Total Environment Microbiology 2 (2026) 100074

Contents lists available at ScienceDirect

Total Environment Microbiology

journal homepage: www.elsevier.com/locate/temicr

Abstraction method shapes microbial ecology and pathogen risk in water from shallow aquifers

Reuben Duncan [a], Umer Zeeshan Ijaz [b], Fabien Cholet [b], Ciara Keating [b,c], Orges Koci [b], Ruth Quinn [d], Julie Russell [b], Francis Hassard [a], Alison Parker [a,*], Cindy J. Smith [b]

[a] Faculty of Engineering and Applied Sciences, Cranfield University, Vincent Building, Cranfield, MK 43 0AL, UK
[b] James Watt School of Engineering, University of Glasgow, Advanced Research Centre, Glasgow, G11 6EW, UK
[c] Department of Engineering, Durham University, Durham DH1 3LE, UK
[d] Department of Civil Engineering and Construction Studies, Atlantic Technological University, Sligo F91 YW50, Ireland

ARTICLE INFO

Keywords:
Microbial community assembly
Hand pumps
Scoop holes
Pathogen
16S rRNA Gene Sequencing
Shallow Groundwater System

ABSTRACT

Shallow groundwater in semi-arid regions provides a critical year-round water source, naturally filtered and protected from evaporation. However, many communities continue to suffer from preventable diseases caused by waterborne pathogens. This study applies ecological theory to investigate the microbial ecology of shallow aquifers in southeast Kenya. We used quantitative PCR and 16S rRNA gene sequencing to characterise pathogenic indicator species and bacterial communities in three isolated sand dam aquifers across seasons and between various abstraction methods used by local communities, including sealed hand pumps (classified by the WHO as an improved source), open wells, scoop holes, and downstream surface water.

Both faecal and opportunistic environmental pathogens were detected in all samples. However, our models indicate that contamination largely originates from unsealed abstraction points such as open wells and cracked hand pumps which expose water to the environment. In contrast, natural sand filtration effectively limited the dispersal of species between sample points.

Abstraction method was the dominant factor shaping microbial community structure, exceeding seasonal or locational effects. Unimproved sources exhibited unstable, stressed ecosystems with reduced diversity and resilience, consistent with strong environmental selection pressures. Across all abstraction points, community composition reflected adaptation to local environmental niches that remained consistent between sites and seasons. Groundwater contained significantly fewer total microbes (16S rRNA copies) than surrounding surface water, and hand pump samples had the lowest abundance of pathogen indicators, suggesting growth-limiting conditions and effective in-dam filtration.

Our findings reinforce the value of sealed, sanitary hand pumps for water abstraction from shallow sand aquifers. Ensuring robust construction and regular maintenance of such hand pumps is important for low-income regions relying on shallow groundwater for water with minimal secondary treatment. Potential contamination routes identified at all sites indicate the need for targeted infrastructure improvements and continual maintenance.

1. Introduction

Limited access to clean water, sanitation, and hygiene (WASH) remains a major global health challenge that is responsible for around 1.4 million deaths and 74 million disability-adjusted life years (DALYs) annually. Half a billion people worldwide use unimproved water sources, around 180 million of whom are located in rural sub-saharan Africa – often in semi-arid, water-stressed climates (WHO and UNICEF, 2019). Second-order effects on communities can include chronic poverty, lower educational achievement, social conflict, and increased vulnerability to climate change (Wolf et al., 2023). Groundwater can provide a reliable supply since it is largely protected from evaporation and pathogen contamination. However, the use of unimproved abstraction points such as open wells, scoop holes, and poorly maintained hand pumps can expose groundwater to the environment and open pathways for pathogens to enter water bodies (Parker et al., 2010; Quinn et al., 2018;

* Corresponding author.
E-mail address: a.parker@cranfield.ac.uk (A. Parker).

https://doi.org/10.1016/j.temicr.2026.100074
Received 6 November 2025; Received in revised form 23 January 2026; Accepted 6 March 2026
Available online 10 March 2026
3050-6417/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Sutton, 2002).

The application of microbial ecological frameworks can yield valuable insights into groundwater systems, indicating for example the prevalence of pathogenic species, antimicrobial resistance genes, or pathogenicity factors (Acharya et al., 2020; Semenza, 2020; Sterk et al., 2013; Wahome et al., 2014). Ecological models can be used to describe the structure of microbial communities, including the diversity (which reflects the complexity of metabolic processes and environmental niches) and evenness (which reflects the balance of competition, dominance, or disturbance within the community), and the evolution of communities with respect to their environment. These models can be applied to environmental systems to elucidate microbial source tracking (Gayner, 2018; McGhee et al., 2020), nutrient cycling (Zhang et al., 2024), and community assembly processes (Chase and Myers, 2011; Sloan et al., 2006; Stegen et al., 2012; Zhou and Ning, 2017).

This study addresses a critical gap in our understanding of groundwater microbiology by investigating how community practices shape microbial ecology in shallow aquifers. We investigated several groundwater systems in rural southeast Kenya, which is a semi-arid region known for an abundance of engineered water harvesting structures known as sand dams. These promote groundwater recharge by capturing sand behind a dam wall constructed across a seasonal river, thus creating an artificial shallow aquifer in the riverbed (Lasage et al., 2008; Quinn et al., 2019). Successful sand dams capture sufficient water during the wet seasons to provide local communities with a cost-effective, year-round water supply, and provide ideal examples of shallow groundwater systems due to well-characterised recharge patterns and relative isolation from larger systems (Ritchie et al., 2021). However, previous work has identified a number of contamination risks around the sites, including nearby livestock and agriculture, stagnant water, and damaged drainage channels (Quinn et al., 2018). We apply ecological theory in this context in order to explore how the structure and usage patterns of groundwater systems impact microbial community composition, contamination pathways, and potential health risks, thereby enabling proactive, evidence-based risk assessment and water safety planning for rural groundwater systems.

We tested three linked hypotheses: that sealed and protected abstraction points (hand pumps) present with a lower abundance of faecal and opportunistic pathogen gene markers and lower total bacterial loads than unprotected abstraction points (open wells and scoop holes) and downstream surface water; that abstraction method explains more variation in microbial community composition and diversity than season or site; and that exposed abstraction points show stronger deterministic assembly and phylogenetic clustering than protected points, consistent with stronger environmental selection and disturbance.

Findings from this work will guide the development of targeted engineering interventions, including improved abstraction infrastructure and management practices, ultimately supporting safer and more sustainable groundwater supplies for vulnerable rural communities.

2. Materials and methods

2.1. Sample sites

The study was conducted in Makueni County, Kenya, a semi-arid region characterised by two distinct wet seasons (March–May and November–December), each averaging approximately 150 mm of rainfall per month, separated by prolonged dry intervals (CHIRPS, 2021). The area experiences annual average temperatures ranging between 20 and 30 °C and faces persistent challenges associated with water scarcity. Sand dams have been constructed across ephemeral riverbeds to improve community access to water, storing seasonal runoff in accumulated sand sediments. Water is typically abstracted from hand dug wells fitted with sealed hand pumps or via open, shallow "scoop holes" dug directly by communities. In some locations, water also leaks under dam walls, creating downstream pools that serve as supplementary water sources. Water from sand dams is an improved water source when abstracted via a sealed hand pump with well-maintained concrete apron and effective drainage system (WHO, 2017). Conversely, contamination has been observed in up to 89 % of samples from unimproved abstraction points such as open wells, scoop holes, and poorly maintained hand pumps (Quinn et al., 2018).

The three sand dams examined in this study were selected because they were representative structures built by the African Sand Dam Foundation, contained both sand and water at the end of the wet season, and were accessible for daily monitoring from our base in Mtito Andei, approximately 30 km away. The dams, range from 250 to 480 m in length and 3 to 5 m in height, were situated in areas primarily used for small-scale agriculture and grazing. They were located at the following coordinates, as shown in Fig. 1:

Dam 106: Latitude 2.5504° S, Longitude 38.0424° E
Dam 167: Latitude 2.6033° S, Longitude 38.1021° E
Dam 211: Latitude 2.5690° S, Longitude 38.0954° E

2.2. Sample collection

Samples were collected from the three sites in Makueni County, Kenya, during the early dry season (May) and the middle of the dry season (July) of 2017 to see how the microbial ecology evolved through the season (Table 1). Sample points at each location were included to cover the main abstraction methods that local communities use: sealed hand pumps (improved sources), open wells, community scoop holes (maintained seasonally by the community), downstream surface water pools, and excavated scoop holes - temporary sampling points dug by researchers using sterile equipment. At each point, water samples (500 ml to 2 L depending on availability) were collected and immediately filtered through sterile 0.2 μm pore-size cellulose nitrate membrane filters (Thermo Scientific, USA). Filters were stored in DNA/RNA Shield (Zymo Research, USA) and shipped to the UK where they were stored at −80 °C prior to DNA extraction. To prevent contamination, all non-sterilised sampling equipment was sterilised with 70 % ethanol before use, and field personnel wore disposable nitrile gloves throughout sample handling and processing. Physicochemical parameters including temperature, pH, electrical conductivity, and turbidity were measured in situ using a calibrated multiparameter probe (YSI Inc., USA). Sand grain size was measured using a stacked sieve.

2.3. DNA extraction and sequencing

Total genomic DNA was extracted from frozen membrane filters using the PowerWater® DNA Isolation Kit (Qiagen, Germany), following the manufacturer's standard protocol. DNA concentration and purity were assessed using both a Qubit 4 fluorometer (Thermo Fisher Scientific, USA) and a NanoDrop One spectrophotometer (Thermo Fisher Scientific, USA).

16S rRNA amplicon libraries were prepared for microbial community characterization using the Illumina Nextera XT DNA Library Preparation Kit (Illumina, USA), following the manufacturer's protocol. The V4 hypervariable region of the 16S rRNA gene was amplified using primers 515F (5′-GTGCCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACHVGGGTWTCTAAT-3′), incorporating Illumina-specific overhang adapters. PCR amplification was carried out under the following optimised conditions: initial denaturation at 95 °C for 3 minutes; followed by 25 cycles at 95 °C for 30 seconds, 55 °C for 30 seconds, and 72 °C for 30 seconds; and a final extension at 72 °C for 5 minutes. Library quality and fragment size were evaluated prior to sequencing using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Sequencing was performed on the Illumina MiSeq platform (Illumina, USA), generating paired-end reads of 2 × 300 bp length.

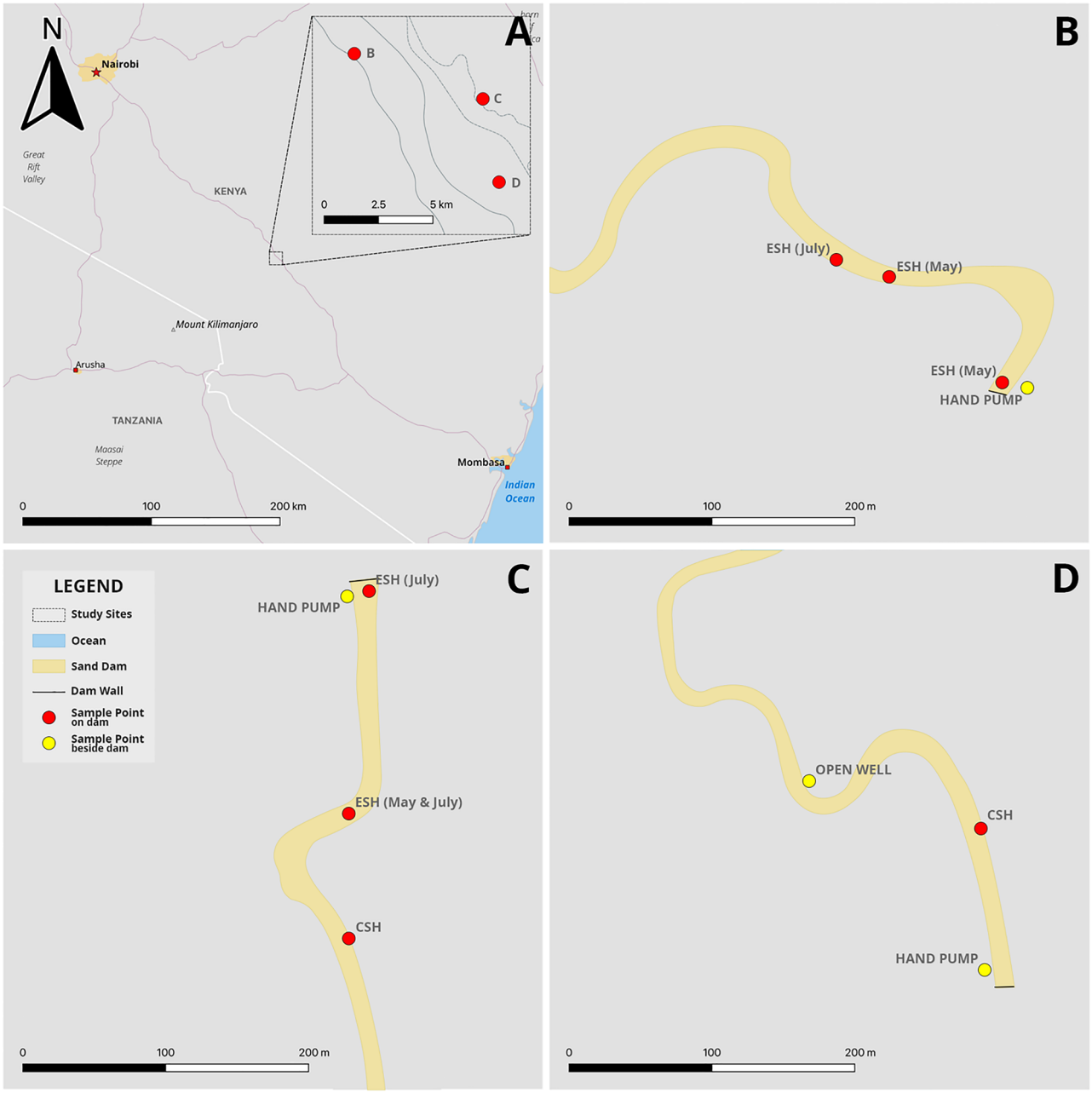

Fig. 1. Sample Point and Site Locations.
Samples were collected from three sites in Makueni County, Kenya, across two site visits in May and July of 2017.
A: regional map of site locations; B: sand dam 1 (ID 106); C: sand dam 2 (ID 167); D: sand dam 3 (ID 211); CSH: community scoop hole; ESH: excavated scoop hole.

2.4. Microbial water quality

Quantitative PCR (qPCR) assays were carried out using the Quanti-Tect Probe Kit (Qiagen) in a 25 µl final volume to elucidate the copy number for total bacteria (16S rRNA gene [1369F/1492R/1389P]), coliform bacteria (lacZ gene [lacZF/lacZR/lacZP]), *Escherichia coli* (rodA gene [rodAF/rodAR/rodAP]), Legionella (ssrA gene [lsppF/lsppR/lsppP]), Mycobacterium (atpE gene [FatpE/RatpE/PatpE]), and Pseudomonas aeruginosa (gyrB gene [PgyrBF/PgyrBR/PgyrBP]) (Anuj et al., 2009; Benitez and Winchell, 2013; Chern et al., 2011; Radomski et al., 2013; Thom et al., 2025). TaqMan probes with a 5′ FAM fluorophore were used for the 16S, rodA, ssrA, atpE, and gyrB assays, while SYBR Green was used for the coliform lacZ assay. Probe and primer sequences with reaction volumes are included in Supplementary Table 1. Thermal cycling conditions included an initial denaturation at 95 °C for 15 min, followed by 40 cycles of 94 °C for 10 s and 60 °C annealing for 60 s. Assay efficiencies were calculated from standard curves generated for each target gene using serial dilutions of G-blocks (Integrated DNA Technologies, UK) containing gene inserts flanked by an additional M13-T7 primer sequences ranging from 10^1 to 10^8 copies per reaction. The lowest calibrated standard (10^1 copies per reaction) therefore provided a practical limit of quantification for each assay; targets were considered not quantified when amplification was absent or below this range. Gene copy numbers in each sample were quantified by comparing mean threshold cycle (Ct) values to the respective standard curves. Quantitative PCR results were standardised according to sample volume, while sequencing data were normalised by equimolar amounts of input DNA. Final bacterial counts were adjusted for total DNA concentration and water volume collected to yield bacterial counts per 100 ml.

Table 1
Sample point details.

Dam Number	SD1 (106)	SD2 (167)	SD3 (211)
Name	Ngulai	Athiani Farmers	Kipico Self Help Group
Year of construction	2013	2014	2015
Location (lat/long)	−2.5504 38.0424	−2.6033 38.1021	−2.5690 38.0954
Length monitored (m) (from final piezometer to dam wall)	250	480	380
Sample collection points			
Excavated scoop hole	May, July	May, July	May
Community scoop hole		May, July	May, July
Hand pump	May, July	May, July	May, July
Open well	May, July		
Downstream	July	May, July	

Samples were collected from three sites in Makueni County, Kenya in May and July of 2017.
SD1: sand dam 1 (ID 106); SD2: sand dam 2 (ID 167); SD3: sand dam 3 (ID 211).

2.5. Bioinformatics

Raw paired-end 16S rRNA gene sequencing reads were first subjected to error correction using BayesHammer within SPAdes v2.5.0 (Nikolenko et al., 2013), producing corrected forward and reverse reads. Corrected reads were then merged with PANDAseq v2.4, which aligns overlapping paired-end reads to generate high-quality consensus sequences (Masella et al., 2012). The merged sequences were subsequently processed with DADA2 v1.26.0, which performs quality filtering, denoising, and removal of chimeric sequences (Callahan et al., 2016). This workflow resulted in the generation of amplicon sequence variants (ASVs), enabling single-nucleotide resolution in downstream analyses of microbial community composition.

Taxonomic assignment of ASVs was performed using the SILVA reference database v138 (Quast et al., 2013) with a minimum bootstrap confidence threshold of 80 %. A multiple sequence alignment was constructed using Kalign v2.04 (Lassmann and Sonnhammer, 2005), and a phylogenetic tree was generated using FastTree v2.1.7 (Price et al., 2010). In total, 26,916 ASVs were identified across 67 samples, with sequencing depth ranging from 7265 to 185,204 reads per sample (median: 46,120; mean: 50,527). Processed sequencing outputs (ASV abundance table and taxonomic assignments) are available in a public repository on figshare (DOI: 10.6084/m9.figshare.30218107). Raw paired-end 16S rRNA gene amplicon sequencing reads are available in the NCBI SRA under accession number PRJNA1333909.

2.6. Alpha and beta diversity

All statistical analyses were performed using R software (v4.3.2; R Core Team, 2023). Alpha diversity was assessed for each sample by calculating rarefied species richness and evenness (Shannon entropy) using the vegan R package (Oksanen et al., 2007). Richness values represent counts of observed ASVs at the rarefaction depth and are reported as whole numbers for clarity. Beta diversity, reflecting differences in microbial communities between samples, was calculated using three complementary metrics: (1) Bray-Curtis, which assesses dissimilarity based solely on taxa abundance counts; (2) Unweighted UniFrac, which measures phylogenetic differences based on presence or absence of taxa; and (3) Weighted UniFrac, which integrates both phylogenetic distances and relative abundances, thus emphasizing dominant taxa. Beta diversity analyses were performed using the phyloseq package (McMurdie and Holmes, 2013). Sources of community variation were statistically evaluated using permutational multivariate analysis of variance (PERMANOVA) with the vegan adonis() function (Oksanen et al., 2007). This method applies linear models to distance matrices (Bray-Curtis, Unweighted UniFrac, and Weighted UniFrac) and utilises permutation tests with pseudo-F ratios to determine significant sources of variation.

To quantify the role of environmental filtering in shaping microbial communities (phylogenetic overdispersion versus clustering), phylogenetic relationships within each sample were characterised by calculating the nearest taxon index (NTI) and the net relatedness index (NRI). This approach helps discern whether community structure is predominantly stochastic (overdispersion, indicating random assembly) or deterministic (clustering, driven by environmental selection pressures). The NTI was calculated using mntd() and ses.mntd() functions, and NRI was calculated using mpd() and ses.mpd() functions from the picante package (Kembel et al., 2010). NTI and NRI represent the negative outputs from ses.mntd() and ses.mpd(), respectively, quantifying how many standard deviations the observed values differ from the mean of a null distribution generated through 999 randomizations (null.model = 'richness'). For these calculations, only the top 1000 most abundant ASVs were included, based on established recommendations (Stegen et al., 2012).

2.7. Core microbiome

The core microbiome was defined as the taxa that exert the greatest ecological influence on microbial community structure and function, and which are therefore important in shaping the overall microbial dynamics (Shade and Stopnisek, 2019). To identify core microbiome taxa at each abstraction point, taxa were rank-ordered based on occupancy (the frequency of occurrence across samples). We then quantified their relative contribution to beta diversity (Bray-Curtis dissimilarity), selecting only those taxa whose inclusion provided at least a 2 % increase in explanatory power. In this manner, the core microbiome for each abstraction point comprised taxa characterised by both high occupancy and substantial influence on community composition, providing insight into stable and functionally important microbial groups. Core microbiome analyses were performed using the microbiome R package (Lahti and Shetty, 2017).

2.8. Community assembly processes

To elucidate microbial community assembly mechanisms, we applied two complementary null-modelling approaches that quantify the relative contributions of stochastic and deterministic processes. Firstly, we utilised the normalised stochasticity ratio (NST) method (Ning et al., 2019), which generates a stochasticity ratio (SR) for each community. The SR indicates the relative proportion of stochastic (ecological drift and random dispersal) versus deterministic processes (environmental selection) shaping each community. NST calculations involved 1000 randomizations utilizing both Ruzicka and Jaccard distance metrics. Taxon occurrence frequencies were constrained proportionally, with richness constrained using both proportional and fixed approaches.

Secondly, we employed the Quantitative Process Estimates (QPE) approach (Stegen et al., 2013), which classifies microbial assembly mechanisms into five distinct ecological processes: (i) variable selection, where changing selective environmental conditions result in high community turnover; (ii) homogenous selection, where stable environmental pressures consistently shape community composition; (iii) dispersal limitation, where limited microbial migration promotes high compositional turnover driven by ecological drift; (iv) homogenizing dispersal, characterised by high dispersal rates resulting in uniform community composition; and (v) “undominated”, where variation cannot be clearly attributed to either selection or dispersal (Stegen et al., 2015). The QPE method was selected due to its capability to quantify the simultaneous contributions of selection and dispersal processes while explicitly incorporating phylogenetic relationships, allowing robust

identification and comparison of assembly mechanisms under different environmental scenarios.

2.9. Correlation and regression analyses

A best subsets regression approach was used to identify predictors most strongly associated with microbiome diversity metrics from the set of pH, temperature, conductivity, phosphate, grain size, catchment gradient, season, and abstraction method. The function regsubsets() from the leaps R package (Plot Lumley and Miller, 2009) was utilised to systematically construct and evaluate optimal regression models for each diversity metric, selecting subsets of explanatory variables that best predicted observed diversity patterns. Model performance was evaluated by computing the 5-fold cross-validation error using the train() function from the caret R package (Kuhn, 2008). The results of the top-performing models were summarised and visualised using the tab_model() function from the sjPlot package (Lüdecke, 2018).

We also applied a compositional balance approach to identify a minimal set of microbial genera which correlated with the pathogen abundances as determined by qPCR. We performed CODA LASSO regression using the top 100 most abundant genera. The resulting regression model took the form:

$$y_i = \beta_0 + \beta_1 \log(x_{1i}) + \ldots + \beta_j \log(x_{ji}) + \epsilon_i$$

where y_i represents the pathogen abundance from qPCR (continuous variable), and x_{ji} is the abundance of the j^{th} genus in the i^{th} sample. The CODA LASSO regression was implemented using the coda_glmnet() function from the coda4microbiome package (Calle et al., 2023). This approach imposed two constraints on the model: (a) the sum of all β-coefficients was constrained to zero, separating taxa into positively and negatively correlated groups relative to the pathogen outcomes; and (b) a LASSO shrinkage was applied, forcing insignificant β-coefficients to zero, thereby producing a parsimonious subset of taxa significantly correlated with pathogen abundance.

3. Results

Results are presented in relation to the study hypotheses: we first compare pathogen indicator gene markers and bulk microbial loads across abstraction points, then assess how abstraction method, season, and site shape community composition and diversity, and finally use null-model analyses to infer the relative roles of selection and dispersal in community assembly.

3.1. Indicators of contamination across abstraction points

Physicochemical parameters including temperature, pH, electrical conductivity, and turbidity were measured in situ using a calibrated multiparameter probe (YSI Inc., USA). Sample pH ranged from 7.0 to 8.2; downstream samples were slightly more alkaline (median: 8.0) than other abstraction points (median: ~7.5). Downstream samples also had the highest electrical conductivity (median: 3595 μS/cm). Sample temperature ranged from 23.6 °C to 30.6 °C. Turbidity was much lower in hand pump samples (0 – 5 NTU) compared to the other abstraction points (20 – >1000 NTU).

Quantitative PCR results (presented as gene copies per 100 ml) revealed significant variation in total bacteria (*16S rRNA*) across abstraction methods (ANOVA, $F(4, 61) = 8.37$, $p < 0.001$). Downstream pool samples exhibited the greatest number of bacteria (median = 1.3×10^8 copies/100 ml), while protected hand pump samples exhibited the lowest (median = 9.7×10^7 copies/100 ml). *E. coli* abundance varied among abstraction points (hand pump median = 12 copies/100 ml vs. open well median = 78 copies/100 ml), although these differences were not significant (ANOVA, $F(4, 4) = 1.39$, $p = 0.38$) and do not link directly to water quality compliance of 0 culturable *E. coli* per 100 ml. In contrast, coliform bacteria exhibited significant differences (ANOVA, $F(4, 24) = 8.11$, $p < 0.001$), with downstream samples showing the highest abundance (median = 400 copies/100 ml) and hand pump samples the lowest (median = 56 copies/100 ml). Similarly, *Legionella* concentrations differed significantly (ANOVA, $F(4, 37) = 19.45$, $p < 0.001$), with downstream samples again showing the highest abundance (median = 234 copies/100 ml) compared to significantly lower counts in hand pump samples (median = 63 copies/100 ml). *Mycobacterium* was detected only in excavated scoop holes and hand pump samples, where values were significantly different (ANOVA, $F(1, 9) = 10.73$, $p < 0.001$); excavated scoop holes exhibited higher abundance (median = 209 copies/100 ml) compared to hand pumps (median = 38 copies/100 ml). *Pseudomonas aeruginosa* (gyrB) was not detected above the assay quantification limit in any samples. These findings are presented in full, along with pairwise ANOVA test results between each abstraction method, in Fig. 2.

3.2. Contamination indicators correlate with environmental taxa

To explore associations between microbial community composition and the qPCR indicator panel, we fitted compositional balance (CODA LASSO) regressions using the top 100 genera. The resulting balances highlight genera whose relative abundances covary with higher or lower indicator gene copies, and we treat these taxa as putative indicators of association. Genera positively associated with faecal-indicator gene copies included taxa commonly reported from wastewater, sediments or anoxic niches (e.g., Chryseomicrobium, Methanobacterium and Methanoperedens), which could reflect a combination of shared contamination pathways, environmental conditions conducive to persistence, or symbiotic co-occurrence behaviours. Several genera associated with Legionella/Mycobacterium markers (e.g., Dongia) have been reported from biofilm communities, suggesting that biofilm-associated assemblages may co-occur with opportunistic pathogens. Conversely, genera negatively associated with indicator abundance included taxa typically reported from more oxygenated and dynamic freshwater environments (e.g., Methylosinus, Denitratisoma and Duganella). Because 16S rRNA amplicon data do not directly measure metabolic function, viability, or activity, these associations should be interpreted cautiously and validated with targeted measurements such as hydrochemistry, functional genomics, or cultivation.

3.3. Microbial diversity distinguishes abstraction method

Significant differences in microbial community richness (ANOVA, $F(4, 58) = 24.60$, $p < 0.001$) and Shannon evenness (ANOVA, $F(4, 58) = 34.83$, $p < 0.001$) were observed among the different abstraction methods. Specifically, researcher-excavated scoop holes exhibited the highest richness (median =1094) and evenness (median = 6.33) values while downstream pools exhibited the lowest richness (median =204) and evenness (median = 4.02) values (Fig. 3). Alpha diversity measures (richness, evenness, NRI, and NTI) were all statistically similar between community scoop holes and hand pumps, although distinct differences were observed between the two groups in terms of both taxa abundance (Bray-Curtis dissimilarity) and phylogenetic composition (Unweighted UniFrac distance) (Fig. 3). We performed a sensitivity analysis by calculating Bray-Curtis dissimilarity after filtering out the lowest 0.01 %, 0.1 %, and 1 % of ASVs by abundance. This left us with a subset of 17,375, 2924, and 255 ASVs, respectively. We found that ordination structure was almost entirely influenced by high-abundance features, and that the long tail of low-abundance features does not materially change sample relationships (**Supplementary Figure 6**).

PERMANOVA analyses confirmed that the abstraction method consistently explained the greatest proportion of variability in microbial community composition across all beta diversity metrics tested: Bray-Curtis ($R^2 = 0.297$, $p < 0.001$), Unweighted UniFrac ($R^2 = 0.223$, $p < 0.001$), and Weighted UniFrac ($R^2 = 0.367$, $p < 0.001$). By comparison,

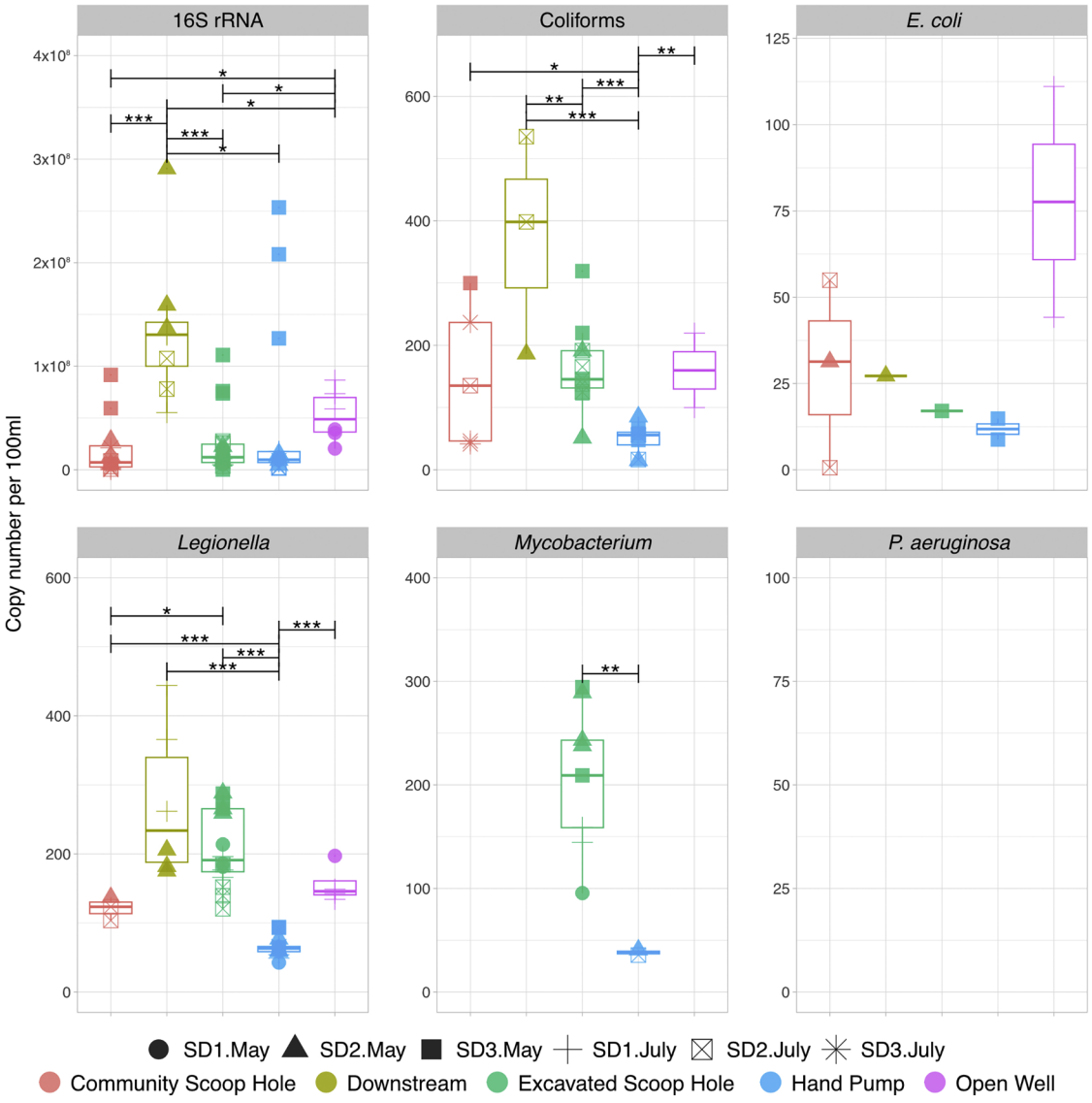

Fig. 2. Microbial water quality was assessed via quantitative PCR. Standard curves were generated for each target gene using serial dilutions of G-block standards containing the corresponding gene inserts ranging from 1 × 108 to 1 × 101 copies per reaction. Quantification of target genes in each sample was calculated by comparing the threshold cycle (Ct) values to the standard curve. Standard curve descriptors were as follows: 16S rRNA (efficiency: 100.08, R2: 1.0), coliforms (efficiency: 100.08, R2: 1.0), *E. coli* (efficiency: 98.84, R2: 1.0), Legionella (efficiency: 110.17, R2: 0.99), and Mycobacterium (efficiency: 106.28, R2: 1.0). Results were adjusted to account for total DNA concentration and volume of water collected for each sample in order to obtain bacterial counts per 100 ml.
SD1: sand dam 1 (ID 106); SD2: sand dam 2 (ID 167); SD3: sand dam 3 (ID 211).

differences associated with sampling sites or seasonal factors were substantially less influential. This result highlights that the abstraction method is the predominant factor shaping microbial communities within these shallow aquifer systems (Figs. 4–8).

3.3. Ecological core microbiome varies between abstraction points

The ecological core microbiome differed greatly between abstraction points, reflecting the differing environmental niches occupied by each one. Downstream samples had a relatively simple core, constituted of just 4 taxa all with a prevalence of 100 %. These included *Sporichthyaceae, Cyanobium, Hydrogenophaga*, and *Flavobacterium*, suggesting eutrophic conditions and, of particular concern, algal blooms. Community scoop hole samples were characterised by a high abundance of aerobic heterotrophs including *Vogesella, Acidovorax, Cellvibrio*, and *Methylophilus*, again with a prevalence of 100 %. Hand pump samples were more varied with just one taxon, the family of aerobic heterotrophs *Vicinamibacteraceae*, exhibiting 100 % prevalence. Open well samples were characterised by aerobic taxa such as *Novosphingobium*, carbohydrate degraders such as *Chitinophagaceae*, and marine groups including *Ilumatobacteraceae*,which may reflect elevated ionic strength (as indicated by higher electrical conductivity) in some samples. By contrast, the excavated scoop holes contained a number of anaerobic taxa such as *Omnitrophus*. Excavated scoop hole and open well samples had the most complex core microbiomes, which included 25 and 23 taxa respectively. Hand pump samples had 16 taxa, community scoop holes had 11, and downstream samples had 4.

3.4. Impact of selection pressure on community structure

We used a combination of null modelling procedures – including NRI, NTI, NST, and QPE – to analyse phylogenetic relationships between and within samples, in order to provide insight into the relative contributions of deterministic (environmental selection) and stochastic

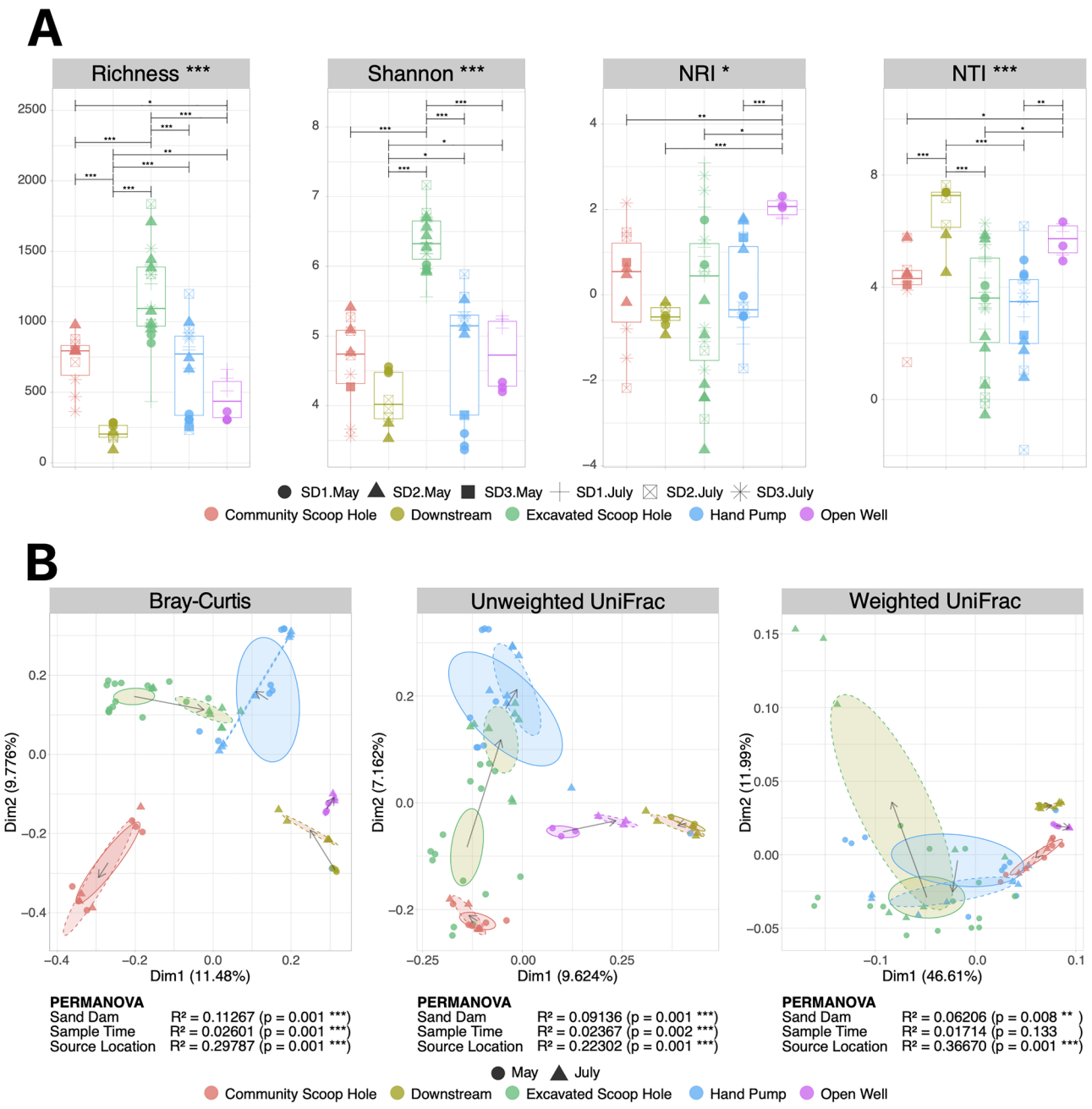

Fig. 3. Alpha and Beta Diversity Metrics.(A) includes alpha diversity measures including richness, evenness, and NRI and NTI. ANOVA was used to determine significance between groups, where *: $p < 0.05$; **: $p < 0.01$, and ***: $p < 0.001$. (B) includes beta diversity measures utilising three different distance metrics: Bray-Curtis measures dissimilarity in species abundance, Unweighted UniFrac measures dissimilarity in species taxonomy, and Weighted UniFrac measures dissimilarity in both species abundance and taxonomy. Ellipses are drawn at 95 % confidence interval of standard error with lines going from May to July plotted at mean of the ordination values at each time point. PERMANOVA statistics are provided against Sand Dams, Sampling Time, and Source Location is provided giving the percentage variability (R2) explained by these predictors.
SD1: sand dam 1 (ID 106); SD2: sand dam 2 (ID 167); SD3: sand dam 3 (ID 211).

(ecological drift and dispersal) processes in microbial community assembly. Across most abstraction points, deterministic processes predominantly shaped microbial community composition, as evidenced by consistently low NST values (ranging from 0.21 to 0.32), indicating limited influence from random dispersal or drift. A notable exception was researcher-excavated scoop holes, which exhibited a balanced influence from both stochastic and deterministic processes (NST ≈ 0.5), suggesting a more complex community structure influenced equally by environmental selection and ecological drift or random dispersal.

To characterise the nature of deterministic selection pressures, we compared phylogenetic patterns across samples collected from multiple abstraction points and across seasons. Abstraction points directly exposed to environmental influences (i.e., community scoop holes, open wells, and downstream pools) exhibited homogenous selection patterns across both spatial and temporal scales, indicating persistent, uniform environmental conditions. By contrast, researcher-excavated scoop holes and sealed hand pump samples exhibited roughly equal proportions of homogenous and variable selection pressures.

Strong environmental filtering (NTI > 2) was observed across all abstraction points, with open wells and downstream pools—sites continuously exposed to environmental contamination—displaying the highest phylogenetic clustering. This finding highlights a robust

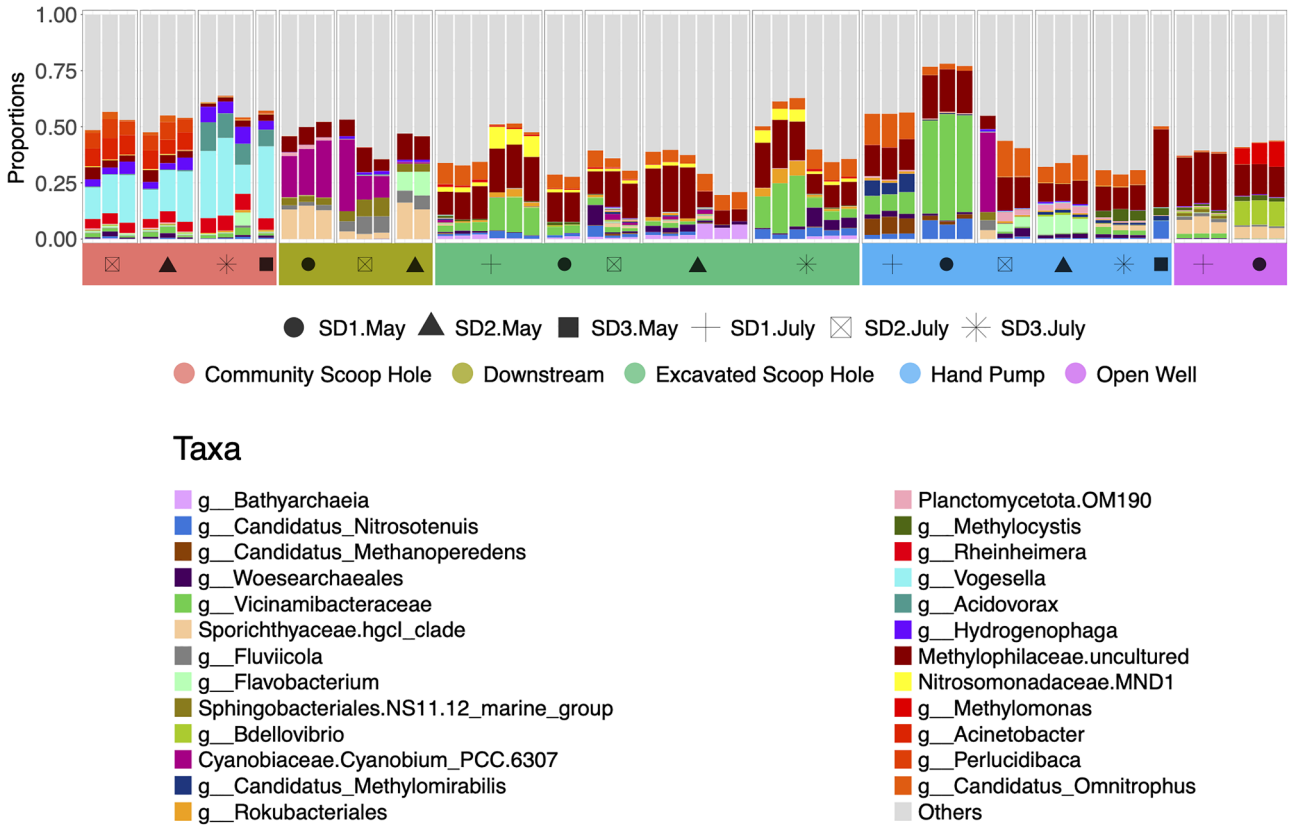

Fig. 4. Most Abundant Genera.
Community structure based on relative abundance of the top 25 most abundant taxa identified at different taxonomic groups, where 'others' refer to all taxa not included in the top 25.
SD1: sand dam 1 (ID 106); SD2: sand dam 2 (ID 167); SD3: sand dam 3 (ID 211).

correlation between environmental exposure and microbial selection pressures. Notably, the comparison between sealed hand pumps and open wells revealed significant differences in phylogenetic clustering metrics (NRI, $p < 0.001$; NTI, $p < 0.01$) and proportion of homogenous selection (24.17 % for hand pumps vs. 40 % for open wells). This observation strongly suggests that protective sanitary infrastructure (sealed hand pumps) substantially reduces environmental selection pressure.

3.5. *Impact of environmental variables on community structure*

Multiple regression analyses were conducted to identify environmental factors influencing microbial alpha diversity metrics (richness, Shannon entropy) and environmental filtering (NRI and NTI). Models were ranked according to cross-validation error, and key trends from the top models are summarised in Fig. 8, with full details in **Supplementary Tables 2–5**. In the highest-ranking models, richness was positively correlated with sediment grain size ($\beta = 6.05$, SE = 2.79, $p < 0.05$) and researcher-excavated scoop holes ($\beta = 492.23$, SE = 188.64, $p < 0.001$), but negatively correlated with downstream samples ($\beta = -492.23$, SE = 188.64, $p < 0.05$). Shannon entropy showed a strong positive correlation with researcher-excavated scoop hole samples ($\beta = 1.80$, SE = 0.30, $p < 0.001$) and with water sampled in May, at the start of the dry season. Environmental filtering metrics (NRI and NTI) were positively correlated with community scoop holes (NRI: $\beta = 3.40$, SE = 0.50, $p < 0.001$; NTI: $\beta = 1.19$, SE = 0.54, $p < 0.05$), indicating intensified selection pressures consistent with homogenous selection observed in QPE analyses. Conversely, downstream samples exhibited negative correlations with both richness and evenness, characteristic of eutrophic or frequently disturbed conditions which favour opportunistic, r-selected species.

Because our null models indicated a dominance of deterministic rather than stochastic processes, we expected clear associations between microbial community structure and measured physicochemical parameters. However, few significant relationships were detected, suggesting that other unmeasured variables contributed significantly to deterministic community assembly.

4. Discussion

Groundwater is an important water source in low-income, semi-arid regions due to improved water quality and greater protection from evaporation compared with surface water. Sand dams aim to increase groundwater availability by collecting surface runoff and promoting aquifer recharge, and can be a cost-effective mechanism of sustainably supporting local communities and stimulating local ecosystems (Neufeld et al., 2020; Ritchie et al., 2021). However, their efficacy depends on the quantity and quality of water they can sustainably provide over a decades-long lifespan. This requires an ongoing ability to protect communities from pathogens and chemical contaminants.

Water quality in these regions is a growing concern due to increasing levels of contamination. Risk factors for microbial contamination include proximity to unsanitary waste, agriculture, and livestock, as well as poor well construction and maintenance; all of these risks were highlighted at the study sites (Kanoti et al., 2019; Quinn et al., 2018). Water quality varied significantly between different abstraction points sampled in this study, where water from unimproved sources was found to exhibit higher pH, conductivity, and turbidity than water abstracted from hand pumps. We used qPCR to detect multiple targets, including both faecal contamination indicators (coliforms and *E. coli*) and

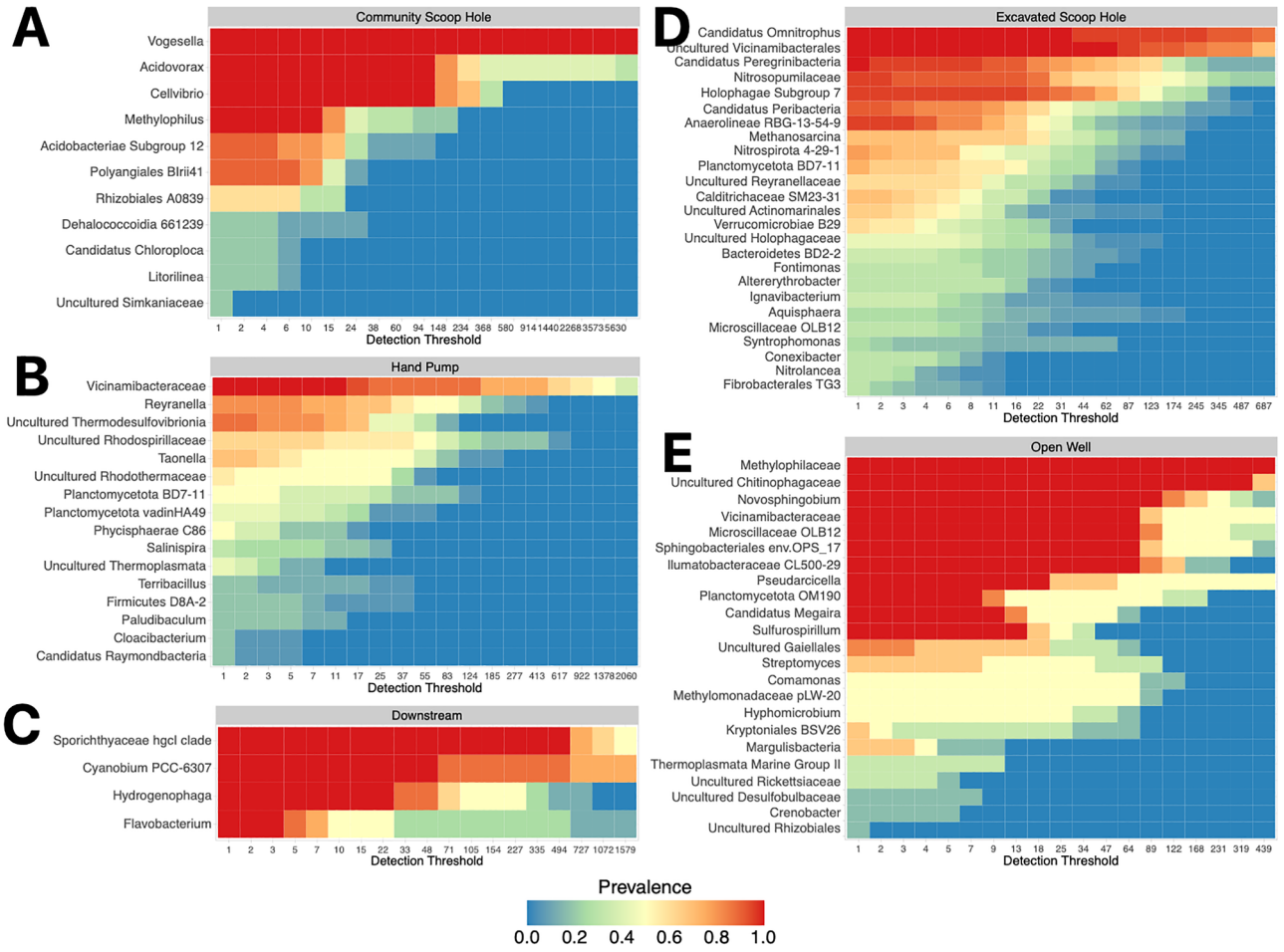

Fig. 5. Abstraction Point Core Microbiomes. The core microbiome for each abstraction point was determined by first selecting the top 500 most prevalent genera by occupancy, and then selecting the subset of these whose explanatory value in Bray-Curtis beta diversity exceeded 2 %.

opportunistic environmental pathogens (*Legionella* and *Mycobacterium*), which suggested that a multitude of contamination routes were present at each abstraction point. While hand pumps improved microbial water quality relative to unimproved sources, the presence of faecal indicator bacteria DNA in all of these abstraction points is concerning. Kenyan water quality standards are grounded in culture-based assays that do not directly map onto qPCR measurements, therefore one should be careful to interpret these results in terms of between-source trends rather than as evidence of regulatory failure (WASREB, 2023). Notably, the observed patterns align with prior evidence of pathogen risk at these sites (Quinn et al., 2018).

Interpretation of 16S rRNA amplicon and qPCR data requires caution. Amplicon sequencing provides relative-abundance profiles that can be influenced by primer bias, variation in rRNA gene copy number, PCR/sequencing artefacts, and reagent or low-biomass contamination; alpha-diversity estimates can also be affected by the composition of other samples in multiplexed runs (Bissett and Brown, 2018; Salter et al., 2014). We made efforts to minimise the impact of such biases, and found indication in our sensitivity analysis that our results are robust to potential sequencing artefacts, but results should nevertheless be interpreted with caution. Our conclusions focus on consistent qualitiative differences between sources and community-level patterns rather than absolute diversity values or fine-scale functional claims from taxonomy alone. Moreover, it is important to note that qPCR quantifies gene copies and can detect DNA from viable and non-viable cells as well as extracellular DNA; non-detection may also reflect concentrations below the effective detection capacity for the filtered volumes, imperfect extraction efficiency, PCR inhibition, or marker/primer coverage limitations. For example, the coliform *LacZ* assay we used achieves a limit of quantification of ~100 copies with >3 replicates, and correlates relatively weakly with culture-based methods ($R^2 = 0.72$) (Thom et al., 2025). Future work might wish to replicate the molecular sensing approach with propidium monoazide (PMA) and/or ddPCR to address some of these limitations. We therefore interpret the qPCR panel as a comparative indicator of contamination potential, and recommend that water safety planning and quantitative microbial risk assessments combine molecular assays with culture-based measurements.

Microbial community structure differed significantly among abstraction methods, as confirmed by beta diversity metrics (Bray-Curtis, Unweighted UniFrac, and Weighted UniFrac). PERMANOVA analyses indicated that abstraction method explained the highest proportion of community variability (Bray-Curtis: $R^2 = 0.297$; Unweighted UniFrac: $R^2 = 0.223$; Weighted UniFrac: $R^2 = 0.367$, all $p < 0.001$), surpassing the influence of location and seasonal variations. Microbial community structure was predominantly driven by deterministic forces over stochastic, in line with the observation that microbial communities are highly governed by micro-scale habitat changes such as chemical profiles and redox conditions in sand aquifer systems (Yamamoto et al., 2019). The result implies strong niche differentiation across abstraction methods; the one exception was excavated scoop holes, where the less extreme micro-environment within the dams produced an equal balance between deterministic and stochastic assembly processes (NST = 0.51).

The other abstraction points were characterised by distinct environmental niches, which remained largely homogenous between sites and throughout the season. The higher richness and evenness observed

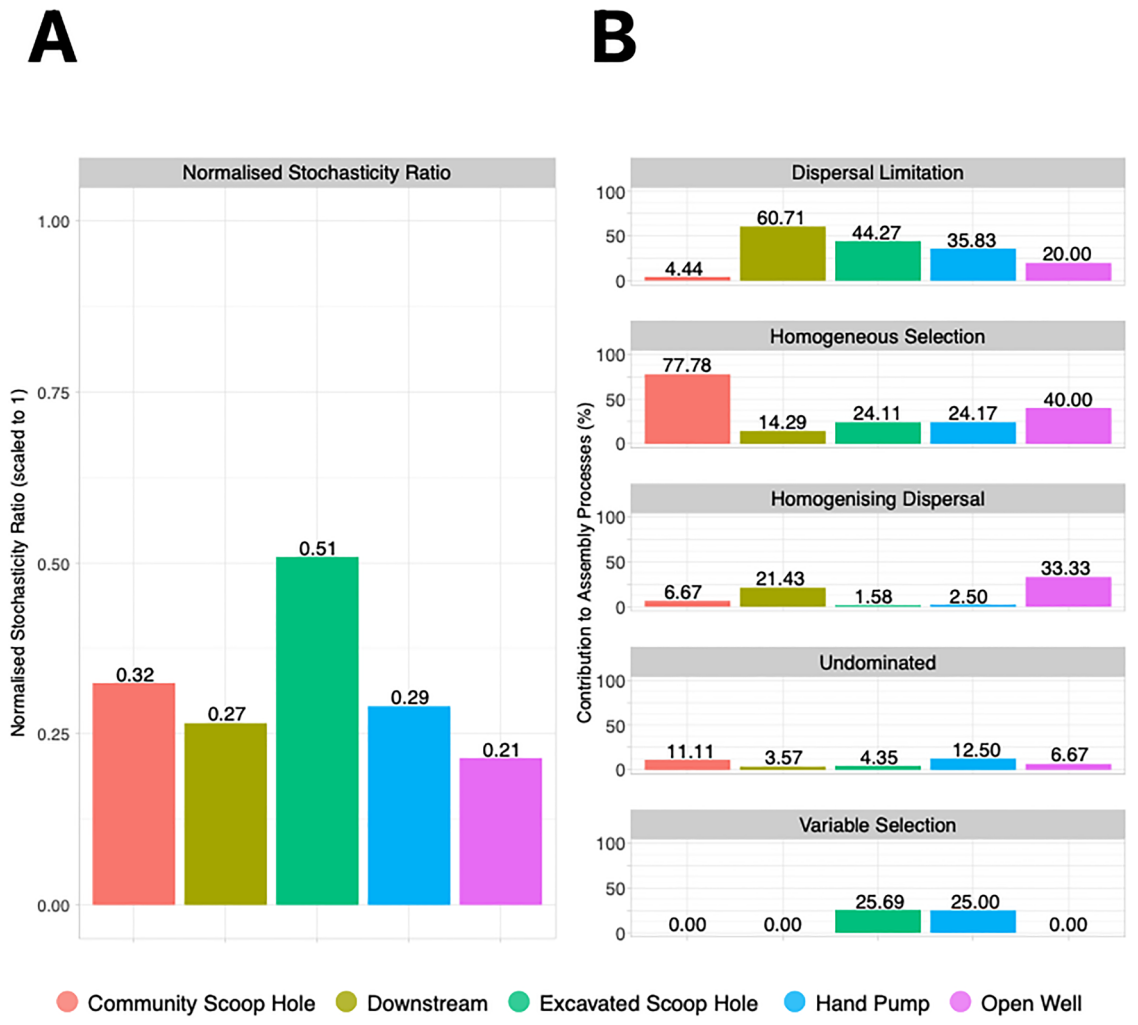

Fig. 6. Community Assembly Processes.
Various null modelling approaches were taken to elucidate community assembly processes. (A) normalised stochasticity ratio (NST) compares beta diversity to a randomised null model to determine the contribution of stochastic processes (ecological drift) to quantify community turnover, and (B) quantitative process estimate (QPE) determines the relative contribution of deterministic processes to community assembly by comparing βNTI (for homogenous and variable selection) and Raup-Crick (for dispersal) indices to a randomised null model.

at the beginning of the dry season – indicative of more diverse and resilient communities – may suggest that selection pressures increase throughout the dry season, leaving behind only more stress-tolerant species. We hypothesise that such selection pressures may include UV or osmotic stress, ion toxicity, or nutrient imbalances, as have been observed previously in samples from this region (Gevera et al., 2020). More research is required, however, to substantiate these hypotheses. The abundance of taxa involved in nitrogen cycling, including Nitrosopira, Nitrosotenius, Hydrogenophaga, and Methylomonas could be consistent with nutrient inputs and other solutes introduced via surface runoff and human/animal waste as reported previously in the region (Gevera et al., 2020). While we were unable to measure dissolved oxygen, redox potential and nutrient concentrations in situ, which limits our ability to draw direct conclusions concerning viability or activity, this would align with ongoing public health risk factors previously identified at the study sites, and further investigation of these factors would be a valuable avenue for future research (Parker et al., 2010; Quinn et al., 2019). Where we discuss low-oxygen or nutrient-rich niches, this is inferred indirectly and should be validated with paired hydrochemical measurements in future work.

Our compositional-balance analyses identified sets of genera whose relative abundances covaried with the qPCR indicator panel. Several genera that were positively associated with higher indicator gene copies are commonly reported from wastewater, sediments or anoxic microenvironments, suggesting that these taxa and markers may share transport pathways and/or occur under similar environmental conditions. Again, limitations inherent to 16S sequencing – in this case including rRNA gene copy number and large uncertainties in resolving functional capacity – mean that these compositional balance coefficients should be interpreted as directions of association rather than an accurate representation of underlying biology. These associations can be treated as hypotheses requiring targeted validation in future work to reveal definitive mechanistic insights. Across sites, hand pump samples had lower overall bacterial loads and lower indicator gene copies than exposed abstraction points, consistent with reduced exposure to environmental contamination during abstraction. Downstream pools showed elevated relative abundance of cyanobacteria, suggesting more eutrophic conditions that may increase the risk of harmful algal blooms and cyanotoxins (Cheng et al., 2023; Metcalf and Souza, 2019).

Stochastic influences on microbial assembly include dispersal processes and ecological drift. To assess dispersal efficiency between abstraction points, we calculated Raup-Crick (RC) beta diversity, quantifying deviations in species turnover relative to a randomised null model. Our results revealed a large proportion of community assembly is attributable to dispersal limitation in most of the abstraction points, which can be well explained by the layers of sand and concrete that

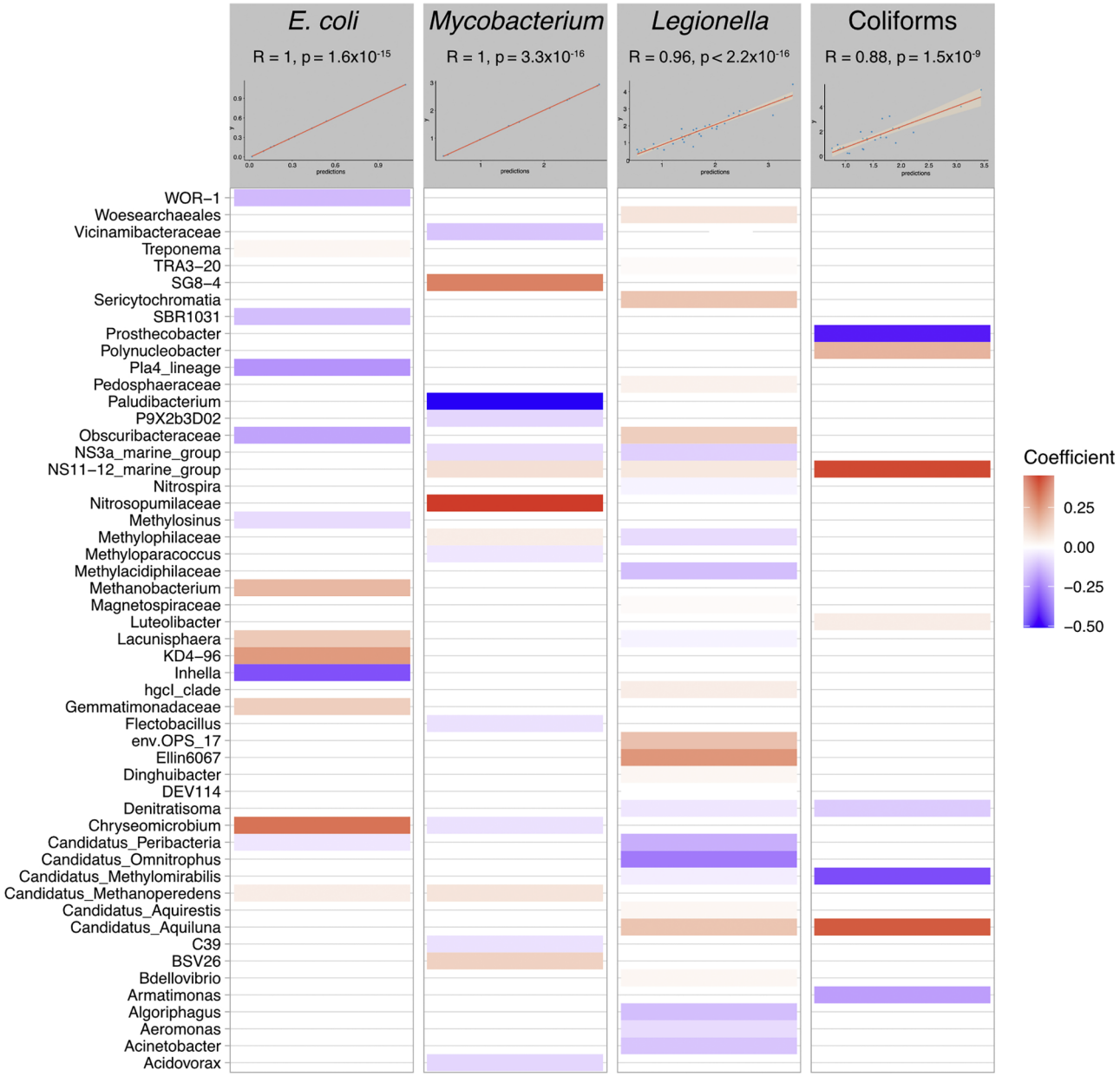

Fig. 7. Compositional Balance Regressions against Water Quality Indicators.
A compositional balance approach was used to find a minimal subset of genera whose abundance was correlated with our pathogen qPCR panel, accounting for potentially spurious correlations due to the compositional nature of the data. The model constraint is that all β-coefficients sum to zero, thereby producing two sets of features: one positively and one negatively associated with the outcome. The optimization function incorporated a LASSO shrinkage that forced some of the β-coefficients to go to zero, in order to produce a minimal subset.

separate abstraction points from one another. We observed that sediment grain size (D50) correlated positively with richness and negatively with environmental filtering (NRI). This corroborates previous literature demonstrating that grain size and ionic strength are key predictors of colloid fate (Mitropoulou et al., 2013; Vasiliki I. V.I. Syngouna and Chrysikopoulos, 2011).

Our findings indicate effective microbial dispersal limitation between sample points around sand dams, suggesting that pathogens are primarily introduced during abstraction when adequate protections are not in place. Sites continuously exposed to the environment exhibit persistent contamination risks, while sealed hand pumps reduce the risk of pathogen ingress. We found that microbial community assembly at abstraction points is shaped predominantly by environmental selection and dispersal limitation, and we found evidence that pathogen abundance correlates with the abundance of environmental taxa and perhaps other chemical contaminants, likely through shared transport pathways (Borchardt et al., 2021). Chemical contamination is also a major concern in this region. Groundwater from Kenya has been demonstrated to exhibit high levels of total dissolved solids, fluoride, salinity, and nitrate, exceeding WHO and Kenyan standards (Gevera et al., 2020). Thus, investigating these links further may provide valuable evidence towards improving water quality across a number of different dimensions.

While sand dams provide a useful example of shallow groundwater systems, future modelling studies should aim to better account for factors including sediment heterogeneity, biological activity, and microbial transport complexity, which are known to have a significant impact on microbial communities. Future research on natural sand filtration efficiency could refine pathogen transport models and quantitative microbial risk assessments (QMRA) by including sand grain single-collector efficiency, biofilm modelling, and advection-diffusion-reaction models

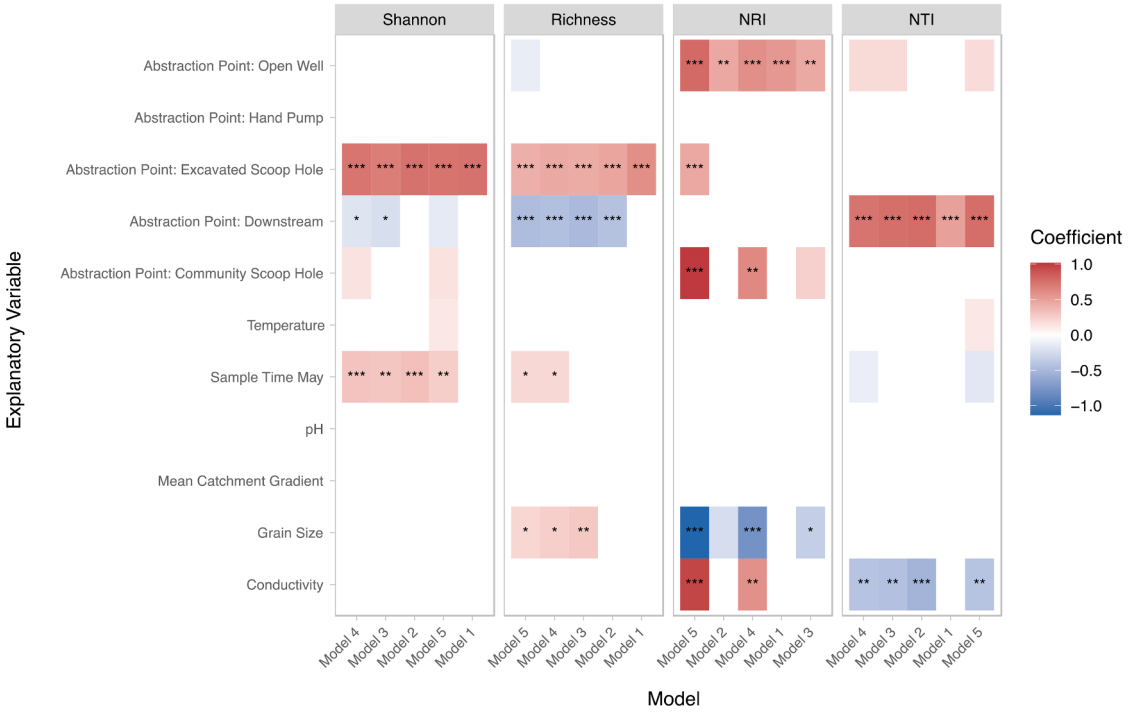

Fig. 8. Multiple Linear Regression Models.
A best subsets multiple regression approach was used to find the minimal set of explanatory variables that modelled each alpha diversity metric. Models for each response variable were rank ordered by cross-validation error and the top 5 models are summarised in the table. Positive coefficients represent features positively correlated with the response variable, and negative coefficients represent negative correlations.
*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; NRI: net relatedness index; NTI: nearest taxon index; TTC: thermotolerant coliforms; CSH: community scoop hole; HP: hand pump; ESH: excavated scoop hole; DS: downstream; OW: open well.

(V.I. Syngouna and Chrysikopoulos, 2011; Trikannad et al., 2023; Tufenkji and Elimelech, 2004), while emerging viability-focused molecular techniques may improve models of pathogen fate (Codony et al., 2019; Zhang et al., 2021). Since drinking-water standards primarily focus on culture-based enumeration of faecal indicator bacteria (e.g., CFU or MPN and the guideline of 0 culturable *E. coli* per 100 ml), future work should aim to validate conclusions against these data to strengthen water safety planning and QMRA.

From a public health perspective, prioritizing sealed sanitary infrastructure and regular maintenance is essential. Water safety plans for the region would be a positive step towards identifying and managing contamination risks proactively at every stage from source protection to abstraction, thereby strengthening community resilience and safeguarding public health.

5. Conclusion

Our study demonstrates that sealed hand pumps significantly reduce microbial loads and pathogen contamination indicators in shallow aquifers. Water filtered through sand was of comparable quality to other groundwater sources. Indicators of faecal contamination were predominantly introduced during abstraction, particularly from unimproved abstraction points such as open wells and scoop holes, highlighting the importance of routine hand pump maintenance and consistent water quality monitoring. Modelling results suggest high sand-filtration efficiency, although more research is needed to fully understand the complexities of microbial transport through the sand, and a field-based QMRA would be beneficial to guide water safety interventions.

Data availability

Raw reads are available in the NCBI Sequence Read Archive under accession PRJNA1333909. Other data used in this study is hosted on figshare at https://doi.org/10.6084/m9.figshare.30218107 and includes the feature table, metadata, and phylogenetic tree used in this study.

CRediT authorship contribution statement

Reuben Duncan: Writing – original draft, Visualization, Formal analysis, Data curation. **Umer Zeeshan Ijaz:** Formal analysis. **Fabien Cholet:** Formal analysis. **Ciara Keating:** Formal analysis. **Orges Koci:** Formal analysis. **Ruth Quinn:** Writing – review & editing, Investigation. **Julie Russell:** Resources, Investigation. **Francis Hassard:** Writing – review & editing, Supervision. **Alison Parker:** Writing – review & editing, Supervision, Investigation, Funding acquisition, Conceptualization. **Cindy J. Smith:** Writing – review & editing, Investigation, Conceptualization.

Declaration of competing interest

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.

Acknowledgements

This work was funded by an award from the Royal Academy of Engineering Frontiers of Engineering (FoESF/1617/216); the Engineering and Physical Sciences Research Council (Grant Nos.: EP/N009711/1, EP/V030515/1, and EP/S022066/1); and a Royal Academy of Engineering Research Chair (RCSRF1718/6/43). The authors greatly appreciate the support of Justus Nyangoka for his assistance in field data collection.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.temicr.2026.100074.

References

Acharya, K., Blackburn, A., Mohammed, J., Haile, A.T., Hiruy, A.M., Werner, D., 2020. Metagenomic water quality monitoring with a portable laboratory. Water Res. 184, 116112. https://doi.org/10.1016/j.watres.2020.116112.

Anuj, S.N., Whiley, D.M., Kidd, T.J., Bell, S.C., Wainwright, C.E., Nissen, M.D., Sloots, T. P., 2009. Identification of *Pseudomonas aeruginosa* by a duplex real-time polymerase chain reaction assay targeting the *ecfX* and the *gyrB* genes. Diagn. Microbiol. Infect. Dis. 63, 127–131. https://doi.org/10.1016/j.diagmicrobio.2008.09.018.

Benitez, A.J., Winchell, J.M., 2013. Clinical application of a multiplex real-time PCR assay for simultaneous detection of Legionella species, Legionella pneumophila, and Legionella pneumophila serogroup 1. J. Clin. Microbiol. 51, 348–351. https://doi.org/10.1128/jcm.02510-12.

Bissett, A., Brown, M.V., 2018. Alpha-diversity is strongly influenced by the composition of other samples when using multiplexed sequencing approaches. Soil Biol. Biochem. 127, 79–81. https://doi.org/10.1016/j.soilbio.2018.09.017.

Borchardt, M.A., Stokdyk, J.P., Kieke, B.A., Muldoon, M.A., Spencer, S.K., Firnstahl, A.D., Bonness, D.E., Hunt, R.J., Burch, T.R., 2021. Sources and risk factors for nitrate and microbial contamination of private household wells in the fractured Dolomite Aquifer of Northeastern Wisconsin. Environ. Health Perspect. 129, 067004. https://doi.org/10.1289/EHP7813.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869.

Calle, M.L., Pujolassos, M., Susin, A., 2023. coda4microbiome: compositional data analysis for microbiome cross-sectional and longitudinal studies. BMC Bioinform. 24, 82. https://doi.org/10.1186/s12859-023-05205-3.

Chase, J.M., Myers, J.A., 2011. Disentangling the importance of ecological niches from stochastic processes across scales. Philos. Trans. R. Soc. B Biol. Sci. 366, 2351–2363. https://doi.org/10.1098/rstb.2011.0063.

Cheng, W., Hwang, S., Guo, Q., Qian, L., Liu, W., Yu, Y., Liu, L., Tao, Y., Cao, H., 2023. The special and general mechanism of cyanobacterial harmful algal blooms. Microorganisms 11. https://doi.org/10.3390/microorganisms11040987.

Chern, E.C., Siefring, S., Paar, J., Doolittle, M., Haugland, R.A., 2011. Comparison of quantitative PCR assays for *Escherichia coli* targeting ribosomal RNA and single copy genes. Lett. Appl. Microbiol. 52, 298–306. https://doi.org/10.1111/j.1472-765X.2010.03001.x.

CHIRPS, 2021. Chirps-Gefs. https://doi.org/10.15780/G2PH2M.

Codony, F., Dinh-Thanh, M., Agustí, G., 2019. Key factors for removing bias in viability PCR-based methods: a review. Curr. Microbiol. 77, 682–687. https://doi.org/10.1007/s00284-019-01829-y.

Gayner, N.J., 2018. River Bank Inducement Influence on a Shallow Groundwater Microbial Community and Its Effects on Aquifer Reactivity. Theses Diss. Univ. Milwaukee, 1990.

Gevera, P., Cave, M., Dowling, K., Gikuma-Njuru, P., Mouri, H., 2020. Naturally occurring potentially harmful elements in groundwater in Makueni County, South-Eastern Kenya: effects on drinking Water quality and agriculture. Geosciences. https://doi.org/10.3390/geosciences10020062.

Kanoti, J., Olago, D., Opiyo, N., Nyamai, C., Dulo, S., Ayah, R., 2019. Microbial and physical chemical indicators of groundwater contamination in Kenya: a case study of Kisumu Aquifer System. Kenya. J. Water Resour. Prot. https://doi.org/10.4236/JWARP.2019.114024.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. Bioinforma. Oxf. Engl. 26, 1463–1464. https://doi.org/10.1093/bioinformatics/btq166.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Software. 28 (5), 1–26. https://doi.org/10.18637/jss.v028.i05.

Lüdecke, D., 2018. sjPlot: Data Visualisation for Statistics in Social Science. UKE, Hamburg. https://doi.org/10.32614/CRAN.package.sj.

Lahti, L., Shetty, S.. Tools for Microbiome Analysis in R. Bioconductor. https://doi.org/10.18129/B9.bioc.microbiome.

Lasage, R., Aerts, J., Mutiso, G.-C.M., de Vries, A., 2008. Potential for community based adaptation to droughts: sand dams in Kitui, Kenya. Phys. Chem. Earth Parts ABC, Hydrological Assessment and Integrated Water Resources Management with Special Focus on Developing Countries 33, 67–73. https://doi.org/10.1016/j.pce.2007.04.009.

Lassmann, T., Sonnhammer, E.L., 2005. Kalign – an accurate and fast multiple sequence alignment algorithm. BMC Bioinform. 6, 298. https://doi.org/10.1186/1471-2105-6-298.

Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G., Neufeld, J.D., 2012. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinform. 13, 31. https://doi.org/10.1186/1471-2105-13-31.

McGhee, J.J., Rawson, N., Bailey, B.A., Fernandez-Guerra, A., Sisk-Hackworth, L., Kelley, S.T., 2020. Meta-SourceTracker: application of bayesian source tracking to shotgun metagenomics. PeerJ 8, e8783. https://doi.org/10.7717/peerj.8783.

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217.

Metcalf, J.S., Souza, N.R., 2019. Chapter 6 - cyanobacteria and their toxins. In: Ahuja, S. (Ed.), Separation Science and Technology, Evaluating Water Quality to Prevent Future Disasters. Academic Press, pp. 125–148. https://doi.org/10.1016/B978-0-12-815730-5.00006-5.

Mitropoulou, P.N., Syngouna, V.I., Chrysikopoulos, C.V., 2013. Transport of colloids in unsaturated packed columns: role of ionic strength and sand grain size. Chem. Eng. J. 232, 237–248. https://doi.org/10.1016/j.cej.2013.07.093.

Neufeld, D.G., Muendo, B., Muli, J., Kanyari, J., 2020. Coliform bacteria and salt content as drinking water challenges at sand dams in Kenya. J. Water Health 18, 602–612. https://doi.org/10.2166/WH.2020.192.

Nikolenko, S.I., Korobeynikov, A.I., Alekseyev, M.A., 2013. BayesHammer: bayesian clustering for error correction in single-cell sequencing. BMC Genomics 14 (Suppl 1), S7. https://doi.org/10.1186/1471-2164-14-S1-S7.

Ning, D., Deng, Y., Tiedje, J.M., Zhou, J., 2019. A general framework for quantitatively assessing ecological stochasticity. Proc. Natl. Acad. Sci. USA. 116, 16892–16898. https://doi.org/10.1073/pnas.1904623116.

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'hara, R., et al., 2007. vegan: Community Ecology Package. San Francisco, CA. https://doi.org/10.32614/CRAN.package.vegan.

Parker, A.H., Youlten, R., Dillon, M., Nussbaumer, T., Carter, R.C., Tyrrel, S.F., Webster, J., 2010. An assessment of microbiological water quality of six water source categories in north-east Uganda. J. Water Health 8, 550–560. https://doi.org/10.2166/wh.2010.128.

Plot Lumley, T., Miller, A., 2009. leaps: Regression Subset Selection. University of Auckland. https://doi.org/10.32614/CRAN.package.leaps.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLOS ONE 9490. https://doi.org/10.1371/journal.pone.0009490.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219.

Quinn, R., Avis, O., Decker, M., Parker, A., Cairncross, S., 2018. An assessment of the microbiological water quality of sand dams in southeastern Kenya. Water 10, 708. https://doi.org/10.3390/w10060708.

Quinn, R., Rushton, K., Parker, A., 2019. An examination of the hydrological system of a sand dam during the dry season leading to water balances. J. Hydrol. X 4, 100035. https://doi.org/10.1016/j.hydroa.2019.100035.

R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Radomski, N., Roguet, A., Lucas, F.S., Veyrier, F.J., Cambau, E., Accrombessi, H., Moilleron, R., Behr, M.A., Moulin, L., 2013. atpE gene as a new useful specific molecular target to quantify mycobacteriumin environmental samples. BMC Microbiol. 13, 277. https://doi.org/10.1186/1471-2180-13-277.

Ritchie, H., Eisma, J.A., Parker, A., 2021. Sand dams as a potential solution to rural water security in drylands: existing research and future opportunities. Front. Water 3.

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 12, 87. https://doi.org/10.1186/s12915-014-0087-z.

Semenza, J.C., 2020. Cascading risks of waterborne diseases from climate change. Nat. Immunol. 21, 484–487. https://doi.org/10.1038/s41590-020-0631-7.

Shade, A., Stopnisek, N., 2019. Abundance-occupancy distributions to prioritize plant core microbiome membership. Curr. Opin. Microbiol. 49, 50–58. https://doi.org/10.1016/j.mib.2019.09.008.

Sloan, W.T., Lunn, M., Woodcock, S., Head, I.M., Nee, S., Curtis, T.P., 2006. Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environ. Microbiol. 8, 732–740. https://doi.org/10.1111/j.1462-2920.2005.00956.x.

Stegen, J.C., Lin, X., Konopka, A.E., Fredrickson, J.K., 2012. Stochastic and deterministic assembly processes in subsurface microbial communities. ISME J. 6, 1653–1664. https://doi.org/10.1038/ismej.2012.22.

Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J., Rockhold, M.L., Konopka, A., 2013. Quantifying community assembly processes and identifying features that impose them. ISME J. 7, 2069–2079. https://doi.org/10.1038/ismej.2013.93.

Stegen, J.C., Lin, X., Fredrickson, J.K., Konopka, A.E., 2015. Estimating and mapping ecological processes influencing microbial community assembly. Front. Microbiol. 6, 370. https://doi.org/10.3389/fmicb.2015.00370.

Sterk, A., Schijven, J., de Nijs, T., de Roda Husman, A.M., 2013. Direct and indirect effects of climate change on the risk of infection by water-transmitted pathogens. Environ. Sci. Technol. 47, 12648–12660. https://doi.org/10.1021/es403549s.

Sutton, S., 2002. Community Led Improvements of Rural Drinking Water Supplies. Knowledge and Research Project.

Syngouna, V.I., Chrysikopoulos, C., 2011a. Transport of pathogens in water saturated sand columns 417–425. https://doi.org/10.1007/978-3-642-24076-8_49.

Syngouna, Vasiliki I., Chrysikopoulos, C.V., 2011b. Transport of biocolloids in water saturated columns packed with sand: effect of grain size and pore water velocity. J. Contam. Hydrol. 126, 301–314. https://doi.org/10.1016/j.jconhyd.2011.09.007.

Thom, C., Ijaz, U., Moore, G., Weir, P., Smith, C.J., 2025. Evaluation of existing lacZ primers and de novo design of an optimized qPCR assay to quantify coliform bacteria in drinking water. J. Appl. Microbiol. 136, lxaf156. https://doi.org/10.1093/jambio/lxaf156.

Trikannad, S., van Halem, D., Foppen, J., van der Hoek, J.P., 2023. The contribution of deeper layers in slow sand filters to pathogens removal. Water Res. 237. https://doi.org/10.1016/j.watres.2023.119994.

Tufenkji, N., Elimelech, M., 2004. Correlation equation for predicting single-collector efficiency in physicochemical filtration in saturated porous Media. Environ. Sci. Technol. 38, 529–536. https://doi.org/10.1021/es034049r.

Wahome, C.N., Okemo, P.O., Nyamache, A.K., 2014. Microbial quality and antibiotic resistant bacterial pathogens isolated from groundwater used by residents of Ongata Rongai. Kenya. Int. J. Biol. Chem. Sci. 8, 134–143. https://doi.org/10.4314/ijbcs.v8i1.13.

WASREB, 2023. Drinking Water Quality and Effluent Monitoring Guideline. WASREB. URL. https://wasreb.go.ke/drinking-water-quality-guidelines/ (accessed 4.17.25).

WHO and UNICEF, 2017. JMP Methodology: 2017 Update and SDG Baselines. WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP). Available at: https://washdata.org/report/jmp-methodology-2017-update.

WHO, UNICEF, 2019. Progress on drinking water, sanitation and hygiene 2000–2017 | UNICEF [WWW Document]. URL. https://www.unicef.org/reports/progress-on-drinking-water-sanitation-and-hygiene-2019 (accessed 4.14.25).

Wolf, J., Johnston, R.B., Ambelu, A., Arnold, B.F., Bain, R., Brauer, M., Brown, J., Caruso, B.A., Clasen, T., Colford, J.M., Mills, J.E., Evans, B., Freeman, M.C., Gordon, B., Kang, G., Lanata, C.F., Medlicott, K.O., Prüss-Ustün, A., Troeger, C., Boisson, S., Cumming, O., 2023. Burden of disease attributable to unsafe drinking water, sanitation, and hygiene in domestic settings: a global analysis for selected adverse health outcomes. Lancet Lond. Engl. 401, 2060–2071. https://doi.org/10.1016/S0140-6736(23)00458-0.

Yamamoto, K., Hackley, K.C., Kelly, W.R., Panno, S.V., Sekiguchi, Y., Sanford, R.A., Liu, W.-T., Kamagata, Y., Tamaki, H., 2019. Diversity and geochemical community assembly processes of the living rare biosphere in a sand-and-gravel aquifer ecosystem in the Midwestern United States. Sci. Rep. 9, 13484. https://doi.org/10.1038/s41598-019-49996-z.

Zhang, Y., Thompson, K.N., Branck, T., Yan, Y., Nguyen, L.H., Franzosa, E.A., Huttenhower, C., 2021. Metatranscriptomics for the Human microbiome and microbial community functional profiling. Annu. Rev. Biomed. Data Sci. 4, 279–311. https://doi.org/10.1146/annurev-biodatasci-031121-103035.

Zhang, R., Ye, Z., Guo, X., Yang, Y., Li, G., 2024. Microbial diversity and metabolic pathways linked to benzene degradation in petrochemical-polluted groundwater. Environ. Int. 188, 108755. https://doi.org/10.1016/j.envint.2024.108755.

Zhou, J., Ning, D., 2017. Stochastic Community Assembly: does it matter in microbial ecology? Microbiol. Mol. Biol. Rev. MMBR 81, e00002–e00017. https://doi.org/10.1128/MMBR.00002-17.

DSPACE

https://dspace.org/

Abstraction method shapes microbial ecology and pathogen risk in water from shallow aquifers

Duncan, Reuben; Ijaz, Umer Zeeshan; Cholet, Fabien; Keating, Ciara; Koci, Orges ...

2026-06
Elsevier
https://doi.org/10.1016/j.temicr.2026.100074
Duncan R, Ijaz UZ, Cholet F, et al., (2026) Abstraction method shapes microbial ecology and pathogen risk in water from shallow aquifers. Total Environment Microbiology, Volume 2, Issue 2, June 2026, Article number 100074

Downloaded from a DSpace repository. https://dspace.org/