Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Aquifer geochemistry shapes microbial diversity and carbon and nitrogen cycling potential

Karen M. Houghton

karen.houghton@earthsciences.nz

Earth Sciences New Zealand

Magali Moreau

Earth Sciences New Zealand

Conny Tschritter

Earth Sciences New Zealand

Research Article

Keywords: Aquifer, Diversity, Geochemistry, Groundwater, Hydrochemistry, Metabolism, Metagenome

Posted Date: October 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11342114/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Despite the importance of groundwater and its ecosystem functions, including nitrogen cycling, subsurface microbial biodiversity and ecology remain poorly understood globally. We investigated how aquifer geochemistry shapes microbial diversity and metabolic potential across Aotearoa New Zealand using shotgun metagenomics of 39 groundwater samples from 11 regions. We recovered 494 high- and medium-quality metagenome-assembled genomes (MAGs), which clustered into four groups (A, B, C1 and C2), based on hydrochemistry, and differed significantly in taxonomy and metabolic pathways. Cluster A had the highest microbial diversity and was dominated by uncultured taxa, particularly *Patescibacteria*. Carbon fixation pathways, predominantly the Calvin-Benson-Bassham and reductive tricarboxylic acid cycles, varied among clusters and were associated with different redox metabolisms. Aerobic respiration pathways also differed, particularly the microaerophilic cbb_3 cytochrome *c* oxidase. Differences were also observed in nitrogen, sulfur, carbohydrate and organic acid metabolism, with several MAGs strongly correlated with nitrate, sulfate and potassium. Our findings demonstrate that aquifer geochemistry is a major determinant of microbial community structure and metabolic potential across Aotearoa New Zealand groundwater ecosystems.

INTRODUCTION

Groundwater is Earth's largest freshwater resource, sustaining subterranean and surface-water ecosystems through springs and river baseflow (Taylor et al., 2013, White et al., 2015). In Aotearoa New Zealand, groundwater is the only source of drinking water for about 40% of New Zealanders (Davies, 2001), and it is used to irrigate ~ 370,000 ha of land (KC, 2019).

Aotearoa New Zealand groundwater harbours novel and highly diverse microbial communities, with high proportions of as yet uncultured microbes (Houghton, 2025, Gios et al., 2023, Mosley et al., 2022). Ecosystem services such as nitrogen and sulfur cycling, and degradation of toxic substances, result directly from this biodiversity and its biological processes (Griebler and Avramov, 2015). Microbial community diversity has previously been shown to correlate with pH, temperature, salinity and conductivity (Fierer and Jackson, 2006, Thompson et al., 2017, Power et al., 2018, Bas-Lorillot et al., 2025). Other hydrochemical parameters (e.g. dissolved oxygen, arsenic, iron and sulfate) may also drive diversity, either by acting as an energy source or through their toxic effects on micro-organisms (Sirisena et al., 2018, Houghton and Santamaria Cerrutti, 2022). Some have been reported at naturally occurring high concentrations in Aotearoa New Zealand (Moreau et al., 2025). High biodiversity is often linked to functional redundancy and ecosystem resilience (Fetzer et al., 2015), and changes in an environment can lead to a decrease in or loss of specific ecosystem processes. Understanding groundwater ecosystems, and particularly their sensitivity to contaminants such as nitrate or microplastics, is a critical knowledge gap within Aotearoa New Zealand (Ministry for the Environment and Stats NZ, 2026).

This study investigates the effect of hydrochemistry on microbial diversity and metabolic functions in subsurface ecosystems. We recovered 494 metagenome-assembled genomes (MAGs) from 39

groundwater sites in 11 regions distributed across the country, covering a wide range of hydrochemistry, lithology, depths, groundwater age and land use types. Land use may be expected to alter microbial communities through the addition of nutrients, pesticides or other pollutants such as heavy metals (Watson et al., 2023, Liang et al., 2026). However, previous studies in Aotearoa New Zealand have not detected significant relationships between aquifer characteristics, hydrochemistry or microbial diversity (Daughney and Reeves, 2005, Sirisena et al., 2018). This may be because land-use information is limited to the immediate bore area rather than the entire recharge zone and does not account for the longer residence times of many groundwaters.

To the best of our knowledge, this is the first national study of groundwater metagenomes in Aotearoa New Zealand, although regional studies on specific metabolisms have previously been carried out (Gios et al., 2023, Mosley et al., 2022). Results give insights into environmental factors influencing the presence, co-occurrence, and metabolic potential of sub-surface micro-organisms which determine the fate of nutrients and toxic substances in Aotearoa New Zealand aquifers.

METHODS

Sample collection

Groundwater was collected from 37 wells and 2 springs spanning 33 aquifers in 11 regions nationwide (Fig. 1). Wells were 3-372 m deep (median 32 m), and groundwater ages ranged between 1 and approximately 350 years mean resident time (MRT) (median 38 years). To calculate median age, data given as > *x* years was listed as *x* + 1 years (e.g. >200 years mean residence time was converted to 201 years). Samples encompassed a range of lithologies (mostly gravel, *n* = 9, and basalt, *n* = 7) and land use types (mostly pasture, *n* = 12, and residential, *n* = 6) from Regional Council data; full details including hydrochemistry are given in Table S1.

A) Regional boundaries of New Zealand, with locations of samples. Regions are coloured if at least one sample was taken. B). Simplified geology and bore depth of samples.

There were some noticeable differences between the regions. Bores from the Bay of Plenty and Auckland regions were generally deeper (medians 220 and 69 m, respectively), and older (3/4 samples, and 6/8 samples over 190 years old, respectively). Five of the seven basalt bores were from the Auckland region, while most of the gravel samples came from the South Island (Marlborough, Canterbury and Tasman regions). Land use types were more evenly distributed across the bores.

Wells were purged (at least three bore volumes) until field measurements stabilised (temperature, pH, and electrical conductivity), following the National Protocol for State of the Environment Groundwater Sampling (Daughney et al., 2006). Water for chemistry analysis was collected in pre-prepared bottles prior to collection of water for microbial DNA analysis in 2 L pre-autoclaved bottles, using nitrile gloves. All samples were transported to and stored at the Earth Sciences New Zealand Microbiology Laboratory (Wairakei Research Centre) at 4°C. All laboratory hydrochemistry analyses were performed at the New

Zealand Geothermal Analytical Laboratory (NZGAL; Wairakei, Aotearoa New Zealand) following national standards (Milne et al., 2019).

DNA extraction and sequencing

DNA extraction and sequencing
Methods are detailed in (Houghton, 2025). Briefly, groundwater samples were filtered through a 0.1 membrane, DNA extracted using DNeasy PowerSoil Pro and sequenced using Illumina NovaSeq for whole genome shotgun sequencing. Taxonomy was assigned using the Genome Taxonomy Database Toolkit (GTDB-tk) (Chaumeil et al., 2019). Genome completeness and contamination were identified as defined by the Genomic Standards Consortium (GSC): completeness as the ratio of observed single-copy marker genes to total single-copy marker genes in the chosen marker gene set (bac120 for Bacteria and ar53 for Archaea); contamination as the ratio of observed single-copy marker genes present in ≥ 2 copies to total single-copy marker genes in that set.

Raw, unprocessed DNA sequences were uploaded to the NCBI Sequence Read Archive (SRA) following the Minimum Information about a Metagenomic Sequence (MIMS) standards for metadata (Field et al., 2008); sequences have been deposited under BioProject PRJNA1230396.

Metagenomic metabolic predictions

Specific genes involved in carbon fixation, oxidative phosphorylation, nitrogen metabolism, and sulfur metabolism were identified from Kegg Orthology (KO) classifications by DRAM (Shaffer et al., 2020). The completeness of metabolic pathways in each MAG was determined based on the presence or absence of these genes. Carbohydrate and organic acid metabolism, and arsenate and mercury reduction potential were defined as present or absent using DRAM (Shaffer et al., 2020).

Hierarchical cluster analysis (HCA)

Following best practice in Aotearoa New Zealand (Daughney et al., 2006), HCA was conducted on log-transformed and normalised median concentrations, temperatures and conductivities, using Ward's linkage rule (Ward and Hook, 1963) and squared Euclidean distance as the similarity measure, across 16 parameters: Ca, Cl, DRP, Fe, Mg, Mn, K, Na, NH_3-N, NO_3-N, Si, SO4, Total Alkalinity, field electrical conductivity, pH, and temperature.

Statistical analysis

Welch's ANOVA (base R; (R Core Team, 2025)) was used to test for significant differences between HCA clusters in metabolic pathways — either percentages of complete pathways (carbon fixation, oxidative phosphorylation, nitrogen and sulfur metabolism) or presence/absence of pathways

(carbohydrate/organic acid metabolism, arsenate and mercury reduction) — reflecting non-normal data with unequal variance. ANOVAs were followed by Games-Howell post-hoc tests (rstatix package; (Kassambara, 2023)). Correlations between hydrochemical parameters and microbial communities were investigated using the microeco R package (Liu et al., 2020).

RESULTS

General description of the assembled genomes

In total, 494 metagenome-assembled genomes (MAGs) were recovered from the 39 groundwater samples, which were ≥ 70% complete and with ≤ 10% contamination (Table S2). Of these, 43 were classified as high-quality (≥ 90% complete and with ≤ 5% contamination (Bowers et al., 2017)).

Numbers of recovered MAGs varied from 2 (GW18WARK) to 59 (GW158TN50). There were 472 bacterial MAGs (95.5%), spanning 19 phyla, and 22 archaeal MAGs from 6 phyla. Archaeal MAGs were less likely to be classified to genus level; 3/22 (13.6%), compared to 320/472 bacterial MAGs (67.8%), although all archaeal MAGs were classified to at least class level. Nineteen bacterial MAGs were classified only to domain (1) or phylum (18) level (Table S2).

Predictions of metabolic pathways

Although only one MAG had a complete pathway for carbon fixation, a MAG was considered capable of autotrophy if 60% of the genes for a carbon fixation pathway were present (Rogers et al., 2023). 270 MAGs (54.7%) met this requirement for the Calvin-Benson-Bassham (CBB) pathway, although only 66 MAGs had genes for both chains of RuBisCo. 244 MAGs (49.4%) met the criteria for the Reductive Tricarboxylic Acid (rTCA) cycle, and only 11 MAGs (2.2%) and 3 MAGs (0.6%) had at least 60% of the genes for the Dicarboxylate-Hydroxybutyrate (DC/HB) cycle or Wood-Ljungdahl (WL)/reductive acetyl-CoA pathways, respectively. Microaerophilic respiratory oxidases were more common than aerobic types, with 247 MAGs (50.0%) encoding complete cbb_3 cytochrome *c* oxidases and 81 (16.4%) encoding cytochrome *bd* oxidases, compared to cytochrome *c* oxidase (92, 18.6%) and cytochrome *o* oxidase (69, 14.0%).

Complete pathways for dissimilatory nitrate reduction were found in 159 (32.2%) of MAGs, but nitrogen fixation (66 MAGs, 13.4%), and full denitrification (29 MAGs, 5.9%) were rare. No MAGs contained complete pathways for ammonia removal via nitrification or anammox. Sulfate oxidation (SOX system) was present in 79 MAGs (16.0%). No MAGs contained complete methanogenic pathways, the methyl coenzyme M reductase gene, or photosystems.

Hierarchical clustering analysis

HCA grouped sites by hydrochemistry; eight samples (92 MAGs) lacked chemistry data and were excluded. At the highest threshold, the data is divided into three clusters, A, B and C, with the third cluster being further divided into clusters C1 and C2 (Fig. 2).

The vertical axis represents the distance or dissimilarity between clusters. Regions are coloured as in Fig. 1 (North Island blues and greens, South Island reds and pinks).

Cluster A (8 samples, 142 MAGs) comprised Mg-HCO3 waters with elevated conductivity and reducing conditions, indicative of prolonged residence (median age 53 years). Clusters B and C were more oxygenated with lower solute concentrations: Cluster B (12 samples, 145 MAGs) was also Mg-HCO3, while Cluster C2 (8 samples, 81 MAGs) was mixed. Cluster C1 (3 samples, 34 MAGs; two Na-HCO3, one mixed) had higher pH (median 7.51) and elevated DRP and sodium; Cluster C2 had higher median temperatures, lower median pH, and higher silica than other clusters (Fig. 3).

The number displayed above each box plot indicates the sample size. The horizontal lines in the middle of each box are median values, and the upper and lower limit of the box are the 75th and 25th percentiles, respectively. The vertical lines extend from the 95th to the 5th percentiles.

There was some geographical clustering; Cluster A contained all Gisborne samples and two of the three Taranaki samples, while Cluster B contained all of the South Island samples. Clusters C1 and C2 came from several regions (Fig. 2).

For cluster A, there were a range of land-uses from pasture to an industrial ammonia-manufacturing plant, and lithologies were mixed (Table S1). Cluster B samples were from shallow (median: 7 m deep) and young (median: 15 years MRT) groundwater bores, and half of the samples had gravel lithologies. Land-use types ranged from native forest to industrial areas. Cluster C1 bores were deep (median: 161 m) and contained older waters (median: 320 years MRT), despite the low concentrations of solutes such as calcium, magnesium and sulfate.

C2 samples, in addition to the differentiating parameters outlined above, also contained the highest median arsenic (0.014 mg/L) and lithium concentrations (0.09 mg/L, Table S1). This is consistent with the five samples in this cluster from Waikato and the Bay of Plenty from known geothermal areas (GW10BP, GW217WAWP, GW218WACB, GW239BPTT and GW240BPKP). However, boron concentrations, another common geothermal marker, were highest in Cluster A (median 0.07 mg/L), primarily due to GW233BPHR (1.18 mg/L B).

Taxonomic descriptions of the clusters

Cluster A (142 MAGs) contained the highest microbial diversity across 22 phyla and 83 families, with 45/142 MAG classifications found only in this cluster. Many MAGs were most similar to uncultured or unclassified groups, with only 52 MAGs classified to a named genus, largely due to the Taranaki samples GW156TN12 (16/17 MAGs not classified to genus) and GW158TN50 (56/59 MAGs). Over a quarter of MAGs (41) were from the *Patescibacteria* phylum, which consistently co-occurred with one another, or

with members of *Micrarchaeota*, *Nanoarchaeota*, *Nitrospirota*, and *Omnitrophota* (Chaudhari et al., 2021, Gios et al., 2023). Samples without *Patescibacteria* were dominated by *Pseudomonadota*, with common genera including *Methylotenera*, *Rhodoferax*, and *Undibacterium* (6 MAGs each). With one exception, all archaeal MAGs were from Cluster A, with 17/21 of these MAGs from GW158TN50.

Cluster B (145 MAGs) consisted of 34 families from 11 phyla, with the majority of MAGs (111) being classified as *Pseudomonadota* and only 11 from *Patescibacteria*. 16 MAG classifications were found only in this cluster. MAGs were commonly classified as genera prevalent in groundwater (Houghton, 2025, Watson et al., 2023), including *Aquabacterium* (16 MAGs), *Polaromonas* (13 MAGs), and *Phenylobacterium* (12 MAGs). The only other archaeal MAG, from the phylum *Nanoarchaeota*, was identified in GW210WAFS in this cluster.

Cluster C1 (34 MAGs) consisted of 15 families from just three phyla, with 29 MAGs from *Pseudomonadota*. Three MAG classifications were found only in this cluster, the order *Sphingomonadales* and the genera *Algoriphagus* and *Alkanindiges*. Cluster C2 (81 MAGs) comprised taxa from 24 families across 8 phyla, primarily common groundwater microbes from *Pseudomonadota* (65 MAGs, including *Aquabacterium*, 11 MAGs), with 10 MAGs from uncultured phyla including *Patescibacteria* (7 MAGs). 14 MAG classifications were found only in this cluster.

MAGs not assigned to an HCA cluster (*n* = 92) were from 31 families across 9 phyla, with 14 MAG classifications not found in the other clusters. In this group, methane-oxidising bacteria (methanotrophs) *Methylobacter* (4 MAGs) were identified as well as methylotrophs (11 MAGs), which oxidise methanol formed by methanotrophs.

Some samples carried MAGs from taxa capable of anoxygenic photosynthesis (e.g. *Rhodoferax*, *Novosphingobium*) or otherwise more typical of surface water than groundwater (Imhoff et al., 2019, Yurkov and Messner, 2025); notably 3/7 MAGs (43%) from GW215WAEW and 4/10 (40%) from GW217WARP were potential surface-water bioindicators. Samples with these bioindicators did not cluster together by hydrochemistry.

Metabolic descriptions of the clusters

Metabolic profiles by cluster are shown in Fig. 4 (oxidative phosphorylation, carbon fixation, nitrogen and sulfur metabolism), Fig. 5 (carbohydrate/organic acid metabolism, mercury and arsenate reduction), and Tables S4–S5 (per-MAG detail). Significant differences existed between clusters (Table S3) for all carbon fixation types, cytochromes *o*, *c*(cbb_3) and *bd*, nitrate reduction, nitrogen fixation, and sulfate reduction Fig. 4; and for carbohydrate metabolism (chitin, glucans, polyphenolics), organic acid metabolism (pyruvate dehydrogenase, alcohol production), and arsenate reduction (Fig. 5).

The heatmap represents metabolic pathway completeness of the MAGs based on the presence or absence of genes (as Kegg Orthologs) identified using the DRAM package (Shaffer, 2020), with the median for each sample shown. Results for individual MAGs are shown in Table S4, and statistical test results in Table S3.

The heatmap represents the percentage of MAGs within each sample which contain metabolic pathways identified using the DRAM distillation package (Shaffer, 2020). Results for individual MAGs are shown in Table S5, and statistical test results in Table S3.

Cluster A contained the highest proportion of MAGs (54/142, 38%) which encoded cytochrome *c* oxidase, associated with aerobic respiration, and the lowest proportions with cytochrome *o* oxidase (6, 4%) or cbb_3 cytochrome *c* (36, 25%). Cluster A had less carbon fixation gene content identified than other clusters, with under 30% of MAGs meeting the criteria for either the CBB pathway or the rTCA cycle. Some complete dissimilatory nitrate reduction (32, 23%), nitrogen fixation (16, 11%) and assimilatory sulfate reduction (21, 15%) pathways were identified, consistent with cluster hydrochemistry having reducing conditions and low nitrate (Fig. 3). Heterotrophic pathways for carbohydrate or organic acid metabolism were more common (e.g. chitin degradation, 77 MAGs, 54%, and polyphenolics degradation, 65 MAGs, 46%), though these pathways were still found at lower proportions than in other clusters. Cluster A contained 17 MAGs with no cytochrome *o*, *c* or *bd*-type oxidases and no complete pathways for nitrate, sulfur, carbohydrate or organic acid metabolism, mostly classified as *Patescibacteria* (12 MAGs).

Cluster B (145 MAGs) contained 97 MAGs (67%) with complete pathways for the aerobic respiratory oxidase cbb_3 cytochrome *c* oxidase, consistent with oxygen concentrations across the cluster (Table S1). The cluster also contained 93 MAGs (64%) with at least 60% of the genes required for the CBB carbon fixation pathway and 87 MAGs with at least 60% of rTCA cycle genes (many of these MAGs contained both).

Complete dissimilatory nitrate reduction pathways were identified in 63 MAGs (43%), consistent with slightly higher median nitrate concentrations in this cluster (Fig. 3). Cluster B contained 14 MAGs with no cytochromes identified, and no complete pathways for carbon fixation, nitrate, sulfur, carbohydrate, or organic acid metabolism, with ten classified as *Patescibacteria*.

Clusters B, C1 and C2, plus many non-clustered MAGs, were broadly similar to each other across carbohydrate/organic-acid pathways and arsenate/mercury degradation: chitin (86–92% of MAGs), polyphenolics (76–94%), pyruvate (62–74%) and arsenate reduction (64–72%) metabolism were all complete in higher proportions than in Cluster A. Non-clustered samples had the highest proportions capable of mixed-linkage glucan (55, 60%) and xylan (47, 51%) metabolism.

In cluster C1 (34 MAGs), more than a third of MAGs contained at least 60% of the genes for either the CBB (23 MAGs, 68%) or rTCA (25 MAGs, 74%) carbon fixation pathways. This included the only complete carbon fixation pathway: all CBB genes present in a *Polaromonas* MAG from GW5TA. Cytochromes were mostly *c* oxidase (28, 82%) and cbb_3 *c* (25, 74%). Dissimilatory nitrate reduction (15, 44%) and sulfur oxidation (10, 29%) proportions were relatively high, but nitrogen fixation was the lowest of all clusters (6, 4%).

In Cluster C2, around half had ≥ 60% of CBB (47, 58%) or rTCA (36, 44%) genes. Cytochromes were detected in 55 MAGs (68%), from cytochrome *o* (30, 37%) to cytochrome *c* oxidase (64, 79%). Nitrogen fixation was highest of all clusters (23, 28%), proportions of sulfur oxidation and sulfate reduction pathways were not noticeably different from other clusters.

Chemistry correlations

Redundancy Analysis (RDA) and Mantel tests of correlation indicated that the groundwater samples were associated with the environmental parameters potassium and sulfate, (both $p = 0.036$), and overall, water chemistry explained 26% of variance across microbial taxa (not shown). Lithology, region and land use were not significantly correlated ($p > 0.25$ for all).

In addition, at family level classification, some MAGs showed high correlation with environmental variables (Fig. 6). Some of these are only found in one cluster. For example, *Brocadiaceae*, *Nitrospiraceae*, *Fredricksoniimonadaceae*, and *Nitrosopumilaceae*, which were associated with nitrate, potassium, and sulfate (all $p < 0.001$), were all only found in cluster A, although one MAG in the non-clustered samples was also classified as *Gemmatimonadaceae*.

Ammonia was correlated with the presence of *Methylomonadaceae* (14 MAGs across 3 clusters, $p < 0.001$). Iron was correlated with the presence of *Gallionellaceae* (6 MAGs, clusters A and B, $p < 0.001$), and the single uncultured *Deferrimicrobiaceae* and *Profunditerraquicolaceae* MAGs (both cluster A, $p < 0.001$).

Asterisks represent significant correlations: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. Only families with at least one correlation of < 0.001 are shown. Correlations were calculated using the R package microeco [32].

DISCUSSION

Each groundwater sample site yielded multiple MAGs (2–59 per site) with taxonomically diverse populations and varied metabolic potential. High biodiversity is often linked to functional redundancy and ecosystem resilience (Fetzer et al., 2015), although environmental changes can lead to a loss of specific ecosystem processes. MAGs clustered into four distinct groups (A, B, C1 and C2) based on sample hydrochemistry, with some influence of location, age and depth.

The clustering of Gisborne and Taranaki samples in Cluster A, despite being on opposite sides of the North Island and having very different lithologies, depths and ages, is intriguing. Elevated calcium, chloride, magnesium, manganese, sodium and sulfate may reflect extensive water-rock interactions or land-use impact, but saltwater intrusion is unlikely as these sites were not coastal. High boron concentrations, primarily in GW233BPHR, may be due to peat and organic material in clay sediments (Gordon, 2002).

This cluster's abundant *Patescibacteria*, consistent with other studies (Tian et al., 2020), routinely co-occurred with one another, or with members of *Micrarchaeota*, *Nanoarchaeota*, *Nitrospirota*, and *Omnitrophota* (Chaudhari et al., 2021, Gios et al., 2023). Their reduced genomes and lack of complete metabolic pathways are consistent with symbiotic or parasitic relationships with bacterial hosts (Srinivas et al., 2024). Their absence of aerobic (cbb_3, *bd*) respiration matches prior reports for *Patescibacteria* and the archaeal superphylum DPANN (*Diapherotrites*, *Parvarchaeota*, *Aenigmarchaeota*, *Nanoarchaeota* and *Nanohaloarchaeota*), with predictions of a fermentative lifestyle for these organisms (Beam et al., 2020).

However, these microbes were all identified in either GW156TN12, GW158TN50 (both Taranaki) or GW211WAFD (Waikato), indicating that while Gisborne samples may have similar hydrochemistry to at least GW150TN50 (Table S1), their microbial communities appear to be quite different. Some inconsistencies may be expected due to aquifer boundaries not matching regional boundaries, as hydrochemistry differs within and across borders.

Interestingly, all samples from the South Island were grouped together in Cluster B, sharing shallow depth and young MRT, despite a range of lithologies and land-uses. Cluster B also had ten *Patescibacteria* MAGs lacking cytochromes and complete metabolic pathways (Tasman and Waikato samples), and the *Moraxellaceae* MAG identified with the same characteristics may represent a host-dependent pathogenic lineage with a streamlined genome (Zhao et al., 2023).

Cluster C2 contained five samples from known geothermal areas with some accepted hydrochemistry signatures, such as elevated arsenic and lithium concentrations. Although geothermal influence could affect microbial community composition, no clear geothermal signal was evident from MAG taxonomy or sulfur oxidation and sulfate reduction pathways. While pH is known to be a strong driver of microbial diversity in geothermal systems (Thompson et al., 2017, Power et al., 2018), the pH range (5.64 – 8.76) and limited sample size in this cluster likely obscure any pH-related patterns. Although geothermal (warm) groundwater was not specifically targeted, there appears to be some geothermal influence on groundwater in some areas, and there is scope for further investigation in this area.

Bioindicators of surface-water exchange such as *Rhodoferax*, capable of phototrophic growth and dark fermentation, though no photosystems were found and plant-associated carbohydrate targets (pectin, polyphenolics, xylans, xyloglucan) that differed significantly between clusters (Table S3) both suggest recharge influence from rivers, streams or overlying wetlands at some sites. All sampling occurred in the drier, warmer December–March period, and proportions of groundwater- and surface-water-associated microbes may shift with wetter winter sampling. Repeated temporal sampling would help identify such changes.

Although only one complete carbon fixation pathway was found, MAGs with potential (≥ 60%) CBB or rTCA gene sets occurred across clusters with differential abundance. The energy-efficient Wood-Ljungdahl pathway, typically predominant in deep subsurface environments (Rogers et al., 2023, Chaudhari et al., 2021), was detected in 3 MAGs — unexpectedly from shallow (3–5 m) bores, perhaps

reflecting local anoxic, low-energy conditions favouring this route (Shoemaker et al., 2024, Gong et al., 2026). Chemolithoautotrophy using nitrate or sulfur/sulfate was more prevalent in Clusters B and C1, while potential endosymbiosis was widespread in Cluster A.

The lack of denitrification and ammonia removal pathways (nitrification or anammox) across all clusters may be of concern. Dissimilatory nitrate reduction pathways, which produce ammonia, were identified in 159/494 MAGs, and several samples showed elevated ammonia concentrations (above the drinking-water standard of 1.5 mg/L) in combination with low dissolved oxygen. Oxic conditions are required for the aerobic process of nitrification (although anammox is an anaerobic process), and the build-up of ammonia also inhibits the final step of denitrification, where that pathway was present (29 MAGs; Liang et al., 2026). This leads to a shift towards heterotrophy, and the degradation of organic matter can lead to further ammonia release. However, the use of DNA in this study does not allow us to distinguish between active, dormant or dead microbes. More precise management of groundwater resources may be achieved by incorporating microbiological monitoring to understand biogeochemical processes, with the potential for biostimulation or bioaugmentation at sites impacted by land-use activities or other forms of contamination.

Community composition correlated significantly with potassium and sulfate ($p < 0.05$), suggesting that unlike much deeper aquifers (average 927 m; (Bas-Lorillot et al., 2025)), pH, temperature and conductivity are not the primary drivers of microbial distribution in these subsurface ecosystems. The correlation between ammonia and *Methylomonadaceae* likely reflects pMMO (particulate methane mono-oxygenase), a promiscuous enzyme that also oxidises ammonia given its structural similarity to methane, as well as many methanotrophs' capacity to use ammonium as a nitrogen source (Whittenbury et al., 1970). Methane, not measured here, may be prevalent in groundwater from biogenic processes or migration of thermogenic methane from deeper aquifers (Wang et al., 2018, Humez et al., 2016). Iron was correlated with the presence of known Fe (II) oxidisers *Gallionellaceae* and uncultured *Deferrimicrobiaceae* and *Profunditerraquicolaceae* MAGs. While not yet grown in isolation, genomic analysis indicates *Deferrimicrobiaceae* at least may have the capacity for dissimilatory iron reduction (Begmatov et al., 2022).

This study has a relatively small number of sample sites across the country, and future work will include expanding the dataset to cover a wider range of environmental parameters as well as lithology and land-use to fully understand contributions to microbial diversity and function.

Our results provide insights into microbial biogeochemical cycling within Aotearoa New Zealand's subsurface ecosystems. We conclude that these ecosystems support multiple metabolic processes and associated autotrophic pathways and oxidative phosphorylation mechanisms, closely associated with environmental parameters such as aquifer geochemistry. Microbial diversity could serve as an indicator of groundwater ecosystem condition and help predict responses to contamination or environmental change such as temperature fluctuations: high biodiversity with duplicated or overlapping functional roles confers resilience, while this study shows that some low-diversity aquifers are likely more

vulnerable to contaminant pressure or other change. As human impacts from agriculture and wastewater are projected to increase with climate change, these findings highlight a critical need to investigate and understand aquifer biogeochemical cycling to aid management of these critical resources.

Declarations

1. Data Availability Statement

Raw Illumina sequence reads are available at NCBI SRA under BioProject PRJNA1230396.

2. Acknowledgments

We thank the staff of the following councils who kindly collected groundwater samples for analysis: Auckland Council, Bay of Plenty Regional Council, Environment Canterbury, Gisborne District Council, Hawke's Bay Regional Council, Marlborough District Council, Northland Regional Council, Otago Regional Council, Taranaki Regional Council, and Tasman District Council. Computational resources were provided by New Zealand eScience Infrastructure (NeSI).

3. Funding

This project was supported by the New Zealand Ministry of Business, Innovation & Employment (MBIE) through the Strategic Science Investment Fund Groundwater programme (contract C05X1702) and Nationally Significant Collections and Databases programme (contract C05X1701).

4. Conflict of Interest Statement

“The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.”

5. Ethics statement

The content and authorship of the submitted manuscript have been approved by all authors, and all prevailing local, national and international regulations and conventions, and normal scientific ethical practices, have been respected. No generative AI was used in the production of this manuscript.

6. Author Contributions

KMH: Conceptualisation, Methodology, Investigation, Formal Analysis, Visualization, Writing – original draft, Writing – review & editing.

MM: Formal Analysis, Project administration, Writing – review & editing.

CT: Funding acquisition, Project administration, Writing – review & editing.

References

1. Bas-Lorillot J, Ménez B, Wild B, Borrel G, Le Bihan M, Stefánsson A, Gérard E (2025) Groundwater Microbial Diversity Associated With Icelandic Basaltic Subsurface Environments. Environ Microbiol Rep 17. https://doi.org/10.1111/1758-2229.70238
2. Beam JP, Becraft ED, Brown JM, Schulz F, Jarett JK, Bezuidt O, Ravin NV (2020) Ancestral Absence of Electron Transport Chains in *Patescibacteria* and DPANN. Front Microbiol 11. https://doi.org/10.3389/fmicb.2020.01848
3. Begmatov S, Beletsky AV, Dedysh SN, Mardanov AV, Ravin NV (2022) Genome analysis of the candidate phylum MBNT15 bacterium from a boreal peatland predicted its respiratory versatility and dissimilatory iron metabolism. Front Microbiol 13. https://doi.org/10.3389/fmicb.2022.951761
4. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Woyke T (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. https://doi.org/10.1038/nbt.3893
5. Chaudhari NM, Overholt WA, Figueroa-Gonzalez PA, Taubert M, Bornemann TLV, Probst AJ, Küsel K (2021) The economical lifestyle of CPR bacteria in groundwater allows little preference for environmental drivers. Environ Microbiome 16:24. https://doi.org/10.1186/s40793-021-00395-w
6. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH (2019) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848
7. Daughney CJ, Jones A, Baker T, Hanson C, Davidson P, Zemansky G, Thompson M (2006) A National Protocol for State of the Environment Groundwater Sampling in New Zealand. Ministry for the Environment,, Wellington
8. Daughney CJ, Reeves RR (2005) Definition of hydrochemical facies in the New Zealand National Groundwater Monitoring Programme. J Hydrology New Z 44:105–130
9. Davies H (2001) Groundwater and Health. In: Rosen MR, White PA (eds) Groundwaters of New Zealand. New Zealand Hydrological Society Inc, Wellington, New Zealand, *In*:
10. Fetzer I, Johst K, Schawea R, Banitz T, Harms H, Chatzinotas A (2015) The extent of functional redundancy changes as species' roles shift in different environments. Proc Natl Acad Sci USA 112:14888–14893. https://doi.org/10.1073/pnas.1505587112
11. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Wipat A (2008) The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 26:541–547. https://doi.org/10.1038/nbt1360
12. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci USA 103:626–631. https://doi.org/10.1073/pnas.0507535103
13. Gios E, Mosley OE, Weaver L, Close M, Daughney C, Handley KM (2023) Ultra-small bacteria and archaea exhibit genetic flexibility towards groundwater oxygen content, and adaptations for

attached or planktonic lifestyles. ISME Commun 3:13. https://doi.org/10.1038/s43705-023-00223-x

14. Gordon D (2002) Groundwater Resources of the Bay of Plenty. Environment Bay of Plenty
15. Griebler C, Avramov M (2015) Groundwater ecosystem services: A review. Freshw Sci 34:355–367. https://doi.org/10.1086/679903
16. Houghton KM (2025) Groundwater Ecosystem Monitoring: Microbial diversity and function, Report 2. Lower Hutt (NZ),: GNS Science,. https://doi.org/10.21420/GET6-YX26
17. Houghton KM, Santamaria Cerrutti CE (2022) Groundwater Ecosystem Monitoring: Microbial diversity and function, Annual Report 2022. Lower Hutt, New Zealand: GNS Science,. https://doi.org/10.21420/E0ZE-N191
18. Humez P, Mayer B, Nightingale M, Becker V, Kingston A, Taylor S, Kloppmann W (2016) Redox controls on methane formation, migration and fate in shallow aquifers. Hydrol Earth Syst Sci 20:2759–2777. https://doi.org/10.5194/hess-20-2759-2016
19. Imhoff JF, Rahn T, Künzel S, Neulinger SC (2019) Phylogeny of anoxygenic photosynthesis based on sequences of photosynthetic reaction center proteins and a key enzyme in bacteriochlorophyll biosynthesis, the chlorophyllide reductase. Microorganisms 7. https://doi.org/10.3390/microorganisms7110576
20. Kassambara A (2023) rstatix: Pipe-Friendly Framework for Basic Statistical Tests. 0.7.2 ed
21. KC B (2019) New Zealand Groundwater Atlas: Groundwater use and drinking water supply wells serving more than 100 people. *Water Management Report.* Ministry for the Environment
22. Liang Z, Liu W, Yang Z, Liang Z, Province S, China PR (2026) Appl Water Sci, 16. https://doi.org/10.1007/s13201-026-02823-z
23. Liu C, Cui Y, Li X, Yao M (2020) microeco: an R package for data mining in microbial community ecology. FEMS Microbiol Ecol 97:9. https://doi.org/10.1093/femsec/fiaa255
24. Milne J, Patterson M, Davies-Colley R, Robinson P, Donald R, Moreau M, Ausseil O (2019) Water Quality: Part 1 of 4: Sampling, Measuring, Processing and Archiving of Discrete Groundwater Quality Data. 1.0.0 ed. Ministry for the Environment,, Wellington
25. Ministry for the Environment &, Stats NZ (2026) New Zealand's Environmental Reporting Series: Our Freshwater 2026 |Tō Tātou Wai Māori. Ministry for the Environment,, Wellington, NZ
26. Moreau M, Herpe M, Santamaria Cerrutti M (2025) 2024 Update of the national groundwater quality indicator. GNS Science, Wairakei (NZ)
27. Mosley OE, Gios E, Close M, Weaver L, Daughney C, Handley KM (2022) Nitrogen cycling and microbial cooperation in the terrestrial subsurface. ISME J 16:2561–2573. https://doi.org/10.1038/s41396-022-01300-0
28. Power JF, Carere CR, Lee CK, Wakerley GLJ, Evans DW, Button M, Stott MB (2018) Microbial biogeography of 925 geothermal springs in New Zealand. Nat Commun 9. https://doi.org/10.1038/s41467-018-05020-y

29. R Core Team (2025) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
30. Rogers TJ, Buongiorno J, Jessen GL, Schrenk MO, Fordyce JA, de Moor JM, Lloyd KG (2023) Chemolithoautotroph distributions across the subsurface of a convergent margin. ISME J 17:140–150. https://doi.org/10.1038/s41396-022-01331-7
31. Shaffer M, Borton MA, McGivern BB, Zayed AA, Rosa L, Solden SL, Wrighton LM, K. C (2020) DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res 48:8883–8900. https://doi.org/10.1093/nar/gkaa621
32. Sirisena KA, Daughney CJ, Moreau M, Sim DA, Lee CK, Cary SC, Chambers GK (2018) Bacterial bioclusters relate to hydrochemistry in New Zealand groundwater. FEMS Microbiol Ecol 94. https://10.1093/femsec/fiy170
33. Srinivas P, Brook Peterson S, Gallagher LA, Wang Y, Mougous JD (2024) Beyond genomics in Patescibacteria: A trove of unexplored biology packed into ultrasmall bacteria. Proc Natl Acad Sci USA 121. https://doi.org/10.1073/pnas.2419369121
34. Taylor RG, Scanlon B, Döll P, Rodell M, van Beek R, Wada Y, Treidel H (2013) Ground water and climate change. Nat Clim Change 3:322–329. https://doi.org/10.1038/nclimate1744
35. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Zhao H (2017) A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457. https://doi.org/10.1038/nature24621
36. Tian R, Ning D, He Z, Zhang P, Spencer SJ, Gao S, Zhou J (2020) Small and mighty: adaptation of superphylum *Patescibacteria* to groundwater environment drives their genome simplicity. Microbiome 8:51. https://doi.org/10.1186/s40168-020-00825-w
37. Wang ZP, Zhang L, Wang B, Hou LY, Xiao CW, Zhang XM, Han XG (2018) Dissolved methane in groundwater of domestic wells and its potential emissions in arid and semi-arid regions of Inner Mongolia, China. Sci Total Environ 626:1193–1199. https://doi.org/10.1016/j.scitotenv.2018.01.205
38. Ward JH, Hook ME (1963) Application of an hierarchical grouping procedure to a problem of grouping profiles. Educ Psychol Meas 23:69–81. https://doi.org/10.1177/001316446302300107
39. Watson SJ, Arisdakessian C, Petelo M, Keliipuleole K, Tachera DK, Okuhata BK, Frank KL (2023) Geology and land use shape nitrogen and sulfur cycling groundwater microbial communities in Pacific Island aquifers. ISME Commun 3. https://doi.org/10.1038/s43705-023-00261-5
40. White PA, Tschritter C, Rawlinson Z, Moreau M, Dewes K, Edbrooke S (2015) Groundwater resource characterisation in the Waikato River catchment for Healthy Rivers project. Waikato Regional Council
41. Whittenbury R, Phillips KC, Wilkinson JF (1970) Enrichment, isolation and some properties of methane-utilizing bacteria. J Gen Microbiol 61:205–218. https://doi.org/10.1099/00221287-61-2-205
42. Yurkov V, Messner K (2025) Phenomenal Diversity of the Photosynthetic Apparatus Evolved in Aerobic Anoxygenic Phototrophs. Microorganisms 13.

https://doi.org/10.3390/microorganisms13112446

43. Zhao Y, Wei HM, Yuan JL, Xu L, Sun JQ (2023) A comprehensive genomic analysis provides insights on the high environmental adaptability of *Acinetobacter* strains. Front Microbiol 14. https://doi.org/10.3389/fmicb.2023.1177951

Figures

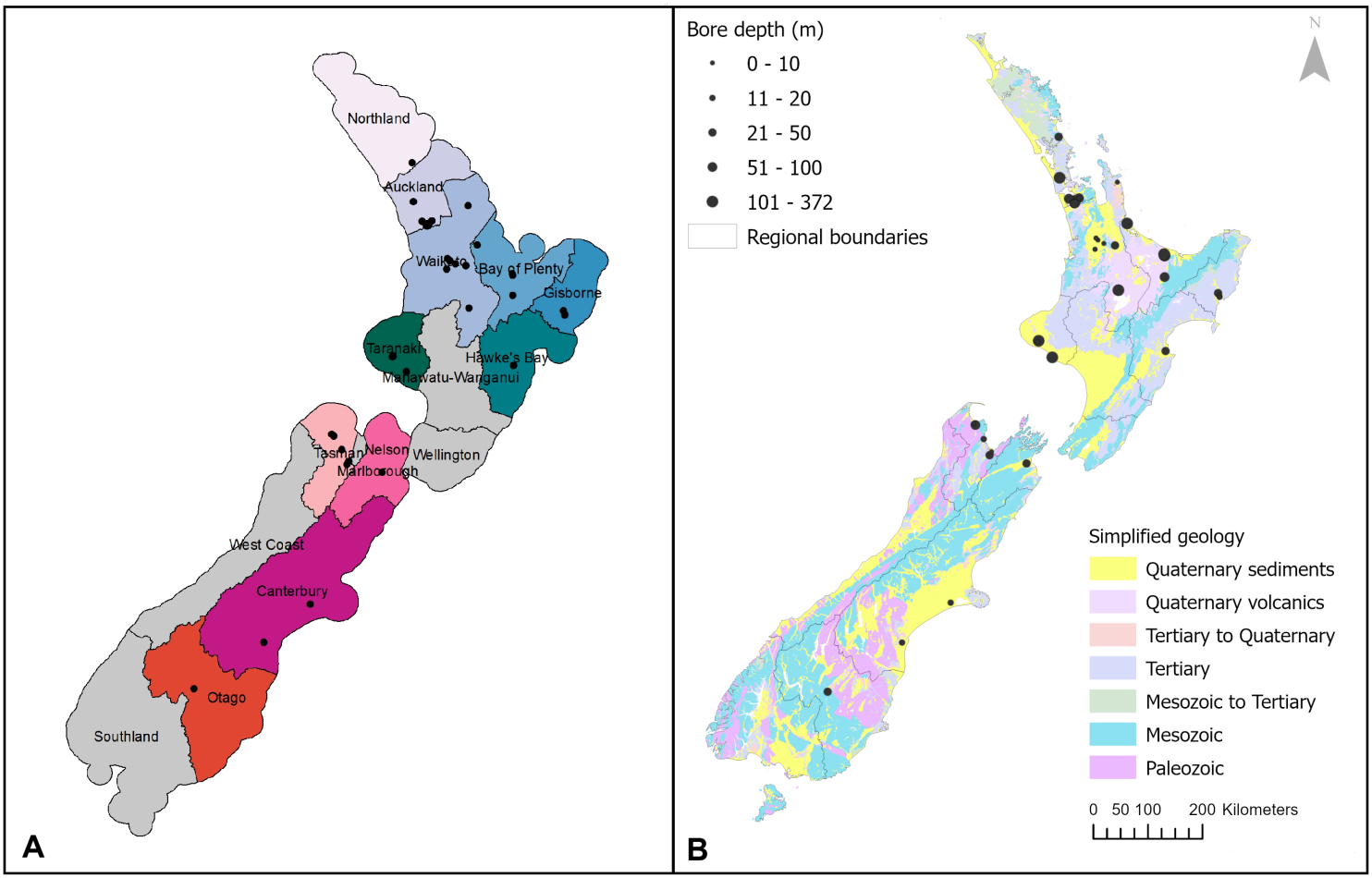


Figure 1

Maps of sample locations.

A) Regional boundaries of New Zealand, with locations of samples. Regions are coloured if at least one sample was taken. B). Simplified geology and bore depth of samples.

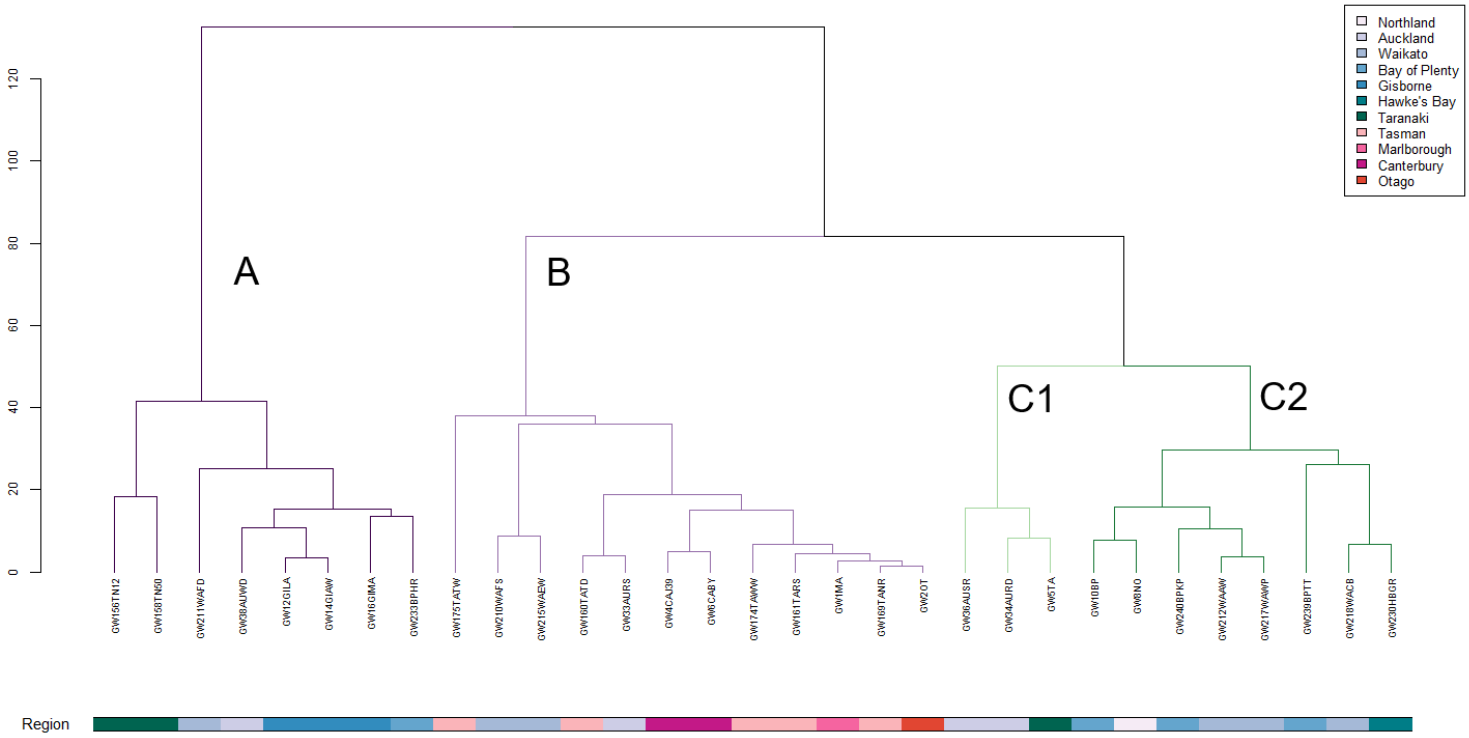


Figure 2

Dendrogram produced by hierarchical cluster analysis (HCA) for all groundwater samples, based on hydrochemistry.

The vertical axis represents the distance or dissimilarity between clusters. Regions are coloured as in Figure 1 (North Island blues and greens, South Island reds and pinks).

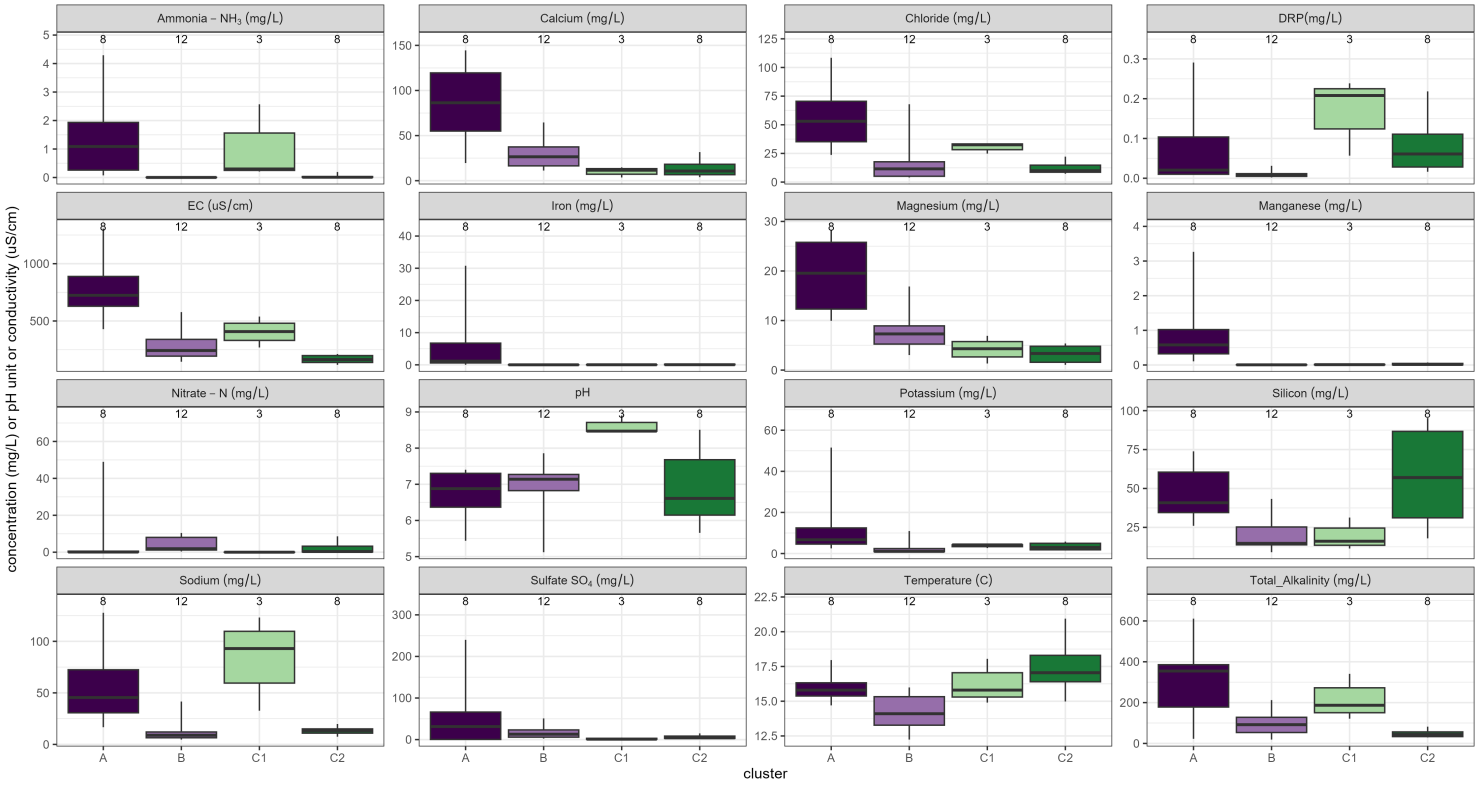


Figure 3

Box plots of hydrochemistry parameters for all samples included in HCA.

The number displayed above each box plot indicates the sample size. The horizontal lines in the middle of each box are median values, and the upper and lower limit of the box are the 75th and 25th percentiles, respectively. The vertical lines extend from the 95th to the 5th percentiles.



Figure 4

Heatmap of metabolic profiling of MAGs within HCA clusters (carbon, nitrogen and sulfur).

The heatmap represents metabolic pathway completeness of the MAGs based on the presence or absence of genes (as Kegg Orthologs) identified using the DRAM package (Shaffer, 2020), with the median for each sample shown. Results for individual MAGs are shown in Table S4, and statistical test results in Table S3.

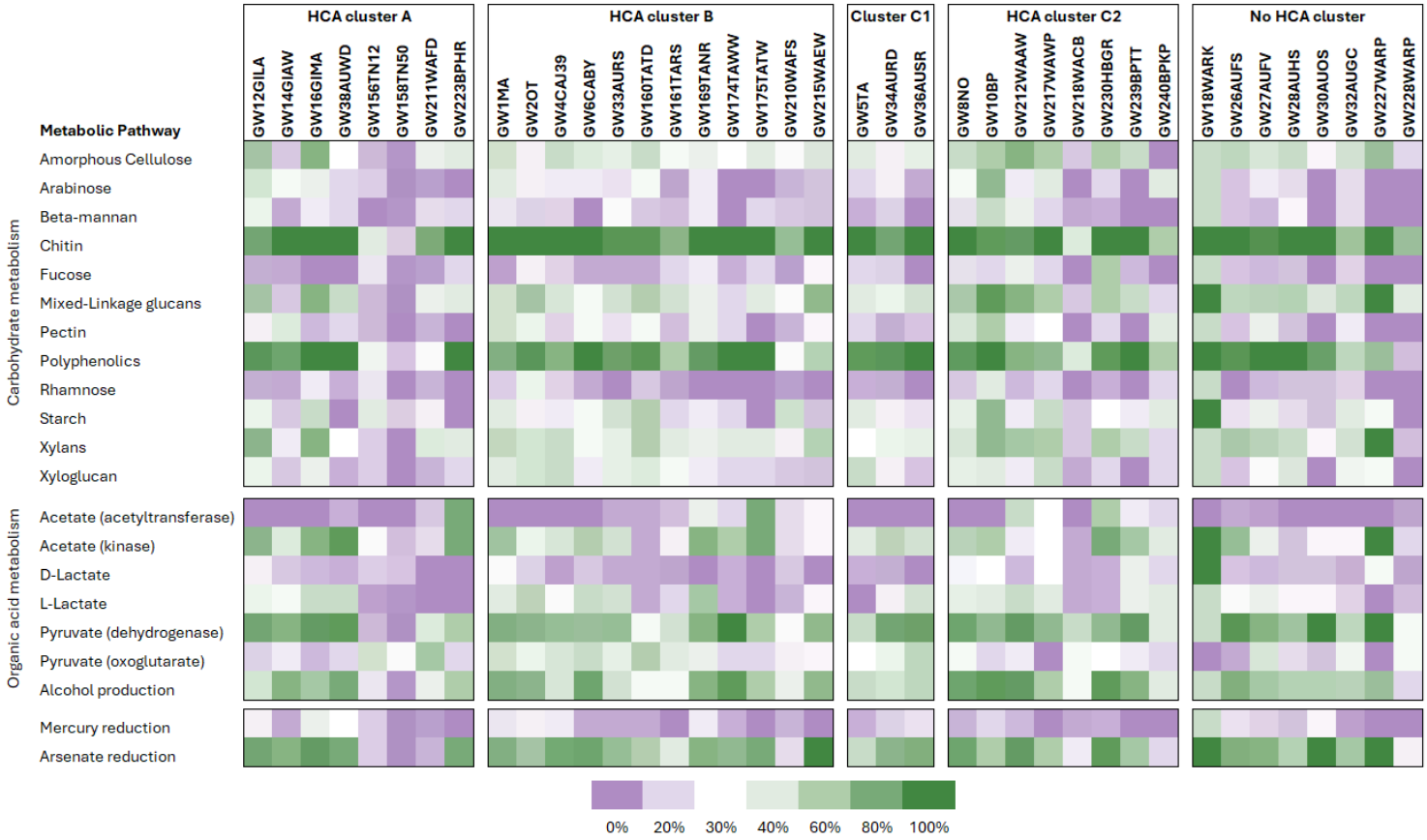


Figure 5

Heatmap of metabolic profiling of MAGs within HCA clusters (carbohydrates and organic acids).

The heatmap represents the percentage of MAGs within each sample which contain metabolic pathways identified using the DRAM distillation package (Shaffer, 2020). Results for individual MAGs are shown in Table S5, and statistical test results in Table S3.

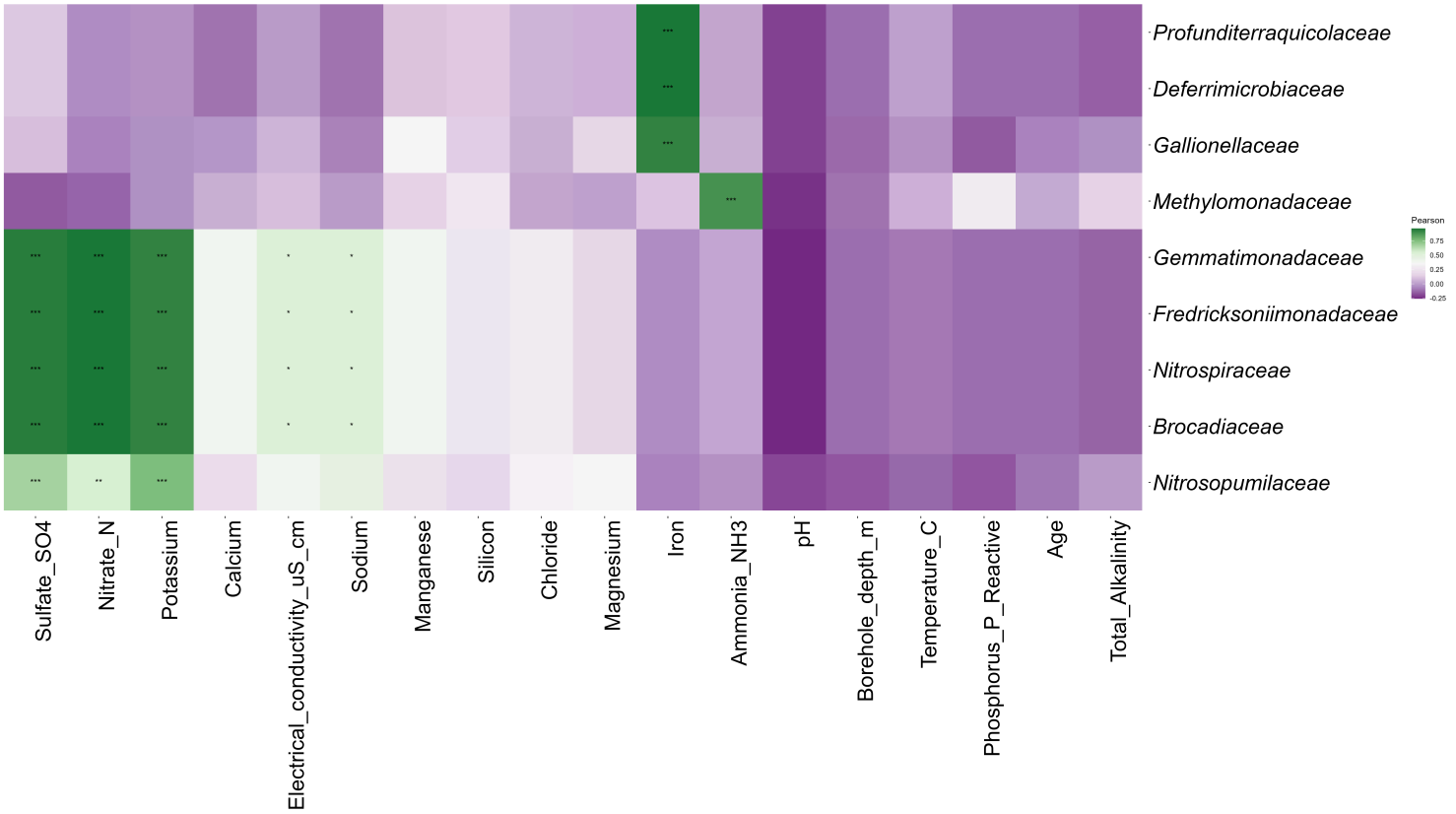


Figure 6

Correlation heat map between hydrochemistry parameters and individual MAGs, grouped by taxonomic family.

Asterisks represent significant correlations: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. Only families with at least one correlation of < 0.001 are shown. Correlations were calculated using the R package microeco [32]

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.xlsx