

Assessing sequencing-based pathogen surveillance of a recreational swimming area in Oslo, Norway

Vegard Eldholm^{1,*}, Daniel Straume² and Ola B. Brynildsrud¹

Abstract

Sequencing-based surveillance can enable rapid and sensitive detection of environmental pathogens. The Oslofjord inlet is relatively narrow and is exposed to substantial human activity, including occasional wastewater contamination. Restricted water exchange also allows for occasional summer heat spells with elevated water temperatures. Thus, infections stemming from wastewater contamination and seasonal opportunistic pathogens are potential health threats to recreational users of the fjord. In this pilot study, we assess the suitability of sequencing-based surveillance for the detection of pathogens at a popular urban location for recreational water activities, employing both long- and short-read sequencing platforms, paired with selective culturing. We find both metagenomic and full-length 16S sequencing to be promising tools for surveillance of seasonal opportunistic *Vibrio* pathogens. Furthermore, we identified *Rhodferax* abundance to be a potentially attractive indicator of sewage contamination using low to medium-depth full-length 16S sequencing. Selective plating revealed minimal abundance of culturable extended-spectrum β -lactam-resistant bacteria, of which none were detected by metagenomic sequencing. Metagenomic analyses did, however, pick up several other β -lactamases in various bacterial taxa, including some that were closely related to those identified by selective plating and sequencing.

DATA SUMMARY

Fastq data are available from the European Nucleotide Archive study PRJEB88820. Figshare datasets (see Tables 1–3 for a sample-wise overview) are available at: <https://doi.org/10.6084/m9.figshare.29234570.v1> and <https://doi.org/10.6084/m9.figshare.29234627.v1>.

In addition, assemblies of three ESBL isolates are available at: <https://github.com/vehuardo/AquaGenomics/>.

INTRODUCTION

Sequencing-based surveillance holds potential for rapid and sensitive detection of pathogens in the environment. Among current sequencing platforms, third-generation nanopore sequencing in particular comes with major advantages: (1) the potential generation of long reads, containing comparatively more information than shorter reads, (2) the production of sequencing results in real time and (3) the availability of light-weight field deployable platforms.

Norway has a temperate but relatively cool climate, which limits the growth of many environmental pathogens. However, during particularly warm summers, increased frequencies of opportunistic *Shewanella* and *Vibrio* infections associated with outdoor water activities have been reported [1]. Furthermore, the inner Oslofjord regularly experiences sewage contamination following bursts of heavy rainfall overwhelming the wastewater treatment systems. In Oslo, a relatively simple surveillance system relying on Entrolert (Idexx), which quantifies enterococci based on substrate metabolism, is used to survey sewage contamination in recreational swimming areas. Yet, the advice to the public is largely empirical, and the city generally advises against swimming

Access Microbiology is an open research platform. Pre-prints, peer review reports, and editorial decisions can be found with the online version of this article. Received 04 June 2025; Accepted 29 January 2026; Published 27 February 2026

Author affiliations: ¹Division of Infection Control, Norwegian Institute of Public Health, Lovisenberggata 6, 0456, Oslo, Norway; ²Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Science, 1430, Ås, Norway.

***Correspondence:** Vegard Eldholm, Vegard.Eldholm@fhi.no

Keywords: eDNA; environmental pathogens; full-length 16S; Illumina sequencing; metagenomics; nanopore sequencing; *Rhodferax*; sewage contamination; *Vibrio*.

Abbreviations: AMR, antimicrobial resistance; ARGs, antimicrobial resistance genes; eDNA, environmental DNA; ESBL, extended-spectrum β -lactamase; Gbp, giga base pairs; GTDB, Genome Taxonomy Database; NCBI, National Center for Biotechnology Information; ONT, Oxford Nanopore Technologies; PCA, plate count agar; SUP, super accurate; TCBS, thiosulfate-citrate-bile salts-sucrose.

001062.v3 © 2026 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

Table 1. List of European Nucleotide Archive accessions for metagenomic eDNA sequence data generated on the Illumina NextSeq platform

Location	Sampling date	Study accession	Run accession
Operastranda	28-05-2024	ERP171883	ERR14872774
Operastranda	04-06-2024	ERP171883	ERR14872775
Operastranda	18-06-2024	ERP171883	ERR14872776
Operastranda	02-07-2024	ERP171883	ERR14872777
Operastranda	09-07-2024	ERP171883	ERR14872778
Operastranda	06-08-2024	ERP171883	ERR14872779
Operastranda	20-08-2024	ERP171883	ERR14872780
Huk	04-06-2024	ERP171883	ERR14872781
Huk	02-07-2024	ERP171883	ERR14872782

within 24 h of particularly heavy rainfall. Beyond coliform bacteria, there is no system in place for the detection of rarer seasonal pathogens, such as *Vibrio* spp.

To assess the potential of sequencing-based environmental surveillance for the detection of human pathogens, a pilot study was designed for the surveillance of a popular recreational swimming area in central Oslo. Classical selective culturing, full-length 16S and metagenomic sequencing on both the Oxford Nanopore (ONT) and Illumina platforms were conducted to assess the power of sequencing-based methods to detect coliform bacteria, *Vibrio* spp. and extended-spectrum β -lactamase (ESBL) producing bacteria.

METHODS

Sampling and culturing

One-litre samples were retrieved from Operastranda, a popular recreational swimming area in the centre of Oslo, at semi-regular intervals. The location is in the immediate vicinity of the outlet of the river Akerselva, occasionally contaminated by wastewater following the intense precipitation. Two samples were also taken from a control location, Huk, which is another popular recreational swimming area further removed from wastewater contaminants. The sampling was conducted from land, with sampling bottles fully immersed at ~50 cm depth.

Volumes of 1 ml and 100 μ l of the samples were plated on the following media: Plate count agar (Merck #1462690020), Thiosulfate-citrate-bile salts-sucrose (TCBS) agar (Merck #86348-500G), ESBL ChromoSelect Agar (Merck: agar base #55806, supplement #61471) and ReadyPlate™ Chromocult Coliform Agar (Merck). The plates were incubated for 18–24 h at 37 °C with 5% CO₂. Colonies growing on ESBL ChromoSelect agar plates were restreaked and incubated for another 18–24 h.

Growth on ESBL re-streak plates, Chromocult Coliform agar and TCBS agar was harvested ('swept') using 10 μ l inoculation loops and frozen before subsequent DNA extraction and sequencing.

Table 2. Full-length 16S data generated from eDNA are available at <https://doi.org/10.6084/m9.figshare.29234570.v1>

Location	Sampling date	File name
Huk	04-06-2024	Huk_06-04.fastq
Huk	02-07-2024	Huk_07-02.fastq
Operastranda	28-05-2024	Opera_05-28.fastq
Operastranda	04-06-2024	Opera_06-04.fastq
Operastranda	18-06-2024	Opera_06-18.fastq
Operastranda	02-07-2024	Opera_07-02.fastq
Operastranda	09-07-2024	Opera_07-09.fastq
Operastranda	06-08-2024	Opera_08-06.fastq
Operastranda	20-08-2024	Opera_08-20.fastq

Table 3. Plate-sweep sequencing data are available at <https://doi.org/10.6084/m9.figshare.29234627.v1>

Location	Sampling date	Agar plate	File name
Operastranda	28-05-2024	Chromocult Coliform (Merck)	CO1_min1000.fastq.tar.gz
Operastranda	04-06-2024	Chromocult Coliform (Merck)	CO2_min1000.fastq.tar.gz
Operastranda	18-06-2024	Chromocult Coliform (Merck)	CO3_min1000.fastq.tar.gz
Operastranda	02-07-2024	Chromocult Coliform (Merck)	CO4_min1000.fastq.tar.gz
Operastranda	06-08-2024	Chromocult Coliform (Merck)	CO6_min1000.fastq.tar.gz
Operastranda	20-08-2024	Chromocult Coliform (Merck)	CO7_min1000.fastq.tar.gz
Operastranda	18-06-2024	ESBL ChromoSelect (Merck)	E01_min1000.fastq.tar.gz
Operastranda	02-07-2024	ESBL ChromoSelect (Merck)	E02_min1000.fastq.tar.gz
Operastranda	08-06-2024	ESBL ChromoSelect (Merck)	E03_min1000.fastq.tar.gz
Operastranda	28-05-2024	TCBS agar (Merck)	T01_min1000.fastq.tar.gz
Operastranda	04-06-2024	TCBS agar (Merck)	T02_min1000.fastq.tar.gz
Operastranda	18-06-2024	TCBS agar (Merck)	T03_min1000.fastq.tar.gz
Operastranda	02-07-2024	TCBS agar (Merck)	T04_min1000.fastq.tar.gz
Huk	04-06-2024	TCBS agar (Merck)	TH2_min1000.fastq.tar.gz
Huk	02-07-2024	TCBS agar (Merck)	TH4_min1000.fastq.tar.gz
Operastranda	09-07-2024	TCBS agar (Merck)	TO5_min1000.fastq.tar.gz
Operastranda	06-08-2024	TCBS agar (Merck)	TO6_min1000.fastq.tar.gz
Operastranda	20-08-2024	TCBS agar (Merck)	TO7_min1000.fastq.tar.gz

DNA extraction

DNA was extracted from the water samples on the same day as sampling, immediately after plating. The water was filtered through Sterivex 0.22 µm filters (Merck #SVGPL10RC) with the aid of a faucet vacuum aspirator. Filtering was stopped if the filter showed signs of clogging (very low rate of flow-through); otherwise, the full ~1 l volume was filtered per sample. DNA was extracted from the filter using the ZymoBIOMICS DNA Miniprep Kit (Zymo #D4300), with the first steps modified as follows: the full content of individual bashing bead tubes was poured into the Sterivex filter through the inlet opening. Subsequently, 750 µl lysis buffer was added to the filter, and the filter was capped with luer combi-locks (B. Braun) at both inlet and outlet openings. The Sterivex filters were subsequently vortexed at maximum intensity on a Vortex-Genie2 fitted with a vortex adapter for 5–15-ml tubes (QIAGEN #13000-V1-5) for 20 min. After vortexing, the sample with bashing beads and lysis buffer was expelled through the inlet opening into a clean Eppendorf tube by applying pressure with a 1-ml pipette at the outlet opening. The remaining steps followed the ZymoBIOMICS DNA Miniprep manual. The DNA concentration in the final extracts ranged between 13 and 100 ng µl⁻¹ in 50 µl. A single control extraction without added water was performed but failed to result in detectable DNA (<0.05 ng µl⁻¹) and was thus not taken further in downstream sequencing steps. DNA extraction from plate-sweeps was performed following the standard ZymoBIOMICS DNA Miniprep manual in full.

DNA sequencing

Full-length 16S sequencing of environmental DNA

Full-length 16S amplicons were generated from environmental total DNA extracts from the water samples using the modified versions of the forward and reverse primers from Matsuo *et al.* [2]. The template targeting portion of the above primers was not modified, but the so-called anchor sequences were excluded. Instead, forward primers contained a short linker sequence (GATC), followed by a barcode for demultiplexing, followed by the primer binding sequence. The reverse primer had no barcode and was the same for all multiplexed samples (Table 4).

PCR was performed using LongAmp *Taq* 2X master mix (New England Biolabs M0287S) in 25 µl reactions containing 0.4 µM of each primer and 50–100 ng template DNA. PCR cycling conditions were as follows: 94 °C for 3 min, followed by 18 cycles of 15 s at 94 °C, 15 s at 52 °C and 1 min 15 s at 65 °C, followed by a final extension step of 10 min at 65 °C. PCR products were purified using AMPure XP beads (Beckman Coulter A63881), quantified with Qubit 1X dsDNA HS kits (ThermoFisher Q33266) and pooled before library preparation with the Ligation Sequencing Kit V14 (Oxford Nanopore SQK-LSK114), following the Flongle branch

Table 4. 16S primers

Primer	Sequence (5'-->3')
BC1_16S_F	GATCGGAGAAGAAGAAAGRGT ^{TT} YGATYMTGGCTCAG
BC2_16S_F	GATCAAGAGGCAAGAAAGRGT ^{TT} YGATYMTGGCTCAG
BC3_16S_F	GATCTCCTCCTAAGAAAGRGT ^{TT} YGATYMTGGCTCAG
Universal_16S_R	CGGYTACCTTGT ^T TACGACTT

The underlined sequences indicate unique barcodes for demultiplexing.

of the manufacturer's protocol. The libraries were sequenced in pools of three on three individual R10.4.1 Flongle flow cells and basecalled using the super accurate (SUP) model. The reads were demultiplexed using a Python script (demultiplex_AquaGenomics_amplicons.py) generated to recognize the in-house barcodes, available at <https://github.com/vehuardo/AquaGenomics>.

Metagenomic sequencing of environmental DNA

All environmental DNA (eDNA) samples ($n=9$) were also subjected to metagenomic sequencing on a single Illumina NextSeq 500 mid-output flowcell. One hundred fifty bp paired-end reads were trimmed using Trimmomatic v0.39 [3]. After trimming, 19–41 million paired-end reads remained per sample, translating to ~2.8–6.1 giga base pairs (Gbp) per sample.

A single sample, Operastranda1, was additionally sequenced on a single MinIon R10.4.1 flow cell and basecalled using the SUP model. The run generated 1,341,875 reads, making up 5.6 Gbp of sequence data. FiltLong (<https://github.com/rrwick/Filtlong>) was used to retain reads $\geq 1,000$ nt long only, resulting in 1,042,861 reads making up 5.4 Gbp.

Plate-sweep metagenomic sequencing

Sequencing libraries were generated from DNA extracted from individual ChromoCult Coliform, TCBS and ESBL ChromoSelect agar plate-sweeps using the Rapid Barcoding Kit 24 V14 (Oxford Nanopore SQK-RBK114.24) following the manufacturer's protocol. These libraries were sequenced on MinIon R10.4.1 flow cells and basecalled using the SUP model.

Sequencing-based profiling

Full-length 16S reads were generated on the ONT platform and profiled using Emu [4] against the default database (generated on 18 August 2022), while specifying 'map-ont' as read type and otherwise using default settings. Individual Emu reports were collected in a single file using the script 'collect_emu_outputs.sh' available on (<https://github.com/vehuardo/AquaGenomics>). 16S full-length reads were filtered to only retain reads 1,300–1,600 nt long, using FiltLong (<https://github.com/rrwick/Filtlong>), whereas Illumina short reads were end-trimmed using Trimmomatic v0.39 [3] with default settings.

Metagenomic Illumina sequences were initially profiled using Kraken2 v 2.0.8-beta, using the 'plusfpf_08_gb_20240112' database [5].

ONT sequencing reads generated from the plate sweeps were profiled using Sourmash [6]. Pre-prepared databases of bacterial, archaeal, viral, protozoan and fungal sequences from Genbank (March 2022) of kmer size 31 were downloaded from <https://sourmash.readthedocs.io/en/latest/databases.html#genbank-genomes-from-march-2022>. Metagenomic sequences to be profiled were sketched using the command `sourmash sketch dna -p k=31,scaled=1000,abund --name-from-first [input].fastq -o [output].sig`

The sig files were subsequently screened against the five databases using `sourmash gather: [input].sig genbank-2022.03-*k31.zip --threshold-bp 1000 -o [output].gather.csv`

Finally, species-level taxonomic summaries were generated for each sample, while retaining higher-order taxonomic information, using the script `sum_tax_GBDB.py` (retrievable at <https://github.com/vehuardo/AquaGenomics/>).

In addition to direct read-profiling using Sourmash, ONT long reads generated from TCBS plate-sweeps were assembled using Flye v.2.9.5 [7] with the -meta option. The contigs were blasted against *Vibrio cholerae* strain RFB16 (NZ_CP043554.1) using BLASTN [8] to identify hits. Subsequently, the best hits were manually searched using National Center for Biotechnology Information (NCBI) nucleotide web BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the nt database, to establish whether *V. cholerae* was indeed the closest hit. ONT reads were also mapped directly against the genes *ctxA* and *ctxB* (*ctxAB* extracted as a single sequence from NC_015209.1), encoding the cholera toxin subunits using bwa v0.7.18.

Sequences generated from re-streaked ESBL-positive colonies ($n=3$) were assembled using Flye v.2.9.5 [7]. For species identification, the assemblies were screened to identify the most closely related genome in Genome Taxonomy DB (GTDB) using the

ani_rep function of GTDB-Tk v2 [9]. Candidate genes for conferring ESBL phenotypes were identified using AMRFinderPlus v4.0.3 [10].

Metagenomic identification of antimicrobial resistance genes

Reads were washed for phiX using bowtie2 v2.4.2 [11], trimmed for adapters with fastp v0.23.4 [12], and each sample assembled with metaSPAdes v4.0.0. Binning into pseudo-genomes was done with MetaBat2 v2.15 [13], and bin completeness was assessed using BUSCO v5.8.2 [14] in genome mode and using auto-lineage. Prokka v1.14.6 [15] was used to predict and annotate ORFs. Finally, sequence bins and proteins were screened for antimicrobial resistance genes using NCBI AMRFinderPlus v4.0.3 (database built 2024-10-22 [10]). Only hits annotated as class and subclass ‘AMR’ were reported, which excludes stress and metal-resistance-associated genes. Additionally, we searched for ARGs at the read level against the CARD database [16] using the RGI v6.0.6 subcommand bwt with default settings (<https://github.com/arpcard/rgi>). To filter out spurious results, we only report ARGs with 90% or higher coverage across their gene length.

RESULTS

Environmental parameters and sampling

Hourly water body temperatures, measured at 60 cm depth by means of a buoy located ~400 m away from the actual sampling location of Operastranda, were retrieved from <https://badetassen.no/>. Hourly precipitation measured at the main meteorological station at Blindern, Oslo, was retrieved from the meteorological institute (<http://www.met.no>). Sampling at Operastranda was conducted at seven time points from late May to late August (Fig. 1). In addition, a control location (‘Huk’) was sampled twice. The control location is further removed from major sources of sewage contamination and is generally characterized by better water quality compared to inner-city swimming areas such as Operastranda.

The variation in water body temperature was relatively modest during the summer of 2024, which was not ideal for the detection of potential pathogens requiring high temperatures. The first sampling was conducted <12 h after a heavy downpour, which typically leads to sewage contamination of the inner fjord basin. None of the other samplings was immediately preceded by heavy precipitation.

eDNA profiling

eDNA was extracted from the water samples as described in the ‘Methods’ section and characterized by metagenomic sequencing on the Illumina platform and full-length 16S sequencing on the ONT platform. The 16S sequencing was relatively shallow, with a median of 74,381 full-length 16S reads (11,252–102,549) per sample. The metagenomic sequencing generated 19–41 million paired-end reads per sample.

The 16S reads were profiled using Emu [4] with the default database, whereas the Illumina metagenomic data were profiled using Kraken2 [5] with the PlusPFP database.

The profiling results differed substantially between the two approaches (Fig. 2). Across all time points, the five most abundant genera at the main location (Operastranda) were *Hydrogenophaga*, *Flavobacterium*, *Rhodoferrax*, *Pseudorhodobacter* and *Sediminicola* based on 16S sequencing and *Pelagibacter*, *Flavobacterium*, *Thalassiosira*, *Lentibacter* and *Synechococcus* based on

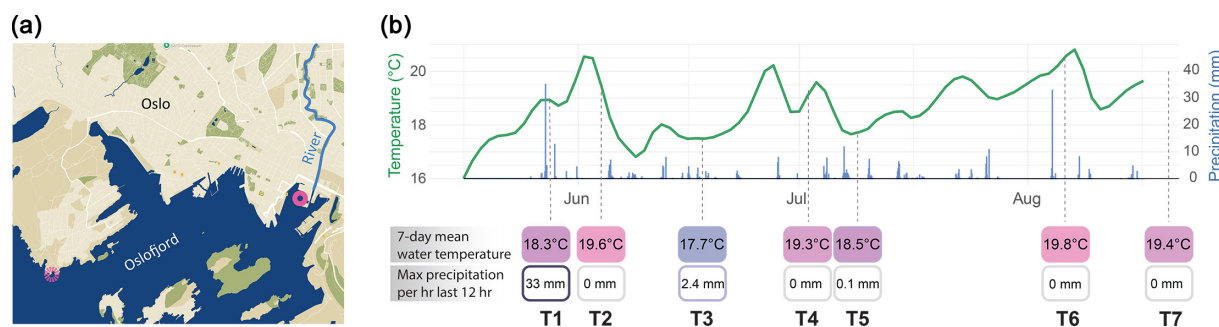


Fig. 1. Sampling location and environmental parameters. (a) The primary sampling location, Operastranda, is highlighted by a pink circle; the secondary control location, Huk, further removed from the inner city, is marked with a lighter dotted circle. The Akerselva river has also been highlighted on the map. (b) Smoothed hourly water body temperatures and hourly precipitation at Operastranda. The boxes below the graph indicate the seven sampling time points (T1–T7) and summarize the average water temperature in the 7-day periods leading up to sampling, as well as maximum hourly precipitation in the 12-h period leading up to sampling.



Fig. 2. Abundance estimates based on full-length 16S ONT reads (top) and Illumina metagenomics (bottom). Only genera with abundance >2% at a minimum of one time point were assigned individual colours.

metagenomic characterization. That is, the only common genus among the top five genera was *Flavobacterium*, but it should be noted that *Thalassiosira*, picked up by metagenomic sequencing, is a eukaryotic diatom which is not detectable by 16S sequencing. When extending the comparison to the top 20 genera identified with each method, 4 were overlapping, namely *Flavobacterium*, *Limnohabitans*, *Marivivens* and *Rhodoferrax*.

Plate-sweep sequencing

In parallel with direct eDNA extraction and sequencing of the water samples, samples were plated on plate count agar (PCA), ChromoCult agar and TCBS agar, in order to obtain comparable abundance estimates of total viable bacteria, coliform bacteria and *Vibrio* bacteria, respectively, at each time point. Following counting, the growth on each plate was harvested and underwent metagenomic plate-sweep sequencing on the ONT platform.

Growth on ChromoCult coliform plates was dominated by *Aeromonas*, *Enterobacter*, *Escherichia*, *Pseudomonas* and *Klebsiella*, whereas growth on TCBS plates was dominated by *Vibrio* and *Aeromonas*. Substantial variation in taxon composition was observed between time points on both growth media (Fig. 3). As we were particularly interested in seasonal pathogens belonging to the genus *Vibrio*, we separately plotted the *Vibrio* species composition on TCBS plates. *Vibrio anguillarum*, *Vibrio diazotrophicus* and *Vibrio navarrensis* were the most abundant species overall, with *V. navarrensis* as the overall most common across all time points.

At time point 4, 3.3% of the reads generated from the TCBS plate were assigned to *V. cholerae*. Although *V. cholerae* has been detected in Nordic countries, including sewage samples from Copenhagen, Denmark [17] and Norwegian blue mussels [18], this was a surprising finding. We thus performed some additional control steps to assess the veracity of the hit. Following metagenomic assembly, we observed that the intersection between the plate-sweep assemblies and the two *V. cholerae* strains in the Sourmash gather output was among the shortest of all intersections (290 kbp and 310 kbp vs. median intersect of 2.76 Mbp). A blast search of the plate-sweep metagenomic assembly, containing 3569 contigs with an N50 of 78,580 bp, against a *V. cholerae* reference genome, resulted in numerous hits, with a maximum hit length of 31,603 bp. A manual web blast of this contig against the NCBI nt database confirmed *V. cholera* as the closest hit with a 98.25% sequence identity. None of the ONT reads mapped against the cholera toxin subunit genes *ctxAB*. Taken together, our results suggest the possible presence of nontoxigenic *V. cholerae*, or perhaps more likely, a closely related undescribed *Vibrio* sp.

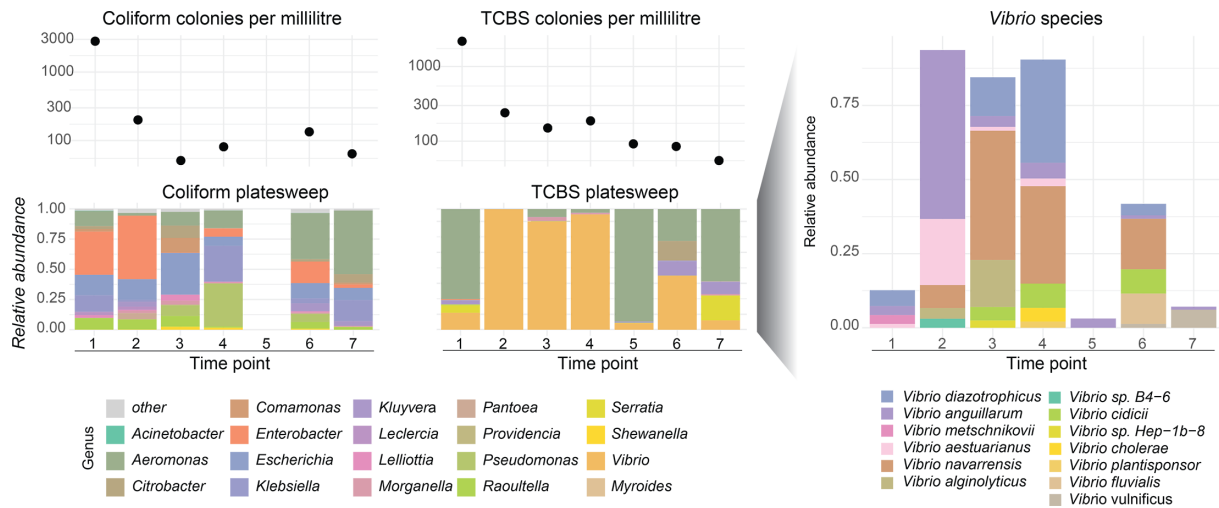


Fig. 3. Plate-sweep sequencing results from location Operastranda. From the first and second panels, visualize colony counts and genus abundance of growth on coliform and TCBS plates, respectively. The rightmost panel visualizes the relative abundance of *Vibrio* species only from the TCBS plates. Millilitre is abbreviated 'ml'.

Sequencing-based detection of *Vibrio* and coliforms in eDNA

Next, we investigated the relative abundance of *Vibrio* and coliform genera in the metagenomic eDNA sequences. Again, we analysed both full-length 16S ONT sequences and Illumina metagenomic sequences. The abundances of viable bacteria in general, and coliforms specifically, were $>10\times$ higher at the urban location Operastranda compared to the control location Huk (Fig. 4). Despite the high abundance of coliforms at the first sampling time point ($2,820\text{ colonies ml}^{-1}$), the total relative abundance of coliforms (*Citrobacter*, *Enterobacter*, *Escherichia* and *Klebsiella*) as assessed by metagenomic sequencing was only 0.275% in the sample (Fig. 4b). The abundance was even lower at the subsequent time points, as expected. The abundance of coliforms was below the detection level of the full-length 16S sequencing, which is probably at least partially related to the low depth of sequencing.

Vibrio spp. were detected by metagenomic sequencing throughout the study period, at $\sim 0.35\text{--}1.0\%$ abundance, across both locations (Fig. 4b). The relatively stable abundance of *Vibrio* spp. is likely a result of relatively stable waterbody temperatures during the period.

Indicator taxa for the detection of wastewater contamination

Relative coliform abundance was well below 1%, as assessed by eDNA sequence analysis, even when sewage contamination was substantial, as assessed by culturing. We thus set out to identify taxa that could serve as indicators of sewage contamination. We started with the 16S sequencing data, as the full-length 16S approach is both affordable and, to some extent, field deployable. A simple Pearson correlation with genus abundance and c.f.u. ml^{-1} on coliform agar as the two variables resulted in four genera

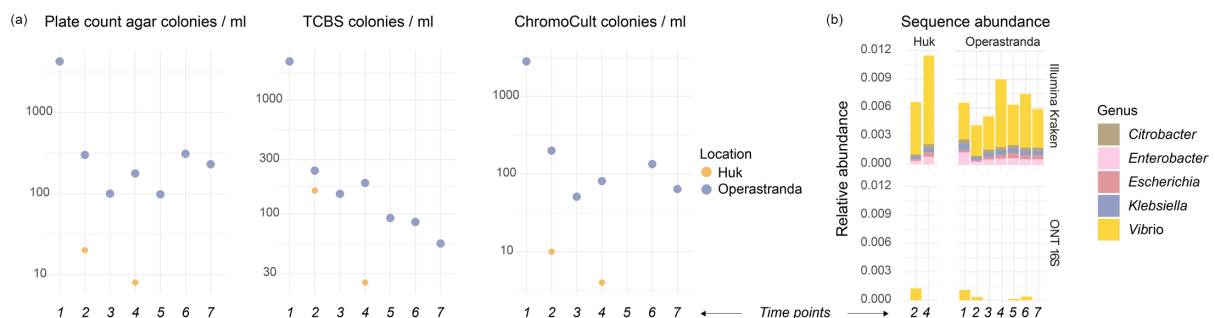


Fig. 4. (a) Number of colonies on PCA, TCBS agar and ChromoCult coliform agar per millilitre of seawater for each time point. (b) Relative abundance of *Vibrio* and coliform genera as determined by Illumina sequencing (profiled using Kraken and the PlusPFP database) and full-length 16S sequencing on the ONT platform, profiled with Emu with the default database.

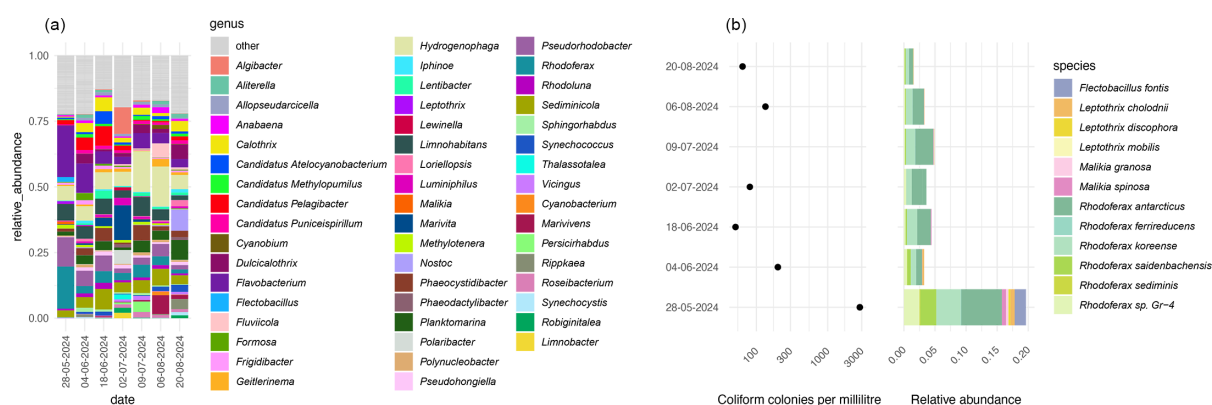


Fig. 5. Genus-level 16S profiles and identification of indicator species for sewage contamination (a) Bacterial abundance at Operastranda during the summer of 2024 (7 time points). Genera not reaching a minimum of 1% abundance at at least one time point were combined in the group 'other'. (b) Coliform bacterial count per millilitre (note log-scale) and abundance of species belonging to genera whose abundance was found to correlate with coliform counts.

exhibiting correlation coefficients >0.95 , namely *Flectobacillus*, *Malikia*, *Rhodoferrax* and *Leptothrix*. Among these potential indicator genera, *Rhodoferrax* was the most abundant (Fig. 5).

That *Rhodoferrax* abundance could serve as an indicator of sewage contamination was also supported by independent metagenomic profiling, which estimated substantially higher abundance of the genus in the first sampling compared to all other time points (Fig. 2). Similar patterns were observed for *Leptothrix* but at much lower abundance (maximum $\sim 1\%$ abundance with 16S profiling and 0.13% abundance with metagenomic profiling), whereas *Malikia* and *Flectobacillus* were undetected with metagenomic sequencing.

Antibiotic resistance

Water samples were plated on ESBL ChromoSelect Agar in order to identify and characterize ESBL bacteria. Across the seven time points and two locations, a total of three colonies grew from 1 ml of water samples. The colonies were cultivated further and sequenced on the ONT platform. The three strains belonged to the species *Pseudomonas tohonis*, *Flavobacterium odoratimimum* and *Acinetobacter venetianus*. In each of the strains, a single β -lactamase gene was identified, of which two were annotated as carbapenem resistance genes.

Independently, eDNA sequences were assembled and searched for resistance genes using AMRFinderPlus [10]. This search was not restricted to ESBL genes. To maximize sensitivity, sequencing reads were also screened against the CARD database using RGI. This resulted in 1628 hits with at least one mapped read (maximum read number mapped to a single gene=354). Restricting the hits to genes covered by reads across $\geq 90\%$ of their length resulted in eight high-confidence antimicrobial resistance (AMR) genes. Of these, three overlapped with findings from the AMRFinderPlus analyses of assemblies. A total of 20 resistance genes were identified (Table 5). Nineteen of the resistance genes were identified at Operastranda, the urban main sampling location, and one at the control location, Huk.

The β -lactamase genes identified in the genomes of ESBL-producing colonies represented unambiguous AMR identifications. We, therefore, specifically searched for these resistance determinants in the eDNA sequence data, both at the assembly- and read-level. There was no overlap between ESBL genes in the metagenomic assemblies, as identified by AMRFinderPlus, and those identified in ESBL-producing colonies. However, both methods detected functionally related MUS and BlaB/IND/MUS metallo- β -lactamase genes carried by *Flavobacterium* spp. at the Operastranda location (Table 5), but not at the same time points.

Next, we searched for the β -lactamase genes identified in the ESBL-producing colonies in the read-level outputs from RGI. The *P. tohonis* PAM-3 gene was not recovered in the read-level analyses. In fact, not a single read mapped to any PAM beta-lactamase in any sample.

The *F. odoratimimum* MUS family subclass B1 metallo- β -lactamase conferring carbapenem resistance, identified at Operastranda, was not perfectly recapitulated in the RGI-CARD analysis. However, two eDNA reads from the same sample were mapped to a *Flavobacterium johnsoniae* JOHN-1 carbapenemase. Two other samples, one from Operastranda and the other from Huk, had 5 and 16 reads, respectively, mapping to this gene.

The *A. venetianus* OXA-266 β -lactamase identified in the Operastranda sample from 8 August was not recovered in the eDNA reads from the same time point, but a single read from the same location was mapped to the gene 2 weeks later.

Table 5. AMR determinants identified by selective plating and sequence analysis

Date	Location	AMR gene	Subclass	Taxon*	AMRFinder (assemblies)	RGI (reads)	ESBL agar
28 May	Operastranda	Mph(E) family macrolide 2'-phosphotransferase	ERM	<i>Spirosomataceae</i> (f) ¹ (AF)/ <i>A. baumannii</i> (RGI)	X	X	
	Operastranda	ABC-F type ribosomal protection protein Msr(E)	AZM/ERM	<i>Spirosomataceae</i> (f) ¹ (AF)/ <i>Acinetobacter</i> sp. (RGI)	X	X	
	Operastranda	Tetracycline efflux MFS transporter Tet(39)	TET	<i>Spirosomataceae</i> (f) ¹ (AF)/uncultured bacterium (RGI)	X	X	
	Operastranda	BlaB/IND/MUS family subclass B1 metallo- β -lactamase	β -Lactam	<i>Flavobacterium</i> (g)	X		
	Operastranda	qacL	Multiple (efflux)	<i>V. cholerae</i>		X	
	Operastranda	ErmB	MLS	<i>Enterococcus faecium</i>		X	
	Operastranda	YajC	Multiple (efflux)	<i>Pseudomonas aeruginosa</i>		X	
	Operastranda	MexF	Multiple (efflux)	<i>Pseudomonas aeruginosa</i>		X	
	Operastranda	sulI	Sulphonamides	<i>Vibrio fluvialis</i>		X	
04 June	Huk	Metallo-beta-lactamase bla	β -Lactam	Azotimanducaceae (f) ²	X		
18 June	Operastranda	Subclass B1 metallo- β -lactamase	β -Lactam	Bacteroidia (c) ³	X		
		Subclass B3 metallo- β -lactamase PAM-3	Carbapenem	<i>P. tohonis</i>			X
02 July	Operastranda	Subclass B1 metallo- β -lactamase	β -Lactam	<i>Lewinella</i> (g)	X		
	Operastranda	NAD(+)-rifampin ADP-ribosyltransferase	Rifamycin	<i>Lewinella</i> (g)	X		
	Operastranda	MUS family subclass B1 metallo- β -lactamase	Carbapenem	<i>F. odoratimum</i>			X
08 August	Operastranda	Class A β -lactamase	β -Lactam	Sphingorhabdus_B (g)	X		
	Operastranda	OXA-266 family class D beta-lactamase	β -Lactam	<i>A. venetianus</i>			X
	Operastranda	Class A β -lactamase	β -Lactam	Porticoccaceae (f) ⁴	X		
	Operastranda	Subclass B1 metallo- β -lactamase	β -Lactam	<i>Aquirufa</i> (g)	X		
20 August	Operastranda	Class A β -lactamase	β -Lactam	<i>Aequoribacter</i> (g)	X		

*Reported taxon for method identifying the resistance gene. For genes found with >1 method, 'AF', AMRFinder; 'RGI', RGI.

Lowest reported GTDB taxon for hits without standing in nomenclature: ¹JALRIJ01 sp029865665; ²UBA4421 sp009886205; ³JACPUW01 (MAG);

⁴HTCC2207 sp001438605.

DISCUSSION

Sequencing-based surveillance is a potentially powerful and effective method for surveillance and early identification of pathogens and has been applied across a wide range of settings, including urban waterways [19], drinking water in high-risk settings [20] and marine recreational water bodies [21]. We found sequence-based profiles generated by metagenomic (Illumina) and full-length 16S sequencing (Oxford Nanopore) to differ substantially and suspect that this is mainly driven by the very complex composition of the microbiomes in the urban recreational swimming area under study, combined with database effects [22, 23]. *In lieu of* ground truths, selective culturing was applied to anchor the findings from sequence-based profiling to real-world abundance metrics. Our findings suggest that both metagenomic and full-length 16S sequencing are promising tools for surveillance of

seasonal opportunistic *Vibrio* pathogens, as *Vibrio* species are detected at low abundance with both approaches. However, a lack of any prolonged heating spells during the study period did not allow us to assess rigorously our ability to detect seasonal upswings.

Furthermore, *Flectobacillus*, *Malikia*, *Rhodofex* and *Leptothrix* 16S abundances were found to correlate strongly with sewage contamination as measured by coliform colony counts. This finding is likely robust, as all four have been identified as members of sewage, wastewater and/or activated sludge ecosystems [24–27]. Indeed, the most abundant genus among the four, *Rhodofex*, was recently identified as a major indicator of organic pollution of a river system [24]. The high abundance of *Rhodofex* (16% of the total 16S abundance in the most contaminated sample) makes it a potentially attractive indicator of sewage contamination. In comparison, the most abundant coliform genus (*Enterobacter*) never exceeded 0.13% 16S abundance.

Plating of water samples on ESBL-selective plates revealed minimal levels of resistance at all time points, including the one sample containing relatively high levels of sewage contamination. Four samples resulted in zero colonies from 1 ml of water, whereas three samples produced a single colony each. Whole-genome sequencing of the isolates revealed the presence of carbapenem resistance determinants in two of them and β -lactam resistance in one. In parallel, culture-independent identification of resistance genes not restricted to ESBL was performed, resulting in the identification of 17 additional antibiotic resistance determinants, none of which were carbapenemases.

The identification of resistance determinants in metagenomic assemblies is a relatively conservative endeavour, as demonstrated by the low number of identified resistance genes. The mapping of raw reads against resistance gene databases is, however, a different beast altogether. Here, the number of identified resistance genes was largely a function of filtering criteria, with a total of 1,628 resistance genes identified, of which 427 (26%) had a single read mapped, and 1,135 (70%) had 1–5 mapped reads. We did find traces of closely related genes in the read-level analysis for two out of three culture-confirmed ESBL-conferring genes. These were, however, represented by only a handful of reads per sample (1–16 reads) and would, under no circumstance, be regarded as confident hits *in lieu* of additional evidence.

Taken together, both culturing and eDNA sequencing demonstrated low levels of antibiotic resistance in the environment. Our results also demonstrate that the identification of AMR determinants from raw sequencing reads is extremely dependent on filtering criteria.

Our study demonstrates that sequencing-based surveillance could be a useful tool for pathogen surveillance of recreational swimming areas. However, the implementation of sequencing-based surveillance of seasonal *Vibrio* pathogens and sewage contamination would require additional work to identify relevant abundance thresholds.

Funding information

The study was partially funded by a Norwegian Research Council FRIPRO grant (grant ID: 314720).

Acknowledgements

We would like to thank the City of Oslo Agency for Urban Environment and the Agency for Water and Sewerage Works for their helpful advice and sampling for the project.

Author contributions

Conceptualization: V.E. Data curation: V.E. Formal analysis: V.E., D.S. and O.B.B. Funding acquisition: V.E. and D.S. Investigation: V.E., D.S. and O.B.B. Methodology: V.E. and O.B.B. Project administration: V.E. Resources: V.E. and D.S. Software: V.E. and O.B.B. Writing – original draft: V.E. Writing – review and editing: V.E., D.S. and O.B.B.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Amato E, Riess M, Thomas-Lopez D, Linkevicius M, Pitkänen T, et al. Epidemiological and microbiological investigation of a large increase in vibriosis, northern Europe, 2018. *Euro Surveill* 2022;27:2101088.
- Matsuo Y, Komiya S, Yasumizu Y, Yasuoka Y, Mizushima K, et al. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinIONTM nanopore sequencing confers species-level resolution. *BMC Microbiol* 2021;21:35.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
- Curry KD, Wang Q, Nute MG, Tyshaieva A, Reeves E, et al. Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat Methods* 2022;19:845–853.
- Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20:257.
- Irber L, Pierce-Ward NT, Abuelanin M, Alexander H, Anant A, et al. sourmash v4: a multitool to quickly search, compare, and analyze genomic and metagenomic data sets. *JOSS* 2024;9:6830.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 2022;38:5315–5316.
- Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, et al. AMRfinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep* 2021;11:12728.

11. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10:1–10.
12. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* 2023;2:e107.
13. Kang DD, Li F, Kirton E, Thomas A, Egan R, *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019;7:e7359.
14. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;31:3210–3212.
15. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
16. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, *et al.* CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2023;51:D690–D699.
17. Brinch C, Otani S, Munk P, van den Beld M, Franz E, *et al.* Discovery of *Vibrio cholerae* in urban sewage in Copenhagen, Denmark. *Microb Ecol* 2024;87:102.
18. Bauer A, Østensvik Ø, Florvåg M, Ørmen Ø, Rørvik LM. Occurrence of *Vibrio parahaemolyticus*, *V. cholerae*, and *V. vulnificus* in Norwegian Blue Mussels (*Mytilus edulis*). *Appl Environ Microbiol* 2006;72:3058–3061.
19. Logan LK, Zhang L, Green SJ, Dorevitch S, Arango-Argoty GA, *et al.* A pilot study of Chicago waterways as reservoirs of multidrug-resistant *Enterobacteriaceae* (MDR-Ent) in a high-risk region for community-acquired MDR-ent infection in children. *Antimicrob Agents Chemother* 2020;64:e02310-19.
20. Sevillano M, Vosloo S, Cotto I, Dai Z, Jiang T, *et al.* Spatial-temporal targeted and non-targeted surveys to assess microbiological composition of drinking water in Puerto Rico following Hurricane Maria. *Water Res X* 2021;13:100123.
21. An X-L, Wang J-Y, Pu Q, Li H, Pan T, *et al.* High-throughput diagnosis of human pathogens and fecal contamination in marine recreational water. *Environ Res* 2020;190:109982.
22. Xu R, Rajeev S, Salvador LCM. The selection of software and database for metagenomics sequence analysis impacts the outcome of microbial profiling and pathogen detection. *PLOS ONE* 2023;18:e0284031.
23. Smith RH, Glendinning L, Walker AW, Watson M. Investigating the impact of database choice on the accuracy of metagenomic read classification for the rumen microbiome. *Anim Microbiome* 2022;4:57.
24. Exton B, Hassard F, Medina Vaya A, Grabowski RC. Polybacterial shift in benthic river biofilms attributed to organic pollution – a prospect of a new biosentinel? *Hydrol Res* 2023;54:348–359.
25. Kunoh T, Yamamoto T, Sugimoto S, Ono E, Nomura N, *et al.* *Leptothrix cholodnii* response to nutrient limitation. *Front Microbiol* 2021;12:691563.
26. Leiviskä T, Risteelä S. Analysis of pharmaceuticals, hormones and bacterial communities in a municipal wastewater treatment plant – comparison of parallel full-scale membrane bioreactor and activated sludge systems. *Environ Pollut* 2022;292:118433.
27. Spring S, Wagner M, Schumann P, Kämpfer P. *Malikia granosa* gen. nov., sp. nov., a novel polyhydroxyalkanoate- and polyphosphate-accumulating bacterium isolated from activated sludge, and reclassification of *Pseudomonas spinosa* as *Malikia spinosa* comb. nov. *Int J Syst Evol Microbiol* 2005;55:621–629.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org