



Dual-platform metagenomic surveillance distinguishes pathogen and resistome hotspots across agricultural and mixed-use watersheds

Yichao Shi^a, Haley Sanderson^b, Jiacheng Chuan^c, Izhar U.H. Khan^a, Mark Sunohara^a, Emilia Craiovan^a, David R. Lapen^a, Moussa Diarra^d, Wen Chen^{a,e,*}

^a Ottawa Research & Development Centre, Science & Technology Branch, Agriculture & Agri-Food Canada (AAFC), 960 Carling Avenue, Ottawa, ON K1A 0C6, Canada

^b Biological Informatics Center of Excellence, Information Systems Branch, AAFC, 107 Science Pl, Saskatoon, SK S7N 5E2, Canada

^c The Charlottetown Laboratory, Canadian Food & Inspection Agency (CFIA), 93 Mt Edward Rd, Charlottetown, PEI C1A 5T1, Canada

^d Guelph Research and Development Centre, AAFC, 93 Stone Rd W, Guelph, ON N1G 5C9, Canada

^e Department of Biology, University of Ottawa, Marie-Curie Private, Ottawa, ON K1N 9A7, Canada

ARTICLE INFO

Keywords:

Freshwater microbiome
agricultural watersheds
metagenomic sequencings
pathogens
antimicrobial resistance

ABSTRACT

Freshwater systems embedded in agricultural landscapes serve as dynamic reservoirs and conduits for fecal-associated microbes, zoonotic pathogens, and antimicrobial resistance (ARG) and virulence factor (VF) genes. Yet factors that govern their densities and diversity remain a research challenge. From 2016 to 2021, we conducted a longitudinal water surveillance in an agriculturally dominated river basin in eastern Ontario, Canada; characterizing fecal-associated bacterial communities using 16S rRNA gene amplicon and shotgun metagenomic sequencing. Agricultural drainage ditches consistently harbored higher fecal-associated bacterial diversity with pronounced seasonal shifts; i.e., higher levels during larger flow periods in spring and fall. Elevated discharge was associated with enrichment of genera containing zoonotic or opportunistic pathogens, such as those in *Pseudomonas*, *Sphingomonas*, and *Massilia*. Conditionally rare taxa (CRTs), although typically low in abundance, accounted for ~12.6% of all pathogen-associated genera and disproportionately contributed to community turnover, highlighting their role as transient reservoirs of microbial risk. Shotgun metagenomics detected 27 ARGs, primarily at mixed-use sites, and 14 VFs, mainly in agricultural ditches. Clinically relevant β -lactamase genes (e.g., *oxa*, *imp*, *sme*) co-occurred with metal-resistance operons, a pattern suggestive of possible co-selection, although selective agents were not directly measured. Although the prevalence of ARG and VF was low (<5% of samples), their ecological context indicates potential transmission pathways. Limited overlap in ARGs between short-read and metagenome-assembled genome (MAG)-based profiling reflects their complementary strength: gene-level sensitivity versus host-resolved analysis. Together, these findings demonstrate the utility of integrated amplicon and shotgun metagenomic surveillance for proactive One Health risk assessment in agricultural watersheds.

1. Introduction

The One Health framework emphasizes the interconnectedness of human, animal, and environmental health, aiming to improve, for example, outcomes and challenges as associated with microbial pathogen emergence and antimicrobial resistance (AMR) through coordinated, cross-sectoral action [1,2]. Zoonotic pathogens, which account for approximately 60% of infectious diseases in humans, can harbor and disseminate AMR genes and virulence factors (VFs), posing significant

public health risks [3]. The widespread use of antibiotics in animal husbandry exerts strong selective pressure on gut microbiota, facilitating the emergence and spread of antibiotic-resistant bacteria (ARBs) [4]. These ARBs can further transfer their resistance genes among native flora through horizontal gene transfer events, vectors, and quorum sensing mechanisms [5]. Moreover, unmetabolized antibiotics often enter the environment and accumulate in soils, sediments, and aquatic systems, where they promote the persistence and propagation of resistance [6]. Consequently, aquatic environments receiving inputs from

* Corresponding author at: Ottawa Research & Development Centre, Science & Technology Branch, Agriculture & Agri-Food Canada (AAFC), 960 Carling Avenue, Ottawa, ON K1A 0C6, Canada.

E-mail address: wen.chen@agr.gc.ca (W. Chen).

<https://doi.org/10.1016/j.onehlt.2026.101384>

Received 27 November 2025; Received in revised form 4 March 2026; Accepted 5 March 2026

Available online 8 March 2026

2352-7714/Crown Copyright © 2026 Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

agricultural runoff and drainage, livestock operations, and municipal wastewater [7,8], can concurrently act as reservoirs and sources for AMR, VFs, and fecal-associated pathogens, amplifying exposure risks for humans and animals [9–11].

Aquatic systems are highly dynamic, and their microbial composition and resistome profiles are strongly shaped by land use. Agricultural landscapes with intensive livestock production and/or seasonal manure applications to land, are especially susceptible to AMR gene (ARGs) accumulation and the proliferation of fecal-associated zoonotic pathogens, including *Escherichia coli* O157:H7, *Salmonella* spp., and *Campylobacter* spp. [12–14]. Urban and other anthropogenic land uses further compound selective pressures through wastewater and industrial effluent discharges [15–17] for example. Importantly, resistance and virulence determinants are also naturally present in relatively undisturbed ecosystems, where they provide ecological functions within competition, stress tolerance, and survival niches [18]. Monitoring ARGs, VFs, and fecal-associated pathogens in variably influenced waterways helps identify hotspots of contamination and to establish ecological baselines that distinguish anthropogenic impacts from natural background levels [19].

Conventional monitoring approaches, such as culture-based assays and quantitative PCR (qPCR), are valuable for detecting specific fecal indicator bacteria (FIB) and targeted pathogens, ARGs or VF genes. However, such methods capture only a fraction of microbial diversity which can overlook broader ecological interactions and environmental drivers [20–22]. In Canadian watersheds, studies have shown that pathogen and ARG distributions are shaped by land use and hydrology [23,24], yet most investigations are constrained by limited spatial or temporal coverage, impeding comprehensive landscape-scale understanding. Advancements in high-throughput sequencing technologies, including both the amplicon-based and shotgun metagenomics, have transformed microbial ecology by enabling comprehensive, systematic snapshots of microbial diversity, taxonomic composition, and functional potential over time in complex environments [25,26]. Previous studies have shown that combining both approaches was essential to fully capture microbial and functional diversity, in thermal spring ecosystems for example, because each provides complementary insights [27]. The 16S rRNA gene metabarcoding enables sensitive and cost-effective taxonomic profiling of bacteria, making it well-suited for large-scale ecological observatories and detecting compositional shifts across environmental gradients [28,29]. In contrast, shotgun metagenomics provides an untargeted, genome-wide view of community DNA, allowing simultaneous characterization of ARGs, VFs, mobile genetic elements, and broader functional attributes [30].

While both amplicon and shotgun metagenomic approaches have proven invaluable in environmental microbiome research, each entails methodological limitations that shape data interpretation. Amplicon-based approach, though sensitive and scalable, can be affected by primer bias, limited taxonomic resolution, and constrained functional inference [31]. Shotgun metagenomics, while functionally comprehensive, can yield community structure patterns that differ from marker-gene profiling due to uneven genome coverage, reference database limitations, and taxonomic classification challenges, particularly at finer taxonomic resolutions [30,32,33]. Linking functional genes to their taxonomic origins also remains difficult due to assembly fragmentation, horizontal gene transfer (HGT), and incomplete reference databases [25,34,35]. Both approaches share a common limitation in resolving bacterial communities at the species level. In short-read 16S rRNA gene sequencing, the conserved nature of the amplified region constrains taxonomic resolution, often preventing reliable differentiation among closely related species within the same genus [36]. Similarly, in shallow shotgun metagenomics, species-level classification is hindered by incomplete reference genomes, HGT, and sequence conservation across taxa [34,37]. These challenges underscore the importance of cautious interpretation and, when feasible, the integration of complementary targeted methods such as qPCR to improve pathogen identification

accuracy.

Despite substantial advances in microbial ecology, significant knowledge gaps remain. Long-term, multi-seasonal investigations that link microbial dynamics to land use pressures, environmental gradients, as well as targeted assessments of the interplay between core and conditionally rare taxa as ecosystem health indicators are still limited [38–40]. Moreover, few studies have concurrently examined fecal-associated pathogens and ARGs across landscapes using integrated multi-omic frameworks, though global, regional, and watershed-scale investigations have revealed strong associations between anthropological interventions, such as agricultural and urban developments, and fecal pollution, pathogen dissemination, and ARG dynamics [41–43]. To address these gaps, we conducted a five-year longitudinal study (2016–2021) within an agriculturally dominated but mixed-use river basin in eastern Ontario, Canada. Combining 16S rRNA gene amplicon sequencing and shotgun metagenomic sequencing, we characterized fecal-associated bacterial communities, resistomes, and virulomes across spatial and temporal gradients in surface waters.

This study was guided by the central question: *How do land use and environmental conditions shape the diversity and structure of fecal-associated bacterial communities, particularly the dynamics of potential zoonotic pathogens, ARGs and VFs, in agricultural and mixed-use freshwater ecosystems?* To address this overarching question, we pursued four specific objectives: 1) characterize how land use and seasonal conditions influence the diversity and structure of fecal-associated microbiota, with emphasis on the ecological roles of core and conditionally rare taxa (CRT) in freshwater pathogen reservoirs as quantified health indicators; 2) compare the resolution of 16S rRNA gene amplicon sequencing and shotgun metagenomics for profiling fecal-associated communities and detecting potential pathogens; 3) assess the spatial and temporal dynamics of bacterial genera containing pathogenic species; and 4) map the distribution of ARGs and VFs across land use types to identify their environmental hotspots. Collectively, this study endeavors to advance our understanding of fecal contamination and environmental resistome and virulome determinants in agroecosystems, and by extension, contribute to the advancement of risk-factor based microbial water quality surveillance within a One Health framework.

2. Materials & methods

2.1. Study site and sampling design

This study was conducted as part of the Environmental Change OneHealth Observatory (ECO²), a long-term field-based research initiative focused on natural capital in agroecosystems and its influence on animal, environment, and public health, with due consideration of maintaining or augmenting agricultural productivity [44,45]. Fieldwork was conducted in the South Nation River (SNR) basin in eastern Ontario, Canada, which covers approximately 3,900 km². Land use in the basin is primarily agricultural (>60% of land area), including dairy farming operations and associated field cropping systems. Stream sampling locations were selected to represent gradients in upstream land use and stream order [38].

This study included eight water sampling sites representing different upstream land uses (Fig. 1). Four agricultural drainage ditch sampling sites (SN_18, SN_19, SN_20, and SN_21), representative of the ubiquity of such features in the region, have been documented in earlier studies [39,44]. Sites SN_18 and SN_20 were monitored continuously from 2016 to 2021, while SN_19 and SN_21 were included since 2017. The first ditch (SN_18 and SN_19) remained unmanaged, whereas the second (SN_20 and SN_21) underwent vegetation brushing in February 2018 followed by channel dredging in November 2018 [44]. The managed and unmanaged ditches drain catchment areas are 225 ha and 467 ha, respectively [46]. The drainage ditches are shallow, open-channel systems with nominal water depths around 0.2 to 0.5m (occasionally 1m during high rainfall) depth. Across the 2016–2021 monitoring period for

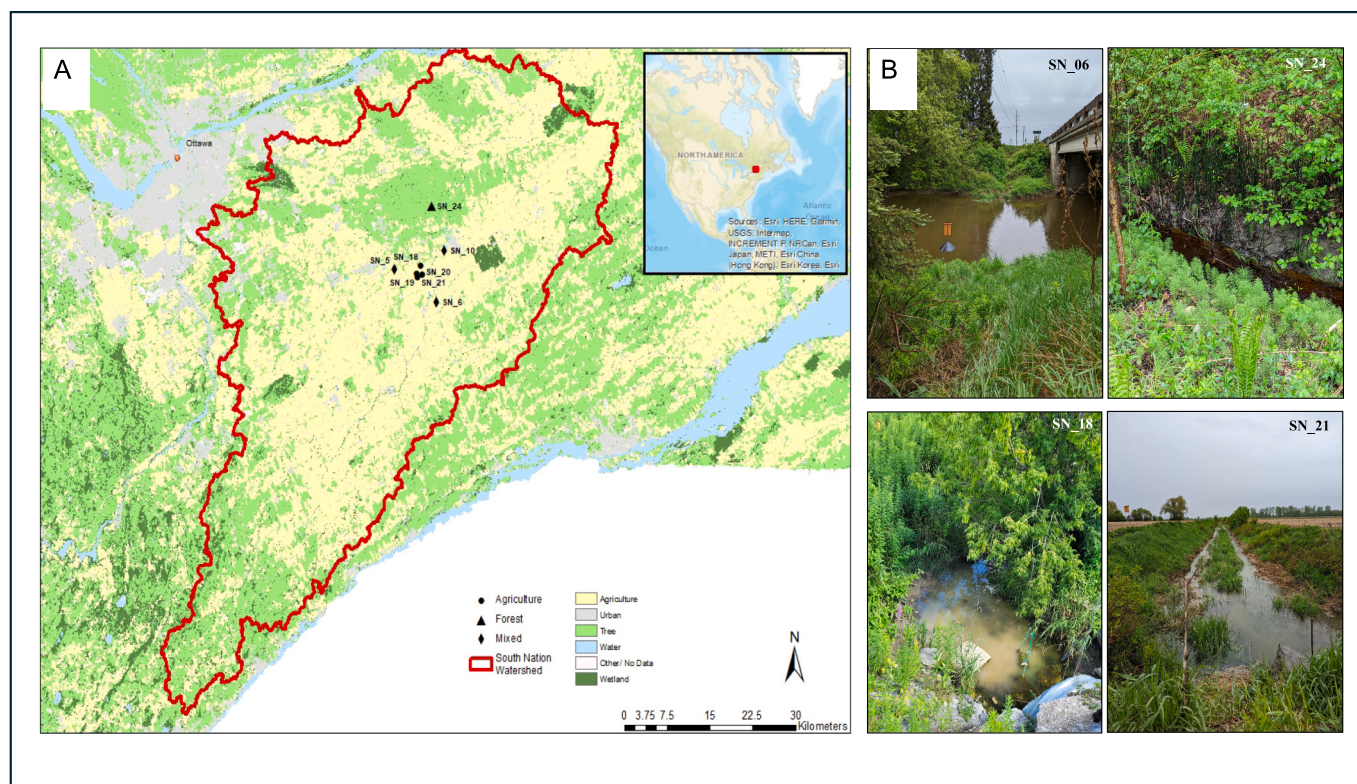


Fig. 1. Geological map of the study site (A) and photographs of representative sampling locations across each land use type (B).

all ditch sites, specific conductivity averaged 0.84 mS cm^{-1} (range: $0.38\text{--}2.27 \text{ mS cm}^{-1}$), and turbidity averaged 46 NTU (range: $0.5\text{--}944 \text{ NTU}$). While these site pairs served as upstream-downstream monitoring points within their respective watersheds, they were not replicates in the strict sense. Prior to 2018, neither ditch had experienced major maintenance for over 40 years, aside from occasional removal of in-stream debris dams etc. by farmers. Three additional stream monitoring sites (SN_5, SN_6, and SN_10) were influenced by mixed-land uses, which included agriculture, urban development, and forested areas. A single forested reference site (SN_24), situated at outflow of a treed wetland with a $<5 \text{ km}^2$ catchment area, was also included in the analysis as a baseline/control given there were no anthropogenic land uses upstream of the site [45].

Surface water samples were collected biweekly during ice free conditions from April to November over six years (2016–2021). Sampling in the ditches was interrupted during dry-spell periods which induced limited (discontinuous water pools) or no-flow conditions, to avoid sample contamination by sediments. Due to logistical constraints imposed by the COVID-19 pandemic, only a limited number of samples were collected in 2020; however, a total of 499 samples were collected across the eight sampling sites over the entire study period. Samples were collected from 0–50 cm depth (avoiding sediment in sample), transported on ice, and processed within 24 hours at the Ottawa Research and Development Centre, Agriculture and Agri-Food Canada (AAFC). Water physicochemical parameters and land use characteristics followed established methods [39]. On-site measurements included water temperature, pH, dissolved oxygen, specific conductivity, and turbidity. These parameters were measured using YSI multi-parameters sondes; a YSI 6600 multi-parameter sonde was used from 2016 to 2018 and a YSI ProDSS multi-parameter digital water quality meter was used from 2019 to 2021. Nutrient analyses (e.g., ammonia, nitrite, nitrate, total Kjeldahl nitrogen, dissolved reactive phosphorus, and total phosphorus) were conducted at the City of Ottawa Laboratory Services located at the Robert O. Pickard Environmental Centre (Ottawa, ON).

Organic carbon (total and dissolved) was measured via APHA Method 2540D [47].

Land use and stream order were determined using GIS-based catchment analyses (2, 5, and 10 km upstream) following Wilkes et al. [48] and Lyautey et al. [49]. Meteorological data (air temperature, precipitation, solar radiation) were recorded near SN_20 using a HOBO data logger (Bourne, MA, USA) from 2016 to 2018 and then a Forest Technology Services (FTS) weather data logger (Victoria, BC, Canada) from 2018 to 2021, with metrics calculated for the day of sampling and up to seven days prior. Due to differences in stream order among the sites, river discharge data for the Castor River (Russell Station 02LB006) were retrieved from the Water Survey of Canada [50] as a standardized streamflow proxy as per approach described in [50].

2.2. DNA extraction, amplicon, and shotgun metagenomic sequencing

To capture a broad range of microbial taxa, including bacteria and fungi, 500 mL of each surface water sample underwent sequential filtration. Samples were first passed through $0.7 \mu\text{m}$ borosilicate glass fibre filters (Thermo Fisher, Ottawa, ON) to retain larger particles and fungal spores, followed by filtration through sterile $0.22 \mu\text{m}$ nitrocellulose membranes (Millipore, Billerica, MA, USA) to collect smaller microbial cells such as bacteria and archaea. DNA was extracted from both filter types using the DNeasy PowerSoil® Kit (Qiagen), and quantified with a Qubit 3.0 fluorometer (Invitrogen). DNA integrity was verified by 1% agarose gel electrophoresis, and samples were stored at -80°C until sequencing.

For amplicon sequencing, DNA from both filter types was used to amplify the V4–V5 region of the 16S rRNA gene with primers 515F-Y and 926R. PCR was conducted using Qiagen HotStar MasterMix under standard cycling conditions (95°C for 3 min; 25 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 30 s; final extension at 72°C for 5 min). Amplicons were purified (NucleoMag beads, Macherey-Nagel), pooled in equimolar concentrations, and sequenced on an Illumina MiSeq using a

500-cycle v2 kit (2 × 250bp reads). To ensure consistency with the shotgun metagenomic dataset, only amplicon data derived from the 0.22 µm filters were used for downstream analysis.

Of the 499 total water samples collected, a subset of 270 samples was selected for shotgun metagenomic sequencing. Selection was based on multiple criteria to ensure spatial, temporal, and land use representation across the five-year period. Priority was given to samples with high-quality and high-concentration DNA, as verified by Qubit quantification and agarose gel electrophoresis. Additionally, representative samples were chosen to capture seasonal variability, land use gradients (agriculture, mixed-use, forest), and key hydrological events (e.g., spring runoff, post-harvest). Resource limitations in sequencing capacity and cost also necessitated a focused yet informative subset for deep metagenomic profiling, while ensuring broad coverage of microbial functional potential. Shotgun metagenomic sequencing was performed on DNA from the 0.22 µm filters. Libraries were prepared using the Illumina Nextera XT kit. Samples from 2016–2019 were sequenced on an Illumina HiSeq 2500 (2 × 150 bp) at NRC Montreal, while 2020–2021 samples were sequenced on the NovaSeq 6000 at NRC Saskatoon. No mock communities were included for comparison the recovery between the two instruments. All DNA was stored at –80 °C before and after library preparation to maintain quality.

While mock communities were not included to benchmark recovery efficiency, strict adherence to standardized DNA extraction and library preparation protocols was maintained across all samples to ensure methodological consistency. Consequently, the comparison between platforms relies on the detection of robust ecological patterns within complex environmental matrices rather than synthetic standards.

2.3. Data processing for 16S rRNA amplicon sequencing data

Adapter and primer sequences were trimmed from raw FASTQ files using Atria (v4.1.1) [51] and Cutadapt v4.1 [52]. Paired-end reads were processed in QIIME2 [53] using the DADA2 plugin (v1.14) [54] for quality filtering, denoising, chimera removal, and amplicon sequence variant (ASV) inference. Forward and reverse reads were truncated at 229 nt and 187 nt, respectively, based on quality profiles. Taxonomic classification was conducted using the q2-feature-classifier plugin in QIIME2 [53], with a Naive Bayes classifier trained on the Greengenes2 reference database [55]. A minimum bootstrap confidence threshold of 70% was applied for taxonomic assignments.

2.4. Data processing for shotgun sequencing data

2.4.1. Short read-based taxonomic profiling and detection of antimicrobial resistance and virulence factor genes

Raw reads from shotgun sequencing were first subjected to adapter removal using BBduk (<https://sourceforge.net/projects/bbmap/>), followed by quality trimming with Trimmomatic v0.39 [56] using the parameters: LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, MINLEN:36, and HEADCROP:6. The quality and quantity of trimmed reads were assessed using FastQC v0.12.1 [57], and results were summarized with MultiQC v1.2 [58].

Taxonomic profiling was performed using Kraken2 v2.1.3 [59] with the May 17, 2021 version of the Kraken2 database, followed by Bracken v2.9 [60] to estimate relative abundances at species, genus, family, and order levels. A pre-built Bracken database, based on the NCBI nucleotide (nt) database dated May 2, 2023, was used with a 150 bp k-mer specification. Post-processing was automated with a Snakemake v7.32.4 [61] pipeline, which parsed and merged Bracken reports, retaining sample ID, taxonomy name, and estimated read counts (new_est_reads column) for downstream analyses. This classification workflow was selected for its computational efficiency on large datasets and its demonstrated high accuracy and sensitivity for pathogen detection and abundance estimation in complex metagenomic communities [62].

Short read-based ARG and VF detection were performed by mapping

trimmed reads to the MEGARes v3.0 [63] and VFDB v12-06-2023 (complete set) [64] databases using KMA v1.3.23 [65]. Presence or absence of ARGs, as determined by KMA, was used for downstream analyses.

2.4.2. Assembly of metagenome assembled genomes (MAGs) and detection of their antimicrobial resistance genes and virulence factors

The trimmed FASTQ files were subjected to co-assembly using MEGAHIT v1.2.9, with a minimum contig length of 1 kb [66,67]. Gene prediction on assembled contigs was performed with Prodigal v2.6.3 [68]. Quality-trimmed reads were then mapped back to the assembled contigs using Bowtie2 v2.4.5 [69], and alignments were stored in BAM files with Samtools v1.7 [70].

Sample-specific contig binning was executed based on tetranucleotide frequencies using CONCOCT v1.1.0 [71] and MetaBAT2 v2.12.1 [72]. Consensus bins were generated with DAS Tool v1.1.3 [73]. MAG completeness and contamination were assessed using CheckM2 v1.0.1 [74], retaining bins with >50% completeness and <10% contamination per MIMAG standards [75], resulting in 41 medium- to high-quality MAGs.

Taxonomic classification was conducted with GTDB-Tk v2.3.2 [76] against GTDB Release R214 [77]. Unclassified bins lacking bacterial or archaeal marker genes were further identified using BLAST v2.15.0 [78] against the NCBI nucleotide database (July 2023), accepting matches with >90% identity and e-value <1e-5.

Antimicrobial resistance genes (ARGs) and virulence factors (VFs) within MAGs were detected by BLAST against MEGARes v3.0 [63] and VFDB v12-06-2023 [64], considering hits with >80% identity. MEGARes hits requiring further SNP analysis to confirm resistance were excluded to allow comparison with read-based ARG detection.

2.5. Curated reference databases for fecal-associated and opportunistic pathogen taxa

To characterize fecal-associated bacterial communities, we curated the *Fecal-Associated Microbiome Reference (FAMR)* by integrating genus-level taxonomic information from two established sources: the Animal Gut Microbiome Database (AMDB) [79] and the Annotated Database of Domestic Animal Gut Microbiota (ADDAGMA) [80]. AMDB provides a broad collection of bacterial taxa and samples across diverse animal species, while ADDAGMA focuses specifically on gut microbiomes of four key domestic animals (cattle, horse, pig, and chicken). This combined database enabled the identification of fecal-associated bacterial genera from both 16S rRNA and shotgun metagenomic sequencing datasets.

For opportunistic pathogen detection, we developed an *Opportunistic and Zoonotic Pathogen Reference (OZPR)* curated a separate database by consolidating genus-level taxonomic information from several authoritative sources. The core reference was a list of 1,513 bacterial species with confirmed pathogenicity to humans [81], which was expanded by incorporating species from the NCBI Pathogen Detection database (97 species) [82], the PATRIC pathogen database (<https://www.bv-brc.org/>) [83], and the Voluntary 2025 U.S. National Animal Health Reporting System (NAHRS) Reportable Diseases, Infections, and Infestations List [84]. We also included taxa from the WHO 2024 bacterial priority pathogen list [85] and the CDC zoonotic pathogen list (<https://www.cdc.gov/healthy-pets/diseases/index.html>) [86]. The final pathogen reference list comprised 336 genera and was used to extract and compare pathogen-associated taxa from the sequencing data. Because taxonomic resolution from sequencing data is typically limited to the genus level, the OZPR reference includes genera that encompass both pathogenic and commensal species; therefore, detected genera should be interpreted as potentially pathogen-associated rather than definitively pathogenic.

2.6. Statistical analysis

All statistical analyses were performed in R (ver. 4.5.1) [87]. Alpha diversity indices were used to quantify within-sample ASV diversity of fecal-associated bacterial communities. The Shannon Index (H), Simpson Index (D), and Chao1 Index were calculated using the *vegan* (v2.6-6.1) [88] and *biodiversityR* (v2.16-1) [89] packages. Shannon-based True Diversity (Shannon-TD = $\exp(H)$) and Simpson-based True Diversity (Simpson-TD = $1/(1-D)$) were computed following the approach described by Jost [90]. Normality of alpha diversity distributions was assessed using the *shapiro.test* function, and Box-Cox transformations were applied when needed using the *boxcox* function in the MASS package (v7.3-65) [91]. General linear mixed-effects models (GLMMs) were implemented via the *glmmTMB* function in *glmmTMB* package (v1.1.12) [92], to evaluate the effects of site type, year, and their interactions on alpha diversity and physicochemical properties. Site type and year were treated as fixed effects, block as a random effect, and repeated measures were nested within year-week. Post hoc comparisons were performed using the *emmeans* package (v1.10.2) [93] with Bonferroni-adjusted P-values. Environmental drivers of diversity were identified via random forest analysis using the *randomForest* package (v4.7-1.1) [94]. Random Forest regression was selected for its ability to model complex, nonlinear relationships and interactions among environmental variables without assuming normality or linearity, which are common limitations in ecological datasets [95]. Finally, Spearman's rank-order correlation between Shannon-TD and physicochemical parameters were evaluated using *rcorr.adjust* function in *RcmdrMisc* (v2.9-1) [96].

To characterize the temporal stability of bacterial communities, we identified core taxa and conditionally rare taxa (CRT) across all sampling sites using methods described by Shade and Stopnisek [97] and Shade et al. [98], with detailed procedures described by Shi et al. [40]. In brief, core ASVs were selected based on abundance–occupancy distributions and their contributions to overall community heterogeneity. CRTs were identified by assessing ASVs with $\geq 0.1\%$ abundance that exhibited bimodal abundance distributions over time, using a coefficient of bimodality (b) threshold of 0.90 [98]. The percent contributions of core taxa and CRT to overall community beta-diversity were calculated separately for agricultural drainage ditch sites and mixed-use sites, but not for the single forested site (SN_24). To evaluate the effect of land use on the contributions of core taxa and CRT, linear mixed-effects models were used with land use class as a fixed factor and sampling date as a repeated measure. Additionally, to identify key environmental drivers of temporal variation in Shannon-based true diversity, core and CRT contributions to community dissimilarity, variable importance was ranked using the *randomForest* function (*randomForest* package v4.7–1.1) [94].

The *glmmTMB* models were used to assess the effect of site_type on robust centered log ratio (rCLR)-transformed abundances of the 10 most abundant genera recovered by both sequencing methods, with site_type included as a fixed factor and Year–Month (YW) as a random effect [92]. In addition, *glmmTMB* was applied to evaluate the correlation between amplicon and shotgun sequencing data, using the amplicon abundance of a specific genus as the response variable (Y) and the corresponding shotgun abundance as the predictor (X, fixed factor), while accounting for site_type and Year–Month as random effects. To investigate temporal pattern similarity between amplicon and shotgun sequencing data, dynamic time warping (DTW) distances were calculated using the *dtw* package [99] after normalization by genus and site type.

Heatmaps were generated using the *heatmap* package (v1.0.13) [100] to visualize the relative abundance patterns of fecal-associated taxa known to carry AMR carriers and AMR gene presence. The data were row-scaled (z-score normalized), so that for each taxon (row), the mean relative abundance across samples was centered at zero and values were scaled by the standard deviation [101]. Hierarchical clustering was performed using Euclidean distance and complete linkage for both rows and columns [100]. To evaluate differences in AMR gene presence

across site types, a negative binomial generalized linear model was fitted using the *glm.nb* function in the MASS package [102].

2.7. Availability of bioinformatic scripts

The Bioinformatic pipeline and data analysis scripts are available here: <https://github.com/AAFC-Bioinfo-AAC/snra-metagenomic-statistics-amrvf-ordc/tree/main>.

3. Results

3.1. Relative strength of amplicon and shotgun metagenomics sequencing in detecting fecal-associated bacterial communities

A total of 499 freshwater samples were analyzed by 16S rRNA gene amplicon sequencing, yielding over 21 million high-quality reads (152,076 amplicon sequence variants, ASVs). A subset of 270 (54.1%) was also subjected to shotgun metagenomic sequencing (average 16.6 million reads per sample).

To compare recovery of fecal-associated bacterial taxa, we queried a curated database (Section 2.5) against the paired 270 samples. We identified 730 genera (26 phyla) in the amplicon dataset and 878 genera (30 phyla) in the shotgun dataset. Of these, 14 phyla and 443 genera were shared between the two datasets (Fig. 2A, B). Dominant genera, like *Flavobacterium*, *Limnohabitans*, *Polynucleobacter*, and *Rhodoluna* (Fig. 2C, D), were consistent across methods. However, abundance distributions differed significantly: while non-dominant genera accounted for less than 20% of the relative abundance in amplicon profiles, they contributed over 35% in the shotgun dataset (Fig. 2C, D), indicating that shotgun sequencing recovered a larger proportion of the lower-abundance community fraction in this study.

3.2. Fecal-associated bacterial community diversity and pathogen detection across land-use types

3.2.1. Ecological patterns of fecal-associated microbiota based on amplicon sequencing (n = 499)

Using the 16S rRNA gene amplicon dataset (n = 499), we characterized fecal-associated bacterial communities across different land use classes to help identify potential health risk factors. Across 499 samples, we detected 19,492 ASVs representing 947 fecal-associated genera. Alpha-diversity matrices (Simpson-based true diversity, Shannon-based true diversity, and Chao1 richness) varied significantly with upstream land use ($P < 0.05$) (Fig. 3A, supplementary Fig. S1). Agricultural drainage ditch sites consistently exhibited the highest diversity, followed by mixed-use and forest sites, with Shannon-based true diversity significantly higher in mixed-use than forested sites (Fig. 3A). Temporal dynamics revealed distinct seasonal peaks in Shannon-based true diversity at the beginning and end of the sampling seasons, particularly in agriculturally dominated ditch sites, except for years with atypical sampling (2016 and 2017) (Fig. 3B). Random Forest (RF) regression identified water temperature as the primary driver of temporal variation in agricultural and mixed-use sites, while total organic carbon (TOC) and nitrate significantly contributed to predictability for agricultural ditches (Supplementary Fig. S2). RF models explained a substantial proportion of diversity variation in agricultural ($R^2 = 0.54$) and mixed-use ($R^2 = 0.33$) sites but had negligible power at forested sites (Fig. 3C).

We identified 80 core fecal-associated ASVs (70.6% of total abundance) and 1,383 conditionally rare taxa/ASVs (CRTs; 8.6% of abundance), totalling 7.5% of total ASVs across all samples. The remaining (83%) ASVs were predominantly rare (<10 reads). The core microbiota was dominated by *Flavobacterium*, *Limnohabitans*, *Polynucleobacter*, and *Rhodoluna*, with *Flavobacterium* including many known zoonotic pathogens. CRT was dominated by *Flavobacterium*, *Legionella*, *Haematospirillum*, *Mycobacterium*, and *Acinetobacter*, all are pathogen-associated genera (See Section 2.5), highlighting the potential public

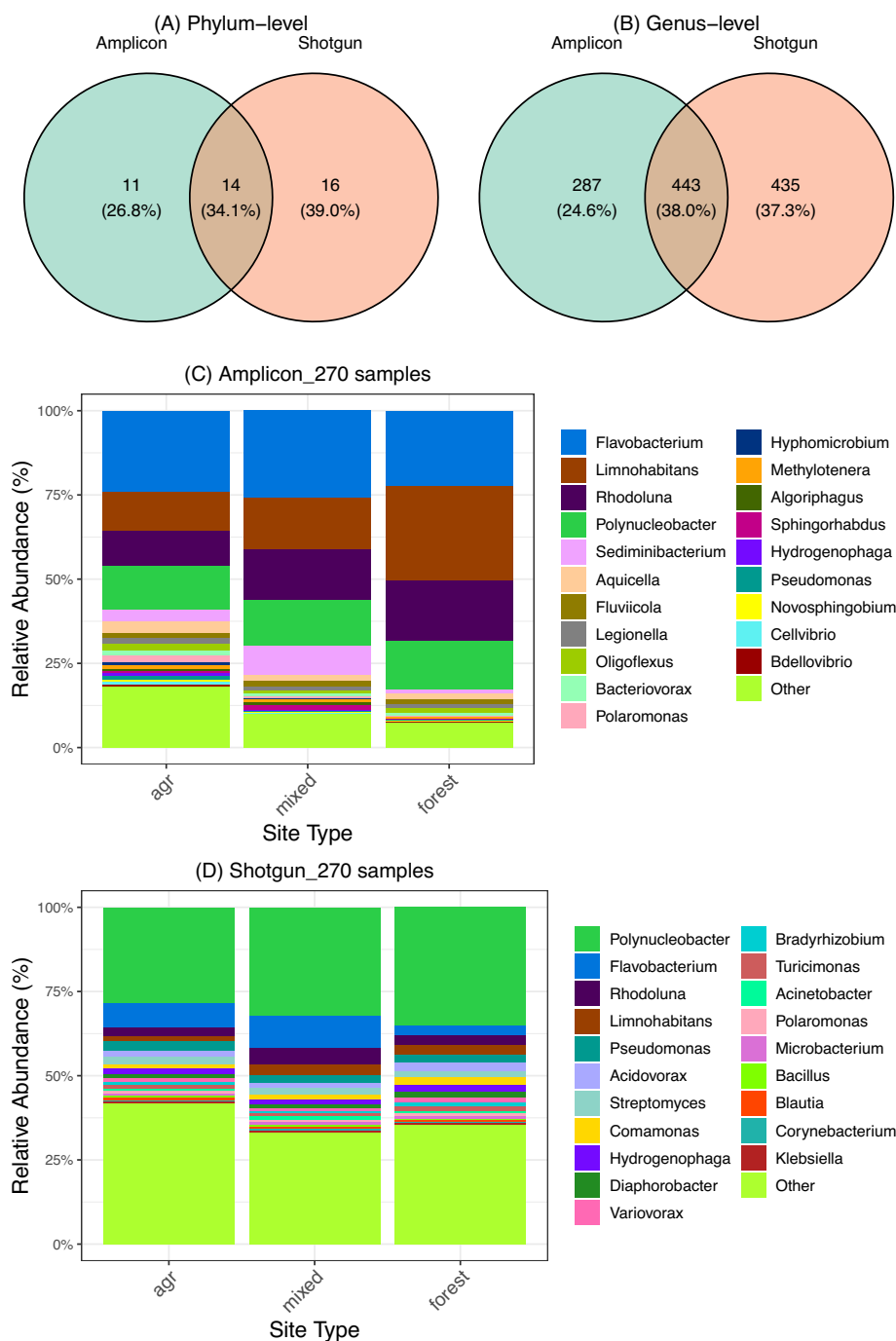


Fig. 2. Comparison of fecal-associated bacterial communities identified by amplicon and shotgun metagenomic sequencing. (A-B) Venn diagrams showing overlap of taxa detected at the phylum (A) and genus (B) levels. (C-D) Relative abundance of the 20 most abundant genera in the amplicon (C) and shotgun (D) datasets.

health relevance of this transient but dynamic fraction (supplementary Fig. S3).

Beta diversity analysis indicated that agricultural ditches had significantly lower core but higher CRT contributions to Bray–Curtis dissimilarity compared to mixed-use sites ($P < 0.001$) (Fig. 3D). Environmental drivers of these fractions differed across different land use classes. RF models explained more variance in agricultural systems (core $R^2 = 0.58$; CRT $R^2 = 0.37$) than mixed-use sites (core $R^2 = 0.23$; CRT $R^2 = 0.08$), with dissolved oxygen and TOC emerging as primary predictors (Supplementary Fig. S4). Notably, CRT contributions to fecal-community dynamics in agricultural drainage ditches increased from 6.1% in 2017, 25.9% in 2018, to 20.8% in 2019 ($P < 0.05$) (Fig. 3D), a

trend absent at mixed-use sites. This increase coincided with ditch management (vegetation brushing and dredging); managed sites (SN_20, SN_21) exhibited reduced core contribution (−49.0%) and elevated CRT contribution (+27.9%) in 2019 compared to unmanaged ditches (SN_18, SN_19) (Supplementary Fig. S5). Such observations suggest, either directly or indirectly, physical disturbance and vegetation loss intensified bacterial community turnover in the managed drainage systems.

These pronounced shifts in diversity and CRT contributions within agricultural ditches establish an ecological baseline for evaluating the comparative resolution of shotgun metagenomics in Section 3.2.2.

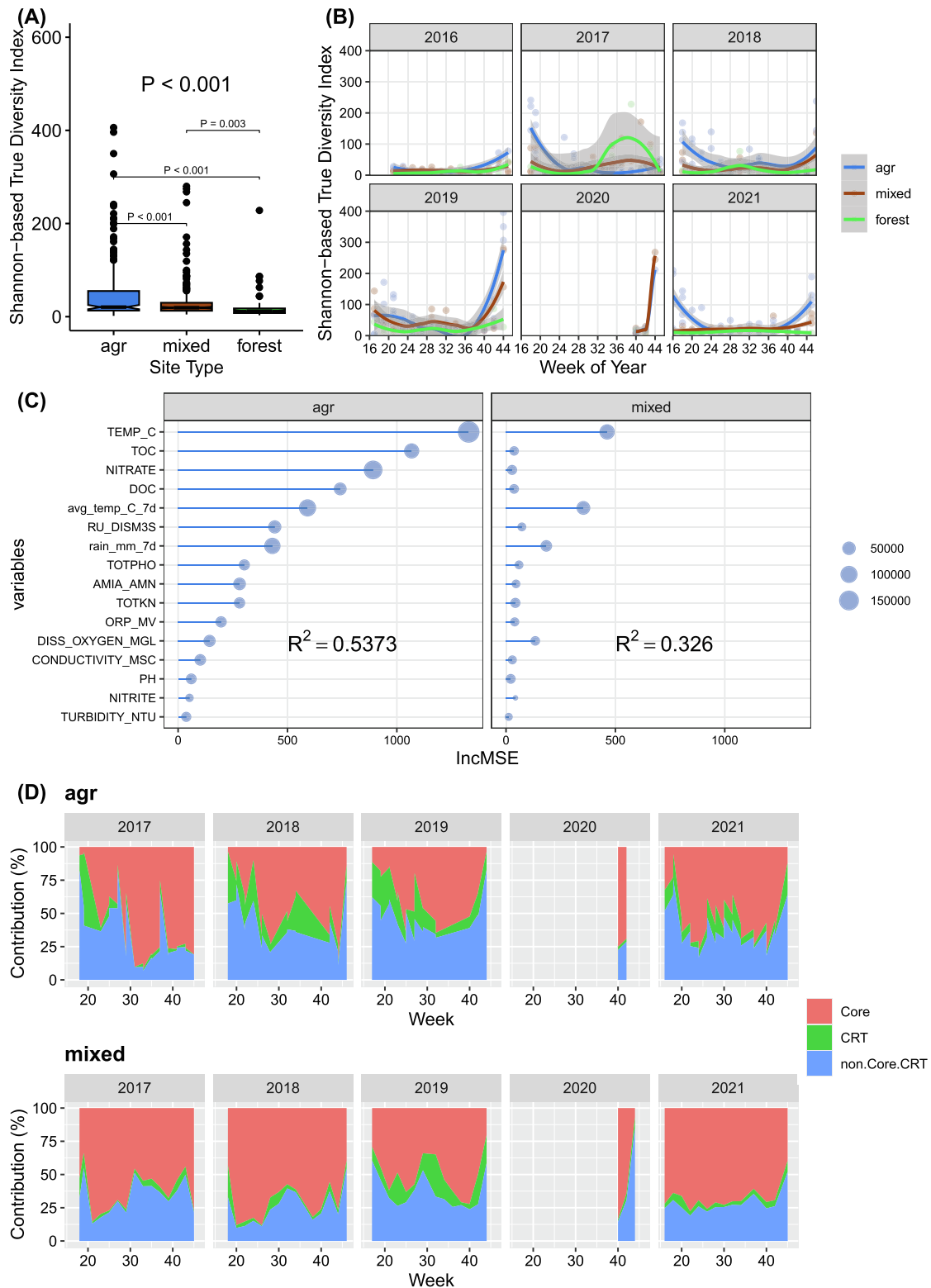


Fig. 3. Diversity and dynamics of fecal-associated bacterial communities across land use types. (A) Mean Shannon-Wiener true diversity (SW-TD) index across agricultural drainage ditches, mixed-use, and forested sites. Differences among land-use types were assessed using generalized linear mixed-effects models with Bonferroni-adjusted P -values for pairwise contrasts. (B) Seasonal SW-TD dynamics fitted by local polynomial regression (2016–2021). (C) Importance of environmental variables explaining alpha diversity, ranked by random-forest models for agricultural drainage ditches and mixed-use sites. The coefficient of determination (R^2) for each land-use class is shown. (D) Relative contributions of core and conditionally rare taxa (CRTs) to Bray–Curtis dissimilarity (%) from 2016–2021. Red areas represent core taxa; green areas represent CRTs. Fewer samples were collected in 2020 due to pandemic restrictions. Descriptions of the measured water physicochemical variables are provided in Supplementary Table S1.

3.2.2. Cross-validation of fecal-associated microbiota using amplicon and shotgun metagenomics (n = 270)

Using a subset of 270 samples with paired amplicon and shotgun metagenomics datasets, we evaluated cross-method consistency, temporal concordance, and the relative sensitivity for low-abundance or taxonomically complex taxa.

Comparison of the 270 paired samples revealed that while 138 pathogen-associated genera were shared (accounting for >88% of pathogen abundance in both datasets) (supplementary Fig. S6), detection profiles differed for specific taxa. Shotgun metagenomics identified 230 pathogen-associated genera (29.8% of fecal-associated taxa) compared to 157 in the amplicon dataset (37.1%). Among the 10 most abundant genera that were detected by both methods, shotgun sequencing consistently recovered all targets (100% frequency), whereas amplicon detection frequencies varied markedly (e.g., *Corynebacterium* was detected in only 4.4% of samples) (Supplementary Fig. S7).

Both methods detected 18 pathogen-associated bacterial genera listed in the latest CDC zoonotic disease database and 2024 WHO bacterial priority list (Table 1). However, six high-priority genera (*Salmonella*, *Shigella*, *Capnocytophaga*, *Haemophilus*, *Chlamydia*, and *Streptobacillus*), collectively representing 0.39% of the fecal-associated taxa, were detected only by shotgun sequencing (Table 1). Similarly, *Listeria* presence was corroborated by shotgun sequencing (100% samples) and qPCR (86.9% samples, unpublished data) but was undetected in the amplicon dataset (Supplementary Fig. S8), likely due to primer mismatch (supplementary Table S2).

Ecological resolution also varied between methods. Amplicon profiles showed significant differentiation among different land use classes, with *Clostridium*, *Massilia*, *Pseudomonas*, *Rickettsia*, and *Sphingomonas* significantly enriched in agricultural ditches ($P < 0.001$) (Fig. 4). Furthermore, *Brucella*, *Campylobacter*, and *Neisseria* detection were restricted to agricultural ditches in the amplicon dataset, while *Bartonella* and *Yersinia* were specific to mixed-use sites. *Erysipelothrix* and *Enterococcus* were detected in both agricultural and mixed-use sites but not in the forested reference site (Table 1). In contrast, shotgun profiling

detected these genera across most samples (>90% frequency) with no significant abundance differences between site types (Table 1).

Despite these spatial discrepancies, temporal dynamics showed broad concordance. GLMMs revealed strong positive correlations (slope = 0.9–1.2; $P < 0.001$) for *Flavobacterium* and *Legionella*, as well as moderate correlations (slope < 0.5) for *Sphingomonas* ($P < 0.001$), *Pseudomonas* ($P = 0.001$), and *Massilia* ($P = 0.004$). Dynamic time warping (DTW) confirmed this synchrony, particularly for *Acinetobacter*, which peaked synchronously in October 2019 at mixed-use sites in both datasets (Fig. 4; Supplementary Fig. S9 and S10).

Overall, cross-validation confirmed high concordance for dominant taxa while highlighting a critical trade-off: amplicon profiling resolved stronger ecological differentiation along land-use gradients, whereas shotgun metagenomics provided superior coverage of rare, clinically relevant genera more relevant for functional One Health surveillance.

3.3. Prevalence and distribution of resistome and virulome (n = 270)

To characterize reservoirs of antimicrobial resistance and virulence, we analyzed 270 water samples using both short-read metagenomics and metagenome-assembled genomes (MAGs).

Short-read metagenomic profiling detected 27 distinct ARGs in 12 samples and 14 VFs in 15 samples (Table 2). Most genes were rare, of which, eight ARGs and one VF (*csrA*) occurred in multiple samples, while the remainder appeared only once (Table 2; Fig. 5). Spatial distributions of these genes were markedly distinct: mixed-use sites were hotspots for resistance, harboring 24 of the 27 detected ARGs (19 specifically at site SN_5 during 2018–2019), whereas agricultural ditches were reservoirs for virulence, containing 13 of the 14 identified VFs (Fig. 5). Consequently, ARG richness was significantly higher at mixed-use sites compared to agricultural ditches (negative binomial model, $P < 0.05$).

Genome-resolved analysis of 41 medium- to high-quality MAGs provided complementary host-linkage data (Supplementary Table S3). Five bacterial MAGs carried ARGs (Table 2) including *vmeF* (RND efflux, *Algoriphagus*), *srmB* (ABC efflux, CAJAU01), *oqxB* (multidrug

Table 1

Genera containing CDC- and WHO-listed pathogens detected by amplicon and shotgun metagenomic sequencing across different land-use types.

Genera	amplicon						shotgun					
	Relative abundance (%)			Detection frequency			Relative abundance (%)			Detection frequency		
	agr	mixed	forest	agr	mixed	forest	agr	mixed	forest	agr	mixed	forest
<i>Pseudomonas</i>	0.993	0.282	0.190	0.602	0.418	0.432	2.810	2.100	2.300	1	1	1
<i>Rickettsia</i>	0.246	0.056	0.171	0.585	0.236	0.595	0.079	0.048	0.064	1	1	1
<i>Acinetobacter</i>	0.203	0.461	0.015	0.333	0.345	0.216	0.762	1.100	0.719	1	1	1
<i>Mycobacterium</i>	0.148	0.166	0.177	0.520	0.573	0.378	0.319	0.353	0.278	1	1	1
<i>Coxiella</i>	0.086	0.028	0.014	0.390	0.155	0.270	0.060	0.032	0.057	1	1	1
<i>Aeromonas</i>	0.056	0.015	0.016	0.358	0.227	0.270	0.274	0.174	0.196	1	1	1
<i>Leptospira</i>	0.023	0.092	0.024	0.236	0.382	0.162	0.082	0.054	0.069	1	1	1
<i>Escherichia</i>	0.012	0.001	0.006	0.065	0.018	0.081	0.281	0.246	0.218	1	1	1
<i>Erysipelothrix</i>	0.008	0.008	0.000	0.073	0.055	0.000	0.035	0.027	0.023	1	1	1
<i>Staphylococcus</i>	0.005	0.000	0.008	0.024	0.009	0.054	0.489	0.330	0.326	1	1	1
<i>Enterococcus</i>	0.002	0.001	0.000	0.016	0.018	0.000	0.169	0.136	0.121	1	1	1
<i>Brucella</i>	0.001	0.000	0.000	0.016	0.000	0.000	0.073	0.073	0.074	1	1	1
<i>Neisseria</i>	0.001	0.000	0.000	0.016	0.000	0.000	0.191	0.162	0.228	1	1	1
<i>Streptococcus</i>	0.001	0.001	0.000	0.024	0.027	0.000	0.313	0.238	0.245	1	1	1
<i>Campylobacter</i>	0.0002	0.000	0.000	0.016	0.000	0.000	0.254	0.173	0.197	1	1	1
<i>Borrelia</i>	0.000	0.000	0.0002	0.008	0.000	0.027	0.037	0.023	0.025	1	1	1
<i>Bartonella</i>	0.000	0.001	0.000	0.0000	0.0091	0.0000	0.110	0.102	0.101	1	1	1
<i>Yersinia</i>	0.000	0.0001	0.000	0.0000	0.0091	0.0000	0.097	0.080	0.094	1	1	1
<i>Salmonella</i>	0.000	0.000	0.000	0.0000	0.0000	0.0000	0.207	0.193	0.126	1	1	1
<i>Capnocytophaga</i>	0.000	0.000	0.000	0.0000	0.0000	0.0000	0.081	0.055	0.074	1	1	1
<i>Haemophilus</i>	0.000	0.000	0.000	0.0000	0.0000	0.0000	0.076	0.072	0.061	1	1	1
<i>Chlamydia</i>	0.000	0.000	0.000	0.0000	0.0000	0.0000	0.065	0.060	0.033	1	1	1
<i>Shigella</i>	0.000	0.000	0.000	0.0000	0.0000	0.0000	0.017	0.016	0.032	0.984	0.973	0.946
<i>Streptobacillus</i>	0.000	0.000	0.000	0.0000	0.0000	0.0000	0.005	0.003	0.004	1	1	1

Relative abundance (%) represents the proportion of each genus within the total fecal-associated bacterial community. Detection frequency indicates the proportion of samples in which each genus was detected. agr, agricultural drainage ditch sites; mixed, mixed-use sites; forest, forested reference site.

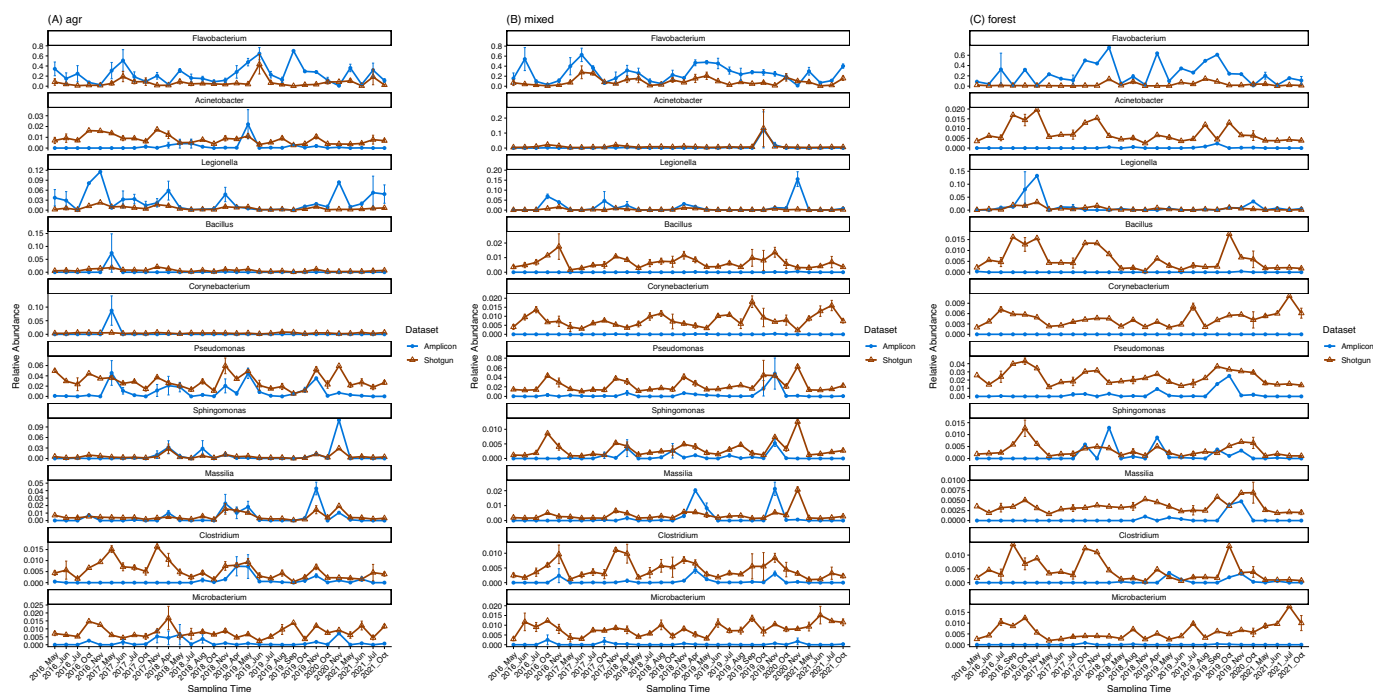


Fig. 4. Temporal variation of ten shared genera among the twenty most abundant taxa identified in both sequencing approaches. Mean relative abundances are shown over time for (A) agricultural drainage ditches, (B) mixed-use sites, and (C) forested sites. Error bars represent standard errors.

resistance, *Zwartia*), *ileS* (mupirocin resistance, JALRLI01), and *hpcopA* (copper resistance). Notably, these MAG-derived ARGs did not overlap with those found in short-read profiles, highlighting the independent value of each assembly method. Additionally, five MAGs encoded VF genes (Table 2), including loci encoding a multispecies isoprenyl transferase (in MAG 209), response regulator (in MAG 369), a metal-dependent transcriptional regulator (*kasB*; MAG 530), a polysaccharide biosynthesis enzyme *rfbA* (MAG 648), and ATP-binding proteins (MAG 665). These genes represent potential virulence determinants involved in stress response, cell-surface modification, and environmental persistence rather than acute pathogenicity [103].

The detected ARGs spanned multiple antimicrobial classes. According to Canadian categorization of antimicrobial importance, three genes, *oxa* (SN_5, 2019), *imp* (SN_6, 2020), and *sme* (SN_19, 2017), include alleles conferring resistance to Category I antimicrobials of very high importance, including beta-lactams and carbapenems (Fig. 5B). Fourteen ARGs confer resistance to Category II antimicrobials of high importance, including aminoglycosides [*ant(3'')-I*, *ant(6)*, *ant(9)*, *aph(3'')-I*, *aph(6)*], macrolides (*ermF*, *mphG*, *mefA*, *mefB*, *mefC*), trimethoprim (*dhfrB*), and lincosamides (*lncC*, *lsa*, *lsae*). Ten additional ARGs conferred resistance to Category III antimicrobials of moderate importance, primarily tetracyclines (*tet39*, *tet44*, *tetM*, *tetQ*, *tetT*, *tetW*), chloramphenicol (*cat*), and mupirocin (*ileS*) (Table 2). These genes represent diverse resistance mechanisms, like enzymatic drug modification, target protection, and efflux pumps (Table 2), highlighting the functional breadth of AMR in watershed environments.

Consistent with these patterns, known ARG-carrying bacteria such as *Arcobacter* spp. and *Acinetobacter* spp., were abundant at mixed-use SN_5 across multiple years (Fig. 5C). In addition to canonical ARGs, several metal- and stress-response genes were detected, including components of the mercury resistance operon (*merC*, *merF*, *merP*, *merT*), oxidative-stress enzymes (*sodA*, *sodB*), and the DNA-repair helicase *ruvB*.

Collectively, these results reveal spatially distinct reservoirs: mixed-use sites functioned as ARG hotspots, whereas agricultural ditches served as VF reservoirs. Although overall prevalence was low, the functional diversity captured by combining short-read and MAG analyses reflects complex, site-specific selective pressures. These findings

underscore the necessity of integrating gene-level and genome-resolved perspectives to accurately and more holistically assess the ecological persistence and dissemination potential of resistance and virulence traits.

4. Discussion

This five-year longitudinal study illustrates how integrated metagenomic surveillance can characterize microbial risk factors in agriculturally influenced watersheds. By combining the scalability of 16S rRNA gene amplicon sequencing with the functional resolution of shotgun metagenomics, we established a complementary framework for tracking fecal-associated taxa, ARGs, and VFs. Beyond method performance, we identified ecological sentinels, specifically conditionally rare taxa (CRTs) and pathogen-associated genera, that link land-use pressures to potential health risk-factors. Collectively, these findings underscore the value of tiered monitoring to characterize environmental exposure pathways within a One Health continuum.

4.1. Complementarity of amplicon and shotgun metagenomics in one health surveillance

Evaluating sequencing trade-offs is critical for optimizing One Health surveillance frameworks. While 16S amplicon sequencing offers the scalability required for long-term ecological monitoring [104], shotgun metagenomics offers broader taxonomic and functional resolution necessary to track antimicrobial resistance and virulence potential [26,42,105].

In this study, both approaches successfully captured dominant fecal-associated genera, revealing strong temporal concordance for key taxa. Across the five-year dataset, each method detected more than 700 fecal-associated genera, with substantial taxonomic overlap (> 50%) within the core community. These results contrast with earlier reports suggesting poor recovery in shallow shotgun datasets [105,106], demonstrating that optimized wet-lab workflow and bioinformatics pipelines can effectively profile pathogen-associated communities in complex environmental matrices.

Table 2

Antimicrobial resistance genes (ARG) and virulence factors (VFs) detected by short-read and metagenome-assembled genome (MAG) analyses from shotgun metagenomic data.

Category	Gene name	Associated drug class/function	AMR relevance/ VFs	MAGs #	Taxonomy
I	<i>imp</i>	beta-lactamase and carbapenem resistance genes	AMR	-	-
I	<i>oxa</i>	beta-lactamase and carbapenem resistance genes	AMR	-	-
I	<i>sme</i>	beta-lactamase and carbapenem resistance genes	AMR	-	-
II	<i>ant3-Dprime</i>	aminoglycoside resistance genes	AMR	-	-
II	<i>ant6</i>	aminoglycoside nucleotidyltransferase gene	AMR	-	-
II	<i>ant9</i>	aminoglycoside resistance genes	AMR	-	-
II	<i>aph3-Dprime</i>	aminoglycoside resistance genes	AMR	-	-
II	<i>aph6</i>	aminoglycoside resistance genes	AMR	-	-
II	<i>dftrB</i>	plasmid-borne trimethoprim-resistant dihydrofolate reductase	AMR	-	-
II	<i>ermF</i>	erythromycin resistance genes	AMR	-	-
II	<i>lnuC</i>	transposon-mediated lincosamide nucleotidyltransferase	AMR	-	-
II	<i>lsa</i>	resistance to numerous antibiotics, including lincosamides, pleuromutilin, and streptogramins	AMR	-	-
II	<i>lsae</i>	resistance to numerous antibiotics, including lincosamides, pleuromutilin, and streptogramins	AMR	-	-
II	<i>mphG</i>	macrolide resistance gene	AMR	-	-
II	<i>mefA</i>	macrolide efflux pump and transporter genes	AMR	-	-
II	<i>mefB</i>	macrolide efflux pump and transporter genes	AMR	-	-
II	<i>mefC</i>	macrolide efflux pump and transporter genes	AMR	-	-
III	<i>cat</i>	chloramphenicol acetyltransferase gene	AMR	-	-
III	<i>tet39</i>	tetracycline resistance genes	AMR	-	-
III	<i>tet44</i>	tetracycline resistance gene	AMR	-	-
III	<i>tetM</i>	tetracycline resistance gene	AMR	-	-
III	<i>tetQ</i>	tetracycline resistance genes	AMR	-	-
III	<i>tetT</i>	tetracycline resistance genes	AMR	-	-
III	<i>tetW</i>	tetracycline resistance genes	AMR	-	-
not class-specific	<i>emrAsm</i>	efflux pump and transporter genes	AMR	-	-
not class-specific	<i>emrBsm</i>	efflux pump and transporter genes	AMR	-	-
not class-specific	<i>mexF</i>	Part of RND efflux system in <i>Pseudomonas</i>	AMR	-	-
not class-specific	<i>sodA</i>	multidrug resistance gene	Indirect	-	-
not class-specific	<i>sodB</i>	resistance to heavy metals and antimicrobial resistance	Indirect	-	-
not class-specific	<i>merC</i>	mercury resistance genes	co-selection link	-	-
not class-specific	<i>merF</i>	mercury resistance genes	co-selection link	-	-
not class-specific	<i>merP</i>	mercury resistance genes	co-selection link	-	-
not class-specific	<i>merT</i>	mercury resistance gene	co-selection link	-	-
not class-specific	<i>ruvB</i>	tellurium resistance	co-selection link	-	-
	VFG000863	heat-stable enterotoxin 1	VF		
	VFG001225	twitching motility protein PilG	VF		
	VFG001256	flagellar motor switch protein FliM	VF		
	VFG001855	Hsp60, 60K heat shock protein HtpB	VF		
	VFG002070	type VI secretion system tubule-forming protein VipB	VF		
	VFG010906	carbon storage regulator CsrA	VF		
	VFG013948	type 4 fimbrial biogenesis protein PilZ	VF		
	VFG038176	outer membrane protein OmpA	VF		
	VFG038340	Type III secretion system protein, LcrG-like	VF		
	VFG038346	Type III secretion system gatekeeper subunit	VF		
	VFG038659	flagellar motor switch protein FliN	VF		
	VFG038663	flagellar biosynthesis protein, FliO	VF		
	VFG038876	lateral flagellar basal-body rod protein LfgG	VF		
	VFG043016	purine-binding chemotaxis protein	VF		
MAG-detection					
II	<i>srnB</i>	Multi-drug, macrolide, spiramycin resistance gene	AMR	775	CAJAUT01 (Patescibacteria)
II	<i>oqxB</i>	Drug_and_biocide_resistance	AMR	616	Zwartia (Pseudomonadota)
not class-specific	<i>vmeF</i>	Multi-biocide RND efflux pump	AMR	17	Algoriphagus (Bacteroidota)
not class-specific	<i>ileS</i>	Mupirocin	AMR	530	JALRLI01 (Actinomycetota)
not class-specific	<i>hpcopA</i>	copper resistance gene	co-selection link	166	UBA955 (Bacteroidota)

(continued on next page)

Table 2 (continued)

Category	Gene name	Associated drug class/function	AMR relevance/ VFs	MAGs #	Taxonomy
	VFG045688	isoprenyl transferase	VF	209	UBA970 (Bacillota)
	VFG015782	response regulator	VF	369	DMER64 (Bacteroidota)
	VFG047044	glucose-1-phosphate thymidyltransferase	VF	648	Lacibacter (Bacteroidota)
	VFG010916/ VFG010917	two-component system sensor histidine kinase LetS	VF	665	UBA10906 (Pseudomonadota) JALRLI01
	VFG009124	3-oxoacyl-ACP synthase KasB	VF	530	(Actinomycetota) JALRLI01
	VFG009242	3-oxoacyl-ACP synthase KasB	VF	530	(Actinomycetota) JALRLI01
	VFG022567	3-oxoacyl-ACP synthase KasB	VF	530	(Actinomycetota) JALRLI01
	VFG009437	metal-dependent transcriptional regulator	VF	530	(Actinomycetota) JALRLI01
	VFG009805	response regulator transcription factor	VF	530	(Actinomycetota) JALRLI01
	VFG009807	response regulator transcription factor	VF	530	(Actinomycetota) JALRLI01
	VFG009808	response regulator transcription factor	VF	530	(Actinomycetota) JALRLI01
	VFG024163	response regulator transcription factor	VF	530	(Actinomycetota)

Antimicrobial resistance genes (ARGs) were categorized according to the importance of their associated antimicrobial drug classes in human medicine in Canada, following the Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS) classification. Category I includes drugs of very high importance, Category II includes drugs of high importance, and Category III includes drugs of moderate importance. “MAG #” indicates the number of metagenome-assembled genomes in which each gene or factor was detected, and “Taxonomy” denotes the corresponding genus-level (phylum) classification when available.

However, differences emerged in ecological sensitivity and spatial resolution. Several genera with pathogenic potential, such as *Microbacterium*, *Bacillus*, and *Corynebacterium*, exhibited strong seasonal and site-specific patterns in amplicon data but appeared more evenly distributed in the shotgun profiles. These discrepancies likely reflect taxonomic assignment strategies along with amplification bias and sequencing depth [32,33,107,108]. Whole-genome shotgun reads derive from random genomic regions with heterogeneous taxonomic resolution. Lowest Common Ancestor (LCA)-based classifiers such as Kraken/Bracken [109] tend to assign conserved fragments to higher taxonomic levels, leading to taxonomic compression when aggregating across reads. In contrast, 16S amplicon sequencing targets a phylogenetically informative locus with consistent discriminatory power across taxa, enhancing sensitivity to community shifts along environmental gradients. Interestingly, a recent study from Lemonnier et al. [108] demonstrated that extracting 16S or 23S rRNA fragments from shotgun datasets recapitulated the ecological patterns observed in amplicon sequencing, suggesting that discrepancies between approaches largely reflect differences in locus targeting and taxonomic resolution rather than inherent platform limitations.

Conversely, shotgun metagenomics showed higher sensitivity for low-abundance and clinically relevant taxa. Six genera containing CDC- and WHO-listed pathogens (e.g., *Salmonella*, *Listeria*) were detected exclusively by shotgun sequencing. Methodological evaluation revealed three main drivers for these discrepancies. *In silico* evaluation revealed a distinct asymmetry in primer performance against the missing genera (*Salmonella*, *Listeria*, *Capnocytophaga*, *Haemophilus*, and *Streptobacillus*). While the forward primer (S15F-Y) aligned to these targets with moderate fidelity (typically 17/19 bp matches), the reverse primer (926R) exhibited severe mismatches, yielding only non-specific 7–8 bp alignments (Supplementary Table S3). This confirms that reverse primer failure prevented amplification, causing the exclusion of these pathogens from the amplicon dataset despite their presence. Quantitative PCR validation further corroborated the shotgun-based detection of *Listeria* and *Acinetobacter baumannii* [25,36]. Second, template abundance played a role. All genera exclusive to the shotgun dataset were low-abundance taxa. PCR amplification inherently favors abundant templates, potentially masking rare members even in the absence of severe primer mismatches [110]. This aligns with findings by Durazzi et al. [29], who showed that deep shotgun metagenomics often recovers rare

taxa missed by amplicon profiling. Finally, database limitations constrained resolution for specific groups. The 16S rRNA gene cannot reliably distinguish closely related genera; notably, even with the updated Greengenes2 database, *Escherichia* cannot be resolved from *Shigella* due to intrinsic sequence similarity [111]. Collectively, these findings indicate that specific primer mismatches and database constraints, rather than biological factors, account for the detection gaps in the amplicon surveillance.

In summary, these findings support a tiered One Health surveillance strategy. Amplicon sequencing serves as the scalable “ecological baseline” for long-term monitoring of community shifts and sentinel taxa (e.g., CRTs) across broad spatiotemporal scales. Shotgun metagenomics provides the complementary “targeted risk-factor assessment” layer, offering the functional precision needed to characterize resistance reservoirs, virulence factors, and emerging pathogens at high-priority sites. Integrating these approaches balances cost-effective routine monitoring with the diagnostic resolution required for proactive risk assessment in complex agricultural watersheds [29,106].

4.2. Land use and fecal-associated bacterial communities

Fecal-associated bacterial communities were strongly structured by upstream land use, stream function, and seasonal conditions. Agricultural drainage ditches consistently supported higher alpha diversity than mixed-use or forested sites, with distinct early- and late-season peaks. These temporal patterns likely reflect nutrient and bacterial pulses mobilized by snowmelt, surface runoff, and tile drainage [22,112–114]. The emergence of water temperature and nitrate as key predictors links bacterial diversity shifts to the seasonal co-mobilization of nutrients and fecal pollution characteristic of agricultural landscapes [115–118]. Previous work identified livestock manure, wildlife inputs, and septic effluent as key fecal sources at the current study sites [46,113,114,119,120].

Land use also altered the balance between core and conditionally rare taxa (CRTs). Agricultural ditches exhibited reduced core stability and elevated CRT contributions, indicating heightened responsiveness to episodic contaminant transport. This instability aligns with the specific physical attributes of drainage ditches, namely their rapid hydrological response, small volume, shallow and sometimes intermittent nature, and proximity to fertilized fields [40,48,98,121], as well as their

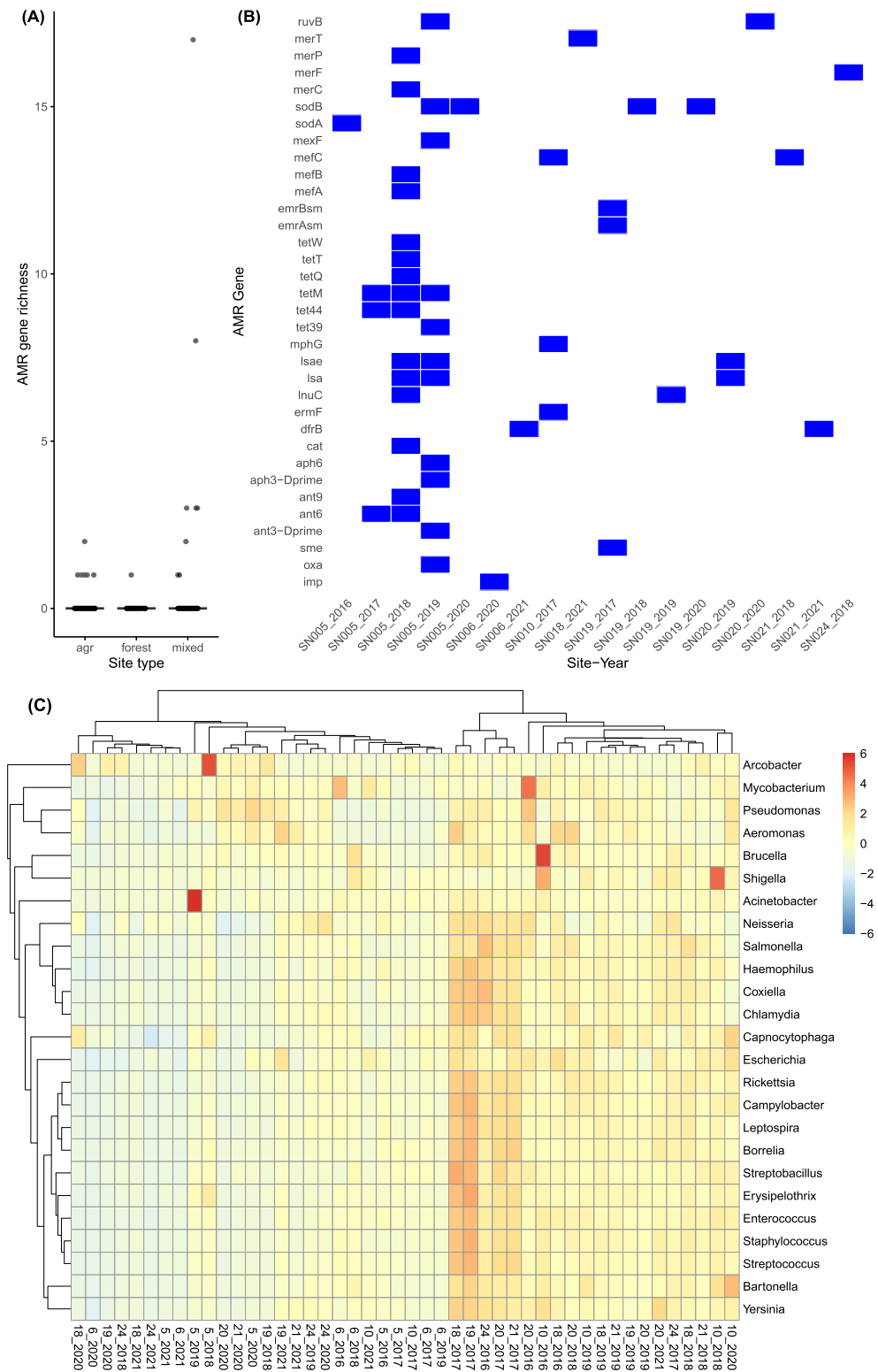


Fig. 5. Detection of antimicrobial resistance genes (ARGs) and associated taxa based on short-read shotgun data. (A) AMR gene detection counts across site types. (B) Heatmap showing presence of AMR genes by site and year. (C) Heatmap of genera known to harbor AMR genes across sites and years; colors indicate row-scaled (z-score) relative abundance (red = higher, blue = lower relative to the genus mean).

critical function as wildlife corridors [46,122]. In this agroecosystem, CRTs functioned as a latent reservoir for relatively high-risk taxa, encompassing 53 genera with zoonotic or opportunistic pathogenic potential (12.6% of all pathogen-associated taxa). Of these, only one genus was identified within the core

microbiome. However, CRT enrichment is context-dependent and correlative: in pristine systems, rare taxa may fluctuate due to natural environmental variability rather than contamination or management-induced alterations [40,98,122,123]. Therefore, CRTs are best interpreted as ecosystem-specific indicators of instability rather than

universal early-warning signals of pollution impact and impairment. Although rare under stable conditions, these taxa contributed disproportionately to community turnover during stress periods [98,124,125]. Given that low-abundance taxa can still harbor ARGs and VFs [126], their episodic dominance represents an underrecognized pathway for pathogen persistence and transmission in hydrologically dynamic watersheds [123,127].

Collectively, land use and hydrology drive fecal-associated bacterial dynamics, establishing agricultural ditches as persistent upstream hotspots [45]. Strong correlations between nitrate and diversity implicate nutrient-laden agricultural runoff/drainage in mobilizing these communities. Consequently, CRTs emerge as ecological sentinels under the One Health framework; monitoring their fluctuations provides a sensitive early-warning mechanism for detecting environmentally engendered risk-factors and pathogen emergence in agricultural watersheds.

4.3. Ecological and public health implications of fecal-associated pathogens

Fecal-associated pathogens serve as direct links between land use practices and One Health risk-factors [128,129]. In this study, agricultural drainage ditches emerged as important sentinel environments, functioning as convergence zones for manure, runoff/drainage, and wildlife inputs. We consistently detected elevated levels of opportunistic genera, such as *Sphingomonas* spp., *Massilia*, *Microbacterium* and *Clostridium* [130–134], within these nutrient enriched, hydrologically responsive systems [135,136].

Spatial patterning further revealed distinct contamination pathways [112]. Livestock-associated taxa such as *Campylobacter* and *Brucella*, wastes from cattle and poultry operations [137–139], were confined to agricultural ditches [140,141]. By contrast, *Bartonella* and *Yersinia* appeared exclusively at mixed-use sites, reflecting complex contamination pathways like urban runoff, wildlife sources, and septic leachate more typical of peri-urban landscapes [142–145]. Many enriched genera, like *Campylobacter*, *Brucella*, *Rickettsia*, and *Acinetobacter*, appear on CDC and WHO priority pathogen lists [85,86], supporting their use as indicators of anthropogenic pollution. Temporal dynamics reinforced this link, with pathogen richness peaking during snowmelt and high-flow events, indicating transport-driven mobilization.

Although drainage ditches are not typically direct exposure points for humans, they act as persistent upstream reservoirs influencing downstream waters used for recreation, irrigation, and drinking water. Consequently, the detection of WHO- and CDC-priority pathogens highlights the necessity of integrated surveillance, coupling taxonomic precision with functional screening, to monitor upstream hazards before they impact downstream receptors.

4.4. Distinct resistome and virulome profiles in agricultural and mixed-use watersheds

Antimicrobial resistance genes (ARGs) and Virulence Factors (VFs) represent latent reservoirs of pathogenic potential bridging environmental and health domains [128,129]. We observed low prevalent but distinct spatial segregation of these determinants: agricultural drainage ditches were reservoirs for VFs, while mixed-use sites were hotspots for ARGs. These spatial contrasts align with global evidence that resistance and virulence determinants respond to local environmental determinants and shifts [44,140,146].

The detection of ARGs at mixed-use sites, although at low frequency, is consistent with the possibility of co-selection [147–151] arising from intersecting urban and agricultural inputs. Although selective agents such as metals or antibiotics were not directly quantified, the co-occurrence of metal-resistance (e.g., *merC*, *merP*) and oxidative-stress (e.g., *sodA*, *ruvB*) determinants alongside high-priority β -lactamases and carbapenemase genes (*oxa*, *imp*) [151–153] may reflect shared ecological stressors rather than confirmed selective mechanisms. While

short-read profiling provides limited host resolution, multi-method integration identified *Acinetobacter* spp. and *Arcobacter* spp. as dominant ARG carriers at the mixed-use site SN_5. Specifically, genome-resolved analysis confirmed *A. baumannii* as a reservoir for blaOXA-type β -lactamase [154], a finding corroborated by qPCR detection of *oxa-51-like* gene in this species at multiple sampling sites included in this study (Supplementary Fig. S11, data and method unpublished). Similarly, the detection of resistance genes such as *cat*, *lhuC*, *lsaE*, and *mefA/B* at SN_5 in 2018 aligns with culture-based isolation of 279 multidrug-resistant *Arcobacter butzleri* strains from the same location [155], establishing a robust link between the resistome profile and viable hosts.

The integration of short-read and genome-resolved (MAG) approaches reveals that each captures distinct dimensions of the resistome and VFs. The minimal overlap between datasets highlights the trade-off between the broad sensitivity of short-read profiling and the genomic context provided by MAGs [25,156]. MAG-associated VFs were predominantly regulatory or metabolic (e.g., response regulators, *rfaA*, *KasB*) rather than toxin-encoding. This suggests that environmental reservoirs function primarily as incubators for adaptive traits, enhancing bacterial persistence and stress tolerance [157–159], which may eventually be co-opted by clinical pathogens [160,161].

Consequently, we interpret the detection of these low-abundance determinants not as evidence of active disease, but as indicators of latent risk-factor pathways. Their epidemiological relevance is defined by their potential mobility and the presence of compatible hosts within nutrient-enriched, managed and/or critically disturbed environments [162]. The observed co-occurrence of ARGs, VFs, and metal-resistance genes further validates the necessity of dual-platform surveillance to characterize these complex exposure risks [149].

4.5. Limitations and future directions

This study offers a longitudinal One Health perspective on fecal-associated microbial dynamics, pathogen occurrence, and resistance reservoirs. However, several methodological and contextual limitations warrant consideration.

First, both sequencing approaches carry inherent biases. Primer-dependent amplification can skew community profiles toward dominant taxa, as observed for *Flavobacterium* (24.7% in amplicon data vs. 7.8% in shotgun profiles) [31], while primer mismatches led to the underrepresentation of specific pathogens (e.g., *Listeria*, *Salmonella*). Conversely, shotgun metagenomics was constrained by sequencing depth (avg. 16.6 million reads/sample in this study vs. > 50–100 million reads per sample in deep sequencing) and high background DNA load [32,33]. Taxonomic assignment represents an additional limitation. While the Kraken2/Bracken framework used here is computationally efficient, alternative marker-gene strategies (e.g., MetaPhlAn/ChocoPhlAn implemented in HUMAnN) can, in principle, infer species-level presence [163]. However, such inference is strictly limited to reference-covered organisms and fails to capture strain-level or plasmid-borne virulence and AMR determinants [164]. Thus, different tools can yield varying levels of taxonomic resolution and may influence between-sample differentiation, particularly in complex environmental matrices. While a full re-analysis using alternative classifiers or rRNA fragment extraction, as recommended by Lemonnier et al. [108], was beyond the scope of this study, it represents a critical direction for future method comparison. Given these bioinformatic constraints combined with shallow sequencing depth, robust species-level resolution was not feasible. Consequently, classification was restricted to the genus level, and we relied on targeted qPCR assays for species-specific confirmation and quantification.

Second, environmental heterogeneity introduces complexity that challenges the resolution of pulse disturbances. While hydrological variability and land use intensity drive microbial mobilization, bi-weekly sampling intrinsically underrepresents transient, high-impact

stressors—including extreme rainfall pulses, storm-driven runoff/discharge, pesticide and manure applications, and point-source discharges like lagoon discharge events [39,165–167]. This constraint limits the resolution of fine-scale disturbance responses. Although the current design provided robust long-term coverage, future surveillance should prioritize event-based sampling during hydrologically sensitive periods (e.g., snowmelt, stormflow) to capture the rapid mobilization of pathogens, ARGs, and VFs.

To fully operationalize these ecological insights, surveillance must also expand analytically. Integrating multi-omic approaches (e.g., long-read sequencing, metatranscriptomics) would resolve the strain-level pathogenicity and gene expression gaps identified here. Furthermore, coupling these high-resolution molecular data with Microbial Source Tracking (MST) [168,169] and Quantitative Microbial Risk Assessment (QMRA) is essential to translate gene detection into more actionable risk probabilities. Embedding this event-driven, multi-omic strategy into watershed management will strengthen the predictive capacity of One Health surveillance, directly informing mitigation policies for sustainable agricultural landscapes.

5. Conclusion

This five-year longitudinal study establishes a critical baseline for integrating microbial ecology into One Health surveillance. We demonstrated that a tiered monitoring strategy, coupling the ecological scalability of amplicon sequencing with the functional precision of shotgun metagenomics, is essential for resolving the complex interplay between land use determinants and microbial risk-factors. Biologically, agricultural drainage ditches functioned as persistent upstream reservoirs of elevated diversity and pathogen-associated taxa, where conditionally rare taxa (CRTs) served as sensitive sentinels of environmental change. Conversely, mixed-use watersheds emerged as relative hotspots for antimicrobial resistance, with patterns compatible with potential co-selection in transitional landscapes characterized by varied source contributions. It is suggested to prioritize these high-risk interfaces and shift toward more event-based sampling, or systematic-event hybrid, to capture pulse-driven mobilization. Ultimately, integrating these multi-omic, risk-based approaches into policy/public health frameworks will transform passive environmental monitoring into proactive, evidence-based mitigation of zoonotic and resistance threats.

Crown copyright

©2026 His Majesty the King in Right of Canada, as represented by the Minister of Agriculture and Agri-Food Canada. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license.

CRedit authorship contribution statement

Yichao Shi: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Haley Sanderson:** Writing – review & editing, Methodology. **Jiacheng Chuan:** Writing – review & editing, Methodology. **Izhar U.H. Khan:** Writing – review & editing, Data curation. **Mark Sunohara:** Writing – review & editing, Data curation. **Emilia Craiovan:** Writing – review & editing, Data curation. **David R. Lapen:** Writing – review & editing, Data curation, Funding acquisition. **Moussa Diarra:** Writing – review & editing. **Wen Chen:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The study was financially funded by the Government of Canada's Environmental

Change Onehealth Observatory Project (ECO2, J-002305), the Genomics Research and Development Initiative-Metagenomics Based Ecosystem Biomonitoring (GRDI-EcoBiomics, J-001263). Salary of bio-informatician was supported by the SCAP-ASC-18 Poultry Cluster Activity 5A - Microbiome in Poultry Production Environment as a Function of Climate and Alternative Use of Litter (J-003360).

Declaration of competing 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.

Acknowledgments

We thank R. Assabgui, M. Cloutier, and students from Agriculture and Agri-Food Canada for their diligent work in collecting and processing water samples. We also thank S. Sanschagrin and D. Konkin from the National Research Council for their assistance in high-throughput sequencing. We thank Dr. Cecilio Valadez Cano for providing computational support. We acknowledge the support of the Canadian federal government's Science Data Centre of Excellence team, the Science Data Storage Solution (SDSS), and the General Purpose Science Cluster (GPSC) for providing the computational infrastructure for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.onehlt.2026.101384>.

Data availability

The raw paired-end 16S rRNA gene amplicon sequencing data have been deposited in the NCBI Sequence Read Archive (SRA), under the Bioproject accession PRJNA858176 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA858176>), Biosample accessions SAMN29671977–2598 and SAMN56369704–56369930. The raw shotgun sequencing data in FASTQ format are accessible from the European Nucleotide Archive under the Study accession PRJEB102733, with the samples accessions ERS27328153–ERS27328689.

References

- [1] R.R. Ghai, R.M. Wallace, J.C. Kile, T.R. Shoemaker, A.R. Vieira, M.E. Negron, S. V. Shadomy, J.R. Sinclair, G.W. Goryoka, S.J. Salyer, C. Barton Behravesh, A generalizable one health framework for the control of zoonotic diseases, *Sci. Rep.* 12 (2022) 8588, <https://doi.org/10.1038/s41598-022-12619-1>.
- [2] D.T.S. Hayman, W.B. Adisasmito, S. Almuhairei, C.B. Behravesh, P. Bilivogui, S. A. Bukachi, N. Casas, N.C. Becerra, D.F. Charron, A. Chaudhary, et al., Developing one health surveillance systems, *One Health* 17 (2023) 100617, <https://doi.org/10.1016/j.onehlt.2023.100617>.
- [3] N.A. Dafale, S. Srivastava, H.J. Purohit, Zoonosis: an emerging link to antibiotic resistance under “one health approach”, *Indian J. Microbiol.* 60 (2020) 139–152, <https://doi.org/10.1007/s12088-020-00860-z>.
- [4] Y. He, Q. Yuan, J. Mathieu, L. Stadler, N. Senehi, R. Sun, P.J.J. Alvarez, Antibiotic resistance genes from livestock waste: occurrence, dissemination, and treatment, *npj Clean Water* 3 (2020) 4, <https://doi.org/10.1038/s41545-020-0051-0>.
- [5] Z. Li, J. Tang, X. Wang, X. Ma, H. Yuan, C. Gao, Q. Guo, X. Guo, J. Wan, C. Dagot, The environmental lifecycle of antibiotics and resistance genes: transmission mechanisms, challenges, and control strategies, *Microorganisms* 13 (2025) 2113.
- [6] P. Barathe, K. Kaur, S. Reddy, V. Shiram, V. Kumar, Antibiotic pollution and associated antimicrobial resistance in the environment, *J. Hazard. Mater. Lett.* 5 (2024) 100105, <https://doi.org/10.1016/j.hazl.2024.100105>.
- [7] N. Lajqi Berisha, A. Poceva Panovska, Z. Hajrulai-Musliu, Antibiotic resistance and aquatic systems: importance in public health, *Water* 16 (2024) 2362.
- [8] S. Frey, N. Gottschall, G. Wilkes, D. Grégoire, E. Topp, K. Pintar, M. Sunohara, R. Marti, D. Lapen, Rainfall-induced runoff from exposed streambed sediments: an important source of water pollution, *J. Environ. Qual.* 44 (2015) 236–247.
- [9] L. Taing, H. Bhatia, R.A. Kaiser, M. Qadir, H. Mehmood, A rapid review of environmental health gaps in antimicrobial resistance and water-related research from 1990–2020, *Int. J. Environ. Res. Public Health* 19 (2022) 6549.

- [10] E. O'Brien, I. Xagoraki, A water-focused one-health approach for early detection and prevention of viral outbreaks, *One Health* 7 (2019) 100094, <https://doi.org/10.1016/j.onehlt.2019.100094>.
- [11] S.A. McEwen, P.J. Collignon, Antimicrobial resistance: a one health perspective, In *Antimicrobial resistance in bacteria from livestock and companion animals*, pp. (2018) 521–547, <https://doi.org/10.1128/9781555819804.ch25>.
- [12] A. Feßler, S. Schwarz, Antimicrobial resistance in *Corynebacterium* spp., *arcanobacterium* spp., and *Trueperella pyogenes*, *Microbiol. Spectrum* 5 (6) (2017), <https://doi.org/10.1128/microbiolspec.ARBA-0021-2017>.
- [13] J.L. Mawdsley, R.D. Bardgett, R.J. Merry, B.F. Pain, M.K. Theodorou, Pathogens in livestock waste, their potential for movement through soil and environmental pollution, *Appl. Soil Ecol.* 2 (1995) 1–15.
- [14] Y.-G. Zhu, T.A. Johnson, J.-Q. Su, M. Qiao, G.-X. Guo, R.D. Stedtfeld, S. A. Hashsham, J.M. Tiedje, Diverse and abundant antibiotic resistance genes in Chinese swine farms, *Proc. Natl. Acad. Sci. USA* 110 (2013) 3435–3440, <https://doi.org/10.1073/pnas.1222743110>.
- [15] S. Reddy, K. Kaur, P. Barathe, V. Shiriram, M. Govarthanan, V. Kumar, Antimicrobial resistance in urban river ecosystems, *Microbiol. Res.* 263 (2022) 127135, <https://doi.org/10.1016/j.micres.2022.127135>.
- [16] H. Li, Q. Luo, S. Zhao, P. Zhao, X. Yang, Q. Huang, J. Su, Watershed urbanization enhances the enrichment of pathogenic bacteria and antibiotic resistance genes on microplastics in the water environment, *Environ. Pollut.* 313 (2022) 120185.
- [17] H.-T. Hwang, S. Frey, Y.-J. Park, K. Pintar, D. Lapen, J. Thomas, J. Spoelstra, S. Schiff, S. Brown, E. Sudicky, Estimating cumulative wastewater treatment plant discharge influences on *acetic acid* and *escherichia coli* in a highly impacted watershed with a fully-integrated modelling approach, *Water Res.* 157 (2019) 647–662.
- [18] K. Hwengwere, H. Paramel Nair, K.A. Hughes, L.S. Peck, M.S. Clark, C.A. Walker, Antimicrobial resistance in Antarctica: is it still a pristine environment? *Microbiome* 10 (2022) 71, <https://doi.org/10.1186/s40168-022-01250-x>.
- [19] T.M. Nolan, L.J. Reynolds, L. Sala-Comera, N.A. Martin, J.H. Stephens, G.M. P. O'Hare, J.J. O'Sullivan, W.G. Meijer, Land use as a critical determinant of faecal and antimicrobial resistance gene pollution in riverine systems, *Sci. Total Environ.* 871 (2023) 162052, <https://doi.org/10.1016/j.scitotenv.2023.162052>.
- [20] K.G. Field, M. Samadpour, Fecal source tracking, the indicator paradigm, and managing water quality, *Water Res.* 41 (2007) 3517–3538.
- [21] V.J. Harwood, C. Staley, B.D. Badgley, K. Borges, A. Korajkic, Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes, *FEMS Microbiol. Rev.* 38 (2014) 1–40.
- [22] C. Staley, K.V. Gordon, M.E. Schoen, V.J. Harwood, Performance of two quantitative PCR methods for microbial source tracking of human sewage and implications for microbial risk assessment in recreational waters, *Appl. Environ. Microbiol.* 78 (2012) 7317–7326.
- [23] G. Wilkes, N.J. Ruecker, N.F. Neumann, V.P.J. Gannon, C. Jokinen, M. Sunohara, E. Topp, K.D.M. Pintar, T.A. Edge, D.R. Lapen, Spatiotemporal analysis of cryptosporidium species/genotypes and relationships with other zoonotic pathogens in surface water from mixed-use watersheds, *Appl. Environ. Microbiol.* 79 (2013) 434–448, <https://doi.org/10.1128/AEM.01924-12>.
- [24] C.C. Jokinen, T.A. Edge, W. Koning, C.R. Laing, D.R. Lapen, J. Miller, S. Mutschall, A. Scott, E.N. Taboada, J.E. Thomas, et al., Spatial and temporal drivers of zoonotic pathogen contamination of an agricultural watershed, *J. Environ. Qual.* 41 (2012) 242–252, <https://doi.org/10.2134/jeq2011.0203>.
- [25] C. Quince, A.W. Walker, J.T. Simpson, N.J. Loman, N. Segata, Shotgun metagenomics, from sampling to analysis, *Nat. Biotechnol.* 35 (2017) 833–844.
- [26] X. Huang, M. Toro, A. Reyes-Jara, A.I. Moreno-Switt, A.D. Adell, R.R. Bonelli, S. Gutiérrez, F.P. Álvarez, A.D. de Lima Rocha, G.B. Kraychete, Integrative genome-centric metagenomics for surface water surveillance: elucidating microbiomes, antimicrobial resistance, and their associations, *Water Res.* 264 (2024) 122208.
- [27] C.S. Chan, K.-G. Chan, Y.-L. Tay, Y.-H. Chua, K.M. Goh, Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing, *Front. Microbiol.* 6 (2015) 177.
- [28] R. Ranjan, A. Rani, A. Metwally, H.S. McGee, D.L. Perkins, Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing, *Biochem. Biophys. Res. Commun.* 469 (2016) 967–977, <https://doi.org/10.1016/j.bbrc.2015.12.083>.
- [29] F. Durazzi, C. Sala, G. Castellani, G. Manfreda, D. Remondini, A. De Cesare, Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota, *Sci. Rep.* 11 (2021) 3030, <https://doi.org/10.1038/s41598-021-82726-y>.
- [30] M.R. Stothart, P.D. McLoughlin, J. Poissant, Shallow shotgun sequencing of the microbiome recapitulates 16S amplicon results and provides functional insights, *Mol. Ecol. Resour.* 23 (2023) 549–564.
- [31] J. He, T. Zhou, X. Shen, N. Zhang, C. Sun, S. Lu, Y. Shao, Primer selection impacts the evaluation of microecological patterns in environmental microbiomes, *Imeta* 2 (2023) e135, <https://doi.org/10.1002/imt2.135>.
- [32] B. Hillmann, G.A. Al-Ghalith, R.R. Shields-Cutler, Q. Zhu, D.M. Gohl, K. Beckman, R. Knight, D. Knights, Evaluating the information content of shallow shotgun metagenomics, *mSystems* 3 (2018), <https://doi.org/10.1128/mSystems.00069-18>.
- [33] J. Zhou, Z. He, Y. Yang, Y. Deng, S.G. Tringe, L. Alvarez-Cohen, High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats, *mBio* 6 (2015), <https://doi.org/10.1128/mbio.02288-02214>.
- [34] A. Sczyrba, P. Hofmann, P. Belmann, D. Koslicki, S. Janssen, J. Dröge, I. Gregor, S. Majda, J. Fiedler, E. Dahms, Critical assessment of metagenome interpretation—a benchmark of metagenomics software, *Nat. Methods* 14 (2017) 1063–1071.
- [35] T. Thomas, J. Gilbert, F. Meyer, Metagenomics—a guide from sampling to data analysis, *Microb. Inf. Exp.* 2 (2012) 3.
- [36] J.S. Johnson, D.J. Spakowicz, B.-Y. Hong, L.M. Petersen, P. Demkowicz, L. Chen, S.R. Leopold, B.M. Hanson, H.O. Agresta, M. Gerstein, Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis, *Nat. Commun.* 10 (2019) 5029.
- [37] S. Nayfach, K.S. Pollard, Toward accurate and quantitative comparative metagenomics, *Cell* 166 (2016) 1103–1116.
- [38] W. Chen, G. Wilkes, I.U. Khan, K.D. Pintar, J.L. Thomas, C.A. Lévesque, J. T. Chapados, E. Topp, D.R. Lapen, Aquatic bacterial communities associated with land use and environmental factors in agricultural landscapes using a metabarcoding approach, *Front. Microbiol.* 9 (2018) 2301.
- [39] P. Pham, Y. Shi, I. Khan, M. Sumarah, J. Renaud, M. Sunohara, E. Craiovan, D. Lapen, S. Aris-Brosou, W. Chen, The functions and factors governing fungal communities and diversity in agricultural waters: insights into the ecosystem services aquatic mycobiota provide, *Front. Microbiol.* 15 (2024) 1460330, <https://doi.org/10.3389/fmicb.2024.1460330>.
- [40] Y. Shi, I.U.H. Khan, D. Radford, G. Guo, M. Sunohara, E. Craiovan, D.R. Lapen, P. Pham, W. Chen, Core and conditionally rare taxa as indicators of agricultural drainage ditch and stream health and function, *BMC Microbiol.* 23 (2023) 62, <https://doi.org/10.1186/s12866-023-02755-7>.
- [41] S. Spänig, L. Eick, J.K. Nuy, D. Beisser, M. Ip, D. Heider, J. Boenigk, A multi-omics study on quantifying antimicrobial resistance in European freshwater lakes, *Environ. Int.* 157 (2021) 106821, <https://doi.org/10.1016/j.envint.2021.106821>.
- [42] R.S. Hendriksen, P. Munk, P. Njage, B. van Bunnik, L. McNally, O. Lukjancenko, T. Röder, D. Nieuwenhuijse, S.K. Pedersen, J. Kjeldgaard, et al., Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage, *Nat. Commun.* 10 (2019) 1124, <https://doi.org/10.1038/s41467-019-08853-3>.
- [43] M.H. Tarek, J.A. Hubbart, E. Garner, Tracking sources and dissemination of indicator antibiotic resistance genes at a watershed scale, *ACS EST Water* 4 (2024) 399–412, <https://doi.org/10.1021/acsestwater.3c00394>.
- [44] G. Guo, D.R. Lapen, I.U. Khan, M. Sunohara, E. Craiovan, W. Chen, Examining the impact of agricultural drainage ditch management on in-stream bacterial communities involved in nitrogen cycling: insights from the environmental change one health observatory (eco2), *Front. Sustain. Food Syst.* 8 (2024) 1329422.
- [45] D. Lapen, P. Schmidt, J. Thomas, T. Edge, C. Flemming, J. Keithlin, N. Neumann, F. Pollari, N. Ruecker, A. Simhon, Towards a more accurate quantitative assessment of seasonal cryptosporidium infection risks in surface waters using species and genotype information, *Water Res.* 105 (2016) 625–637.
- [46] G. Wilkes, J. Brassard, T.A. Edge, V. Gannon, N. Gottschall, C.C. Jokinen, T. H. Jones, I.U. Khan, R. Marti, M.D. Sunohara, Long-term monitoring of waterborne pathogens and microbial source tracking markers in paired agricultural watersheds under controlled and conventional tile drainage management, *Appl. Environ. Microbiol.* 80 (2014) 3708–3720.
- [47] E.W. Rice, L. Bridgewater, A. APH, A. AQQ, F. WE, Standard Methods for the Examination of Water and Wastewater, American Public Health Association, 2012.
- [48] D. Wiemer, U. Loderstaedt, H. von Wulffen, S. Priesnitz, M. Fischer, E. Tannich, R. M. Hagen, Real-time multiplex PCR for simultaneous detection of campylobacter jejuni, salmonella, shigella and yersinia species in fecal samples, *Int. J. Med. Microbiol.* 301 (2011) 577–584, <https://doi.org/10.1016/j.ijmm.2011.06.001>.
- [49] E. Lyautey, Z. Lu, D.R. Lapen, G. Wilkes, A. Scott, T. Berkers, T.A. Edge, E. Topp, Distribution and diversity of *Escherichia coli* populations in the south nation river drainage basin, eastern Ontario, Canada, *Appl. Environ. Microbiol.* 76 (2010) 1486–1496.
- [50] WSC, Historical Hydrometric Data. <https://wateroffice.ec.gc.ca/search/historic.html>, 2025.
- [51] J. Chuan, A. Zhou, L.R. Hale, M. He, X. Li, Atria: an ultra-fast and accurate trimmer for adapter and quality trimming, *GigaByte* 2021 (2021) 1–18.
- [52] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet. J.* 17 (2011) 10–12.
- [53] E. Bolyen, J.R. Rideout, M.R. Dillon, N.A. Bokulich, C.C. Abnet, G.A. Al-Ghalith, H. Alexander, E.J. Alm, M. Arumugam, F. Asnicar, Reproducible, interactive, scalable and extensible microbiome data science using qiime 2, *Nat. Biotechnol.* 37 (2019) 852–857.
- [54] B.J. Callahan, P.J. McMurdie, M.J. Rosen, A.W. Han, A.J.A. Johnson, S.P. Holmes, Dada2: high-resolution sample inference from illumina amplicon data, *Nat. Methods* 13 (2016) 581–583.
- [55] D. McDonald, Y. Jiang, M. Balaban, K. Cantrell, Q. Zhu, A. Gonzalez, J.T. Morton, G. Nicolaou, D.H. Parks, S.M. Karst, Greengenes2 unifies microbial data in a single reference tree, *Nat. Biotechnol.* 42 (2024) 715–718.
- [56] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for illumina sequence data, *Bioinformatics* 30 (2014) 2114–2120.
- [57] A. Simon, FastQC: A Quality Control Tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, 2010.
- [58] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics* 32 (2016) 3047–3048.

- [59] J. Lu, S.L. Salzberg, Ultrafast and accurate 16s rRNA microbial community analysis using kraken 2, *Microbiome* 8 (2020) 124.
- [60] J. Lu, F.P. Breitwieser, P. Thiele, S.L. Salzberg, Bracken: estimating species abundance in metagenomics data, *PeerJ Comput. Sci.* 3 (2017) e104.
- [61] F. Mölder, K.P. Jablonski, B. Letcher, M.B. Hall, C.H. Tomkins-Tinch, V. Sochat, J. Forster, S. Lee, S.O. Twardziok, A. Kunitz, Sustainable data analysis with snakemake, *F1000Research* 10 (2021) 33.
- [62] M. Timilsina, D. Chundru, A.K. Pradhan, R.A. Blaustein, M. Ghanem, Benchmarking metagenomic pipelines for the detection of foodborne pathogens in simulated microbial communities, *J. Food Prot.* 88 (2025) 100583, <https://doi.org/10.1016/j.jfmp.2025.100583>.
- [63] N. Bonin, E. Doster, H. Worley, L.J. Pinnell, J.E. Bravo, P. Ferm, S. Marini, M. Prosperi, N. Noyes, P.S. Morley, Megares and amr++, v3. 0: an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing, *Nucleic Acids Res.* 51 (2023) D744–D752.
- [64] B. Liu, D. Zheng, S. Zhou, L. Chen, J. Yang, Vfdb 2022: a general classification scheme for bacterial virulence factors, *Nucleic Acids Res.* 50 (2022) D912–D917.
- [65] P.T. Clausen, F.M. Aarestrup, O. Lund, Rapid and precise alignment of raw reads against redundant databases with kma, *BMC Bioinform.* 19 (2018) 307.
- [66] D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, Megahit: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph, *Bioinformatics* 31 (2015) 1674–1676.
- [67] D. Li, R. Luo, C.-M. Liu, C.-M. Leung, H.-F. Ting, K. Sadakane, H. Yamashita, T.-W. Lam, Megahit v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices, *Methods* 102 (2016) 3–11.
- [68] D. Hyatt, G.-L. Chen, P.F. LoCascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification, *BMC Bioinform.* 11 (2010) 119.
- [69] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with bowtie 2, *Nat. Methods* 9 (2012) 357–359.
- [70] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, G.P.D.P. Subgroup, The sequence alignment/map format and samtools, *Bioinformatics* 25 (2009) 2078–2079.
- [71] J. Alneberg, B.S. Bjarnason, I. De Bruijn, M. Schirmer, J. Quick, U.Z. Ijaz, L. Lahti, N.J. Loman, A.F. Andersson, C. Quince, Binning metagenomic contigs by coverage and composition, *Nat. Methods* 11 (2014) 1144–1146.
- [72] D.D. Kang, F. Li, E. Kirton, A. Thomas, R. Egan, H. An, Z. Wang, Metabat 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies, *PeerJ* 7 (2019) e7359.
- [73] C.M. Sieber, A.J. Probst, A. Sharrar, B.C. Thomas, M. Hess, S.G. Tringe, J. F. Banfield, Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy, *Nat. Microbiol.* 3 (2018) 836–843.
- [74] A. Chklovskii, D.H. Parks, B.J. Woodcroft, G.W. Tyson, Checkm2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning, *Nat. Methods* 20 (2023) 1203–1212.
- [75] R.M. Bowers, N.C. Kyrpides, R. Stepanauskas, M. Harmon-Smith, D. Doud, T. Reddy, F. Schulz, J. Jarett, A.R. Rivers, E.A. Elze-Fadrosch, Minimum information about a single amplified genome (misag) and a metagenome-assembled genome (mimag) of bacteria and archaea, *Nat. Biotechnol.* 35 (2017) 725–731.
- [76] P.-A. Chaumeil, A.J. Mussig, P. Hugenholtz, D.H. Parks, Gtdb-tk: a toolkit to classify genomes with the genome taxonomy database, *Bioinformatics* 36 (2019) 1925–1927, <https://doi.org/10.1093/bioinformatics/btz848>.
- [77] D.H. Parks, M. Chuvochina, C. Rinke, A.J. Mussig, P.-A. Chaumeil, P. Hugenholtz, Gtdb: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy, *Nucleic Acids Res.* 50 (2022) D785–D794.
- [78] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis, T.L. Madden, Ncbi blast: a better web interface, *Nucleic Acids Res.* 36 (2008) W5–W9.
- [79] J. Yang, J. Park, Y. Jung, J. Chun, Amdb: a database of animal gut microbial communities with manually curated metadata, *Nucleic Acids Res.* 50 (2021) D729–D735, <https://doi.org/10.1093/nar/gkab1009>.
- [80] Y. Xu, B. Lei, Q. Zhang, Y. Lei, C. Li, X. Li, R. Yao, R. Hu, K. Liu, Y. Wang, et al., Addgma: a database for domestic animal gut microbiome atlas, *Comput. Struct. Biotechnol. J.* 20 (2022) 891–898, <https://doi.org/10.1016/j.csbj.2022.02.003>.
- [81] A. Bartlett, D. Padfield, L. Lear, R. Bendall, M. Vos, A comprehensive list of bacterial pathogens infecting humans, *Microbiology* 168 (2022), <https://doi.org/10.1099/mic.0.001269>.
- [82] NCBI, Ncbi Pathogens. <https://www.ncbi.nlm.nih.gov/pathogens/>, 2025.
- [83] R.D. Olson, R. Assaf, T. Brettin, N. Conrad, G. Cucinelli, J.J. Davis, D.M. Dempsey, A. Dickerman, E.M. Dietrich, R.W. Kenyon, et al., Introducing the bacterial and viral bioinformatics resource center (bv-brc): a resource combining patric, ird and vipr, *Nucleic Acids Res.* 51 (2023) D678–D689, <https://doi.org/10.1093/nar/gkac1003>.
- [84] USDA-APHIS, National List of Reportable Animal Diseases. <https://www.aphis.usda.gov/livestock-poultry-disease/surveillance/reportable-diseases>, 2025.
- [85] WHO, Who Bacterial Priority Pathogens List, 2024, p. 2024. <https://www.who.int/publications/i/item/9789240093461>.
- [86] CDC, Facts About Diseases that Can Spread Between Animals and People. <https://www.cdc.gov/healthy-pets/diseases/index.html>, 2025.
- [87] R.C. Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, 2025.
- [88] J. Oksanen, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R. O'Hara, G. L. Simpson, P. Solymos, M.H.H. Stevens, H. Wagner, Package 'vegan', in: *Community Ecology Package*, Version 2, 2013, pp. 1–295.
- [89] R. Kindt, R. Coe, *Tree Diversity Analysis: A Manual and Software for Common Statistical Methods for Ecological and Biodiversity Studies*, World Agroforestry Centre, 2005.
- [90] L. Jost, Entropy and diversity, *Oikos* 113 (2006) 363–375, <https://doi.org/10.1111/j.2006.0030-1299.14714.x>.
- [91] W.N. Venables, B.D. Ripley, *Modern Applied Statistics with s*, Springer Science & Business Media, 2013.
- [92] B. Bolker, *Getting Started with the Glimmmb Package*, Cran. R-Project Vignette, 2019, p. 9.
- [93] R. Lenth, M.R. Lenth, Package 'lsmeans', *Am. Stat.* 34 (2018) 216–221.
- [94] S. RColorBrewer, M.A. Liaw, Package 'Randomforest', University of California, Berkeley, Berkeley, CA, USA, 2018.
- [95] D.R. Cutler, T.C. Edwards Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, J. J. Lawler, Random forests for classification in ecology, *Ecology* 88 (2007) 2783–2792.
- [96] J. Fox, R. Muenchen, D. Putler, M.J. Fox, Package 'rmdrmisc', CRAN, 2022.
- [97] A. Shade, N. Stopnisek, Abundance-occupancy distributions to prioritize plant core microbiome membership, *Curr. Opin. Microbiol.* 49 (2019) 50–58, <https://doi.org/10.1016/j.mib.2019.09.008>.
- [98] A. Shade, S.E. Jones, J.G. Caporaso, J. Handelsman, R. Knight, N. Fierer, J. A. Gilbert, Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity, *mBio* 5 (2014) e01371–14, <https://doi.org/10.1128/mBio.01371-14>.
- [99] T. Giorgino, Computing and visualizing dynamic time warping alignments in R: the dtw package, *J. Stat. Softw.* 31 (2009) 1–24.
- [100] R. Kolde, M.R. Kolde, Package 'pheatmap', R Package, 2015, p. 790.
- [101] Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data, *Bioinformatics* 32 (2016) 2847–2849.
- [102] B. Ripley, B. Venables, D.M. Bates, K. Hornik, A. Gebhardt, D. Firth, M.B. Ripley, Package 'mass' 538, Cran r, 2013, p. 822.
- [103] S. Tripathi, D. Purchase, M. Govarthanan, R. Chandra, S. Yadav, Regulatory and innovative mechanisms of bacterial quorum sensing-mediated pathogenicity: a review, *Environ. Monit. Assess.* 195 (2022) 75, <https://doi.org/10.1007/s10661-022-10564-0>.
- [104] M.G. Langille, J. Zaneveld, J.G. Caporaso, D. McDonald, D. Knights, J.A. Reyes, J. C. Clemente, D.E. Burkpile, R.L. Vega Thurber, R. Knight, Predictive functional profiling of microbial communities using 16s rRNA marker gene sequences, *Nat. Biotechnol.* 31 (2013) 814–821.
- [105] K.K. Chau, W. Matlock, B. Constantines, S. Lipworth, L. Newbold, H. Tipper, T. Goodall, H. Brett, J. Hughes, D.W. Crook, et al., Evaluation of metagenomic, 16s rRNA gene and ultra-plexed PCR-based sequencing approaches for profiling antimicrobial resistance gene and bacterial taxonomic composition of polymicrobial samples, *bioRxiv* (2022), <https://doi.org/10.1101/2022.05.12.491637>.
- [106] M. Tessier, J.S. Neumann, E. Afshinnekoo, M. Pineda, R. Hersch, L.F.M. Velho, B. T. Segovia, F.A. Lansac-Toha, M. Lemke, R. DeSalle, et al., Large-scale differences in microbial biodiversity discovery between 16s amplicon and shotgun sequencing, *Sci. Rep.* 7 (2017) 6589, <https://doi.org/10.1038/s41598-017-06665-3>.
- [107] W. Zuo, B. Wang, X. Bai, Y. Luan, Y. Fan, S. Michail, F. Sun, 16s rRNA and metagenomic shotgun sequencing data revealed consistent patterns of gut microbiome signature in pediatric ulcerative colitis, *Sci. Rep.* 12 (2022) 6421, <https://doi.org/10.1038/s41598-022-07995-7>.
- [108] C. Lemonnier, B. Alric, I. Domaizon, F. Rimet, Comparison of metabarcoding and shotgun sequencing confirms the relevance of chloroplast rRNA genes to assess community structure of lake phytoplankton, *Mol. Ecol. Resour.* 26 (2026) e70077, <https://doi.org/10.1111/1755-0998.70077>.
- [109] D.E. Wood, S.L. Salzberg, Kraken: ultrafast metagenomic sequence classification using exact alignments, *Genome Biol.* 15 (2014) R46.
- [110] R. Sipos, A.J. Székely, M. Palatinszky, S. Révész, K. Márialigeti, M. Nikolausz, Effect of primer mismatch, annealing temperature and PCR cycle number on 16s rRNA gene-targeting bacterial community analysis, *FEMS Microbiol. Ecol.* 60 (2007) 341–350.
- [111] I. Abellan-Schneyder, M.S. Machado, S. Reitmeier, A. Sommer, Z. Sewald, J. Baumbach, M. List, K. Neuhaus, Primer, pipelines, parameters: issues in 16s rRNA gene sequencing, *mSphere* 6 (2021), <https://doi.org/10.1128/msphere.01202-01220> doi:10.1128/msphere.01202-20.
- [112] B.D. Badgley, M.K. Steele, C. Cappellin, J. Burger, J. Jian, T.P. Neher, M. Orentas, R. Wagner, Fecal indicator dynamics at the watershed scale: variable relationships with land use, season, and water chemistry, *Sci. Total Environ.* 697 (2019) 134113, <https://doi.org/10.1016/j.scitotenv.2019.134113>.
- [113] G. Wilkes, J. Brassard, T.A. Edge, V. Gannon, C.C. Jokinen, T.H. Jones, R. Marti, N.F. Neumann, N.J. Ruecker, M. Sunohara, Coherence among different microbial source tracking markers in a small agricultural stream with or without livestock exclusion practices, *Appl. Environ. Microbiol.* 79 (2013) 6207–6219.
- [114] S.K. Frey, E. Topp, T. Edge, C. Fall, V. Gannon, C. Jokinen, R. Marti, N. Neumann, N. Ruecker, G. Wilkes, Using swat, bacteroidales microbial source tracking markers, and fecal indicator bacteria to predict waterborne pathogen occurrence in an agricultural watershed, *Water Res.* 47 (2013) 6326–6337.
- [115] S. Zhang, S. Pang, P. Wang, C. Wang, C. Guo, F.G. Addo, Y. Li, Responses of bacterial community structure and denitrifying bacteria in biofilm to submerged macrophytes and nitrate, *Sci. Rep.* 6 (2016) 36178, <https://doi.org/10.1038/srep36178>.
- [116] X. Cao, Y. Wang, Y. Xu, G. Duan, M. Huang, J. Peng, Adaptive variations of sediment microbial communities and indication of fecal-associated bacteria to nutrients in a regulated urban river, *Water* 12 (2020) 1344.

- [117] L.H. Zeglin, Stream microbial diversity in response to environmental changes: review and synthesis of existing research, *Front. Microbiol.* 6 (2015) 454.
- [118] G. Wilkes, N.J. Ruecker, N.F. Neumann, V.P. Gannon, C. Jokinen, M. Sunohara, E. Topp, K.D. Pintar, T.A. Edge, D.R. Lapen, Spatiotemporal analysis of cryptosporidium species/genotypes and relationships with other zoonotic pathogens in surface water from mixed-use watersheds, *Appl. Environ. Microbiol.* 79 (2013) 434–448.
- [119] R. Marti, V.P. Gannon, C. Jokinen, M. Lanthier, D.R. Lapen, N.F. Neumann, N. J. Ruecker, A. Scott, G. Wilkes, Y. Zhang, Quantitative multi-year elucidation of fecal sources of waterborne pathogen contamination in the south nation river basin using bacteroidales microbial source tracking markers, *Water Res.* 47 (2013) 2315–2324.
- [120] L. Liu, M. Cloutier, E. Craiovan, M. Edwards, S.K. Frey, N. Gottschall, D.R. Lapen, M. Sunohara, E. Topp, I.U. Khan, Quantitative real-time pcr-based assessment of tile drainage management influences on bacterial pathogens in tile drainage and groundwater, *Sci. Total Environ.* 624 (2018) 1586–1597.
- [121] M.D. Sunohara, N. Gottschall, E. Craiovan, G. Wilkes, E. Topp, S.K. Frey, D. R. Lapen, Controlling tile drainage during the growing season in eastern canada to reduce nitrogen, phosphorus, and bacteria loading to surface water, *Agric. Water Manag.* 178 (2016) 159–170, <https://doi.org/10.1016/j.agwat.2016.08.030>.
- [122] N.K. Rideout, D.R. Lapen, D.L. Peters, D.J. Baird, Ditch the low flow: agricultural impacts on flow regimes and consequences for aquatic ecosystem functions, *Ecohydrology* 15 (5) (2022) e2364, <https://doi.org/10.1002/eco.2364>.
- [123] A. Jousset, C. Bienhold, A. Chatzinotas, L. Gallien, A. Gobet, V. Kurm, K. Küsel, M. C. Rillig, D.W. Rivett, J.F. Salles, Where less may be more: how the rare biosphere pulls ecosystems strings, *ISME J.* 11 (2017) 853–862.
- [124] M.D. Lynch, J.D. Neufeld, Ecology and exploration of the rare biosphere, *Nat. Rev. Microbiol.* 13 (2015) 217–229.
- [125] Y. Wang, F. Ye, S. Wu, J. Wu, J. Yan, K. Xu, Y. Hong, Biogeographic pattern of bacterioplanktonic community and potential function in the yangtze river: roles of abundant and rare taxa, *Sci. Total Environ.* 747 (2020) 141335, <https://doi.org/10.1016/j.scitotenv.2020.141335>.
- [126] K.M. Pärnänen, C. Narciso-da-Rocha, D. Kneis, T.U. Berendonk, D. Cacace, T. Do, C. Elpers, D. Fatta-Kassinos, I. Henriques, T. Jaeger, Antibiotic resistance in european wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence, *Sci. Adv.* 5 (2019) eaau9124.
- [127] S.D. Allison, J.B. Martiny, Resistance, resilience, and redundancy in microbial communities, *Proc. Natl. Acad. Sci. USA* 105 (2008) 11512–11519.
- [128] O.H.H.-L.E. Panel, D.T. Hayman, W.B. Adisasmito, S. Almuhairei, C.B. Behraves, P. Biliogui, S.A. Bukachi, N. Casas, N.C. Becerra, D.F. Charron, Developing one health surveillance systems, *One Health* 17 (2023) 100617.
- [129] S.J. Pitt, A. Gunn, The one health concept, *Br. J. Biomed. Sci.* 81 (2024) 12366.
- [130] G.G. Ozenen, Z.S. Bal, N.M. Bilen, S.Y. Arslan, S. Aydemir, Z. Kurugol, F. Ozkinay, The first report of sphingomonas yanoikuyae as a human pathogen in a child with a central nervous system infection, *Pediatr. Infect. Dis. J.* 40 (2021) e524.
- [131] K.B. Laupland, D.L. Paterson, A.G. Stewart, F. Edwards, P.N.A. Harris, Sphingomonas paucimobilis bloodstream infection is a predominantly community-onset disease with significant lethality, *Int. J. Infect. Dis.* 119 (2022) 172–177, <https://doi.org/10.1016/j.ijid.2022.03.060>.
- [132] A.L. Truant, R. Gulati, O. Giger, V. Satishchandran, J.G. Caya, Methylobacterium species: an increasingly important opportunistic pathogen, *Lab. Med.* 29 (1998) 704–710.
- [133] K. Gneiding, R. Frodl, G. Funke, Identities of microbacterium spp. Encountered in human clinical specimens, *J. Clin. Microbiol.* 46 (2008) 3646–3652.
- [134] E.C. Chern, L. Wymer, K. Brenner, K. Oshima, R.A. Haugland, Persistence of fecal indicator bacteria and associated genetic markers from wastewater treatment plant effluents in freshwater microcosms, *J. Water Health* 20 (2021) 205–215, <https://doi.org/10.2166/wh.2021.152>.
- [135] J.G. Bartlett, Clostridium difficile: history of its role as an enteric pathogen and the current state of knowledge about the organism, *Clin. Infect. Dis.* 18 (1994) S265–S272.
- [136] R.O.S. Silva, F.C.F. Lobato, Clostridium perfringens: a review of enteric diseases in dogs, cats and wild animals, *Anaerobe* 33 (2015) 14–17.
- [137] D.J. Trott, S. Abraham, B. Adler, Antimicrobial resistance in leptospira, brucella, and other rarely investigated veterinary and zoonotic pathogens, *Microbiol. Spectrum* 6 (2018), <https://doi.org/10.1128/microbiolspec.arba-0029-2017> doi: 10.1128/microbiolspec.arba-0029-2017.
- [138] J.K. Lama, D.S. Bachoon, Detection of brucella suis, campylobacter jejuni, and escherichia coli strains in feral pig (sus scrofa) communities of georgia, *Vector Borne Zoonotic Dis.* 18 (2018) 350–355.
- [139] Q. Zhang, A.F. Beyi, Y. Yin, Zoonotic and antibiotic-resistant campylobacter: a view through the one health lens, *One Health Adv.* 1 (2023) 4, <https://doi.org/10.1186/s44280-023-00003-1>.
- [140] P. Schmidt, K. Pintar, A. Fazil, C. Flemming, M. Lanthier, N. Laprade, M. Sunohara, A. Simhon, J. Thomas, E. Topp, Using campylobacter spp. And escherichia coli data and bayesian microbial risk assessment to examine public health risks in agricultural watersheds under tile drainage management, *Water Res.* 47 (2013) 3255–3272.
- [141] I.U.H. Khan, V. Gannon, C.C. Jokinen, R. Kent, W. Koning, D.R. Lapen, D. Medeiros, J. Miller, N.F. Neumann, R. Phillips, et al., A national investigation of the prevalence and diversity of thermophilic campylobacter species in agricultural watersheds in canada, *Water Res.* 61 (2014) 243–252, <https://doi.org/10.1016/j.watres.2014.05.027>.
- [142] M. Thakur, S. Dolker, L.K. Wangmo, A. Ghosh, N. Dhankhar, V.K. Singh, M. Shukla, A. Rameshkumar, M. Biswal, D. Banerjee, Zoonotic surveillance of bartonella spp.: exploring the public health risks in human settlements, *Zoonoses Public Health* 72 (2025) 284–292.
- [143] A. Cevidan, L. Altet, A.D. Chirife, T. Proboste, J. Millán, Drivers of bartonella infection in micromammals and their fleas in a mediterranean peri-urban area, *Vet. Microbiol.* 203 (2017) 181–188.
- [144] L.O. Rouffier, D. Strubbe, A. Teyssier, N. Sallé Hudin, A.-M. Van den Abeele, I. Cox, R. Haesendonck, M. Delmée, F. Haesebrouck, F. Pasmans, Effects of urbanization on host-pathogen interactions, using yersinia in house sparrows as a model, *PLoS One* 12 (2017) e0189509.
- [145] R.D. Arnone, J. Perdek Walling, Waterborne pathogens in urban watersheds, *J. Water Health* 5 (2007) 149–162.
- [146] G. Wilkes, J. Brassard, T.A. Edge, V. Gannon, C.C. Jokinen, T.H. Jones, N. Neumann, K.D.M. Pintar, N. Ruecker, P.J. Schmidt, et al., Bacteria, viruses, and parasites in an intermittent stream protected from and exposed to pasturing cattle: prevalence, densities, and quantitative microbial risk assessment, *Water Res.* 47 (2013) 6244–6257, <https://doi.org/10.1016/j.watres.2013.07.041>.
- [147] S. Mindlin, I. Bass, E. Bogdanova, Z.M. Gorlenko, E. Kalyaeva, M. Petrova, V. Nikiforov, Horizontal transfer of mercury resistance genes in environmental bacterial populations, *Mol. Biol.* 36 (2002) 160–170.
- [148] G. Liu, L.E. Thomsen, J.E. Olsen, Antimicrobial-induced horizontal transfer of antimicrobial resistance genes in bacteria: a mini-review, *J. Antimicrob. Chemother.* 77 (2022) 556–567.
- [149] C. Pal, J. Bengtsson-Palme, E. Kristiansson, D.J. Larsson, Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential, *BMC Genomics* 16 (2015) 964.
- [150] M.W. Van Goethem, R. Pierneef, O.K. Bezuidt, Y. Van De Peer, D.A. Cowan, T. P. Makhallanyane, A reservoir of 'historical' antibiotic resistance genes in remote pristine antarctic soils, *Microbiome* 6 (2018) 1–12.
- [151] B.F. Gillieatt, N.V. Coleman, Unravelling the mechanisms of antibiotic and heavy metal resistance co-selection in environmental bacteria, *FEMS Microbiol. Rev.* 48 (2024), <https://doi.org/10.1093/femsre/fuae017>.
- [152] L.M. Murray, A. Hayes, J. Snape, B. Kasprzyk-Hordern, W.H. Gaze, A.K. Murray, Co-selection for antibiotic resistance by environmental contaminants, *npj Antimicrob. Resist.* 2 (2024) 9, <https://doi.org/10.1038/s44259-024-00026-7>.
- [153] S. Mokni-Tlili, S. Hechmi, H.-I. Ouzari, N. Mechergui, M. Ghorbel, N. Jedidi, A. Hassen, H. Hamdi, Co-occurrence of antibiotic and metal resistance in long-term sewage sludge-amended soils: influence of application rates and pedo-climatic conditions, *Environ. Sci. Pollut. Res.* 30 (2023) 26596–26612, <https://doi.org/10.1007/s11356-022-23802-2>.
- [154] C. Valadez-Cano, I.U.H. Khan, J. Chuan, X. Li, D.R. Lapen, S. Gerdis, J. Dettman, W. Chen, Draft genomes of multidrug-resistant acinetobacter baumannii isolates from an agriculturally-dominated watershed and dairy cattle feces, *Microbiol. Resour. Announc.* 0 (2025) e00829-25, <https://doi.org/10.1128/mra.00829-25>.
- [155] I.U. Khan, W. Chen, M. Cloutier, D.R. Lapen, E. Craiovan, G. Wilkes, Pathogenicity assessment of acrobacter butzleri isolated from canadian agricultural surface water, *BMC Microbiol.* 24 (2024) 17.
- [156] A. Crits-Christoph, S. Diamond, C.N. Butterfield, B.C. Thomas, J.F. Banfield, Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis, *Nature* 558 (2018) 440–444.
- [157] C. Vilchère, V. Molle, S. Carrère-Kremer, J. Leiba, L. Mourey, S. Shenai, G. Baronian, J. Tufariello, T. Hartman, R. Veyron-Churlet, et al., Phosphorylation of kasb regulates virulence and acid-fastness in mycobacterium tuberculosis, *PLoS Path.* 10 (2014) e1004115, <https://doi.org/10.1371/journal.ppat.1004115>.
- [158] Y. Ma, J.A. Mills, J.T. Belisle, V. Vissa, M. Howell, K. Bowlin, M.S. Scherman, M. McNeil, Determination of the pathway for rhamnose biosynthesis in mycobacteria: cloning, sequencing and expression of the mycobacterium tuberculosis gene encoding α -d-glucose-1-phosphate thymidyltransferase, *Microbiology* 143 (1997) 937–945, <https://doi.org/10.1099/00221287-143-3-937>.
- [159] G.A. Somerville, R.A. Proctor, At the crossroads of bacterial metabolism and virulence factor synthesis in staphylococci, *Microbiol. Mol. Biol. Rev.* 73 (2009) 233–248, <https://doi.org/10.1128/mmbr.00005-09>.
- [160] E.M. Wellington, A.B. Boxall, P. Cross, E.J. Feil, W.H. Gaze, P.M. Hawkey, A. S. Johnson-Rollings, D.L. Jones, N.M. Lee, W. Otten, The role of the natural environment in the emergence of antibiotic resistance in gram-negative bacteria, *Lancet Infect. Dis.* 13 (2013) 155–165.
- [161] J.A. Perry, G.D. Wright, The antibiotic resistance “mobilome”: searching for the link between environment and clinic, *Front. Microbiol.* 4 (2013) 138.
- [162] U. Klümper, P. Fang, B. Li, Y. Xia, D. Frigon, K.A. Hamilton, H. Quon, T. U. Berendonk, M. de la Cruz Barron, Towards the integration of antibiotic resistance gene mobility into environmental surveillance and risk assessment, *npj Antimicrob. Resist.* 3 (2025) 81, <https://doi.org/10.1038/s44259-025-00154-8>.
- [163] E.A. Franzosa, L.J. McIver, G. Rahnnavard, L.R. Thompson, M. Schirmer, G. Weingart, K.S. Lipson, R. Knight, J.G. Caporaso, N. Segata, Species-level functional profiling of metagenomes and metatranscriptomes, *Nat. Methods* 15 (2018) 962–968.
- [164] F. Beghini, L.J. McIver, A. Blanco-Míguez, L. Dubois, F. Asnicar, S. Maharjan, A. Mailyan, P. Manghi, M. Scholz, A.M. Thomas, Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with biobakery 3, *elife* 10 (2021) e65088.
- [165] W. Ahmed, K. Hamilton, S. Toze, S. Cook, D. Page, A review on microbial contaminants in stormwater runoff and outfalls: potential health risks and mitigation strategies, *Sci. Total Environ.* 692 (2019) 1304–1321, <https://doi.org/10.1016/j.scitotenv.2019.07.055>.

- [166] D.R. Urycki, S.P. Good, B.C. Crump, N.C. Ceperley, J.R. Brooks, Microbial community storm dynamics signal sources of “old” stream water, *PLoS One* 19 (2024) e0306896, <https://doi.org/10.1371/journal.pone.0306896>.
- [167] K.E. Williamson, J.V. Harris, J.C. Green, F. Rahman, R.M. Chambers, Stormwater runoff drives viral community composition changes in inland freshwaters, *Front. Microbiol.* 5 (2014) 105, <https://doi.org/10.3389/fmicb.2014.00105>.
- [168] A.B. Boehm, L.C. Van De Werfhorst, J.F. Griffith, P.A. Holden, J.A. Jay, O. C. Shanks, D. Wang, S.B. Weisberg, Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study, *Water Res.* 47 (2013) 6812–6828.
- [169] L. Shenhav, M. Thompson, T.A. Joseph, L. Briscoe, O. Furman, D. Bogumil, I. Mizrahi, I. Pe'er, E. Halperin, Feast: fast expectation-maximization for microbial source tracking, *Nat. Methods* 16 (2019) 627–632.