



# Anaerobes and methanogens dominate the microbial communities in water harvesting ponds used by Kenyan rural smallholder farmers



Benjamin H. Gregson<sup>a</sup>, Alessia Bani<sup>a</sup>, Laurel Steinfield<sup>b</sup>, Diane Holt<sup>c</sup>, Corinne Whitby<sup>a,\*</sup>

<sup>a</sup> School of Life Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

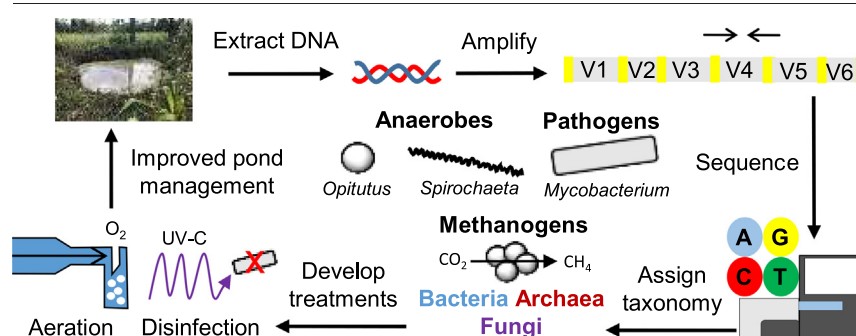
<sup>b</sup> Department of Marketing, Bentley University, Waltham, MA 02452, USA

<sup>c</sup> Center for Enterprise and Entrepreneurship, Leeds University Business School, Leeds LS2 9JT, UK

## HIGHLIGHTS

- Amplicon sequencing determined the microbial community in water harvesting ponds.
- Anaerobes (e.g. *Spirochaeta* and *Opiritatus*) dominated the bacterial community.
- Woese archaeota and methanogens dominated the archaeal community.
- A potentially pathogenic *Mycobacterium* spp. was also detected.
- Highlights need to reduce anaerobic conditions and screen for potential pathogens.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 19 November 2021

Received in revised form 6 January 2022

Accepted 6 January 2022

Available online 10 January 2022

Editor: Frederic Coulon

### Keywords:

Water harvesting  
Microbial ecology  
High Throughput Sequencing  
Pathogens  
Microbial quality  
Chemical quality  
Amplicon

## ABSTRACT

Many rural smallholder farmers in Kenya use water-harvesting ponds, to collect rainwater, as sustainable sources of water for domestic and agricultural purposes. There is currently limited information regarding the microbial ecology in these ponds. Here, we used High Throughput Sequencing (HTS) to characterize the microorganisms present (including potential pathogens and indicator species) alongside ion chromatography to measure water chemistry (anion and cation concentration). Fluoride and magnesium concentration were the strongest predictor variables of the microbial community. Obligately or facultatively anaerobic bacterial genera (e.g. *Spirochaeta* and *Opiritatus*) were abundant within the bacterial community, whilst Woese archaeota and methanogens dominated the archaeal community. This suggests the water in the ponds is hypoxic or anoxic, and if used for irrigation, may potentially impact crop yield and viability. In addition, the opportunistic pathogen non-tuberculous mycobacteria (NTM), *Mycobacterium fortuitum* was found, comprising >1% of the bacterial community, suggesting a potential human health risk. Here we suggest low-cost changes to pond management, to improve or ameliorate pond anoxia and remove pathogens to benefit the livelihoods and welfare of these farms. This study also shows the applicability of HTS to broadly screen the microbial communities, assess water quality, and identify potentially pathogenic groups.

**Abbreviations:** RHS, roof harvesting system; PHS, pond harvesting system; HTS, High Throughput Sequencing; FIB, faecal indicator bacteria; DNA, deoxyribonucleic acid; PCR, polymerase chain reaction; qPCR, real-time quantitative polymerase chain reaction; LOD, limit of detection; LEfSe, linear discriminant analysis effect size; OTU, operational taxonomic unit; rRNA, ribosomal ribonucleic acid; NCBI, National Center for Biotechnology Information; BLAST, basic local alignment search tool; ANOVA, analysis of variance; NMDS, non-metric multidimensional scaling; ANOSIM, analysis of similarity; PERMANOVA, permutational multivariate analysis of variance; VIF, variance inflation factor; CCA, canonical correspondence analysis; NTM, non-tuberculosis mycobacteria.

\* Corresponding author at: School of Life Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK.

E-mail address: [cwhitby@essex.ac.uk](mailto:cwhitby@essex.ac.uk) (C. Whitby).

## 1. Introduction

The majority of rural smallholders in sub-Saharan Africa, currently have no access to surface water or groundwater (Rockstrom, 2000). In addition, climate change is adversely affecting the availability of water in countries, such as Kenya, through increased temperatures, erratic rainfall, and drought (Boelee et al., 2013). Therefore, ex-situ rainwater-harvesting systems, especially the use of on-farm water storage structures, represents the sole alternative for accessing sustainable quantities of water for domestic and agricultural purposes (Pachpute et al., 2009). These systems can be divided into two categories, a roof harvesting system (RHS) or pond harvesting system (PHS) (Zabidi et al., 2020). A typical RHS consists of a sloped catchment system (a roof) which drains via a gutter and attached pipe into a storage container (cistern or tank) (Hamilton et al., 2019). A PHS consists of small runoff storage structures ranging from small manually dug ponds to large community earth and sand dams, that directly collect rainwater, used mainly for irrigation of kitchen gardens, and sometimes for domestic and livestock water supply (Biazin et al., 2012). Although RHSs are more widely utilised around the world (Cook et al., 2013; Despins et al., 2009; Farreny et al., 2011; Hafizi Md Lani et al., 2018), PHSs offer several advantages including: (i) the capacity to store larger volumes of water (Wu et al., 2018); (ii) simplicity in terms of construction and maintenance, meaning an overall lower capital cost to the farmer, leading to greater benefit-cost ratio (Berhane, 2018). Despite these advantages, these ponds are openly exposed, with potential risk to human health where they may act as reservoirs for both human, animal and plant pathogens (Brodie et al., 2007; dos Santos and de Farias, 2017). If the water is to be used for potable purposes and/or crop irrigation (especially if the produce is consumed fresh) (Jongman and Korsten, 2016), then water quality needs to be assessed and ensure it meets the minimum drinking water guidelines (World Health Organisation, 2017) prior to use.

Historically, the microbial quality of water has been evaluated on culture-based enumeration of pathogens and the monitoring of faecal indicator bacteria (FIB), such as *E. coli* and total coliforms (Ahmed et al., 2010, 2011; Hamilton et al., 2019). Culture-dependent methods for pathogen detection are impractical as (i) each pathogenic group requires its own unique isolation technique (Acharya et al., 2019); (ii) <1% of microorganisms are culturable (Hugenholz et al., 1998); (iii) you can only detect the presence of a few microbial groups at a time with limited taxonomic resolution (Chan et al., 2019). The reliability of using FIB to assess water quality has also been questioned due to poor correlations with the occurrence of pathogens, especially non-faecal pathogens (Ahmed et al., 2010; Wu et al., 2011). This may be due to ecological or environmental interactions that compromise their predictive power as proxies for pathogens (Tan et al., 2015).

Molecular techniques, such as qPCR, have been used as an alternative to these culture-dependent methods, for quantification of microbial load and direct detection of pathogens in harvested rainwater (Ahmed et al., 2014). High Throughput Sequencing (HTS) provides one approach for elucidating the fine-scale taxonomic composition of microbial communities, including the identification of potential pathogenic groups (Tan et al., 2015; Vierheilig et al., 2015). Indeed, HTS has been successfully used to detect potential pathogenic groups in RHS systems (Ahmed et al., 2017; Chidamba and Korsten, 2015; Strauss et al., 2019; Zhang et al., 2020); with some of the most prevalent potential pathogenic genera detected including: *Acinetobacter*, *Burkholderia*, *Chromobacterium*, *Clostridium*, *Legionella*, *Mycobacterium*, *Nocardia*, *Serratia*, and *Yersinia*. Although Archaea and Fungi play an important role in human/plant pathogenicity (Bell et al., 2006; Hageskal et al., 2009), and in the biogeochemical cycling of nutrients (Offre et al., 2013), these microorganisms are often overlooked in HTS studies. Consequently, there is limited information regarding the microbial ecology, across all microbial domains in PHS.

The main aim of this study was to apply HTS to characterize the microbial communities present in 16 ponds in Kenya, that are used for kitchen garden crop production. Specifically, this study aims to better understand the relative abundance of microorganisms across taxonomic domains, in relation to the physico-chemical drivers of the communities. This study

therefore provides a thorough assessment of harvested rainwater quality including both the presence of potential pathogens and indicator species, and water chemistry (anion and cation concentrations), which may have potential implications for long-term human health and sustainability.

## 2. Materials and methods

### 2.1. Sample collection

We conducted field surveys at 40 smallholder farms in the County of Laikipia, in Kenya to determine if they had working water harvesting ponds. Of these 40 farms, working ponds were found at 16 sites (mapped in Fig. 1). GPS coordinates are available for each pond but are not listed here for farm anonymity. Triplicate surface water (50 ml) was collected from each pond on the same day. All water samples were filtered through sterile 0.2 µm pore-size syringe filters (Minisart NML, Sartorius). Replicates from each site were filtered using different filters. The volume of water filtered ranged from 6.54 to 15.3 ml. The filtrates were stored at -20 °C prior to chemical analysis and filters stored in their plastic casing at -80 °C prior to DNA extraction.

### 2.2. Chemical analysis

The filtrate was diluted in MilliQ® water and filtered through a Whatman® GF/F filter by vacuum. Anion and cation concentration was analysed using a Dionex ICS-3000 (Thermo Scientific) against a set of standards ranging from 0 to 500 µM (cations) or 0 to 200 µM (anions). For cation analysis, the eluent was 20 nM methyl sulphonic acid run at a flow rate of 1 ml min<sup>-1</sup> for 30 min on a Dionex Ionpac 4 mm column (Column temp 30 °C). The eluted cations were detected with a CSRS 300, 4 mm Suppressor onto a conductivity cell detector. For the anion analysis, the eluent was MilliQ® water and 100 mM KOH. Potassium hydroxide was run on a gradient with a flow of 0.25 ml min<sup>-1</sup>. This was run through an Ionpac AS 18, 2 mm column onto a conductivity cell detector. The water chemistry was assessed by comparison to local and international guidelines on irrigation and drinking water standards (Ayers and Westcot, 1994; National Environment Management Agency, 2006; Kenya Bureau of Standards, 2015; World Health Organisation, 2017).

### 2.3. DNA extraction

Filters were removed from their plastic casing using a sterile scalpel and sterile tweezers, cut into small pieces using sterile scissors and added to PowerBead Tubes (Qiagen) containing 750 µl of PowerBead Solution and 0.7 mm garnet beads. Filters were subsequently homogenised using a Precellys Evolution Homogenizer (Bertin Technologies) fitted with a 2 ml tube holder and white blocking plate (two 30 s cycles at 10,000 rpm). DNA was extracted from homogenised filters using with a DNeasy PowerSoil Kit (Qiagen) according to the manufacturer's instructions and stored at -80 °C until use.

### 2.4. qPCR analysis of phylogenetic marker genes

qPCR of archaeal and bacterial 16S rRNA genes and the fungal internal transcribed spacer (ITS) region was performed using the following primer pairs: Archaea, 344f-ACGGGGYGCAGCAGCGCGA and 915r-GTGCTC CCCCCTAATTCCT (Raskin et al., 1994; Stahl and Amann, 1991); Bacteria, 341f-CCTACGGGNGGCWGCAG and 785r-GACTACHVGGGTATCTA ATCC (Klindworth et al., 2013); Fungi, P5-giTS7-GTGARTCATCGAATCT TTG and P7-ITS4-TCCTCCGCTTATTGATATGC (Ihrmark et al., 2012). These primers have previously been used to quantify bacterial and archaeal 16S rRNA genes (Clark et al., 2020; Huby et al., 2021) and fungal ITS region (Zhang et al., 2018) in environmental samples. qPCR standards of known quantity were created using DNA previously extracted from sediment samples (obtained from the University of Essex). Standards were generated by PCR amplification of the target genes in 50 µl reactions using 25 µl



Fig. 1. Map of smallholder farms in the County of Laikipia, Kenya, which had working Pond Harvesting Systems (PHS).

REDTaq® ReadyMix™ (Sigma-Aldrich), 2 µl of each (10 µM) primer, 20 µl of PCR water (Bioline Reagents Ltd) and 1 µl of DNA template. PCR thermal cycling conditions are summarised in the Table 1:

DNA standards were purified using a GenElute PCR Clean-Up Kit (Sigma-Aldrich), prior to quantification using PicoGreen® dsDNA quantification assays (Thermo Fisher Scientific) on a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific). The target gene abundance for DNA standards were calculated assuming a molecular mass of 660 Da for double-stranded DNA using the following formula: Target abundance =  $6.023 \times 10^{23}$  (copies mol<sup>-1</sup>) × standard concentration (g µl<sup>-1</sup>) / molecular weight (g/mol<sup>-1</sup>) (McKew and Smith, 2015; Tatti et al., 2016). Standard curves for each gene were created using ten-fold dilution series ranging from 10<sup>2</sup> to 10<sup>7</sup> gene copies per µl. For each of the genes, the DNA standards, triplicate samples and three no template controls were amplified in triplicate technical replicates on a CFX 96 Real Time System (Bio-Rad) using SensiFast SYBR No-ROX Kit (Bioline) in 20 µl reactions (10 µl of 2 × mastermix, 0.4 of forward and reverse primers (10 µM), 8.2 µl PCR grade water (Bioline) and 1 µl of template DNA) using a 2-step cycle programme (initial denaturation/polymerase activation for 3 min at 95 °C, followed by 40 cycles of denaturation at 95 °C for 15 s and combined annealing and extension at 60 °C for 30 s). A dissociation curve was run at the end of each assay to verify that only the expected amplification product was generated in addition to confirming by agarose gel electrophoresis. Gene abundances were quantified against their respective standard curves

**Table 1**  
Summary of PCR conditions used to generate DNA standards.

Target	Gene	Cycling conditions	Reference
Archaea	16S rRNA	Initial denaturation: 95 °C 5 min Followed by 35 cycles of: 95 °C 30 s; 60 °C 30 s; 72 °C 45 s Final extension: 72 °C 5 min	(Raskin et al., 1994; Stahl and Amann, 1991)
Bacteria	16S rRNA	Initial denaturation: 95 °C 5 min Followed by 30 cycles of: 95 °C 30 s; 55 °C 30 s; 72 °C 30 s Final extension: 72 °C 5 min	(Klindworth et al., 2013)
Fungi	ITS	Initial denaturation: 95 °C 5 min Followed by 35 cycles of: 95 °C 30 s; 57 °C 30 s; 72 °C 45 s Final extension: 72 °C 10 min	(Ihrmark et al., 2012)

with CFX Manager software (Bio-Rad) using automatic analysis settings for the Cq values and baseline settings. The limit of detection (LOD) for all genes was set at 3.3 cycles lower than the Cq value of the no-template controls (McKew and Smith, 2015; Tatti et al., 2016). Analysis of the standard curves from all qPCR assays showed high efficiency (efficiency >91.5%, R<sup>2</sup> > 0.99) (Table S1).

#### 2.5. Amplicon library preparation and sequencing

Amplicon libraries were prepared and sequenced using the methods previously described (Thomas et al., 2020, 2021). Amplicons were generated by a 25-cycle (Bacteria) and 31-cycle (Archaea and Fungi) PCR. The additional cycles were performed to account for lower abundances of Archaea and Fungi. The 16S rRNA genes and ITS region were amplified using the same primers previously described for qPCR, modified to contain Illumina-specific overhang adapters. Amplification occurred in 25 µl reactions using 12.5 µl REDTaq® ReadyMix™ (Sigma-Aldrich), 10.5 µl of PCR grade water (Bioline), 0.5 µl of each (10 µM) primer and 1 µl of DNA template. Thermal cycling consisted of an initial DNA denaturation step of 3 min at 95 °C followed by 25 or 31 cycles each of 30 s at 95 °C, 30 s at 55 °C and 30 s at 72 °C with a final extension step of 5 min at 72 °C, on an Applied Biosystems Veriti 96-well thermal cycler. The resulting PCR products were purified using Agencourt AMPure XP PCR Purification beads (Beckman Coulter), following the manufacturer's instructions. 5 µl of purified PCR product was used in a short secondary PCR, to attach Nextera XT indices, in the presence of 5 µl of Nextera i5 and i7 index, 25 µl REDTaq® ReadyMix™ (Sigma-Aldrich) and 10 µl of PCR water (Bioline Reagents Ltd). Thermal cycling conditions consisted of an initial denaturation step of 3 min at 95 °C followed by 8 cycles each of 30 s at 95 °C, 30 s at 55 °C and 30 s at 72 °C followed by a final extension step of 5 min at 72 °C. PCR products were purified using Agencourt AMPure XP PCR Purification beads and quantified using PicoGreen® dsDNA quantification assays (Thermo Fisher Scientific), on a POLAR star Omega (BMG Labtech) plate reader. Nextera XT amplicons were then pooled in equimolar concentration. The length of amplicons was verified on SYBR® Safe DNA stain 2% (w/v) agarose E-gels (Invitrogen). Final quantification of the pooled amplicon library was determined with the NEBNext® Library Quant Kit for Illumina® (New England BioLabs) prior to sequencing on the Illumina MiSeq® platform, using a Miseq® 600 cycle v3 reagent kit.

Sequences have been submitted to the NCBI SRA archive under the accession number PRJNA699859.

## 2.6. Bioinformatic analysis

Sequence reads were processed as detailed in Dumbrell et al. (2016). First quality filtering was carried out with Sickle (Joshi and Fass, 2011) with trimming when the average Q score dropped under Q20 across a sliding window of 35 base pairs. Error correction was carried out in SPAdes (Bankevich et al., 2012) which corrects misidentified bases during the sequencing process by implementing the BayesHammer algorithm (Nikolenko et al., 2013). Paired-end reads were merged into single contigs in PANDAsq (Masella et al., 2012) using the PEAR algorithm (Zhang et al., 2014). Further quality filtering was carried out in MOTHUR (Schloss et al., 2009) to remove homopolymer inserts longer than 8. VSEARCH was used to dereplicate sequences, remove singleton sequences, sort by abundance and cluster sequences around centroids at a 97% similarity threshold (Rognes et al., 2016). Chimeric sequences were removed with UCHIME (Edgar et al., 2011) using both de novo and reference-based chimera detection against the 16S rRNA RDP Release 11.5 sequences. Taxonomy was assigned to the OTU centroids using the RDP Classifier (Wang et al., 2007) for bacteria/archaea and the UNITE database for fungi (Kõljalg et al., 2013) (OTU tables are available in Tables S2, S3, S4). To identify potential pathogenic groups, we used the methods previously described (Ferguson et al., 2021). We compared the OTUs generated to a list of known human pathogens (Kembel et al., 2012). We obtained representative 16S rRNA sequences for each pathogen from the NCBI reference sequence database (RefSeq) and search the representative sequence for each OTU against this database with the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990). We defined a pathogenic group as any OTU that shared >99% sequence identity with a strain in the human pathogen reference database.

## 2.7. Statistical analysis

All statistical analyses were carried out in R Studio version 4.0.2. All figures were generated using the 'ggplot2' (Wickham, 2016) and 'cowplot' (Wilke, 2020) packages. Data were first tested for normality using a Shapiro-Wilks test (Shapiro and Wilk, 1965). Normally distributed data were tested for significance using ANOVAs, with p-values adjusted for multiple comparisons using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), followed by a Tukey's HSD post hoc test (Tukey, 1953) within the 'agricolae' package (De Mendiburu, 2020). Non-normally distributed data were tested for significance using a Kruskal-Wallis test, with p-values adjusted for multiple comparisons using Bonferroni correction

**Table 2**

Anion concentrations ( $\mu\text{mol/l}$ ; means  $\pm$  SE;  $n = 3$ ) of water collected from ponds used by Kenyan rural smallholder farmers. Values of concentration that share the same letter (a, b, c) are not significantly different between ponds ( $P < 0.05$ ; Kruskal-Wallis test). \* indicates concentration below the limit of detection. Anions include:  $\text{F}^-$  - Fluoride;  $\text{C}_2\text{H}_3\text{O}^-$  - Acetate;  $\text{HCO}_2^-$  - Formate;  $\text{Cl}^-$  - Chloride;  $\text{NO}_3^-$  - Nitrate;  $\text{SO}_4^{2-}$  - Sulphate;  $\text{PO}_4^{3-}$  - Phosphate;  $\text{CO}_3^{2-}$  - Carbonate.

Pond identity	$\text{F}^-$ ( $\mu\text{mol/l}$ )	$\text{C}_2\text{H}_3\text{O}^-$ ( $\mu\text{mol/l}$ )	$\text{HCO}_2^-$ ( $\mu\text{mol/l}$ )	$\text{Cl}^-$ ( $\mu\text{mol/l}$ )	$\text{NO}_3^-$ ( $\mu\text{mol/l}$ )	$\text{SO}_4^{2-}$ ( $\mu\text{mol/l}$ )	$\text{PO}_4^{3-}$ ( $\mu\text{mol/l}$ )	$\text{CO}_3^{2-}$ ( $\mu\text{mol/l}$ )
1	3.89 $\pm$ 0.17 (ab)	1.48 $\pm$ 0.15 (a)	0.19 $\pm$ 0.01 (a)	49.54 $\pm$ 0.29 (ab)	0.12 $\pm$ 0.06 (ab)	0.12 $\pm$ 0.06 (ab)	0.52 $\pm$ 0.03 (a)	83.07 $\pm$ 9.57 (ab)
2	28.57 $\pm$ 1.25 (abc)	2.34 $\pm$ 0.23 (a)	0.50 $\pm$ 0.06 (a)	600.35 $\pm$ 35.99 (b)	422.59 $\pm$ 24.94 (b)	86.49 $\pm$ 3.48 (c)	1.37 $\pm$ 0.08 (a)	245.51 $\pm$ 21.72 (ab)
3	9.47 $\pm$ 0.16 (abc)	1.92 $\pm$ 0.23 (a)	0.26 $\pm$ 0.16 (a)	64.56 $\pm$ 7.02 (ab)	1.19 $\pm$ 0.29 (ab)	36.58 $\pm$ 5.29 (abc)	15.13 $\pm$ 13.78 (a)	345.97 $\pm$ 68.39 (ab)
4	4.55 $\pm$ 0.33 (abc)	4.19 $\pm$ 1.13 (a)	0.50 $\pm$ 0.43 (a)	71.49 $\pm$ 2.11 (ab)	0.64 $\pm$ 0.23 (ab)	37.14 $\pm$ 3.93 (abc)	1.67 $\pm$ 0.16 (a)	339.65 $\pm$ 6.76 (ab)
5	21.16 $\pm$ 1.36 (abc)	2.13 $\pm$ 0.43 (a)	0.16 $\pm$ 0.08 (a)	234.65 $\pm$ 26.95 (ab)	1.51 $\pm$ 0.72 (ab)	36.45 $\pm$ 10.00 (abc)	1.71 $\pm$ 0.37 (a)	331.42 $\pm$ 61.17 (ab)
6	32.16 $\pm$ 0.24 (abc)	4.02 $\pm$ 0.11 (a)	0.37 $\pm$ 0.03 (a)	87.64 $\pm$ 4.98 (ab)	1.14 $\pm$ 0.15 (ab)	45.93 $\pm$ 1.44 (abc)	8.49 $\pm$ 6.01 (a)	330.48 $\pm$ 9.25 (ab)
7	21.31 $\pm$ 0.06 (abc)	3.40 $\pm$ 0.25 (a)	0.21 $\pm$ 0.07 (a)	37.56 $\pm$ 1.46 (ab)	3.80 $\pm$ 2.48 (ab)	47.40 $\pm$ 3.15 (abc)	2.19 $\pm$ 0.23 (a)	298.46 $\pm$ 8.44 (ab)
8	22.14 $\pm$ 0.34 (abc)	3.65 $\pm$ 0.11 (a)	0.42 $\pm$ 0.06 (a)	166.76 $\pm$ 2.14 (ab)	1.65 $\pm$ 0.60 (ab)	53.52 $\pm$ 1.70 (bc)	5.55 $\pm$ 2.20 (a)	429.57 $\pm$ 3.95 (b)
9	8.67 $\pm$ 0.08 (abc)	2.74 $\pm$ 0.37 (a)	0.07 $\pm$ 0.05 (a)	67.79 $\pm$ 1.32 (ab)	0.28 $\pm$ 0.15 (ab)	19.89 $\pm$ 2.09 (abc)	0.57 $\pm$ 0.57 (a)	356.44 $\pm$ 33.18 (ab)
10	20.40 $\pm$ 0.46 (abc)	4.00 $\pm$ 1.26 (a)	0.35 $\pm$ 0.27 (a)	54.45 $\pm$ 3.36 (ab)	0.94 $\pm$ 0.10 (ab)	15.88 $\pm$ 1.86 (abc)	2.27 $\pm$ 0.17 (a)	323.94 $\pm$ 9.16 (ab)
11	10.53 $\pm$ 0.23 (abc)	3.28 $\pm$ 0.25 (a)	0.11 $\pm$ 0.06 (a)	39.30 $\pm$ 0.71 (ab)	1.93 $\pm$ 1.42 (ab)	18.81 $\pm$ 1.95 (abc)	2.14 $\pm$ 0.12 (a)	266.30 $\pm$ 10.55 (ab)
12	6.28 $\pm$ 0.53 (abc)	6.83 $\pm$ 1.96 (a)	0.12 $\pm$ 0.08 (a)	58.74 $\pm$ 8.12 (ab)	0.13 $\pm$ 0.13 (ab)	29.80 $\pm$ 5.65 (abc)	0.83 $\pm$ 0.83 (a)	78.66 $\pm$ 78.66 (ab)
13	34.05 $\pm$ 0.25 (ac)	3.89 $\pm$ 1.54 (a)	0.18 $\pm$ 0.09 (a)	44.36 $\pm$ 0.85 (ab)	* (a)	34.31 $\pm$ 5.81 (abc)	* (a)	* (a)
14	8.75 $\pm$ 0.82 (abc)	* (a)	0.15 $\pm$ 0.09 (a)	20.76 $\pm$ 1.58 (a)	1.03 $\pm$ 0.18 (ab)	8.28 $\pm$ 3.14 (ab)	* (a)	28.10 $\pm$ 28.10 (a)
15	3.08 $\pm$ 0.06 (b)	0.81 $\pm$ 0.21 (a)	0.14 $\pm$ 0.02 (a)	37.39 $\pm$ 1.87 (ab)	* (a)	16.89 $\pm$ 3.74 (abc)	* (a)	161.18 $\pm$ 29.82 (ab)
16	60.33 $\pm$ 6.84 (c)	* (a)	0.11 $\pm$ 0.07 (a)	477.68 $\pm$ 77.99 (b)	1.16 $\pm$ 0.23 (ab)	29.25 $\pm$ 4.84 (abc)	* (a)	306.57 $\pm$ 20.06 (ab)

(Bonferroni, 1936), followed by a Dunn's post hoc test (Dunn, 1964) within the 'FSA' package (Ogle et al., 2020).

The 'Phyloseq' package (McMurdie and Holmes, 2013) was used to rarefy OTUs, transform counts into relative abundances, analyse the taxonomic composition of communities and calculate beta diversity. Only abundant taxa (>1% phyla and >10% genera relative abundance) were used to generate bubble plots. Linear discriminant analysis effect size (LEfSe; LDA score > 2; P-value < 0.05) (Segata et al., 2011) was used to identify OTUs that were differentially abundant between sampling sites.

To evaluate the alpha diversity, Hill numbers (Hill, 1973) were calculated using the 'PhenoFlow' package (Props et al., 2016). Hill numbers are a unified family of diversity indices that compensate for the disproportionate impact of rare taxa by weighting them based on abundance, making them appropriate for analysis of large datasets produced by amplicon sequencing (Vidal-Durà et al., 2018). The degree of weighting is controlled by the order of diversity, q, where increasing q places more weight on the high abundance species and discounts rare species (Chao et al., 2014). Beta diversity was calculated using Bray-Curtis dissimilarity (Bray and Curtis, 1957). The resulting dissimilarity scores were visualised using non-metric multidimensional scaling (NMDS) to observe overall patterns in microbial community structure among the different ponds. Differences in community structure were further tested by analysis of similarities (ANOSIM) (Clarke, 1993) and permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2001), using the 'anosim' and 'adonis' functions in the 'vegan' package (Oksanen et al., 2020), respectively.

Variance Inflation Factor (VIF) values for the chemical variables were calculated in the 'usdm' package (Naimi, 2017). Continuous variables with VIF values greater than 10, and those that were above the multicollinearity threshold of 0.7, were removed. Canonical correspondence analysis (CCA) between community structure and chemical variables was performed in the 'vegan' package (Oksanen et al., 2020). The significance of the CCA model, axes and terms were assessed by a Monte Carlo permutation test (Legendre et al., 2011), using 999 permutations.

## 3. Results

### 3.1. Nutrient analysis

Anion and cation concentrations are summarised in Tables 2 and 3, respectively. Generally, Pond 2 had higher concentrations of anions in comparison to the other ponds, whereas Pond 16 had higher concentrations of cations. Sources of nitrogen (e.g. ammonium and nitrate) were either undetectable or low (<1  $\mu\text{mol/l}$ ) in Ponds 1 and 9 indicating they may be severely nutrient limited, which will impact microbial growth. No anion or cation concentration exceeded the maximum permissible concentration

**Table 3**

Cation concentrations ( $\mu\text{mol/l}$ ; means  $\pm$  SE;  $n = 3$ ) of water collected from ponds used by Kenyan rural smallholder farmers. Values of concentration that share the same letter (a,b,c) are not significantly different between ponds ( $P < 0.05$ ; Kruskal-Wallis test). \* indicates concentration below the limit of detection. Cations include:  $\text{Li}^+$  - Lithium;  $\text{Na}^+$  - Sodium;  $\text{NH}_4^+$  - Ammonium;  $\text{K}^+$  - Potassium;  $\text{Mg}^{2+}$  - Magnesium;  $\text{Mn}^{2+}$  - Manganese;  $\text{Ca}^{2+}$  - Calcium.

Pond identity	$\text{Li}^+$ ( $\mu\text{mol/l}$ )	$\text{Na}^+$ ( $\mu\text{mol/l}$ )	$\text{NH}_4^+$ ( $\mu\text{mol/l}$ )	$\text{K}^+$ ( $\mu\text{mol/l}$ )	$\text{Mg}^{2+}$ ( $\mu\text{mol/l}$ )	$\text{Mn}^{2+}$ ( $\mu\text{mol/l}$ )	$\text{Ca}^{2+}$ ( $\mu\text{mol/l}$ )
1	*(a)	201.35 $\pm$ 4.83 (a)	*(a)	84.32 $\pm$ 1.81 (abc)	46.27 $\pm$ 1.60 (abc)	0.17 $\pm$ 0.17 (ab)	112.00 $\pm$ 0.17 (ab)
2	*(a)	1678.17 $\pm$ 85.04 (b)	13.95 $\pm$ 1.65 (a)	108.87 $\pm$ 4.21 (abc)	67.25 $\pm$ 7.39 (abc)	*(b)	173.03 $\pm$ 9.80 (abc)
3	*(a)	248.46 $\pm$ 3.44 (ab)	4.05 $\pm$ 0.37 (a)	90.09 $\pm$ 5.95 (abc)	57.54 $\pm$ 2.49 (abc)	*(b)	194.04 $\pm$ 2.82 (abc)
4	*(a)	259.33 $\pm$ 13.20 (ab)	5.05 $\pm$ 3.61 (a)	122.82 $\pm$ 3.85 (abc)	92.70 $\pm$ 2.45 (abc)	*(b)	266.61 $\pm$ 7.55 (abc)
5	*(a)	343.84 $\pm$ 18.96 (ab)	8.13 $\pm$ 3.92 (a)	289.67 $\pm$ 13.34 (bc)	197.59 $\pm$ 12.40 (bc)	*(b)	565.48 $\pm$ 20.87 (bc)
6	*(a)	426.57 (ab)	7.41 $\pm$ 7.41 (a)	62.41 $\pm$ 3.68 (abc)	58.46 $\pm$ 2.34 (abc)	*(b)	214.66 $\pm$ 8.98 (abc)
7	*(a)	258.17 $\pm$ 16.01 (ab)	3.30 $\pm$ 0.17 (a)	40.35 $\pm$ 2.59 (abc)	53.51 $\pm$ 3.83 (abc)	*(b)	179.57 $\pm$ 8.78 (abc)
8	*(a)	295.08 $\pm$ 37.55 (ab)	5.09 $\pm$ 2.55 (a)	181.31 $\pm$ 19.18 (abc)	105.17 $\pm$ 11.95 (abc)	*(b)	369.68 $\pm$ 35.85 (abc)
9	*(a)	416.23 $\pm$ 8.79 (ab)	1.30 $\pm$ 1.30 (a)	62.32 $\pm$ 1.54 (abc)	43.54 $\pm$ 3.13 (abc)	*(b)	117.47 $\pm$ 2.58 (abc)
10	*(a)	249.29 $\pm$ 15.80 (ab)	3.05 $\pm$ 1.67 (a)	78.31 $\pm$ 6.04 (abc)	71.55 $\pm$ 4.69 (abc)	1.13 $\pm$ 0.32 (a)	272.54 $\pm$ 8.68 (abc)
11	*(a)	188.13 $\pm$ 11.07 (a)	2.08 $\pm$ 1.08 (a)	20.50 $\pm$ 0.92 (ab)	28.40 $\pm$ 1.66 (a)	0.04 $\pm$ 0.04 (ab)	82.00 $\pm$ 4.81 (a)
12	*(a)	197.99 $\pm$ 7.62 (a)	4.08 $\pm$ 2.15 (a)	75.80 $\pm$ 1.89 (abc)	47.57 $\pm$ 0.76 (abc)	*(b)	154.18 $\pm$ 3.32 (abc)
13	*(a)	390.01 $\pm$ 10.13 (ab)	8.36 $\pm$ 5.02 (a)	67.54 $\pm$ 0.52 (abc)	108.23 $\pm$ 1.25 (abc)	*(b)	364.77 $\pm$ 18.62 (abc)
14	*(a)	370.86 $\pm$ 7.09 (ab)	5.91 $\pm$ 1.25 (a)	19.15 $\pm$ 1.47 (a)	38.48 $\pm$ 1.09 (ab)	*(b)	140.26 $\pm$ 3.04 (abc)
15	*(a)	233.94 $\pm$ 4.86 (ab)	4.08 $\pm$ 0.31 (a)	92.97 $\pm$ 0.76 (abc)	55.52 $\pm$ 1.63 (abc)	*(b)	162.04 $\pm$ 4.58 (abc)
16	*(a)	1419.63 $\pm$ 50.07 (ab)	6.51 $\pm$ 3.27 (a)	331.28 $\pm$ 6.20 (c)	218.82 $\pm$ 4.79 (c)	*(b)	834.30 $\pm$ 30.00 (c)

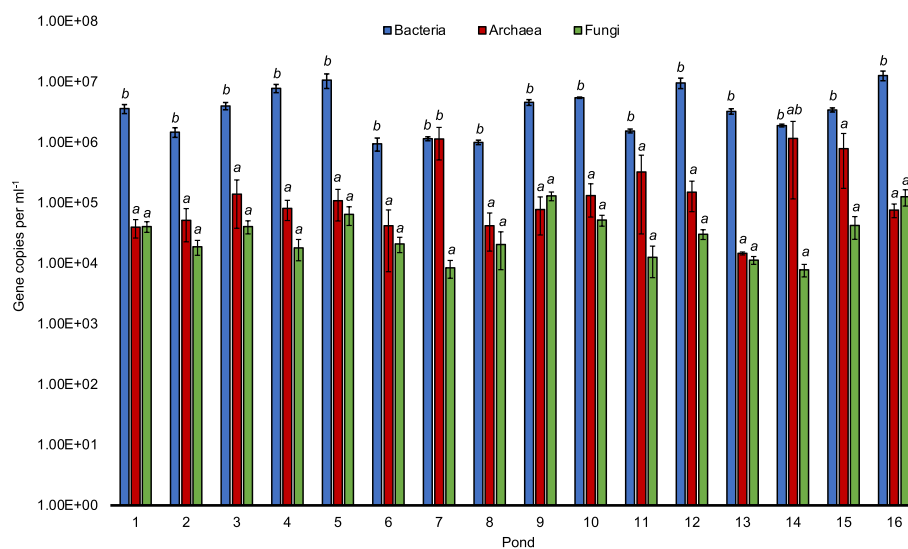
(MPC) set for drinking water quality, based on local (Kenya Bureau of Standards, 2015) or international (World Health Organisation, 2017) standards. In terms of irrigation water quality, thirteen of the ponds exceeded the MPC for potassium (51.15  $\mu\text{mol/l}$ ) and Pond 16 exceeded the MPC for fluoride (52.65  $\mu\text{mol/l}$ ), based on international standards (Ayers and Westcot, 1994). All ponds exceeded the MPC for chloride (0.28  $\mu\text{mol/l}$ ), based on local guidelines (National Environment Management Authority, 2006) for irrigation water quality. However, this value is much lower than the internationally accepted standard (9872.23  $\mu\text{mol/l}$ ), which no ponds exceed.

A Shapiro-Wilk test determined that the chemical data was not normally distributed. With the exception of ammonium, formate and lithium (lithium concentrations below the limit of detection) there were significant differences in ion concentration between ponds (Kruskal-Wallis test,  $P < 0.05$ ). However, no significant differences between ponds were detected for acetate and phosphate after a post-hoc Dunn's test with Bonferroni adjustment for multiple comparisons. Pond 2 had the highest concentration of chloride (600.35  $\pm$  35.99  $\mu\text{mol/l}$ ), nitrate (422.59  $\pm$  24.94  $\mu\text{mol/l}$ ), sodium (1678.17  $\pm$  85.04  $\mu\text{mol/l}$ ) and sulphate (86.49  $\pm$  3.48  $\mu\text{mol/l}$ ). Nitrate concentration had the most variability (coefficient of variation = 385%) driven by the higher concentrations

found in Pond 2, which was at least 111-fold greater than any other pond. Pond 8 had the highest concentration of carbonate (429.57  $\pm$  3.95  $\mu\text{mol/l}$ ). Pond 16 had the highest concentration of calcium (834.40  $\pm$  30.00  $\mu\text{mol/l}$ ), fluoride (60.33  $\pm$  6.84  $\mu\text{mol/l}$ ), magnesium (218.82  $\pm$  4.79  $\mu\text{mol/l}$ ) and potassium (331.28  $\pm$  6.20  $\mu\text{mol/l}$ ). Manganese concentration was very low (<1.13  $\mu\text{mol/l}$ ) and was only detected in ponds 1, 10, 11. Lithium was not detected in any of the ponds tested.

3.2. Microbial community analysis of PHS

The mean bacterial 16S rRNA gene abundance ranged from 9.48  $\times 10^5$  copies  $\text{ml}^{-1}$  (Pond 6) to 1.28  $\times 10^7$  copies  $\text{ml}^{-1}$  (Pond 16) (Fig. 2). Overall, archaeal 16S rRNA gene copies (1.46  $\times 10^4$  to 1.17  $\times 10^6$  copies  $\text{ml}^{-1}$ ) were around two to three-fold lower than the bacterial gene copy abundance (with the exception of Pond 7 where archaeal and bacterial 16S rRNA gene abundance were almost equal 1.14  $\times 10^6$  and 1.15  $\times 10^6$  copies  $\text{ml}^{-1}$ , respectively). Fungal ITS gene copies were consistently lower than that for bacterial 16S rRNA genes, typically by two to three orders of magnitude (7.79  $\times 10^3$  to 1.30  $\times 10^5$  copies  $\text{ml}^{-1}$ ).



**Fig. 2.** 16S rRNA gene and fungal internal transcribed spacer (ITS) region abundance (means  $\pm$  SE;  $n = 3$ ) in water collected from Pond Harvesting Systems (PHS) used by Kenyan rural smallholder farmers, as determined by qPCR analysis. Different letters (a, b) indicate a significant difference in gene copy number between the genes targeted in each pond ( $P < 0.05$ , ANOVA followed by a Tukey's HSD post hoc test).

From the 48 water samples (16 sites in triplicate), in which PCR products were successfully obtained, we recovered 1,111,140 archaeal sequences (~23,000 per sample), 4,433,155 bacterial sequences (~92,000 per sample) and 404,705 fungal sequences (~8400 per sample). After stringent quality filtering and pair-end alignment there was a reduction to 789,851 archaeal sequences (~16,000 per sample), 3,936,285 bacterial sequences (~82,000 per sample) and 357,114 fungal sequences (7400 per sample). Following rarefaction, the number of unique OTUs obtained were 4640 bacterial OTUs, which consisted of 33 phyla, and 779 genera; 627 archaeal OTUs, made up of 8 phyla and 47 genera; and 740 fungal OTUs, across 6 phyla and 248 genera (Tables S2, S3, S4). Irrespective of pond location, Proteobacteria were dominant across ponds (44.2 ± 1.7%) followed by Bacteroidetes (12.8 ± 1.2%), Spirochaetes (12.6 ± 1.9%), Verrucomicrobia (10.3 ± 1.8%), Actinobacteria (9.4 ± 1.3%) and Planctomycetes (4.6 ± 0.7%), which were also present in all sites (Fig. 3). Nitrospirae (1.5%) and Chlamydiae (1.4%) were only detected in Pond 2.

At the genus level several of the most abundant taxa (>5% relative abundance) were obligately or facultatively anaerobic including: *Spirochaeta* (14.1 ± 1.9%), *Opitutus* (10.5 ± 1.8%), *Mycobacterium* (7.6%), *Thermogutta* (7.3%), *Rhodoferrax* (7.2%) and *Salinispira* (6.7 ± 1.3%) (Fig. 4). These anaerobic genera dominated half the sites with *Spirochaeta* exhibiting the highest relative abundance in five ponds (Pond 5, 9, 13, 14, 16) and *Opitutus* in three (Pond 1, 11, 12). In addition, both of these genera were prevalent (>10% relative abundance) in 12 out of the 16 ponds. The remainder of the predominant taxa included strictly aerobic genera: *Chryseolinea* (10.9 ± 3.9%), *Asticcacaulis* (10.9 ± 1.1%), *Polynucleobacter* (8.5 ± 0.9%), *Terracoccus* (7.6 ± 0.5%), *Cellvibrio* (7.6 ± 1.5%), *Haliscomenobacter* (7.5 ± 1.6%), *Ornithinibacter* (7.4%), *Gemmata* (6.4%), *Lacibacter* (6.2%), *Chlorophyta* (5.1%) and *Curvibacter* (5.1%) (Fig. 4). None of these bacterial genera were present in all sites.

At the OTU level fourteen bacterial OTUs had a match of >99% nucleotide identity with a strain in a human pathogen database and were

identified as potential pathogens (Table 4). Although overall, potential bacterial pathogens were not abundant, one notable OTU (OTU97) shared 99.5% identity with the rapidly growing non-tuberculous mycobacteria (NTM) *Mycobacterium fortuitum* and had a significantly higher relative abundance in Pond 1 (LEfSe, LDA = 3.73, P = 3.28 × 10<sup>-3</sup>) equating to >1% of the total community.

The most abundant Archaea present in all ponds, belonged to the phyla Woesearchaeota (42.5 ± 6.7%), a member of the DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota) superphylum, and Euryarchaeota (33.8 ± 5.8%) (Fig. 3). At the genus level, taxonomic assignments are *incertae sedis* within the phyla Woesearchaeota and have been grouped together (Fig. 4). The relative abundance was highly variable between ponds ranging from 8.3% up to 97.3% of the archaeal community. Genera within Euryarchaeota mainly consisted of strictly anaerobic methanogenic archaea which produce methane as a metabolic by-product in hypoxic conditions (Fig. 4). These include *Methanospirillum* (14.6%), *Methanotherix* (14.6 ± 2.7%), *Methanocella* (14.4 ± 9.1%), *Methanobacterium* (11.7 ± 4.8%), *Methanoregula* (11.4 ± 2.6%), *Methanopyrus* (9.7 ± 2.4%), *Methanosarcina* (9.0 ± 2.0%), *Methanomassiliococcus* (7.0 ± 0.6%) and *Methanimicrococcus* (6.9 ± 0.1%). In addition, *Halonotius* an extremely halophilic haloarchaeal genera was highly abundant in Pond 13 and 16, both of which are more saline than the other ponds as it has the highest concentrations of sodium (Table 3).

Within the fungi, Ascomycota (72.6 ± 4.4%) dominated across all ponds with the exception of Pond 9 and Pond 12, in which Zygomycota (12.4 ± 4.4%) and Basidiomycota (13.1 ± 2.4%) had the highest relative abundance, respectively (Fig. 3). The majority of Zygomycota in Pond 9 consisted of the genus *Mortierella* (71.3%), whereas for Pond 12 *Scleroderma* (17.9%) made up most of the detected Basidiomycota (Fig. 4). *Candida* was the most prevalent genus and showed the highest relative abundance in 12 out of 16 ponds. None of the *Candida* affiliated OTUs shared >99% sequence identity with pathogenic species in a human pathogen database, such as *C. albicans*, *C. glabrata*, *C. parapsilosis* and *C. tropicalis*. No human

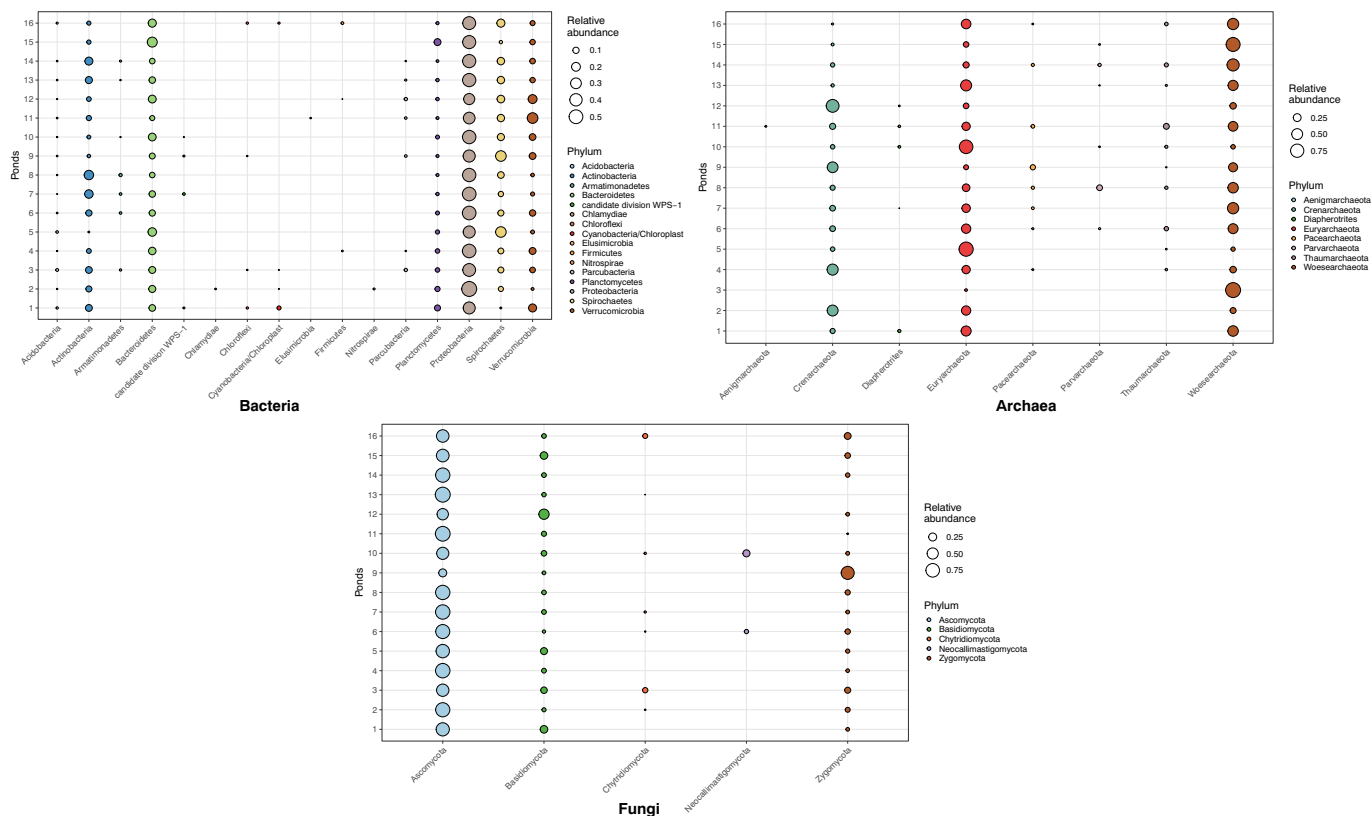


Fig. 3. Bubble plot showing the relative abundance (depicted by size) of the most abundant phyla in each Pond Harvesting System (PHS) within the bacterial, archaeal and fungal communities. Only phyla which account for >1% (0.01) of the relative abundance in a pond is included.

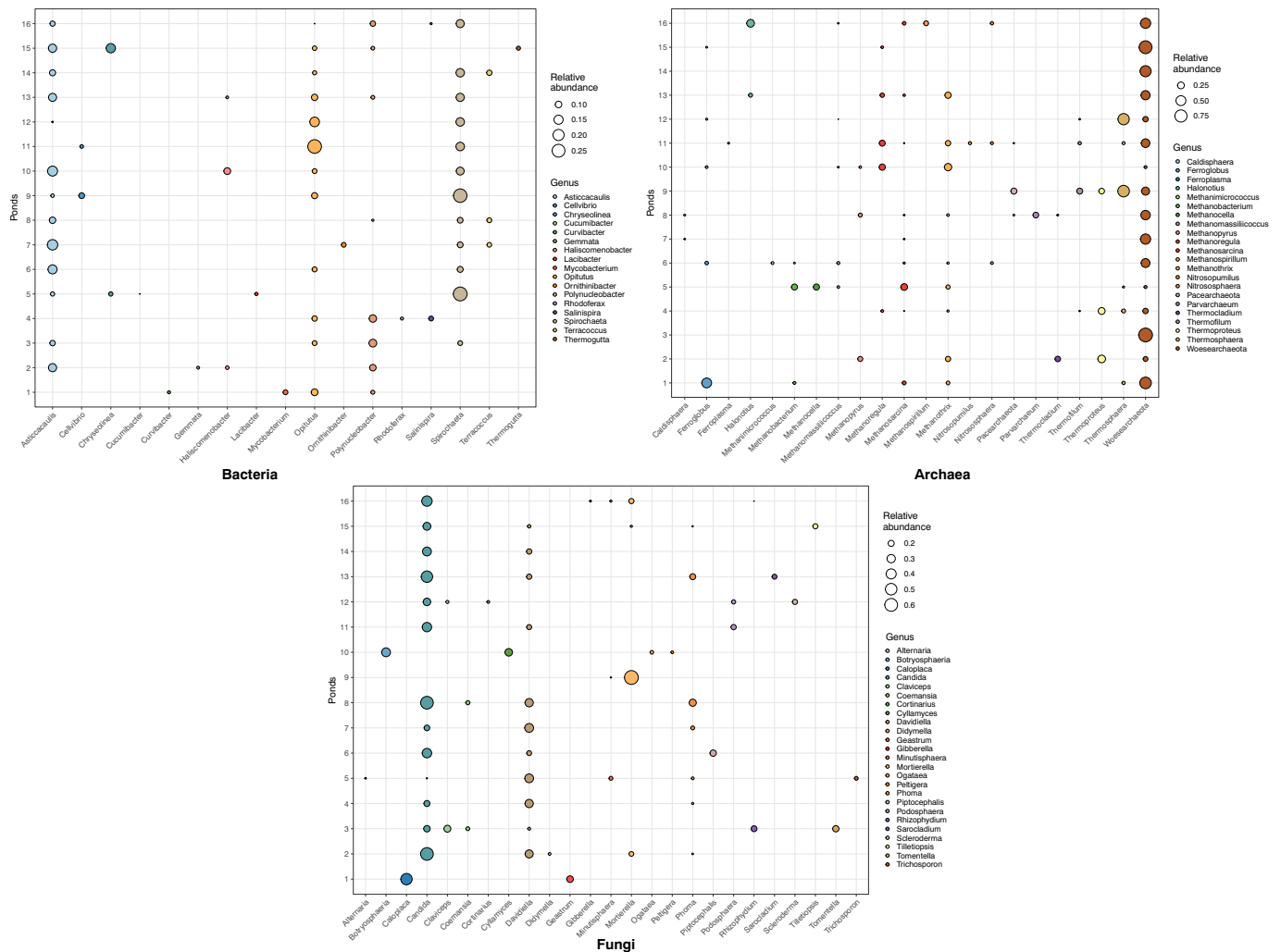


Fig. 4. Bubble plot showing the relative abundance (depicted by size) of the most abundant genera in each Pond Harvesting System (PHS) within the bacterial, archaeal and fungal communities. Only genera which account for >10% (0.10) of the relative abundance in a pond is included.

fungal pathogens were detected in our database. However, a number of genera, whose species include major plant pathogens, were abundant, such as *Botryosphaeria* (28.7%), *Rhizophyidium* (20.7%), *Claviceps* (17.9 ± 3.3%), *Phoma* (15.3 ± 2.1%), *Didymella* (12.8%), *Podosphaera* (12.8%), *Alternaria* (12.6%) and *Phyllosticta* (10.1%).

3.3. Differences in microbial community structure between ponds

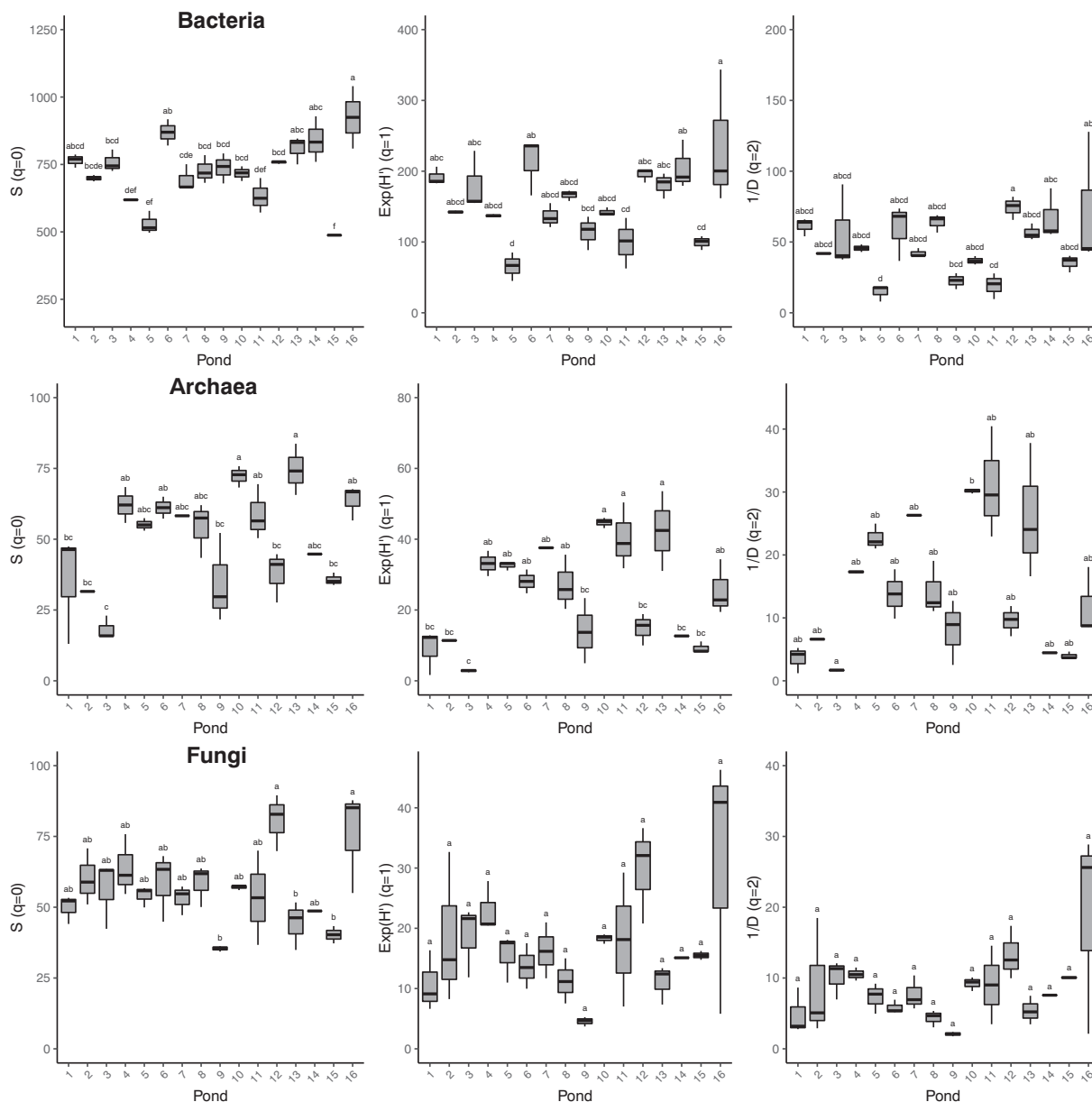
Within sample/alpha diversity was expressed using Hill numbers with the parameter order of diversity (q) from 0 to 2. For the bacterial

communities significant differences were observed for all orders of diversity; q = 0 (OTU richness/S) (ANOVA,  $F(15,30) = 14.38, P = 1.50 \times 10^{-9}$ ), q = 1 (exponential of Shannon entropy/Exp(H')) (ANOVA,  $F(15,30) = 5.851, P = 2.05 \times 10^{-5}$ ) and q = 2 (inverse Simpson index/1/D) (ANOVA,  $F(15,30) = 3.657, P = 0.001$ ) (Fig. 5). The average richness was 726 ± 18 OTUs. Post hoc comparisons with Tukey's HSD showed the differences for q = 0 were driven by a higher richness in Pond 16 (925 ± 67 OTUs) and a lower richness in Pond 5 (530 ± 25 OTUs) and Pond 15 (488 ± 1 OTUs). However, q = 0 does not take into account relative abundance of OTUs. The diversity in

Table 4

Operational taxonomic units (OTUs) detected in ponds which share >99% nucleotide identity with a human pathogen's 16S rRNA sequence.

OTU	Matching pathogen	Accession number	Nucleotide identity (%)	Potential diseases
OTU97	<i>Mycobacterium fortuitum</i>	NR_042912.1	99.51	Osteomyelitis, Skin and Joint Infections
OTU1226	<i>Pseudomonas stutzeri</i>	NR_118798.1	100	Endocarditis, Arthritis
OTU2770	<i>Staphylococcus epidermidis</i>	NR_113957.1	100	Wound Infections, Bacteraemia
OTU23248	<i>Cutibacterium acnes</i>	NR_040847.1	100	Blepharitis, Endophthalmitis
OTU39346	<i>Mycobacterium simiae</i>	NR_117227.1	100	Pulmonary Disease
OTU51780	<i>Neisseria elongata</i>	NR_025893.1	100	Endocarditis, Osteomyelitis
OTU58114	<i>Bartonella quintana</i>	NR_044748.1	100	Trench Fever
OTU104409	<i>Acinetobacter calcoaceticus</i>	NR_042387.1	100	Ophthalmic Infection, Chronic Synovitis
OTU111570	<i>Escherichia coli</i>	J01859.1	99.77	Urinary and Gastrointestinal Infections,
OTU193369	<i>Aeromonas bivalvium</i>	NR_043885.1	99.30	Gastroenteritis, Muscle infections
OTU92953	<i>Klebsiella oxytoca</i>	AJ871855.1	99.06	Colitis, Sepsis
OTU93931	<i>Streptococcus mutans</i>	NR_114726.1	99.53	Tooth Decay, Cardiovascular Disease



**Fig. 5.** Alpha diversity at each Pond Harvesting System (PHS) measured with Hill numbers (means  $\pm$  SE;  $n = 3$ ) of different orders of diversity ( $q = 0, 1$  and  $2$ ). These represent operational taxonomic unit (OTU) richness ( $S$ ), exponential of Shannon entropy ( $\text{Exp}(H')$ ) and inverse Simpson index ( $1/D$ ), respectively. As  $q$  increases, rare OTUs are given less weight and therefore contribute less toward “effective number of OTUs”. Hill numbers are shown for the bacterial, archaeal and fungal communities. Different letters (a, b, c, d, e) indicate a significant difference in Hill number between the ponds ( $P < 0.05$ , ANOVA followed by a Tukey’s HSD post hoc test or a Kruskal-Wallis test followed by Dunn’s post hoc test).

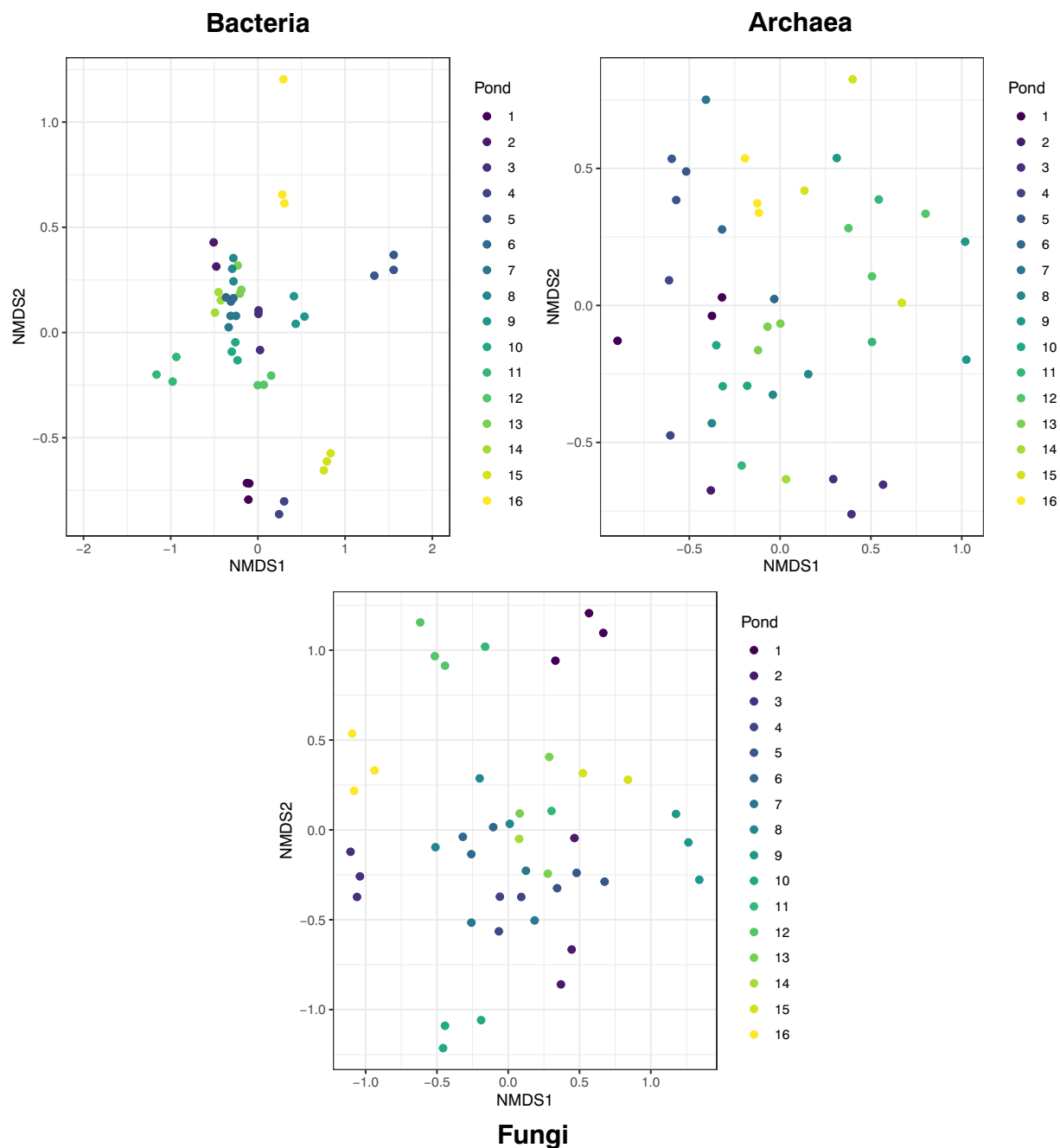
Pond 5 was also significantly lower at the orders  $q = 1$  and  $q = 2$  when we are considering OTUs that are common ( $66 \pm 12$  OTUs) and dominant ( $15 \pm 3$  OTUs), respectively. This indicates the bacterial community in that pond is not very diverse, being made up of, and dominated, by very few members. A similar pattern was also seen for the archaeal community where there differences in diversity between ponds at all orders;  $q = 0$  (ANOVA,  $F(15, 24) = 7.595, P = 7.67 \times 10^{-6}$ ),  $q = 1$  (ANOVA,  $F(15, 24) = 11.37, P = 1.76 \times 10^{-7}$ ) and  $q = 2$  (Kruskal-Wallis,  $\chi^2 = 34.621, P = 2.78 \times 10^{-3}$ ) (Fig. 5). These differences were mainly driven by Pond 3 which had significantly lower Hill numbers at all orders.

For the fungal communities there was also a difference, but only at the order of  $q = 0$  (ANOVA,  $F(15, 28) = 3.74, P = 1.27 \times 10^{-3}$ ). These differences were between the two ponds with the highest OTU

richness, Pond 12 ( $81 \pm 6$  OTUs) and Pond 16 ( $76 \pm 11$  OTUs), and the three ponds with the lowest, Pond 9 ( $36 \pm 1$  OTUs), Pond 13 ( $45 \pm 5$  OTUs) and Pond 15 ( $41 \pm 4$  OTUs). At the orders of  $q = 1$  or  $q = 2$ , which place increasing weight on common and dominant OTUs, there was no significant difference between ponds (Fig. 5).

Non-metric multidimensional scaling (NMDS) of Bray-Curtis dissimilarities showed the microbial communities were clearly separated by site with the separation of bacterial communities being the most distinguishable (Fig. 6). ANOSIM confirmed significant separation of the bacterial ( $R = 0.9981, P < 0.001$ ), archaeal ( $R = 0.936, P < 0.001$ ) and fungal ( $R = 0.7732, P < 0.001$ ) communities by site. PERMANOVA also demonstrated bacterial ( $R^2 = 0.058, P = 0.001$ ), archaeal ( $R^2 = 0.049, P = 0.006$ ) and fungal communities ( $R^2 = 0.058, P = 0.009$ ) were significantly differentiated by site.

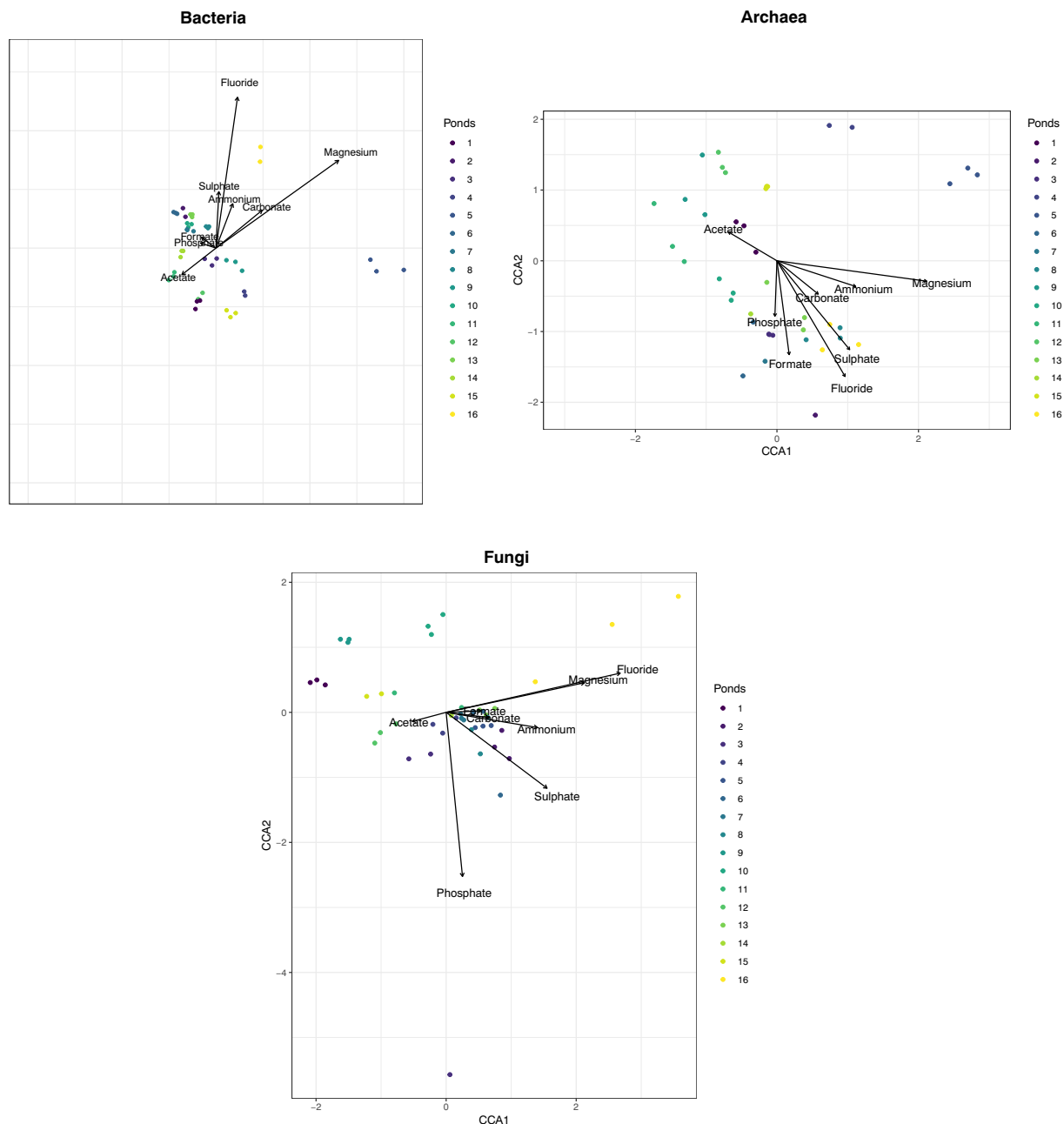




**Fig. 6.** Non-metric multidimensional scaling (NMDS) ordination of the microbial communities present among the different Pond Harvesting Systems (PHS). The plot is based on rarefied operational taxonomic unit (OTU) abundance and calculated using Bray-Curtis dissimilarity between sites.

Canonical correspondence analysis (CCA) was also performed to discern the possible relationship between community structure in each pond and nutrient concentrations (Fig. 7). Eight environmental variables were included in the CCA biplot, as their variance inflation factors were below 10, and they did not exceed the multicollinearity threshold of 0.7. Acetate, ammonium, carbonate, fluoride, formate, magnesium, phosphate and sulphate concentration were included. The models were significant between bacterial (Monte Carlo test,  $P = 0.001$ , 999 permutations), archaeal ( $P = 0.008$ ) and fungal ( $P = 0.001$ ) community structures and the chemical variables. For the bacterial community these chemical variables explain 28% of variance and the strongest predictor variables were fluoride and magnesium concentration ( $P = 0.001$ ). Other significant variables include acetate ( $P = 0.014$ ), carbonate ( $P = 0.022$ ) and sulphate ( $P = 0.040$ ). All

significant variables, with the exception of acetate, appear to strongly influence the bacterial community in Pond 16. Acetate influenced the bacterial community in Pond 11. Only 24% of archaeal community variations could be explained by the eight environmental variables. The amount of variability explained by the chemical variables is only significant for the first canonical axis ( $P = 0.02$ ) and three chemical variables including fluoride ( $P = 0.008$ ), magnesium ( $P = 0.001$ ) and sulphate ( $P = 0.017$ ). The chemical variables explained 25% of fungal community structure variations. As with the bacterial community, fluoride concentration was the strongest predictor variable ( $P = 0.001$ ), and strongly influenced Pond 16, which has the highest fluoride concentration ( $60.33 \pm 6.84 \mu\text{mol/l}$ ; Table 2). All other variables were significant with the exception of ammonium and formate.



**Fig. 7.** Canonical correspondence analysis (CCA) ordinations of the microbial community structure and chemical variables (acetate, ammonium, carbonate, fluoride, formate, magnesium, phosphate and sulphate concentration) for each Pond Harvesting System (PHS). Sites that are close together on the ordination diagram are more similar (in terms of their microbial community structure) than sites that are farther apart. Vectors affiliated with each of the axes represent a chemical variable. The length of a vector in the ordination plot is indicative of the strength of the relationship of that parameter to community composition.

## 4. Discussion

### 4.1. Microbial community trends

The composition of the bacterial and archaeal communities in the ponds suggest they are either hypoxic or anoxic environments, supporting the growth of anaerobic microorganisms. For example, in the bacterial community OTUs affiliated with the genera *Spirochaeta* and *Opitutus* dominated. Members of the genus *Spirochaeta* are obligate or facultative anaerobes frequently isolated and detected in diverse anoxic environments including anchialine sinkholes (Davis and Garey, 2018), soda lakes (Vavourakis et al., 2018), microbial mats (Spring et al., 2015), marine intertidal muds (Harwood and Canale-Parola, 1983) and marine sediments (Miyazaki et al., 2014). In these anoxic environments *Spirochaeta* spp. are the trophic

intermediate between hydrolytic bacteria and secondary anaerobes, as the main compounds produced by spirochetes are acetate, H<sub>2</sub> and CO<sub>2</sub>, which are normally consumed by methanogens (Berlanga et al., 2008; Blazejak et al., 2005). The genus *Opitutus* currently only consists of one cultured representative, *Opitutus terrae*, an obligatory anaerobic member of the phylum Verrucomicrobia (van Passel et al., 2011). Whilst this strain was originally isolated in anoxic rice paddy soil, *Opitutus* spp. have also been detected in aquatic habitats (Arnds et al., 2010; Bai et al., 2019). The metabolism of *Opitutus* spp. is suited for growth on plant-derived polysaccharides and for interaction with methanogens with its metabolic by-products (Chin and Janssen, 2002; Janssen, 1998).

With regard to archaeal communities, representatives of the phylum Woesearchaeota and methanogens were dominant. Woesearchaeota mainly grow anaerobically as they lack a complete electron transport chain, a

complete TCA cycle and continuous glycolysis, which may explain their ecological distribution pattern of being found in primarily inland anoxic environments (Castelle et al., 2015; Liu et al., 2018). Methanogens are strict anaerobes and most methanogenic niches are in oxygen-free aqueous systems that provide a supply of organic matter (Hoehler et al., 2010). These include aquatic sediments, wetlands, agricultural or natural soils subject to inundation, sewage digesters and anoxic portions of animal digestive tracts (Chaban et al., 2006; Liu and Whitman, 2008). Lineage abundance distribution and co-occurrence network analyses across diverse biotopes confirmed metabolic complementation and revealed a potential syntrophic relationship between Woesearchaeota and methanogens, indicating Woesearchaeota may also impact methanogenesis in inland ecosystems (Liu et al., 2018). This relationship, along with support of growth provided by the dominant anaerobic bacteria, may explain the prevalence and dominance of methanogens in our study.

Fluoride and magnesium were found to be strong predictor variables for the microbial communities. The occurrence of fluoride in these ponds are likely due to natural processes, such as weathering and dissolution of minerals (García and Borgnino, 2015). Fluoride has been shown to have a strong inhibitory effect on the growth of anaerobic microorganisms (Ji et al., 2016; Ochoa-Herrera et al., 2009). In addition, magnesium can interact with fluoride, forming metal complexes, which also impacts bacterial growth and metabolic activity by inhibiting phosphate transfer enzymes (Baxter et al., 2008). Given that anaerobes are prevalent in this study, distinct communities between ponds could form due to the differences in fluoride and magnesium fluoride complex concentrations.

#### 4.2. Prevention of anaerobic conditions in ponds

As the PHSs are openly exposed, large amounts of organic matter can get into them. Decomposition of organic matter exerts an oxygen demand, and accumulation of excess organic matter may lead to dissolved oxygen depletion, promoting anaerobic growth conditions (Dai et al., 2018). Oxygen is an important irrigation water quality parameter that can be the limiting factor in some agricultural systems (Bhattarai et al., 2004, 2005). Low oxygen in irrigation water can lead to root oxygen deficiency, which in turn causes agronomic problems such as crop stress, slow plant growth and low yields (Bhattarai et al., 2008; Maestre-Valero and Martínez-Alvarez, 2010). Therefore, to overcome this impact on water quality, adjustments can be made to the pond's design preventing oxygen being depleted and hypoxic/anoxic conditions developing. The amount of organic matter going into the pond must be minimised in order to reduce oxygen consumption. Ensure that prior to the pond's construction a location is chosen that reduces this risk. For example, position the pond away from trees to stop leaves falling in, and fence animals to prevent them from defecating in it. A low wall could also be constructed to protect the pond from surface runoff. Once the pond is built, a filtration system could be used to keep debris from entering the pond. The pond could be covered with some form of netting, mesh, silt trap or strainer, positioned either directly on top of or beside it. If material does enter the water, a pond skimmer or fishing net could also be used to remove it.

Another method would be to regularly oxygenate the water. Water aeration is commonly used to increase oxygen concentration in lakes and rivers and increases water quality (Alp and Melching, 2011; Mahmud et al., 2020). There are many different types of aerators that could be used in the pond, such as fine pore aerators (Shammas, 2007), or surface aerators (Rosso et al., 2008). However, venturi aeration may be a simpler and more cost-effective solution for farmers. When a minimal amount of pressure exists between the inlet and outlet sides of a venturi tube, a vacuum occurs. Air that is entrained into the water is instantly forced downstream in the form of small air bubbles and so increases the oxygen levels in the water (Baylar et al., 2010; Baylar and Ozkan, 2006). The reduced cost of purchase and maintenance of venturi aerators compared to others may make them viable alternatives, in low resource areas, such as Africa (Mahmud et al., 2020; Therrien et al., 2019). The use of venturi aeration has successfully been applied to irrigation water for agriculture production and increased yields of the crops (Bagatur, 2014; Dahrazma et al., 2019). Alternatively,

if the sole purpose of the water is for irrigation of crops, hydrogen peroxide can be added to the water. Hydrogen peroxide should decompose into water and molecular oxygen within a few hours in natural waters (Spool et al., 2020). The addition of hydrogen peroxide to irrigation water has been shown to be an effective method to increase oxygen concentration and improve plant parameters in Africa, and other parts of the world (Abd Elhady et al., 2021; Bhattarai et al., 2004). However, if too high concentrations of hydrogen peroxide is used this can lead to oxidative stress on the plants (Ben-Noah and Friedman, 2016; Gil M et al., 2009), which will reduce overall yield, and may pose a health risk to the farmers. Additionally, the use of water plants that are native to the area could be introduced into the pond to supply oxygen via photosynthesis such as water hyacinths.

#### 4.3. Removal of human pathogens in PHS

Although the relative abundance of potential bacterial human health pathogens was low in the ponds, our data indicates a potential human health risk. Specifically, the non-tuberculous mycobacteria (NTM), *Mycobacterium fortuitum*, was found to be present in one of the ponds, comprising >1% of the bacterial community. In terms of absolute abundance, our results suggest over 36,000 pathogenic *Mycobacterium* 16S rRNA copies per ml of water. Non-tuberculosis mycobacteria (NTM) are ubiquitous environmental organisms increasingly recognised as important human pathogens (Kasperbauer and De Groote, 2015; Sfeir et al., 2018). *M. fortuitum* is considered rapidly growing and is distinguished from other NTM by their ability to form colonies in less than one week and their in vitro resistance to antimycobacterials (Daley and Griffith, 2002; Park et al., 2008). It can cause skin disease, osteomyelitis, joint infections, infections of the eye, and pulmonary infections (Griffith et al., 2007; Okamori et al., 2018). Harvested rainwater can be used for many household purposes, such as, drinking, swimming and bathing (Domènech and Sauri, 2011; Hofman-Caris et al., 2019). It is these activities where humans are likely to be exposed to waterborne NTM (Thomson et al., 2013). Furthermore, aerosols generated during these activities can be inhaled potentially resulting in disease (Falkinham, 2003; Lever et al., 2000). Given low doses of *Mycobacterium* are required to establish an infection (Johnson et al., 2007; Saini et al., 2012) this suggests usage of the harvested water may be a human health risk. Furthermore, NTM are extremely resistant to many chemical disinfectants used to treat water, such as chlorine, chlorine dioxide, monochloramine and ozone (Taylor et al., 2000). Given this resistance to chemical disinfection, propensity to form biofilms (Faria et al., 2015) and ability to survive in low carbon environments (Falkinham, 2003), complete removal of all pathogenic NTM from harvested water will be difficult. Physical methods of water disinfection, such as Pall filtration and UV-C disinfection have proved effective in significantly reducing the viable NTM population in drinking water (Norton et al., 2020). There are a number of commercially available water purification systems which uses these methods and could provide a simple way for individuals to minimise their exposure to waterborne NTM. However, due to the low incomes of the rural smallholder farmers these options may not be feasible without financial support of non-governmental or charitable organisations.

#### 4.4. Limitations of the study

A limitation of this study to evaluate potential human health risks is the taxonomic resolution achieved through amplicon sequencing, which can only give identification to the genus level. Furthermore, with amplicon sequencing it remains challenging to differentiate between viable/metabolically active, and non-viable microorganisms (Emerson et al., 2017; Wang et al., 2021). Indeed in this study, although opportunistic pathogenic groups were detected, information is lacking on their viability. However, amplicon sequencing represents the most efficient and cost-effective solution to provide a taxonomic overview and screen for potentially pathogenic groups, compared to PCR-based identification on a pathogen-by-pathogen basis, which would be time-consuming and requires prior knowledge of which groups to target.

Shotgun metagenomic sequencing could be an alternative approach, allowing for strain-level identification of pathogens as well as associations between phylogeny and function (Quince et al., 2017). However, shotgun metagenomics has much higher input DNA requirements compared to amplicon sequencing (Sui et al., 2020), which may be difficult to achieve with the filtered harvested rainwater samples used in this study. A further limitation of metagenomics for pathogen detection is that abundant taxa get most of the coverage, meaning rare pathogens could be missed or not sequenced with enough depth for identification (Ferguson et al., 2021).

To fully validate our results, it would be necessary to use a combination of amplicon sequencing, to provide a broad identification of taxa, shotgun metagenomics, to identify strains and virulence genes, and PCR-based methods for specific confirmation of pathogen presence. Once the presence of a pathogen has been confirmed using molecular methods, viability can be determined using for example culture-dependent methods (e.g. plate counts), or alternative techniques such as flow cytometry, or viability dyes coupled with qPCR. This would provide an effective method to fully assess potential human health risks.

## 5. Conclusions

In conclusion, this study characterised the microbial ecology of PHS across taxonomic domains. We identified the presence of microbial pathogens, notably, *Mycobacterium fortuitum* in one of the ponds, which is a human health concern as mycobacteria are rarely targeted in screening of microbial water quality with standard techniques (e.g. heterotrophic plate counts) and would be missed. This study highlights the need for a broader screening of microbial community composition through HTS, to identify all potential pathogens, followed by a more sensitive molecular method (e.g. qPCR) to confirm their presence and abundance. The abundance of anaerobes and methanogens, which are driven by fluoride and magnesium concentrations, indicates poor water quality of supplies being used to irrigate crops, which in turn could reduce crop yield and viability. Here we suggest changes to PHS management (e.g. pond design, aeration) to ameliorate hypoxic or anoxic conditions and allow the transport of aerated water to roots, promoting plant water uptake and maximising crop yield. We also suggest physical methods of water disinfection, such as Pall filtration and UV-C disinfection, to mitigate pathogen contamination in the ponds. Overall, this will be beneficial to the livelihoods and welfare of the rural smallholders in Kenya.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2022.153040>.

## Funding

We thank Global Challenges Research Fund 'Investigating the effectiveness of a low technology planter for subsistence farming' (Project A004, GCRF74) and NERC (NE/S005560/1 and NE/S006958/1) for funding this work.

## CRedit authorship contribution statement

CW and DH conceived and designed the experiment. CW and DH acquired funding. LH carried out the sampling campaign. BG performed the molecular analysis. BG and AB performed the data and statistical analysis. BG wrote the manuscript, supplementary information, and tables. BG and AB produces the figures. All authors reviewed and contributed to the submitted version.

## Declaration of competing interest

All authors declare no conflict of interest.

## Acknowledgements

We thank Mr. John Green for technical assistance.

## References

- Abd Elhady, S.A., El-Gawad, H.G.A., Ibrahim, M.F.M., Mukherjee, S., Elkelish, A., Azab, E., Gobouri, A.A., Farag, R., Ibrahim, H.A., El-Azm, N.A., 2021. Hydrogen peroxide supplementation in irrigation water alleviates drought stress and boosts growth and productivity of potato plants. *Sustainability* 13, 899. <https://doi.org/10.3390/su13020899>.
- Acharya, K., Khanal, S., Pantha, K., Amaty, N., Davenport, R.J., Werner, D., 2019. A comparative assessment of conventional and molecular methods, including MinION nanopore sequencing, for surveying water quality. *Sci. Rep.* 9, 15726. <https://doi.org/10.1038/s41598-019-51997-x>.
- Ahmed, W., Goonetilleke, A., Gardner, T., 2010. Implications of faecal indicator bacteria for the microbiological assessment of roof-harvested rainwater quality in Southeast Queensland, Australia. *Can. J. Microbiol.* 56, 471–479. <https://doi.org/10.1139/W10-037>.
- Ahmed, W., Gardner, T., Toze, S., 2011. Microbiological quality of roof-harvested rainwater and health risks: a review. *J. Environ. Qual.* 40, 13–21. <https://doi.org/10.2134/jeq2010.0345>.
- Ahmed, W., Brandes, H., Gyawali, P., Sidhu, J.P.S., Toze, S., 2014. Opportunistic pathogens in roof-captured rainwater samples, determined using quantitative PCR. *Water Res.* 53, 361–369. <https://doi.org/10.1016/j.watres.2013.12.021>.
- Ahmed, W., Staley, C., Hamilton, K.A., Beale, D.J., Sadowsky, M.J., Toze, S., Haas, C.N., 2017. Amplicon-based taxonomic characterization of bacteria in urban and peri-urban roof-harvested rainwater stored in tanks. *Sci. Total Environ.* 576, 326–334. <https://doi.org/10.1016/j.scitotenv.2016.10.090>.
- Alp, E., Melching, C.S., 2011. Allocation of supplementary aeration stations in the Chicago waterway system for dissolved oxygen improvement. *J. Environ. Manag.* 92, 1577–1583. <https://doi.org/10.1016/j.jenvman.2011.01.014>.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>.
- Arnds, J., Knittel, K., Buck, U., Winkel, M., Amann, R., 2010. Development of a 16S rRNA-targeted probe set for verrucomicrobia and its application for fluorescence in situ hybridization in a humic lake. *Syst. Appl. Microbiol.* 33, 139–148. <https://doi.org/10.1016/j.syapm.2009.12.005>.
- Ayers, R.S., Westcot, D.W., 1994. *Water quality for agriculture*. FAO Irrigation and Drainage Paper 29 Rev. 1. M-56. Food and Agricultural Organization (FAO), Rome (Italy).
- Bagatur, T., 2014. Evaluation of plant growth with aerated irrigation water using venturi pipe part. Arab. J. Sci. Eng. 39, 2525–2533. <https://doi.org/10.1007/s13369-013-0895-4>.
- Bai, Y., Ruan, X., Xie, X., Yan, Z., 2019. Antibiotic resistome profile based on metagenomics in raw surface drinking water source and the influence of environmental factor: a case study in Huaihe River Basin, China. *Environ. Pollut.* 248, 438–447. <https://doi.org/10.1016/j.envpol.2019.02.057>.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Pribelaski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Baxter, N.J., Blackburn, G.M., Marston, J.P., Hounslow, A.M., Cliff, M.J., Bermel, W., Williams, N.H., Hollfelder, F., Wemmer, D.E., Waltho, J.P., 2008. Anionic charge is prioritized over geometry in aluminum and magnesium fluoride transition state analogs of phosphoryl transfer enzymes. *J. Am. Chem. Soc.* 130, 3952–3958. <https://doi.org/10.1021/ja078000n>.
- Baylar, A., Ozkan, F., 2006. Applications of venturi principle to water aeration systems. *Environ. Fluid Mech.* 6, 341–357. <https://doi.org/10.1007/s10652-005-5664-9>.
- Baylar, A., Ozkan, F., Unsal, M., 2010. Effect of air inlet hole diameter of venturi tube on air injection rate. *KSCIE J. Civ. Eng.* 14, 489–492. <https://doi.org/10.1007/s12205-010-0489-6>.
- Bell, T., Freckleton, R.P., Lewis, O.T., 2006. Plant pathogens drive density-dependent seedling mortality in a tropical tree. *Ecol. Lett.* 9, 569–574. <https://doi.org/10.1111/j.1461-0248.2006.00905.x>.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. <https://doi.org/10.2307/2346101>.
- Ben-Noah, I., Friedman, S.P., 2016. Oxygation of clayey soils by adding hydrogen peroxide to the irrigation solution: lysimetric experiments. *Rhizosphere* 2, 51–61. <https://doi.org/10.1016/j.rhisp.2016.08.002>.
- Berhane, G., 2018. Benefits and challenges of dugout rainwater harvesting ponds in Tigray Region, Ethiopia. *Rainwater-smart Agriculture in Arid and Semi-arid Areas*. Springer International Publishing, Cham, pp. 259–280. [https://doi.org/10.1007/978-3-319-66239-8\\_14](https://doi.org/10.1007/978-3-319-66239-8_14).
- Berlanga, M., Aas, J.A., Paster, B.J., Boumenna, T., Dewhurst, F.E., Guerrero, R., 2008. Phylogenetic diversity and temporal variation in the spirochaeta populations from two Mediterranean microbial mats. *Int. Microbiol.* 11, 267. <https://doi.org/10.2436/20.1501.01.71>.
- Bhattarai, S.P., Huber, S., Midmore, D.J., 2004. Aerated subsurface irrigation water gives growth and yield benefits to zucchini, vegetable soybean and cotton in heavy clay soils. *Ann. Appl. Biol.* 144, 285–298. <https://doi.org/10.1111/j.1744-7348.2004.tb00344.x>.
- Bhattarai, S.P., Su, N., Midmore, D.J., 2005. Oxygation unlocks yield potentials of crops in oxygen-limited soil environments. *Adv. Agron.* 88, 313–377. [https://doi.org/10.1016/S0065-2113\(05\)88008-3](https://doi.org/10.1016/S0065-2113(05)88008-3).
- Bhattarai, S.P., Midmore, D.J., Pendergast, L., 2008. Yield, water-use efficiencies and root distribution of soybean, chickpea and pumpkin under different subsurface drip irrigation depths and oxygation treatments in vertisols. *Irrig. Sci.* 26, 439–450. <https://doi.org/10.1007/s00271-008-0112-5>.
- Biazin, B., Sterk, G., Temesgen, M., Abdulkedir, A., Stroosnijder, L., 2012. Rainwater harvesting and management in rainfed agricultural systems in sub-Saharan Africa – a review. *Phys. Chem. Earth A/B/C* 47–48, 139–151. <https://doi.org/10.1016/j.pce.2011.08.015>.

- Blazejak, A., Erséus, C., Amann, R., Dubilier, N., 2005. Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru margin. *Appl. Environ. Microbiol.* 71, 1553–1561. <https://doi.org/10.1128/AEM.71.3.1553-1561.2005>.
- Boelee, E., Yohannes, M., Poda, J.-N., McCartney, M., Cecchi, P., Kibret, S., Hagos, F., Laamrani, H., 2013. Options for water storage and rainwater harvesting to improve health and resilience against climate change in Africa. *Reg. Environ. Chang.* 13, 509–519. <https://doi.org/10.1007/s10113-012-0287-4>.
- Bonferroni, C., 1936. *Teoria statistica delle classi e calcolo delle probabilità*. *Pubbl. del R. Ist. Super. di Sci. Econ. e Commerciali di Firenze*, 8, pp. 3–62.
- Bray, J.R., Curtis, J.T., 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27, 325–349. <https://doi.org/10.2307/1942268>.
- Brodie, E.L., DeSantis, T.Z., Parker, J.P.M., Zubietta, I.X., Piceno, Y.M., Andersen, G.L., 2007. Urban aerosols harbor diverse and dynamic bacterial populations. *Proc. Natl. Acad. Sci.* 104, 299–304. <https://doi.org/10.1073/PNAS.0608255104>.
- Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J., Frischkorn, K.R., Tringe, S.G., Singh, A., Markillie, L.M., Taylor, R.C., Williams, K.H., Banfield, J.F., 2015. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* 25, 690–701. <https://doi.org/10.1016/J.CUB.2015.01.014>.
- Chaban, B., Ng, S.Y., Jarrell, K.F., 2006. Archaeal habitats — from the extreme to the ordinary. *Can. J. Microbiol.* 52, 73–116. <https://doi.org/10.1139/w05-147>.
- Chan, A.W., Naphtali, J., Schellhorn, H.E., 2019. High-throughput DNA sequencing technologies for water and wastewater analysis. *Sci. Prog.* 102, 351–376. <https://doi.org/10.1177/0036850419881855>.
- Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., Ellison, A.M., 2014. Rarefaction and extrapolation with hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* 84, 45–67. <https://doi.org/10.1890/13-0133.1>.
- Chidamba, L., Korsten, L., 2015. Pyrosequencing analysis of roof-harvested rainwater and river water used for domestic purposes in luthengele village in the eastern Cape Province of South Africa. *Environ. Monit. Assess.* 187, 41. <https://doi.org/10.1007/s10661-014-4237-0>.
- Chin, K.-J., Janssen, P.H., 2002. Propionate formation by *Opiritatus terrae* in pure culture and in mixed culture with a hydrogenotrophic methanogen and implications for carbon fluxes in anoxic rice paddy soil. *Appl. Environ. Microbiol.* 68, 2089. <https://doi.org/10.1128/AEM.68.4.2089-2092.2002>.
- Clark, D.R., McKew, B.A., Dong, L.F., Leung, G., Dumbrell, A.J., Stott, A., Grant, H., Nedwell, D.B., Trimmer, M., Whitty, C., 2020. Mineralization and nitrification: archaea dominate ammonia-oxidising communities in grassland soils. *Soil Biol. Biochem.* 143, 107725. <https://doi.org/10.1016/J.SOILBIO.2020.107725>.
- Clarke, K.R., 1993. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117–143. <https://doi.org/10.1111/j.1442-9993.1993.tb00438.x>.
- Cook, S., Sharma, A., Chong, M., 2013. Performance analysis of a communal residential rainwater system for potable supply: a case study in Brisbane, Australia. *Water Resour. Manag.* 27, 4865–4876. <https://doi.org/10.1007/s11269-013-0443-8>.
- Dahrazma, B., Naghedinia, A., Ghasemian Gorji, H., Saghravani, S.F., 2019. Morphological and physiological responses of *Cucumis sativus* L. to water with micro-nanobubbles. *J. Agric. Sci. Technol.* 21, 181–192.
- Dai, L., Liu, C., Yu, L., Song, C., Peng, L., Li, X., Tao, L., Li, G., 2018. Organic matter regulates ammonia-oxidizing bacterial and archaeal communities in the surface sediments of *Ctenopharyngodon idellus* aquaculture ponds. *Front. Microbiol.* 9. <https://doi.org/10.3389/FMICB.2018.02290>.
- Daley, C.L., Griffith, D.E., 2002. Pulmonary disease caused by rapidly growing mycobacteria. *Clin. Chest Med.* 23, 623–632. [https://doi.org/10.1016/S0272-5231\(02\)00021-7](https://doi.org/10.1016/S0272-5231(02)00021-7).
- Davis, M., Garey, J., 2018. Microbial function and hydrochemistry within a stratified anchialine sinkhole: a window into coastal aquifer interactions. *Water* 10, 972. <https://doi.org/10.3390/w10080972>.
- De Mendiburu, F., 2020. R Package ‘Agricolae’. Available online <https://cran.r-project.org/web/packages/agricolae/index.html>.
- Despins, C., Farahbaksh, K., Leidl, C., 2009. Assessment of rainwater quality from rainwater harvesting systems in Ontario, Canada. *J. Water Supply Res. Technol.* 58, 117–134. <https://doi.org/10.2166/aqua.2009.013>.
- Doménech, L., Saurí, D., 2011. A comparative appraisal of the use of rainwater harvesting in single and multi-family buildings of the metropolitan area of Barcelona (Spain): social experience, drinking water savings and economic costs. *J. Clean. Prod.* 19, 598–608. <https://doi.org/10.1016/J.JCLEPRO.2010.11.010>.
- dos Santos, S.M., de Farias, M.M.M.W.E.C., 2017. Potential for rainwater harvesting in a dry climate: assessments in a semiarid region in northeast Brazil. *J. Clean. Prod.* 164, 1007–1015. <https://doi.org/10.1016/J.JCLEPRO.2017.06.251>.
- Dumbrell, A.J., Ferguson, R.M.W., Clark, D.R., 2016. Microbial Community Analysis by Single-amplicon high-throughput Next Generation Sequencing: Data Analysis – From Raw Output to Ecology. Springer, Berlin, Heidelberg, pp. 155–206. <https://doi.org/10.1007/978-3-662-22016-228>.
- Dunn, O.J., 1964. Multiple comparisons using rank sums. *Technometrics* 6, 241–252. <https://doi.org/10.1080/00401706.1964.10490181>.
- Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. <https://doi.org/10.1093/bioinformatics/btr381>.
- Emerson, J.B., Adams, R.L., Román, C.M.B., Brooks, B., Coil, D.A., Dahlhausen, K., Ganz, H.H., Hartmann, E.M., Hsu, T., Justice, N.B., Paulino-Lima, I.G., Luongo, J.C., Lympyropoulou, D.S., Gomez-Silvan, C., Rothschild-Mancinelli, B., Balk, M., Huttenhower, C., Nocker, A., Vaishampayan, P., Rothschild, L.J., 2017. Schrödinger’s microbes: tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome* 5, 86. <https://doi.org/10.1186/s40168-017-0285-3>.
- Falkingham III, J.O., 2003. Mycobacterial aerosols and respiratory disease. *Emerg. Infect. Dis.* 9, 763. <https://doi.org/10.3201/EID0907.02-0415>.
- Faria, S., Joao, I., Jordao, L., 2015. General overview on nontuberculous mycobacteria, biofilms, and human infection. *J. Pathog.* 2015. <https://doi.org/10.1155/2015/809014>.
- Farreny, R., Morales-Pinzón, T., Guisasaola, A., Tayá, C., Riederavall, J., Gabarrell, X., 2011. Roof selection for rainwater harvesting: quantity and quality assessments in Spain. *Water Res.* 45, 3245–3254. <https://doi.org/10.1016/J.WATRES.2011.03.036>.
- Ferguson, R.M.W., Neath, C.E.E., Nasir, Z.A., Garcia-Alcega, S., Tyrrell, S., Coulon, F., Dumbrell, A.J., Colbeck, I., Whitty, C., 2021. Size fractionation of bioaerosol emissions from green-waste composting. *Environ. Int.* 147, 106327. <https://doi.org/10.1016/J.ENVINT.2020.106327>.
- García, M.G., Borgnino, L., 2015. Chapter 1. Fluoride in the Context of the Environment. *Royal Society of Chemistry*, pp. 3–21. <https://doi.org/10.1039/9781782628507-00003>.
- Gil M, P.M., Ferreyra, E.R., Barrera, M.C., Zúñiga, E.C., Gurovich, R.L., 2009. Effect of injecting hydrogen peroxide into heavy clay loam soil on plant water status, NET CO<sub>2</sub> assimilation, biomass, and vascular anatomy of avocado trees. *Chil. J. Agric. Res.* 69, 97–106. <https://doi.org/10.4067/S0718-58392009000100012>.
- Griffith, D.E., Aksamit, T., Brown-Elliott, B.A., Catanzaro, A., Daley, C., Gordin, F., Holland, S.M., Horsburgh, R., Huit, G., Iademarco, M.F., Iseman, M., Olivier, K., Ruoss, S., von Reyn, C.F., Wallace, R.J., Winthrop, K., 2007. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am. J. Respir. Crit. Care Med.* 175, 367–416. <https://doi.org/10.1164/rccm.200604-571ST>.
- Hafizi Md Lani, N., Yusop, Z., Syafuddin, A., 2018. A review of rainwater harvesting in Malaysia: prospect and challenges. *Water* 10, 506. <https://doi.org/10.3390/w10040506>.
- Hageskal, G., Lima, N., Skaar, I., 2009. The study of fungi in drinking water. *Mycol. Res.* 113, 165–172. <https://doi.org/10.1016/J.MYCRE.2008.10.002>.
- Hamilton, K., Reyneke, B., Waso, M., Clements, T., Ndlovu, T., Khan, W., DiGiovanni, K., Rakestraw, E., Motalto, F., Haas, C.N., Ahmed, W., 2019. A global review of the microbiological quality and potential health risks associated with roof-harvested rainwater tanks. *nj CleanWater 2*, 7. <https://doi.org/10.1038/s41545-019-0030-5>.
- Harwood, C.S., Canale-Parola, E., 1983. *Spirochaeta isovaleric* sp. nov., a marine anaerobe that forms branched-chain fatty acids as fermentation products. *Int. J. Syst. Bacteriol.* 33, 573–579. <https://doi.org/10.1099/00207713-33-3-573>.
- Hill, M.O., 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54, 427–432. <https://doi.org/10.2307/1934352>.
- Hoehler, T., Gunsalus, R.P., McInerney, M.J., 2010. Environmental constraints that limit methanogenesis. *Handbook of Hydrocarbon And Lipid Microbiology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 635–654. [https://doi.org/10.1007/978-3-540-77587-4\\_51](https://doi.org/10.1007/978-3-540-77587-4_51).
- Hofman-Caris, R., Bertelkamp, C., de Waal, L., van den Brand, T., Hofman, J., van der Aa, R., van der Hoek, J., 2019. Rainwater harvesting for drinking water production: a sustainable and cost-effective solution in the Netherlands? *Water* 11, 511. <https://doi.org/10.3390/w11030511>.
- Huby, T.J.C., Clark, D.R., McKew, B.A., McGenity, T.J., 2021. Extremely halophilic archaeal communities are resilient to short-term entombment in halite. *Environ. Microbiol.* 23, 3370–3383. <https://doi.org/10.1111/1462-2920.14913>.
- Hugenholtz, P., Goebel, B.M., Pace, N.R., 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180, 4765–4774. <https://doi.org/10.1128/JB.180.18.4765-4774.1998>.
- Ihmark, K., Bödeker, I.T.M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M., Clemmensen, K.E., Lindahl, B.D., 2012. New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* 82, 666–677. <https://doi.org/10.1111/j.1574-6941.2012.01437.x>.
- Janssen, P.H., 1998. Pathway of glucose catabolism by strain VeGlC2, an anaerobe belonging to the verrucomicrobiales lineage of bacterial descent. *Appl. Environ. Microbiol.* 64, 4830–4833. <https://doi.org/10.1128/AEM.64.12.4830-4833.1998>.
- Ji, Y.-X., Zhu, M.-H., Ye, X.-Q., Peng, T.-Y., Jin, R.-C., 2016. The inhibitory effect of fluoride on anaerobic ammonium oxidation (anammox). *J. Chem. Technol. Biotechnol.* 91, 640–646. <https://doi.org/10.1002/jctb.4616>.
- Johnson, L., Dean, G., Rhodes, S., Hewinson, G., Vordermeier, M., Wangoo, A., 2007. Low-dose *Mycobacterium bovis* infection in cattle results in pathology indistinguishable from that of high-dose infection. *Tuberculosis* 87, 71–76. <https://doi.org/10.1016/J.TUBE.2006.04.002>.
- Jongman, M., Korsten, L., 2016. Microbial quality and suitability of roof-harvested rainwater in rural villages for crop irrigation and domestic use. *J. Water Health* 14, 961–971. <https://doi.org/10.2166/wh.2016.058>.
- Joshi, N.A., Fass, J.N., 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. <https://github.com/najoshi/sickle>.
- Kasperbauer, S.H., De Groot, M.A., 2015. The treatment of rapidly growing mycobacterial infections. *Clin. Chest Med.* 36, 67–78. <https://doi.org/10.1016/J.CCCM.2014.10.004>.
- Kembel, S.W., Jones, E., Kline, J., Northcutt, D., Stenson, J., Womack, A.M., Bohannon, B.J., Brown, G.Z., Green, J.L., 2012. Architectural design influences the diversity and structure of the built environment microbiome. *ISME J.* 6, 1469–1479. <https://doi.org/10.1038/ismej.2011.211>.
- Kenya Bureau of Standards, 2015. *Potable Water – Specification*. KS EAS 12:2014. ICS 13.060.20.
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, e1. <https://doi.org/10.1093/nar/gks808>.
- Köjljal, U., Nilsson, R.H., Abarenkov, K., Tedersoo, L., Taylor, A.F.S., Bahram, M., Bates, S.T., Bruns, T.D., Bengtsson-Palme, J., Callaghan, T.M., Douglas, B., Drenkhan, T., Eberhardt, U., Duenas, M., Grebenc, T., Griffith, G.W., Hartmann, M., Kirk, P.M., Kohout, P., Larsson, E., Lindahl, B.D., Lücking, R., Martin, M.P., Matheny, P.B., Nguyen, N.H., Niskanen, T., Oja, J., Peay, K.G., Peintner, U., Peterson, M., Pöhlmaa, K., Saag, L., Saar, I., Schüßler, A., Scott, J.A., Senés, C., Smith, M.E., Suija, A., Taylor, D.L., Telleria, M.T.,

- Weiss, M., Larsson, K.-H., 2013. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22, 5271–5277. <https://doi.org/10.1111/mec.12481>.
- Legendre, P., Oksanen, J., ter Braak, C.J.F., 2011. Testing the significance of canonical axes in redundancy analysis. *Methods Ecol. Evol.* 2, 269–277. <https://doi.org/10.1111/j.2041-210X.2010.00078.x>.
- Lever, M.S., Williams, A., Bennett, A.M., 2000. Survival of mycobacterial species in aerosols generated from artificial saliva. *Lett. Appl. Microbiol.* 31, 238–241. <https://doi.org/10.1046/j.1365-2672.2000.00807.x>.
- Liu, Y., Whitman, W.B., 2008. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. N. Y. Acad. Sci.* 1125, 171–189. <https://doi.org/10.1196/annals.1419.019>.
- Liu, X., Li, M., Castelle, C.J., Probst, A.J., Zhou, Z., Pan, J., Liu, Y., Banfield, J.F., Gu, J.-D., 2018. Insights into the ecology, evolution, and metabolism of the widespread weesearchaeotal lineages. *Microbiome* 6, 102. <https://doi.org/10.1186/s40168-018-0488-2>.
- Maestre-Valero, J.F., Martínez-Alvarez, V., 2010. Effects of drip irrigation systems on the recovery of dissolved oxygen from hypoxic water. *Agric. Water Manag.* 97, 1806–1812. <https://doi.org/10.1016/j.agwat.2010.06.018>.
- Mahmud, R., Erguvan, M., MacPhee, D.W., 2020. Performance of closed loop venturi aspirated aeration system: experimental study and numerical analysis with discrete bubble model. *Water* 12, 1637. <https://doi.org/10.3390/w12061637>.
- Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G., Neufeld, J.D., 2012. PANDAseq: paired-end assembler for Illumina sequences. *BMC Bioinforma.* 13, 31. <https://doi.org/10.1186/1471-2105-13-31>.
- McKew, B.A., Smith, C.J., 2015. Real-time PCR approaches for analysis of hydrocarbon-degrading bacterial communities. In: McGenty, T.J., Timmis, K.N., Nogales, B. (Eds.), *Hydrocarbon And Lipid Microbiology Protocols: Microbial Quantitation, Community Profiling And Array Approaches*, Springer Protocols Handbooks. Springer, Berlin, Heidelberg, pp. 45–64.
- 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>.
- Miyazaki, M., Sakai, S., Yamanaka, Y., Saito, Y., Takai, K., Imachi, H., 2014. Spirochaeta psychrophila sp. nov., a psychrophilic spirochaete isolated from seafloor sediment, and emended description of the genus Spirochaeta. *Int. J. Syst. Evol. Microbiol.* 64, 2798–2804. <https://doi.org/10.1099/ijs.0.062463-0>.
- Naimi, B., 2017. R package 'usdm'. <https://cran.r-project.org/web/packages/usdm/usdm.pdf> Available online.
- National Environment Management Authority, 2006. *Environment Management And Coordination (Water Quality) Regulations, Legal Notice No. 121, Republic of Kenya*, Government Printers, Kenya.
- Nikolenko, S.I., Korobeynikov, A.I., Alekseyev, M.A., 2013. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* 14, S7. <https://doi.org/10.1186/1471-2164-14-S1-S7>.
- Norton, G.J., Williams, M., Falkingham, J.O., Honda, J.R., 2020. Physical measures to reduce exposure to tap water-associated nontuberculous mycobacteria. *Front. Public Health* 8, 190. <https://doi.org/10.3389/fpubh.2020.00190>.
- Ochoa-Herrera, V., Banihani, Q., León, G., Khatri, C., Field, J.A., Sierra-Alvarez, R., 2009. Toxicity of fluoride to microorganisms in biological wastewater treatment systems. *Water Res.* 43, 3177–3186. <https://doi.org/10.1016/j.watres.2009.04.032>.
- Offre, P., Spang, A., Schleper, C., 2013. Archaea in biogeochemical cycles. *Annu. Rev. Microbiol.* 67, 437–457. <https://doi.org/10.1146/annurev-micro-092412-155614>.
- Ogle, D., Wheeler, P., Dinno, A., 2020. R package 'FSA'. <https://cran.r-project.org/web/packages/FSA/index.html> Available online.
- Okamori, S., Asakura, T., Nishimura, T., Tamizu, E., Ishii, M., Yoshida, M., Fukano, H., Hayashi, Y., Fujita, M., Hoshino, Y., Betsuyaku, T., Hasegawa, N., 2018. Natural history of Mycobacterium fortuitum pulmonary infection presenting with migratory infiltrates: a case report with microbiological analysis. *BMC Infect. Dis.* 18, 1. <https://doi.org/10.1186/s12879-017-2892-9>.
- Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2020. R package 'vegan'. <https://cran.r-project.org/web/packages/vegan/vegan.pdf> Available online.
- Pachpute, J.S., Tumbo, S.D., Sally, H., Mul, M.L., 2009. Sustainability of rainwater harvesting systems in rural catchment of sub-Saharan Africa. *Water Resour. Manag.* 23, 2815–2839. <https://doi.org/10.1007/s11269-009-9411-8>.
- Park, S., Suh, G.Y., Chung, M.P., Kim, H., Kwon, O.J., Lee, K.S., Lee, N.Y., Koh, W.-J., 2008. Clinical significance of Mycobacterium fortuitum isolated from respiratory specimens. *Respir. Med.* 102, 437–442. <https://doi.org/10.1016/j.rmed.2007.10.005>.
- Props, R., Monsieurs, P., Mysara, M., Clement, L., Boon, N., 2016. Measuring the biodiversity of microbial communities by flow cytometry. *Methods Ecol. Evol.* 7, 1376–1385. <https://doi.org/10.1111/2041-210X.12607>.
- Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., Segata, N., 2017. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844. <https://doi.org/10.1038/nbt.3935>.
- Raskin, L., Stromley, J.M., Rittmann, B.E., Stahl, D.A., 1994. Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl. Environ. Microbiol.* 60, 1232–1240.
- Rockstrom, J., 2000. Water resources management in smallholder farms in Eastern and Southern Africa: an overview. *Phys. Chem. Earth B Hydrol. Ocean. Atmos.* 25, 275–283. [https://doi.org/10.1016/S1464-1999\(00\)00015-0](https://doi.org/10.1016/S1464-1999(00)00015-0).
- Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584. <https://doi.org/10.7717/peerj.2584>.
- Rosso, D., Larson, L.E., Stenstrom, M.K., 2008. Aeration of large-scale municipal wastewater treatment plants: state of the art. *Water Sci. Technol.* 57, 973–978. <https://doi.org/10.2166/wst.2008.218>.
- Saini, D., Hopkins, G.W., Seay, S.A., Chen, C.-J., Perley, C.C., Click, E.M., Frothingham, R., 2012. Ultra-low dose of Mycobacterium tuberculosis aerosol creates partial infection in mice. *Tuberculosis* 92, 160–165. <https://doi.org/10.1016/j.tube.2011.11.007>.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., Huttenhower, C., 2011. Metagenomic biomarker discovery and explanation. *Genome Biol.* 12, R60. <https://doi.org/10.1186/GB-2011-12-6-R60>.
- Sfeir, M., Walsh, M., Rosa, R., Aragon, L., Liu, S.Y., Cleary, T., Worley, M., Frederick, C., Abbo, L.M., 2018. Mycobacterium abscessus complex infections: a retrospective cohort study. *Open Forum Infect. Dis.* 5. <https://doi.org/10.1093/OFID/OFY022>.
- Shammas, N.K., 2007. Fine pore aeration of water and wastewater. *Advanced Physicochemical Treatment Technologies*. Humana Press, Totowa, NJ, pp. 391–448. [https://doi.org/10.1007/978-1-59745-173-4\\_8](https://doi.org/10.1007/978-1-59745-173-4_8).
- Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 591. <https://doi.org/10.2307/2333709>.
- Spoof, L., Jaakkola, S., Vazić, T., Häggqvist, K., Kirkkala, Terhi, Ventelä, A.-M., Kirkkala, Teija, Svirčev, Z., Meriluoto, J., 2020. Elimination of cyanobacteria and microcystins in irrigation water—effects of hydrogen peroxide treatment. *Environ. Sci. Pollut. Res. Int.* 27, 8638. <https://doi.org/10.1007/s11356-019-07476-X>.
- Spring, S., Brinkmann, N., Murrja, M., Spröer, C., Reither, J., Klenk, H.-P., 2015. High diversity of culturable prokaryotes in a lithifying hypersaline microbial mat. *Geomicrobiol. J.* 32, 332–346. <https://doi.org/10.1080/01490451.2014.913095>.
- Stahl, D.A., Amann, R., 1991. Development and application of nucleic acid probes in bacterial systematics. In: Stackebrandt, E., Goodfellow, M. (Eds.), *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley & Sons Ltd, Chichester, UK, pp. 205–248.
- Strauss, A., Reyneke, B., Waso, M., Ndlovu, T., Brink, C.J., Khan, S., Khan, W., 2019. EMA-amplicon-based taxonomic characterisation of the viable bacterial community present in untreated and SODIS treated roof-harvested rainwater. *Environ. Sci. Water Res. Technol.* 5, 91–101. <https://doi.org/10.1039/C8EW00613J>.
- Sui, H., Weil, A.A., Nuwagira, E., Qadri, F., Ryan, E.T., Mezzari, M.P., Phipatanakul, W., Lai, P.S., 2020. Impact of DNA extraction method on variation in human and built environment microbial community and functional profiles assessed by shotgun metagenomics sequencing. *Front. Microbiol.* 11, 953. <https://doi.org/10.3389/fmicb.2020.00953>.
- Tan, B., Ng, C., Nshimiyimana, J.P., Loh, L.L., Gin, K.Y.-H., Thompson, J.R., 2015. Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. *Front. Microbiol.* 6, 1027. <https://doi.org/10.3389/fmicb.2015.01027>.
- Tatti, E., McKew, B.A., Whitby, C., Smith, C.J., 2016. Simultaneous DNA-RNA extraction from coastal sediments and quantification of 16S rRNA genes and transcripts by real-time PCR. *J. Vis. Exp.* <https://doi.org/10.3791/54067>.
- Taylor, R.H., Falkingham III, J.O., Norton, C.D., LeChevallier, M.W., 2000. Chlorine, chloramine, chlorine dioxide, and ozone susceptibility of Mycobacterium avium. *Appl. Environ. Microbiol.* 66, 1702. <https://doi.org/10.1128/AEM.66.4.1702-1705.2000>.
- Therrien, J.-D., Vanrolleghem, P.A., Dorea, C.C., 2019. Characterization of the performance of venturi-based aeration devices for use in wastewater treatment in low-resource settings. *Water SA* 45, 251–258. <https://doi.org/10.4314/wsa.v45i2.12>.
- Thomas, G.E., Cameron, T.C., Campo, P., Clark, D.R., Coulon, F., Gregson, B.H., Hepburn, L.J., McGenity, T.J., Miliou, A., Whitby, C., McKew, B.A., 2020. Bacterial community legacy effects following the Agia Zoni II Oil-spill, Greece. *Front. Microbiol.* 11, 1706. <https://doi.org/10.3389/fmicb.2020.01706>.
- Thomas, G.E., Brant, J.L., Campo, P., Clark, D.R., Coulon, F., Gregson, B.H., McGenity, T.J., McKew, B.A., 2021. Effects of dispersants and biosurfactants on crude-oil biodegradation and bacterial community succession. *Microorganisms* 9, 1200. <https://doi.org/10.3390/microorganisms9061200>.
- Thomson, R., Tolson, C., Sidjabat, H., Huygens, F., Hargreaves, M., 2013. Mycobacterium abscessus isolated from municipal water – a potential source of human infection. *BMC Infect. Dis.* 13, 241. <https://doi.org/10.1186/1471-2334-13-241>.
- Tukey, J.W., 1953. Some selected quick and easy methods of statistical analysis. *Trans. N. Y. Acad. Sci.* 16. <https://doi.org/10.1111/J.2164-0947.1953.TB01326.X>.
- van Passel, M.W.J., Kant, R., Palva, A., Copeland, A., Lucas, S., Lapidus, A., del Rio, T.G., Pitluck, S., Goltzman, E., Clum, A., Sun, H., Schmutz, J., Larimer, F.W., Land, M.L., Hauser, L., Kyrpides, N., Mikhailova, N., Richardson, P.P., Janssen, P.H., de Vos, W.M., Smidt, H., 2011. Genome sequence of the verrucomicrobium oputtus terrae PB90-1, an abundant inhabitant of rice paddy soil ecosystems. *J. Bacteriol.* 193, 2367–2368. <https://doi.org/10.1128/JB.00228-11>.
- Vavourakis, C.D., Andrei, A.-S., Mehrshad, M., Ghai, R., Sorokin, D.Y., Muyzer, G., 2018. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome* 6, 168. <https://doi.org/10.1186/s40168-018-0548-7>.
- Vidal-Durà, A., Burke, I., Mortimer, R., Stewart, D., 2018. Diversity patterns of benthic bacterial communities along the salinity continuum of the Humber estuary (UK). *Aquat. Microb. Ecol.* 81, 277–291. <https://doi.org/10.3354/ame01875>.
- Vierheilig, J., Savio, D., Ley, R.E., Mach, R.L., Farnleitner, A.H., Reischer, G.H., 2015. Potential applications of next generation DNA sequencing of 16S rRNA gene amplicons in microbial water quality monitoring. *Water Sci. Technol.* 72, 1962–1972. <https://doi.org/10.2166/wst.2015.407>.
- Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. <https://doi.org/10.1128/AEM.00062-07>.
- Wang, Y., Yan, Y., Thompson, K.N., Bae, S., Accorsi, E.K., Zhang, Y., Shen, J., Vlamakis, H., Hartmann, E.M., Huttenhower, C., 2021. Whole microbial community viability is not quantitatively reflected by propidium monoazide sequencing approach. *Microbiome* 9, 17. <https://doi.org/10.1186/s40168-020-00961-3>.

- Wickham, H., 2016. Ggplot2: elegant graphics for data analysis. *J. Stat. Softw.* 35.
- Wilke, C., 2020. R package 'cowplot'. <https://cran.r-project.org/web/packages/cowplot/index.html> Available online.
- World Health Organisation (WHO), 2017. *Guidelines for Drinking-water Quality, 4th edition with 1st addendum*, Geneva.
- Wu, J., Long, S.C., Das, D., Dorner, S.M., 2011. Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. *J. Water Health* 9, 265–278. <https://doi.org/10.2166/wh.2011.117>.
- Wu, R.-S., Molina, G.L.L., Hussain, F., 2018. Optimal sites identification for rainwater harvesting in northeastern Guatemala by analytical hierarchy process. *Water Resour. Manag.* 32, 4139–4153. <https://doi.org/10.1007/s11269-018-2050-1>.
- Zabidi, H.A., Goh, H.W., Chang, C.K., Chan, N.W., Zakaria, N.A., 2020. A review of roof and pond rainwater harvesting systems for water security: the design, performance and way forward. *Water* 12, 3163. <https://doi.org/10.3390/w12113163>.
- Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620. <https://doi.org/10.1093/bioinformatics/btt593>.
- Zhang, B., Zhang, Yongqing, Li, X., Zhang, Yuanming, 2018. Successional changes of fungal communities along the biocrust development stages. *Biol. Fertil. Soils* 54, 285–294. <https://doi.org/10.1007/s00374-017-1259-0>.
- Zhang, X., Xia, S., Zhao, R., Wang, H., 2020. Effect of temperature on opportunistic pathogen gene markers and microbial communities in long-term stored roof-harvested rainwater. *Environ. Res.* 181, 108917. <https://doi.org/10.1016/J.ENVRES.2019.108917>.