arsenic metabolisms, transmembrane transportation, cytosolic pH moderation, and comparison of Hydrogenothermaceae can be found below, with a full list of genes annotated with predicted function outlined in Table D.12.

D.2.1.1 Carbon assimilation

Presence of the Type 1 reductive TCA (rTCA) cycle was evident by the annotation of ATP-citrate lyase (ACL; gene *aclAB*), succinate dehydrogenase/fumarate reductase (*sdhABC*), and 2-oxoglutarate:ferredoxin oxidoreductase (*korAB*) genes[384]. Citryl-coa synthetase (*ccsAB*) and 2-oxoglutarate carboxylase (*cfiAB*), necessary for the alternate Type II rTCA cycle used by the Aquificaceae, were not present[384,385]. ACL is thought to have been acquired by the Hydrogenothermaceae through horizontal gene transfer[384], initially formed from a fusion of citryl-coA lyase (*ccl*) and citryl-coA synthetase (*ccsAB*), both used in the older Type II cycle. There was also no evidence of citrate synthase (*gltA*), found in many autotrophic bacteria, which can be used in both the oxidative TCA cycle for the production of citrate and the reversed oxidative cycle (roTCA) to fix carbon dioxide[386]. An entire Embden-Meyerhof-Parnas (EMP) pathway for glycolysis/gluconeogenesis was annotated[387], with no evidence found of the alternate Entner-Doudoroff (ED) pathway or the oxidative branch of the pentose phosphate pathway (oxPPP), traditional sources of NADPH for cell metabolism[388]. A non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase gene (*gapN*) was annotated,

which can produce NADPH by irreversibly oxidising glyceraldehyde-3-phosphate (GAP) straight to 3-phosphoglycerate (3-PG) in the EMP pathway[388]. There was also no evidence of genes for carbon monoxide dehydrogenase, or ribulose-bisphosphate carboxylase (RuBisCo) from the Calvin-Benson-Bassham cycle.

D.2.1.2 Electron transport

The genome had 13 subunits of the proton-translocating NADH:ubiquinone oxidoreductase (*nuoA-N*) for complex I of the electron transport chain. Succinate dehydrogenase (*sdhABC* and *frdB*) was annotated for complex II. Three subunits for the cytochrome bc1 complex (cytochrome c oxidoreductase) were encoded for complex III (*petABC*), with cytochrome bd being the respiratory terminal oxidase (*cydAB*; complex IV). Research has shown that cytochrome bd increases expression in response to a range of environmental stressors, including pH and temperature extremes[389]. The genome had membrane-bound F-type ATPases (*atpA-H*) for complex V.

D.2.1.3 Sulfur, nitrogen, & arsenic

The only gene found in the CP.B2^T genome from the SOX pathway (*soxD*) for sulfur/thiosulfate oxidation is a subunit in cytochrome c of the electron transport chain, and may also be involved in arsenic cycling[354]. Cytochrome c is reduced as an intermediate between complexes III and IV[390], and a proton gradient is created for ATP synthesis. There was no evidence of sulfite dehydrogenase (*sorAB* or *soeA*) in the genome, even though genes were present for the biosynthesis of the cofactor molybdopterin[391]. No further genes from other sulfur-metabolising pathways including sulfur oxygenase reductase (SOR), thiosulfate dehydrogenase (*tqoAB*), dissimilatory sulfite reductase (*dsrABC*), and heterodisulfide reductase subunits *hdrC1* and *hdrB2*[392] were detected. Nitrogen assimilation via the uptake of ammonia by glutamine synthetase (*glnA*) and glutamate synthase (*gltBD*) was noted, along with nitronate monooxygenase (NMO) which putatively generates nitrite from nitroalkane. No genes associated with nitrogen-dissimilatory pathways were identified. There was no evidence of genes encoding nitric oxide reductase (*norB*) for denitrification, which is commonly found in other Hydrogenothermaceae. Congruently, CP.B2^T was the only Hydrogenothermaceae analysed to not encode a nicotinamidase gene (*pncA*). While the arsenic resistance operon *arsRBC* was annotated in the CP.B2^T genome, there was no indication of the arsenic ABC transporter ATPase gene (*arsA*; as part of the more complex *arsRDABC* operon)[324]. Additionally, no putative genes involved with dissimilatory arsenic

metabolism were found, such as arsenate reductases (*arrAB*), including those found in *Pyrobaculum arsenaticum*[393] and *Geobacter lovleyi*[394], and arsenite oxidase (*aioAB*)[395], which is prevalent in *Sulfurihydrogenibium*[396].

D.2.1.4 Transmembrane transportation

Genes associated with transport systems for the facilitated diffusion of ions across the membrane were found for sodium, iron, calcium, magnesium, and ammonium, with multiple transport systems found for potassium. There were also several ABC transporters for zinc, cobalt, nickel, phosphate and a range of molecules involved in cell membrane formation (e.g., phospholipids, lipopolysaccharides, and lipoproteins). While a gene for the molybdate transport system regulatory protein (*modE*) was present, the rest of the high-affinity molybdate uptake system (*modABCD*) was missing. This is contrary to the rest of the Hydrogenothermaceae. There was also one copper exporting ATPase (*copB*) encoded in the genome, whereas all other Hydrogenothermaceae analysed had a range of two to five.

D.2.1.5 Ability to moderate cytosolic pH

Two genes associated with the *aguBDAC* operon for increasing alkalinity within the cell, agmatine deiminase (*aguA*) and putrescine amidase (*aguB*), were detected in the genome. Potential for malolactic fermentation was also noted by the presence of a lactate dehydrogenase gene (*dld*). No copies of glutamate, arginine or lysine decarboxylase, urease, or tryptophanase were observed, indicating that *Venenivibrio* does not possess these well-known mechanisms to manage cell pH homeostasis[397]. However, it should be noted that all Hydrogenothermaceae family members analysed appear to have no significant differences in genomic capabilities when it comes to cytosolic pH modulation.

D.2.1.6 Hydrogenothermaceae genomes

Along with CP.B2[T], a group 2d hydrogenase was found in three *Persephonella* spp., *H. marinus* VM1[T], and *S. subterraneum* HGMK1[T]. *Sulfurihydrogenibium* sp. Y03AOP1 and *S. yellowstonense* SS-5[T] had no hydrogenases, and instead rely on the oxidation of reduced sulfur species[292]. An arsenite oxidase gene (*aioA*, formally *aroA* or *aoxA*)[395], which was found in four *Sulfurihydrogenibium* spp. analysed, was not annotated in the CP.B2[T] genome.

D.2.2   Growth range reassessment of *V. stagnispumantis* CP.B2^T

Moderate growth of CP.B2^T was observed from 40.9-79.7 °C (opt 70.4 °C), with limited growth noted at both 38.5 °C and 80 °C. Cells at 80 °C were not viable when subcultured to new DSMZ Medium 1149 and cultivated at 70 °C. Growth was observed between pH 4.5-8.0, with the most significant growth observed at pH 5.5-7.0 (opt pH 6.0). A reduced number of viable cells were also visualised in medium from pH 3.0-4.0 and pH 8.5. Salinity tolerance was confirmed at 0.0-8.0 % NaCl (opt 0.0-0.2 %). Cells of CP.B2^T tolerated O$_2$ concentrations of up to 10 % v/v and as low as 1.25 % v/v of the headspace. Growth of CP.B2^T in Champagne Pool water was determined by the presence of typical aggregates in all three preparations. Within 24 hours, growth was also observed in all hybrid Obsidian Pool medium preparations. Microscopy confirmed increased numbers of CP.B2^T cells were viable in subsequent subculturing of unaltered spring water. Both the Champagne Pool and Obsidian Pool cultures were sequentially subcultured five times to ensure growth was not a result of essential element(s) carryover. Results are summarised in Table D.13.

D.2.3   Global search for 16S rRNA genes reported as, or closely related to, *Venenivibrio*

D.2.3.1 Nucleotide Collection (NCBI)

The Nucleotide collection (nt) of NCBI had 71,983,277 non-redundant DNA sequences at the time of analysis. Two entries were deposited for the 16S rRNA gene of *V. stagnispumantis* CP.B2^T (1506 bp in length); the first discovery of the gene in Champagne Pool, Aotearoa-New Zealand in 2007 (GenBank accession DQ989208)[138], and the characterisation of the type strain in 2008 (NR_044029)[179]. There was only one result with both ≥95 % sequence similarity and ≥95 % query coverage to the near full length 16S rRNA gene, a clone deposited in 2003 from a hot spring in Kuirau Park, Aotearoa-New Zealand (AF402979; 98.6 % sequence similarity, 1441 bp)[331]. The next most similar result across the full gene was *Sulfurihydrogenibium azorense* Az-Fu1^T (CP001229; 94.5 % sequence similarity). Six other partial entries (536-899 bp) were found with ≥98.0 % similarity, which were all clones from two separate studies of Champagne Pool; EF101539 and EF101540[138], and FN429034, FN429035, FN429036 and, FN429037[187]. The next closest results to CP.B2^T were 29 clones (538-629 bp) from the same study[398], with a similarity range of 95.2-95.7 %. These were most closely related to *S. azorense* Az-Fu1^T (98.9-99.8 %) and were from a hot spring in the

Nagano Prefecture, Japan. All other results from NCBI, both full and partial length, were <95 % sequence similarity to *V. stagnispumantis* CP.B2[T].

D.2.3.2 Sequence Read Archive (NCBI)

Three samples from SRA had total k-mer counts of 1672, 3218 and 7146 that assigned to the genus *Venenivibrio* when searched using the STAT program (Table D.16), which searched 208.5 gb of metadata from a possible 12.2 petabytes (or 27.3 quadrillion bases) of open access sequence data (06/Dec/2021; https://www.ncbi.nlm.nih.gov/sra/docs/sragrowth/). One of these samples was a metagenome from a geothermal spring in Aotearoa-New Zealand that was already included in this study (P1.0019, Radiata Pool; SRR14702244). The other two (SRR15830908 and SRR15830907) were 16S rRNA gene amplicon samples from seafloor hydrothermal vents near the Baja California Peninsula, Mexico. The greatest sequence similarity to the full length 16S rRNA gene of *V. stagnispumantis* CP.B2[T] found across both of these samples was 92 %, with 16 % of the gene covered by the query. The remaining samples from this SRA search (*n*=85) that produced k-mers assigned to *Venenivibrio* had counts that ranged between 25 and 430 (Table D.16).

D.2.3.3 SILVA database

The SILVA SSU r138.1 database contained a total of 33 entries classified to *Venenivibrio*, including the 16S rRNA gene of *V. stagnispumantis* CP.B2[T], from a total of 9,469,124 aligned rRNA sequences (Table D.14). Seven of these entries were clones originating from Aotearoa-New Zealand hot springs and all had ≥98.0 % pairwise sequence similarity to *V. stagnispumantis* CP.B2[T] (GenBank accessions AF402979, EF101539, EF101540, FN429034, FN429035, FN429036, and FN429037)[138,187,331]. The remaining 25 entries composed of one isolate and 24 clones, and had a similarity range of 79.3-94.7 % to *V. stagnispumantis* CP.B2[T] (Table D.14). An approximate maximum-likelihood phylogenetic tree of all 33 aligned sequences clustered only the seven Aotearoa-New Zealand clones together with *V. stagnispumantis* CP.B2[T] (Figure D.6). SILVA identified one closest neighbour (≥95 % identity) to *V. stagnispumantis* CP.B2[T] in the Ref NR database, the Aotearoa-New Zealand clone AF402979[331].

D.2.3.4 Ribosomal Database Project (RDP)

RDP (release 11_6) had a total of 3,351,829 sequences screened. Seven of these sequences matched the 16S rRNA gene sequences of *V. stagnispumantis* CP.B2[T] (98.6-99.6 % sequence

similarity), which were the same clones originating from Aotearoa-New Zealand that were found in the GenBank and SILVA databases (GenBank accessions AF402979, EF101539, EF101540, FN429034, FN429035, FN429036 and FN429037)[138,187,331].

D.2.3.5 Greengenes database

The latest version of the Greengenes 16S rRNA gene database v13.8 (August 2013) had 1,262,986 unique sequences. These were clustered into 203,452 and 99,321 representative OTUs at 99% and 97% similarity, respectively. Only one of these OTUs classified as *Venenivibrio stagnispumantis* (Greengenes OTU ID 1142935) in the 99 % representative set, with one additional sequence in the database (ID 189417) also found mapped to this cluster. Both sequences corresponded to the 16S rRNA gene sequence of the type strain, CP.B2^T (GenBank accessions DQ989208 and NR_044029). No OTU in the 97 % representative set was assigned to *Venenivibrio*, with only 25 classifying to the family Hydrogenothermaceae. From these 25 OTUs, OTU ID 32720 had 98.6 % sequence similarity to *V. stagnispumantis* CP.B2^T and the corresponding GenBank accession was AF402979, a clone found in a hot spring from Kuirau Park, Aotearoa-New Zealand[331]. The next highest similarity from the 97% OTUs was OTU ID 242647 at 94.2 %, with all others at ≤92.7 %. OTU ID 1142935 was also mapped to OTU 32720 cluster in the 97 % reference set.

D.2.3.6 Integrated Microbial Next Generation Sequencing (IMNGS) platform

From a total of 500,048 samples (24,835,679,746 reads), 31 OTUs were found across 29 samples with ≥95 % sequence similarity to the 16S rRNA gene of *V. stagnispumantis* CP.B2^T (Table D.15). None of these OTUs had ≥99 % similarity, with only eight having ≥97 % across ≤34 % of the CP.B2^T 16S rRNA gene. One of these OTUs (with ≥97 % similarity) was sourced from a sample of sugarcane root soil in Australia (SRA run accession SRR1924223)[399], and two were from peat soil adjacent to cold temperature springs in Canada (SRR1029457 and SRR2026416)[400]. However, the abundance of these OTUs ranged from one to five reads which contributed to only ≤0.01 % of the sample communities and were present in ecosystems not conducive to supporting *Venenivibrio* populations (e.g., soils, cold temperature). The remaining five OTUs with ≥97 % similarity to CP.B2^T were sourced from synthetic samples (Table D.15). Samples containing reads at ≥95 % sequence similarity to *Venenivibrio* which accounted for ≥0.1 % of the sample community were all from Aotearoa-New Zealand geothermal springs (*n*=10).

D.2.3.7 Integrated Microbial Genomes and Microbiomes (IMG/M) database

IMG/M had a total of 147,328 datasets at the time of analysis, including 26,097 distinct public non-redundant genomes and 83,287 metagenomic bins. Thirteen genomes were found classified to the Hydrogenothermaceae from 11 isolates and two MAGs. Two of these genomes were found assigned to the genus *Venenivibrio*: *Venenivibrio stagnispumantis* DSM 18763 (IMG Genome ID 2724679818; Gold Analysis ID Ga0170441), which was sequenced from the type strain stored in the DSMZ culture collection; and *Venenivibrio stagnispumantis* CP.B2 (IMG Genome ID 2799112217; Gold Analysis ID Ga0311387), which was sequenced by this study from the type strain stored in the original laboratory that isolated the microorganism. No other genomes in the entire collection contained genes that matched the 16S rRNA gene of *V. stagnispumantis* CP.B2^T. Likewise, no metagenome bins were found assigned to the genus. There were 20 bins classified to the Hydrogenothermaceae, using GTBD-Tk lineage, and these were either *Sulfurihydrogenibium* (*n*=11) or *Persephonella* (*n*=9).

D.2.3.8 Earth Microbiome Project (EMP) & Qiita database

No samples from the first release of the Earth Microbiome Project (EMP) had sequences that matched *Venenivibrio* in the analysis classified using SILVA taxonomy. This dataset had a total of 27,411 samples, 126,730 OTUs, and 1,754,319,647 sequence reads. The reference OTUs for this analysis were checked and two OTUs were classified to the genus *Venenivibrio* (AF402979.1.1441 and KM221400.1.1484). To confirm this result, the redbiom search function was enabled for both taxon (g__Venenivibrio) and features (OTU ID 32720 and 1142935) in the Greengenes-assigned taxonomy. Again, no hits were found. The EMP subset used to create trading cards had 155,002 unique sequences (ASVs; amplicon sequence variants) from a total of 10,000,000 reads, 2,000 samples and 95 studies. *Venenivibrio* was also not found in this subset and there were only two Hydrogenothermaceae-assigned ASVs, one found in five samples with a total of 238 reads, the second found in one sample with six reads (or observations). These six samples were from the Lost City Hydrothermal Field, Mid-Atlantic Ridge, and no significant similarity was found to the 16S rRNA gene of *V. stagnispumantis* CP.B2^T. Finally, no results were found for *Venenivibrio* within publicly available studies (*n*=599) in the Qiita platform. A total of 276,184 samples were searched by taxon name, OTU IDs and sequence.

D.2.3.9 NCBI & Google Scholar word search

A word search for *Venenivibrio* in all NCBI databases highlighted four entries in SRA and four entries in GenBank. Two of the SRA entries belonged to whole genome sequencing of the type strain CP.B2^T from the culture collection DSMZ (DSM 18763; SRA runs SRR5889102 and SRR5889103). The other two entries were amplicon sequences added in 2020 from an unpublished study of hot spring microbial communities in SiChuan, China (SRR10580885 and SRR10580889). The highest sequence similarity found with the full length 16S rRNA gene of *V. stagnispumantis* CP.B2^T was 93.1 %, with 16 % of the gene covered. The GenBank results included three accession numbers from the type strain CP.B2^T (DQ989208, NR_044029 and EF581124)[138,179], with the fourth result from an environmental clone labelled as uncultured *Venenivibrio* sp. CCB8131 (GenBank accession KY480601). This sequence had only 78.3 % sequence similarity to *V. stagnispumantis* CP.B2^T, with 85 % of the query covered.

There were eight published manuscripts that contained the word '*Venenivibrio*' from NCBI's PubMed Central database (PMC; accessed 28/Apr/2022). These included five studies that used samples from Aotearoa-New Zealand geothermal springs[156,185,195,401,402], and three publications that referenced the type strain *V. stagnispumantis* CP.B2^T and/or associated characterisation[293,403,404]. A similar search was also conducted in Google Scholar (accessed 28/Apr/2022) which highlighted an additional nine manuscripts that referenced the type strain[189,370,377,405–410], plus two that reported *Venenivibrio* taxa in amplicon sequencing of Chinese hot spring and wetland microbial communities[411,412]. The first of these studies described an average of 8.7 % *Venenivibrio* in the microbial communities across 16 hot springs[411]; however, the greatest sequence similarity to CP.B2^T from all of these samples was 93.5 % over 16 % of the 16S rRNA gene. The second study reported trace amounts of *Venenivibrio* (0.03-2.15 %) in three microbial communities of a voltage-applied wetland plant[412]. The 16S rRNA gene sequences from this study were not deposited in a database for review, so this result could not be verified.

D.2.4 Screening metagenomes for *Venenivibrio*

Similar to *V. stagnispumantis* CP.B2^T, low coverage breadth was observed from the six global metagenomes mapped to *P. hydrogeniphila* 29W^T, *S. yellowstonense* SS-5^T and *Sulfurihydrogenibium* sp. Y03AOP1 (Figure 4.6, Table D.18). This could be due in part to low numbers of Hydrogenothermaceae originally present in the sites analysed, enrichment

strategies excluding preferred Aquificota growth conditions, or perhaps the samples contained reads from an uncharacterised close relative of the genomes tested. The three samples from Japan had the highest coverage breadth of *P. hydrogeniphila* 29W^T at 2.8-6.1 % (Table D.18).

To test the sensitivity of Kraken2 at classifying reads as *Venenivibrio*, eight mock communities (samples A-H) were created with varying concentrations of *V. stagnispumantis* CP.B2^T, *Sulfurihydrogenibium* sp. Y03AOP1, *P. hydrogeniphila* 29W^T and a selection of random genomes (Table D.20). The software correctly classified concentrations of both *V. stagnispumantis* (±0.93 %) and *Sulfurihydrogenibium* sp. Y03AOP1 (±1.39 %), but failed to classify *P. hydrogeniphila* whatsoever. Non-equivalent concentrations were reported, however, for *P. marina* EX-H1 in those mock communities with *P. hydrogeniphila*. Interestingly, the sample that consisted of *P. hydrogeniphila* reads only (sample B) returned 1.42 % of the community as *Venenivibrio*, suggesting that previous traces of *Venenivibrio* found globally in both metagenomic and amplicon samples could have been inflated by similar family members being present in these sites.

The alignment of synthetic metagenomes back onto their respective genomes did not yield any anomalies (Figure D.7; Table D.21). In the three mock communities where *V. stagnispumantis* concentration was either 100, 10, or 33 %, the type strain genome was predominantly covered by each sample (98.7, 98.3, and 98.6 % respectively). Even in the samples with 1 % of the species, 83.2 % of the genome was recovered. While classification of samples with *P. hydrogeniphila* proved unreliable, this was not the case with aligning synthetic reads back onto this genome (>99.4 % coverage breadth with initial concentrations of 33, 45, or 100 %). *Sulfurihydrogenibium* sp. Y03AOP1 also had near complete coverage breadth across the genome (>99.96 %), with only <0.9 % of another species in the same genus (*S. yellowstonense* SS-5^T) being covered. The depth of coverage from the recovered genomes varied across samples, depending on taxon concentration that comprised the initial community makeup (Table D.21). Average coverage depth of *V. stagnispumantis* was 100x in the sample with 100 % of reads from that species. This reduced to 2.3x for samples with just 1 % *V. stagnispumantis*. A similar trend was observed for coverage depth for *P. hydrogeniphila* and *Sulfurihydrogenibium* sp. Y03AOP1.
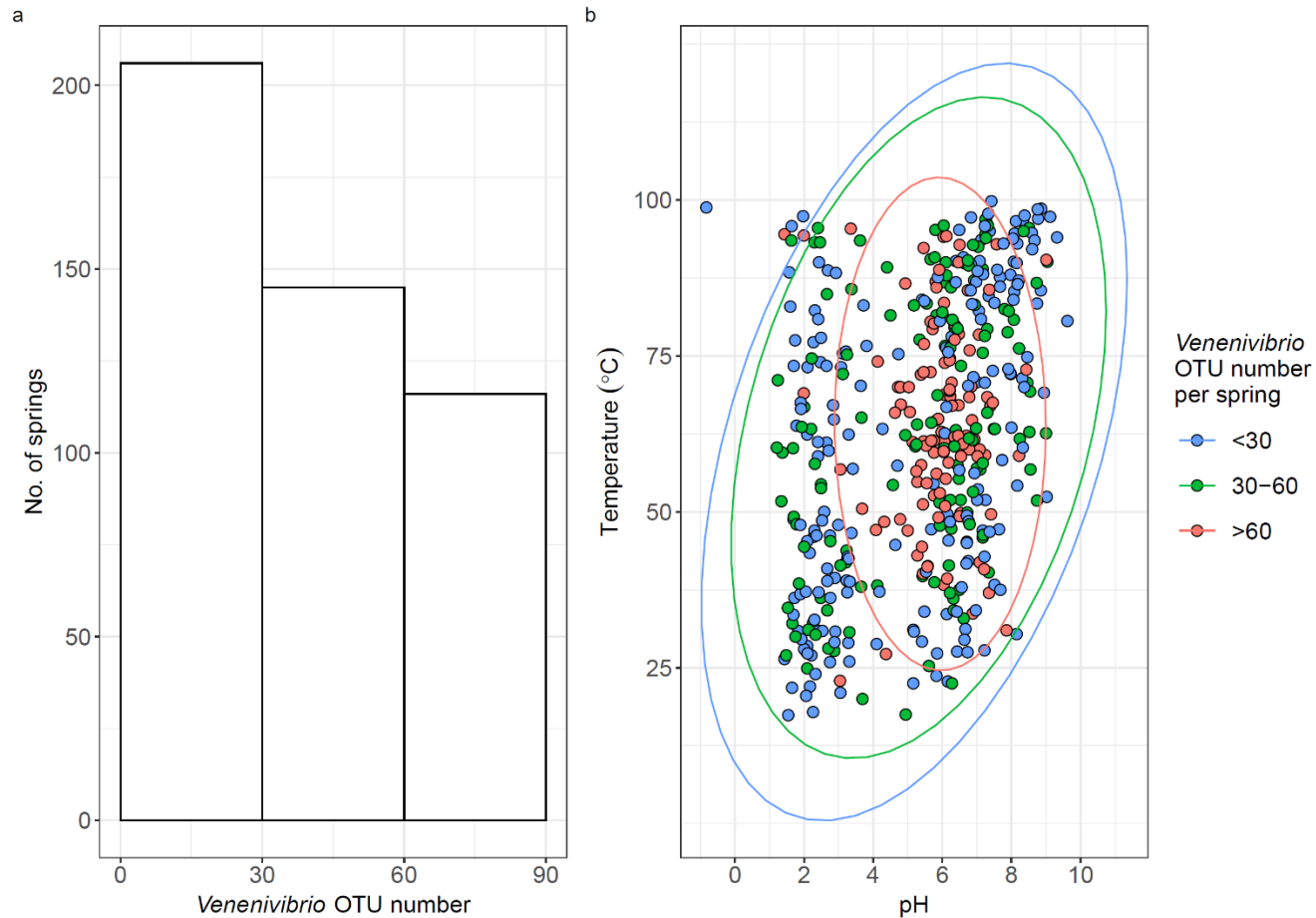
Figure D.1 - *Venenivibrio* 16S rRNA gene diversity in the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand. **(a)** The number of operational taxonomic units (OTUs) found across 467 geothermal springs that assigned to the genus *Venenivibrio* are shown (post filtering; n=99). **(b)** Springs are plotted as a function of environmental pH and temperature conditions. The number of *Venenivibrio*-assigned OTUs in each spring is represented by blue (<30), green (30-60) or red (>60), with data ellipses assuming multivariate t-distribution and a 95 % confidence interval.
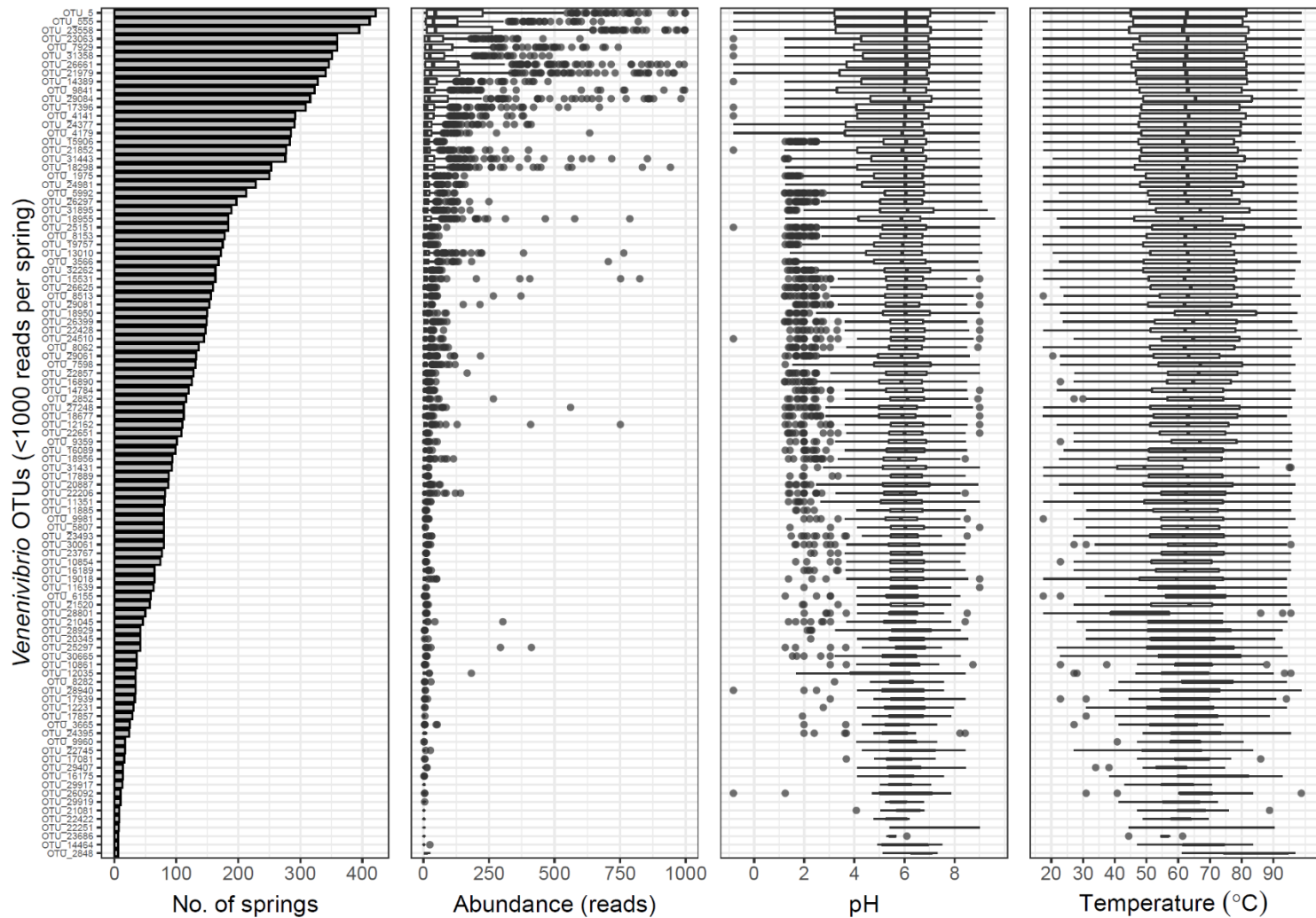
Figure D.2 - Prevalence, read abundance, pH and temperature ranges of low abundance *Venenivibrio* operational taxonomic units (OTUs). Low abundance OTUs (<1,000 reads; 10.5 % of the spring community) that assigned to the genus *Venenivibrio* are shown (post filtering; *n*=99). These are ordered by the number of springs where each OTU was found (i.e., prevalence). Median pH and temperature for low abundance OTUs were 6.0 (IQR 1.7) and 62.9 °C (IQR 23.4), respectively.

192
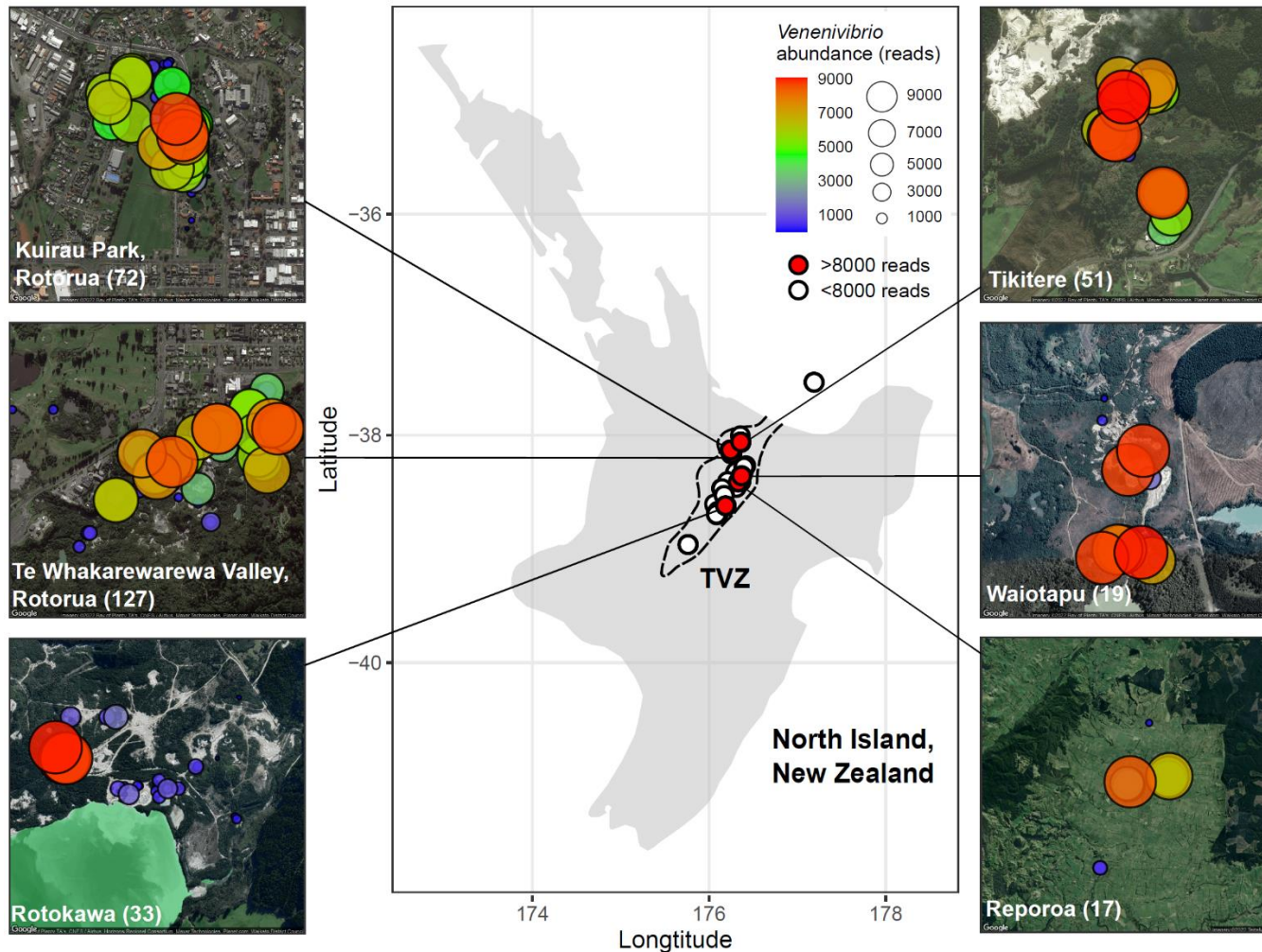
Figure D.3 - *Venenivibrio* hotspots in the Taupō Volcanic Zone (TVZ), Aotearoa-New Zealand. Geothermal springs containing *Venenivibrio* 16S rRNA genes (post filtering; *n*=467) are shown in the centre map, with springs containing ≥8,000 reads (≥85 % of the microbial community; *n*=20) highlighted in red. These springs are also presented in their respective geothermal fields, with circles coloured and sized according to read abundance. Total number of springs that contained *Venenivibrio* per geothermal field are in brackets. Median pH and temperature of these hotspots were 5.5 (IQR 0.8) and 66.0 °C (IQR 18.1) respectively.
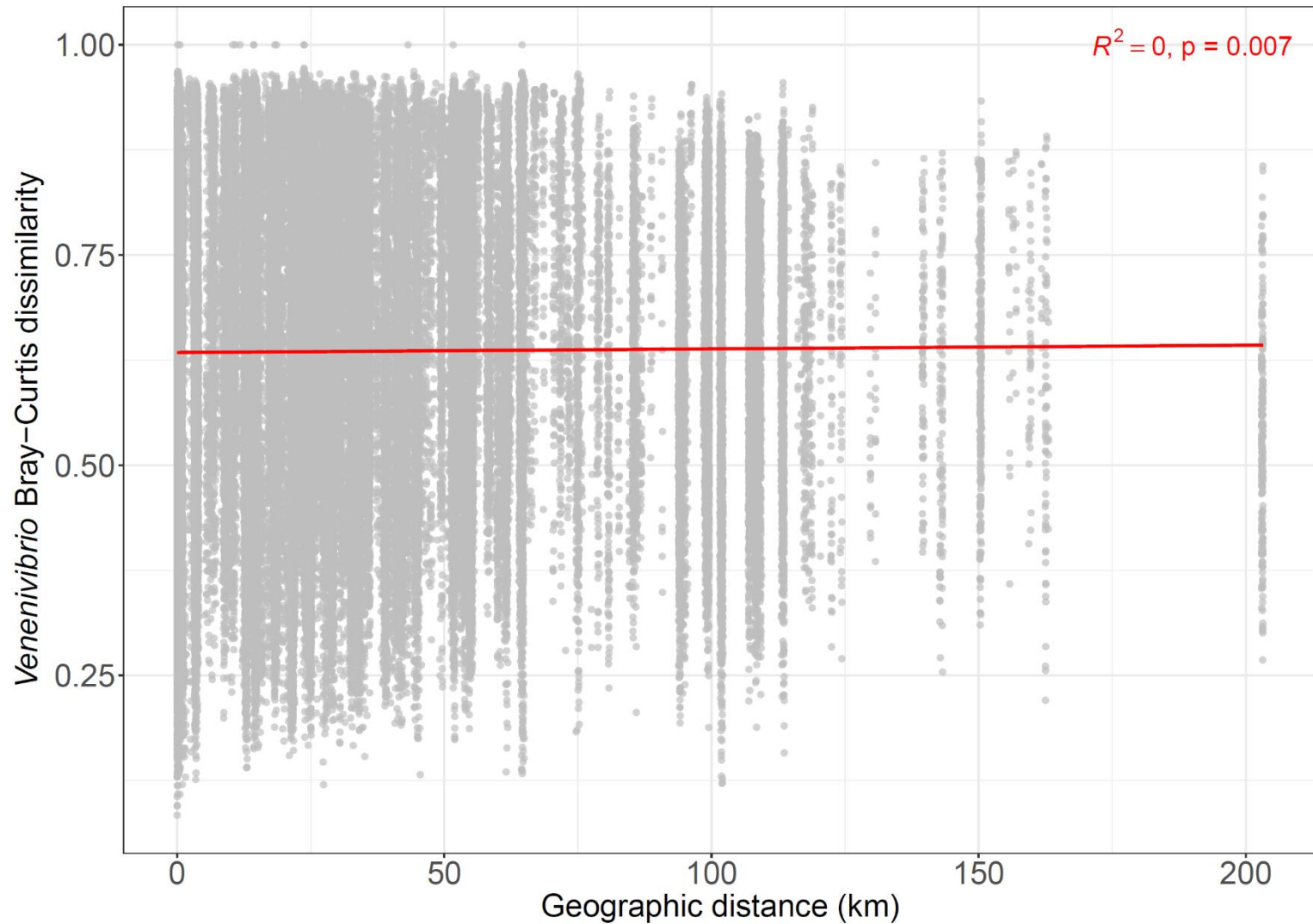
Figure D.4 - Distance-decay pattern of *Venenivibrio* populations in Aotearoa-New Zealand. Bray-Curtis dissimilarity was calculated between *Venenivibrio* populations (at operational taxonomic unit [OTU]-level) in geothermal springs (post-filtering, *n*=467) from across the TVZ, and plotted against pairwise geographic distance. A linear regression model was applied, highlighted in red (slope=4.36 x 10^{-5}).

Figure D.5 - *Venenivibrio* read abundance per geothermal spring in Aotearoa-New Zealand. Springs that containing *Venenivibrio*-assigned reads (post filtering; *n*=467) are shown, split by respective geothermal field. Operational taxonomic units (OTUs) with the greatest read abundance (*n*=20) are represented by colour in each spring. Maximum reads per spring community was 9,500.
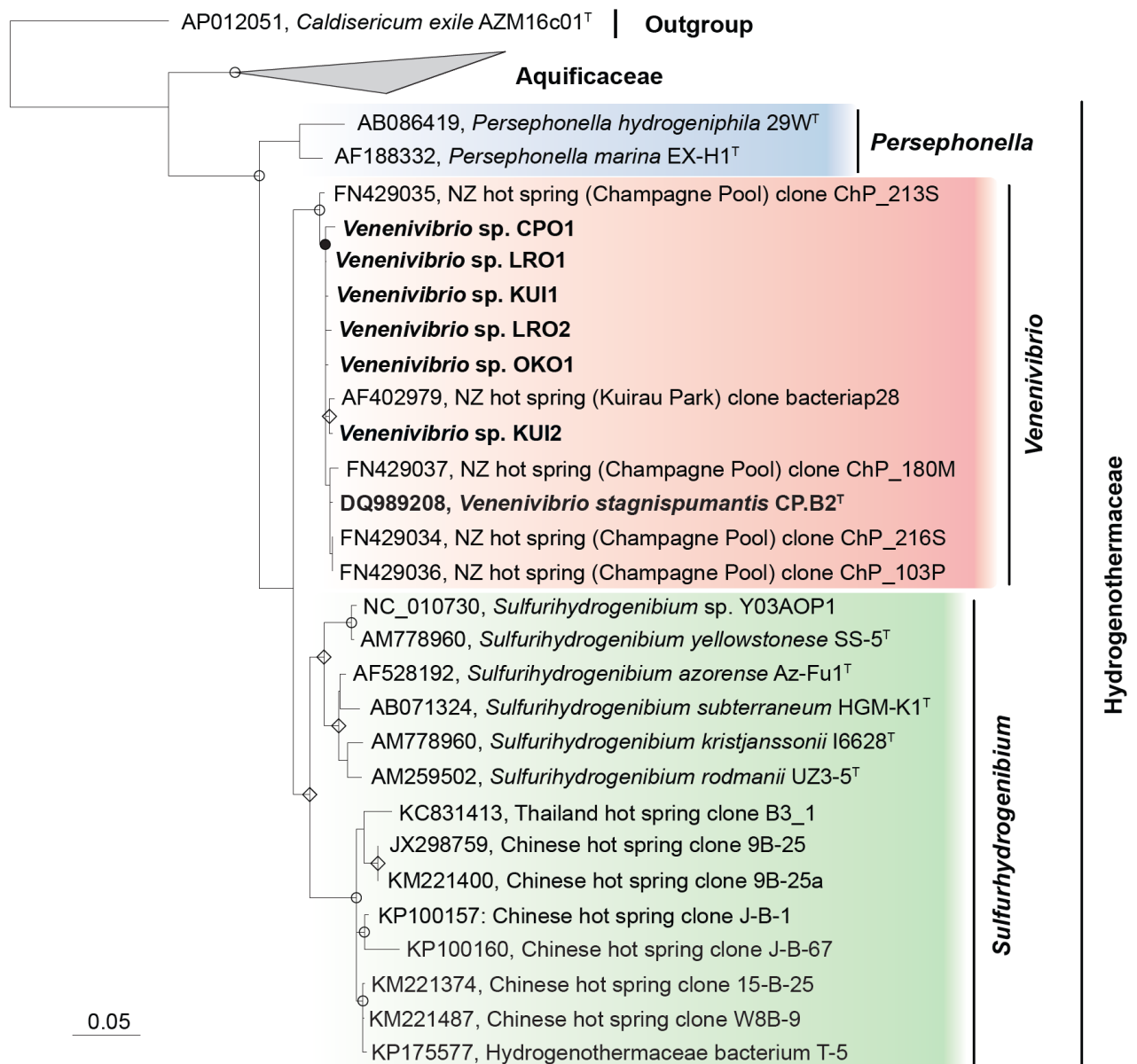
Figure D.6 - Phylogenetic tree of 16S rRNA gene sequences assigned to *Venenivibrio*. Maximum-likelihood quartet-puzzling phylogenetic tree showing the position of the type strain *Venenivibrio stagnispumantis* CP.B2[T], with near full-length environmental 16S rRNA gene clones that have been reported as belonging to or are closely related to the genus *Venenivibrio* from the SILVA database (SSU r138.1). Six new *Venenivibrio* strains recently isolated from the TVZ (CPO1, KUI1, KUI2, LRO1, LRO2 and OKO1), and type strains from closely related genera *Sulfurihydrogenibium* and *Persephonella* have also been included, with type strains from the family Aquificaceae collapsed for clarity. *Caldisericum exile* AZM16c01[T] was used as an outgroup. Quartet-puzzling support values (10,000 resamples) are represented by the following symbols: [open circles] >90 %, [closed circles] >80 %, and [open diamonds] >70 % at each internal branch. Multifurcations are drawn where the support value for a bifurcation is <50 %. The scale bar represents 0.05 substitutions per nucleotide position.
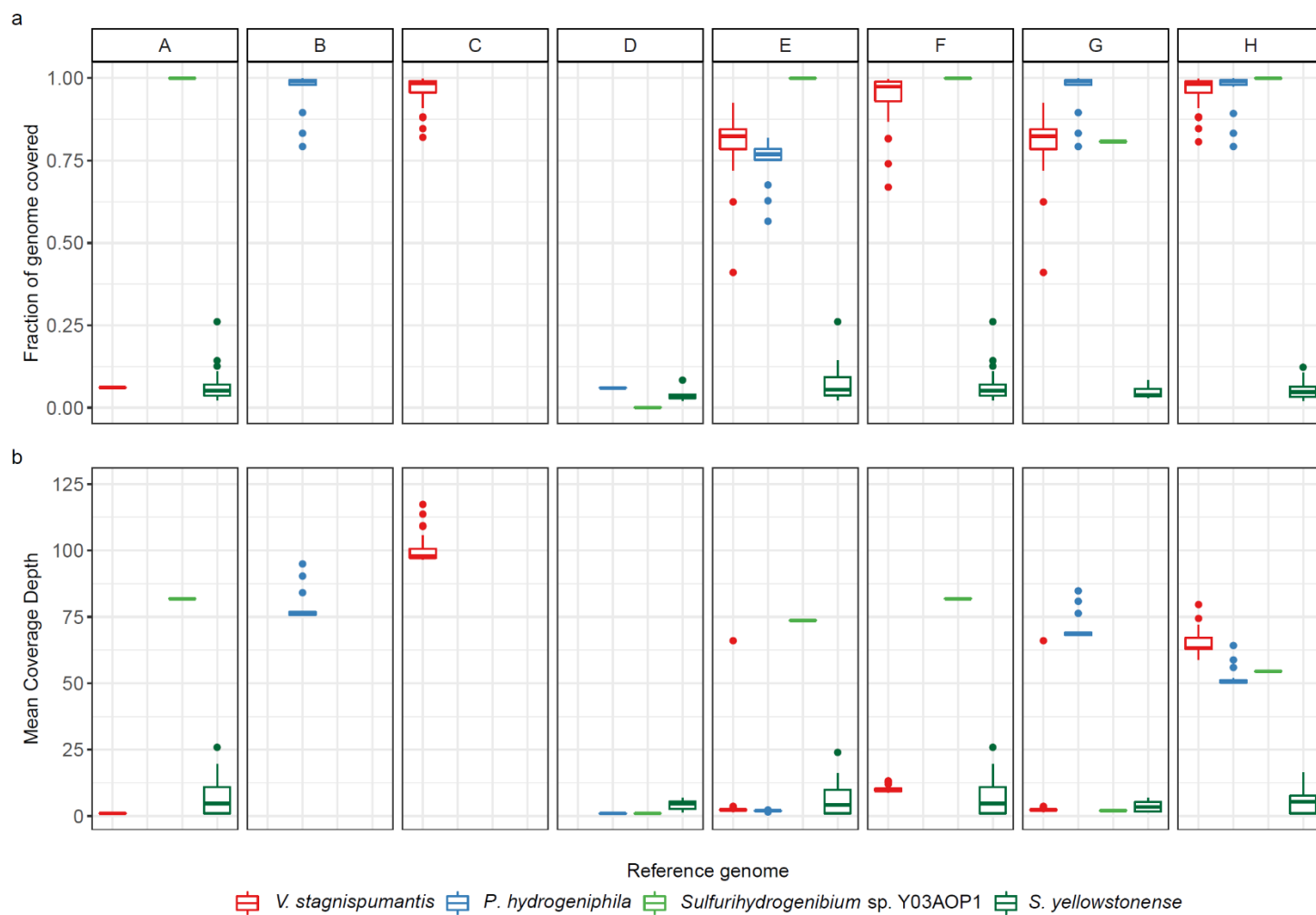
Figure D.7 - Hydrogenothermaceae genomes recovered from synthetic metagenomes. **(a)** Average coverage breadth (i.e., fraction) of reference genome contigs is displayed from alignment against eight mock metagenomic communities (A-H). **(b)** Average coverage depth of the reference genome contigs from the synthetic samples is shown. The four reference genomes used were *V. stagnispumantis* CP.B2^T (GOLD Analysis ID Ga0170441), *P. hydrogeniphila* 29W^T (GenBank Accession GCA_900215515.1), *Sulfurihydrogenibium yellowstonense* SS-5^T (GCF_000173615), and *Sulfurihydrogenibium* sp. Y03AOP1 (GCA_000020325.1).

D.4 SUPPLEMENTARY TABLES

Table D.1 - Aquificota-assigned taxa in Aotearoa-New Zealand geothermal springs. Unfiltered data from the 1,000 Springs Project[195] shows total read abundance of each genus, mean relative abundance across all 925 springs with standard deviation, maximum relative abundance (Max) found in a single spring, the number of springs where each genus was found, percentage prevalence across the dataset, and the number of OTUs assigned to each genus.

Family	Genus	Total read abundance	Mean relative abundance (%)	SD (%)	Max (%)	No. of springs	Prevalence (%)	No. of OTUs
Hydrogenothermaceae	*Venenivibrio*	879350	10.0	21.3	95.1	686	74.2	111
Aquificaceae	*Hydrogenobaculum*	840786	9.6	22.8	99.9	562	60.8	80
Aquificaceae	*Aquifex*	603635	6.9	18.0	95.0	460	49.7	72
Aquificaceae	n.d.	391681	4.5	13.0	96.8	496	53.6	42
Hydrogenothermaceae	*Sulfurihydrogenibium*	37926	0.4	3.7	60.8	134	14.5	13
Aquificaceae	*Hydrogenivirga*	25175	0.3	2.0	34.1	195	21.1	15
Aquificaceae	*Thermocrinis*	9556	0.1	0.8	14.4	241	26.1	2
Aquificaceae	*Hydrogenobacter*	9102	0.1	0.7	11.5	192	20.8	3
Aquificales (o)	*Thermosulfidibacter*	1364	0.0	0.2	4.1	83	9.0	1
Aquificaceae	n.d.	214	0.0	0.0	0.8	30	3.2	1

n.d.: not determined, SD: standard deviation; o: taxonomic order

Table D.2 - Geothermal spring metagenomes from the Taupō Volcanic Zone, Aotearoa-New Zealand, with associated metadata and *V. stagnispumantis* read abundance. Sixteen metagenomic samples from ten geothermal springs were screened for the presence of *Venenivibrio*, with read abundance counts and taxonomic classification performed using Kraken2. Sample characteristics presented are location details, description, temperature and pH. Samples are ordered by descending relative abundance of *V. stagnispumantis*. Metagenomes deposited in the Aotearoa Genomic Data Repository (AGDR) can be found under the Project ID TAONGA-AGDR00025 at https://doi.org/10.57748/vpk8-zp44. Samples NZ1 and NZ8 can be found in the Genomes Online Database (GOLD) at https://gold.jgi.doe.gov/. All other samples are in the Sequence Read Archive (SRA) at https://www.ncbi.nlm.nih.gov/sra.

| | | | | | | | | *Venenivibrio* | | |
Accession	Sample ID	TVZ spring	Geothermal field	Description	Collection date	T (°C)	pH	(%)	Reads	Reference
Ga0079637	NZ1*	Waiotapu No.15 Feature 1	Waiotapu	sediment	n.r.	74.0	4.5	45.1	3497580	Unpublished
AGDR00025	P2.0050*	Whakarewarewa Feature 51	Rotorua	water	2016-08-02	60.0	3.0	36.5	1167284	Unpublished
AGDR00025	CPp*	Champagne Pool	Waiotapu	pool	n.r.	75.0	5.5	27.7	1191790	Hug *et al.*, 2014[188]
AGDR00025	CPr	Champagne Pool	Waiotapu	rim	n.r.	68.0	5.5	18.2	656205	Hug *et al.*, 2014[188]
AGDR00025	AP	Champagne Pool	Waiotapu	terrace	n.r.	45.0	6.9	9.3	518220	Hug *et al.*, 2014[188]
AGDR00025	CPc	Champagne Pool	Waiotapu	channel	n.r.	75.0	5.8	8.6	295357	Hug *et al.*, 2014[188]
SRR14702244	P1.0019*	Radiata Pool	Ngatamariki	water	2013-08-14	63.3	7.5	0.8	340235	Zaremba-Niedzwiedzka *et al.*, 2017[413]
AGDR00025	P1.0103	Inferno Crater Lake	Waimangu	stage 2 oscillation	2013-10-10	55.1	2.4	0.6	153461	Unpublished
AGDR00025	P1.0037	Inferno Crater Lake	Waimangu	stage 4 retreat	2013-08-30	41.0	2.4	0.2	34457	Unpublished
Ga0079638	NZ8	Waimangu No.16 Feature 3	Waimangu	sediment	n.r.	90.0	7.5	0.1	3974	Unpublished
AGDR00025	IMP.15	Fred+Maggie's Pool	Orakei Korako	water	2017-02-23	96.4	6.9	0.0	12377	Unpublished
AGDR00025	P1.0108	Inferno Crater Lake	Waimangu	stage 3 overflow	2013-10-18	69.1	2.6	0.0	196	Unpublished
AGDR00025	IMP.14	Fred+Maggie's Pool	Orakei Korako	sediment	2017-02-23	96.4	6.9	0.0	1661	Unpublished
SRR10063241	RIES	Eastern Spring	Raoul Island	water	2016-03-31	27.2	5.4	0.0	24	Unpublished
SRR10063241	RIMB1	Marker Bay	Raoul Island	water	2016-03-31	29.9	5.9	0.0	312	Unpublished
SRR10063240	RIWS	Western Springs, Green Lake	Raoul Island	water	2016-03-31	34.5	7.0	0.0	263	Unpublished

n.r.: not reported

*Samples aligned to Hydrogenothermaceae reference genomes

Table D.3 - Single linear regressions of Aquificota diversity against 46 environmental variables. Both OTU number and total read abundance in each spring where Aquificota was found (post filtering; n=783) were analysed against 46 physicochemical parameters. Results are ordered by descending R^2 value (OTU number), with no transformations performed on any parameter prior to analysis. The false discovery rate (FDR) Benjamini-Hochberg correction was applied to all p-values.

	OTU number			Read abundance		
	R^2 value	p-value	Corrected p-value	R^2 value	p-value	Corrected p-value
OTU number	n.a.	n.a.	n.a.	0.441	<0.001	<0.001
Read abundance	0.441	<0.001	<0.001	n.a.	n.a.	n.a.
TEMP	0.260	<0.001	<0.001	0.225	<0.001	<0.001
ORP	0.149	<0.001	<0.001	0.116	<0.001	<0.001
pH	0.101	<0.001	<0.001	0.033	<0.001	<0.001
COND	0.047	<0.001	<0.001	0.043	<0.001	<0.001
Ag	0.038	<0.001	<0.001	0.013	0.002	0.004
HCO_3^-	0.037	<0.001	<0.001	0.025	<0.001	<0.001
SO_4^{2-}	0.025	<0.001	<0.001	0.031	<0.001	<0.001
S	0.023	<0.001	<0.001	0.030	<0.001	<0.001
U	0.022	<0.001	<0.001	0.021	<0.001	<0.001
Rb	0.013	0.001	0.005	0.007	0.020	0.035
As	0.012	0.002	0.006	0.012	0.003	0.007
B	0.012	0.002	0.006	0.012	0.002	0.006
Mn	0.012	0.002	0.007	0.018	<0.001	0.001
Al	0.011	0.004	0.012	0.014	0.001	0.003
Cs	0.010	0.006	0.016	0.008	0.015	0.027
NO_3^-	0.009	0.007	0.017	0.024	<0.001	<0.001
Tl	0.009	0.007	0.019	0.003	0.118	0.157
Si	0.009	0.009	0.021	0.013	0.001	0.004
Cu	0.007	0.018	0.042	0.004	0.065	0.105
dO	0.006	0.028	0.061	0.024	<0.001	<0.001
Sr	0.006	0.031	0.065	0.011	0.003	0.007
NO_2^-	0.005	0.043	0.086	0.008	0.012	0.023
K	0.005	0.055	0.105	0.002	0.169	0.213
Se	0.005	0.060	0.111	0.013	0.001	0.004
Pb	0.004	0.089	0.158	0.001	0.346	0.396
Ca	0.004	0.097	0.166	0.010	0.006	0.013
Cl^-	0.003	0.111	0.184	0.005	0.055	0.091
Li	0.003	0.137	0.219	0.004	0.075	0.113
Cd	0.002	0.163	0.252	0.001	0.448	0.489
Fe^{2+}	0.002	0.170	0.255	0.011	0.004	0.009
Fe	0.002	0.234	0.341	0.009	0.009	0.019
Ba	0.002	0.254	0.359	0.001	0.382	0.426
Br	0.002	0.266	0.365	0.002	0.255	0.314
Mg	0.001	0.342	0.456	0.006	0.029	0.050
Zn	0.001	0.357	0.463	0.000	0.535	0.571
Hg	0.001	0.427	0.540	0.000	0.573	0.598
Ni	0.000	0.610	0.725	0.004	0.094	0.137

PO_4^{3-}	0.000	0.617	0.725	0.008	0.014	0.026
Co	0.000	0.628	0.725	0.003	0.113	0.154
TURB	0.000	0.634	0.725	0.004	0.069	0.108
Mo	0.000	0.700	0.781	0.000	0.671	0.685
H_2S	0.000	0.757	0.826	0.002	0.275	0.330
Na	0.000	0.828	0.883	0.000	0.713	0.713
NH_4^+	0.000	0.988	0.995	0.001	0.322	0.377
V	0.000	0.992	0.995	0.003	0.102	0.144
Cr	0.000	0.995	0.995	0.003	0.121	0.157

COND: conductivity, dO: dissolved oxygen, ORP: oxidation-reduction potential, n.a.: not applicable, TEMP: temperature, TURB: turbidity

Table D.4 - Multiple linear regression model of Aquificota OTU number against metadata. Physicochemical variables were measured from each spring where Aquificota reads were present (post filtering; n=783). Samples with missing or aberrant data were not included in this analysis (n=14). An Akaike information criterion (AIC) was applied to find the model of best fit, with the final R^2 value being 0.389 ($p<0.001$). Results are ordered by ascending p-value.

	Estimate	Std. Error	t-value	p-value	FDR corrected p-value
(Intercept)	23.690	4.072	5.817	<0.001	<0.001
TEMP	0.665	0.050	13.375	<0.001	<0.001
ORP	-0.038	0.006	-5.993	<0.001	<0.001
Ag	-8473	1846	-4.589	<0.001	<0.001
Fe	0.831	0.252	3.296	0.001	0.004
U	-12220	3717	-3.287	0.001	0.004
Br	3.817	1.186	3.218	0.001	0.005
NO_2^-	-35.930	11.330	-3.172	0.002	0.005
Cs	16.070	5.168	3.110	0.002	0.005
B	-0.699	0.233	-2.999	0.003	0.006
Rb	-15.710	5.260	-2.986	0.003	0.006
Cd	1923	652	2.950	0.003	0.007
Cl^-	-0.016	0.006	-2.760	0.006	0.011
NH_4^+	0.052	0.020	2.566	0.010	0.018
Cr	-105.500	42.520	-2.481	0.013	0.021
Mn	-7.604	3.114	-2.442	0.015	0.022
Fe^{2+}	-0.180	0.076	-2.359	0.019	0.026
dO	1.365	0.598	2.284	0.023	0.030
Ca	0.090	0.041	2.167	0.031	0.039
S	-0.017	0.008	-2.145	0.032	0.039
Na	0.019	0.010	1.924	0.055	0.063
As	-2.026	1.153	-1.757	0.079	0.087
Ni	170.300	109.900	1.549	0.122	0.126
Hg	-824.700	538.500	-1.531	0.126	0.126

dO: dissolved oxygen, FDR: false discovery rate, ORP: oxidation-reduction potential, TEMP: temperature

Table D.5 - Multiple linear regression model of Aquificota read abundance against metadata. Physicochemical variables were measured from each spring where Aquificota reads were present (post filtering; n=783). Samples with missing or aberrant data were not included in the analysis (n=14). An Akaike information criterion (AIC) was applied to find the model of best fit, with the R^2 value for the final model being 0.358 (p<0.001). Results are ordered by ascending p-value.

	Estimate	Std. Error	t-value	p-value	FDR corrected p-value
(Intercept)	1174	545.300	2.153	0.032	0.040
TEMP	63.440	4.947	12.823	<0.001	<0.001
ORP	-3.799	0.706	-5.382	<0.001	<0.001
pH	-299.800	67.130	-4.466	<0.001	<0.001
SO_4^{2-}	-2.180	0.496	-4.395	<0.001	<0.001
Mn	-1477	323.700	-4.565	<0.001	<0.001
Fe^{2+}	-27.940	7.949	-3.515	<0.001	<0.001
Mg	44.070	9.752	4.519	<0.001	0.002
Fe	92.360	30.780	3.001	0.003	0.007
Hg	159500	53610	2.975	0.003	0.007
Na	2.775	0.935	2.968	0.003	0.007
Sr	2671	905.700	2.949	0.003	0.007
TURB	1.411	0.479	2.947	0.003	0.007
Al	32.740	11.450	2.860	0.004	0.008
Tl	89740	37660	2.383	0.017	0.031
Cl^-	-1.403	0.595	-2.356	0.019	0.031
Ba	2427	1058	2.293	0.022	0.035
Cr	-9509	4273	-2.225	0.026	0.039
Co	-260100	120400	-2.160	0.031	0.040
NH_4^+	4.952	2.309	2.145	0.032	0.040
Ni	22110	11140	1.984	0.048	0.057
Li	-178.600	96.280	-1.855	0.064	0.073
NO_3^-	-568.700	323.200	-1.759	0.079	0.084
U	-729600	416600	-1.751	0.080	0.084
NO_2^-	-2160	1297	-1.666	0.096	0.096

FDR: false discovery rate, ORP: oxidation-reduction potential, TEMP: temperature, TURB: turbidity

Table D.6 - Correlations between Aquificota diversity and environmental variables. Both OTU number and read abundance in each spring (post filtering; $n=783$) was correlated against 46 physicochemical parameters using Pearson's and Spearman's coefficient. Due to missing or aberrant data, only 771, 781 and 782 springs were used to analyse conductivity, HCO_3^-, and H_2S respectively. The false discovery rate (FDR) Benjamini-Hochberg correction was applied to all p-values. Results are ordered by descending Pearson's coefficient for correlations with OTU number.

	OTU number				Read abundance			
	Pearson	Corrected p-value	Spearman	Corrected p-value	Pearson	Corrected p-value	Spearman	Corrected p-value
OTU number	n.a.	n.a.	n.a.	n.a.	0.379	<0.001	0.371	<0.001
Read abundance	0.379	<0.001	0.371	<0.001	n.a.	n.a.	n.a.	n.a.
TEMP	0.211	<0.001	0.183	<0.001	0.169	<0.001	0.159	<0.001
ORP	0.103	<0.001	0.082	<0.001	0.102	<0.001	0.089	<0.001
HCO_3^-	0.091	<0.001	0.079	<0.001	0.031	<0.001	0.043	<0.001
pH	0.079	<0.001	0.100	<0.001	0.011	<0.001	0.019	<0.001
dO	0.054	<0.001	0.055	<0.001	0.025	<0.001	0.038	<0.001
Cd	0.047	<0.001	-0.012	<0.001	-0.016	<0.001	-0.012	<0.001
Zn	0.045	<0.001	-0.021	<0.001	-0.016	<0.001	-0.014	<0.001
Co	0.038	<0.001	-0.013	<0.001	-0.022	<0.001	-0.024	<0.001
Cu	0.031	<0.001	0.046	<0.001	0.007	<0.001	0.013	<0.001
Pb	0.031	<0.001	-0.011	<0.001	0.002	0.384	-0.003	0.143
Ni	0.031	<0.001	0.035	<0.001	-0.015	<0.001	0.007	<0.001
Cr	0.025	<0.001	0.052	<0.001	-0.023	<0.001	0.005	0.014
Ag	0.020	<0.001	0.008	<0.001	0.009	<0.001	0.005	0.006
NO_3^-	0.019	<0.001	0.020	<0.001	-0.026	<0.001	-0.033	<0.001
V	0.015	<0.001	-0.018	<0.001	-0.019	<0.001	-0.018	<0.001
NO_2^-	-0.005	0.011	0.012	<0.001	-0.010	<0.001	-0.020	<0.001
Fe	-0.005	0.004	-0.036	<0.001	-0.020	<0.001	-0.026	<0.001
Mo	-0.007	<0.001	-0.006	0.002	0.005	0.003	0.022	<0.001
U	-0.013	<0.001	-0.059	<0.001	-0.010	<0.001	-0.030	<0.001
Br	-0.018	<0.001	-0.013	<0.001	0.002	0.265	0.000	0.901
PO_4^{3+}	-0.019	<0.001	-0.030	<0.001	-0.019	<0.001	-0.019	<0.001
SO_4^{2-}	-0.021	<0.001	-0.016	<0.001	-0.028	<0.001	-0.027	<0.001
Tl	-0.023	<0.001	-0.042	<0.001	0.009	<0.001	-0.006	0.001
NH_4^+	-0.023	<0.001	-0.047	<0.001	-0.015	<0.001	0.007	<0.001
COND	-0.024	<0.001	-0.008	<0.001	-0.003	0.135	-0.002	0.341
H_2S	-0.024	<0.001	0.000	0.827	0.005	0.006	0.055	<0.001
Mn	-0.026	<0.001	-0.033	<0.001	-0.018	<0.001	-0.020	<0.001
S	-0.028	<0.001	-0.040	<0.001	-0.026	<0.001	-0.028	<0.001
As	-0.031	<0.001	-0.028	<0.001	-0.002	0.267	-0.004	0.024
Si	-0.032	<0.001	-0.034	<0.001	0.007	<0.001	0.011	<0.001
Sr	-0.034	<0.001	-0.027	<0.001	-0.013	<0.001	0.009	<0.001
TURB	-0.036	<0.001	-0.045	<0.001	0.029	<0.001	0.024	<0.001
Na	-0.041	<0.001	-0.022	<0.001	-0.012	<0.001	-0.008	<0.001
Rb	-0.041	<0.001	-0.038	<0.001	-0.007	<0.001	0.001	0.477

K	-0.044	<0.001	-0.038	<0.001	-0.005	0.009	-0.006	0.001
B	-0.046	<0.001	-0.028	<0.001	-0.014	<0.001	-0.006	0.001
Cs	-0.047	<0.001	-0.033	<0.001	-0.010	<0.001	0.004	0.027
Li	-0.047	<0.001	-0.027	<0.001	-0.015	<0.001	-0.003	0.145
Se	-0.048	<0.001	-0.032	<0.001	-0.020	<0.001	-0.017	<0.001
Ca	-0.003	0.107	-0.024	<0.001	-0.012	<0.001	0.002	0.309
Cl$^-$	0.003	0.155	-0.020	<0.001	-0.009	<0.001	-0.012	<0.001
Hg	0.002	0.220	0.072	<0.001	0.001	0.590	0.100	<0.001
Mg	0.001	0.446	0.003	0.112	-0.011	<0.001	-0.015	<0.001
Al	-0.001	0.537	-0.032	<0.001	-0.015	<0.001	-0.025	<0.001
Ba	0.001	0.551	0.008	<0.001	-0.001	0.650	0.026	<0.001
Fe^{2+}	-0.001	0.713	-0.042	<0.001	-0.019	<0.001	-0.033	<0.001

COND: conductivity, dO: dissolved oxygen, n.a.: not applicable, ORP: oxidation-reduction potential, TEMP: temperature, TURB: turbidity

Table D.7 - Characteristics of *Venenivibrio*-assigned OTUs. OTUs (post filtering; *n*=99) are ordered by descending mean relative abundance across the original 925 geothermal springs analysed. The maximum relative abundance in an individual spring (max) is also shown.

Genus	OTU ID	Prevalence	Prevalence (%)	Mean relative abundance (%)	SD (%)	Max (%)
Venenivibrio	OTU_23558	451	48.76	3.23	5.72	41.97
Venenivibrio	OTU_5	456	49.30	2.64	6.00	76.63
Venenivibrio	OTU_26661	410	44.32	2.55	5.93	62.29
Venenivibrio	OTU_21979	396	42.81	1.89	4.75	38.33
Venenivibrio	OTU_29084	382	41.30	1.06	3.10	26.64
Venenivibrio	OTU_555	447	48.32	1.05	1.95	22.63
Venenivibrio	OTU_18955	249	26.92	0.82	3.98	34.63
Venenivibrio	OTU_7929	391	42.27	0.72	1.33	7.82
Venenivibrio	OTU_9841	380	41.08	0.57	1.97	17.17
Venenivibrio	OTU_31443	343	37.08	0.56	2.14	19.06
Venenivibrio	OTU_23063	399	43.14	0.48	0.84	6.28
Venenivibrio	OTU_31358	399	43.14	0.47	0.80	5.05
Venenivibrio	OTU_18298	300	32.43	0.40	1.66	20.93
Venenivibrio	OTU_14389	369	39.89	0.35	0.71	5.00
Venenivibrio	OTU_17396	368	39.78	0.31	0.78	7.06
Venenivibrio	OTU_4141	348	37.62	0.24	0.51	4.02
Venenivibrio	OTU_24377	350	37.84	0.22	0.54	4.33
Venenivibrio	OTU_4179	340	36.76	0.17	0.43	6.67
Venenivibrio	OTU_21852	321	34.70	0.16	0.40	4.22
Venenivibrio	OTU_13010	231	24.97	0.11	0.50	8.04
Venenivibrio	OTU_15906	337	36.43	0.10	0.15	0.84
Venenivibrio	OTU_24981	283	30.59	0.10	0.23	1.67
Venenivibrio	OTU_15531	226	24.43	0.10	0.61	8.68
Venenivibrio	OTU_1975	309	33.41	0.10	0.20	1.64
Venenivibrio	OTU_26297	251	27.14	0.09	0.25	2.63

Genus	OTU ID	Prevalence	Prevalence (%)	Mean relative abundance (%)	SD (%)	Max (%)
Venenivibrio	OTU_3566	227	24.54	0.08	0.39	7.42
Venenivibrio	OTU_5992	276	29.84	0.08	0.17	1.26
Venenivibrio	OTU_31895	253	27.35	0.07	0.19	1.87
Venenivibrio	OTU_12162	173	18.70	0.06	0.43	7.91
Venenivibrio	OTU_8513	238	25.73	0.05	0.23	3.92
Venenivibrio	OTU_25151	247	26.70	0.05	0.10	0.94
Venenivibrio	OTU_26399	208	22.49	0.04	0.11	0.95
Venenivibrio	OTU_27248	176	19.03	0.04	0.29	5.91
Venenivibrio	OTU_32262	228	24.65	0.04	0.10	0.75
Venenivibrio	OTU_7598	207	22.38	0.04	0.12	1.27
Venenivibrio	OTU_29061	190	20.54	0.04	0.16	2.29
Venenivibrio	OTU_29081	216	23.35	0.03	0.14	2.27
Venenivibrio	OTU_8153	238	25.73	0.03	0.07	0.63
Venenivibrio	OTU_8062	192	20.76	0.03	0.11	1.00
Venenivibrio	OTU_24510	206	22.27	0.03	0.08	0.79
Venenivibrio	OTU_19757	237	25.62	0.03	0.06	0.58
Venenivibrio	OTU_26625	232	25.08	0.03	0.07	0.54
Venenivibrio	OTU_22428	212	22.92	0.03	0.07	0.81
Venenivibrio	OTU_18950	230	24.86	0.03	0.08	0.92
Venenivibrio	OTU_2852	181	19.57	0.02	0.14	2.81
Venenivibrio	OTU_18956	150	16.22	0.02	0.10	1.21
Venenivibrio	OTU_14784	175	18.92	0.02	0.06	0.45
Venenivibrio	OTU_22206	147	15.89	0.02	0.12	1.49
Venenivibrio	OTU_22857	186	20.11	0.02	0.09	1.76
Venenivibrio	OTU_25297	80	8.65	0.02	0.25	4.34
Venenivibrio	OTU_20887	131	14.16	0.02	0.07	0.67
Venenivibrio	OTU_16890	180	19.46	0.02	0.04	0.51
Venenivibrio	OTU_16089	164	17.73	0.02	0.04	0.36
Venenivibrio	OTU_18677	195	21.08	0.02	0.04	0.41
Venenivibrio	OTU_9359	163	17.62	0.01	0.04	0.56
Venenivibrio	OTU_21045	89	9.62	0.01	0.15	3.19
Venenivibrio	OTU_22651	171	18.49	0.01	0.03	0.24
Venenivibrio	OTU_23493	128	13.84	0.01	0.03	0.34
Venenivibrio	OTU_31431	150	16.22	0.01	0.03	0.22
Venenivibrio	OTU_19018	116	12.54	0.01	0.05	0.54
Venenivibrio	OTU_16189	117	12.65	0.01	0.03	0.32
Venenivibrio	OTU_17889	151	16.32	0.01	0.03	0.21
Venenivibrio	OTU_30061	134	14.49	0.01	0.03	0.29
Venenivibrio	OTU_12035	67	7.24	0.01	0.09	1.93
Venenivibrio	OTU_11351	143	15.46	0.01	0.03	0.31
Venenivibrio	OTU_9981	148	16.00	0.01	0.02	0.23
Venenivibrio	OTU_23767	136	14.70	0.01	0.02	0.13
Venenivibrio	OTU_21520	119	12.86	0.01	0.02	0.20
Venenivibrio	OTU_11885	149	16.11	0.01	0.02	0.14

Genus	OTU ID	Prevalence	Prevalence (%)	Mean relative abundance (%)	SD (%)	Max (%)
Venenivibrio	OTU_5807	148	16.00	0.01	0.01	0.08
Venenivibrio	OTU_11639	136	14.70	0.01	0.02	0.12
Venenivibrio	OTU_10854	124	13.41	0.01	0.02	0.14
Venenivibrio	OTU_6155	124	13.41	0.01	0.01	0.14
Venenivibrio	OTU_28801	91	9.84	0.01	0.02	0.15
Venenivibrio	OTU_3665	75	8.11	0.00	0.04	0.56
Venenivibrio	OTU_28929	112	12.11	0.00	0.01	0.06
Venenivibrio	OTU_8282	99	10.70	0.00	0.02	0.31
Venenivibrio	OTU_30665	71	7.68	0.00	0.01	0.15
Venenivibrio	OTU_20345	92	9.95	0.00	0.01	0.19
Venenivibrio	OTU_10861	89	9.62	0.00	0.01	0.11
Venenivibrio	OTU_17939	80	8.65	0.00	0.01	0.17
Venenivibrio	OTU_24395	47	5.08	0.00	0.02	0.15
Venenivibrio	OTU_12231	97	10.49	0.00	0.01	0.06
Venenivibrio	OTU_28940	72	7.78	0.00	0.01	0.08
Venenivibrio	OTU_17857	70	7.57	0.00	0.01	0.06
Venenivibrio	OTU_22745	52	5.62	0.00	0.01	0.28
Venenivibrio	OTU_17081	56	6.05	0.00	0.01	0.06
Venenivibrio	OTU_9960	59	6.38	0.00	0.01	0.03
Venenivibrio	OTU_29407	35	3.78	0.00	0.01	0.15
Venenivibrio	OTU_16175	58	6.27	0.00	0.01	0.04
Venenivibrio	OTU_29917	51	5.51	0.00	0.01	0.04
Venenivibrio	OTU_2848	7	0.76	0.00	0.02	0.27
Venenivibrio	OTU_14464	31	3.35	0.00	0.01	0.26
Venenivibrio	OTU_26092	33	3.57	0.00	0.01	0.05
Venenivibrio	OTU_22422	34	3.68	0.00	0.01	0.06
Venenivibrio	OTU_22251	39	4.22	0.00	0.00	0.04
Venenivibrio	OTU_29919	31	3.35	0.00	0.01	0.05
Venenivibrio	OTU_23686	38	4.11	0.00	0.00	0.03
Venenivibrio	OTU_21081	28	3.03	0.00	0.00	0.04

SD: standard deviation

Table D.8 - Single linear regressions of *Venenivibrio* OTU number and read abundance against 46 metadata. The physicochemical variables were measured from each spring where *Venenivibrio* was present (post filtering; $n=467$). Results are ordered by descending R^2 value for OTU number, with no transformations performed on any parameter prior to analysis. The false discovery rate (FDR) Benjamini-Hochberg correction was applied to all p-values.

	OTU number			Read abundance		
	R^2 value	p-value	Corrected p-value	R^2 value	p-value	Corrected p-value
OTU number	n.a.	n.a.	n.a.	0.698	<0.001	<0.001
Read abundance	0.698	<0.001	<0.001	n.a.	n.a.	n.a.
ORP	0.052	<0.001	<0.001	0.093	<0.001	<0.001
TURB	0.026	<0.001	0.005	0.049	<0.001	<0.001
pH	0.019	0.003	0.025	0.009	0.044	0.211
COND	0.016	0.007	0.054	0.014	0.012	0.095
NO_2^-	0.015	0.008	0.057	0.011	0.024	0.130
NO_3^-	0.014	0.010	0.059	0.017	0.005	0.049
HCO_3^-	0.011	0.027	0.142	0.005	0.112	0.353
Cu	0.010	0.034	0.152	0.005	0.140	0.353
NH_4^+	0.010	0.035	0.152	0.005	0.133	0.353
Ag	0.007	0.076	0.288	0.000	0.850	0.912
Tl	0.007	0.078	0.288	0.005	0.131	0.353
U	0.006	0.090	0.289	0.003	0.260	0.446
SO_4^{2-}	0.006	0.090	0.289	0.005	0.113	0.353
Al	0.005	0.131	0.388	0.002	0.391	0.521
Na	0.005	0.141	0.388	0.000	0.937	0.957
S	0.005	0.146	0.388	0.003	0.221	0.442
Mn	0.004	0.158	0.400	0.006	0.105	0.353
Mo	0.003	0.267	0.603	0.002	0.321	0.482
Si	0.003	0.277	0.603	0.000	0.923	0.957
Fe^{2+}	0.003	0.280	0.603	0.003	0.232	0.442
Mg	0.002	0.289	0.603	0.004	0.186	0.429
Fe	0.002	0.310	0.603	0.003	0.249	0.442
Rb	0.002	0.325	0.603	0.003	0.239	0.442
TEMP	0.002	0.327	0.603	0.005	0.111	0.353
Cs	0.002	0.365	0.637	0.005	0.119	0.353
Ca	0.002	0.376	0.637	0.004	0.196	0.429
As	0.002	0.397	0.637	0.012	0.018	0.109
Ni	0.002	0.398	0.637	0.002	0.377	0.521
Br	0.001	0.439	0.637	0.000	0.704	0.785
Hg	0.001	0.444	0.637	0.001	0.563	0.693
Cl^-	0.001	0.447	0.637	0.001	0.435	0.564
Cr	0.001	0.452	0.637	0.002	0.314	0.482
H_2S	0.001	0.465	0.637	0.012	0.016	0.109
Ba	0.001	0.525	0.700	0.000	0.855	0.912
Se	0.001	0.547	0.710	0.000	0.674	0.770
V	0.001	0.574	0.726	0.002	0.357	0.519
Pb	0.001	0.612	0.753	0.000	0.982	0.982

PO_4^{3-}	0.000	0.694	0.833	0.003	0.271	0.449
Li	0.000	0.795	0.908	0.005	0.140	0.353
Sr	0.000	0.796	0.908	0.003	0.239	0.442
dO	0.000	0.813	0.908	0.001	0.492	0.621
K	0.000	0.840	0.914	0.002	0.312	0.482
B	0.000	0.857	0.914	0.004	0.188	0.429
Cd	0.000	0.906	0.928	0.000	0.647	0.757
Co	0.000	0.917	0.928	0.002	0.383	0.521
Zn	0.000	0.928	0.928	0.001	0.592	0.711

COND: conductivity, dO: dissolved oxygen, n.a.: not applicable, ORP: oxidation-reduction potential, TEMP: temperature, TURB: turbidity

Table D.9 - Multiple linear regression model of *Venenivibrio* OTU number against metadata. The metadata values were measured from each spring where *Venenivibrio* was present (post filtering; $n=467$). Two samples with missing HCO_3^- and H_2S data were removed from the analysis. An Akaike information criterion (AIC) was also applied to find the model of best fit. Results are ordered by ascending p-value. The R^2 value for the final model was 0.192 ($p<0.001$).

	Estimate	Std. Error	t-value	p-value	FDR corrected p-value
(Intercept)	32.120	2.720	11.810	<0.001	<0.001
Na	0.059	0.011	5.565	<0.001	<0.001
TURB	0.018	0.005	4.061	<0.001	<0.001
ORP	-0.026	0.007	-3.637	<0.001	0.002
NH_4^+	0.064	0.019	3.360	0.001	0.004
Sr	33.670	11.960	2.815	0.005	0.016
SO_4^{3-}	-0.013	0.005	-2.808	0.005	0.016
Cu	-24.250	9.079	-2.671	0.008	0.021
Li	-3.122	1.214	-2.572	0.010	0.023
Fe^{2+}	-0.374	0.146	-2.562	0.011	0.023
Rb	-13.250	5.811	-2.280	0.023	0.044
Cl^-	-0.016	0.007	-2.137	0.033	0.058
NO_3^-	-6.414	3.675	-1.745	0.082	0.123
dO	1.033	0.600	1.720	0.086	0.123
Ca	0.075	0.044	1.712	0.088	0.123
Pb	12.370	7.992	1.547	0.122	0.141
Ag	-2959	1913	-1.547	0.123	0.141
NO_2^-	-27.920	18.210	-1.533	0.126	0.141
S	0.020	0.013	1.519	0.129	0.141
Co	-1643	1094	-1.501	0.134	0.141
Se	401.500	275.400	1.458	0.146	0.146

dO: dissolved oxygen, FDR: false discovery rate, ORP: oxidation-reduction potential, TURB: turbidity

Table D.10 - Multiple linear regression model of *Venenivibrio* read abundance against metadata. The metadata values were measured from each spring where *Venenivibrio* was present (post filtering; n=467). Two samples with missing HCO_3 and H_2S data were removed from the analysis. An Akaike information criterion (AIC) was also applied to find the model of best fit. Results are ordered by ascending p-value. The R^2 value for the final model was 0.183 ($p<0.001$).

	Estimate	Std. Error	t-value	p-value	FDR corrected p-value
(Intercept)	1529.000	251.800	6.075	<0.001	<0.001
ORP	-3.799	0.752	-5.049	<0.001	<0.001
TURB	2.343	0.490	4.783	<0.001	<0.001
Li	-444.500	112.900	-3.939	<0.001	<0.001
Na	3.727	1.014	3.674	<0.001	0.001
Al	33.610	10.260	3.277	0.001	0.003
SO_4^{2-}	-1.428	0.451	-3.163	0.002	0.003
NH_4^+	5.705	2.386	2.391	0.017	0.030
Fe^{2+}	-34.410	15.560	-2.212	0.028	0.042
Cu	-2171	995.800	-2.180	0.030	0.042
NO_2^-	-3566	1772	-2.012	0.045	0.057
Sr	2328	1256	1.854	0.064	0.075
Ca	7.682	4.801	1.600	0.110	0.119
Co	-178500	118100	-1.511	0.131	0.131

FDR: false discovery rate, ORP: oxidation-reduction potential, TURB: turbidity

Table D.11 - Correlations between *Venenivibrio* diversity and environmental variables. Both OTU number and read abundance in each spring where *Venenivibrio* was found (post filtering; n=467) were correlated against 46 physicochemical parameters using Pearson's and Spearman's coefficient. Due to missing or aberrant data, only 455, 465 and 466 springs were used to analyse conductivity, HCO_3^- and H_2S respectively. The false discovery rate (FDR) Benjamini-Hochberg correction was applied to all p-values. Results are ordered by descending Pearson's coefficient for correlations with OTU number.

	OTU number				Read abundance			
	Pearson	Corrected p-value	Spearman	Corrected p-value	Pearson	Corrected p-value	Spearman	Corrected p-value
OTU number	n.a.	n.a.	n.a.	n.a.	0.674	<0.001	0.714	<0.001
Read abundance	0.674	<0.001	0.714	<0.001	n.a.	n.a.	n.a.	n.a.
TURB	0.061	<0.001	0.045	<0.001	0.164	<0.001	0.141	<0.001
H_2S	0.040	<0.001	0.042	<0.001	0.081	<0.001	0.129	<0.001
ORP	0.037	<0.001	0.033	<0.001	0.070	<0.001	0.052	<0.001
HCO_3^-	0.033	<0.001	0.044	<0.001	-0.002	0.567	0.017	<0.001
Ag	0.022	<0.001	0.031	<0.001	-0.002	0.626	0.019	<0.001
NH_4^+	0.015	<0.001	0.048	<0.001	0.040	<0.001	0.088	<0.001
Cu	-0.009	0.002	-0.009	0.007	-0.035	<0.001	-0.037	<0.001
Ba	-0.010	0.002	0.008	0.015	-0.011	<0.001	0.041	<0.001
pH	-0.010	0.001	0.017	<0.001	-0.094	<0.001	-0.077	<0.001
Na	-0.012	<0.001	0.006	0.079	-0.029	<0.001	-0.003	0.296

Hg	-0.013	<0.001	0.007	0.038	-0.020	<0.001	0.043	<0.001
Rb	-0.013	<0.001	-0.009	0.003	-0.021	<0.001	-0.023	<0.001
PO_4^{3-}	-0.015	<0.001	-0.013	<0.001	-0.037	<0.001	0.001	0.723
Zn	-0.016	<0.001	-0.010	0.002	-0.022	<0.001	-0.020	<0.001
Br	-0.016	<0.001	0.000	0.877	-0.032	<0.001	0.000	0.974
Cl^-	-0.016	<0.001	-0.017	<0.001	-0.025	<0.001	-0.032	<0.001
NO_3^-	-0.017	<0.001	-0.002	0.542	-0.072	<0.001	-0.069	<0.001
NO_2^-	-0.017	<0.001	-0.017	<0.001	-0.052	<0.001	-0.105	<0.001
Cd	-0.017	<0.001	0.014	<0.001	-0.020	<0.001	0.012	<0.001
Si	-0.018	<0.001	-0.006	0.041	-0.043	<0.001	-0.018	<0.001
K	-0.019	<0.001	-0.008	0.014	-0.027	<0.001	-0.008	0.006
Ni	-0.019	<0.001	-0.014	<0.001	-0.025	<0.001	-0.039	<0.001
Tl	-0.022	<0.001	-0.038	<0.001	-0.039	<0.001	-0.067	<0.001
Se	-0.024	<0.001	-0.013	<0.001	-0.014	<0.001	0.000	0.995
Fe^{2+}	-0.025	<0.001	-0.022	<0.001	-0.038	<0.001	-0.066	<0.001
Co	-0.026	<0.001	-0.020	<0.001	-0.035	<0.001	-0.041	<0.001
dO	-0.026	<0.001	0.001	0.750	-0.015	<0.001	0.030	<0.001
Mg	-0.026	<0.001	-0.002	0.526	-0.043	<0.001	-0.044	<0.001
Mn	-0.027	<0.001	-0.004	0.203	-0.052	<0.001	-0.039	<0.001
Cr	-0.028	<0.001	-0.015	<0.001	-0.035	<0.001	-0.007	0.022
Fe	-0.028	<0.001	-0.030	<0.001	-0.038	<0.001	-0.068	<0.001
V	-0.029	<0.001	-0.023	<0.001	-0.030	<0.001	-0.063	<0.001
B	-0.030	<0.001	0.007	0.034	-0.056	<0.001	-0.017	<0.001
Li	-0.030	<0.001	-0.003	0.299	-0.056	<0.001	-0.022	<0.001
Cs	-0.030	<0.001	-0.009	0.007	-0.045	<0.001	-0.026	<0.001
Ca	-0.031	<0.001	-0.017	<0.001	-0.043	<0.001	-0.034	<0.001
TEMP	-0.032	<0.001	-0.011	0.001	-0.089	<0.001	-0.072	<0.001
COND	-0.040	<0.001	-0.021	<0.001	-0.053	<0.001	-0.080	<0.001
As	-0.041	<0.001	0.001	0.877	-0.075	<0.001	-0.027	<0.001
Al	-0.042	<0.001	-0.045	<0.001	-0.016	<0.001	-0.109	<0.001
Sr	-0.043	<0.001	-0.020	<0.001	-0.040	<0.001	-0.008	0.006
U	-0.047	<0.001	-0.036	<0.001	-0.026	<0.001	-0.083	<0.001
S	-0.051	<0.001	-0.028	<0.001	-0.043	<0.001	-0.041	<0.001
SO_4^{2-}	-0.052	<0.001	-0.029	<0.001	-0.051	<0.001	-0.070	<0.001
Pb	-0.006	0.050	0.011	0.001	-0.001	0.650	0.025	<0.001
Mo	-0.006	0.052	0.001	0.774	-0.025	<0.001	-0.023	<0.001

COND: conductivity; dO: dissolved oxygen; n.a.: not applicable, ORP: oxidation-reduction potential, TEMP: temperature, TURB: turbidity

Table D.12 - Predicted function and annotation source of the *V. stagnispumantis* CP.B2^T genome. Genome assembly of the type strain was deposited in the Genomes Online Database (GOLD Analysis ID Ga0311387), with annotation performed by the Integrated Microbial Genomes platform v4.16.4 (IMG Genome ID 2799112217). Genes are ordered by predicted pathway or metabolism.

Gene ID	Locus tag	Gene	Predicted function	Annotation
Glycolysis/Gluconeogenesis/EMP				
2799181671	Ga0311387_10062	*glk*	Glucokinase (GK)	K00845, EC:2.7.1.2
2799181405	Ga0311387_100356	*pgi*	Glucose-6-phosphate isomerase (GPI)	K01810, EC:5.3.1.9
2799181071	Ga0311387_10018	*pfkA*	6-phosphofructokinase (PFK)	K00850, EC:2.7.1.11
2799181965	Ga0311387_1008101	*fbp*	Fructose-1,6-biphoshatase (FBP)	K03841, EC:3.1.3.11
2799182515	Ga0311387_101718	*fbaB*	Fructose-biphosphate aldolase (FBA)	K11645, EC:4.1.2.13
2799182173	Ga0311387_101119	*tpiA*	Triosephosphate isomerase (TPI)	K01803, EC:5.3.1.1
2799182067	Ga0311387_10101	*gapA*	Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)	K00134, EC:1.2.1.12
2799181708	Ga0311387_100639	*gapN*	Glyceraldehyde-3-phosphate dehydrogenase (NADP+)	K00131, EC:1.2.1.9
2799181324	Ga0311387_100294	*pgk*	Phosphoglycerate kinase (PGK)	K00927, EC:2.7.2.3
2799182172	Ga0311387_101118	*gpmA*	Phosphoglycerate mutase (PGAM)	K01834, EC:5.4.2.11
2799181974	Ga0311387_10099	*gpmB*	Probable phosphoglycerate mutase	K15634, EC:5.4.2.11
2799182594	Ga0311387_102111	*gpmB*	Probable phosphoglycerate mutase	K15634, EC:5.4.2.11
2799181329	Ga0311387_100299	*eno*	Enolase	K01689, EC:4.2.1.11
2799181825	Ga0311387_100762	*pyk*	Pyruvate kinase (PK)	K00873, EC:2.7.1.40
2799181317	Ga0311387_100287	*porA*	Pyruvate ferredoxin oxidoreductase alpha subunit	K00169, EC:1.2.7.1
2799181316	Ga0311387_100286	*porB*	Pyruvate ferredoxin oxidoreductase beta subunit	K00170, EC:1.2.7.1
2799181318	Ga0311387_100288	*porD*	Pyruvate ferredoxin oxidoreductase delta subunit	K00171, EC:1.2.7.1
2799181319	Ga0311387_100289	*porG*	Pyruvate ferredoxin oxidoreductase gamma subunit	K00172, EC:1.2.7.1
2799182580	Ga0311387_102017	*korA, oorA*	2-oxoglutarate ferredoxin oxidoreductase subunit alpha	K00174, EC:1.2.7.3
2799182581	Ga0311387_102018	*korB, oorB*	2-oxoglutarate ferredoxin oxidoreductase subunit beta	K00175, EC:1.2.7.3
2799181752	Ga0311387_100683	*acs*	Acetyl-CoA synthetase (ACSS1_2)	K01895, EC:6.2.1.1
2799182097	Ga0311387_101031	*acs*	Acetyl-CoA synthetase (ACSS1_2)	K01895, EC:6.2.1.1
2799181618	Ga0311387_100538	*lpd*	Dihydrolipoamide dehydrogenase (DLD)	K00382, EC:1.8.1.4
Pentose phosphate				
2799181349	Ga0311387_1002119	*rpe*	Ribulose-phosphate 3-epimerase (RPE)	K01783, EC:5.1.3.1
2799181409	Ga0311387_100360	*rpiB*	Ribose-5-phosphate isomerase	K01808, EC:5.3.1.6
2799182370	Ga0311387_101419	*tktA, tktB*	Transketolase	K00615, EC:2.2.1.1
2799182371	Ga0311387_101420	*tktA, tktB*	Transketolase	K00615, EC:2.2.1.1
2799181940	Ga0311387_100876	*talA, talB*	Transaldolase	K00616, EC:2.2.1.2

Gene ID	Locus tag	Gene	Predicted function	Annotation
rTCA cycle				
2799181709	Ga0311387_100640	*aclA*	ATP-citrate lyase alpha subunit	K15230, EC:2.3.3.8
2799181710	Ga0311387_100641	*aclB*	ATP-citrate lyase beta subunit	K15231, EC:2.3.3.8
2799182575	Ga0311387_102012	*mdh*	Malate dehydrogenase	K00024, EC:1.1.1.37
2799182576	Ga0311387_102013	*fumA*	Fumarate hydratase subunit alpha	K01677, EC:4.2.1.2
2799182577	Ga0311387_102014	*fumB*	Fumarate hydratase subunit beta	K01678, EC:4.2.1.2
2799182580	Ga0311387_102017	*korA, oorA*	2-oxoglutarate ferredoxin oxidoreductase subunit alpha	K00174, EC:1.2.7.3
2799182581	Ga0311387_102018	*korB, oorB*	2-oxoglutarate ferredoxin oxidoreductase subunit beta	K00175, EC:1.2.7.3
2799182537	Ga0311387_101810	*sdhA, frdA*	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	K00239, EC:1.3.5.1/4
2799181805	Ga0311387_100742	*sdhB, frdB*	Succinate dehydrogenase/fumarate reductase, iron-sulfur subunit	K00240, EC:1.3.5.1/4
2799181845	Ga0311387_100782	*sdhB, frdB*	Succinate dehydrogenase/fumarate reductase, iron-sulfur subunit	K00240, EC:1.3.5.1/4
2799181846	Ga0311387_100783	*sdhC, frdC*	Succinate dehydrogenase/fumarate reductase, cytochrome b subunit	K00241
2799182579	Ga0311387_102016	*sucD*	Succinyl-CoA synthetase alpha subunit	K01902, EC:6.2.1.5
2799182578	Ga0311387_102015	*sucC*	Succinyl-CoA synthetase beta subunit	K01903, EC:6.2.1.5
2799181711	Ga0311387_100642	*icd*	Isocitrate dehydrogenase (IDH)	K00031, EC:1.1.1.42
2799181759	Ga0311387_100690	*acnB*	Aconitate hydratase	K01682, EC:4.2.1.3
2799181317	Ga0311387_100287	*porA*	Pyruvate ferredoxin oxidoreductase alpha subunit	K00169, EC:1.2.7.1
2799181316	Ga0311387_100286	*porB*	Pyruvate ferredoxin oxidoreductase beta subunit	K00170, EC:1.2.7.1
2799181318	Ga0311387_100288	*porD*	Pyruvate ferredoxin oxidoreductase delta subunit	K00171, EC:1.2.7.1
2799181319	Ga0311387_100289	*porG*	Pyruvate ferredoxin oxidoreductase gamma subunit	K00172, EC:1.2.7.1
2799181822	Ga0311387_100759	*ppsA*	Pyruvate, water dikinase	K01007, EC:2.7.9.2
2799181328	Ga0311387_100298	*pycA*	Pyruvate carboxylase subunit A	K01959, EC:6.4.1.1
2799181726	Ga0311387_100657	*pycB*	Pyruvate carboxylase subunit B	K01960, EC:6.4.1.1
2799181332	Ga0311387_1002102	*ccl*	Citryl-CoA lyase	K15234, EC:4.1.3.34
Complex I				
2799181192	Ga0311387_1001129	*nuoA*	NADH-quinone oxidoreductase subunit A	K00330, EC:7.1.1.2
2799181193	Ga0311387_1001130	*nuoB*	NADH-quinone oxidoreductase subunit B	K00331, EC:7.1.1.2
2799181194	Ga0311387_1001131	*nuoC/D*	NADH-quinone oxidoreductase subunit C/D	K13378, EC:7.1.1.2
2799181181	Ga0311387_1001118	*nuoC/D*	NADH-quinone oxidoreductase subunit C/D	K13378, EC:7.1.1.2
2799181943	Ga0311387_100879	*nuoE*	NADH-quinone oxidoreductase subunit E	K00334, EC:7.1.1.2
2799181944	Ga0311387_100880	*nuoF*	NADH-quinone oxidoreductase subunit F	K00335, EC:7.1.1.2
2799182303	Ga0311387_10133	*nuoG*	NADH-quinone oxidoreductase subunit G	K00336, EC:7.1.1.2
2799181195	Ga0311387_1001132	*nuoH*	NADH-quinone oxidoreductase subunit H	K00337, EC:7.1.1.2
2799181183	Ga0311387_1001120	*nuoI*	NADH-quinone oxidoreductase subunit I	K00338, EC:7.1.1.2

Gene ID	Locus tag	Gene	Predicted function	Annotation
2799181196	Ga0311387_1001133	*nuoI*	NADH-quinone oxidoreductase subunit I	K00338, EC:7.1.1.2
2799181197	Ga0311387_1001134	*nuoJ*	NADH-quinone oxidoreductase subunit J	K00339, EC:7.1.1.2
2799181198	Ga0311387_1001135	*nuoK*	NADH-quinone oxidoreductase subunit K	K00340, EC:7.1.1.2
2799181199	Ga0311387_1001136	*nuoL*	NADH-quinone oxidoreductase subunit L	K00341, EC:7.1.1.2
2799181200	Ga0311387_1001137	*nuoM*	NADH-quinone oxidoreductase subunit M	K00342, EC:7.1.1.2
2799181201	Ga0311387_1001138	*nuoN*	NADH-quinone oxidoreductase subunit N	K00343, EC:7.1.1.2
Complex II				
2799182537	Ga0311387_101810	*sdhA, frdA*	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	K00239, EC:1.3.5.1/4
2799181805	Ga0311387_100742	*sdhB, frdB*	Succinate dehydrogenase/fumarate reductase, iron-sulfur subunit	K00240, EC:1.3.5.1/4
2799181845	Ga0311387_100782	*sdhB, frdB*	Succinate dehydrogenase/fumarate reductase, iron-sulfur subunit	K00240, EC:1.3.5.1/4
2799181846	Ga0311387_100783	*sdhC, frdC*	Succinate dehydrogenase/fumarate reductase, cytochrome b subunit	K00241
Complex III (Cytochrome bc1)				
2799182378	Ga0311387_101427	*petA*	Ubiquinol-cytochrome c reductase iron-sulfur subunit	K00411, EC:7.1.1.8
2799182379	Ga0311387_101428	*petB*	Ubiquinol-cytochrome c reductase cytochrome b subunit	K00412
2799182380	Ga0311387_101429	*petC*	Ubiquinol-cytochrome c reductase cytochrome c1 subunit	K00413
Complex IV (Cytochrome bd)				
2799182465	Ga0311387_10161	*cydA*	Cytochrome bd ubiquinol oxidase subunit I	K00425, EC:7.1.1.7
2799182466	Ga0311387_10162	*cydB*	Cytochrome bd ubiquinol oxidase subunit II	K00426, EC:7.1.1.7
Complex V (ATPases)				
2799181829	Ga0311387_100766	*atpA*	F-type H+/Na+-transporting ATPase subunit alpha (ATPF1A)	K02111, EC:7.1.2.2
2799182662	Ga0311387_102511	*atpA*	F-type H+/Na+-transporting ATPase subunit alpha (ATPF1A)	K02111, EC:7.1.2.2
2799181831	Ga0311387_100768	*atpD*	F-type H+/Na+-transporting ATPase subunit beta (ATPF1B)	K02112, EC:7.1.2.2
2799182664	Ga0311387_102513	*atpD*	F-type H+/Na+-transporting ATPase subunit beta (ATPF1B)	K02112, EC:7.1.2.2
2799182663	Ga0311387_102512	*atpG*	F-type H+/Na+-transporting ATPase subunit gamma (ATPF1G)	K02115
2799181830	Ga0311387_100767	*atpG*	F-type H+/Na+-transporting ATPase subunit gamma (ATPF1G)	K02115
2799182661	Ga0311387_102510	*atpH*	F-type H+/Na+-transporting ATPase subunit delta (ATPF1D)	K02113
2799181832	Ga0311387_100769	*atpC*	F-type H+/Na+-transporting ATPase subunit epsilon (ATPF1E)	K02114
2799182170	Ga0311387_101116	*atpC*	F-type H+/Na+-transporting ATPase subunit epsilon (ATPF1E)	K02114
2799181129	Ga0311387_100166	*atpB*	F-type H+/Na+-transporting ATPase subunit a (ATPF0A)	K02108
2799181826	Ga0311387_100763	*atpB*	F-type H+/Na+-transporting ATPase subunit a (ATPF0A)	K02108
2799182659	Ga0311387_10258	*atpF*	F-type H+/Na+-transporting ATPase subunit b (ATPF0B)	K02109
2799182660	Ga0311387_10259	*atpF*	F-type H+/Na+-transporting ATPase subunit b (ATPF0B)	K02109
2799181828	Ga0311387_100765	*atpF*	F-type H+/Na+-transporting ATPase subunit b (ATPF0B)	K02109
2799181131	Ga0311387_100168	*atpE*	F-type H+/Na+-transporting ATPase subunit c (ATPF0C)	K02110

Gene ID	Locus tag	Gene	Predicted function	Annotation
2799181827	Ga0311387_100764	*atpE*	F-type H+/Na+-transporting ATPase subunit c (ATPF0C)	K02110
2799182707	Ga0311387_102814	*ppa*	Inorganic pyrophosphatase	K01507, EC:3.6.1.1
Cytochrome c				
2799181979	Ga0311387_100914	*soxD*	Cytochrome c (CYC)	K08738
2799181251	Ga0311387_100221	cytochrom_C_asm	Cytochrome c assembly protein	pfam01578
2799182595	Ga0311387_102112	*ccsB*	Cytochrome c-type biogenesis protein	TIGR03144
2799182583	Ga0311387_102020	*dsbD_2*	Cytochrome c biogenesis protein	pfam13386
2799181069	Ga0311387_10016	*resB, css1*	Cytochrome c biogenesis protein	K07399
2799182130	Ga0311387_101064	cytochrom_C_2	Cytochrome c'	pfam01322
2799182376	Ga0311387_101425	*ccdA*	Cytochrome c-type biogenesis protein	K06196
2799181456	Ga0311387_1003107	c554	Cytochrome c554/c'-like protein	pfam13435
Hydrogen				
2799181768	Ga0311387_10075	*hynB, hydB*	Quinone-reactive Ni,Fe-hydrogenase large subunit (Group 1b)	HydDB, K05922, EC:1.12.5.1
2799181769	Ga0311387_10076	*hynA, hydA*	Quinone-reactive Ni,Fe-hydrogenase small subunit (Group 1b)	HydDB, K05927, EC:1.12.5.1
2799181774	Ga0311387_100711	*huaL, hyaB*	Ni,Fe-hydrogenase I large subunit (Group 2d)	HydDB, COG0374, EC:1.12.99.6
2799181775	Ga0311387_100712	*huaS, hyaA*	Ni,Fe-hydrogenase I small subunit (Group 2d)	HydDB, COG1740, EC:1.12.99.6
2799181771	Ga0311387_10078	*hypE*	Hydrogenase expression/formation protein	K04655
2799181764	Ga0311387_10071	*hypF*	Hydrogenase maturation protein	K04656
2799181765	Ga0311387_10072	*hypF*	Hydrogenase maturation protein	K04656
2799181766	Ga0311387_10073	*hyaD, hybD*	Hydrogenase maturation protease	K03605
2799181773	Ga0311387_100710	*hypC*	Hydrogenase expression/formation protein	K04653
2799181772	Ga0311387_10079	*hypD*	Hydrogenase expression/formation protein	K04654
2799181780	Ga0311387_100717	*hypB*	Hydrogenase nickel incorporation protein	K04652
2799181770	Ga0311387_10077	*hypA, hybF*	Hydrogenase nickel incorporation protein	K04651
2799182199	Ga0311387_101145	*hyfE*	Hydrogenase-4 component E	K12140
2799182200	Ga0311387_101146	*hyfF*	Hydrogenase-4 component F	K12141
2799182201	Ga0311387_101147	*hycE*	Formate hydrogenlyase subunit 5	K15830
2799182202	Ga0311387_101148	*hycG*	Formate hydrogenlyase subunit 7	K15832
2799181767	Ga0311387_10074	*hyaC*	Ni,Fe-hydrogenase 1 B-type cytochrome subunit	K03620
Nitrogen				
2799181706	Ga0311387_100637	*glnA*	Glutamine synthetase	K01915, EC:6.3.1.2
2799181707	Ga0311387_100638	*glnK*	Nitrogen regulatory protein P-II	COG0347
2799182556	Ga0311387_101910	*gltB*	Glutamate synthase (NADPH) large chain	K00265, EC:1.4.1.13
2799181562	Ga0311387_1004101	*gltD*	Glutamate synthase (NADPH) small chain	K00266, EC:1.4.1.13

Gene ID	Locus tag	Gene	Predicted function	Annotation
2799182447	Ga0311387_101540	*ncd2, npd*	Nitronate monooxygenase	K00459, EC:1.13.12.16
Sulfur				
2799181720	Ga0311387_100651	*sqr*	Sulfide:quinone oxidoreductase	K17218, EC:1.8.5.4
2799182727	Ga0311387_102915	*sqr*	Sulfide:quinone oxidoreductase	K17218, EC:1.8.5.4
2799181402	Ga0311387_100353	*tusA*	TusA-related sulfurtransferase	COG0425
2799182283	Ga0311387_101246	*tusA*	TusA-related sulfurtransferase	COG0425
2799182308	Ga0311387_10138	*tusA*	TusA-related sulfurtransferase	COG0425
2799181400	Ga0311387_100351	*sseA*	Thiosulfate/3-mercaptopyruvate sulfurtransferase (TST/MPST)	K01011, EC:2.8.1.1/2
2799182138	Ga0311387_101072	*pspE*	Rhodanese-related sulfurtransferase	COG0607
2799181481	Ga0311387_100420	*cysM*	Cysteine synthase B	K12339, EC:2.5.1.144
2799182712	Ga0311387_102819	*nrnA*	Bifunctional oligoribonuclease and PAP phosphatase	K06881, EC:3.1.3.7, EC:3.1.13.3
2799182487	Ga0311387_101623	*thiS*	Sulfur carrier protein	K03154
2799182658	Ga0311387_10257	*thiS*	Sulfur carrier protein	K03154
2799181338	Ga0311387_1002108	*moaD, cysO*	Molybdopterin synthase sulfur carrier subunit	K03636
2799181818	Ga0311387_100755	*moaD, cysO*	Molybdopterin synthase sulfur carrier subunit	K03636
2799181149	Ga0311387_100186	*moaE*	Molybdopterin synthase catalytic subunit	K03635, EC:2.8.1.12
2799182282	Ga0311387_101245	*moeB,* MOCS3	Adenylyltransferase and sulfurtransferase	K11996, EC:2.8.1.11
2799182119	Ga0311387_101053	*iscS*	Cysteine desulfurase	K04487, EC:2.8.1.7
2799182286	Ga0311387_101249	*tusB, dsrH*	tRNA 2-thiouridine synthesizing protein B	K07237
2799182285	Ga0311387_101248	*tusC, dsrF*	tRNA 2-thiouridine synthesizing protein C	K07236
2799182284	Ga0311387_101247	*tusD, dsrE*	tRNA 2-thiouridine synthesizing protein D	K07235
2799181663	Ga0311387_100583	*mnmA, trmU*	tRNA-uridine 2-sulfurtransferase	K00566, EC:2.8.1.13
2799182307	Ga0311387_10137	*mnmA, trmU*	tRNA-uridine 2-sulfurtransferase	K00566, EC:2.8.1.13
Arsenic resistance				
2799181352	Ga0311387_10033	*arsB*	Arsenical pump membrane protein	K03893
2799181351	Ga0311387_10032	*arsC*	Arsenate reductase (thioredoxin)	K03741, EC:1.20.4.4
2799181354	Ga0311387_10035	*arsR*	ArsR family transcriptional regulator	K03892
Stress response				
2799182192	Ga0311387_101138	*aguA*	Agmatine deiminase	K10536, EC:3.5.3.12
2799182193	Ga0311387_101139	*aguB*	N-carbamoylputrescine amidase	K12251, EC:3.5.1.53
2799181397	Ga0311387_100348	*speE*	Spermidine synthase	K00797, EC:2.5.1.16
2799182076	Ga0311387_101010	*glcD, dld*	Glycolate oxidase/D-lactate dehydrogenase	K00104, K03777, EC:1.1.3.15, EC:1.1.5.12
2799182366	Ga0311387_101415	*ldhA*	D-lactate dehydrogenase	K03778, EC:1.1.1.28
2799181752	Ga0311387_100683	*acs*	Acetyl-coa synthetase (ACSS1_2)	K01895, EC:6.2.1.1

Gene ID	Locus tag	Gene	Predicted function	Annotation
2799182097	Ga0311387_101031	*acs*	Acetyl-coA synthetase (ACSS1_2)	K01895, EC:6.2.1.1
2799182184	Ga0311387_101130	*mtaD*	Adenosine/cysteine deaminase	K12960, EC:3.5.4.31/28
2799181124	Ga0311387_100161	*dcd*	Deoxycytidine-3P (dCTP) deaminase	K01494, EC:3.5.4.13
2799182720	Ga0311387_10298	*ribD*	Diaminohydroxyphosphoribosylaminopyrimidine deaminase	K11752, EC:3.5.4.26
Pilus assembly				
2799181382	Ga0311387_100333	*tadD*	Flp pilus assembly protein	COG5010
2799182368	Ga0311387_101417	*tadD*	Flp pilus assembly protein	COG5010
2799181285	Ga0311387_100255	*gspF, pilC*	General secretion pathway protein F/Type IV pilus assembly protein	K02455, K02653
2799182051	Ga0311387_100986	*gspF, pilC*	General secretion pathway protein F/Type IV pilus assembly protein	K02455, K02653
2799182701	Ga0311387_10288	*fimC*	P pilus assembly protein (chaperone PapD)	COG3121
2799181615	Ga0311387_100535	*pilO*	Pilus assembly protein	pfam04350
2799182037	Ga0311387_100972	*pilA*	Type IV pilus assembly protein	K02650
2799181286	Ga0311387_100256	*pilB*	Type IV pilus assembly protein	K02652
2799181660	Ga0311387_100580	*pilB*	Type IV pilus assembly protein	K02652
2799181662	Ga0311387_100582	*pilC*	Type IV pilus assembly protein	K02653
2799181613	Ga0311387_100533	*pilM*	Type IV pilus assembly protein	K02662
2799181617	Ga0311387_100537	*pilQ*	Type IV pilus assembly protein	K02666
2799182042	Ga0311387_100977	*pilV*	Type IV pilus assembly protein	K02671
2799182041	Ga0311387_100976	*pilW*	Type IV pilus assembly protein	K02672
2799182044	Ga0311387_100979	*pilY1*	Type IV pilus assembly protein	K02674
2799181858	Ga0311387_100795	*pilW*	Type IV pilus biogenesis/stability protein	TIGR02521
Flagellar assembly				
2799181419	Ga0311387_100370	*fliD*	Flagellar hook-associated protein 2	K02407
2799181417	Ga0311387_100368	*fliC*	Flagellin	K02406
2799182137	Ga0311387_101071	*flgL*	Flagellar hook-associated protein 3	K02397
2799181380	Ga0311387_100331	*flgK*	Flagellar hook-associated protein 1	K02396
2799181302	Ga0311387_100272	*flgD*	Flagellar basal-body rod modification protein	K02389
2799181387	Ga0311387_100338	*flgG*	Flagellar basal-body rod protein	K02392
2799181486	Ga0311387_100425	*flgH*	Flagellar L-ring protein precursor	K02393
2799181487	Ga0311387_100426	*flgI*	Flagellar P-ring protein precursor	K02394
2799181306	Ga0311387_100276	*fliE*	Flagellar hook-basal body complex protein	K02408
2799181308	Ga0311387_100278	*flgB*	Flagellar basal-body rod protein	K02387
2799181307	Ga0311387_100277	*flgC*	Flagellar basal-body rod protein	K02388
2799181305	Ga0311387_100275	*fliF*	Flagellar M-ring protein	K02409

Gene ID	Locus tag	Gene	Predicted function	Annotation
2799181383	Ga0311387_100334	*fliG*	Flagellar motor switch protein	K02410
2799181384	Ga0311387_100335	*fliM*	Flagellar motor switch protein	K02416
2799181299	Ga0311387_100269	*fliN*	Flagellar motor switch protein	K02417
2799181385	Ga0311387_100336	*fliN*	Flagellar motor switch protein	K02417
2799181489	Ga0311387_100428	*flhA*	Flagellar biosynthesis protein	K02400
2799181729	Ga0311387_100660	*flhB*	Flagellar biosynthetic protein	K02401
2799181303	Ga0311387_100273	*fliI*	Flagellum-specific ATP synthase	K02412, EC:7.4.2.8
2799181296	Ga0311387_100266	*fliQ*	Flagellar biosynthetic protein	K02420
2799181297	Ga0311387_100267	*fliP*	Flagellar biosynthetic protein	K02419
2799181730	Ga0311387_100661	*fliR*	Flagellar biosynthetic protein	K02421
2799181420	Ga0311387_100371	*fliS*	Flagellar protein	K02422
Chemotaxis				
2799182624	Ga0311387_10237	*mcp*	Methyl-accepting chemotaxis protein	K03406
2799182622	Ga0311387_10235	*cheA*	Two-component system chemotaxis sensor kinase	K03407, EC:2.7.13.3
2799182621	Ga0311387_10234	*cheW*	Purine-binding chemotaxis protein	K03408
2799182623	Ga0311387_10236	*cheV*	Two-component system chemotaxis response regulator	K03415
2799182619	Ga0311387_10232	*cheY*	Two-component system chemotaxis response regulator	K03413
2799182620	Ga0311387_10233	*cheZ*	Chemotaxis protein	K03414
2799181256	Ga0311387_100226	*motA*	Chemotaxis protein	K02556
2799181255	Ga0311387_100225	*motB*	Chemotaxis protein	K02557
Transporters				
2799181479	Ga0311387_100418	*kefB*	K+ transport system	COG0475
2799181579	Ga0311387_1004118	*kefB*	K+ transport system	COG0475
2799181834	Ga0311387_100771	*nhaP*	Na+ (or K+)/H+ antiporter	COG0025
2799181840	Ga0311387_100777	*trkA*	K+ uptake protein	K03499
2799181841	Ga0311387_100778	*trkH*	K+ uptake protein	K03498
2799181839	Ga0311387_100776	*kup*	K+ uptake protein	K03549
2799182601	Ga0311387_10223	*trkA*	K+/H+ antiporter	K07228, COG0490
2799182600	Ga0311387_10222	TC.KEF	Monovalent cation/H+ antiporter-2 (cpa2 family)	K03455
2799181275	Ga0311387_100245	*trkA*	Voltage-gated potassium channel	K10716
2799181518	Ga0311387_100457	*trkA*	Voltage-gated potassium channel	K10716
2799182094	Ga0311387_101028	*actP*	Cation/acetate symporter	K14393
2799181475	Ga0311387_100414	*putP*	Na+/proline symporter	COG0591, K03307
2799182197	Ga0311387_101143	*hyfB*	Na+/H+ antiporter/hydrogenlyase-4 component B	COG0651, K12137

Gene ID	Locus tag	Gene	Predicted function	Annotation
2799181799	Ga0311387_100736	*ccc1*	Fe2+/Mn2+ transporter (VIT family)	COG1814, K22736
2799181505	Ga0311387_100444	*yrbG*	Cation/H+ antiporter	K07301
2799181395	Ga0311387_100346	*amt*	Ammonium transporter	K03320
2799182776	Ga0311387_10382	*mgtE*	Magnesium transporter	K06213
2799182779	Ga0311387_10385	*corA*	Magnesium transporter	K03284
2799181101	Ga0311387_100138	*fieF*	Divalent metal cation (Fe/Co/Zn/Cd) transporter	COG0053
2799181454	Ga0311387_1003105	TC.APA	Amino acid/polyamine antiporter (APA family)	K03294
2799181390	Ga0311387_100341	*proP*	Proline/betaine MFS transporter (MHS family)	K03762
2799181236	Ga0311387_10026	*acrA*	Multidrug efflux pump	COG0845
2799181602	Ga0311387_100522	*acrA*	Multidrug efflux pump	COG0845
2799182256	Ga0311387_101219	*acrA*	Multidrug efflux pump	COG0845
2799181722	Ga0311387_100653	*emrA*	Multidrug efflux pump	COG1566
2799181597	Ga0311387_100517	*emrA*	multidrug efflux pump/membrane fusion protein	COG1566, K03543
2799182355	Ga0311387_10144	*emrA*	Multidrug efflux pump/secretion protein	COG1566, K01993
2799181237	Ga0311387_10027	*acrB*	Multidrug efflux pump subunit	COG0841
2799181601	Ga0311387_100521	*acrB*	Multidrug efflux pump subunit	COG0841, K03296
2799181721	Ga0311387_100652	*acrB*	Multidrug efflux pump subunit	COG0841
2799182255	Ga0311387_101218	*acrB*	Multidrug efflux pump subunit	COG0841, K15726
2799182110	Ga0311387_101044	*citT*	Sodium-dependent dicarboxylate transporter	K14445, COG0471
2799182262	Ga0311387_101225	UMF1	Major Facilitator Superfamily (MFS) transporter	K06902
2799181532	Ga0311387_100471	*modE*	Molybdate transport system regulatory protein	K02019
2799182470	Ga0311387_10166	*czcB*	Cobalt-zinc-cadmium efflux system membrane fusion protein	K15727
2799182471	Ga0311387_10167	*czcA*	Cobalt-zinc-cadmium resistance protein	K15726
2799182469	Ga0311387_10165	*czcC*	Cobalt-zinc-cadmium efflux system outer membrane protein	K15725
Nutrient ABC transporters				
2799181207	Ga0311387_1001144	*pstS*	Phosphate transport system substrate-binding protein	K02040
2799181208	Ga0311387_1001145	*pstC*	Phosphate transport system permease protein	K02037
2799181209	Ga0311387_1001146	*pstA*	Phosphate transport system permease protein	K02038
2799181210	Ga0311387_1001147	*pstB*	Phosphate transport system ATP-binding protein	K02036, EC:7.3.2.1
2799181233	Ga0311387_10023	*mlaD, linM*	Phospholipid/cholesterol/gamma-HCH substrate-binding protein	K02067
2799182585	Ga0311387_10212	*mlaD, linM*	Phospholipid/cholesterol/gamma-HCH substrate-binding protein	K02067
2799182288	Ga0311387_101251	*mlaE, linK*	Phospholipid/cholesterol/gamma-HCH permase protein	K02066
2799182587	Ga0311387_10214	*mlaE, linK*	Phospholipid/cholesterol/gamma-HCH permase protein	K02066
2799181075	Ga0311387_100112	*mlaF, linL*	Phospholipid/cholesterol/gamma-HCH ATP-binding protein	K02065

Gene ID	Locus tag	Gene	Predicted function	Annotation
2799182586	Ga0311387_10213	*mlaF, linL*	Phospholipid/cholesterol/gamma-HCH ATP-binding protein	K02065
2799181913	Ga0311387_100849	*rfbA*	Lipopolysaccharide transport system permease protein	K09690
2799181914	Ga0311387_100850	*rfbB*	Lipopolysaccharide transport system ATP-binding protein	K09691
2799182571	Ga0311387_10208	*lolC/E*	Lipoprotein-releasing system permease protein	K09808
2799181406	Ga0311387_100357	*lolD*	Lipoprotein-releasing system ATP-binding protein	K09810
2799182438	Ga0311387_101531	*lptF*	Lipopolysaccharide export system permease protein	K07091
2799181516	Ga0311387_100455	*lptB*	Lipopolysaccharide export system ATP-binding protein	K06861, EC:7.5.2.5
2799182522	Ga0311387_101725	*lptC*	Lipopolysaccharide ABC transporter	pfam06835
2799181086	Ga0311387_100123	*msbA*	ATP-binding cassette (subfamily B)	K11085, EC:7.5.2.6
2799181230	Ga0311387_1001167	ABC-2	ABC2-membrane transporter	pfam01061
2799182352	Ga0311387_10141	ABC-2	ABC2-membrane-3 transporter	pfam12698
2799182353	Ga0311387_10142	ABC-2.P	ABC2 type transport system permease subunit	K01992
Metal ABC transporters				
2799182157	Ga0311387_10113	*znuA*	Zinc transport system substrate-binding protein	K09815
2799182088	Ga0311387_101022	*znuB*	Zinc transport system permease protein	K09816
2799182086	Ga0311387_101020	*znuC*	Zinc transport system ATP-binding protein	K09817, EC:7.2.2.20
2799181783	Ga0311387_100720	*cbiM*	Cobalt/nickel transport system permease protein	K02007
2799181782	Ga0311387_100719	*cbiQ*	Cobalt/nickel transport system permease protein	K02008
2799181781	Ga0311387_100718	*cbiO*	Cobalt/nickel transport system ATP-binding protein	K02006
2799181786	Ga0311387_100723	*cbiK*	Nickel transport protein	K10094
2799181785	Ga0311387_100722	*cbiL*	Nickel transport protein	K16915
2799181248	Ga0311387_100218	*oppD, ddpD*	Oligopeptide/nickel transport system ATP-binding protein	K15583, K02031
2799181366	Ga0311387_100317	*copB*	Cu2+ exporting ATPase (P type)	K01533, EC:7.2.2.9
2799181126	Ga0311387_100163	*mgtA, mgtB*	Mg2+ importing ATPase (P type)	K01531, EC:7.2.2.14
Glycogen synthesis				
2799182335	Ga0311387_101335	*glgA*	Glycogen synthase	K00703, EC:2.4.1.21
2799182780	Ga0311387_10391	*glgB*	1,4-alpha-glucan branching enzyme	K16149, EC:2.4.1.18

Table D.13 - Growth tolerances of *V. stagnispumantis* CP.B2[T]. Four growth characteristics of CP.B2[T] were reanalysed to clarify widespread distribution of the genus in Aotearoa-New Zealand. These results were compared to findings from the original characterisation (Hetzer *et al.*, 2008)[179]. Growth was also tested in Obsidian Pool (OP) spring water, from Yellowstone National Park, USA.

Parameter	Unit	Original CP.B2[T]	Original Tested	Current CP.B2[T]	Current Tested
Temperature (opt)	°C	45-75 (70)	n.r.	38.5-79.7 (70.4)	38.5-80
pH (opt)	n.a.	4.8-5.8 (5.4)	4.5-6.5	3.0-8.5 (6.0)	3.0-8.5
Salinity (opt)	% (w/v)	0.0-0.8 (0.4)	0-2	0.0-8.0 (0-0.2)	0-10
Oxygen	% (v/v)	1-10 (4-8)	0-20	<1.25-10	0-25
OP growth	n.a.	n.d.	n.d.	Yes	Yes

n.a.: not applicable, n.d.: not determined, n.r.: not reported, OP: Obsidian Pool, opt: optimum

Table D.14 - 16S rRNA gene sequences assigned to the genus *Venenivibrio* in the SILVA SSU r138.1 database. Pairwise sequence similarity to *V. stagnispumantis* CP.B2[T] is shown.

Accession	Sample ID	Country	Definition	Sequence length	Similarity to CP.B2[T] (%)	Reference
DQ989208	CP.B2	New Zealand	*Venenivibrio stagnispumantis*	1506	100	Hetzer *et al.*, 2008[179]
FN429034	ChP_216S	New Zealand	Hot spring sinter clone	896	100	Childs *et al.*, 2008[187]
AF402979	bacteriap28	New Zealand	Hot spring clone	1441	99	Sunna & Bergquist, 2003[331]
FN429037	ChP_180M	New Zealand	Hot spring sinter clone	899	99	Childs *et al.*, 2008[187]
EF101539	OTU_Bac04	New Zealand	Hot spring water clone	715	99	Hetzer *et al.*, 2007[138]
EF101540	OTU_Bac12	New Zealand	Hot spring water clone	536	99	Hetzer *et al.*, 2007[138]
FN429036	ChP_103P	New Zealand	Hot spring sediment clone	896	99	Childs *et al.*, 2008[187]
FN429035	ChP_213S	New Zealand	Hot spring sinter clone	891	98	Childs *et al.*, 2008[187]
KM221374	15-B-25	China	Hot spring clone	1483	93	Li *et al.*, 2016[162]
KM221487	W8B-9	China	Hot spring clone	1483	93	Li *et al.*, 2016[162]
KP100157	J-B-1	China	Hot spring clone	1483	93	Jiang *et al.*, 2017[414]
KP175577	T-5	China	Hydrogenothermaceae bacterium T-5	1432	93	Hedlund *et al.*, 2015[293]

HE657259	TC13	China	Hot spring mat clone	605	93	Pagaling *et al.*, 2012[415]
KM221400	9B-25	China	Hot spring clone	1484	92	Li *et al.*, 2016[162]
JX298759	9B-25	China	Hot spring mat clone	1483	92	Peng *et al.*, 2013[416]
KC831413	B3_1	Thailand	Hot spring sediment clone	1481	92	Unpublished
HE657263	TC17	China	Hot spring mat clone	750	92	Pagaling *et al.*, 2012[415]
HE657264	TC18	China	Hot spring mat clone	750	92	Pagaling *et al.*, 2012[415]
HE657266	TC20	China	Hot spring mat clone	750	92	Pagaling *et al.*, 2012[415]
FJ821651	109	India	Hot spring sediment clone	669	92	Unpublished
FM210909	BJ67	Argentina	Salt lake clone	663	92	Pagaling *et al.*, 2009[417]
HE657256	TC10	China	Hot spring mat clone	621	92	Pagaling *et al.*, 2012[415]
HE657260	TC14	China	Hot spring mat clone	606	92	Pagaling *et al.*, 2012[415]
HE657262	TC16	China	Hot spring mat clone	605	92	Pagaling *et al.*, 2012[415]
KP100160	J-B-67	China	Hot spring clone	1484	91	Jiang *et al.*, 2017[414]
HE657267	TC21	China	Hot spring mat clone	750	91	Pagaling *et al.*, 2012[415]
DQ813026	DSK5-13-75	Thailand	Hot spring clone	506	90	Purcell *et al.*, 2007[143]
LC149934	DTR14_V_36	Indonesia	Travertine clone	376	90	Sugihara *et al.*, 2016[418]
KP100159	J-B-25	China	Hot spring clone	1336	89	Jiang *et al.*, 2017[414]
HQ416840	OTU7_1	Thailand	Hot spring clone	874	89	Cuecas *et al.*, 2014[160]
HQ416841	OTU7_2	Thailand	Hot spring clone	841	89	Cuecas *et al.*, 2014[160]
KP100161	J-B-68	China	Hot spring clone	1480	89	Jiang *et al.*, 2017[414]
MH394142	2113_24899_14329	Algeria	Hot spring clone	372	76	Adjeroud *et al.*, 2020[419]

Table D.15 - Search for *Venenivibrio* in the IMNGS database. The 16S rRNA gene of *V. stagnispumantis* CP.B2^T was queried against the full database of SRA amplicon samples (*n*=500,048) and the number of reads at 95, 97, and 99 % sequence similarity are shown. Sample size, description, and country of origin are also presented, along with the corresponding OTU for each read. Samples are ordered by decreasing percent identity to *V. stagnispumantis* CP.B2^T.

Sample ID	Sample size	Description	Country	OTU ID	OTU length	≥95 % similarity	≥97 % similarity	≥99 % similarity	Identity (%)	Reference
SRR1924223	334328	root	Australia	SRR1924223.500542.2	376	5	5	0	98.4	Yeoh *et al.*, 2016[399]
SRR1029457	8444	soil	Canada	SRR1029457.7257.2	459	1	1	0	97.8	Sharp *et al.*, 2014[153]
SRR2026416	9481	soil	Canada	SRR2026416.727.2	457	1	1	0	97.4	Unpublished
ERR493751	179484	synthetic	NA	ERR493751.827.1	238	366	366	0	97.1	Unavailable
ERR493753	141599	synthetic	NA	ERR493753.347.1	238	187	187	0	97.1	Unavailable
ERR493752	32658	synthetic	NA	ERR493752.2046.1	238	54	54	0	97.1	Unavailable
ERR493749	26420	synthetic	NA	ERR493749.148.1	238	50	50	0	97.1	Unavailable
ERR493748	9819	synthetic	NA	ERR493748.2870.1	238	16	16	0	97.1	Unavailable
ERR493750	38814	synthetic	NA	ERR493750.4245.1	238	78	0	0	96.6	Unavailable
ERR493713	1458039	synthetic	NA	ERR493713.715487.1	244	21	0	0	96.3	Unavailable
ERR777710	843578	synthetic	NA	ERR777710.23179.1	355	2336	0	0	96.1	Schirmer *et al.*, 2015[420]
ERR777710	843578	synthetic	NA	ERR777710.4514.1	347	1945	0	0	96	Schirmer *et al.*, 2015[420]
ERR493713	1458039	synthetic	NA	ERR493713.2679.1	244	8040	0	0	95.9	Unavailable
ERR493712	959518	synthetic	NA	ERR493712.374.1	241	1457	0	0	95.9	Unavailable
ERR2240329	6757	hot springs	New Zealand	ERR2240329.4881.1	273	334	0	0	95.6	Power *et al.*, 2018[195]
ERR2240436	6736	hot springs	New Zealand	ERR2240436.2803.1	271	767	0	0	95.6	Power *et al.*, 2018[195]
SRR4039594	103832	aquatic	Germany	SRR4039594.79187.1	292	1	0	0	95.5	Unpublished
ERR2240891	12612	hot springs	New Zealand	ERR2240891.6.1	264	7379	0	0	95.5	Power *et al.*, 2018[195]
SRR1103269	6871	wastewater	China	SRR1103269.1135.2	334	1	0	0	95.5	Unpublished
SRR1106683	6849	wastewater	China	SRR1106683.5319.2	334	1	0	0	95.5	Unpublished
SRR6664286	282671	hot springs	New Zealand	SRR6664286.67117.1	292	13	0	0	95.2	Stewart *et al.*, 2018[189]
ERR2240457	32421	hot springs	New Zealand	ERR2240457.216.1	271	21809	0	0	95.2	Power *et al.*, 2018[195]
ERR2240443	29458	hot springs	New Zealand	ERR2240443.161.1	273	16672	0	0	95.2	Power *et al.*, 2018[195]
ERR2240455	24519	hot springs	New Zealand	ERR2240455.49.1	273	12097	0	0	95.2	Power *et al.*, 2018[195]

ERR2240604	23587	hot springs	New Zealand	ERR2240604.1494.1	271	628	0	0	95.2	Power *et al.*, 2018[195]
ERR2240397	22335	hot springs	New Zealand	ERR2240397.280.1	270	261	0	0	95.2	Power *et al.*, 2018[195]
ERR2240384	14651	hot springs	New Zealand	ERR2240384.20.1	272	12732	0	0	95.2	Power *et al.*, 2018[195]
SRR090539	6990	hydrocarbon	Canada	SRR090539.5381.2	335	3	0	0	95.2	Unpublished
SRR2064977	31497	freshwater	China	SRR2064977.24656.1	407	1	0	0	95.1	Unpublished
SRR7086599	68629	freshwater	India	SRR7086599.13131.1	440	36	0	0	95	Unpublished
ERR2240263	25770	hot springs	New Zealand	ERR2240263.25693.1	280	40	0	0	95	Power *et al.*, 2018[195]

Table D.16 - *Venenivibrio* search results from the Sequence Read Archive (SRA). Both amplicon and metagenomic samples in SRA were screened for similarity to the 16S rRNA gene sequence of *V. stagnispumantis* CP.B2^T (tax_id=407997), using the STAT tool in the Google Cloud BigQuery platform. Only samples with k-mer counts >25 are shown.

Accession	SRA study	Bioproject	k-mer count
SRR15830908	SRP336352	PRJNA761929	7146
SRR15830907	SRP336352	PRJNA761929	3218
SRR14702244	SRP314913	PRJNA417281	1672
SRR12963992	SRP290760	PRJNA644733	430
ERR3841956	ERP119639	PRJEB36446	379
SRR13052607	SRP292461	PRJNA678141	351
SRR12963985	SRP290760	PRJNA644733	324
ERR3841954	ERP119639	PRJEB36446	311
SRR12963981	SRP290760	PRJNA644733	309
SRR12963974	SRP290760	PRJNA644733	296
ERR3841955	ERP119639	PRJEB36446	277
SRR12963993	SRP290760	PRJNA644733	259
SRR13413827	SRP301353	PRJNA691419	240
SRR14771944	SRP323465	PRJNA736448	211
SRR16914159	SRP345512	PRJNA776698	204
SRR15830909	SRP336352	PRJNA761929	180
SRR13720069	SRP306614	PRJNA702085	172
SRR13151504	SRP294353	PRJNA681007	150
SRR12963983	SRP290760	PRJNA644733	140
SRR12963959	SRP290760	PRJNA644733	140
SRR13052603	SRP292461	PRJNA678141	103
ERR5900061	ERP128879	PRJEB44790	85
SRR15688994	SRP335227	PRJNA759289	82
SRR15688978	SRP335227	PRJNA759289	78
SRR12963990	SRP290760	PRJNA644733	77
SRR12963991	SRP290760	PRJNA644733	73
SRR13808339	SRP308588	PRJNA705495	71
SRR13495290	SRP302655	PRJNA693537	71
SRR13808341	SRP308588	PRJNA705495	71
SRR13997373	SRP311198	PRJNA715357	71
SRR12852619	SRP287871	PRJNA669531	70
SRR12963915	SRP290760	PRJNA644733	68
SRR12963976	SRP290760	PRJNA644733	66
SRR14310727	SRP316062	PRJNA724715	61
SRR13808342	SRP308588	PRJNA705495	60
SRR13495289	SRP302655	PRJNA693537	58
ERR4171976	ERP121855	PRJEB38431	58
SRR13151505	SRP294353	PRJNA681007	58
SRR15688983	SRP335227	PRJNA759289	57
SRR14865839	SRP303567	PRJNA694730	57
ERR4467403	ERP123473	PRJEB39903	56

Accession	SRA study	Bioproject	k-mer count
SRR14148522	SRP313593	PRJNA666502	56
SRR14250241	SRP315032	PRJNA722295	52
SRR13495288	SRP302655	PRJNA693537	52
ERR5900060	ERP128879	PRJEB44790	51
SRR12963973	SRP290760	PRJNA644733	49
SRR13005775	SRP291575	PRJNA674461	48
SRR14310722	SRP316062	PRJNA724715	47
SRR13997375	SRP311198	PRJNA715357	47
SRR13120032	SRP293751	PRJNA680348	47
ERR4467406	ERP123473	PRJEB39903	46
SRR13808340	SRP308588	PRJNA705495	45
SRR14148517	SRP313593	PRJNA666502	43
SRR12825386	SRP287391	PRJNA668816	42
SRR12963971	SRP290760	PRJNA644733	41
SRR12849258	SRP287782	PRJNA670083	40
SRR13766296	SRP307608	PRJNA694468	40
ERR4171977	ERP121855	PRJEB38431	40
SRR12825385	SRP287391	PRJNA668816	40
ERR5900778	ERP128879	PRJEB44790	40
SRR12963987	SRP290760	PRJNA644733	38
SRR15688995	SRP335227	PRJNA759289	38
SRR14250243	SRP315032	PRJNA722295	37
SRR14148838	SRP313601	PRJNA719925	37
ERR3478686	ERP115921	PRJEB33151	36
SRR15927520	SRP337539	PRJNA763043	35
ERR3501975	ERP116119	PRJEB33335	34
SRR13120108	SRP293751	PRJNA680348	34
ERR4467392	ERP123473	PRJEB39903	33
SRR12963984	SRP290760	PRJNA644733	32
ERR3895453	ERP119835	PRJEB36623	32
SRR13052609	SRP292461	PRJNA678141	32
SRR16070367	SRP338671	PRJNA766181	31
SRR13997378	SRP311198	PRJNA715357	31
SRR12730671	SRP285642	PRJNA665748	31
SRR12963972	SRP290760	PRJNA644733	30
SRR12990404	SRP291248	PRJNA645054	29
SRR13872966	SRP309628	PRJNA681299	28
SRR13313541	SRP299468	PRJNA686226	28
SRR12963978	SRP290760	PRJNA644733	28
ERR4467387	ERP123473	PRJEB39903	27
SRR13247020	SRP297763	PRJNA684762	26
SRR13108237	SRP293970	PRJNA679303	26
SRR16014960	SRP253926	PRJNA614995	26
SRR12963948	SRP290760	PRJNA644733	26

Table D.17 - *Venenivibrio* read abundance in global spring metagenomes and associated metadata. Twenty-one samples returned traces of *V. stagnispumantis* CP.B2^T from community characterisation of 188 metagenomes. The samples are ordered by descending read number that assigned to the genus, with percentage of the whole community also shown. Samples that were subsequently aligned to Hydrogenothermaceae reference genomes are highlighted.

									Venenivibrio	
Accession	Sample ID	Spring	Region	Country	Description	Collection date	T (°C)	pH	(%)	Reads
DRR163686*	Y1	Yamanoshiro	Kirishima	Japan	water	2013-09	90.4	4.8	1.08	594730
DRR163687*	N1	Nonoyukoya	Kirishima	Japan	water	2013-09	68	6.9	0.34	202871
SRR7905022*	JP1L	Jinata	Shikinejima Island	Japan	precipitates	2016-09	63	5.4	0.03	26312
DRR163685	K1	Kinyu	Kirishima	Japan	water	2013-09	84.5	6.6	0.03	18500
SRR6049666*	OP-RAMG-02	Obsidian	YNP	USA	sediment	2012-10-12	68-75	6-6.5	0.01	10930
SRR8554444	YNP-CB-019-1	n.r.	YNP	USA	sediment	2017-10-29	n.r.	n.r.	0	7478
SRR7039829	US80-3rd	Ulu Slim	Perak	Malaysia	water	2016-08-04	80	~7	0	5900
SRR5650826	BY_YNP_MoundSpring	Mound	YNP	USA	sediment	2009-09-14	55	9	0.01	3853
SRR7039830	US89-3rd	Ulu Slim	Perak	Malaysia	water	2016-08-04	89	~7	0	2898
SRR5247111	JC3_ASED	Joseph's Coat	YNP	USA	sediment	n.r.	n.r.	n.r.	0	2032
SRR1297204	SHP fosmid	Shi-Huang-Ping	Tatun	Taiwan	water	2010/12-2011/03	69	2.5	0	1361
SRR7905025	JP3	Jinata	Shikinejima Island	Japan	mat	2016-01	37.3-46	6.7	0	1135
SRR5580900*	DC2_2012	Dewar Creek	British Columbia	Canada	sediment	2012-09-26	66.4	7.93	0	634
SRR5580903	DC16B_2010	Dewar Creek	British Columbia	Canada	sediment	2010-09-15	44.5	8.15	0	385
SRR4027810*	Enrichment CS 85C	Great Boiling	Nevada	USA	sediment	2010-02-24	72	7.05	0	344
SRR5650740	C32_wellF02	Mound	YNP	USA	sediment	2009-09-14	55	9	0.01	148
SRR5581334	LN4_2012	Larsen	British Columbia	Canada	sediment	2012-10-01	33.1	7.16	0	96
SRR5650733	C18_wellC07	Mound	YNP	USA	sediment	2009-09-14	55	9	0.01	80
SRR3961740	TAT-1	Tattapani	Chhattisgarh	India	water	2016-01-26	98	7.5	0	78
DRR163690	I1	Ioudani	Kirishima	Japan	water	2013-09	88	2.9	0	52
DRR163691	T1	Torigigoku	Kirishima	Japan	water	2013-09	84	2.4	0	20

*Samples aligned to *Hydrogenothermaceae* reference genomes
YNP: Yellowstone National Park, n.r.: not reported

Table D.18 - Alignment of metagenomic samples to Hydrogenothermaceae genomes. Percentage of total genome coverage breadth per sample is shown, along with coverage depth, for each of the four genomes used in the alignment. Local metagenomes to Aotearoa-New Zealand are sample IDs CPp, P1.0019, NZ1, and P2.0050. The remaining six metagenomes are globally sourced.

Sample ID	*V. stagnispumantis* CP.B2^T		*P. hydrogeniphila* 29W^T		*S. yellowstonense* SS-5^T		*Sulfurihydrogenibium* sp. Y03AOP1	
	Coverage breadth (%)	Coverage depth (x)	Coverage breadth (%)	Coverage depth (x)	Coverage breadth (%)	Coverage depth (x)	Coverage breadth (%)	Coverage depth (x)
CPp	96.82	277.76	0.19	0.10	0.59	0.17	0.44	0.08
P1.0019	94.44	35.37	0.11	0.01	0.41	0.03	0.26	0.01
NZ1	93.01	1005.61	0.21	0.02	0.81	0.17	0.68	0.14
P2.0050	91.80	100.66	0.16	0.00	0.45	0.01	0.50	0.01
DRR163686	10.17	16.21	3.19	5.50	12.86	25.43	13.09	27.39
DRR163687	8.52	4.82	2.81	3.91	10.85	7.43	11.69	7.78
SRR7905022	2.18	0.30	6.12	1.38	1.73	0.24	1.40	0.23
SRR6049666	0.15	0.06	0.14	0.08	4.99	0.93	4.82	0.64
SRR5580900	0.12	0.03	0.16	0.07	0.41	0.08	0.16	0.02
SRR4027810	0.11	0.00	0.14	0.06	0.49	0.04	0.23	0.01

Table D.19 - Metagenome-assembled genomes (MAGs). Four local (CPp, NZ1, P1.0019 and P2.0050) and two global (DRR163687 and DRR163686) metagenomic samples were *de novo* assembled and binned into MAGs. Average nucleotide identity (ANI) was confirmed to *V. stagnispumantis* CP.B2^T. Bins are ordered by descending ANI. MAGs deposited in the Aotearoa Genomic Data Repository (AGDR) can be found under the Project ID TAONGA-AGDR00025 at https://doi.org/10.57748/vpk8-zp44.

MAG ID	MAG Accession	Sample ID	Sample Accession	Geothermal feature	Geothermal field	Country	ANI to CP.B2^T (%)	Completeness (%)	Contamination (%)
CPp-MAG1	AGDR00025	CPp	AGDR00025	Champagne Pool	Waiotapu	NZ	97.9	98.8	0.2
NZ1-MAG2	AGDR00025	NZ1	Ga0079637	Waiotapu No.15 Feature 1	Waiotapu	NZ	96.9	99.2	0.6
P1.0019-MAG3	AGDR00025	P1.0019	SRR14702244	Radiata Pool	Ngatamariki	NZ	96.1	96.1	0.6
P2.0050-MAG4	AGDR00025	P2.0050	AGDR00025	Whakarewarewa Feature 51	Rotorua	NZ	94.8	92.7	7.9
DRR163687-MAG5	By request	N1	DRR163687	Nonoyukoya	Kirishima	Japan	77.8	91.9	0.0
DRR163686-MAG6	By request	Y1	DRR163686	Yamanojo	Kirishima	Japan	76.9	96.7	0.0

Table D.20 - Synthetic metagenomes composition and classification. The makeup of the mock communities (sample A-H) is presented in the first column per genome, with the percentage classified using Kraken2 in the second column.

Sample	*V. stagnispumantis* CP.B2^{T} Synthetic (%)	Classified (%)	*P. hydrogeniphila* 29W^{T} Synthetic (%)	Classified (%)	*Sulfurihydrogenibium* sp. Y03AOP1 Synthetic (%)	Classified (%)	Random genomes Synthetic (%)
A	0	0.00	0	0	100	99.97	0
B	0	1.42	100	0	0	0.8	0
C	100	99.78	0	0	0	0	0
D	0	0.00	0	0	0	0	100
E	1	1.04	1	0	45	46.39	52
F	10	9.07	0	0	90	90.89	0
G	1	1.69	45	0	1	1.4	50
H	33	33.74	33	0	33	33.59	0

Table D.21 - Alignment of synthetic metagenomic samples to Hydrogenothermaceae genomes. Average coverage breadth and depth are shown for contigs from each sample across the four genomes used in the alignment.

Sample	*V. stagnispumantis* CP.B2^{T} Coverage breadth (%)	Coverage depth (x)	*P. hydrogeniphila* 29W^{T} Coverage breadth (%)	Coverage depth (x)	*S. yellowstonense* SS-5^{T} Coverage breadth (%)	Coverage depth (x)	*Sulfurihydrogenibium* sp. Y03AOPI Coverage breadth (%)	Coverage depth (x)
A	0.02	0.00	0.00	0.00	0.74	0.99	99.97	81.82
B	0.00	0.00	99.46	75.72	0.00	0.00	0.00	0.00
C	98.66	100.44	0.00	0.00	0.00	0.00	0.00	0.00
D	0.00	0.00	0.02	0.00	0.14	0.18	0.03	1.00
E	83.21	2.31	77.28	1.95	0.85	0.92	99.97	73.64
F	98.27	9.80	0.00	0.00	0.74	0.99	99.97	81.82
G	83.21	2.31	99.45	68.15	0.19	0.18	80.77	2.03
H	98.62	65.01	99.44	50.44	0.60	0.66	99.96	54.50

APPENDIX E

DRAFT GENOME SEQUENCE OF *VENENIVIBRIO STAGNISPUMANTIS* CP.B2^T, ISOLATED FROM CHAMPAGNE POOL, WAIOTAPU, AOTEAROA-NEW ZEALAND

Jean F. Power, Holly E. Welford, Carlo R. Carere, Ian R. McDonald, S. Craig Cary, & Matthew B. Stott

E.1   ABSTRACT

Venenivibrio stagnispumantis strain CP.B2^T is a thermophilic, chemolithoautotrophic bacterium from family Hydrogenothermaceae (phylum Aquificota), isolated from Champagne Pool in the Waiotapu geothermal field, Aotearoa-New Zealand. The genome consists of 1.6 Mbp in 47 contigs with a G+C content of 29.6 % mol.

E.2   ANNOUNCEMENT

Venenivibrio is a thermophilic bacterial genus (phylum Aquificota) found in numerous geothermal springs (*n*=686) in the Taupō Volcanic Zone (TVZ) of Aotearoa-New Zealand[195]. The type strain, *Venenivibrio stagnispumantis* CP.B2^T, was isolated from Champagne Pool, Waiotapu[138]; a geothermal system under the *kaitiakitanga* (guardianship) of Māori *iwi* (tribe) Ngāti Tahu – Ngāti Whaoa, whom assert *mana whenua* (customary rights) over the bacterium, its genome, and the location. Consistent with other genera within the Hydrogenothermaceae, CP.B2^T is characterised as a thermophilic microaerophilic hydrogenotroph[179]. Recently, *Venenivibrio* was reported as the most abundant taxon across TVZ geothermal spring water columns in Aotearoa-New Zealand[195]. Analysis of the CP.B2^T genome provides important insight into the metabolic and phenotypic traits that facilitate *Venenivibrio* spp. being a dominant taxon within Aotearoa-New Zealand geothermal environments.

Venenivibrio stagnispumantis CP.B2^T was cultivated using a modified MSH medium (70 °C; pH 5.5)[179] within a headspace consisting of $N_2:H_2:CO_2:O_2$ at ratios of 50:40:7.5:2.5 (v/v) respectively. A total DNA extract (3.5 ng µl^{-1}) of CP.B2^T, was performed using a NucleoSpin Soil kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocols. Whole genome sequencing (Macrogen Inc, Seoul, South Korea) was undertaken using the Illumina HiSeq 2500 platform with TruSeq DNA Nano (550) library preparation for 101 bp reads (Illumina Inc, San Diego, CA, USA). A total of 3.8 Gbp raw sequences were assembled using SPAdes v3.12.0[421] with the '–careful' option to minimise mismatches and k-mer sizes of 21, 33 and 55. The assembly was evaluated using QUAST v5.0.0[422]. The genome was registered with the Genomes Online Database[423] (GOLD Analysis ID Ga0311387), and annotated using the Integrated Microbial Genomes annotation pipeline v4.16.4[295] (IMG Genome ID 2799112217).

The *V. stagnispumantis* CP.B2^T genome comprised of 1,609,122 bp containing 1,701 protein coding genes, with 1,409 having predicted function. The genome consisted of 47 contigs, with a total G+C content of 29.64 % mol. The N_{50} value was 91,033 and the L_{50} was 8. Encoded within CP.B2^T were enzymes involved in the fixation of CO_2 via the reductive tricarboxylic acid cycle (rTCA Type I). Also annotated was a complete Embden-Meyerhof-Parnas (EMP) pathway for glycolysis. A cytochrome *bd* oxidase (CydAB) was the sole respiratory cytochrome oxidase, confirming observations of O_2 use as a terminal electron acceptor. Genes encoding two [NiFe]-hydrogenases, belonging to groups 1b (HynAB; oxygen-sensitive respiratory) and 2d (HuaSL; oxygen-tolerant uptake), were identified[296,320], and reflect the hydrogenotrophic, microaerophilic phenotype of *V. stagnispumantis*. Despite an obligate growth requirement for sulfur in the nutrient medium[179], no metabolic pathways reflecting hydrogenotrophic sulfur respiration could be identified. The presence of glutamate synthase (GltBD) and glutamine synthetase (GlnA) confirm the assimilation of nitrogen via the uptake of ammonia. Putative genes involved with arsenic resistance (*arsRBC*), flagellar assembly, and chemotaxis were also annotated.

A pairwise average nucleotide identity (ANI) comparison using FastANI v0.1.3 of the *V. stagnispumantis* CP.B2^T, *Sulfurihydrogenibium yellowstonense* SS-5^T, and *Persephonella marina* EX-H1^T genomes gave <80 % similarity for each, confirming the designation of *Venenivibrio* as a separate genus to sister genera *Sulfurihydrogenibium* and *Persephonella*.

E.3 DATA AVAILABILITY

Raw sequences for the *V. stagnispumantis* CP.B2^T genome were deposited in the European Nucleotide Archive under the Bioproject accession PRJEB55610. The assembled genome was deposited in the Genomes Online Database (GOLD Analysis ID Ga0311387), for associated annotation with the Integrated Microbial Genomes & Microbiomes platform (IMG Genome ID 2799112217).

APPENDIX F

CO-AUTHORSHIP FORMS

Please find attached signed co-authorship forms for Chapters 2, 3, and 4, and Appendix E, which were written as manuscripts that have either been published or are currently submitted to journals for peer-review.

Co-Authorship Form

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 2 | Microbial biogeography of 925 geothermal springs in Aotearoa-New Zealand | Published in *Nature Communications* (2018) 9:2876 | DOI: 10.1038/s41467-018-05020-y

Nature of contribution by PhD candidate	Experimental design, field work, sample & sequence processing, data analysis, and manuscript writing
Extent of contribution by PhD candidate (%)	90

CO-AUTHORS

Name	Nature of Contribution
Carlo Carere	Experimental design, field work assistance, manuscript editing
Charles Lee	Sequence processing for website
Georgia Wakerley	DNA extractions & sequencing
David Evans	Field work assistance
Mathew Button	Field application, database, & website
Duncan White	Field application, database, & website
Melissa Climo	Grant writing
Annika Hinze	Field application, database, & website
Xochitl Morgan	Data analysis assistance, manuscript editing
Ian McDonald	Experimental design, manuscript editing
Craig Cary	Experimental design, manuscript editing
Matthew Stott	Experimental design, field work assistance, manuscript editing
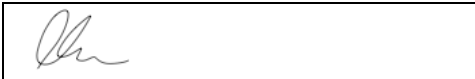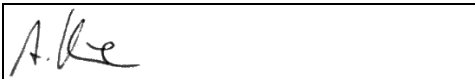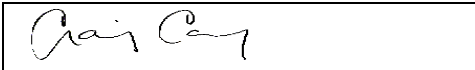
Certification by Co-Authors

The undersigned hereby certify that:
❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
❖ in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
Carlo Carere		12/10/2022
Charles Lee		12th October, 2022
Georgia Wakerley		12th Oct. 22

David Evans		19/10/2022
Mathew Button		13/10/2022
Duncan White		17th October, 2022
Melissa Climo		11th October, 2022
Annika Hinze		12 October, 2022
Xochitl Morgan		12 Oct 22
Ian McDonald		11th October, 2022
Craig Cary		12/10/22
Matthew Stott		11 October, 2022

Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3 | Temporal dynamics of geothermal microbial communities in Aotearoa-New Zealand | Published in *Frontiers in Microbiology* (2023) 12:1094311 | DOI: 10.3389/fmicb.2023.1094311

Nature of contribution by PhD candidate	Experimental design, field work, sample & sequence processing, data analysis, and manuscript writing
Extent of contribution by PhD candidate (%)	90

CO-AUTHORS

Name	Nature of Contribution
Caitlin Lowe	Field work, sample processing, DNA extractions & sequencing
Carlo Carere	Experimental design, field work assistance, manuscript editing
Ian McDonald	Experimental design, manuscript editing
Craig Cary	Experimental design, manuscript editing
Matthew Stott	Experimental design, field work assistance, manuscript editing

Certification by Co-Authors

The undersigned hereby certify that:
- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
Caitlin Lowe		17/10/2022
Carlo Carere		12/10/2022
Ian McDonald		11th October, 2022
Craig Cary		12/10/22
Matthew Stott		11 October, 2022

Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 4 | Allopatric speciation in the bacterial phylum Aquificota enables genus-level endemism in Aotearoa-New Zealand | Submitted to *Nature Communications* (2023)

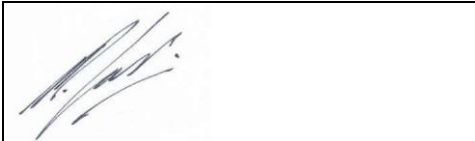|	
Nature of contribution by PhD candidate	Experimental design, sample processing, data analysis, and manuscript writing
Extent of contribution by PhD candidate (%)	80

CO-AUTHORS
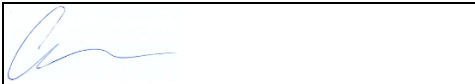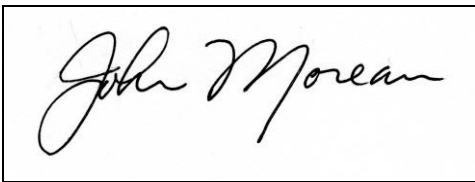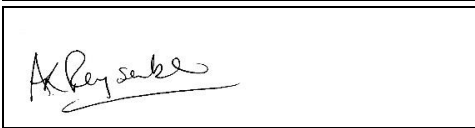
Name	Nature of Contribution
Carlo Carere	Experimental design, manuscript editing
Holly Welford	Culture experiments
Daniel Hudson	Metagenomic analysis
Kevin Lee	Sample provision, manuscript editing
John Moreau	Sample provision, manuscript editing
Thijs Ettema	Sample provision, manuscript editing
Anna-Louise Reysenbach	Sample provision, manuscript editing
Charles Lee	Sample provision
Daniel Colman	Sample provision, manuscript editing
Eric Boyd	Sample provision, manuscript editing
Xochitl Morgan	Experimental design, metagenomic analysis, manuscript editing
Ian McDonald	Manuscript editing
Craig Cary	Manuscript editing
Matthew Stott	Experimental design, culture experiments, manuscript editing

Certification by Co-Authors

The undersigned hereby certify that:
❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
❖ in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
Carlo Carere		12/10/22
Holly Welford		11/10/2022

Daniel Hudson		14th October, 2022
Kevin Lee		15/10/2022
John Moreau		27 October, 2022
Thijs Ettema		18/10/22
Anna-Louise Reysenbach		19/10/22
Charles Lee		12th October, 2022
Daniel Colman		12th October 2022
Eric Boyd		12th October 2022
Xochitl Morgan		12 Oct 22
Ian McDonald		11th October, 2022
Craig Cary		12/10/22
Matthew Stott		11 October, 2022

Co-Authorship Form

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Appendix E | Draft genome sequence of *Venenivibrio stagnispumantis* CP.B2^T, isolated from Champagne Pool, Waiotapu, Aotearoa-New Zealand | Published in *Microbiology Resource Announcements* (2023) 12:e01074-22 | DOI: 10.1128/mra.01074-22

Nature of contribution by PhD candidate	Genome assembly, quality control, annotation, and manuscript writing
Extent of contribution by PhD candidate (%)	90

CO-AUTHORS

Name	Nature of Contribution
Holly Welford	Secondary annotation, manuscript writing
Carlo Carere	DNA extraction, annotation guidance, manuscript editing
Ian McDonald	Manuscript editing
Craig Cary	Manuscript editing
Matthew Stott	Annotation guidance, manuscript writing

Certification by Co-Authors

The undersigned hereby certify that:
- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
Holly Welford		11/10/2022
Carlo Carere		12/10/2022
Ian McDonald		11th October, 2022
Craig Cary		12/10/22
Matthew Stott		11 October, 2022