



## OPEN ACCESS

EDITED BY  
Abdelazim Negm,  
Zagazig University,  
Egypt

REVIEWED BY  
Wensheng Lan,  
Independent Researcher, Shenzhen,  
China  
Kihyuck Choi,  
Dong-A University, Republic of Korea

\*CORRESPONDENCE  
Mohammad Arif  
✉ arif@hawaii.edu

SPECIALTY SECTION  
This article was submitted to  
Systems Microbiology,  
a section of the journal  
Frontiers in Microbiology

RECEIVED 13 September 2022  
ACCEPTED 16 January 2023  
PUBLISHED 17 February 2023

CITATION  
Klair D, Dobhal S, Ahmad A, Hassan ZU,  
Uyeda J, Silva J, Wang K-H, Kim S,  
Alvarez AM and Arif M (2023) Exploring  
taxonomic and functional microbiome of  
Hawaiian stream and spring irrigation water  
systems using Illumina and Oxford Nanopore  
sequencing platforms.  
*Front. Microbiol.* 14:1039292.  
doi: 10.3389/fmicb.2023.1039292

COPYRIGHT  
© 2023 Klair, Dobhal, Ahmad, Hassan, Uyeda,  
Silva, Wang, Kim, Alvarez and Arif. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted which  
does not comply with these terms.

# Exploring taxonomic and functional microbiome of Hawaiian stream and spring irrigation water systems using Illumina and Oxford Nanopore sequencing platforms

Diksha Klair<sup>1</sup>, Shefali Dobhal<sup>1</sup>, Amjad Ahmad<sup>2</sup>, Zohaib Ul Hassan<sup>3,4,5</sup>,  
Jensen Uyeda<sup>2</sup>, Joshua Silva<sup>2</sup>, Koon-Hui Wang<sup>1</sup>, Seil Kim<sup>3,4,5</sup>,  
Anne M. Alvarez<sup>1</sup> and Mohammad Arif<sup>1\*</sup>

<sup>1</sup>Department of Plant and Environmental Protection Sciences, University of Hawaii at Manoa, Honolulu, HI, United States, <sup>2</sup>Department of Tropical Plant and Soil Sciences, University of Hawaii at Manoa, Honolulu, HI, United States, <sup>3</sup>Group for Biometrology, Korea Research Institute of Standards and Science (KRISS), Daejeon, Republic of Korea, <sup>4</sup>Convergent Research Center for Emerging Virus Infection, Korea Research Institute of Chemical Technology (KRICT), Daejeon, Republic of Korea, <sup>5</sup>Department of Bio-Medical Measurement, University of Science and Technology (UST), Daejeon, Republic of Korea

Irrigation water is a common source of contamination that carries plant and foodborne human pathogens and provides a niche for proliferation and survival of microbes in agricultural settings. Bacterial communities and their functions in irrigation water were investigated by analyzing samples from wetland taro farms on Oahu, Hawaii using different DNA sequencing platforms. Irrigation water samples (stream, spring, and storage tank water) were collected from North, East, and West sides of Oahu and subjected to high quality DNA isolation, library preparation and sequencing of the V3–V4 region, full length 16S rRNA, and shotgun metagenome sequencing using Illumina iSeq100, Oxford Nanopore MinION and Illumina NovaSeq, respectively. Illumina reads provided the most comprehensive taxonomic classification at the phylum level where Proteobacteria was identified as the most abundant phylum in the stream source and associated water samples from wetland taro fields. Cyanobacteria was also a dominant phylum in samples from tank and spring water, whereas Bacteroidetes were most abundant in wetland taro fields irrigated with spring water. However, over 50% of the valid short amplicon reads remained unclassified and inconclusive at the species level. In contrast, Oxford Nanopore MinION was a better choice for microbe classification at the genus and species levels as indicated by samples sequenced for full length 16S rRNA. No reliable taxonomic classification results were obtained while using shotgun metagenome data. In functional analyzes, only 12% of the genes were shared by two consortia and 95 antibiotic resistant genes (ARGs) were detected with variable relative abundance. Full descriptions of microbial communities and their functions are essential for the development of better water management strategies aimed to produce safer fresh produce and to protect plant, animal, human and environmental health. Quantitative comparisons illustrated the importance of selecting the appropriate analytical method depending on the level of taxonomic delineation sought in each microbiome.

## KEYWORDS

irrigation water, microbiome, taxonomic diversity, Illumina, metagenomics, shotgun analyses

## Introduction

Irrigation water quality is a growing concern for agriculture as drainage is contaminated with agricultural runoff, wastewater overflows, and polluted storm or rainwater runoff, and irrigation waters are a potential source of plant and food-borne pathogens resulting in economic crop losses and human health risks (Hintz et al., 2010; Uyttendaele et al., 2015; Redekar et al., 2019). The microbial populations sharing the same niche may be commensal, symbiotic, or pathogenic. Many pathogenic bacteria can survive and proliferate in contaminated water and agricultural settings for long duration under favorable biotic and abiotic conditions (Ravva et al., 2006; Van der Linden et al., 2013; Cevallos-Cevallos et al., 2014). Studies have revealed that contaminated water splash can be a potential carrier of plant and food-borne pathogens (Paul et al., 2004; Cevallos-Cevallos et al., 2012) that can enter plants through stomata, hydathodes and wounds (Gu G. et al., 2013). Also, antibiotics introduced through contaminated water are a continuing challenge as they may result in high selection pressure for antibiotic-resistant bacteria (Szczepanowski et al., 2009; Zhang and Li, 2011; Luczkiewicz et al., 2015) and can persist even after water treatment.

Because of water scarcity and a simultaneous need to increase food production, there has been a shift from freshwater to alternative sources of irrigation water such as reclaimed or recycled water. However, potential health and environmental impact concerns are associated with the use of alternative water sources for irrigating the crops (Qin et al., 2015). Therefore, uncovering the bacterial composition and its associated functions in irrigation water will provide insight into formulating new disease management strategies and preventing major economic and public health risks. High-throughput sequencing has facilitated the identification of complex bacterial communities (Diaz et al., 2012) independently of bacterial culture (Tringe and Hugenholtz, 2008; Rinke et al., 2014). The bacterial microbiota is identified by analyzing the prokaryotic 16S ribosomal RNA (rRNA; ~1,500 bp long) with nine variable regions interspaced between conserved regions. The 16S rRNA region selected for sequencing depends on the experimental objectives, design, and sample type. Sequencing of variable regions of the 16S rRNA gene using the most popular sequencing platforms, such as Illumina technology, uncovers the majority of bacterial microbiota (Sanz-Martin et al., 2017). Illumina technology only permits sequencing of short variable regions of the 16S rRNA gene (Goodwin et al., 2016), and therefore, taxonomic assignment of reads at the species level may be elusive. Different species within a genus possess different phenotypic and virulence characteristics, therefore, accurate speciation of bacterial species is of utmost importance for formulating effective disease management strategies against pathogenic bacterial communities.

With the advancement in next generation sequencing technologies (NGS), 3rd generation NGS technology, Oxford Nanopore enables generation of long sequence read lengths, possibly sequencing full length 16S rRNA genes (Matsuo et al., 2021). Full length sequences covering maximum nucleotide heterogeneity and discriminatory power allow better identification at the genus and species level. Comparative studies for Oxford Nanopore and Illumina 16S rRNA gene sequencing demonstrated similar bacterial composition at the genus level, although significant differences were observed at the species level (Heikema et al., 2020). However, this technology complicates accurate species classification, particularly for bacterial species with a high sequence similarity in the 16S rRNA gene, owing to higher sequencing error rates (Laver et al., 2015).

Although Polymorphic marker gene (e.g., 16S rRNA, ITS) based analyzes are useful for broad community taxonomical analysis, it did not

provide functionality nor resolve the complexity of a microbiome. The shotgun metagenomic sequencing using advanced Illumina sequencing platforms have been proven to be a more reliable approach for these purposes (Peng et al., 2021). Metagenomic sequencing is a powerful tool for investigating occurrence, abundance, and distribution of ARGs in the natural environment and is suitable for discovery of novel ARGs that remain unidentified in culture- and amplicon-based analyzes (Schmieder and Edwards, 2012; Xu et al., 2015).

This study aimed to investigate bacterial microbiota and associated gene function of different irrigation systems, mainly associated with wetland taro across the island of Oahu, Hawaii. Mountain streams are the major source of irrigation waters used by farmers to irrigate crops. The overall goal of this project is to reveal the bacterial microbiota from different water source used for irrigation, in addition to field water, which is released back into the stream after use, carrying excess fertilizer, agricultural waste, ARGs and diverse unidentified bacteria. Bacterial communities were investigated based on 16S rRNA amplicon analysis using two principally different sequencing technologies and platforms— Illumina iSeq100 and Oxford Nanopore MinION and their taxonomic compositions were compared. The functionality of all the genes in complex samples and the distribution of ARGs were also investigated using shotgun metagenomic analyses. We aim to compare different technologies and approaches considered for microbiome studies such as shotgun metagenome, short- and long-amplicon read based to provide the desired level of accuracy in resolving the microbial taxonomic composition of the samples.

## Materials and methods

### Sample collection

Irrigation source and associated taro field water samples were collected in September–November 2020, across the Island of Oahu, Hawaii (Supplementary Table S1). Irrigation water samples—R-S1-E, R-S2-W, R-S4-SE, and R-S5-SE—collected from natural streams which were sources of irrigation water for taro fields. Two water samples R-S7-N (stream emerging from the main reservoir on Oahu) and T-S6-N (tank storage water) were sources of irrigation for horticultural crops and other agricultural practices. Taro field water samples, R-F1-E, R-F2-W, R-F4-SE, and R-F5-SE, associated with R-S1-E, R-S2-W, R-S4-SE, and R-S5-SE, respectively, were collected to analyze bacterial microbiota. Two water samples, S-S3-N and S-F3-N were collected from a spring water source and an associated taro field, respectively. From each sampling site, three replicate water samples (21 per sample; DNA from these three replicates were merged for library preparation) were collected in sterile glass bottles, submerged 10–15 cm below the water surface. Samples were transported in an ice-cooler and processed in the laboratory for DNA isolation.

### Sample processing

Water samples collected from each site were vacuum filtered using the Millipore All-Glass Filter Holder kit (EMD Millipore Corporation, Billerica, MA, United States). Collected water from each replicate was filtered through Whatman filter membrane to remove coarse to medium debris, followed by filtration through a MF-Millipore 8 µm sterile mixed cellulose ester (MCE) membrane (Merck Millipore Ltd., Tullagreen

Carrigtwohill, Co. Cork, Ireland), and finally, filtered *via* MF-Millipore 0.22  $\mu\text{m}$  sterile MCE membrane to trap the maximum bacterial community. The 0.22  $\mu\text{m}$  membrane was used for bacterial DNA isolation using NucleoMag DNA/RNA Water Kit (MACHEREY-NAGEL Inc., Bethlehem, PA, United States) following manufacturer's instructions, with a few minor modifications to improve the DNA quantity and quality. The mechanical lysis was performed in lysis buffer MWA1 for 20 min using a vortex at full speed, followed by the addition of 25  $\mu\text{l}$  of RNase (12 mg/ml stock solution); the tubes were incubated for 15 min at room temperature (RT). A lysate of 450  $\mu\text{l}$  was transferred to a 1.5 ml sterile Eppendorf tube and 25  $\mu\text{l}$  of NucleoMag B-beads were added, mixed and shaken for 5 min, and kept on a magnetic rack at RT. The supernatant was removed, and the pellet was washed twice with buffer MWA3, followed by a single final wash with buffer MWA4. The magnetic beads were air dried for 15 min at RT; 70  $\mu\text{l}$  RNase free water was used to elute DNA from the magnetic beads. Qubit dsDNA HS kit and Qubit 4 (Thermo Fisher Scientific, Waltham, MA, United States) were used to quantify the genomic DNA. The DNA replicates from each sample were pooled, due to low DNA yield from each replicate, for downstream processes and stored at  $-80^{\circ}\text{C}$ .

## Illumina 16S rRNA library preparation, sequencing, and analysis

The polymerase chain reaction (PCR) was performed to amplify the V3–V4 hypervariable region of 16S rRNA gene following the reaction conditions:  $94^{\circ}\text{C}$  for 5 min; 40 cycles at  $94^{\circ}\text{C}$  for 20 s,  $58^{\circ}\text{C}$  for 30 s, and  $72^{\circ}\text{C}$  for 1 min; and the final extension at  $72^{\circ}\text{C}$  for 3 min. Primers 341F (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGG GNGGCWGCAG-3') and 805R (5'-GTCTCGTGGGCTCGGAGATGTG TATAAGAGACAGGACTACHVGGGTATCTAATCC-3') were used for PCR amplification (Herlemann et al., 2011). The amplified PCR amplicons were enzymatically cleaned using ExoSAP-IT (Affymetrix, Santa Clara, CA, United States) and quantified using Qubit dsDNA HS Kit and Qubit 4. A secondary bead-linked transposome (BLT) PCR was performed using i5 and i7 adapters, provided in Nextera DNA Flex Library Prep Kit (Illumina, Inc., San Diego, CA, United States), for barcode attachment (Supplementary Table S2). Each sample's library was prepared in duplicate. The BLT PCR conditions were initial denaturation at  $98^{\circ}\text{C}$  for 3 min, followed by X cycles of  $98^{\circ}\text{C}$  for 45 s,  $62^{\circ}\text{C}$  for 30 s, and  $68^{\circ}\text{C}$  for 2 min, with a final extension at  $68^{\circ}\text{C}$  for 1 min. The number of cycles of BLT PCR's (X) was decided based on the amplicon concentration from the previous PCR as recommended by the manufacturer. Samples with concentrations ranging from 1 to 9 ng/ $\mu\text{l}$  and 9–21 ng/ $\mu\text{l}$  were subjected to 8 and 12 cycles BLT PCR, respectively. The amplicon libraries were cleaned using double-sided bead purification protocol following the manufacturer's instructions. The purified libraries were quantified, normalized to 1 nM concentration and pooled. The pooled library was spiked with 2% using Phix control and loaded to Illumina iSeq100 for sequencing with a total of 302 run cycles to generate paired-end 150-bp reads. The total data yield was 717 MB with Q30 value of 88.1 and 89.6% for Read 1 and Read 2, respectively. The obtained raw sequences were submitted in the National Center for Biotechnology Information (NCBI) as Sequence Read Archive (SRA) under the BioProject ID PRJNA856319. The sequenced data was base called and analyzed using BaseSpace sequence hub and EzBioCloud, respectively (Yoon et al., 2017). The paired-end reads were used as a quality control to filter out low-quality (average quality value < 25) and

merged using PandaSeq (Masella et al., 2012); primers were trimmed at a similarity cut-off of 0.8. The pipeline uses EzBioCloud database for taxonomic assignment and sequence similarity was calculated *via* pairwise alignment. The chimeric reads with less than a 97% best hit similarity rate were removed using EzBioCloud non-chimeric 16S rRNA database through UCHIME (Edgar et al., 2011). The sequenced data was clustered using CD-Hit7 and UCLUST with 97% similarity (Li and Godzik, 2006). Bacterial diversity was also analyzed and compared among the samples. For alpha diversity—OTUs, richness, and diversity were calculated, while for beta diversity—principal coordinate analysis (PCoA) and UPGMA clustering analyzes were performed. Permutational multivariate analysis of variance was performed utilizing 'beta set-significance analysis' in EzBioCloud, to test the significant difference between clusters using distance measures of the beta diversity.

Valid reads were normalized for each sample to eliminate the bias produced because of variation in total number of reads. The Wilcoxon rank-sum test was used to calculate differences between the replicates. The differences in relative abundance in phyla and genera among the samples were determined using one-way ANOVA (single factor) with the least significant difference (LSD) test at  $\alpha = 0.05$ .

## Oxford Nanopore 16S rRNA library preparation, sequencing, and analysis

The genomic DNA of sample R-F1-E and S-F3-N was diluted to 1 ng/ $\mu\text{l}$ , and a total 10  $\mu\text{l}$  gDNA was used for full-length 16S rRNA library preparation using 16S Barcoding Kit 1–24 (SQK-16S024; Oxford Nanopore Technologies, Oxford Science Park, United Kingdom) according to the manufacturer's protocol. Ten  $\mu\text{l}$  of input DNA (10 ng) was mixed well with 25  $\mu\text{l}$  LongAmp hot Start Taq 2  $\times$  Master Mix and 5  $\mu\text{l}$  of nuclease free water, afterward, 10  $\mu\text{l}$  of each 16S barcode was added. The PCR was performed using following conditions: Initial denaturation at  $95^{\circ}\text{C}$  for 1 min, 25 cycles of  $95^{\circ}\text{C}$  for 20 s,  $55^{\circ}\text{C}$  for 30 s and  $65^{\circ}\text{C}$  for 2 min, with a final extension at  $65^{\circ}\text{C}$  for 5 min. Each amplified sample was purified and washed with AMPure XP beads and 70% ethanol, respectively. For each sample, barcoded libraries were prepared in duplicate and quantified using Qubit 4; libraries were pooled to a desired ratio of 50–100 fmol in 10  $\mu\text{l}$  of 10 mM Tris-HCl (pH 8.0) with 50 mM NaCl, and 1  $\mu\text{l}$  of Rapid adapter (RAP) was added. The pooled library was loaded on to MinION vR9.4 flow cell and sequenced following manufacturer's instruction. The generated sequencing data were monitored in real-time using the MinKNOW software (version 4.0.20). The obtained FAST5 files were base called using MinKNOW (version 4.0.20) embedded with Guppy version 3.2.10 pipeline. The attained raw long reads were deposited as SRA data in the NCBI database under BioProject ID PRJNA856390. The generated full-length 16S rRNA sequence data were analyzed using cloud based EPI2ME (Oxford Nanopore) workflow for the identification of microbial community composition; EP2ME uses the NCBI GenBank database for taxonomic identification. The minimum and maximum read length of 1,500 and 1,600, respectively, were assigned as a quality control parameter, and Blastn was run using parameters max\_target seqs = 3 (finds the top three hits that are statistically significant) with blast e-value assigned as default 0.01. Per read coverage was calculated as the number of identical matches/query length. All classified reads were filtered for > 77% accuracy and > 30% coverage, which removed invalid alignments and were normalized for analysis. Results were obtained as comma-separated values (CSV) file *via* web report generated by EPI2ME workflow.

## Metagenomic library preparation, sequencing, and analysis

DNA from two samples (R-F1-E and S-F3-N) were used for preparing DNA metagenome libraries using NEBNext Ultra DNA Library Prep Kit (NEB, Ipswich, MA) following manufacturer's instructions. The sonication-based method was used for fragmenting gDNA to the size of 350 bp. The obtained DNA fragments were end-polished, A-tailed, and ligated with full-length indexing adapters to the ends of the DNA fragments, followed by PCR amplification. The PCR products were purified using AMPure XP, and libraries were analyzed for size distribution and quantified using Agilent 2,100 Bioanalyzer (Agilent, Santa Clara, CA, United States) and real-time qPCR, respectively. The quantified libraries were pooled and sequenced on an Illumina NovaSeq 6,000 platform to generate paired end reads. All the procured reads were submitted in the NCBI database as SRA file under BioProject PRJNA856335. The obtained raw reads were pre-processed to trim low-quality bases with quality value ( $Q$ -value  $\leq 38$ ), reads with  $N$  nucleotides over 10 bp, and reads that overlapped with adapters over 15 bp. The obtained clean reads after quality control were assembled into scaffolds using MEGAHIT (Li et al., 2015). The quality of the assembled data was predicted by N50 length. Scaffolds ( $\geq 500$  bp) were used for ORF (Open reading Frame) prediction using MetaGeneMark (Zhu et al., 2010) and the ORF's less than 100 nt were removed. Non-redundant gene catalog, generated using CD-HIT (Fu et al., 2012), was further used to map clean reads using SoapAligner (Gu S. et al., 2013). Each metagenomic homolog was taxonomically annotated against NR database (Buchfink et al., 2015) for classification of microbial community at different taxonomic levels. For functional analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG), evolutionary genealogy of genes: Non-Supervised Orthologous Groups (eggNOG), and Carbohydrate-Active enzymes (CAZy) databases were used for mapping functionally annotated unigenes. For Antibiotic Resistance Genes (ARGs) analysis, all the unique genes were BLASTp against the CARD (Comprehensive Antibiotic Research Database) database ( $e$ -value  $\leq 1e^{-5}$ ). To identify the biologically relevant differences between two samples, statistical analyses were performed using STAMP v 2.1.3 (Parks et al., 2014), employing Fisher's exact test with Newcombe-Wilson CI method (0.95 confidence interval) and Benjamini-Hochberg FDR correction factors and visualized using extended error bar plots.

## Data availability

The obtained raw sequences were submitted in the National Center for Biotechnology Information (NCBI) as Sequence Read Archive (SRA) under the BioProject IDs PRJNA856319, PRJNA856335 and PRJNA856390.

## Results

### Short length amplicon-based analysis— Illumina

The paired end 16S rRNA encoding gene sequences were obtained using Illumina iSeq100. After the data was pre-filtered and passed the quality check to remove low-quality, non-chimeric and non-target amplicons, the total number of valid reads with an average read length was computed (Supplementary Table S3) for each sample. Each sample was successfully sequenced in duplicate, except sample S-S3-N that encountered sequencing biasness in the 2nd replicate run and failed to produce enough valid reads. After quality control, an average of 43,599

and 41,163 valid reads from the first and second replicate run, respectively, were obtained. In both the replicates, the highest and lowest number of valid reads were observed in sample R-S2-W (61,272 and 67,325) and R-F4-SE (22,274 and 26,908), respectively.

Based on phylum comparison performed using valid reads obtained from two sequencing replicates, no differences were observed, therefore the first replicate (barcode1-12) was considered for further taxonomic and diversity analysis (Supplementary Figure S1). The valid reads generated from each sample were normalized to the least number of obtained valid reads (22,274; R-F4-SE) to overcome biasness in analysis outcomes. The reads were further clustered into operational taxonomic units (OTUs) at 97% identity ranging from 1,410 to 4,897. The OTU number remained higher in river stream sources, R-S1-E (3,416), R-S2-W (4,059), R-S4-SE (2,817), and R-S5-SE (4,897), compared with associated field water, R-F1-E (1,570), R-F2-W (2,753), R-F4-SE (1,978), and R-F5-SE (2,946). However, in spring source and field water samples, the OTU count remained comparable (Table 1). Furthermore, sample T-S6-N had the lowest count of 1,077 identified OTUs, followed by sample R-S7-N with 1,410 OTU numbers.

### Taxonomic classification at phylum, genus, and species levels

Based on Good's coverage index, the sequencing covered more than 94% of the taxonomic richness except for sample R-S1-E (92.35%), R-S2-W (91.02%) and R-S5-SE (89.97%; Table 1). A total of 18 phyla with relative abundance of  $>1\%$  were compared after being identified in at least one sample (Figure 1A). Proteobacteria, a phylum with major plant and food-borne pathogens, was significantly the most abundant phylum in 12 different samples (Supplementary Table S4). The relative abundance of Proteobacteria was higher in river stream source samples, R-S1-E (76.99%), R-S2-W (71.28%), R-S4-SE (83.71%), and R-S5-SE (52.04%), and associated field samples, R-F1-E (66.57%), R-F2-W (78.64%), R-F4-SE (89.08%), and R-F5-SE (75.70%). Considering samples collected from North Oahu, Cyanobacteria was the topmost abundant phylum identified from the spring water sample S-S3-N (35.86%) and stored tank water sample T-S6-N (58.39%). Bacteroidetes was the most dominant phylum in spring water irrigated field with relative abundance of 48.82% and interestingly this phylum was also higher in the stream water irrigated field sample, R-F1-E (31.63%), whereas it remained  $<6.9\%$  of relative abundance in other river stream source and associated field water samples. Phylum Actinobacteria was relatively higher in the reservoir stream source, R-S7-N (26.82%) compared with other samples. Other identified phyla varied in their relative abundance among all the samples, as shown in Figure 1A.

The normalized valid reads from all the 12 samples were classified and compared at the genus level (Figure 1B). The taxonomic classifier used to classify valid reads identified uncultured genera and best hit genera classified with high and low confidence values, while the rest remained unclassified at a taxonomic level (genus–species). The genera within the family Comamonadaceae were classified as significantly most abundant among all the other identified genera and named as Comamonadaceae\_uc by the taxonomic classifier (Supplementary Table S5). The taxonomic classifier could not differentiate the genera within the family Comamonadaceae owing to low confidence value in assigning the best hit to the reference database—indicating that short amplicon reads might not be efficient in classifying valid reads with high accuracy. The abundance of Comamonadaceae\_uc was relatively higher in natural stream sources and associated field samples. *Prochlorococcus* was the most abundant genus identified in samples T-S6-N (58.3%) and S-S3-N (35.65%)

TABLE 1 List of total number of OTUs and calculated diversity analysis.

Sample	OTUs	ACE	CHAO	Jackknife	Shannon	Simpson	Phylogenetic Diversity	Good's Coverage of Library (%)
R-F1-E	1,570	2,712.05	2,458.62	3,062.1	3.97	0.12	1,635	96.57
R-S1-E	3,416	5,924.43	5,273.62	6,074.57	5.35	0.08	4,165	92.35
R-S2-W	4,059	6,950.62	6,139.12	7,115.85	6.21	0.03	4,241	91.02
R-F2-W	2,753	3,513.98	3,231.47	3,649	5.17	0.08	3,230	95.98
S-S3-N	2,153	3,149.55	2,905.06	3,240.36	4.34	0.14	2,378	96.14
S-F3-N	2,157	3,526.22	3,221.37	3,746.8	4.35	0.1	1,435	95.49
R-S4-SE	2,817	4,123.62	3,730.12	3,994.1	5.28	0.04	3,516	94.86
R-F4-SE	1,978	2,408.57	2,214.96	2,521	4.42	0.11	2,473	97.56
R-S5-SE	4,897	7,684.84	6,782.04	7,219.57	6.86	0.01	4,739	89.97
R-F5-SE	2,946	4,257.84	3,804.92	4,166.01	4.91	0.09	3,116	94.56
T-S6-N	1,077	1,656.5	1,522.44	1,763.94	2.88	0.34	1,106	97.93
R-S7-N	1,410	1,818.37	1,697.27	1,859.62	4.97	0.02	1,359	97.99

collected from North Oahu. Spring field water sample S-F3-N was dominated by the genus *Flavobacterium* with relative abundance of 37.14%, while 16.99% *Flavobacterium* abundance was calculated in sample R-F1-E—the abundance remained <1% in all the other river stream and associated field water samples. The classified reads at the genus level, with a relative abundance of <1%, ranged between 22.32–61.87% among all samples, indicating diverse microbiota associated with different samples. The percentage of valid reads that remained unclassified varied between 4.83% (T-S6-N) and 21.55% (R-S4-SE) among all the samples (Supplementary Table S6).

At the species level, valid reads that remained unclassified among all the 12 samples ranged from 11.2% (T-S6-N) to 62.23% (R-F4-SE; Supplementary Table S6). A total of 34 species classified at species level using EzBioCloud with relative abundance of more than 1%, only three species, *Flavobacterium fontis*, *F. hydatis* and *F. shanxiense*, remained classified with a high confidence value—indicating that the short length reads-based approach for classifying at species level is an inadequate approach for attaining species level resolution (Supplementary Figure S2).

## Alpha and Beta diversity analyzes

Non-parametric analysis of diversity indices, such as ACE, CHAO, and Jackknife, indicated higher bacterial diversity in river stream compared to associated field water samples, followed by sample S-F3-N, S-S3-N, R-S7-N, and T-S6-N (Table 1). The higher Shannon diversity indices of river stream source field water indicated an increased abundance and bacterial community than associated field water; however, a negligible difference between spring source S-S3-N (4.34) and field water S-F3-N (4.35) was observed (Table 1). The Shannon diversity calculated for sample T-S6-N and R-S7-N was 2.88 and 4.97, respectively. Taken together, natural stream source water contaminated with fertilizer runoff, wastewater runoff and other agricultural waste showed higher diversity in the bacterial community.

To compare the relationship between bacterial communities in all the samples at the genus level, Principal Coordinate Analysis (PCoA) and unweighted pair group method with arithmetic mean (UPGMA) clustering based on the Bray–Curtis dissimilarity index were performed. The beta diversity indices, based on PCoA, revealed clear distinctions between different water samples forming three distinctive clusters (Figure 1C). Cluster one was formed exclusively by natural stream sources and

associated with wet taro field water samples irrespective of the sampling site except for sample R-F1-E. The second distinctive cluster was formed by water samples collected from North Oahu, S-S3-N, T-S6-N, and R-S7-N, except S-F3-N. Interestingly, the 3rd cluster was formed by field water samples R-F1-E and S-F3-N indicating a close association between their bacterial communities, despite having been surveyed from different geographical locations and irrigated by different water sources (spring and river sources). PERMANOVA analysis revealed the significant difference between cluster 1 and 2 (value of  $p=0.009$ ) and cluster 1 and 3 (value of  $p=0.021$ ), however, no significant difference observed on pairwise comparison of cluster 2 and 3 (value of  $p=0.101$ ; Supplementary Table S7).

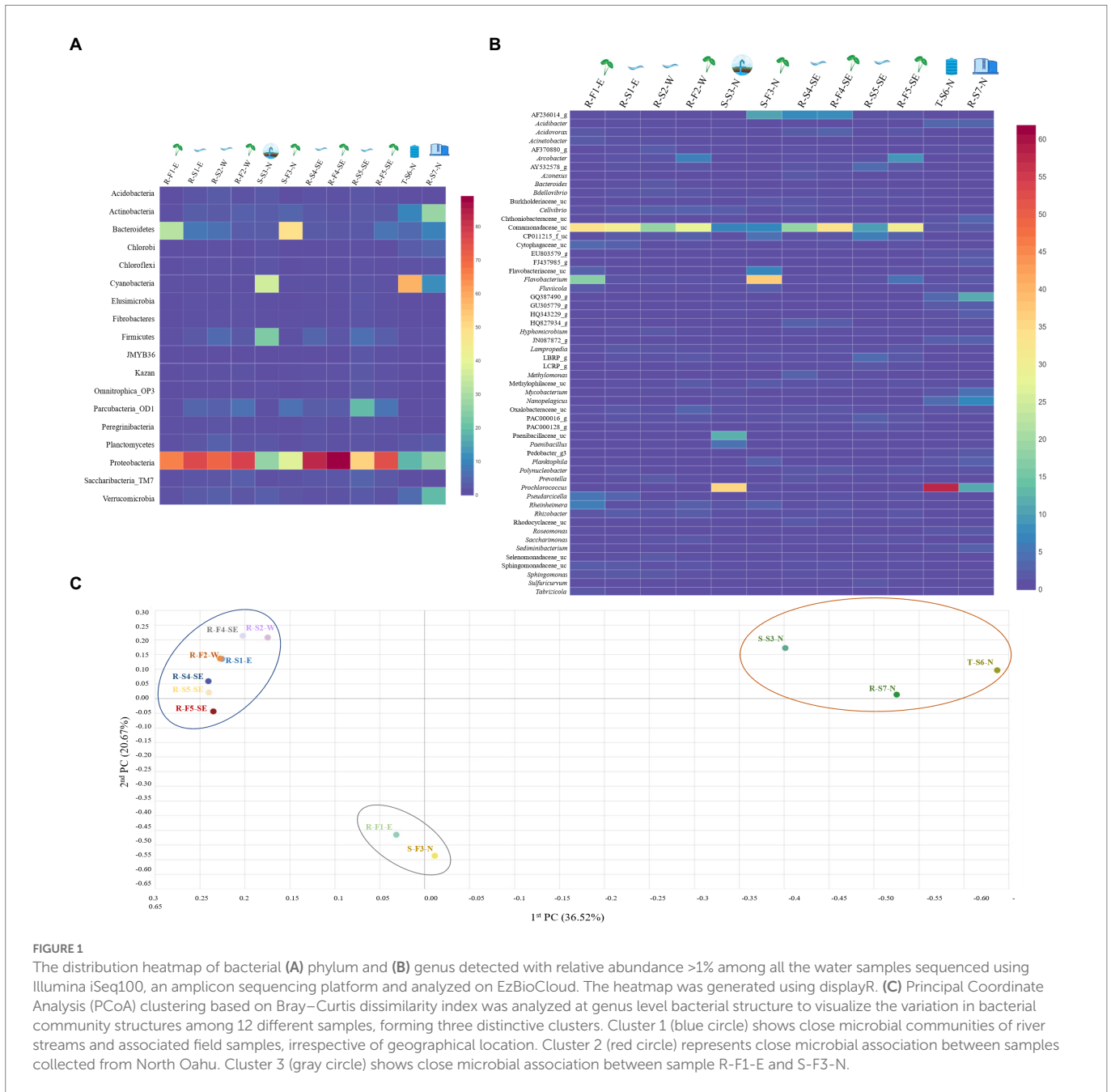
Furthermore, UPGMA clustering revealed a similar clustering pattern in the dissimilarity of relative abundance of the bacterial communities (Supplementary Figure S3). To unravel the close microbial association between R-F1-E and S-F3-N, these two samples were further sequenced to obtain full length 16S RNA and metagenomes using Oxford Nanopore MinION and Illumina NovaSeq, respectively, for amplicon and functional analyzes.

## Full length 16S RNA amplicon analysis—Oxford Nanopore MinION

Samples R-F1-E and S-F3-N were sequenced in duplicate to attain confidence and reliability in the obtained data (Supplementary Table S8). Replicate 1 of sample S-F3-N failed to sequence and no reads were generated; nevertheless, the other replicate generated 87,818 reads with ~1,500 bp length. In contrast, sample R-F1-E sequenced in two repeats validly sequenced 1,27,647 and 5,57,290 reads ranging from 1,500 to 1,600 bp length, and the comparative analyzes between replicates at the genus and species levels were comparable, comprising almost similar bacterial composition (Supplementary Figure S4). Therefore, for further comparative analysis, reads from one sequencing replicate of sample R-F1-E were used.

## Taxonomic classification at phylum, genus, and species levels

At the phylum level, sample R-F1-E showed Bacteroidetes and Proteobacteria with relative abundance of >1%, while sample S-F1-E was dominated with 3 phyla-Bacteroidetes, Proteobacteria and Verrucomicrobia



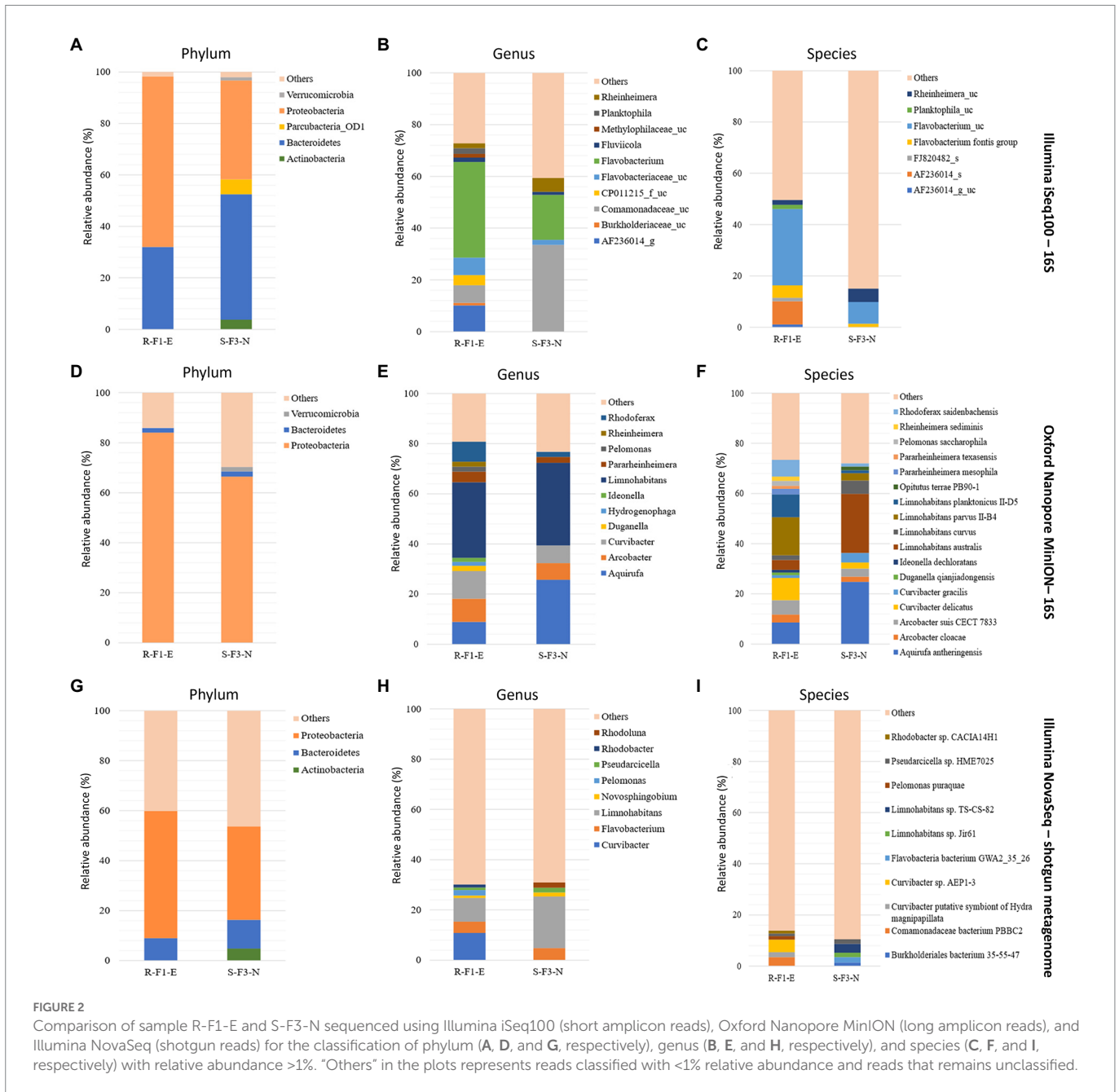
(Figure 2). Classification at the genus level uncovered a total of 11 and 6 genera from samples R-F1-E and S-F3-N, respectively, with relative abundance >1% (Figure 2). The most abundant genus classified in both the samples was *Limnohabitans* belonging to the family Comamonadaceae. Within the family Comamonadaceae, the genera *Arcobacter*, *Curvibacter*, *Limnohabitans*, and *Rhodoferrax* were identified in both samples, with an additional two genera—*Hydrogenophaga* and *Pelomonas*—exclusively in sample R-F1-E with >1% relative abundance. Furthermore, genus *Aquirufa* was recognized in sample S-F3-N with relative abundance of 25.71%, while 8.86% remained in sample R-F1-E. The bacterial genera classified with relative abundance of <1% in total comprised 33.21 and 22.41% of bacterial community in sample R-F1-E and S-F3-N, respectively.

At the species level, 16 and 11 species were classified from samples R-F1-E and S-F3-N, respectively, with >1% relative abundance (Figure 2F). Samples R-F1-E and S-F3-N were dominated with species

*Limnohabitans parvus* II-B4 and *Aquirufa anthreingensis*, respectively. Four species belonging to genus *Limnohabitans*—*L. australis*, *L. curvus*, *L. parvus* II-B4, and *L. planktonicus*—were identified in both the samples with variable abundance. Furthermore, 73.38 and 72.06% of the bacterial diversity was composed of the bacterial population identified with relative abundance >1% in samples R-F1-E and S-F3-N, respectively. Full length amplicon reads that remained unclassified in samples R-F1-E and S-F3-N were 1.1 and 0.82% of the total analyzed reads, respectively.

### Taxonomic classification comparison with short and long reads 16S rRNA-based data sets

Short and full length 16S rRNA amplicon reads were obtained using Illumina iSeq100 and Oxford Nanopore MinION sequencers. The taxonomic classification results at phylum, genus and species levels



**FIGURE 2** Comparison of sample R-F1-E and S-F3-N sequenced using Illumina iSeq100 (short amplicon reads), Oxford Nanopore MinION (long amplicon reads), and Illumina NovaSeq (shotgun reads) for the classification of phylum (A, D, and G, respectively), genus (B, E, and H, respectively), and species (C, F, and I, respectively) with relative abundance >1%. “Others” in the plots represents reads classified with <1% relative abundance and reads that remains unclassified.

were compared with different input reads (10, 20, 30, 40, and 50 K), randomly extracted from total obtained valid reads—for samples R-F1-E and S-F3-N (Figure 3).

At phylum level classification, Illumina sequenced samples R-F1-E and S-F3-N identified a greater number of phyla than MinION at different input reads (Figure 3A). In sample R-F1-E, an increase in the number of identified phyla was observed from 10 to 20K reads sequenced using Illumina (25 and 28, respectively) and MinION (13 and 15, respectively). With an increase in Illumina and MinION reads from 30 to 50K, a uniform number of phyla were identified, except for Illumina sequenced input read of 50 K (Figure 3A). A similar trend in the number of identified phyla was observed in sample S-F3-N, with an exception that uniformity in the number of identified phyla (31) was observed in Illumina sequenced reads from 30 to 50 K (Figure 3B). However, MinION sequenced input reads of 30 to 40 K identified 16 phyla with a slight increase to 18 at 50K reads. Proteobacteria and

Bacteroidetes were two major phyla identified in sample R-F1-E with >1% relative abundance, sequenced using both the techniques (Figure 3A). However, in sample S-F3-N, total 5-Actinobacteria, Bacteroidetes, Parcubacteria\_OD1, Proteobacteria and Verrucomicrobia and 3-Bacteroidetes, Proteobacteria and Verrucomicrobia were identified with relative abundance >1% from Illumina and MinION sequenced reads, respectively, at different input reads (Figure 3B). The number of genera and the genera classified with relative abundance >1% and remaining unclassified reads formed a uniform trend using both short-and long-amplicons at different input reads. The number of genera identified using Illumina input reads from 10 to 50 K ranged from 339 to 675 for sample R-F1-E, whereas ranged from 338 to 561 for sample S-F3-N (Figures 3C,D). In contrast, MinION sequenced reads identified comparatively fewer genera ranging from 311 to 627 and 265 to 581 for sample R-F1-E and S-F3-N, respectively (Figures 3C,D). However, most genera classified using short amplicon

reads were identified with low confidence values against the database, whereas long amplicon reads had comparatively better resolution for classified genera (Figures 3C,D). For both samples, the unclassified reads were fewer than 8 and 2% of the total input reads using short and long amplicon reads, respectively.

The number of species classified using long amplicon reads was higher than when using short amplicon reads (Figure 3). The number of identified species ranged from 619 to 1,421 and 551 to 1,306 for MinION sequenced samples R-F1-E and S-F3-N, respectively (Figures 3E,F). Whereas Illumina sequenced samples R-F1-E and S-F3-N identified species ranging from 464 to 1,089 and from 408 to 722, respectively (Figures 3E,F). At the species level classification, ~50% and ~33% of the total input reads remained unclassified using short amplicon reads for sample R-F1-E and S-F3-N, respectively, whereas long amplicon reads were classified with high accuracy comprising >98% classified reads (Figures 3E,F). In sample R-F1-E and S-F3-N, the species identified with relative abundance >1%, utilizing long amplicon reads at different inputs comprehends >70% of the identified bacterial microbiota.

In term of relative abundance, almost similar abundance patterns were obtained with each technique when 10–50 K reads were used as an input data—indicated that minimum input of 10 K reads from either Illumina iSeq100 or Oxford Nanopore MinION, can provide similar resolution with 5 times more input reads. However, with respect to the number of classified phyla, Illumina provided better outcomes compared to Oxford Nanopore, and there was no dramatic increase in number of

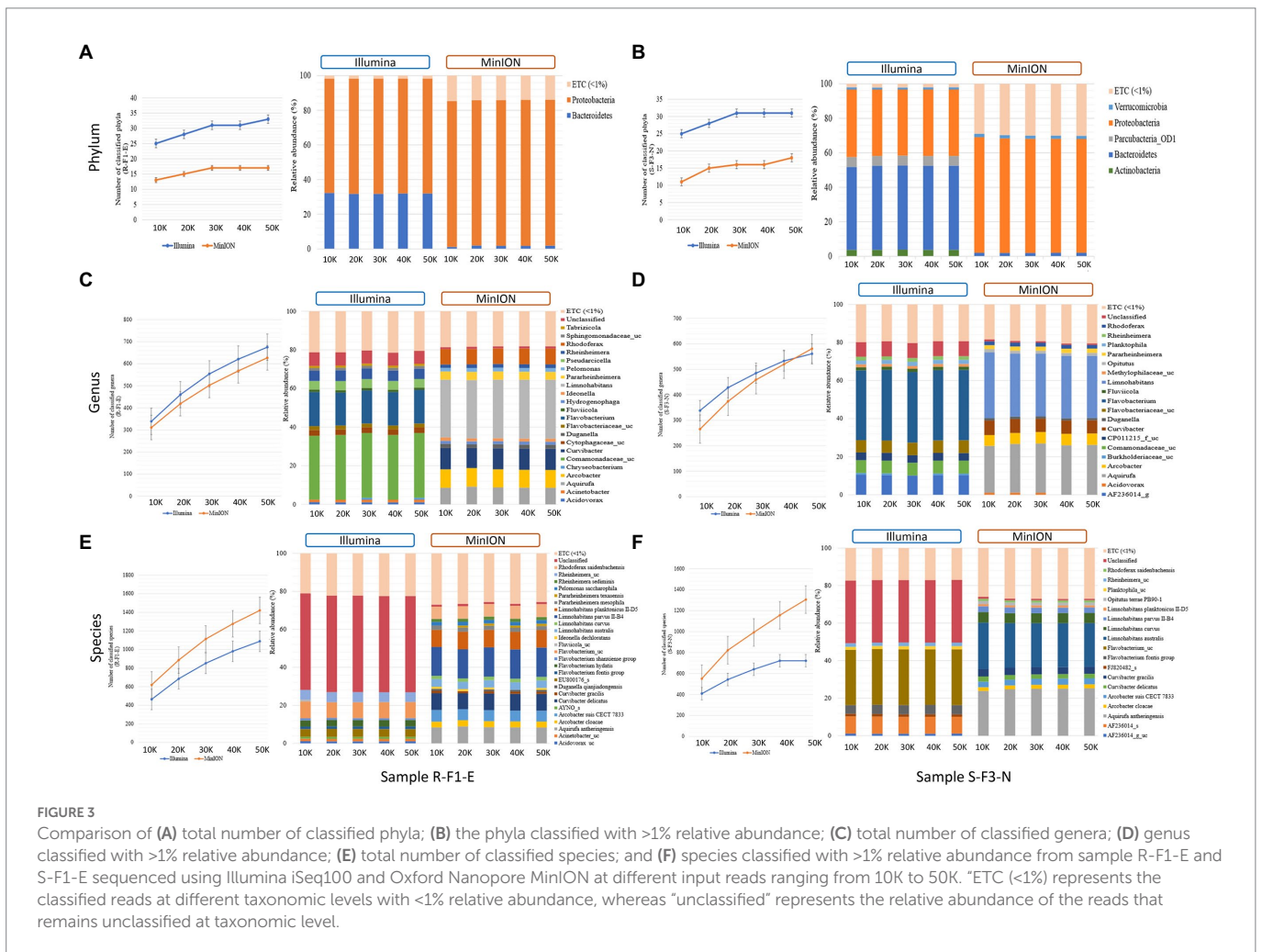
phyla when the input reads were increased from 10 to 50 K by either sequencing technology (Figures 3A,B). The analyzes indicated that Oxford Nanopore MinION is a better choice for higher resolution at genus and species levels (Figures 3C-F). To identify number of genera or species, it is important to include higher number of reads (~> 20 K).

### Shotgun metagenome analysis

A total of 5,61,183 and 4,91,726 non-redundant genes were identified from sample R-F1-E and S-F3-N, respectively, while sharing 1,24,661 (12%) unigenes between both. Despite having close microbial association indicated by PCoA analysis, the samples R-F1-E and S-F3-N were distinctively differentiated based on unique genes composition of 78 and 75%, respectively.

### Taxonomic classification of metagenomics (shotgun) data

According to the obtained abundance table of each taxonomic level, the bar plots were plotted for the top 10 classified phyla, genera, and species (Figures 2G-I). At the phylum level, the most abundant phyla, in both the samples, were Proteobacteria, followed by Bacteroidetes with relative abundance >1%. Additionally, Actinobacteria was also classified in sample S-F3-N with >1% relative abundance, differentiating this from sample R-F1-E in which seven genera—*Curvibacter*, *Limnohabitans*,





*Flavobacterium*, *Pelomonas*, *Rhodobacter*, *Pseudarcicella*, and *Novosphingobium*—were classified with more than 1% relative abundance, whereas only five genera—*Limnohabitans*, *Flavobacterium*, *Rhodoluna*, *Pseudarcicella*, and *Novosphingobium*—were classified in sample S-F3-N. Species level classification revealed 10 species with relative abundance of >1% from both the samples. A high percentage of “others” in the metagenomic analysis could result from an incomplete database.

“Others” representing the relative abundance of the reads that remain unclassified and classified with relative abundance of <1% was higher at phylum, genus, and species level classification for both the samples sequenced using Illumina NovaSeq (shotgun reads) than Illumina iSeq100 and Oxford Nanopore MinION (Figure 2). Sample R-F1-E represented 40.12, 69.88, and 86.12% of reads as “others” at phylum, genus, and species level classification, respectively. Sample S-F3-N at phylum, genus, and species level represented 46.25, 69.04, and 89.53%, respectively, as “others.”

## Functional profiling of active bacterial community

For better insight into the physiology of a bacterial community, the assembled metagenomic protein coding sequences were mapped against three functional databases—eggNOG, KEGG, and CAZy (Supplementary Figure S5). Both samples (R-F1-E and S-F3-N) revealed similarity in annotated gene function profiles and were clustered together.

Annotation based on eggNOG database revealed (Supplementary Figures S5A,B) that highest number genes in sample R-F1-E were associated with inorganic ion, amino acid, carbohydrate, nucleotide, and lipid transport and metabolism, cell motility, and transcription with the relative abundance >1% for each function. Whereas in sample S-F3-N, the maximum number of genes were associated with seven functions and having >1% relative abundance—replication, recombination, and repair, translation, ribosomal structure, and biogenesis, nucleotide transport and metabolism, cell wall/membrane/envelope biogenesis, post-translational modification, protein turnover, chaperons, coenzyme transport and metabolism, and energy production and conversion.

Most of the genes represented in the KEGG pathway analysis were associated with metabolic pathways (Supplementary Figures S5C,D), and particularly dominant in the category of amino acid transport and metabolism having 28,924 and 19,900 associated genes in samples R-F1-E and S-F3-N, respectively. Statistically differential features of functional categories based on KEGG analysis between the two samples were analyzed using STAMP, indicating metabolism, genetic information processing, human diseases, and organismal system dominant in sample S-F3-N, whereas environmental information and cellular processing were enriched in sample R-F1-E (Supplementary Figure S5D).

As per CAZy database-based analysis, glycoside hydrolases (GH) associated genes were most abundant with the relative abundance of 49.33 and 51.87% in sample R-F1-E and S-F3-N, respectively, followed by glycosyl transferase (GT), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), auxiliary activities (AA), polysaccharide lyases (PL; Supplementary Figure S5E). STAMP analysis revealed GH was significantly different with a  $q$ -value of  $4.37e-3$  and was enriched in sample S-F3-N (Supplementary Figure S5F). Whereas glycosyl transferase (GT), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), auxiliary activities (AA), polysaccharide lyases (PL) were higher in sample R-F1-E, with no significant differences observed among these functions.

## Occurrence, abundance, and diversity of ARGs

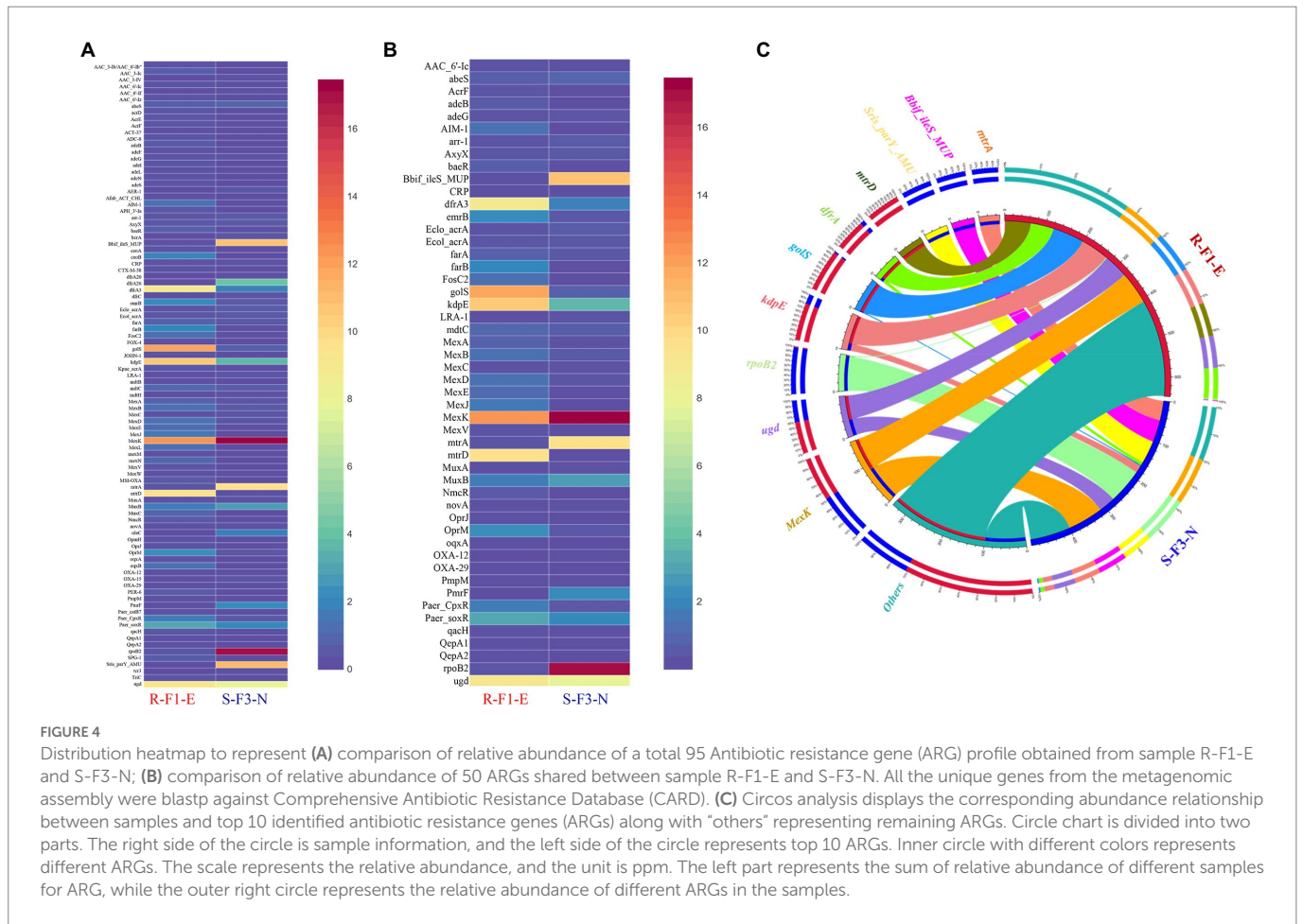
To explore and compare the ARGs profile in sample R-F1-E and S-F3-N, all unique genes obtained from the samples were BLASTp against the CARD database. This analysis revealed the presence of 83 and 62 ARGs in sample R-F1-E and S-F3-N, respectively (Figure 4A), while sharing 50 ARGs between each other with variable relative abundance (Figure 4B). *MexK*, a resistance nodulation cell division (RND) antibiotic efflux pump gene, was the most abundant ARG present in both the samples (Figure 4C).

Furthermore, the top 10 most abundant ARGs out of 95 ARGs, annotated collectively from both samples, were represented in Circos for observing overall proportion and distribution of the resistance genes in both samples (Figure 4C). The top 10 ARGs were: *mexK* (multidrug resistance gene), *ugd* (peptide resistance gene), *rpoB2* (rifamycin resistance gene), *kdpE* (aminoglycoside resistance gene), *gols* (multidrug resistance gene), *dfrA3* (diaminopyrimidine resistance gene), *mtrD* (macrolide resistance gene), *Streptomyces rishiriensis parY* mutant conferring resistance to aminocoumarin (*Sris\_parY\_AMU*; aminocoumarin resistance gene), *Bifidobacterium ileS* conferring resistance to mupirocin (*Bbif\_ileS\_MUP*; mupirocin resistance gene), and *mtrA* (macrolide resistance gene). The relative abundance of gene *ugd*, *kdpE*, *gols*, and *dfrA3* was higher in sample R-F1-E, whereas *mexK*, *rpoB2*, *Bbif\_ileS\_MUP*, and *mtrA* were relatively higher in sample S-F3-N. Interestingly, ARG *mtrD* and *Sris\_parY\_AMU* were only conferred to sample R-F1-E and S-F3-N, respectively.

An additional analysis was performed to reveal the dominant bacterial phyla possessing the most ARG genes with different associated resistance mechanisms. The most abundant resistant mechanism associated with the annotated ARGs corresponded to RND antibiotic efflux pump, followed by major facilitator superfamily (MFS) antibiotic efflux pump, antibiotic target alteration (pmr phosphoethanolamine transferase), protein and two component regulatory system modulating antibiotic efflux (*kdpE*), antibiotic target replacement (*DfrA42\_TMP*), and ABC antibiotic efflux pump. These potential antibiotic mechanisms were associated with the ARG that were affiliated with phyla Proteobacteria (Supplementary Figure S6).

## Discussion

Our study highlighted significant differences and similarities in the bacterial communities of different irrigation water systems from different geographical locations (North, West, and East) on Oahu, Hawaii. Comparative assessment of bacterial communities between samples showed distinctive discriminations based on type of water system and geographical location. It is striking to note that natural stream and associated field water samples were dominated by Proteobacteria, regardless of their geographical locations—there was a close bacterial association between the samples based on beta diversity analysis. These outcomes agreed with the previous studies conducted in Brazil (Godoy et al., 2020) and Tokyo (Reza et al., 2018), which revealed a dominance of Proteobacteria in river water. Samples collected from North Oahu showed close microbial association regardless of different water systems, indicating an influence of geographical locations (topography, water bodies, climatic conditions, natural vegetation, etc.)



in composing the microbial consortia (Aguilar et al., 2020). Field water samples R-F1-E and S-F3-N were clustered based on the microbiota despite being irrigated by different irrigation systems (spring and stream) and different geographical regions (North and East), which prompted us to uncover the complex and diverse microbiota at a higher taxonomic level (Figure 1). The short amplicon reads generated from V3-V4 gene region of 16S rRNA using Illumina iSeq100 was able to detect phyla with high accuracy in addition to classification of most dominant genera as well. However, some genera within the family were not classified with high confidence value and more than 50% of the valid reads were unclassified, indicating a limitation of short amplicon reads for high resolution and accuracy of classification. A study (Nygaard et al., 2020) designed to uncover and compare the microbial consortia of indoor dust sequenced using Illumina and Nanopore MinION revealed significant differences in microbial composition at genus and species levels, with better resolution provided by MinION sequenced reads. Therefore, to investigate the microbiota of sample R-F1-E and S-F3-N at a higher taxonomic level with better resolution, full length 16S rRNA gene region was sequenced using Oxford Nanopore MinION and analyzed. Full length amplicon analysis revealed high abundance of the genus *Limnohabitans* that includes planktonic bacteria and classified other dominant genera within family Comamonadaceae that remained unclassified using short amplicon reads. All the four species within the genus *Limnohabitans* (Hahn et al., 2010; Kasalický et al., 2010) were successfully classified with > 1% relative abundance. Additionally, genus *Aquirufa*, a freshwater bacterium, was identified in spring and stream field water with relative abundance > 1% and *Aquirufa antheringensis*

was the dominant species in spring field water. Another study (Pitt et al., 2019) also found the higher abundance of *A. antheringensis* in fresh water. The resolution obtained for genus and species level classification was better using long amplicon reads with < 2% valid reads that remained unclassified (Figure 2).

Furthermore, we compared the performance of long reads (~ 1,500 bp) obtained from Oxford Nanopore MinION with short reads (~ 300 bp) obtained from Illumina iSeq100 to assess bacterial taxonomic classification at phylum, genus, and species levels with different numbers of input reads. Results from this experimental study showed uniform trends in classification at phylum, genus, and species levels for samples, R-F1-E and S-F3-N, at 10, 20, 30, 40, and 50K input reads (Figure 3). However, when long-and short-read outcomes were compared, dissimilarities in relative abundance at all three taxonomic levels were observed (Figure 3). Short-read-based taxonomic analysis provided the most comprehensive classification at the phylum level compared to 16S rRNA full length reads and shotgun metagenome data (Figures 2, 3). However, 16S rRNA full length reads clearly illustrated its advantage for classification at genus and species levels (Figures 2, 3). In a study (Komiya et al., 2022) proposed Oxford Nanopore MinION as a low cost and rapid technology for revealing microbial communities with higher resolution at the species level which ultimately aids in identifying bacteria potentially pathogenic to human health. In our study, with a high number of unclassified reads at phylum [39.52% (R-F1-E); 45.82% (S-F3-N)], genus [68.04% (R-F1-E); 68.35% (S-F3-N)] and, species [85.37% (R-F1-E); 89.17% (S-F3-N)] levels, we have not observed any advantages of using shotgun metagenome data for taxonomic classification (Figure 2)—this

could be due to the limited and incomplete annotated metagenomic and bacterial genome databases currently available (Pignatelli et al., 2008). With the advancement and improvement in the Nanopore MinION technology, this efficient, cost-effective, and robust technology can be employed for on-field microbiome study of environmental samples with minimum data requirements (Goordial et al., 2017).

The environmental samples consist of complex and diverse microbiota which are better resolved in terms of predication of microbial community's functions. This can be achieved using shotgun metagenomic sequencing with advanced next generation sequencing technologies that generates enormous amounts of genomic data (Meneghini et al., 2017). However, due to different sequencing protocols and annotated databases, metagenome analysis and 16S rRNA gene sequencing cannot provide an identical taxonomic classification, as observed in our study and in (Peterson et al., 2021). Metagenomic functional analysis revealed the presence of 78 and 75% of unique genes in sample R-F1-E and S-F3-N, respectively, while only 12% of the genes were shared between both the samples, but interestingly, were annotated for comparable gene functional profiles (Supplementary Figure S5). The relatively high abundance of genes was related to metabolism of amino acids, nucleotides, carbohydrates, coenzymes, lipids, and inorganic ion metabolism and transport. 'Amino acid metabolism' was enriched in both the samples, which may be due to fertilizer residues that provide a suitable living environment for microbiota that use amino acids. Additionally, environmental samples consist of diverse and abundant complex mixtures of carbohydrates requiring different enzymes for metabolism, mainly supported by glycoside hydrolases (GH; Berlemont and Martiny, 2016). In our study, GH were the most abundant and significantly different among all the other identified enzymes in both samples (Figures 5E,F). This enzyme assists in the enzymatic processing of carbohydrate, ultimately contributing to functioning of an ecosystem, global carbon cycling. The metagenomic data also revealed the prevalence of a variety of ARGs in both the samples. The ubiquity of ARGs in the environmental sample is an emerging concern. A study (Pruden et al., 2006) documented the prevalence of ARGs in irrigation ditch water and urban/agriculturally impacted river sediments leading to the potential spread of ARGs to or from humans. From 95 identified ARGs, only 50 genes were shared between both the samples with variable abundance depending on the microbial consortia and their genome compositions (Figure 4)—the genomic composition can be altered through horizontal gene transfer from environment or other bacteria mediated by mobile genetic elements such as plasmids, transposons, bacteriophages, insertion sequences and integrons (Stalder et al., 2012; Rizzo et al., 2013). The most abundant ARG in both the samples was *MexK*, a resistance nodulation cell division (RND) antibiotic efflux pump gene which can transport multiple classes of antimicrobials, contributing to multidrug resistance (Coldough et al., 2020). Therefore, uncovering the bacterial components, functional analysis, and investigation of the ARGs will resolve the microbial complexity and help to formulate better disease management strategies for water transmitted pathogens.

## Conclusion

The bacterial consortia found in different water source of taro irrigation across the island of Oahu, Hawaii revealed that Proteobacteria is the most dominant phyla, except for a few samples

from storage tank and spring water. The most reliable and comprehensive taxonomic classifications at phylum and genus/species levels were observed with input reads obtained from Illumina and Oxford Nanopore, respectively. The lack of robust and comprehensive annotated metagenome and bacterial genome databases contributed to inconclusive classification using shotgun metagenome reads, particularly at genus and species levels. However, metagenomic data contributed to the understanding of gene distribution of microbiomes and their functions, including ARGs, associated with different microbial consortia. This study provided some appropriate sequencing platforms and pipelines to study irrigation water microbiome.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://www.ncbi.nlm.nih.gov/genbank/>, PRJNA856319, <https://www.ncbi.nlm.nih.gov/genbank/>, PRJNA856335, <https://www.ncbi.nlm.nih.gov/genbank/>, PRJNA856390.

## Author contributions

MA conceived and designed the study. DK performed the experiments, analyzed the data, and wrote the manuscript. SD performed the experiment. ZH helped in the Illumina library preparation. AA, JU, and JS helped in sample collection. MA, SD, AA, ZH, JU, JS, K-HW, SK and AMA revised the manuscript and provided ideas and support for the final submission. All authors contributed to the article and approved the submitted version.

## Funding

This research was funded by NIGMS of the National Institutes of Health under award number P20GM125508. This work was also supported by the USDA National Institute of Food and Agriculture, Hatch project 9038H, managed by the College of Tropical Agriculture and Human Resources.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1039292/full#supplementary-material>

### SUPPLEMENTARY FIGURE 1

Bar plot comparison of phylum level classification, classified with relative abundance of >1% in 11 samples- R-F1-E, R-S1-E, R-S2-W, R-F2-W, S-F3-N, R-S4-SE, R-F4-SE, R-S5-SE, R-F5-E, T-S6-N, and R-S7-N (Replicate 1 and Replicate 2) sequenced for short length amplicon using Illumina iSeq100 and analyzed on EzBioCloud platform. "Others" represents the reads classified with less than <1% relative abundance and remains unclassified in the classification against the database.

### SUPPLEMENTARY FIGURE 2

Distribution heatmap of bacterial species classified with >1% relative abundance among all the 12 water samples—sequenced for V3-V4 region of 16S rRNA gene region using Illumina iSeq100 sequencing platform. The generated short amplicon reads were analyzed using EzBioCloud platform. The heatmap was generated using display R.

### SUPPLEMENTARY FIGURE 3

UPGMA (unweighted pair group method with arithmetic mean) clustering of water samples based on Bray-Curtis dissimilarity index at genus level. Samples were grouped in three distinctive clusters: Cluster 1 (R-F1-E and S-F3-N) irrespective of water system or geographical location, Cluster 2 (R-S1-E, R-F2-W, R-S2-W, R-F4-SE, R-S4-SE, R-F5-SE, and R-S5-SE) based on irrigation source and

associated taro field water, and Cluster 3 (S-S3-N, T-S6-N, and R-S7-N) based on geographical location.

### SUPPLEMENTARY FIGURE 4

Bar plot comparing the (A) genus and (B) species classified with relative abundance of >1% in sample R-F1-E (Replicate 1 and Replicate 2) sequenced for full length amplicon using Oxford Nanopore MinION and analyzed on EPI2ME platform. Input valid reads that were not classified to genus and species levels are represented as "Unclassified", while "ETC (<1%)" represents the bacterial population identified with relative abundance of <1%.

### SUPPLEMENTARY FIGURE 5

Comparison of samples R-F1-E and S-F3-N for relative abundance and statistical differences of annotated gene function profiles based on mapping of assembled metagenomic protein coding sequences to three databases: (A,B) non-supervised Orthologous groups (eggNOG), (C,D) Kyoto Encyclopedia of Genes and Genomes (KEGG), and (E,F) Carbohydrate-Active Enzymes Database (CAZY). Statistical analyses performed using STAMP v 2.1.3 software, employing Fisher's exact test with Newcombe-Wilson CI method and Benjamini-Hochberg FDR correction factors, and visualized using extended error bar plots.

### SUPPLEMENTARY FIGURE 6

Circos analysis displays the corresponding abundance relationship between identified dominant phyla (Proteobacteria and Actinobacteria) along with "other" representation of identified phyla and associated resistance mechanism. Circle chart is divided into two parts. The right side of the circle is phyla information, and the left side of the circle is antibiotic resistance mechanisms. Inner circle with different colors represents different antibiotic resistance mechanisms. The scale represents the relative abundance, and the unit is ppm. The left part represents the sum of relative abundance of different phyla for resistance mechanisms, while the outer right circle vice versa.

## References

- Aguilar, M. O., Gobbi, A., Browne, P. D., Ellegaard-Jensen, L., Hansen, L. H., Semorile, L., et al. (2020). Influence of vintage, geographic location and cultivar on the structure of microbial communities associated with the grapevine rhizosphere in vineyards of San Juan Province, Argentina. *PLoS One* 15:e0243848. doi: 10.1371/journal.pone.0243848
- Berlemont, R., and Martiny, A. C. (2016). Glycoside hydrolases across environmental microbial communities. *PLoS Comput. Biol.* 12:e1005300. doi: 10.1371/journal.pcbi.1005300
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Cevallos-Cevallos, J., Danyluk, M., Gu, G., Vallad, G., and van Bruggen, A. (2012). Dispersal of *salmonella* typhimurium by rain splash onto tomato plants. *J. Food Prot.* 75, 472–479. doi: 10.4315/0362-028X.JFP-11-399
- Cevallos-Cevallos, J. M., Gu, G., Richardson, S. M., Hu, J., and van Bruggen, A. H. C. (2014). Survival of *salmonella enterica* typhimurium in water amended with manure. *J. Food Prot.* 77, 2035–2042. doi: 10.4315/0362-028X.JFP-13-472
- Colclough, A. L., Alav, I., Whittle, E. E., Pugh, H. L., Darby, E. M., Legood, S. W., et al. (2020). RND efflux pumps in gram-negative bacteria; regulation, structure and role in antibiotic resistance. *Future Microbiol.* 15, 143–157. doi: 10.2217/fmb-2019-0235
- Diaz, P. I., Dupuy, A. K., Abusleme, L., Reese, B., Oberfell, C., Choquette, L., et al. (2012). Using high throughput sequencing to explore the biodiversity in oral bacterial communities. *Mol. Oral Microbiol.* 27, 182–201. doi: 10.1111/j.2041-1014.2012.00642.x
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- Godoy, R. G., Marcondes, M. A., Pessôa, R., Nascimento, A., Victor, J. R., Duarte, A. J. D. S., et al. (2020). Bacterial community composition and potential pathogens along the Pinheiros River in the southeast of Brazil. *Sci. Rep.* 10:9331. doi: 10.1038/s41598-020-66386-y
- Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351. doi: 10.1038/nrg.2016.49
- Goordial, J., Altschuler, I., Hindson, K., Chan-Yam, K., Marcoletas, E., and Whyte, L. G. (2017). *In situ* field sequencing and life detection in remote (79°26'N) Canadian high Arctic permafrost ice wedge microbial communities. *Front. Microbiol.* 8:2594. doi: 10.3389/fmicb.2017.02594
- Gu, G., Cevallos-Cevallos, J. M., and van Bruggen, A. H. C. (2013). Ingress of *Salmonella enterica* typhimurium into tomato leaves through hydathodes. *PLoS One* 8:e53470. doi: 10.1371/journal.pone.0053470
- Gu, S., Fang, L., and Xu, X. (2013). Using SOAPaligner for Short Reads Alignment. *Curr. Protoc. Bioinformatics* 44, 11.11.1–11.11.17. doi: 10.1002/0471250953.bi1111s44
- Hahn, M. W., Kasalický, V., Jezbera, J., Brandt, U., and Šimek, K. (2010). *Limnohabitans australis* sp. nov., isolated from a freshwater pond, and emended description of the genus *Limnohabitans*. *Int. J. Syst. Evol. Microbiol.* 60, 2946–2950. doi: 10.1099/ijs.0.022384-0
- Heikema, A. P., Horst-Kreft, D., Boers, S. A., Jansen, R., Hiltmann, S. D., de Koning, W., et al. (2020). Comparison of Illumina versus Nanopore 16S rRNA gene sequencing of the human nasal microbiota. *Genes* 11:1105. doi: 10.3390/genes11091105
- Herlemann, D. P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., and Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5, 1571–1579. doi: 10.1038/ismej.2011.41
- Hintz, L. D., Boyer, R. R., Ponder, M. A., Williams, R. C., and Rideout, S. L. (2010). Recovery of *salmonella enterica* Newport introduced through irrigation water from tomato (*Lycopersicon esculentum*) fruit, roots, stems, and leaves. *HortScience* 45, 675–678. doi: 10.21273/HORTSCI.45.4.675
- Kasalický, V., Jezbera, J., Šimek, K., and Hahn, M. W. (2010). *Limnohabitans planktonicus* sp. nov., and *Limnohabitans parvus* sp. nov., two novel planktonic Betaproteobacteria isolated from a freshwater reservoir. *Int. J. Syst. Evol. Microbiol.* 60, 2710–2714. doi: 10.1099/ijs.0.018952-0
- Komiya, S., Matsuo, Y., Nakagawa, S., Morimoto, Y., Kryukov, K., Okada, H., et al. (2022). MinION, a portable long-read sequencer, enables rapid vaginal microbiota analysis in a clinical setting. *BMC Med. Genet.* 15:68. doi: 10.1186/s12920-022-01218-8
- Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., et al. (2015). Assessing the performance of the Oxford Nanopore technologies MinION. *Biomol. Detect. Quantif.* 3, 1–8. doi: 10.1016/j.bdq.2015.02.001
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Luczkiewicz, A., Kotlarska, E., Artichowicz, W., Tarasewicz, K., and Fudala-Ksiażek, S. (2015). Antimicrobial resistance of *pseudomonas* spp. isolated from wastewater and wastewater-impacted marine coastal zone. *Environ. Sci. Pollut. Res. Int.* 22, 19823–19834. doi: 10.1007/s11356-015-5098-y
- Masella, A. P., Bartram, A. K., Truszowski, J. M., Brown, D. G., and Neufeld, J. D. (2012). PANDaseq: paired-end assembler for illumina sequences. *BMC Bioinform.* 13:31. doi: 10.1186/1471-2105-13-31
- Matsuo, Y., Komiya, S., Yasumizu, Y., Yasuoka, Y., Mizushima, K., Takagi, T., et al. (2021). Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. *BMC Microbiol.* 21:35. doi: 10.1186/s12866-021-02094-5

- Meneghini, A. K., Nielsen, S., Varani, A. M., Thomas, T., and Alves, L. M. C. (2017). Metagenomic analysis of soil and freshwater from zoo agricultural area with organic fertilization. *PLoS One* 12:e0190178. doi: 10.1371/journal.pone.0190178
- Nygaard, A. B., Tunsjø, H. S., Meisal, R., and Charnock, C. (2020). A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA gene sequencing to characterize building-dust microbiomes. *Sci. Rep.* 10:3209. doi: 10.1038/s41598-020-59771-0
- Parks, D. H., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123–3124. doi: 10.1093/bioinformatics/btu494
- Paul, P. A., El-Allaf, S. M., Lipps, P. E., and Madden, L. V. (2004). Rain splash dispersal of *Gibberella zeae* within wheat canopies in Ohio. *Phytopathology* 94, 1342–1349. doi: 10.1094/PHYTO.2004.94.12.1342
- Peng, X., Wilken, S. E., Lankiewicz, T. S., Gilmore, S. P., Brown, J. L., Henske, J. K., et al. (2021). Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. *Nat. Microbiol.* 6, 499–511. doi: 10.1038/s41564-020-00861-0
- Peterson, D., Bonham, K. S., Rowland, S., Pattanayak, C. W., and Klepac-Ceraj, V. RESONANCE Consortium (2021). Comparative analysis of 16S rRNA gene and metagenome sequencing in pediatric gut microbiomes. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.670336
- Pignatelli, M., Aparicio, G., Blanquer, I., Hernández, V., Moya, A., and Tamames, J. (2008). Metagenomics reveals our incomplete knowledge of global diversity. *Bioinformatics* 24, 2124–2125. doi: 10.1093/bioinformatics/btn355
- Pitt, A., Schmidt, J., Koll, U., and Hahn, M. W. (2019). *Aquirufa antheringensis* gen. nov., sp. nov. and *Aquirufa nivalisilvae* sp. nov., representing a new genus of widespread freshwater bacteria. *Int. J. Syst. Evol. Microbiol.* 69, 2739–2749. doi: 10.1099/ijsem.0.003554
- Pruden, A., Pei, R., Storteboom, H., and Carlson, K. H. (2006). Antibiotic resistance genes as emerging contaminants: studies in northern Colorado. *Environ. Sci. Technol.* 40, 7445–7450. doi: 10.1021/es060413l
- Qin, Q., Chen, X., and Zhuang, J. (2015). The fate and impact of pharmaceuticals and personal care products in agricultural soils irrigated with reclaimed water. *Crit. Rev. Environ. Sci. Technol.* 45, 1379–1408. doi: 10.1080/10643389.2014.955628
- Ravva, S. V., Sarreal, C. Z., Duffy, B., and Stanker, L. H. (2006). Survival of *Escherichia coli* O157:H7 in wastewater from dairy lagoons. *J. Appl. Microbiol.* 101, 891–902. doi: 10.1111/j.1365-2672.2006.02956.x
- Redekar, N. R., Eberhart, J. L., and Parke, J. L. (2019). Diversity of *Phytophthora*, *Pythium*, and *Phytophythium* species in recycled irrigation water in a container nursery. *Phytophthora* 3, 31–45. doi: 10.1094/PBIOMES-10-18-0043-R
- Reza, M. S., Mizusawa, N., Kumano, A., Oikawa, C., Ouchi, D., Kobiyama, A., et al. (2018). Metagenomic analysis using 16S ribosomal RNA genes of a bacterial community in an urban stream, the Tama River, Tokyo. *Fish. Sci.* 84, 563–577. doi: 10.1007/s12562-018-1193-6
- Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., et al. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* 9, 1038–1048. doi: 10.1038/nprot.2014.067
- Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M. C., et al. (2013). Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. *Sci. Total Environ.* 447, 345–360. doi: 10.1016/j.scitotenv.2013.01.032
- Sanz-Martin, I., Doolittle-Hall, J., Teles, R. P., Patel, M., Belibasakis, G. N., Hämmerle, C. H. F., et al. (2017). Exploring the microbiome of healthy and diseased peri-implant sites using Illumina sequencing. *J. Clin. Periodontol.* 44, 1274–1284. doi: 10.1111/jcpe.12788
- Schmieder, R., and Edwards, R. (2012). Insights into antibiotic resistance through metagenomic approaches. *Future Microbiol.* 7, 73–89. doi: 10.2217/fmb.11.135
- Stalder, T., Barraud, O., Casellas, M., Dagot, C., and Ploy, M.-C. (2012). Integron involvement in environmental spread of antibiotic resistance. *Front. Microbiol.* 3:119. doi: 10.3389/fmicb.2012.00119
- Szczepanowski, R., Linke, B., Krahn, I., Gartemann, K.-H., Gützkow, T., Eichler, W., et al. (2009). Detection of 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to selected antibiotics. *Microbiology (Reading)* 155, 2306–2319. doi: 10.1099/mic.0.028233-0
- Tringe, S. G., and Hugenholtz, P. (2008). A renaissance for the pioneering 16S rRNA gene. *Curr. Opin. Microbiol.* 11, 442–446. doi: 10.1016/j.mib.2008.09.011
- Uyttendaele, M., Jaykus, L.-A., Amoah, P., Chiodini, A., Cunliffe, D., Jacxsens, L., et al. (2015). Microbial hazards in irrigation water: standards, norms, and testing to manage use of water in fresh produce primary production. *Compr. Rev. Food Sci. Food Saf.* 14, 336–356. doi: 10.1111/1541-4337.12133
- Van der Linden, I., Cottyn, B., Uyttendaele, M., Vlaemynck, G., Maes, M., and Heyndrickx, M. (2013). Long-term survival of *Escherichia coli* O157:H7 and *salmonella enterica* on butterhead lettuce seeds, and their subsequent survival and growth on the seedlings. *Int. J. Food Microbiol.* 161, 214–219. doi: 10.1016/j.ijfoodmicro.2012.12.015
- Xu, J., Xu, Y., Wang, H., Guo, C., Qiu, H., He, Y., et al. (2015). Occurrence of antibiotics and antibiotic resistance genes in a sewage treatment plant and its effluent-receiving river. *Chemosphere* 119, 1379–1385. doi: 10.1016/j.chemosphere.2014.02.040
- Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/ijsem.0.001755
- Zhang, T., and Li, B. (2011). Occurrence, transformation, and fate of antibiotics in municipal wastewater treatment plants. *Crit. Rev. Environ. Sci. Technol.* 41, 951–998. doi: 10.1080/10643380903392692
- Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 38:e132. doi: 10.1093/nar/gkq275