NAR Genomics and Bioinformatics, 2026, **8**, lqag040
https://doi.org/10.1093/nargab/lqag040
High Throughput Sequencing Methods

Long-read nanopore shotgun metagenomic DNA sequencing for river biodiversity, wildlife, pollution, and environmental health monitoring

Orestis Nousias[1,2,3], Fiona G. Duffy[1], Isabelle J. Duffy[1], Mark McCauley[1], Jenny Whilde[1], David J. Duffy [1,4,5,6,*]

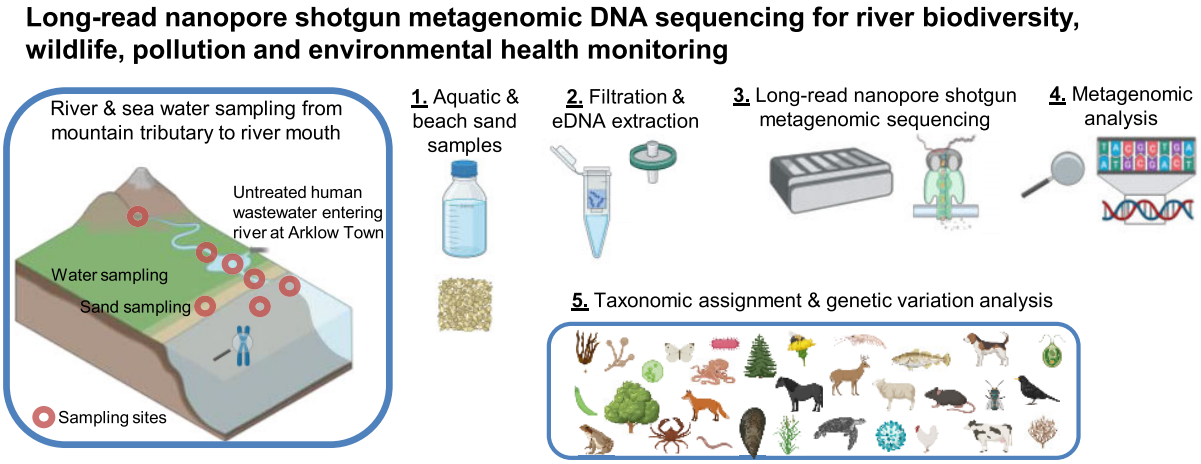
[1]The Whitney Laboratory for Marine Bioscience and Sea Turtle Hospital, University of Florida, St. Augustine, FL 32080, United States
[2]Department of Biostatistics, Yale School of Public Health, New Haven, CT 06510, United States
[3]Biotech Research and Innovation Center (BRIC), University of Copenhagen, 2200 Copenhagen N, Denmark
[4]Department of Biology, College of Liberal Arts and Sciences, University of Florida, Gainesville, FL 32611, United States
[5]Wildlife Rehabilitation Ireland (WRI), Duleek, Co. Meath, A92 R1H5, Ireland
[6]University of Florida Biodiversity Institute, University of Florida, Gainesville, FL 32611, United States

*To whom correspondence should be addressed: Email: duffy@whitney.ufl.edu

Abstract

As the human population expands and global temperatures rise, species, populations, and biodiversity decline at unprecedented rates, while the frequency of infectious disease emergence increases. Therefore, it is more vital than ever to accurately understand the current state of natural habitats and their constituent species. We assess the feasibility of a single assay: long-read shotgun metagenomic sequencing of environmental DNA (eDNA), to monitor species from across the tree of life, from viruses to complex multicellular organisms, across a representative Irish river system (Avoca River, Co. Wicklow). We conducted aquatic eDNA sampling and long-read shotgun metagenomic sequencing from a mountain tributary through to the sea. This approach could detect and quantify organismal DNA present in environmental samples, from microbes (including DNA viruses) to mammals. Rather than the traditional siloing of microbial and multicellular studies of DNA recovered from environmental samples, simultaneously considering viruses, microbes, and eukaryotes (animals, plants, and fungi) can provide deeper insights. This single assay can simultaneously quantify differences in DNA abundance for a broad range of species and pathogens across sites and sample types, enabling wide-ranging biodiversity assessments. This included human, wildlife, plant, and microbial pathogens and parasites with health, agricultural, and economic importance. The environmental genomic data enabled animal phylogeny and transmissible cancer analysis (blue mussel, *Mytilus edulis*) even from natural complex community settings. Oxford Nanopore sequencing provides a quantitative approach for river biodiversity, pollution, and environmental health monitoring. Long-read shotgun metagenomic sequencing of environmental samples offers the means to assess whole ecosystems and the ecological, trophic, and host-pathogen interactions occurring within them.

Graphical abstract



Received: December 17, 2024. **Revised:** February 12, 2026. **Accepted:** February 16, 2026

Introduction

Rapid advances in genomic sequencing technologies and computational analyses are presenting novel opportunities for noninvasive monitoring of environments, microbes, and wild species, including human-impacted habitats. Such technologies are set to play an increasingly important role in studying and responding to the ongoing interrelated international crises of biodiversity loss, habitat loss, climate emergency, infectious disease epidemics, and anthropogenic pollution of biomes [1, 2]. While the microbial research community has long embraced microbial metagenomics [3], the application of shotgun metagenomic techniques to multicellular species environmental DNA (eDNA) studies has been slower, largely due to the assumption that eukaryotic DNA was not sufficiently abundant in environmental samples [4], and that enrichment approaches were the only viable option. Similarly, there is a dearth of studies simultaneously assessing viral, prokaryotic, and eukaryotic DNA from environmental samples [5], including aquatic samples. Eukaryotic organisms have cells with a membrane-bound nucleus, i.e. multicellular animals, plants, fungi, and many unicellular organisms.

Metagenomic sequencing technologies will be crucial for effective management, conservation, and nature restoration, including meeting Target 3 of the Convention on Biological Diversity's Kunming-Montreal Global Biodiversity Framework (GBF). As part of the global effort to halt and reverse biodiversity loss, this target includes the ambitious goal of increasing the global coverage of protected areas and implementing other effective area-based conservation measures to at least 30% by 2030 (referred to as "30 × 30") [6–8]. To ensure the areas most capable of sustaining the richest biodiversity are protected, and to measure the success or otherwise of implemented protections, scalable, rapid, cost-effective tools are required for biodiversity assessment across differing regions of the globe with highly varied flora, fauna, funga, viromes, and microbiomes. Advanced genomic sequencing platforms have been coupled with novel sampling approaches to recover genetic material from environments, providing new tools for ecosystem, pathogen, and biodiversity monitoring and wildlife conservation [9–22].

Despite its ubiquity in fields such as human medical research, there is a relative paucity of multicellular organism studies applying whole-genome (nuclear and organelle DNA) sequencing approaches to environmental samples, particularly with long-read sequencing technologies [5, 17–19, 23–26]. Whole-genome sequencing approaches, in which all genetic material present in a sample is sequenced without enrichment or targeting of specific regions, are known as shotgun sequencing. When applied to environmental samples (bulk samples with multiple individuals and species), this approach is known as shotgun metagenomic sequencing. This paucity of metazoan-focused genome-wide metagenomic shotgun sequencing eDNA studies in the scientific literature occurs despite early comparisons revealing that metabarcoding is less consistent than shotgun whole mitogenome sequencing [26]. When vertebrates were analyzed, shotgun metagenomic sequencing of airborne DNA was also less biased than metabarcoding of the same air eDNA samples [5].

eDNA is beginning to be adopted as a non-invasive tool for biodiversity and pollution monitoring. For example, innovative recent studies demonstrated that eDNA content is correlated with long-term water quality indicators and riverine biodiversity measures [27, 28]. However, to achieve broad taxonomic coverage, 14 separate metabarcoding assays were needed [27], requiring significant laboratory manipulation and processing time. In contrast, single-assay shotgun metagenomic sequencing is rapid, reduces barcode/primer and polymerase chain reaction (PCR) bias, and therefore provides more directly comparable quantitative data across all recovered species groups. Additionally, it can recover species beyond those recoverable even with 14 metabarcoding assays, and importantly, it recovers genomic data beyond the small barcode regions used by metabarcoding approaches. While some pioneering studies have successfully applied long-read sequencing to aquatic systems, these are normally confined to bacterial or fungal analyses and/or metabarcoding approaches [13, 29–31].

As reference genome databases are increasingly being populated with high-quality genomes of species from across the tree of life [7, 32], the species-level resolution of eDNA shotgun metagenomic sequencing will continue to improve. Additionally, improved bioinformatic pipelines to filter out promiscuous reads will help to further increase the specificity of genus and species calls. Continued improvements in sequencing technologies, including higher output, will likely further enhance the already impressive sensitivity of such approaches. Cloud-based automated analysis workflows and artificial intelligence (AI) will improve the useability of such tools, making them more accessible to a wide range of stakeholders.

Here, we conducted environmental sampling and shotgun metagenomic long-read sequencing from water samples taken from the Avoca River watercourse, Co. Wicklow, Ireland, from a mountain tributary of the river through to seawater samples near the river mouth (Fig. 1a and Supplementary Fig. S1). For comparison, we also collected and sequenced DNA recovered from sand from a beach 5 km north of the Avoca River mouth (Fig. 1a and Supplementary Fig. S1). Thirty-seven samples were collected from 2022 to 2024; 31 samples (2022/2023) had their total DNA concentration measured by spectrometry, 31 samples (2022–2024) were assessed by quantitative PCR (qPCR), and 11 samples (2022/2023) were shotgun long-read sequenced. We demonstrate the simultaneous detection of biodiversity from viruses to vertebrates in a single assay: long-read shotgun metagenomic sequencing of eDNA samples. We reveal that this rapid single-assay approach provides a viable complementary or alternative approach to traditional multicellular species eDNA metabarcoding [18]. This approach could successfully generate results related to biodiversity, pollution, environmental health, and viral monitoring, for a representative river system, from mountain tributary to the Irish Sea. As sequencing technologies and associated computational analyses continue to advance, long-read shotgun metagenomic sequencing is poised to become a rapid monitoring tool for changing environments, including aquatic point-source tracing of biotic pollution and for disentangling the relative contributions of different biotic pollution sources (human waste, agricultural, industry, or wildlife). Simultaneously, this biological "everything" assay can monitor changes to biodiversity and species distributions in response to perturbations, including anthropogenic pollution, urbanization, habitat destruction, and beyond.

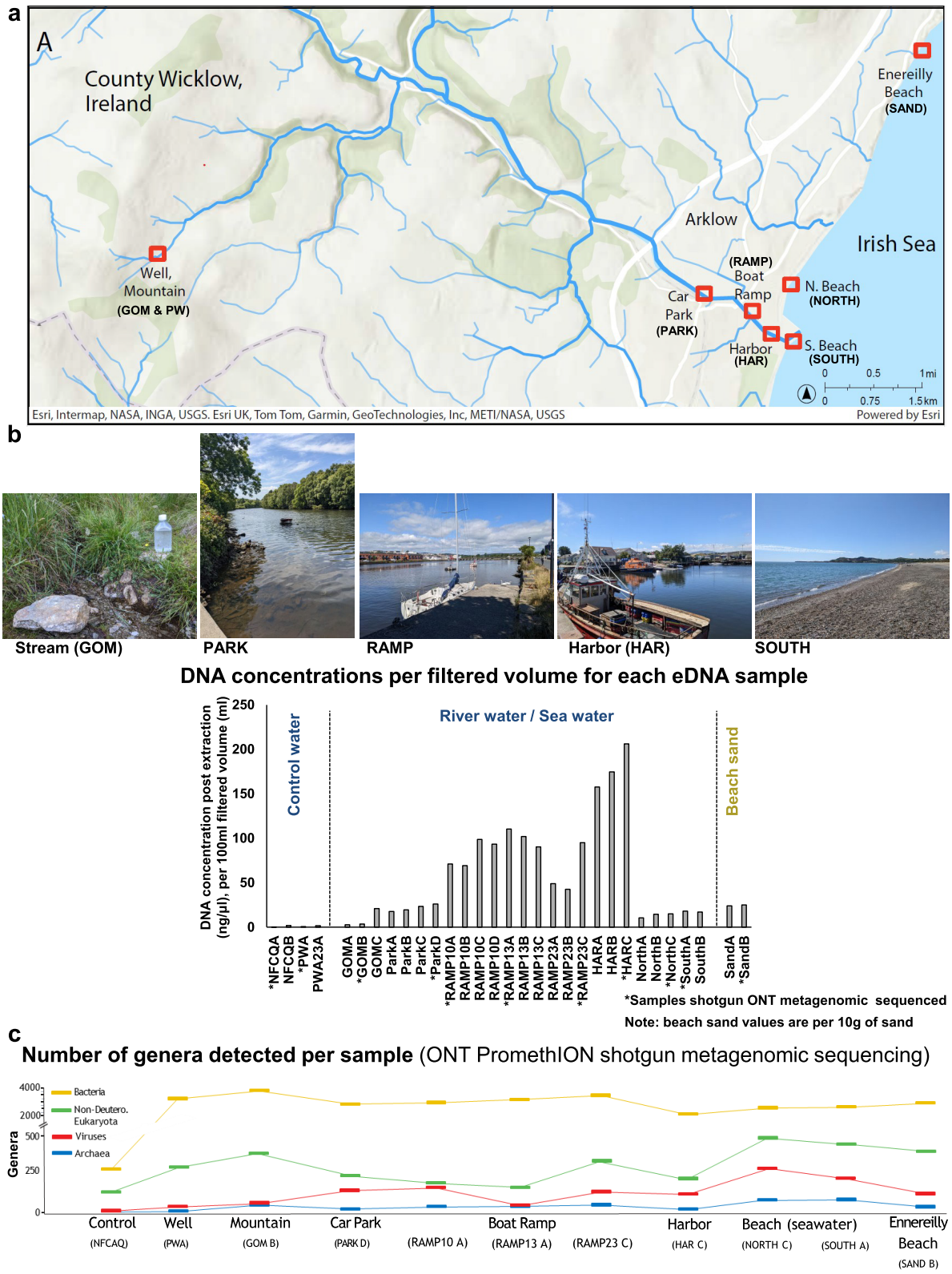


Figure 1. (**a**) Watershed map of the study's sampling areas in Co. Wicklow, Ireland. For satellite imagery of the sampling area, including extent of urbanization (Arklow Town), see Supplementary Fig. S1. Each sampling location is denoted by a red box. Shorthand location names are provided in capital letters within brackets. Maps created in ArcGIS online. (**b**) Graph of eDNA concentrations as measured by Nanodrop Spectrophotometer (Thermo Fisher) per 100 ml filtered volume for each of the 31 2022/2023 samples, with images from the sampling sites inset. Sample names with asterisks denote the 11 samples that were also used for long-read shotgun metagenomic sequencing. For total volumes filtered for each of the sequenced samples, see Supplementary Table S1. For pan-eukaryotic eDNA and human eDNA concentrations per sample (as assessed by qPCR), see Supplementary Fig. S2a and b, respectively. (**c**) Biodiversity assessments of water and sand environmental samples conducted using long-read shotgun metagenomic sequencing. Pan-biodiversity (excl. deuterostomes) genera assessments of each of the 11 samples (minimum of two sequences are required for genera identification).

Materials and methods

Sample collection and DNA extraction

All laboratory procedures from sampling through to final analysis (sequencing or qPCR) were conducted in a way that minimized any human DNA contamination, including any human DNA contamination from investigators. This included no contact of investigators/samplers with substrates being sampled (water or sand), new nitrile gloves being used for each sample collection, and frequent glove changes throughout sample processing, cleaning of equipment and benchtops with bleach prior to use, and negative field controls being treated identically to genuine field samples throughout all processes from extraction through to qPCR/sequencing (see below). All eDNA samples were extracted in a chick/mouse developmental biology lab that does not process human samples: the Murphy Lab, Zoology Department, Trinity College Dublin, Ireland.

Thirty-seven eDNA/negative field control samples were generated for this study. River, estuarine, sea water, and beach sand samples were collected between 2022 and 2024 on summer fieldwork visits to Co. Wicklow, Ireland. Negative field control samples of 500 ml Qiagen nuclease-free water (Cat. no. 129117) were transported from the laboratory to environmental sampling locations and stored in a cool box with the environmental samples to monitor for potential contamination during sampling, transportation, and processing. Negative field controls were filtered and extracted alongside the other collected sand and water samples from each sampling trip and subjected to the same next-generation sequencing conditions and qPCR conditions. A range of 180–600 ml of water was filtered per sample (until the filter was clogged with material, or a maximum of 600 ml was reached) (see Supplementary Table S1 and Whitmore *et al.*, 2023 [23]). Filtration was performed shortly after sampling using 0.22 μm pore Millipore Sterivex-GP pressure filter units (Merk Millipore Cat. no. SVGPL10RC) for each DNA sample [15, 18]. Water samples were pumped (by hand) through 0.22 μm Sterivex-GP pressure filter units and capped with B. Braun luer lock caps (Medline, Cat no. BMGTMR2000B). Hand pumping was carried out with sterile 50 ml BD luer lock syringes (Fisher Scientific, 13-689-8 BD). Seven hundred forty microliters of Qiagen Buffer ATL were then added per filter, and the filters were stored at −20°C until transport to the lab (Trinity College Dublin) and DNA extraction. Filters containing Buffer ATL were stored at −20°C for between 1 and 6 days prior to extraction, with a median storage of 4 days (Supplementary Table S1). For sand eDNA, a 50 ml tube was filled with sand from each sampling event, with 10 ml of this sand used per individual eDNA extraction [18]. Sand samples were transported to the lab (Trinity College Dublin) in a cool box and eDNA extraction initiated on the same day as sample collection (<2 h between collection and extraction). No storage buffer was added to the sand between collection and extraction.

For water samples, filters were thawed, and 60 μl proteinase K from a Qiagen DNeasy Blood and Tissue Kit was added to each sample (filter), and they were placed in sterile 50 ml Falcon conical centrifuge tubes in a rolling incubator overnight at 56°C.

For sand samples, Qiagen nuclease-free water (Cat. no. 129117) was added to each individual 10 ml sand sample in individual 50 ml Falcon conical centrifuge tubes at approximately two times the volume of sand (10 ml sand and 20 ml water). Samples were shaken gently by hand, then set on a rocking platform for 1 h at room temperature, with additional gentle shaking by hand every 15 min. Samples were then rested until sand had sunk to the bottom of each tube, and then the supernatant was immediately pipetted into a 50 ml sterile BD luer lock syringe (Fisher Scientific Cat. no. 136898). Samples were then hand-filtered using 50 ml BD luer lock syringes through 0.22 μm Sterivex-GP pressure filter units (Millipore, Cat. no. SVGPL10RC) and capped with B. Braun luer lock caps (Medline, Cat. no. BMGTMR2000B). Seven hundred forty microliters of Buffer ATL and 60 μl proteinase K from a Qiagen DNeasy Blood and Tissue Kit (Qiagen Cat. no. 69504) were added to each sample and they were placed in sterile 50 ml Falcon conical centrifuge tubes in a rolling incubator overnight at 56°C. For both water and sand samples, after overnight incubation the rest of our eDNA extraction method using a modified Qiagen DNeasy Blood and Tissue kit (Qiagen, Cat. no. 69504) protocol [15, 18, 19, 23, 33] was applied. After overnight incubation, solutions were transferred from the Sterivex-GP Pressure Filter Units to 2 ml microcentrifuge tubes using 1 ml Fisherbrand™ sterile syringes (Fisher Scientific 14-955-462). Equal volume of AL buffer and ice-cold ethanol was added to each sample, and they were vortexed and centrifuged after each addition, according to the rest of the kit manufacturer's protocol, with the following exception: DNA was eluted with 70 μl AE buffer (incubated at 70°C prior to addition to spin column), incubated on the column at room temperature for 7 min, and centrifuged into a 1.5 ml microcentrifuge tube at 6000 × *g* for 1 min. DNA concentration was measured on a Thermo Scientific Nanodrop 2000 Spectrophotometer (at Systems Biology Ireland, University College Dublin) and samples were stored at −20°C until qPCR or shotgun metagenomic sequencing. Prior to filtration and between every sample, laboratory surfaces and equipment were cleaned with 70% ethanol and then 10% bleach (to destroy DNA). Collection bottles were disinfected (washed thoroughly) with 10% bleach and rinsed thoroughly with deionized water.

Previously extracted [34] IMR32 gDNA, extracted using the standard cell line protocol of a Qiagen DNeasy Blood and Tissue kit (Qiagen, Cat. no. 69504), was used as template for generating standard curves (both human and pan-eukaryotic standard curves). This DNA had been extracted from cultured human IMR32 neuroblastoma cells in May 2013 at Systems Biology Ireland, UCD, for the Duffy *et al.* (2018) study [34], and had subsequently been stored at −20°C until being used in this study (2024).

Sequencing & bioinformatics

Of the 31 2022/2023 samples, one replicate sample per site/day was selected for sequencing based on the highest DNA concentration, pan-eukaryotic DNA quantification, and sample volume availability (Fig. 1b and Supplementary Fig. S2a and Supplementary Table S1). In total, 11 samples were shotgun metagenomically sequenced commercially at BMK Gene, with Oxford Nanopore Technologies (ONT) libraries sequenced on a PromethION 48, with allowance for 6 Gb output per sample (Supplementary Table S1). For DNA concentrations of sequenced samples, see Supplementary Table S1; the range was 25.5 ng/μl to 474.4 ng/μl, while for the NFC and private well water it was 1.9 ng/μl and 3.4 ng/μl,

respectively. Extracted DNA was examined by Nanodrop, Qubit, and 0.35% agarose electrophoresis for its concentration, purity, and integrity. gDNA was fragmented with gTube to generate DNA fragments of ~8 kb. Library preparation was conducted with ONT's ligation sequencing gDNA kit (Cat. no. SQK-LSK109) according to manufacturer's protocol. Briefly, harvested DNA fragments were processed for nick repair and end repair, with A-tailing. The products were then purified with magnetic beads (Agencourt AMPure XP beads, Beckman Coulter, A63881). Sequencing adapters were ligated to the end of prepared DNA fragments, and final products were purified with magnetic beads. Library concentration was measured by Qubit. The 11 libraries were sequenced with FLO-PRO002 (R9.4.1) flow cells, raw sequences were processed by BMK Gene for adapter removal, low-quality reads removal, and short-reads removal (threshold: 1000 bp) to generate 50.02 Gb of delivered clean data. All sequencing libraries were pooled equally (including the negative field control) to give equal sequencing depth opportunity. An average of 4.7 million bases and 1.3 million reads were generated per sample, with an average N50 read length of 5434 bp per sample (Supplementary Table S1 and Supplementary Figs S7 and S8). All bioinformatic tools were utilized using default parameters, unless otherwise stated. Negative control data were processed and visualized alongside the genuine eDNA samples to explicitly make clear what portion of findings may have originated from any background reagent contaminome.

Chan Zuckerberg (CZ) ID [35, 36], a free, cloud-based metagenomics platform (https://czid.org/) was utilized for direct biodiversity bioinformatic analysis of the raw reads obtained from the BMK Gene. The CZ ID platform queries genetic information from all species, excluding deuterostomes. The CZ ID nanopore microbial metagenomic pipeline was utilized. As the platform was developed for microbial metagenomics, it does not query deuterostome (bilaterian animal) sequences, as these are identified as host species, and it actively subtracts host reads, such as human reads. Additionally, for samples with over 1M reads, the pipeline actively sub-samples only 1M reads per analysis. Note that for three samples (PW A [0.97 M], RAMP10 A [0.80 M], and HAR C [0.68 M]) fewer than 1 million reads were generated (Supplementary Table S1). However, since the CZ ID nanopore results are reported in nucleotides per million base pairs (nt_bpm) this does not affect data interpretation for these three samples. Genera with at least two sequence reads were utilized. To conduct metazoan metagenomics, including deuterostomes, the CZ ID nanopore metagenomic pipeline was recreated in an HPC environment using bash and Python, without host filtering, without sub-sampling to 1M reads, and aligning to all metazoan species (NCBI Blast), as described in Nousias *et al.* (2025) [5]. To classify metagenomic reads based on taxonomy, DIAMOND v2.1.7 [37] was utilized for its inherent high-speed alignment, sensitivity, and the translation of nucleotide reads for protein alignment capabilities. First, metagenomic reads from eDNA samples were pre-processed for adapter trimming and quality control using Porechop v0.2.4[59]. The two-step DIAMOND process involved database preparation and read alignment. A reference database was constructed using the non-redundant (NR) protein sequences from NCBI. Taxonomic information was integrated using DIAMOND's "makedb" option with the following input files: a protein accession to taxonomy ID mapping file prot.accession2taxid (https://ftp.ncbi.nih.gov/pub/taxonomy/accession2taxid/), a taxonomy nodes file nodes.dmp (https://ftp.ncbi.nih.gov/pub/taxonomy/new_taxdump/), and a taxonomy names file names.dmp (https://ftp.ncbi.nih.gov/pub/taxonomy/new_taxdump/). The processed reads were aligned to the prepared database using DIAMOND's blastx mode with "very-sensitive" option, ensuring comprehensive taxonomic classification as per the most stringent options of the package. This step translated nucleotide sequences into protein sequences, aligning them against the database. The alignment output format was specified to include detailed taxonomic information for each read. Only the best hit for each read was retained using the option "max-target-seqs 1." The results were parsed with in-house Python script to aggregate reads based on the NCBI taxonomy structure for different phyla of protostomes and deuterostomes. Because this reduced version of the CZ ID pipeline was used to overcome the read subset limitations and the exclusive identification of eukaryotes, the contig generation step of the pipeline was not implemented, as it is not feasible to reconstruct species-specific contigs for eukaryotes due to the large genome sizes, the pooled nature, and the unequal representation of species in eDNA NGS data. The metazoan metagenomics bioinformatics pipeline [5] (MetaBio-Tax v1.0) is available via GitHub at https://github.com/nousiaso/MetaBioTax, https://gitfront.io/r/nousiaso/QZwYQLPm9dJQ/MetaBioTax/, and Zenodo (DOI: 10.5281/zenodo.18697671). Based on the metagenomics analysis, the South A seawater eDNA reads were also directly aligned to an *M. edulis* reference mitogenome [38] (Genbank accession number: AY484747.1) with NanoGalaxy [39]. The reads were checked for quality (FastQC, v0.73 [40]), adapters and low-quality reads were trimmed, Porechop v0.2.4 [41], and high-quality reads were aligned to the reference mitogenome minimap2, v2.28 [42]. Mitogenome coverage was visualized with JBrowse2 [43] (v2.17.0). The *M. edulis* mitochondrial aligning reads from the SOUTHA sample shotgun ONT metagenomic sequencing were blasted (NCBI, BLASTn, v2.17.0, https://blast.ncbi.nlm.nih.gov/Blast.cgi) against *M. edulis* nucleotides to determine the closest hits. A radial phylogram was constructed by sequence alignment of the longest *M. edulis* mitochondrial aligning read (11 741 bp) with all *M. edulis* full mitochondrial genomes available in NCBI (see figure for accession numbers) and oyster (*Magallana gigas*) mitochondrial genome (NCBI accession number: PV604936.1) as an outgroup, using Clustal Omega v1.2.4 [44–46]. Full-length mitochondrial sequences were manually trimmed to the ONT eDNA 11.7 kb long-read region. For *Mytilus* transmissible cancer analysis, we built a curated BTN2 marker panel (EF1A H/H′, H4 BTN2 candidates, mtCR C/D, and mtCOI B/Q) and combined it with non-cancer *Mytilus* mitogenomes and a broader mtCOI set, based on the cancer-associated sequences from the Yonemitsu *et al.* 2019 [47] paper. Multiple alignment passes (composite + marker-only) showed no convincing BTN2 signal in the SOUTHA sample sequencing data. Any BTN2-labeled hits only appear when MAPQ = 0 is allowed (i.e. multi-mappers). With standard quality/uniqueness gates (MAPQ ≥ 10–20, locus-specific length, and identity ≥0.85), BTN2 counts are zero. Observed BTN2-labeled matches are non-specific seed

hits consistent with multi-mapping across short, similar fragments. With competitive mapping, all reads aligned to wild-type alignment, i.e. reads matched wild-type sequences rather than mutated cancer-associated marker panel sequences (https://github.com/nousiaso/Transmissible-Cancer-Identification-in-Mussels-from-long-read-eDNA-data) and Zenodo (DOI: 10.5281/zenodo.18697707). Nanopore reads from two FASTQ files (ONT_BMK230829-BO313-006N0006_clean.fq.gz, ONT_BMK230829-BO313-006N0010_clean.fq.gz) were processed. Read identifiers were first extracted using awk from two DAA files (aligned_to_canidae_NORTHC_VS.nanopore.daa, aligned_to_canidae_SANDB_VS.nanopore.daa), previously generated by DIAMOND [37] alignments. These identifiers were used by seqtk subseq [48] to subset the original FASTQ files. The *Canis lupus familiaris* reference genome (GCF_000002285.5_Dog10K_Boxer_Tasha) was indexed using samtools faidx [49]. The subsetted Nanopore reads were then aligned to this reference using minimap2 [42] with options -ax map-ont –secondary = no and 8 threads. Resulting SAM files were converted to BAM, sorted, and indexed using samtools [49]. Alignments were subsequently filtered with samtools view to retain primary alignments with a mapping quality (MAPQ) ≥ 30 (filtering flags: -F 2308). Basic alignment statistics were generated using samtools flagstat and coverage [49]. Assessing our eDNA samples for human reads was conducted with University of Florida Institutional Review Board (IRB-01) ethical approval under project number IRB202201336. Area-proportional Venn diagrams of genus lists were generated using BioVenn [50]. All species silhouette images throughout the figures are from www.phylopic.org.

qPCR

The Applied Biosystems pre-validated TaqMan Gene Expression qPCR assay directed against the human *ZNF285* gene (assay ID Hs00603276_s1) was used as species-specific human assay, based on having no cross-reactivity with over 29 other species from mammals to plants (https://www.thermofisher.com/order/genome-database/ and Whitmore *et al.* 2023 [19, 23]: mouse, rat, *Arabidopsis*, *C. elegans*, fruit fly, bovine, dog, Chinese hamster, goat, white-tufted-ear marmoset, guinea pig, zebrafish, horse, chicken, soybean, cynomolgus monkey, sheep, rabbit, green sea turtle, loggerhead sea turtle, rice, rhesus monkey, baker's yeast, fission yeast, pig, bread wheat, wine grape, western clawed frog, and maize), and having both primers and probe within a single exon (i.e., detects DNA). A pan-eukaryotic 18s rRNA gene (Applied Biosystems, 4319413E) pre-validated TaqMan Gene Expression assay, which also has both primers and probe within a single exon (i.e., detects DNA), was used to quantify the total level of pan-eukaryotic DNA in each of the samples. We have previously applied this pan-eukaryotic assay as a positive eDNA extraction and PCR inhibitor control for a range of eDNA sample types [15, 17, 18, 23].

The qPCR reaction mixtures were performed on 384-well plates in a total volume of 10 µl per well: 5 µl TaqMan Gene Expression Master Mix (Fisher Scientific Cat. no. 4369016); 3.5 µl nuclease-free water (Fisher Scientific); 0.5 µl of the respective assay (primer/probe mix, manufacturer-supplied concentration); 1 µl DNA template (or, for no-template controls, an additional 1 µl of nuclease-free water per well). Each biological sample was run in three technical replicates for eDNA samples and standard curve points. Note that for the previously published qPCR results (2022 samples), three technical replicates (18s assay) or six technical replicates (ZNF285 assay) were conducted [23]. Negative field controls had the same number of technical replicates run as their corresponding eDNA samples. No template controls were run in triplicate on every qPCR plate. qPCR reactions were performed on an Applied Biosystems QuantStudio 7 Flex at The Conway Institute's Genomics Core, University College Dublin, Ireland, with the following cycling parameters: 50°C for 2 min and 95°C for 10 min for one cycle, followed by 45 cycles of 95°C for 15 s and 60°C for 1 min. For absolute quantity qPCRs, a standard curve was generated using six DNA one-in-ten serial dilutions, ranging from 111.3 ng/µl to 0.001113 ng/µl. 1 microliter of template was used per standard curve reaction. The standard curve template used for both pan-eukaryotic and human-specific qPCRs was DNA extracted from cultured human IMR32 neuroblastoma cells in May 2013 for the Duffy *et al.* (2018) study [34]. For each standard curve concentration and each eDNA sample technical triplicate wells/reactions.

qPCR results were plotted with BoxPlotR [51] (http://shiny.chemgrid.org/boxplotr/) with every data point displayed. Tukey whiskers (extending to data points that are $<1.5 \times$ IQR away from 1st/3rd quartile) were utilized for every boxplot. One box is graphed per single sample, consisting of all qPCR technical replicate wells for that sample. Biological replicates are denoted by letters A–D at the end of the sample name. Biological replicates are not pooled on any boxplots, with each sample being denoted by its own box.

Results

Recovery of total DNA from environmental samples

We assessed the ability of a single assay: shotgun metagenomic long-read sequencing of environmental samples, to monitor species DNA from across the tree of life, from viruses to complex multicellular organisms, across a river system (Fig. 1a and Supplementary Fig. S1). Irish eDNA samples were collected for two purposes: (i) to quantify the level of human eDNA (species-specific qPCR) present in each 2022 sample [23], and (ii) to assess the viability of ONT long-read shotgun metagenomic sequencing to monitor all detectable DNA per sample (reported here). The highest DNA yields were obtained from the tidal portion of the Avoca River (boat ramp and harbor samples) that runs through Arklow Town, Co. Wicklow (Fig. 1a and b and Supplementary Fig. S1a and b). Lower quantities of DNA were recovered from the mountain stream and seawater sites, despite larger volumes being filtered (Fig. 1b and Supplementary Table S1). This indicates a higher abundance of biological material at the tidal regions of the river. Despite having lower total DNA abundances (Fig. 1a), the two seawater sites had similar eukaryotic (multicellular animals, plants, fungi, and many unicellular organisms) DNA concentrations to the harbor samples (pan-eukaryotic 18S rRNA gene qPCR assay [23], Supplementary Fig. S2a). When assessed by long-read shotgun metagenomic sequencing, the seawater samples contained a greater diversity of non-deuterostome eukaryotic species (Fig. 1c).

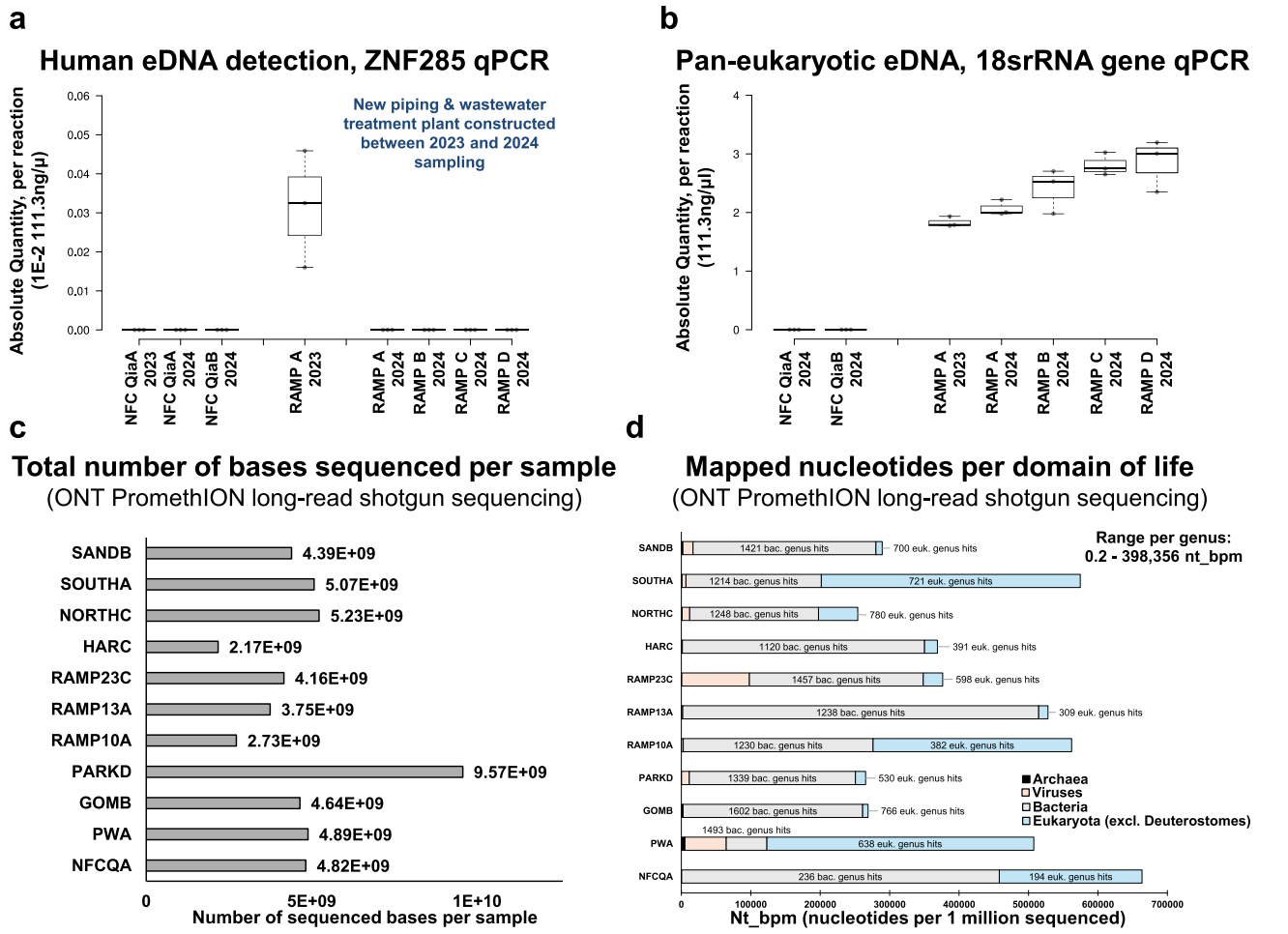


Figure 2. (**a**) qPCR-based species-specific quantification of human eDNA from Avoca River water boat ramp sampling site (2023 and 2024). Quantified with ZNF285 human-specific assay. Each qPCR reaction is a 10 µl reaction containing 1 µl of extracted eDNA template. Absolute quantity of human eDNA per sample was calculated by comparing with a ZNF285 assay on a human cell line (IMR32) DNA standard curve. Each box in the panel is an individual sample, $n = 8$ samples. (**b**) Pan-eukaryotic eDNA levels within ramp site (2023 and 2024) water eDNA samples. Absolute quantity of 18s rRNA pan-eukaryotic eDNA levels, as assessed by qPCR. Each qPCR reaction is a 10 µl reaction containing 1 µl of extracted eDNA template. Absolute quantity of pan-eukaryotic eDNA per sample was calculated by comparison with an 18s rRNA pan-eukaryotic assay on a human cell line (IMR32) DNA standard curve. Each box in the panel is an individual sample, $n = 7$ samples. (**c**) Graph of the total number of nucleotide bases sequenced per sample, by ONT PromethION long-read shotgun metagenomic sequencing. (**d**) Graph of the number of mapped nucleotides per domain of life (by CZ ID analysis, post filtering, and promiscuous read removal) recovered by shotgun long-read ONT metagenomic sequencing, nucleotide bases per 1 million bases sequenced (nt_bpm). The number of bacterial and eukaryotic (excluding deuterostomes) genus hits are shown in or adjacent to relevant domain boxes. Note: all genus hits are reported with no minimum nt_bpm cut-offs applied. The nt_bpm range per genus recovered was from 0.201244 nt_bpm (Philodina rotifers in the PARKD sample) to 398 356 nt_bpm (Bradyrhizobium bacteria DNA in the NFC sample).

Human eDNA for temporal and spatial wastewater release/contamination monitoring

We previously showed by species-specific human qPCR of the 2022 samples (Supplementary Fig. S2b, Whitmore *et al.* 2023 [19, 23]) that the river is heavily contaminated by human eDNA at the boat ramp site in Arklow town. The majority of this human eDNA is suspected to originate from improperly treated Arklow town wastewater entering the river (https://www.epa.ie/publications/monitoring–assessment/waste-water/Urban-Waste-Water-Treatment-in-2022-Report.pdf). Sampling at the same site in 2023 confirmed the continued presence of human eDNA (Fig. 2a). However, in 2024 human eDNA was no longer present at detectable levels (Fig. 2a), despite high eukaryotic eDNA recovery rates (Fig. 2b). This suggests that the extensive wastewater treatment plant and pipe construction works undertaken in Arklow between 2023 and 2024 successfully removed/diverted this pollution source from this site.

Pan-biodiversity assessments from a single assay: shotgun long-read metagenomic sequencing

Of the 31 water and sand eDNA samples collected in 2022 ($n = 25$) and 2023 ($n = 6$), 11 representative samples were selected for long-read sequencing (Figs 1b, c and 2c, d, and Supplementary Table S1). These unbiased shotgun metagenomic sequencing data provide insights into the relative abundance of DNA recovered from each domain of life (Figs 1c and 2d) and animal phylum (Fig. 3a). When normalized by nucleotides per 1 million sequenced bases, bacterial reads were the most abundant in all samples except three: well water (PWA), river water (RAMP10A), and sea water (SOUTHA, Fig. 2d). For these three samples, eukaryotic DNA (excluding deuterostomes) was most abundant (Fig. 2d). However, all samples, including these three, had more bacterial genera than eukaryotic genera (excluding deuterostomes), indicating that bacteria were the most diverse domain of life (Figs 1c and 2d). Archaea were most abundant in private well groundwater

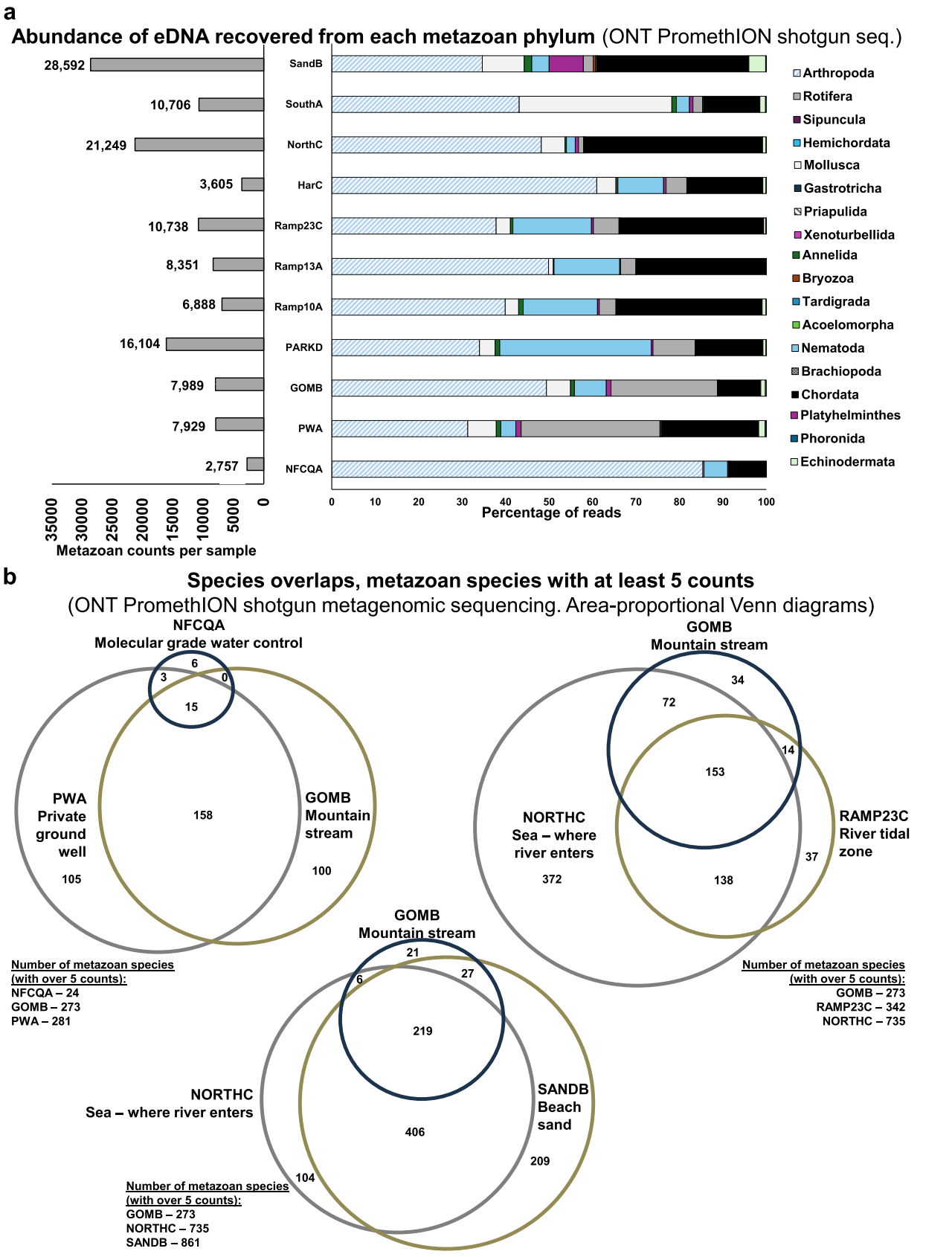


Figure 3. (**a**) Abundance of eDNA recovered from each of the 18 metazoan phyla, by shotgun long-read ONT metagenomic sequencing. Left: total metazoan counts per sample. Right: percentage of metazoan counts per phylum, per sample. Note: no counts were recovered, in any sample, for the deep-sea Xenoturbellida phylum [52, 53]. (**b**) Area-proportional Venn diagrams of metazoan species overlaps between eDNA samples, showing the number and overlap of metazoan species with at least 5 counts per sample, as detected by ONT shotgun metagenomic sequencing. Generated using BioVenn [50]. See Supplementary Fig. 3c for additional sample comparisons.

(5224 nt_bpm), followed by beach sand (2225 nt_bpm, Fig. 2d, and Supplementary Fig. S3a). Across all environmental samples, the metazoan (animal) phylum with the highest read counts was Arthropods (Fig. 3a). Rotifer DNA was highest in the groundwater sample (private well pump, domestic water), while Chordata DNA was highest in the beach sand, followed closely by the northern seawater sample near the Avoca River's outflow into the Irish Sea (Fig. 3a and Supplementary Fig. S1a). No counts were recovered in any sample for the deep sea-restricted Xenoturbellida phylum [52, 53]. With a minimum cut-off of five counts per species, 861 metazoan species were detected in 10 g of beach sand (SOUTHB, Fig. 3b and Supplementary Fig. S3b). The sea water sample (NORTHC) had the second highest diversity of metazoan species detected, with 735 (Fig. 3b and Supplementary Fig. 3b). The negative field control (NFC, Qiagen molecular grade water) had DNA from 24 metazoan species detected (Fig. 3b and Supplementary Fig. 3b), while domestic drinking water from a private groundwater pump (near the mountain stream sampling site, PWA) had DNA from 281 metazoan species detected (Fig. 3b and Supplementary Fig. S3b). Species richness was not a factor of the number of bases generated per sample (Fig. 2c and Supplementary Table S1). There was a mix of overlapping and unique metazoan species detected between the eDNA samples (Fig. 3b and Supplementary Fig. S3c).

Species DNA abundance quantification and eDNA-based genetic analysis

We next compared the top 10 eukaryotic and non-eukaryotic hits in the beach sand (BEACHB) and tidal river water (RAMP23C) samples. For the top hits in both samples, the amount of non-eukaryotic DNA in the environmental substrates was generally an order of magnitude greater than eukaryotic DNA (Fig. 3a). It should be noted that eukaryotic DNA still made up a greater proportion of the total DNA than has been previously estimated for environmental samples [4] (Fig. 3a), with three samples (PWA, RAMP10A, and SOUTHA) even having a greater amount of eukaryotic than prokaryotic aligning nucleotides (Fig. 2d). Our non-eukaryotic category includes bacteria, archaea, and viruses. Note that this analysis omits deuterostomes (an animal subclass of eukaryotes), which are filtered out by the CZ ID tool used as potential host species [36]. The most abundant non-eukaryote in the river water sample was Asteriusvirus, from a family of bacterial and archaeal viruses (Supplementary Fig. S4a). The most abundant non-eukaryote in the beach sand was the bacteria genus Mycobacterium, consisting mostly of *Mycobacterium branderi* (Fig. 3a). Seven of the water sample's (RAMP23C) top 10 non-eukaryotic hits were bacteria; conversely, seven of the beach sand's (SANDB) top 10 non-eukaryotic hits were viruses, including uncultured marine viruses (Supplementary Fig. S4a and b).

Human DNA was also quantifiable by shotgun metagenomic sequencing in environmental samples (Fig. 4a), with the highest eDNA levels being in Avoca River samples taken in Arklow town. This mirrors previous human eDNA assessment using species-specific qPCR assays (Supplementary Fig. S2b); improperly treated wastewater release in Arklow is the suspected primary source of this human eDNA [23]. The boat ramp site sample (RAMP13A) from July 2022 had the highest level of human eDNA (2279 counts), followed by the car park (PARKD, July 2022) and same boat site (RAMP23C, July 2023) sampled one year later (1392 and 1270 counts, respectively). By applying shotgun metagenomic sequencing to these samples, the abundance of human DNA can be directly compared to that of other animals, human pathogens, and gut/fecal-associated microbes. We next compared the level of human eDNA recovered to that of a selected group of human-associated and wild mammal species (Fig. 4b). For this group of mammals, human eDNA was the most abundant in almost all samples, with the main exception being beach sand, which had a higher abundance of otter, rabbit, ferret, and dog eDNA (Fig. 4b). In fact, of the 17 non-human mammal species analyzed, the beach sand had the highest abundance for 13 of these mammals, indicating a role of sand in filtering and accumulating eDNA. Only dog, cat, stoat, and house mouse eDNA were more abundant in at least one of the water samples (Fig. 4b). The three highest abundances of eDNA recovered across these 17 mammals were dog, ferret, and otter (137 to 98 counts). Across all 17 of these mammals, the total count was highest in the sand sample (SANDB, 772 counts, or 175.8 counts per million sequenced bases [cpmb]), followed by a seawater sample (NORTHC, 534 counts, or 102.2 cpmb), while the negative field control and mountain stream only had small numbers of total counts for these mammals, 16 and 20, respectively. NORTHC also had the highest abundance of dog (*Canis lupus familiaris*) eDNA (Fig. 4a) with reads recovered from each nuclear chromosome (Supplementary Fig. S5a).

Next, to test the ability of shotgun metagenomic sequencing long-read eDNA to determine the relative abundance of farmed species eDNA in waterways, we assessed detection of *Oncorhynchus* (fish genus) eDNA (Fig. 5a). There is a rainbow trout (*Oncorhynchus mykiss*) fish farm ~8 km upstream of the boat ramp sampling site, and the mountain stream site (GOMB) is further upstream from the trout farm. No rainbow trout eDNA was detected in the mountain stream site (GOMB). However, almost all downstream samples contained rainbow trout eDNA, with the highest abundance being the site closest to the farm (PARKD, 42 counts), ~8 km downstream from the farm. When all *Oncorhynchus* genus-aligning reads were considered, a seawater sample had the highest read abundance (NORTHC, 81 counts, Fig. 5a). We next assessed the abundance of eDNA from select aquatic metazoan species across the samples. Once again, the beach sand contained the highest abundance of many of these species (10 out of 15 of the detected species), including common octopus, coral, and leatherback sea turtle (Fig. 5b). Of these 15 species, blue mussel (*Mytilus edulis*) in the seawater sample (SOUTHA) had by far the highest abundance (837 counts, Fig. 5b). A radial phylogram was constructed by sequence alignment of the longest *M. edulis* mitochondrial aligning read (11.7 kb) with all available *M. edulis* full mitochondrial genomes in NCBI (Fig. 6a). The eDNA-derived sequence clustered with the F haplogroup clade (Fig. 6a). All *M. edulis* mitochondrial aligning reads from the SOUTHA sample shotgun ONT metagenomic sequencing were blasted (NCBI, BLASTn) against deposited *M. edulis* sequences to determine the closest hits (Fig. 6b). The majority (74%) of the blue mussel mitochondrial eDNA reads recovered from the Irish Sea, Arklow (SOUTHA), clustered most closely to blue mussel tissue DNA samples from the UK (Broadhaven, Pembrokeshire, Fig. 6b). The remainder of the reads most closely matched samples from France (Atlantic Coast) and Wales (Swansea Bay, Irish Sea) (Fig. 6b). The whole *M. edulis* mitochondrial genome was covered by reads from the SOUTHA seawater sample (Supplementary Fig. S5a). The

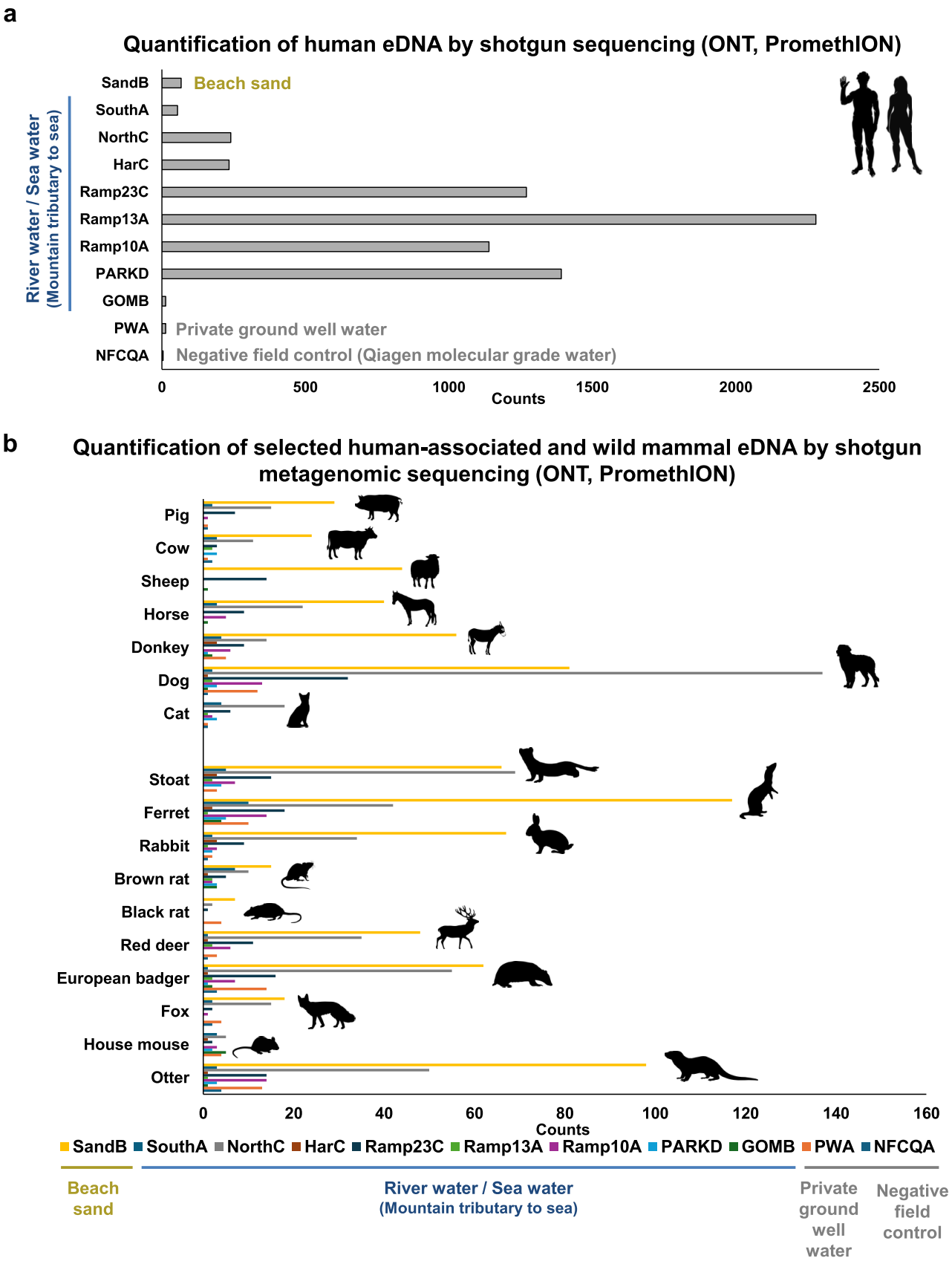


Figure 4. (**a**) Quantification of human eDNA (read counts) across the sample set, as detected by metagenomic analysis of long-read shotgun metagenomic sequencing. (**b**) Quantification of selected human-associated mammals and wild mammals eDNA across the sample set, as detected by metagenomic analysis of long-read shotgun metagenomic sequencing. Overlaid species silhouette images from www.phylopic.org.

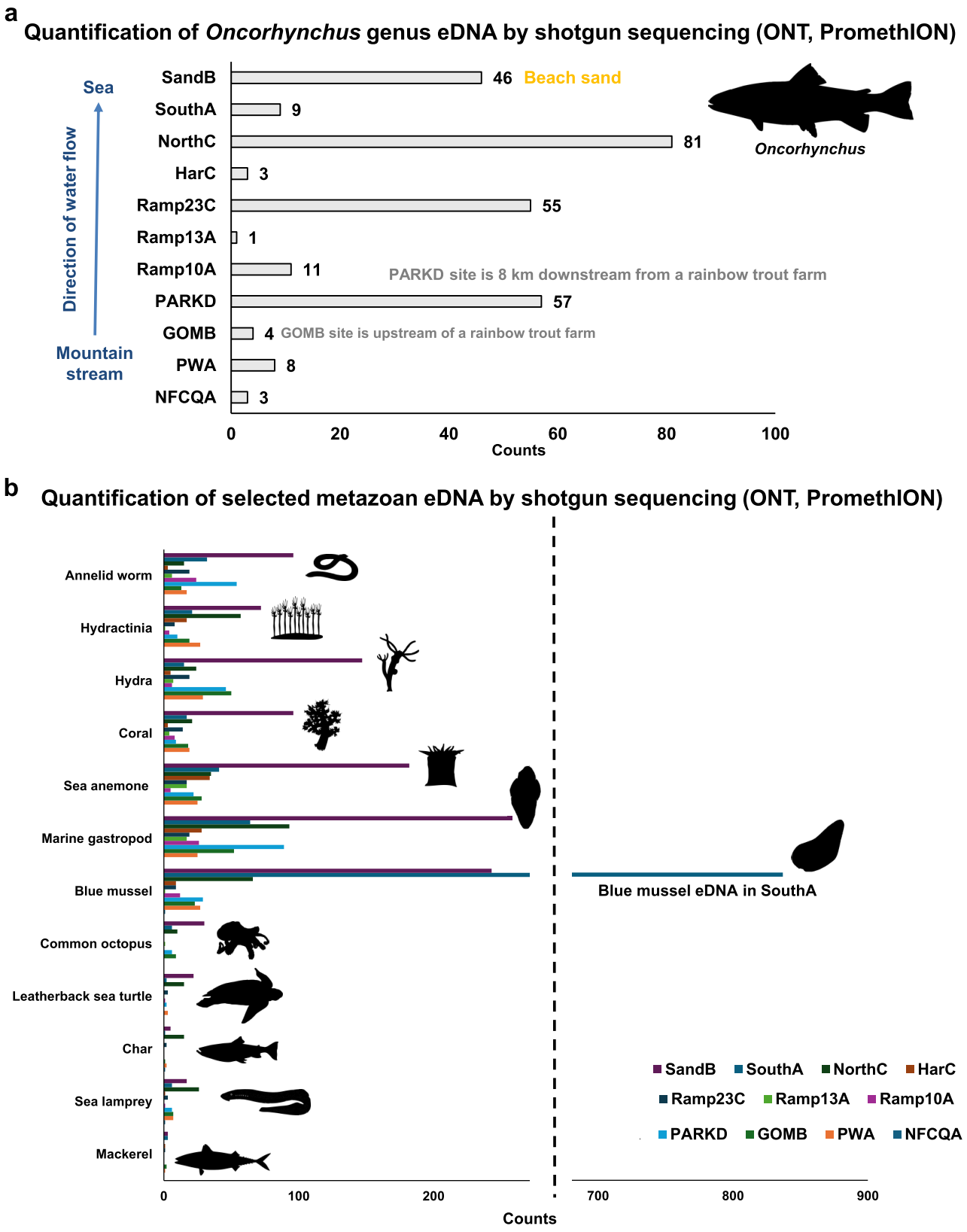


Figure 5. (**a**) Quantification of *Oncorhynchus* (salmon/trout) genus eDNA across the sample set, as detected by metagenomic analysis of long-read shotgun metagenomic sequencing. (**b**) Quantification of selected aquatic or aquatic sediment (Capitella, worm) metazoan eDNA across the sample set, as detected by metagenomic analysis of long-read shotgun metagenomic sequencing.

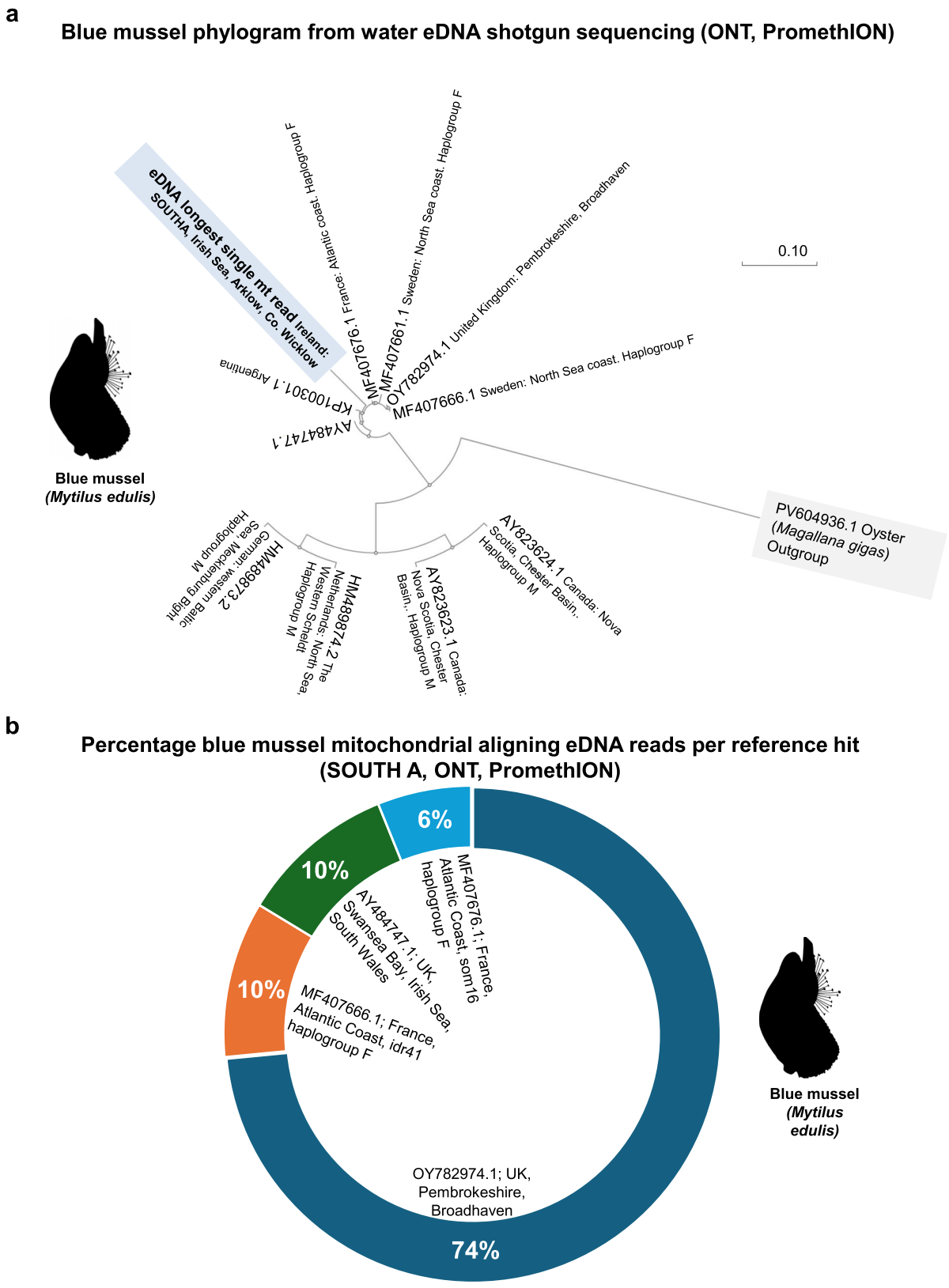


Figure 6. (**a**) Radial phylogram of the longest detected blue mussel (*Mytilus edulis*) eDNA 11.7 kb mitochondrial read, generated by Clustal Omega alignment of the longest *M. edulis* eDNA mitochondrial sequence, against full-length *M. edulis* mitochondrial genomes in the NCBI nucleotide database (trimmed to the 11.7 kb region). NCBI accession numbers of the comparative reference sequences are provided at the start of each branch name. Oyster (*Magallana gigas*) mitochondrial genome (NCBI accession number: PV604936.1) was used as an outgroup. Overlaid species silhouette image from https://www.phylopic.org. (**b**) Percentage of *M. edulis* eDNA mitochondrial reads matching each NCBI reference sequence. N = 49 *M. edulis* mitochondria aligning long reads.

longest single read aligning to the blue mussel mitochondrial genome from this sample was 11 741 bp long. As a single read, this sequence originates from one individual animal and covered 70.14% of the whole mitochondrial genome. This read most closely (BLASTn) hit the Wellcome Sanger Tree of Life Programme *M. edulis* whole mitochondrial genome (Genbank accession number OY782974.1, collected in Broadhaven, Pembrokeshire, UK), followed by the reference mitogenome used (Genbank accession number AY484747.1). *Mytilus* species can harbor a transmissible cancer with unique genetic signatures [47]. Therefore, the SOUTHA seawater sample was also assessed for reads indicative of *Mytilus* transmissible cancer. No transmissible cancer-associated reads were detected in the eDNA sample, with competitive alignment revealing reads matching wild-type rather than mutated cancer-associated sequences (100%, 34 wild-type aligning reads). In addition to mitochondrial reads, shotgun metagenomic sequencing can recover nuclear genetic information, such as dog nuclear sequences recovered from seawater and beach sand (Supplementary Fig. 5b and c). This included high-quality dog-originating reads with MAPQ scores greater than 41 (Supplementary Fig. S5b). Nuclear sequences recovered from environmental samples can potentially enable more in-depth population genetics or genetic trait analyses than mitochondrial sequences alone.

Simultaneous quantification of pathogen DNA from long-read metagenomics

Shotgun metagenomic sequencing also recovered DNA from a variety of pathogens (Figs 1c and 2d, and Supplementary Fig. S4a and b). As expected, marine algae and plankton species also had highest abundances in the seawater samples (Fig. 7a and b). Tree, flowering plant, butterfly, and *Trichoplax* eDNA were also detected, at varying quantities between the samples (Fig. 7c). Human, animal, and plant pathogens and parasites were also detected across the sample set (Fig. 7b and c). This includes the causative agent of global amphibian declines, the pathogenic chytrid fungus *Batrachochytrium dendrobatidis (Bd)* [54, 55] (Fig. 7c and d); and that of crayfish plague, the oomycete *Aphanomyces astaci*. The highest abundance of Bd was from the mountain stream sample, which drains upland bog and conifer plantations, prime amphibian habitat (Fig. 7d). This stream sample also contained the highest level of pine tree eDNA detected (Fig. 7c) and amphibian Ranavirus (2.82 nt_bpm). Of the river samples, this site also had the highest abundance of common frog (*Rana temporaria*) eDNA, which is one of Ireland's only three amphibian species (Fig. 7d). This result highlights the utility of rapid broad-scale environmental screening for organisms, pathogens, and parasites, which could then be followed up with more targeted monitoring of species of concern (species-specific qPCR/dPCR assays, etc.). Targeted study design and site selection could be greatly informed by the rapid generation of exploratory metagenomics data. This also highlights the power of long-read metagenomics for data-driven surveillance and discovery, being able to quantify both pathogen and host without *a priori* knowledge of which pathogens are present at the sampling site.

We next assessed the total number of viruses detected in the beach sand (SANDB) and tidal river water sample (RAMP23C). Both samples had over 300 viruses present, with only 112 of these occurring in both samples (Fig. 8a). DNA from only 15 viruses was detected in the negative field control sample. The control sample also had much lower abundance of viral DNA recovered (range 225–0.64 nt-bpm) than the sand (range 27 519–0.23 nt-bpm) or river water sample (range 94 432–0.24 nt-bpm). The beach sand and river water samples were examined for selected common human sexually transmitted skin, urinary, or genital tract infection pathogens (Fig. 8b). We focused on these pathogens as they are some of the most likely to enter human wastewater systems. Diseases associated with these known human pathogens include genital herpes, periodontitis, urethritis, cervicitis, vaginitis and vulvovaginal candidiasis, gonorrhea, bacterial vaginosis, and Legionnaires' disease. Note, unlike conventional wastewater screening for human pathogens, our results come from environmental samples (beach and river) after DNA has left any wastewater treatment facility and mixed into environmental substrates containing complex DNA mixtures. Despite this dilution effect "uncultured human fecal virus" was detected in both river water (RAMP23C, 503.26 nt-bpm) and beach sand (SANDB, 1.59 nt-bpm) samples. Seven of the beach sand's (SANDB) top 10 non-eukaryotic hits were viruses, including uncultured marine viruses (Supplementary Fig. S4a and b). Human- and animal-infecting pathogen and parasite DNA from different domains of life, such as poxvirus, streptococcus, and cryptosporidium, could also be readily quantified from the data (Supplementary Fig. S6a–c). Archaea were also readily detected by this approach, including human-gut- and acidic soil-associated species (Supplementary Fig. S6d).

Taken together, shotgun metagenomic sequencing and identification of human eDNA and fecal and intestinal-associated microbe DNA offer a powerful tool for the point source identification of human wastewater entering waterways and aquatic bodies. Simultaneously, the information contained within shotgun-metagenomic sequenced environmental samples enables ecosystem health analyses, biodiversity assessments, and even monitoring of human, agricultural, and biodiversity pathogen occurrence and abundance. Therefore, shotgun metagenomics offers cost-effective, non-invasive insights from environmental samples to help address a range of societal and environmental issues.

Discussion

There are clear benefits of utilizing metagenomic approaches to simultaneously study viral, microbial, and multicellular species DNA from environmental samples. Yet, despite the early and innovative application of environmental genome shotgun metagenomic sequencing to microbial DNA of the Sargasso Sea in 2004 by Venter *et al.* [3], the adoption of shotgun metagenomic sequencing approaches for eukaryotic eDNA research has been surprisingly slow, particularly for the investigation of multi-cellular species. This apparent hesitation persists despite the interceding genomics boom of the last 20 years, resulting in dramatic reductions in time and cost associated with shotgun metagenomic sequencing, and simultaneously increases in ease, output, and speed of sequencing and bioinformatic analysis. There has been a precipitous decline in the cost of deep sequencing, from billions of dollars for one human genome in 2000, to <$200 today, making shotgun metagenomic sequencing of environmental samples ever more feasible, even for projects with limited budgets. The pace of advance of genomic technologies has not slowed, being largely driven by medical genomic applications [56]. The

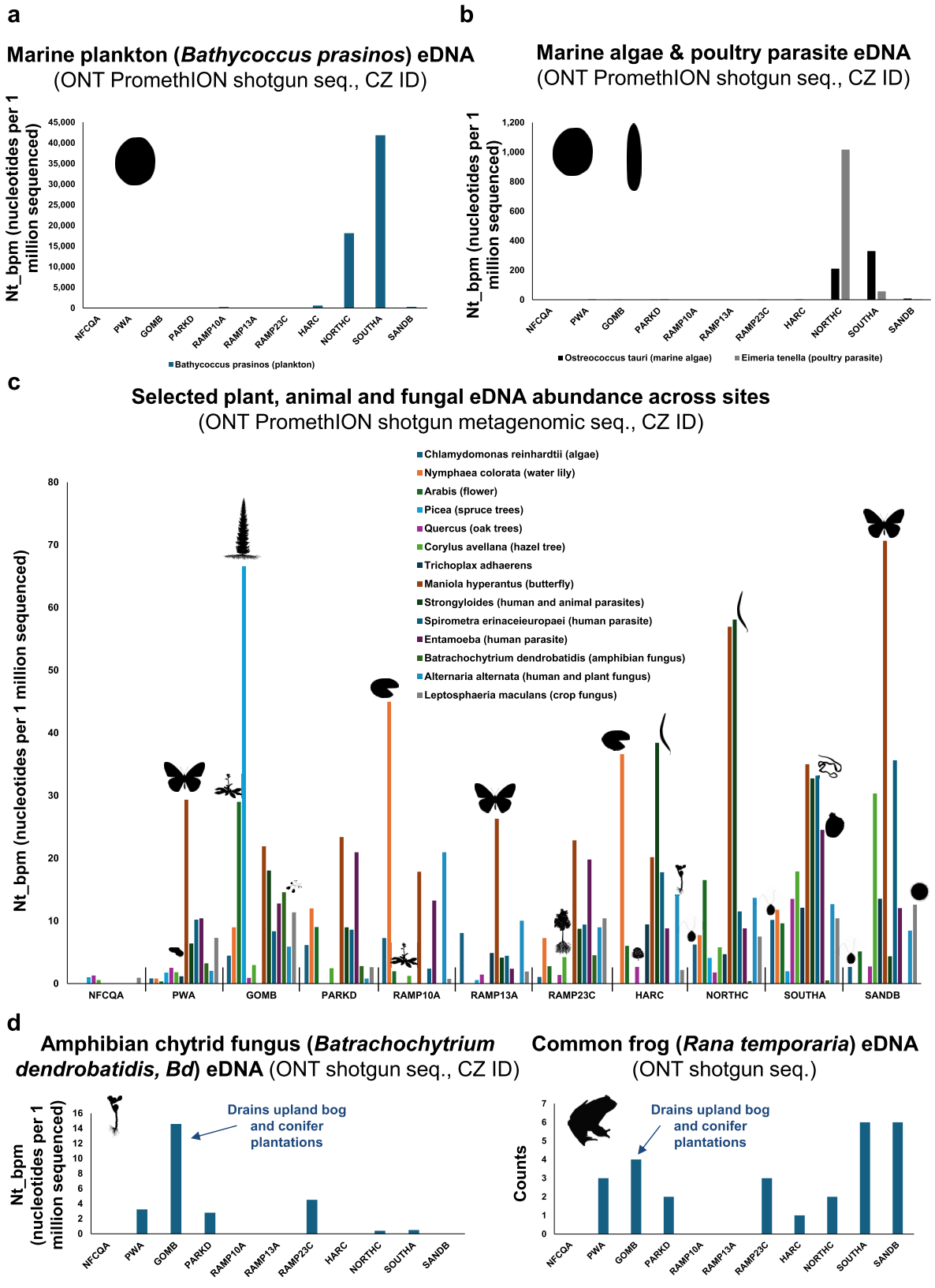


Figure 7. (**a**) Quantification of the marine plankton *Bathycoccus prasinos* throughout the sample set, arranged from freshwater to seawater (and finally beach sand), as detected by CZ ID metagenomic analysis of the long-read shotgun metagenomic sequencing. (**b**) Quantification of a marine alga (*Ostreococcus tauri*) and poultry parasite (*Eimeria tenella*) eDNA throughout the sample set, arranged from freshwater to seawater (and finally beach sand), as detected by CZ ID metagenomic analysis of the long-read shotgun metagenomic sequencing. (**c**) Quantification of selected plant, animal, and fungal species eDNA across the sample set, as detected by metagenomic analysis of long-read shotgun metagenomic sequencing. The list includes parasitic species. (**d**) Left: quantification of the amphibian chytrid disease agent *Batrachochytrium dendrobatidis* throughout the sample set, as detected by CZ ID metagenomic analysis. Right: quantification of the common frog (*Rana temporaria*) throughout the sample set, as detected by metagenomic analysis.

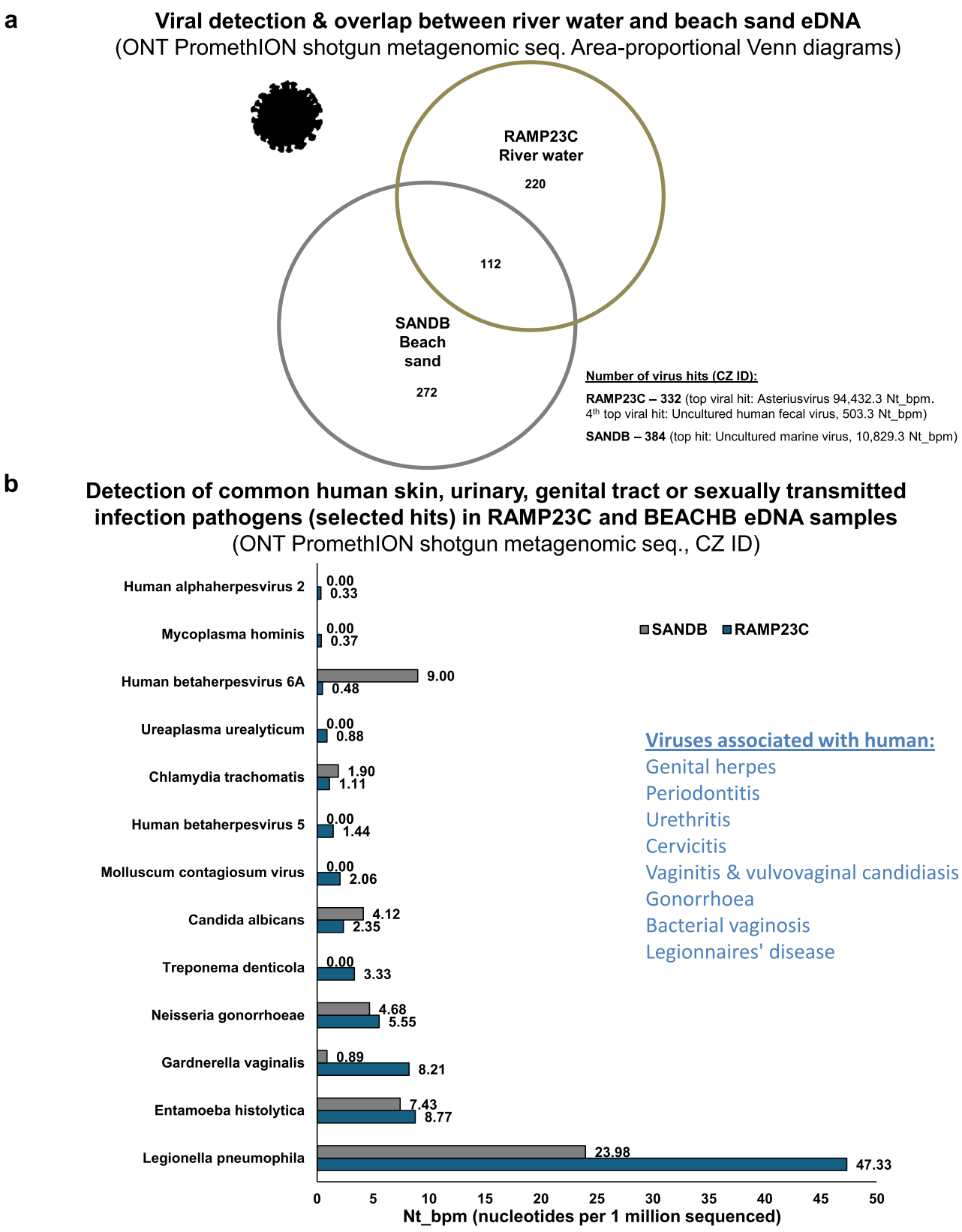


Figure 8. (**a**) Area-proportional Venn diagrams of viral species overlaps between the beach sand and Arklow Town Avoca River water (RAMP23A) eDNA samples, as detected by CZ ID metagenomic analysis of long-read shotgun metagenomic sequencing. Venn diagram was generated using BioVenn [50]. (**b**) Quantification of common human sexually transmitted, skin, urinary, or genital tract infection pathogen eDNA in the beach sand and Arklow Town Avoca River water (RAMP23A) eDNA samples. Note that at the time of sampling this river water sampling site was known to be contaminated with untreated human wastewater (Figs 2a, 4a, and Supplementary Fig. S2b).

development of long-read sequencing should prove particularly beneficial to multicellular eDNA applications [24]; this approach has already gained traction in microbial research with long-read microbial metagenomics utilized increasingly frequently. Environmental DNA for multicellular species has been dominated by less informative metabarcoding approaches. Current cost-effectiveness and advanced readily available computational tools make it timely to re-introduce and re-evaluate the role of modern shotgun metagenomic sequencing technologies in transforming eDNA research capabilities.

Short-read shotgun metagenomic sequencing can outperform metabarcoding approaches for diverse species groups, including plant pollen, macroinvertebrates, reptiles, and fish [5, 26, 57, 58]. We demonstrated here that long-read shotgun metagenomic sequencing can be used as a single assay for simultaneous investigation of biodiversity, pollution, environmental health, and viral monitoring for an example river system from mountain tributary to the sea. This approach will automatically benefit from ongoing drives to increase high-quality reference genomes for eukaryotic and microbial species, as these new references are added to public databases. For example, 1.5 million eukaryotic reference-sequence genomes are being generated by the Earth BioGenome Project (EBP) alone [7, 32]. These approaches will also benefit from continued advances in the accuracy, speed, and portability of long-read sequencing technology [7, 24, 29]. The data reported here showed good agreement between the shotgun metagenomic data, species-specific qPCR, and known species trends, e.g. transition from freshwater to marine species across the samples. However, not all species/genus calls in this proof-of-principle project may be fully correct, due to the current uneven coverage of species reference genomes in databases. For example, if a species does not yet have a reference genome, some reads may accurately align instead to its closest related species present in the queried database, rather than the actual species from which the DNA originates. However, this will improve over time in line with large-scale reference genome generation projects (see above), and assignment to higher taxonomic categories can be useful in such instances. Similar issues arise with metabarcoding data, as barcode databases are far from complete, even for many charismatic species [59–61]. Generally, for metabarcoding, the wider the species net is cast, the less well-resolved the hits. Many metabarcoding studies and barcode sets do not achieve species-level resolution, opting rather to report at genus or higher-level operational taxonomic units (OTUs).

The nature of shotgun metagenomic sequencing data means that they can readily be reanalyzed against updated databases in the future and with further refined bioinformatic pipelines, without requiring any additional laboratory work. Metabarcoding does not enjoy this benefit as widely, as there are limitations to the depth and breadth achievable with a particular barcode set [59–61]. Sufficient sequence differences within the barcode may just not exist, even as more species are added to barcode databases, and any updated barcode sets would require re-PCR with the new barcodes and re-metabarcode sequencing of samples. Furthermore, shotgun metagenomic sequencing is less susceptible to barcode and PCR biases [5, 61], and has much more tolerance of unknown genetic variation within populations, including distant populations from different locations around the globe [18]. Long-read shotgun metagenomic sequencing also does not require specific advanced computational and laboratory effort to develop barcodes or qPCR assays. Additionally, it is becoming cost-effective to sequence deeply and filter out promiscuous reads (reads aligning to multiple locations/genomes) [35, 36]. The cost-effectiveness of sequencing and the high performance of computing make this approach rival barcoding in terms of specificity, as only highly stringent reads can be analyzed. Generating additional information is a boon toward further research beyond mere species identification.

We show here that in a single cost-effective and non-labor-intensive assay, rich information from across the tree of life can be recovered. In each environmental sample, with a single assay, we recovered DNA from all domains of life, from viruses to vertebrates. This approach is unbiased in revealing which groups contribute the most DNA abundance in any given sample. Direct comparison of DNA quantities of different groups and domains of life is not readily achievable with barcoding- or qPCR-based approaches, due to the heterogeneity introduced by the different barcode/primer sets required for each group, and their differing efficiencies and affinities. The only domain of life not recovered by the shotgun metagenomic DNA sequencing is RNA-based viruses. However, shotgun RNA metagenomic sequencing (transcriptomics) approaches, either utilizing reverse transcribed complementary DNA or direct long-read ONT RNA sequencing, can be used to recover RNA viruses from environmental samples. For example, SARS-CoV-2 has already been detected from wastewater [14, 62, 63].

Fecal bacteria have long been used as a proxy in the study of sewage discharge, and, more recently, sequencing-based approaches have been adopted to target fecal bacteria [64, 65]. Here, we show that in a single assay (shotgun long-read metagenomic sequencing), human eDNA, fecal-associated bacteria, fecal-associated archaea, and fecal-associated viruses can be simultaneously detected at a site with known frequent wastewater discharge. In 2022, Arklow was one of the 15 areas in Ireland that did not meet European Union wastewater treatment standards and was also one of the 26 Irish towns and villages discharging raw sewage into the environment every day (https://www.epa.ie/publications/monitoring–assessment/waste-water/Urban-Waste-Water-Treatment-in-2022-Report.pdf). In 2024, pipe improvement and wastewater treatment plant construction had progressed in Arklow town. The fact that human eDNA was no longer detectable (by species-specific qPCR) at the site (https://www.water.ie/projects/local-projects/arklow-wwtp/) highlights that infrastructure investment can effectively remediate some of the environmental damage caused by human wastewater. These eDNA results highlight the utility of human eDNA analysis (targeted qPCR or shotgun metagenomics) as a tool to detect, monitor and point-source locate human wastewater pollution in aquatic systems, and to quantify improvements of remediation works. Either targeted (e.g. human-specific qPCR/dPCR assays [23], field-based CRISPR diagnostic lateral flow eDNA tests [66], or mammal/vertebrate metabarcoding) or untargeted (shotgun metagenomic [23]) human eDNA approaches could be utilized for this purpose. Shotgun metagenomic approaches can simultaneously consider human, agricultural, and wildlife sources of aquatic pollution, as could well-designed multi-assay qPCR or selected metabarcoding approaches. Therefore, eDNA can provide a range of tools that could be tailored to

monitoring and improving pollution impacts on ecosystem health.

For metagenomic shotgun sequencing applications recovering human eDNA, there are ethical considerations to take into account, as discussed in Whitmore *et al.* (2023) [23]. Quantifying the level of human eDNA present in aquatic samples (e.g. by qPCR/dPCR, metabarcoding, CRISPR, or limited analysis metagenomics) does not pose as many ethical considerations [23], and should be sufficient for the vast majority of pollution point source detection and ecosystem health assessments. However, it should be recognized that the generation of potentially identifiable information-rich human eDNA data from metagenomic analysis of environmental samples, whether intentional or as unintentional by-catch from studies targeting other species, raises a number of ethical considerations [23, 67]. Such considerations are particularly pronounced if the resulting sequencing data are to be publicly deposited [23]. In such cases, it may be necessary to avoid depositing in public repositories or to scrub human-aligning reads from deposited data [5, 23]. Given the prevalence of human data in the data generated for this paper, human reads were removed (NCBI Human Read Removal Tool, HRRT; also known as the Human Scrubber) prior to the public release of the data.

Due to their ease of use, unbiased nature, and scalability, long-read shotgun metagenomic sequencing approaches can be used for widespread large-scale biodiversity, pathogen, or pollution surveillance, with key results then being followed up with targeted species-specific approaches, such as qPCR/dPCR. Additionally, as shotgun metagenomic approaches recover genetic information from across the genome (and mitogenome), the data generated can be applied to additional research areas such as phylogenetics, population genetics, pathogen, microbial, and multicellular organism variant analysis [5, 17–19, 22, 23, 68]. Barcodes only recover information from informative, but small, restricted genetic regions, typically smaller than 200 bp. Shotgun metagenomic eDNA sequencing enables a plethora of downstream analyses beyond just species identification, including viral variant analysis, population genetics, human genetic ancestry, and disease-associated variant calling [5, 15, 18, 22, 23, 68]. Mussel (*Mytilus)* transmissible cancer [47] surveillance was also applied to the shotgun metagenomic eDNA data. No transmissible cancer-associated reads were detected, with only wild-type (non-cancer mutated) *Mytilus* sequences present, indicating that either the transmissible cancer was not present in the area or that it occurred at a rate below the assay's detectable threshold. By simultaneously recovering both nuclear and mitochondrial genetic information, shotgun metagenomic sequencing avoids biases in the temporal stability of mitochondrial versus nuclear eDNA [19]. This is an important caveat not considered by the majority of animal/eukaryotic eDNA studies, which predominantly rely on mitochondrial barcodes and mitochondrial qPCR/dPCR targets [18, 19, 69]. Indeed, these differences in nuclear to mitochondrial DNA degradation rates will likely prove to have utility for temporally discriminating the duration since eDNA release from its host species [19].

Of all shotgun metagenomic-sequenced eDNA samples in this study, beach sand had amongst the highest number of genera detected and had a diverse metazoan phylum profile. This mirrors a similar dataset from Florida [5, 19], which indicated that beach sand acts as a natural filter of sea water, accumulating DNA from a wide variety of marine and terrestrial species (likely transported to the ocean by freshwater and air inputs). Therefore, we recommend detailed follow-up studies investigating the utility of beach sand sampling for robust marine (and terrestrial) biodiversity and species distribution surveying.

Environmental DNA research is reaching a crossroads. While metabarcoding approaches have proven incredibly powerful and have served the field well in establishing it as a strong force for biodiversity, pathogen, and wildlife monitoring, metabarcoding only scratches the surface of the rich genomic data present in environmental samples. eDNA can be used to uncover secrets hidden in those parts of genomes that sit outside the limited short barcode region, such as disease-associated variant analysis in vertebrate genomes, population genetics, evolutionary divergence rates, and pathogen variant analysis [5, 17–19, 22, 23, 68]. The pending transition toward more powerful sequencing approaches is akin to the transition that occurred in the medical field from microarrays (DNA/RNA) to next-generation sequencing (genomics/transcriptomics). Despite their unquestionable utility and significance in advancing biomedical research, microarrays were ultimately supplanted by more powerful and less biased shotgun (whole genome) sequencing. This change was initially delayed by investigators' reliance on and trust of microarrays, due to years of use, investment, and more established experience in the associated lab and analytical techniques. At its peak, 11 429 microarray papers were published in 2012, which dropped to 3480 papers in 2024 (PubMed search for "microarray," 01-Nov-2024). Conversely, in 2012 there were already 72 938 genomics papers, with 86 678 in 2024 (PubMed search for "genomics," 01-Nov-2024). Eventually, more experience and trust in new sequencing technologies were built as these technologies were demonstrated again and again to outperform older microarray approaches, and the entire field transitioned to sequencing technologies (including RNA-seq). Having experienced that technological transition firsthand, we suspect that we are on the cusp of a similar technological revolution with shotgun metagenomic sequencing in the field of multicellular eDNA. We also suspect that a similar pattern will play out: initial resistance and skepticism followed by growing acceptance, and then a near-ubiquitous switch.

This change will be aided by the versatility, continued reduction in cost, reduced labor and laboratory time, increased portability, improved analytical pipelines, and richer data provided by shotgun metagenomic sequencing [3, 5, 7, 17–19, 23–26]. This paper aims to contribute to that discussion and adoption, by highlighting the extensive utility of applying shotgun long-read metagenomic sequencing approaches to complex natural community environmental samples.

***Summary*:** Here we demonstrate the feasibility of applying long-read shotgun metagenomic sequencing to aquatic and sand eDNA samples, and its ability to recover genomic information from across the tree of life, including for multicellular species. We show how this information can be utilized to generate an unbiased quantification of the abundance of DNA between different domains of life, phyla, genera, and species. We also demonstrate that, for species with sufficient DNA abundance in the sample, analyses beyond mere species identification are possible. In this case, mitochondrial analyses from complex natural mixed community aquatic eDNA samples were carried out without requiring targeting or amplification of specific DNA targets. Therefore, we call for increased adoption and analysis of shotgun metagenomic sequencing for

multicellular eDNA applications (including further verification and testing of its benefits and limitations). Such adoption should not be restricted to microbial life as is currently the case, but used for the simultaneous study of all domains of life. We also call for targeted investment (funding and research effort) to develop, optimize, and make available eDNA shotgun metagenomic sequencing analysis tools, with an emphasis on ease of access and usability. We stand at a crossroads where, rather than resistance to shotgun metagenomic sequencing approaches for multicellular eDNA applications, adoption, application, and rigorous assessments should be embraced. Shotgun metagenomic sequencing will, at a minimum, complement the current dominant eDNA multi-cellular species approaches and, in all likelihood, exceed the capabilities of amplicon-based approaches, succeeding them as the most frequently applied eDNA tool in the near future. This approach is poised to enable biodiversity and wildlife assessments on the scale required to successfully implement and monitor the Convention on Biological Diversity's GBF conservation targets.

Acknowledgements

Sincerest thanks to Paula Murphy and Rebecca Rolfe for facilitating eDNA extraction of Irish samples in their lab in the Zoology Department, Trinity College Dublin, to Aleksandar Krstic, Walter Kolch, Amaya Garcia Munoz, and the Conway Core Facilities staff for facilitating qPCR of the samples at Systems Biology Ireland and the Conway Institute of Biomolecular and Biomedical Research at University College Dublin, and to BMKGene for conducting the sequencing. Thanks also to Mark Martindale, Nancy Condron, Jennifer Hines, Bogna Griffin, and Catherine Eastman. Some maps in this research paper were created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com. The Graphical Abstract was created in BioRender. Duffy, D. (2026) https://BioRender.com/bffiwb6.

Author contributions: D.J.D. and J.W. designed and supervised the project. D.J.D., F.G.D., I.J.D., and J.W. conducted sampling. D.J.D. generated the data. O.N., M.McC., and D.J.D. conducted bioinformatics analysis. All authors read and approved the final manuscript.

Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

Conflict of interest

None declared.

Funding

This study was funded by David Duffy's (DJD) University of Florida (UF) start-up funds.

Data availability

All sequenced samples are deposited in NCBI (https://www.ncbi.nlm.nih.gov/) under BioProject ID: PRJNA1044147. For individual NCBI SRR accession numbers for each sequenced sample, please see Supplementary Table S1. The metazoan metagenomics bioinformatics pipeline has been deposited in GitHub (https://github.com/nousiaso/MetaBioTax) and Zenodo (DOI: 10.5281/zenodo.18697671). The mussel transmissible cancer bioinformatics pipeline has been deposited in GitHub (https://github.com/nousiaso/Transmissible-Cancer-Identification-in-Mussels-from-long-read-eDNA-data) and Zenodo (DOI: 10.5281/zenodo.18697707).

References

1. Pfenning-Butterworth A, Buckley LB, Drake JM *et al.* Interconnecting global threats: climate change, biodiversity loss, and infectious diseases. *Lancet Planet Health* 2024;8:e270–83. https://doi.org/10.1016/S2542-5196(24)00021-4
2. Wolfe JD, Luther DA, Jirinec V *et al.* Climate change aggravates bird mortality in pristine tropical forests. *Sci Adv* 2025;11:eadq8086. https://doi.org/10.1126/sciadv.adq8086
3. Venter JC, Remington K, Heidelberg JF *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004;304:66–74. https://doi.org/10.1126/science.1093857
4. Stat M, Huggett MJ, Bernasconi R *et al.* Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Sci Rep* 2017;7:12240. https://doi.org/10.1038/s41598-017-12501-5
5. Nousias O, McCauley M, Stammnitz MR *et al.* Shotgun sequencing of airborne eDNA achieves rapid assessment of whole biomes, population genetics and genomic variation. *Nat Ecol Evol* 2025;9:1043–60. https://doi.org/10.1038/s41559-025-02711-w
6. Sandbrook C, Albury-Smith S, Allan JR *et al.* Social considerations are crucial to success in implementing the 30×30 global conservation target. *Nat Ecol Evol* 2023;7:784–5. https://doi.org/10.1038/s41559-023-02048-2
7. Stammnitz MR, Hartman Scholz A, Duffy DJ. Environmental DNA without borders. *EMBO Rep* 2024;25:4095–9. https://doi.org/10.1038/s44319-024-00264-w
8. Convention_on_Biological_Diversity. 2022. https://www.cbd.int/article/cop15-final-text-kunming-montreal-gbf-221222 (5 August 2024, date last accessed).
9. Aucone E, Kirchgeorg S, Valentini A *et al.* Drone-assisted collection of environmental DNA from tree branches for biodiversity monitoring. *Sci Robot* 2023;8:eadd5762. https://doi.org/10.1126/scirobotics.add5762
10. Abrego N, Furneaux B, Hardwick B *et al.* Airborne DNA reveals predictable spatial and seasonal dynamics of fungi. *Nature* 2024;631:835–42. https://doi.org/10.1038/s41586-024-07658-9
11. Lynggaard C, Calvignac-Spencer S, Chapman CA *et al.* Vertebrate environmental DNA from leaf swabs. *Curr Biol* 2023;33:R853–4. https://doi.org/10.1016/j.cub.2023.06.031
12. Aalismail NA, Díaz-Rúa R, Geraldi N *et al.* Diversity and sources of airborne eukaryotic communities (AEC) in the global dust belt over the Red Sea. *Earth Syst Environ* 2021;5:459–71. https://doi.org/10.1007/s41748-021-00219-4
13. Urban L, Holzer A, Baronas JJ *et al.* Freshwater monitoring by nanopore sequencing. *eLife* 2021;10:e61504. https://doi.org/10.7554/eLife.61504
14. Farrell JA, Whitmore L, Duffy DJ. The promise and pitfalls of environmental DNA and RNA approaches for the monitoring of human and animal pathogens from aquatic sources. *Bioscience* 2021;71:609–25. https://doi.org/10.1093/biosci/biab027
15. Farrell JA, Yetsko K, Whitmore L *et al.* Environmental DNA monitoring of oncogenic viral shedding and genomic profiling of sea turtle fibropapillomatosis reveals unusual viral dynamics. *Commun Biol* 2021;4:565. https://doi.org/10.1038/s42003-021-02085-2
16. Bass D, Christison KW, Stentiford GD *et al.* Environmental DNA/RNA for pathogen and parasite detection, surveillance, and

ecology. *Trends Parasitol* 2023;39:285–304. https://doi.org/10.1016/j.pt.2022.12.010
17. Koda SA, McCauley M, Farrell JA *et al.* A novel eDNA approach for rare species monitoring: application of long-read shotgun sequencing to Lynx rufus soil pawprints. *Biol Conserv* 2023;287:110315. https://doi.org/10.1016/j.biocon.2023.110315
18. Farrell JA, Whitmore L, Mashkour N *et al.* Detection and population genomics of sea turtle species via non-invasive environmental DNA analysis of nesting beach sand tracks and oceanic water. *Mol Ecol Resour* 2022;22:2471–93. https://doi.org/10.1111/1755-0998.13617
19. McCauley M, Koda SA, Loesgen S *et al.* Multicellular species environmental DNA (eDNA) research constrained by overfocus on mitochondrial DNA. *Sci Total Environ* 2024;912:169550. https://doi.org/10.1016/j.scitotenv.2023.169550
20. Kjær KH, Winther Pedersen M, De Sanctis B *et al.* A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nature* 2022;612:283–91.
21. Djurhuus A, Closek CJ, Kelly RP *et al.* Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nat Commun* 2020;11:254. https://doi.org/10.1038/s41467-019-14105-1
22. Crits-Christoph A, Levy JI, Pekar JE *et al.* Genetic tracing of market wildlife and viruses at the epicenter of the COVID-19 pandemic. *Cell* 2024;187:5468–82. https://doi.org/10.1016/j.cell.2024.08.010
23. Whitmore L, McCauley M, Farrell JA *et al.* Inadvertent human genomic bycatch and intentional capture raise beneficial applications and ethical concerns with environmental DNA. *Nat Ecol Evol* 2023;7:873–88. https://doi.org/10.1038/s41559-023-02056-2
24. Marx V. Method of the year: long-read sequencing. *Nat Methods* 2023;20:6–11. https://doi.org/10.1038/s41592-022-01730-w
25. Broman E, Bonaglia S, Norkko A *et al.* High throughput shotgun sequencing of eRNA reveals taxonomic and derived functional shifts across a benthic productivity gradient. *Mol Ecol* 2021;30:3023–39. https://doi.org/10.1111/mec.15561
26. Bista I, Carvalho GR, Tang M *et al.* Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Mol Ecol Resour* 2018;18:1020–34. https://doi.org/10.1111/1755-0998.12888
27. Downie AT, Bennett WW, Wilkinson S *et al.* From land to sea: environmental DNA is correlated with long-term water quality indicators in an urbanized estuary. *Mar Pollut Bull* 2024;207:116887. https://doi.org/10.1016/j.marpolbul.2024.116887
28. Perry WB, Seymour M, Orsini L *et al.* An integrated spatio-temporal view of riverine biodiversity using environmental DNA metabarcoding. *Nat Commun* 2024;15:4372. https://doi.org/10.1038/s41467-024-48640-3
29. Agustinho DP, Fu Y, Menon VK *et al.* Unveiling microbial diversity: harnessing long-read sequencing technology. *Nat Methods* 2024;21:954–66. https://doi.org/10.1038/s41592-024-02262-1
30. Liem M, Regensburg-Tuïnk T, Henkel C *et al.* Microbial diversity characterization of seawater in a pilot study using Oxford Nanopore Technologies long-read sequencing. *BMC Res Notes* 2021;14:42. https://doi.org/10.1186/s13104-021-05457-3
31. Benítez-Páez A, Sanz Y. Multi-locus and long amplicon sequencing approach to study microbial diversity at species level using the MinION™ portable nanopore sequencer. *Gigascience* 2017;6:1–12.
32. Lewin HA, Robinson GE, Kress WJ *et al.* Earth BioGenome Project: sequencing life for the future of life. *Proc Natl Acad Sci USA* 2018;115:4325–33. https://doi.org/10.1073/pnas.1720115115
33. Seymour M, Durance I, Cosby BJ *et al.* Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. *Commun Biol* 2018;1:4. https://doi.org/10.1038/s42003-017-0005-3
34. Duffy DJ, Konietzny A, Krstic A *et al.* Identification of a MYCN and Wnt-related VANGL2-ITLN1 fusion gene in neuroblastoma. *Gene Reports* 2018;12:187–200. https://doi.org/10.1016/j.genrep.2018.06.018
35. Kalantar KL, Carvalho T, de Bourcy CFA *et al.* IDseq—an open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *Gigascience* 2020;9:giaa111. https://doi.org/10.1093/gigascience/giaa111
36. Simmonds SE, Ly L, Beaulaurier J *et al.* CZ ID: a cloud-based, no-code platform enabling advanced long read metagenomic analysis. bioRxiv, https://doi.org/10.1101/2024.02.29.579666, 2 March 2024, preprint: not peer reviewed.
37. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;12:59–60. https://doi.org/10.1038/nmeth.3176
38. Hoffmann RJ, Boore JL, Brown WM. A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics* 1992;131:397–412. https://doi.org/10.1093/genetics/131.2.397
39. de Koning W, Miladi M, Hiltemann S *et al.* NanoGalaxy: nanopore long-read sequencing data analysis in Galaxy. *Gigascience* 2020;9:giaa105. https://doi.org/10.1093/gigascience/giaa105
40. Andrews S. FastQC: a quality control tool for high throughput sequence data. Cambridge, United Kingdom: Babraham Bioinformatics, Babraham Institute, 2010 [online, last accessed November 2023]. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
41. Wick R, Volkening J, Loman N. Porechop: a tool for finding and removing adapters from Oxford Nanopore reads. GitHub; 2017 [online, last accessed November 2023]. https://github.com/rrwick/Porechop
42. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–100. https://doi.org/10.1093/bioinformatics/bty191
43. Diesh C, Stevens GJ, Xie P *et al.* JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol* 2023;24:74. https://doi.org/10.1186/s13059-023-02914-z
44. Madeira F, Madhusoodanan N, Lee J *et al.* The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res* 2024;52:W521–5. https://doi.org/10.1093/nar/gkae241
45. Sievers F, Higgins DG. In: Kotah K (ed.), *Multiple Sequence Alignment: Methods and Protocols*. New York, NY: Springer US, 2021, Chapter 8, 3–16.
46. Sievers F, Wilm A, Dineen D *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539. https://doi.org/10.1038/msb.2011.75
47. Yonemitsu MA, Giersch RM, Polo-Prieto M *et al.* A single clonal lineage of transmissible cancer identified in two marine mussel species in South America and Europe. *eLife* 2019;8:e47788. https://doi.org/10.7554/eLife.47788
48. Li H, Murray KD, Langhorst B *et al.* Seqtk: a fast and lightweight tool for processing sequences in FASTA or FASTQ format. 2013. GitHub. Available from: https://github.com/lh3/seqtk. last accessed 14 January 2025
49. Danecek P, Bonfield JK, Liddle J *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* 2021;10:giab008. https://doi.org/10.1093/gigascience/giab008
50. Hulsen T, de Vlieg J, Alkema W. BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* 2008;9:488. https://doi.org/10.1186/1471-2164-9-488
51. Spitzer M, Wildenhain J, Rappsilber J *et al.* BoxPlotR: a web tool for generation of box plots. *Nat Methods* 2014;11:121–2. https://doi.org/10.1038/nmeth.2811
52. Nakano H. What is Xenoturbella? *Zoological Lett* 2015;1:22. https://doi.org/10.1186/s40851-015-0018-z

53. Bourlat SJ, Juliusdottir T, Lowe CJ *et al.* Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* 2006;444:85–8. https://doi.org/10.1038/nature05241
54. Fisher MC, Garner TWJ. Chytrid fungi and global amphibian declines. *Nat Rev Microbiol* 2020;18:332–43. https://doi.org/10.1038/s41579-020-0335-x
55. Scheele BC, Pasmans F, Skerratt LF *et al.* Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. *Science* 2019;363:1459–63. https://doi.org/10.1126/science.aav0379
56. Duffy DJ. Problems, challenges and promises: perspectives on precision medicine. *Brief Bioinform* 2016;17:494–504. https://doi.org/10.1093/bib/bbv060
57. Patin N, Pitz K, Baker J *et al.* Catching small fish in a big pond: targeted vs untargeted sequencing for marine eukaryotic eDNA. *Authorea Preprint* 2024. https://doi.org/10.22541/au.172560271.10683314/v1
58. Bell KL, Petit RA, III, Cutler A *et al.* Comparing whole-genome shotgun sequencing and DNA metabarcoding approaches for species identification and quantification of pollen species mixtures. *Ecology and Evolution* 2021;11:16082–98. https://doi.org/10.1002/ece3.8281
59. Twyford A, Beasley J, Barnes I *et al.* A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life Project. *Wellcome Open Res* 2024;9:339. https://doi.org/10.12688/wellcomeopenres.21143.1.
60. Guo M, Yuan C, Tao L *et al.* Life barcoded by DNA barcodes. *Conservation Genet Resour* 2022;14:351–65. https://doi.org/10.1007/s12686-022-01291-2
61. Antil S, Abraham JS, Sripoorna S *et al.* DNA barcoding, an effective tool for species identification: a review. *Mol Biol Rep* 2023;50:761–75. https://doi.org/10.1007/s11033-022-08015-7
62. Ahmed W, Angel N, Edson J *et al.* First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community. *Sci Total Environ* 2020;728:138764. https://doi.org/10.1016/j.scitotenv.2020.138764
63. Hart OE, Halden RU. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: feasibility, economy, opportunities and challenges. *Sci Total Environ* 2020;730:138875. https://doi.org/10.1016/j.scitotenv.2020.138875
64. Zan R, Blackburn A, Plaimart J *et al.* Environmental DNA clarifies impacts of combined sewer overflows on the bacteriology of an urban river and resulting risks to public health. *Sci Total Environ* 2023;889:164282. https://doi.org/10.1016/j.scitotenv.2023.164282
65. Ahmed W, Hughes B, Harwood VJ. Current status of marker genes of *Bacteroides* and related taxa for identifying sewage pollution in environmental waters. *Water* 2016;8:231. https://doi.org/10.3390/w8060231
66. Leugger F, Lüthi M, Schmidlin M *et al.* Rapid field-based detection of a threatened and elusive species with environmental DNA and CRISPR-Dx. *Global Ecology and Conservation* 2025;59:e03518. https://doi.org/10.1016/j.gecco.2025.e03518
67. Ram N. The ethics of human sequences in environmental samples. *Nat Ecol Evol* 2023;7:796–7. https://doi.org/10.1038/s41559-023-02072-2
68. Liu WJ, Liu P, Lei W *et al.* Surveillance of SARS-CoV-2 at the Huanan Seafood Market. *Nature* 2024;631:402–8. https://doi.org/10.1038/s41586-023-06043-2
69. Nagler M, Podmirseg SM, Ascher-Jenull J *et al.* Why eDNA fractions need consideration in biomonitoring. *Mol Ecol Resour* 2022;22:2458–70. https://doi.org/10.1111/1755-0998.13658

Received: December 17, 2024. **Revised:** February 12, 2026. **Accepted:** February 16, 2026