RESEARCH

Open Access

Primed and ready: nanopore metabarcoding can now recover highly accurate consensus barcodes that are generally indel-free

Jia Jin Marc Chang[1*], Yin Cheong Aden Ip[1,2], Wan Lin Neo[1], Maxine A. D. Mowe[1], Zeehan Jaafar[1,3,4] and Danwei Huang[1,3,4,5]

Abstract

Background DNA metabarcoding applies high-throughput sequencing approaches to generate numerous DNA barcodes from mixed sample pools for mass species identification and community characterisation. To date, however, most metabarcoding studies employ second-generation sequencing platforms like Illumina, which are limited by short read lengths and longer turnaround times. While third-generation platforms such as the MinION (Oxford Nanopore Technologies) can sequence longer reads and even in real-time, application of these platforms for metabarcoding has remained limited possibly due to the relatively high read error rates as well as the paucity of specialised software for processing such reads.

Results We show that this is no longer the case by performing nanopore-based, cytochrome c oxidase subunit I (COI) metabarcoding on 34 zooplankton bulk samples, and benchmarking the results against conventional Illumina MiSeq sequencing. Nanopore R10.3 sequencing chemistry and super accurate (SUP) basecalling model reduced raw read error rates to ~4%, and consensus calling with amplicon_sorter (without further error correction) generated metabarcodes that were ≤ 1% erroneous. Although Illumina recovered a higher number of molecular operational taxonomic units (MOTUs) than nanopore sequencing (589 vs. 471), we found no significant differences in the zooplankton communities inferred between the sequencing platforms. Importantly, 406 of 444 (91.4%) shared MOTUs between Illumina and nanopore were also found to be free of indel errors, and 85% of the zooplankton richness could be recovered after just 12–15 h of sequencing.

Conclusion Our results demonstrate that nanopore sequencing can generate metabarcodes with Illumina-like accuracy, and we are the first study to show that nanopore metabarcodes are almost always indel-free. We also show that nanopore metabarcoding is viable for characterising species-rich communities rapidly, and that the same ecological conclusions can be obtained regardless of the sequencing platform used. Collectively, our study inspires confidence in nanopore sequencing and paves the way for greater utilisation of nanopore technology in various metabarcoding applications.

Keywords Biomonitoring, Illumina, MinION, Next-generation sequencing, Species diversity, Zooplankton

*Correspondence:
Jia Jin Marc Chang
jjmc@u.nus.edu

Full list of author information is available at the end of the article

Background

DNA metabarcoding refers to the high-throughput sequencing of total (and sometimes degraded) DNA from bulk or environmental samples (e.g., air, water, soil, faeces, etc.) with the goal of multispecies identification [1]. It was built upon the DNA barcoding paradigm that has been established for about two decades involving the sequencing of short segments of DNA (termed "barcodes") and matching them to sequence databases to obtain species identities [2]. DNA metabarcoding emerged in the 2010s, and was primarily made possible due to rapid advancements in nucleic acid sequencing technologies—with "next-generation sequencing" (NGS) platforms—which have the ability to generate billions of sequence reads in a single experiment [3]. This development has been groundbreaking due to the sheer ability of NGS platforms to generate sequence reads (i.e., DNA barcodes) in parallel, so multispecies detections and identification from various sample types are now possible. This has led to a meteoric rise in the number of studies that have since performed NGS-based barcoding or metabarcoding for various applications. For instance, 60% of DNA sequencing studies in marine science published yearly between 2013 and mid-2022 generated their sequence reads with Illumina [4]. The release of the MinION in 2014 by Oxford Nanopore Technologies (ONT) became another significant milestone in nucleic acid sequencing for several reasons: [1] its lower entry and per-base sequencing cost (2,000 USD for the entry starter pack) [2], its ability to perform long-read sequencing (now up to ~4 Mb long) [3], its compact size and portability, and [4] its ability to generate data in real-time [5, 6]. All these were perhaps a direct response to common criticisms of Illumina sequencing, which is comparatively more expensive, and limited by its short read-lengths (up to ~500 bp). Since then, nanopore sequencing has been applied in numerous whole-genome sequencing studies [7–10] and metagenomic studies [11, 12].

However, nanopore metabarcoding applications remain relatively uncommon, and this is evident in the handful (but increasing) number of published papers, especially in biodiversity-related fields. Such studies focused on microbes [13–18], and few have paid attention to non-microbial taxa until more recently. Importantly, Krehenwinkel et al. [19] and Baloğlu et al. [20] laid the groundwork with ONT's MinION sequencer by successfully metabarcoding mock communities comprising nine arthropod and 50 aquatic invertebrate species respectively. Other studies have since applied nanopore metabarcoding for biodiversity and community characterisation [13, 21–25], species-specific detections [26, 27], and even gut content analysis [28, 29] with actual samples. The growing consensus from the

abovementioned studies is that nanopore sequencing shows promise in metabarcoding.

We posit that the general lack of nanopore-based metabarcoding studies can be attributed to two main factors. The first is the perception that nanopore reads are highly erroneous. This is unsurprising given that early studies have reported error rates of ~20% [30] to as high as 38% [31]. In contrast, the current error rate of Illumina sequencing is only 0.24% [32]. There is thus concern that the high error rates would hinder accurate species identification in DNA metabarcoding. The second factor could be the lack of programs to process nanopore reads for metabarcoding (but see below), compared to the plethora of pipelines catered to short-read sequencing, like APSCALE [33], DADA2 [34], eDNAflow [35], or OBI-Tools [36]. DADA2 currently supports PacBio circular consensus sequencing but not nanopore reads [37], and even ONT's own EPI2ME platform is intended for microbial sequencing only. Nanopore-specific workflows like ONTrack [38], NGSpeciesID [39] and miniBarcoder [40, 41] were designed mainly for DNA barcoding, although Davidov et al. [13] have successfully applied ONTrack to process their metabarcoding reads. Prior metabarcoding studies have worked around the lack of specialised software by either: (i) conducting BLAST searches of raw nanopore reads with stricter *e*-value settings as low as $1e^{-40}$ to minimise erroneous matches due to chance [21, 26], (ii) using custom reference databases for mapping and processing reads [23], or (iii) using existing programs designed for short reads, like VSEARCH [42] or CD-HIT [43] with more relaxed settings for clustering error-prone nanopore reads [28, 44].

We expect that nanopore metabarcoding studies will become more common, given the release of new nanopore metabarcoding workflows like ASHURE [20], decona [45] and MSI [27], its real-time sequencing capabilities, as well as improvement in flow cell chemistries and base calling models over time. The latter is evidenced in the decreasing raw read error rate to ~6% using R9.4 flow cell chemistry [46], and even lower at ~4% for R10.3 flow cells [47]. Two research groups have since independently confirmed that it is possible to generate highly-accurate, Illumina-like, DNA barcodes without further need for error correction with R10.3 sequencing chemistry [48, 49]. As of writing, raw read accuracy is now ~99% with the latest R10.4.1 sequencing chemistry and base calling models (see https://rrwick.github.io/ for more up-to-date information).

In light of these improvements in sequencing accuracy, we propose that the time is ripe for broader-scale nanopore metabarcoding, and on more complex biological communities. In this study, we performed mitochondrial cytochrome c oxidase subunit I (COI) metabarcoding on species-rich, bulk zooplankton samples collected from

the tropical waters of Singapore. We then benchmarked the relative abundance and community composition of molecular operational taxonomic units (MOTUs) obtained from nanopore sequencing against Illumina sequencing—the current gold standard for metabarcoding sequencing—to investigate if the sequencing platform affects community characterisation of zooplankton communities. We show that processing nanopore reads with available programs like amplicon_sorter [48] produces highly-accurate consensus metabarcodes that are Illumina-like in accuracy. To the best of our knowledge, this is the first study to demonstrate that nanopore consensus metabarcodes are almost always indel-free, even with R10.3 chemistry. This is also an advancement over existing workflows that incorporate clustering and subsequent polishing steps as these sequences would still retain indel errors, thereby reducing confidence in their quality. We further demonstrate that such high-quality metabarcodes can be obtained without the need for complicated wet-laboratory procedures like rolling circle amplification as with the ASHURE workflow, or even error correction programs, like in the MSI and decona pipelines. Moreover, we were able to recover ~85% of zooplankton richness with 12–15 h of sequencing run time. Our study demonstrates the viability of nanopore metabarcoding for analysing complex, biodiverse communities, and we hope this inspires greater confidence in nanopore sequencing for a greater variety of metabarcoding applications.

Methods
Sample collection and processing
The study samples comprised a series of zooplankton collections made during August–September 2020 in Singapore. Collections were permitted by the National Parks Board, Singapore (Permit Number NP/RP18-051). The targeted sites were off Pulau Hantu and Sisters' Islands in the Singapore Strait (See Supplementary File S1 for GPS coordinates). All plankton collections were performed at night (1800–2200 h), and sampling was conducted in two ways. First, triplicate oblique plankton tows were performed from a boat with bongo nets (2 m in length, 500 μm mesh size, 50 cm ring diameter) from a depth of 15 m to the surface at 1 m/s. The plankton net was always rinsed with fresh water before each tow, and its contents were collected as the field negative control. After each tow, the contents from one cod-end were poured through 2 mm and 500 μm sieves to filter excess seawater before bulk preservation in molecular-grade ethanol [50]. Specimens larger than 1 cm were picked out individually. The collections were thus separated into three size fractions—1 cm, 2 mm and 500 μm. Second, a quatrefoil light trap (30 cm diameter by 25 cm height; 5 mm entry slit width) fitted with two GT-AAAs (Glo-Toob) was left at the jetty of each island 1.5 m below the water surface

for two hours (See Supplementary File S1 for GPS coordinates). Light trap samples were processed in the same way as bongo net samples. All bulk samples were brought back to the laboratory and stored at -20 °C prior to DNA extraction.

DNA extraction and PCR amplification
Bulk samples were first ground with pre-sterilized mortar and pestles. Genomic extraction was performed with DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol, except that genomic DNA was eluted in nuclease-free water. To prevent cross-contamination, a fresh set of autoclaved mortar and pestle was used for each tow/light trap. All units were thoroughly washed and autoclaved before the next set of DNA extractions.

We amplified the 313-bp fragment of mitochondrial COI for direct comparison of PCR products across short- and long-read platforms. PCR amplification was performed using the mlCOIintF: 5'-GGW ACW GGW TGA ACW GTW TAY CCY CC-3' [51] and LoboR1: 5'-TAA ACY TCW GGR TGW CCR AAR AAY CA-3' [52] primer combination. This primer combination was also chosen for its high amplification success in marine organisms [53–56], and is approximately four times cheaper than the conventional mlCOIintF and jgHCO2198 [57] metabarcoding primer pair [28, 58]. Furthermore, Yeo et al. [59] have also demonstrated that 313-bp COI sequences performed just as well as 658-bp barcodes for species-level identification. The primers were tagged at the 5' end with custom 13-bp sequences (i.e., "tags") from Srivathsan et al. [41] to allow for downstream demultiplexing of sequence reads to samples. The longer-than-usual tag lengths were necessary to accommodate the error profile of Kit 9 and R10.3 sequencing chemistry [41] (though it was recently reported that shorter 9-bp tags work well for R10.4.1 sequencing kits and flow cells [60]). Each PCR was assigned its own unique forward and reverse tag combination where possible, and if there were overlapping tag combinations, we separated them into different library pools (i.e., Plate A and B).

PCR was carried out in 25 μl triplicate reactions using 2 μl genomic DNA (100× dilution of original extract), 12.5 μl of GoTaq Green Master Mix (Promega), 2 μl of 10 μM 13-bp tagged forward and reverse primers, 1 μl of bovine serum albumin (1 mg/ml; New England Biolabs) and 7.5 μl of nuclease-free water. A step-up thermocycling profile was used: 1 min denaturation at 94 °C; 5 cycles of 30 s at 94 °C; 2 min at 45 °C; 1 min at 72 °C; 30 cycles of 30 s at 94 °C; 2 min at 55 °C; 1 min at 72 °C and a final extension of 3 min at 72 °C. All PCR products were screened on 2% agarose gels stained with GelRed (Biotium Inc.) to ensure appropriate amplification. PCR amplicons were subsequently combined by plate into two

pools and purified with SureClean Plus (Bioline). Plate A and B had 48 and 72 amplicons (including negatives and controls) respectively. In total, 34 samples, four field controls, and two PCR negatives were carried forward for Illumina and nanopore library preparation (40 × 3 PCRs = 120 amplicons).

Illumina metabarcoding and bioinformatics

We prepared two Illumina libraries using NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) following the manufacturer's protocol, up till the adapter ligation step (i.e., PCR-free libraries). Libraries were multiplexed using TruSeq CD Dual Indexes (Illumina). Cleanups were performed using 1.0× AMPure XP beads (Beckman Coulter). The two libraries were pooled together and outsourced for sequencing on a single Illumina MiSeq (2×250-bp) lane at the Genome Institute of Singapore.

Illumina reads were processed according to a modified metabarcoding pipeline from Sze et al. [61] and Ip et al. [62]. First, Illumina paired-end reads were merged using PEAR v0.9.6 [63]. Thereafter, OBITools v1.2.13 [36] was used for downstream processing of assembled reads. Specifically, the ngsfilter module was used to demultiplex reads to respective PCR replicates under default settings, where up to 2-bp mismatch was allowed for primer sequences, but no mismatch allowed for tag sequences. Sequence reads were then dereplicated and sorted to samples with obiuniq and obisubset respectively. We retained sequences with ≥5 counts and between 303- and 323-bp in length using obigrep. Subsequently, the filtered reads were further collapsed with obiclean, where sequences with 1-bp difference from each other were considered sequencing errors and further collapsed, and only reads with 'head' status were retained. We then concatenated all sequences across all samples, and ran cd-hit-est v.4.8.1 [43] to collapse 100% identical sequences. Any sequence that clustered with PCR negatives or control samples at 100% were eliminated.

Nanopore metabarcoding and bioinformatics

The same cleaned amplicon pools were used to prepare two nanopore libraries with the Ligation Sequencing Kit (SQK-LSK109) following the manufacturer's protocol, but end-repair and adapter ligation times were increased to 60 and 15 min respectively [58]. Cleanups were likewise done using 0.9× AMPure XP beads (Beckman Coulter) and the supplied Short Fragment Buffer (SFB). Finally, the two libraries were each sequenced on fresh R10.3 MinION flow cells on MinKNOW v20.10.3 for Ubuntu 16. The R10.3 flow cell chemistry was selected given its improved accuracy and homopolymer resolution [49, 64]. RUN A lasted 20 h and 30 min, while RUN B lasted 41 h.

Raw fast5 reads were exported to the National University of Singapore's High Performance Computing Volta cluster for GPU basecalling on NVIDIA Tesla V100 SXM2 32GB with Guppy v5.0.14+8f53ee9, using the super accurate (SUP) model at default settings. We then performed a length filter with NanoFilt v2.8.0 [65] to retain only sequence reads ≥250-bp in length. Subsequently, the sequences were distributed to respective PCR replicates with the demultiplexing module of ONT-barcoder v0.1.9 [49]. We set 313-bp as the read length threshold, and kept the other settings as default. Only sequences deviating up to 2-bp from the tag sequence were accepted in the demultiplexing process, which was possible as tags were designed to differ by ≥3-bp from each other [41]. Moreover, ONTbarcoder recognises and splits self-ligated reads during demultiplexing, thereby retaining more reads for downstream analysis. Thereafter, we concatenated the reads by sample.

For metabarcoding analysis, we used the amplicon_sorter v2022-03-28 [48] to sort and group the nanopore reads based on length and sequence similarity in order to generate consensus metabarcodes. We selected it for three reasons. First, amplicon_sorter performs reference-free clustering which is extremely useful in our case since we did not have a priori knowledge of the community composition of our zooplankton samples. Second, amplicon_sorter considers all possible clusters when generating consensus sequences, meaning it can be utilised to analyse DNA metabarcoding data. Third, amplicon_sorter corrects for indel errors when calling the majority consensus, thereby generating Illumina-like quality metabarcodes that will almost always be indel-free. This was unachievable with our prior tests of the same dataset using VSEARCH and CD-HIT, and subsequent polishing with RACON [66] and medaka (https://github.com/nanoporetech/medaka), and most nanopore metabarcodes still contained indel-errors after polishing (data not shown).

We adopted a conservative approach where sequences were added into a species group by amplicon_sorter only if they were ≥97% similar (--similar_species), and consensus sequences were combined together only if they were ≥98% similar (--similar_consensus). We also set the minimum and maximum length limits to 293- and 333-bp respectively, and performed 3× random sampling (--maxreads) to increase likelihood of sampling rare reads. We then mapped the sequences of each cluster back to the respective consensus sequence with minimap2 v2.24 [67] and polished the consensus with medaka v1.7.2, using the r103_sup_g507 model. Finally, we removed sequences that were present in our PCR negatives and controls from the samples using the same method described for Illumina metabarcoding.

MOTU delimitation and community analysis

We concatenated both Illumina and nanopore datasets together and aligned the sequences with MAFFT v.7.487 [68], before grouping them into MOTUs with objective clustering (https://github.com/asrivathsan/obj_cluster) at the 3% threshold. This was consistent with distance thresholds applied in past studies on marine invertebrates for Singapore [50, 53, 62]. We then ran blastn (*e*-value: $1e^{-6}$ and 80% identity) with BLAST+v2.12.0 [69] against the NCBI *nt* database (downloaded 13 June 2022), and obtained taxonomic identities for blast hits that had 85% identity match and minimum 250-bp overlap with readsidentifier v1.1.2 [70]. With the taxa identified, we grouped our Illumina MOTUs for a translation check on Geneious Prime v2022.2.2. (http://www.geneious.com/), using codes 2 (Chordata), 4 (Cnidaria), 5 (all other invertebrates), 9 (Echinodermata and Rhabditophora) and 13 (Ascidiacea). Illumina sequences that failed the translation check were considered possible nuclear mitochondrial DNA (NUMT) and discarded. For MOTUs that matched at ≥90%, we also screened the taxonomic identities against World Register of Marine Species (WoRMS; downloaded 8 May 2022) to confirm the MOTUs were marine, and also against past studies [50, 53, 62, 64, 71, 72], as well as SeaLifeBase (https://www.sealifebase.ca/), to confirm each MOTU's geographic ranges were within the Indo-Pacific.

With the final consolidated MOTU dataset, we assessed if and how MOTU communities compared between sequencing types quantitatively using diversity metrics, PERMANOVA, and qualitatively by examining the agreement in MOTU composition in terms of proportion and abundance. All statistical analyses were performed in R v4.3.1 [73], in RStudio (build 2023.03.0) unless otherwise stated, and all relevant plots were generated with the ggplot2 v3.4.2 package [74]. We computed the MOTU richness, Shannon-Wiener, and Simpson indices for each sequencing dataset using the diversity function in vegan v2.6-4 [75] and ran a paired, nonparametric Wilcoxon signed-rank test to test whether differences in the indices were due to different sequencing platforms. We also plotted the rarefaction curves of MOTU richness for each dataset with iNEXT v3.0.0 [76] to examine the relationship between MOTU richness and sampling depth. Community similarities between sequencing types were assessed using: (i) the Jaccard similarity coefficient by converting the MOTU community matrix to binary absence/presence data; and (ii) also with Bray-Curtis distances, where we normalised our MOTUs by relative abundance of sequencing reads [77]. We visualised the distances using nMDS plots (metaMDS in vegan) and heatmaps constructed with pheatmap v1.0.12 package [78]. We also performed PERMANOVA with adonis2 in vegan to test for community differences between Illumina

and nanopore sequencing. Here, sequencing type (Illumina or nanopore) was included as a variable, in addition to site (Pulau Hantu or Sisters' Islands), date (5 August 2020, 19 August 2020, 20 August 2020, 2 September 2020, 3 September 2020 or 16 September 2020), as well as fraction (1 cm, 2–500 μm). We first verified that each variable had a non-significant betadisper result before inclusion into PERMANOVA. We also analysed the datasets separately to confirm the same ecological conclusions would be obtained regardless of sequencing type. For this, we used the same Bray-Curtis distance datasets, and visualised the community dissimilarities with nMDS. For PERMANOVA, we only incorporated the bongo net samples as that sampling method had the most samples. We used the same three variables (site, date, fraction) and groupings as above for PERMANOVA with adonis2.

We also examined MOTU community compositions to determine how consistent they were between nanopore and Illumina platforms. We first looked at MOTU composition based on phyla, and compared the relative proportions of each phylum at the sequencing dataset level, and further at the sample level. In addition, we were also interested to know if a MOTU that was abundant in nanopore sequencing would be similarly so with Illumina sequencing. For each sample, we sorted and ranked the MOTUs by sequencing reads, and then assessed similarity in rank order of MOTUs between sequencing platforms with Kendall rank correlation coefficient (Kendall's τ) [79]. We performed the correlation analysis only for 31 out of 34 samples as the remaining three samples had only one pairwise comparison.

Sequencing accuracy and quality of nanopore reads

A known drawback of nanopore sequencing is its relatively high error rates. A close examination of the error rates of the raw reads and consensus sequences here was thus necessary to allay existing concerns regarding its use. We mapped the nanopore sequences against the cleaned Illumina sequences at the sample-level (e.g., ZPT005 nanopore reads to ZPT005 Illumina reads) with mapPacBio.sh v38.96 in BBTools (script was also recommended for nanopore data; https://sourceforge.net/projects/bbmap/). We maximised mapping sensitivity with the --vslow flag, and mapped two datasets: (i) the demultiplexed reads from ONTbarcoder to estimate raw read error rates and (ii) consensus sequences generated from amplicon_sorter to assess consensus sequence quality. We only considered mappings where the nanopore queries had ≥90% identity match to the Illumina reference sequences, and computed the total error rates, which took into account substitutions, insertions, deletions and ambiguous bases.

Additionally, for each MOTU shared between Illumina and nanopore datasets, we further compared the

constituent Illumina and nanopore member sequences of that MOTU with dnadiff v1.3 [80]. As our Illumina sequences were already confirmed to be translatable, and are thus free of frameshift errors and unlikely NUMTs, this comparison allowed us to assess the frequency of indel errors in our nanopore consensus sequences.

Time sampling of nanopore reads

Given the real-time sequencing properties of the MinION, we also preliminarily examined the relationship between sequencing run time and its effect on the nanopore metabarcoding. It was previously observed that 80–90% of DNA barcodes were obtained within the first few hours of sequencing [40, 49] for DNA barcoding studies. Here, we tested if the observed trends would be similar in a nanopore metabarcoding context. We subsampled the nanopore reads generated from each run for every hour for the first three hours of sequencing, followed by every three hours thereafter, until 18 h for RUN A and 39 h for RUN B. For each time period, we repeated the entire workflow from Guppy basecalling to amplicon_sorter (see section '*Nanopore metabarcoding and bioinformatics*'). For each time point, we noted down (i) the number of raw reads generated, (ii) the number of reads demultiplexed by ONTbarcoder, and (iii) the number of metazoan MOTUs obtained for each time series dataset.

Results

Zooplankton collections

A total of 49 bulk zooplankton samples—24 and 25 from Pulau Hantu and Sisters' Islands respectively—were collected and included in this study (Supplementary File S1). Of the 49 samples, 37 were bongo net samples, seven were light trap samples, and five were field control samples. After sieving and sorting, the 500 μm size fraction was the most common (29 samples), followed by 2 mm (18 samples), with the 1 cm fraction class having the least (2 samples). PCR amplification was successful for 34 samples (28 bongo net and 6 light trap samples), and nanopore and Illumina libraries were prepared for a total of 40 samples for this comparative study (including four field controls and two PCR negatives).

Metabarcoding and MOTU delimitation

For Illumina sequencing, we generated 10,038,735 paired-end reads on a single Illumina MiSeq lane, 7,630,728 reads were successfully assembled with PEAR, 4,218,977 reads were successfully demultiplexed (55.3% demultiplexing success), and 4,162,498 reads remained after the length filter. Most Illumina reads dropped out at the PEAR assembly stage due to Q-score filtering, and during the demultiplexing step due to strict settings (no mismatches allowed in tags). We obtained 10,788 clean

haplotypes after removing sequences present in controls and PCR negatives.

For nanopore sequencing, we generated 20,045,167 raw reads from across two MinION sequencing runs (RUN A and B). We retained 14,123,752 reads after Guppy basecalling and NanoFilt, and 6,918,618 reads after demultiplexing with ONTbarcoder (48.6% demultiplexing success). The low demultiplexing success rate is common for 13-bp tagged primers and sequencing with R10.3 chemistry [41, 64, 81], but will not be a cause for concern as ~60% demultiplexing success rates are obtainable with R10.4.1 chemistry [82]. Consensus calling with amplicon_sorter generated a total of 4,206 sequences from 3,525,077 reads (51% of demultiplexed reads). At the sample level, 57.6% of demultiplexed reads were utilised by the program to generate consensus sequences on average, with a minimum of 47.1–73.3% maximum. The median length was 313-bp (62% of total sequences generated); minimum and maximum sequence lengths were 300- and 339-bp respectively. We also observed that amplicon_sorter very rarely generated consensus sequences from different "gene groups" (two samples had one consensus sequence each while only one sample had five such consensus sequences). These were found to be of non-mitochondrial origin when we conducted nucleotide BLAST searches on NCBI web servers, and were thus excluded from the dataset. After filtering sequences present in the negatives and controls, we retained 3,973 consensus sequences (3,295,247 reads). As polishing with medaka had a minimal impact in reducing error rates (~0.02% decrease), we carried out the analysis using the unpolished dataset instead (see [48]).

From the combined sequencing dataset, we obtained 1,031 molecular operational taxonomic units (MOTUs) at the 3% threshold, with only 688 identified (at 85% identity match with ≥250-bp overlap) via readsidentifier. We discarded 61 MOTUs (four unclassified environmental samples, 35 Rhodophyta, 10 Fungi, eight Bacillarophyta, two Phaeophyceae, one Dinophyceae, and one Oomycota). We further eliminated one Illumina MOTU for failing the translation check, and 10 MOTUs that matched non-marine Insecta. None of the remaining MOTUs' geographic ranges fell outside the Indo-Pacific. Our final dataset comprised 616 Metazoa MOTUs, of which 316 had ≥97% match to a sequence on NCBI *nt* database, and 274 out of 316 obtained a species-level identity (Supplementary File S1).

Comparing nanopore and Illumina metabarcoding

The proportion of demultiplexed reads assigned to each sample was largely consistent across both Illumina and nanopore sequencing for most samples (Fig. 1a). Illumina recovered a higher number of MOTUs (589 vs. 471) than nanopore, but species accumulation curves
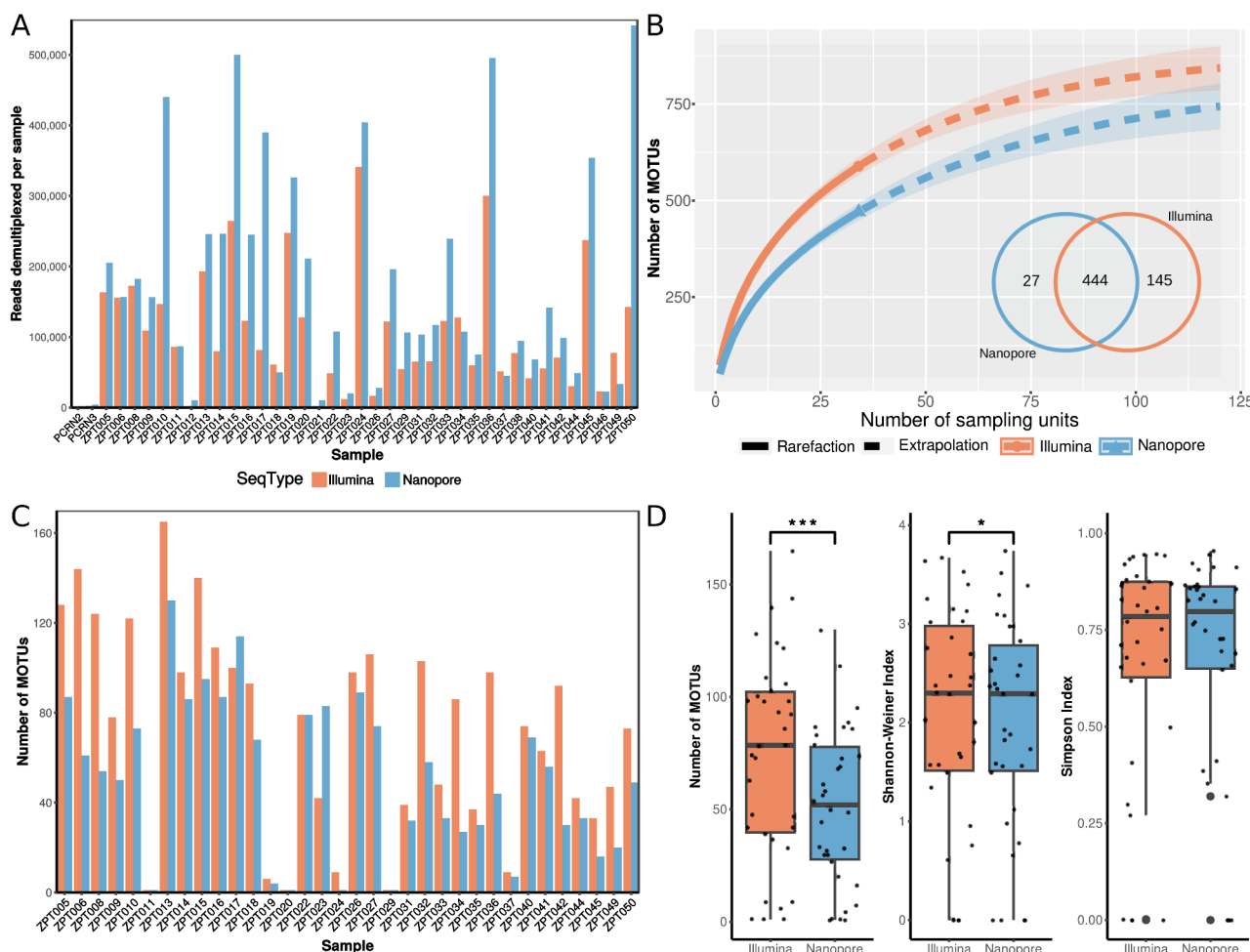
Fig. 1 Sequencing statistics of zooplankton metabarcoding with Illumina MiSeq and Nanopore MinION. (**a**) Bar plot of sequence reads demultiplexed per sample per sequencing dataset. (**b**) Species accumulation curves of molecular operational taxonomic unit (MOTU) richness for each sequencing platform against the number of samples, extrapolated to visualise number of samples needed to capture maximum richness; number of MOTUs obtained (and shared) expressed in Venn (insert). (**c**) Bar plots showing the number of MOTUs obtained per sample per sequencing type. (**d**) Box plots comparing MOTU richness, Simpson index, and Shannon-Weiner index between sequencing platforms; asterisks indicate significant differences for paired Wilcoxon signed-rank tests, and dots represent individual sample points (jittered)

suggested that ~120 samples were needed to fully capture zooplankton diversity for both sequencing types (Fig. 1b). 444 MOTUs were shared (72% overlap) across both sequencing platforms, with more MOTUs unique to Illumina than to nanopore (Fig. 1b, insert). At the sample-level, Illumina metabarcoding also consistently recovered more MOTUs than nanopore, with the exception of ZPT017 and ZPT023 (Fig. 1c). MOTU richness (p-value = 4.056×10^{-5}) and Shannon-Wiener diversity (p-value = 0.03) were found to be significantly different across paired samples, while Simpson diversity was not (p-value = 0.63, Fig. 1d). Even so, we observed clustering by sample on the nonmetric multidimensional scaling (nMDS) plots, especially with the Bray-Curtis distance metric (Fig. S1). This suggested that although MOTU richness differed across paired samples, the relative abundance of MOTUs within each sample were quite similar

across both sequencing platforms. Permutational multivariate analysis of variance (PERMANOVA) revealed significant differences in communities for both Jaccard and Bray-Curtis datasets (Jaccard: df = 27, F = 1.2329, R^2 = 0.4542, p = 0.0014; Bray-Curtis: df = 27, F = 1.6542, R^2 = 0.52754, p = 0.0001), but the differences were driven by the other three variables and not sequencing type (Table 1). When each sequencing dataset was analysed separately, we noted the same ecological conclusions from the nMDS plots and PERMANOVA as well—that the bongo net zooplankton communities were structured by date, fraction and site regardless of the sequencing platform (Fig. 2; Table 2).

Since MOTU richness differed between each sample's Illumina and nanopore datasets, we checked if this difference altered the respective community compositions. Both Illumina and nanopore recovered all 10 metazoan

Chang *et al. BMC Genomics*   (2024) 25:842

Page 8 of 16

Table 1 Permutational multivariate analysis of variance (PERMANOVA) results comparing community differences between nanopore and Illumina metabarcoding datasets, with Jaccard coefficient and bray-Curtis dissimilarity. Variables with significant *p*-values are highlighted in bold

Jaccard

	df	Sum of squares	R²	*F*-value	*p*-value
SeqType	1	0.2065	0.00827	0.6062	0.987
Site	**1**	**0.9791**	**0.03922**	**2.8744**	**0.001**
Date	**4**	**3.0335**	**0.12151**	**2.2264**	**0.001**
Fraction	**2**	**2.4961**	**0.09999**	**3.6639**	**0.001**
SeqType: Site	1	0.0783	0.00314	0.2298	1.000
SeqType: Date	4	0.4419	0.01770	0.3244	1.000
SeqType: Fraction	2	0.2286	0.00916	0.3355	1.000
Site: Fraction	**2**	**1.3542**	**0.05424**	**1.9877**	**0.001**
Date: Fraction	**4**	**2.0055**	**0.08034**	**1.4719**	**0.001**
SeqType: Site: Fraction	2	0.2235	0.00895	0.3281	1.000
SeqType: Date: Fraction	4	0.2916	0.01168	0.2140	1.000
Residual	40	13.6253	0.54580		
Total	67	24.9641	1.00000		
Bray-Curtis					
SeqType	1	0.0607	0.00263	0.2231	0.999
Site	**1**	**1.2093**	**0.05248**	**4.4435**	**0.001**
Date	**4**	**4.0874**	**0.1774**	**3.7547**	**0.001**
Fraction	**2**	**3.7336**	**0.16204**	**6.8594**	**0.001**
SeqType: Site	1	-0.0012	-0.00005	-0.0046	1.000
SeqType: Date	4	0.0419	0.00182	0.0385	1.000
SeqType: Fraction	2	0.0197	0.00085	0.0361	1.000
Site: Fraction	**2**	**1.6076**	**0.06977**	**2.9535**	**0.001**
Date: Fraction	4	1.3671	0.05933	1.2559	0.104
SeqType: Site: Fraction	2	0.0016	0.00007	0.0030	1.000
SeqType: Date: Fraction	4	0.0274	0.00119	0.0252	1.000
Residual	40	10.886	0.47246		
Total	67	23.0409	1.00000		

phyla, with nanopore recovering an additional singleton Platyhelminthes MOTU. Proportions of phyla were found to be consistent across both sequencing datasets, and were largely dominated by Arthropoda (~53%), followed by Chordata (~20%) and then Cnidaria (~12%) (Fig. 3a and Table S1). The differences in MOTU richness were largely from these three dominant groups, with Illumina recovering 1.2 to 1.3× more MOTUs from each of these three phyla compared to nanopore (Table S1). The largest disparity was in Mollusca, for which Illumina recovered twice the number of MOTUs than nanopore. For the remaining six phyla (Echinodermata, Annelida, Porifera, Chaetognatha, Ctenophora, Bryozoa), Illumina and nanopore recovered approximately the same number of MOTUs. At the sample-level, the similar phylum proportions were also consistently observed, albeit with differences in species numbers (Fig. 3b). Only ZPT024 was markedly different in terms of community composition, and this was consistent with the stark dissimilarity

observed with nMDS plots (Fig. S1). When MOTUs were ranked by sequencing read counts between sequencing platforms, we found that Kendall's τ was significantly positive for 30 samples (min: 0.484; max: 0.986; *p*-value<<0.05; Table S2), which suggested a positive correlation in MOTU rank abundance between both sequencing platforms. Kendall's τ was also positive for ZPT024 (0.478), but the *p*-value was insignificant. This meant that if a MOTU was found to be abundant in one sample for one sequencing dataset, it would be highly likely to be abundant in the alternative platform as well. This assessment corroborated with the high pairwise Bray-Curtis similarity observed between samples across both sequencing platforms (Fig. S2), since the metric took into account read count data. This further demonstrated that nanopore metabarcoding could reliably and consistently recover abundant MOTUs; this was similarly corroborated by [28], even though our bioinformatic pipelines differed.

Nanopore metabarcode quality
We found that ~98% of the raw nanopore reads were erroneous when mapped to their respective Illumina samples, with a mean error rate of 4.20% (Fig. S3 and Table S3). This was consistent with the 4% error rate reported by Gunter et al. [47] for R10.3 flow cell chemistry. After consensus calling with amplicon_sorter however, and without further polishing with medaka, the percentage of consensus sequences per sample that remained erroneous dropped to 0–50.0% (average 24.0%), and error rates correspondingly decreased to 0–1.18% (average 0.40%) (Fig. S3 and Table S3).

Furthermore, for the 444 MOTUs shared between Illumina and nanopore, nanopore sequences from 406 MOTUs (91.4%) did not have indel errors when compared to the same MOTU's Illumina sequences (Table S4). For the remaining 38 MOTUs: 22 of them had nanopore sequences with 1 indel-error, five with 2 indel errors. The rest had three or more indel errors, but this only affected 11 MOTUs. Since our Illumina sequences were already confirmed to be translatable, it in turn confirmed that 91.4% of the nanopore consensus sequences were free of any frameshift errors, and thus translatable as well.

Nanopore sequencing with time
We subsampled the fast5 reads of each run for every hour for the first three hours, and every three hours thereafter to investigate the relationship of (i) number of raw reads, (ii) number of demultiplexed reads, and (iii) number of metazoan MOTUs obtained over time. Although the number of samples differed between runs, both runs showed a similar trend in that all three variables increased at a decreasing rate over time (Fig. 4). Raw
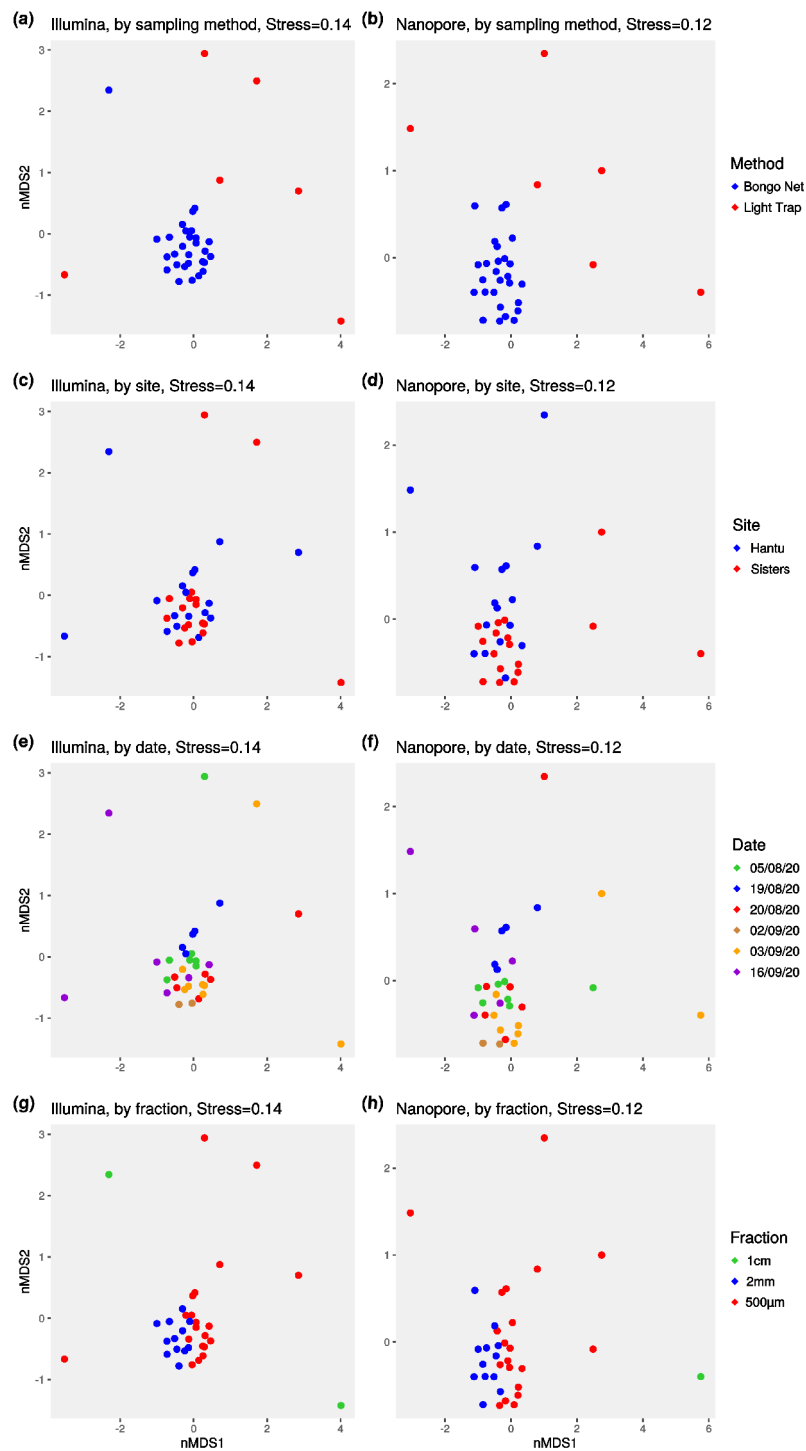
Fig. 2 Two-dimensional nonmetric multidimensional scaling (nMDS) plots based on normalised Bray-Curtis distances for Illumina (**a**, **c**, **e**, **g**), and nanopore (**b**, **d**, **f**, **h**); coloured by sampling method (**a** and **b**), sampling site (**c** and **d**), date (**e** and **f**), and size fraction (**g** and **h**). ZPT024 was removed from the nanopore dataset to better visualise the spread of points; it was similarly distinct from the remaining samples for both sequencing types

reads and demultiplexed reads both increased proportionately with respect to each other, with both variables only starting to plateau near the end of the respective runs. Conversely, metazoan MOTUs largely stabilised by the midway mark of each run, with RUN A and B obtaining 85% of the final MOTU count by the 12- and 15-hour mark respectively (Table S5). Beyond that, however, further increase in reads did not translate to substantial increase in metazoan MOTUs.

Table 2 Permutational multivariate analysis of variance (PERMANOVA) results comparing Bongo net communities for nanopore and Illumina datasets, using Bray-Curtis dissimilarity. Variables with significant *p*-values are highlighted in bold

Nanopore

	df	**Sum of squares**	**R²**	***F*-value**	***p*-value**
Site	1	0.5841	0.08491	3.8896	0.001
Date	4	1.8744	0.27248	3.1203	0.001
Fraction	1	1.2382	0.18000	8.2451	0.001
Site: Fraction	1	0.2479	0.03604	1.6510	0.090
Date: Fraction	4	0.6818	0.09911	1.1350	0.286
Residual	15	2.2526	0.32746		
Total	26	6.8791	1.00000		

Illumina

|	*df*	Sum of squares	R²	F-value	*p*-value
Site	1	0.5412	0.07609	3.3736	0.001
Date	4	2.0928	0.29420	3.2611	0.001
Fraction	1	1.1031	0.15508	6.8759	0.001
Site: Fraction	1	0.2723	0.03828	1.6973	0.082
Date: Fraction	1	0.6974	0.09803	1.0867	0.354
Residual	15	2.4065	0.33831		
Total	26	7.1133	1.00000		

Discussion

Using a set of zooplankton samples as our case study, we performed nanopore-based metabarcoding using ONT's MinION sequencer, and processed the reads with amplicon_sorter to show that nanopore metabarcodes are comparable to Illumina-based metabarcoding, and ready to be incorporated into more projects. Our study is also the first to emphasise that nanopore metabarcodes are nearly indel-free—an aspect that remains unexamined in past studies. We do note that nanopore metabarcoding is not perfect, and so the strengths and weaknesses of nanopore metabarcoding with amplicon_sorter are discussed below.

Nanopore metabarcodes are highly accurate and virtually indel-free

It is now possible to achieve highly accurate nanopore consensus metabarcodes with amplicon_sorter. In our case, nanopore consensus metabarcodes were observed to be ~99.6% accurate when benchmarked against their respective Illumina samples. We note this to be slightly better than the median 99.3% sequencing accuracy
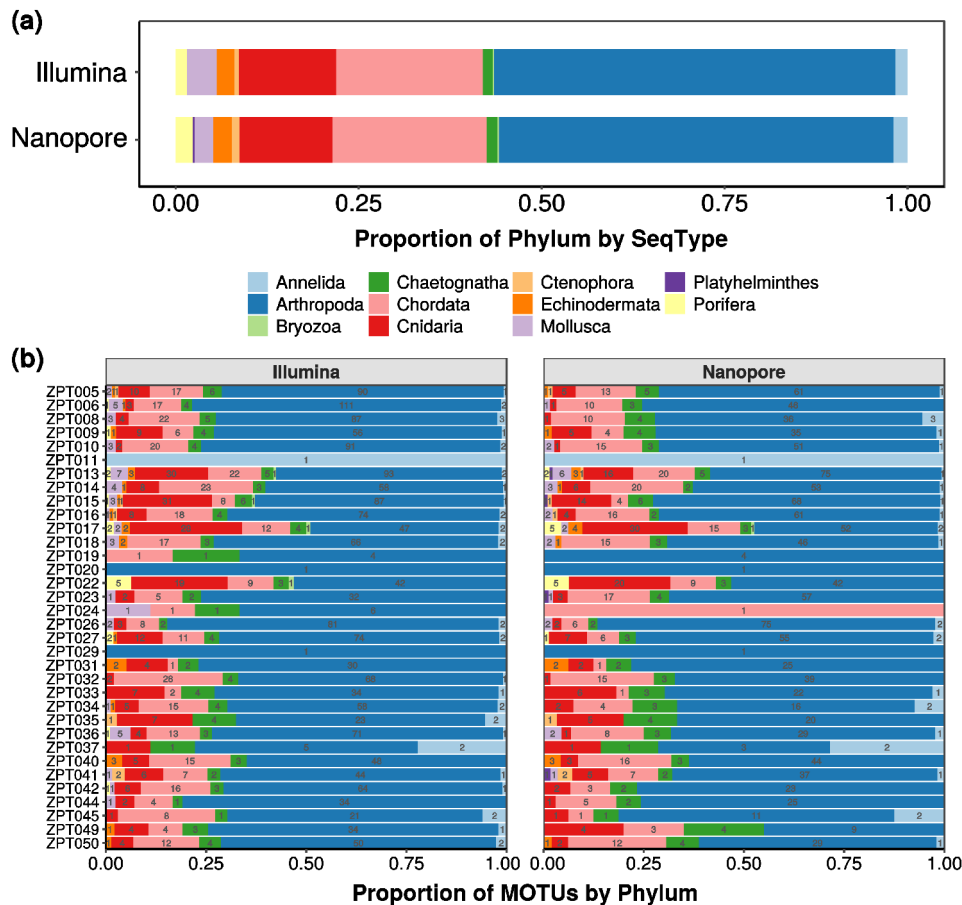
Fig. 3 Bar plots showing the relative proportions of molecular operational taxonomic units (MOTUs, grouped by phylum) by sequencing type (**a**), and by sample (**b**)
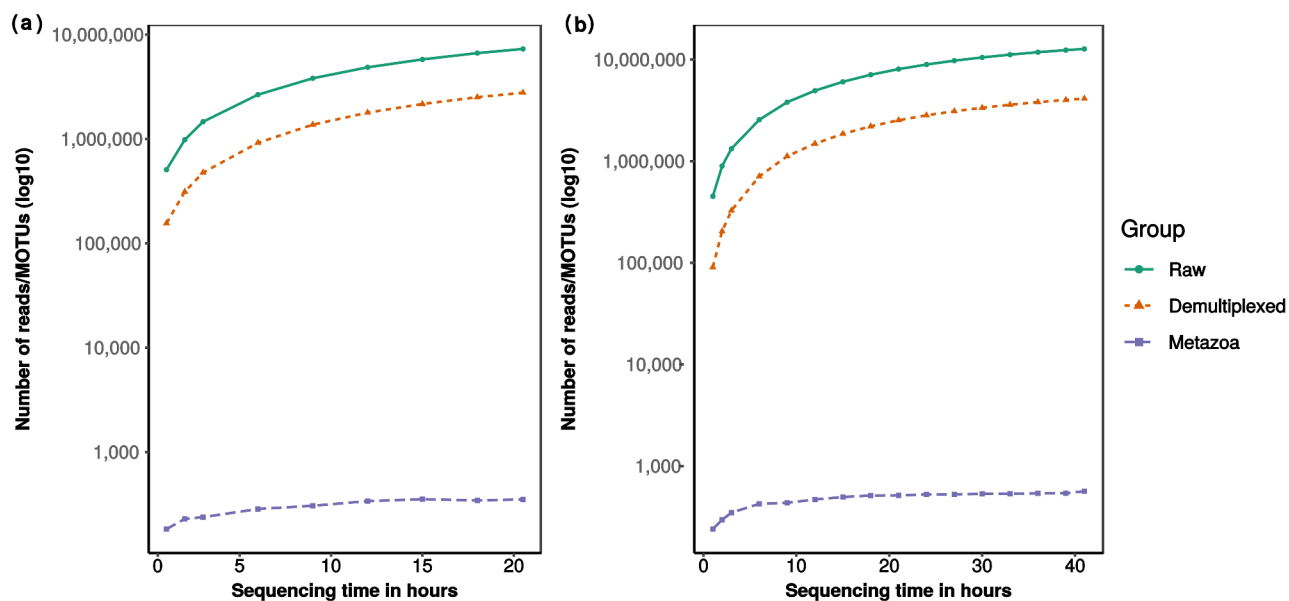
Fig. 4 Line graphs showing the change in number of raw reads (green), demultiplexed reads (orange) and metazoan MOTUs (purple) with sequencing run time, for RUN **A** (left) and RUN **B** (right)

observed by Baloğlu et al. [20], which could be due to our use of the R10.3 sequencing chemistry and SUP base calling model. Furthermore, amplicon_sorter generated consensus metabarcodes that did not require further polishing, mirroring an observation made by Srivathsan et al. [49], and more recently by Wick (https://rrwick. github.io/2023/12/18/ont-only-accuracy-update.html*)* with the most updated sequencing chemistry and base calling models. This is in contrast to prior nanopore metabarcoding pipelines that always included a polishing step, e.g., Egeter et al. [27] polished their sequences with RACON, while decona [45] incorporated medaka for polishing. We observed only a negligible 0.02% improvement in error rates for our nanopore metabarcodes after polishing, which corroborates Wick's findings that polishing is no longer needed (https://rrwick. github.io/2023/12/18/ont-only-accuracy-update.html*).* This is advantageous as it saves on time and computational resources, because each consensus sequence has to be polished individually when running medaka. For our dataset al.one, nearly 4,000 instances of medaka were performed, and this is unlikely to scale well computationally for more diverse, or larger-scale metabarcoding projects, where the number of consensus sequences obtained are expected to increase.

An added advantage was that almost all our unpolished nanopore metabarcodes were indel-free (91.4%) when compared to their Illumina counterparts, with nearly all of the 38 remaining nanopore sequences having only 1−2 indel errors. Existing nanopore metabarcoding benchmarking studies typically investigate sequencing accuracy [20], and unfortunately do not report gap errors, making

it difficult for a direct comparison with our findings. Nevertheless, our workflow presents an improvement over existing pipelines like decona or MSI, as initial tests with our same dataset suggested that polishing programs like RACON and medaka did not greatly improve error rates, and that most nanopore metabarcodes still contained indel-errors. Our validation that nanopore metabarcodes are almost always indel-free means that nanopore metabarcodes can now be subjected to translation checks without error, which would boost the quality of nanopore metabarcodes. Lastly, we were able to achieve clustering and error-correction with just amplicon_sorter alone, and with a single command, which simplifies the analysis workflow.

Lower MOTU richness with nanopore metabarcoding than Illumina

While we have demonstrated that nanopore metabarcoding generated metabarcodes with Illumina-like quality, we recognise that it yielded certain differences in other aspects when benchmarked against Illumina. The most notable difference was in MOTU richness, where we obtained 589 Illumina MOTUs, compared to 471 nanopore MOTUs, with 444 MOTUs shared across both platforms (72% congruence) (Fig. 1b, insert). This was corroborated by a significant difference from the paired Wilcoxon signed-rank test (Fig. 1d).

Based on our Kendall's τ analysis, MOTUs present in Illumina, but missing in nanopore, were MOTUs that generally had very low read depth. This means that MOTUs missed by nanopore sequencing were rarer in the community. The simplest explanation would be that

MOTU differences were a consequence of sequencing effort between platforms, or even stochasticity in the adapter ligation efficiency during respective Illumina and nanopore library preparation steps, but these are oftentimes difficult to account for. We also investigated two potential reasons relating to amplicon_sorter to assess if the MOTU differences could also be program-related.

The first reason was resolution limits of amplicon_sorter, presently at 95–96% [48]. This means that closely-related species, with less than 4% variance in the COI sequence, will be grouped together by amplicon_sorter, resulting in a lower number of MOTUs obtained. This was challenging to determine as our zooplankton samples were not mock communities, and we did not have prior knowledge of closely-related species groups that we could use to evaluate the resolution limits. We screened ZPT024 and ZPT034—samples that had the lowest Jaccard similarity coefficients between Illumina and nanopore. We first searched for a MOTU that was detected in both Illumina and nanopore for that sample, and then checked if there were any congenerics found in Illumina but not in nanopore (we assumed that congenerics had a higher likelihood of being closely-related compared to other taxonomic ranks). We then checked if the pairwise *p*-distance between these sequences differed by ≤4%, but since we did not encounter any such instance, we do not think that the resolution limit of amplicon_sorter was the main contributing factor for differences in MOTU richness for our study. We emphasise that future users pay special heed to this resolution limit when selecting metabarcoding loci. For instance, zooplankton metabarcoding studies have used hypervariable regions in nuclear 18 S rRNA [83–85], nuclear 28 S rRNA [86], and mitochondrial 16 S rRNA [87] in addition to COI [88–91]. The chosen loci must be divergent enough so that the species groups would not be over-collapsed by amplicon_sorter.

The last potential cause for difference in MOTU richness was based on the observation that since amplicon_sorter grouped only ~ 57% of the reads on average for consensus calling, we checked if the MOTUs unique to Illumina could be found in the unsorted nanopore reads. We mapped the ungrouped nanopore reads to the unique Illumina MOTUs with mapPacBio.sh (see Methods), and found that had amplicon_sorter incorporated these reads, 22 ZPT samples would have had a complete overlap with the MOTUs detected by Illumina sequencing. The remaining 10 samples would mostly still lack 1–2 MOTU(s), with only ZPT008 and ZPT049 missing four or five MOTUs respectively. We further found that the unsorted nanopore reads had a comparatively higher total error rate of ~ 4.52%, above the distance or length thresholds for forming and grouping clusters. This implied that bioinformatic processing of reads by amplicon_sorter was the more likely reason for the MOTU

difference. Further tests however, are needed to better optimise consensus calling settings with amplicon_sorter.

In any case, we note that the aforementioned limitations of amplicon_sorter will not pose a major issue to future metabarcoding projects, given that ONT is continuously updating its flow cell chemistry and basecalling algorithms. Its most recent pivot to R10.4.1 flow cell version and v14 kit chemistry (SQK-LSK114) offers Q20+raw read accuracy (i.e., 1 in 100 error rate). Potential implications would most certainly be higher-quality raw reads that allow for more precise formation and merging of species groups by amplicon_sorter, which in turn will likely improve the resolution limits of the algorithm. For instance, Ni et al. [92] and Sereika et al. [93] have reported ~ 99.1% modal raw read accuracy when using the latest R10.4 sequencing chemistry—a considerable improvement compared to the v9+R10.3 sequencing chemistry we used. In addition, with ONT's latest duplex basecalling capabilities, ~ 99% accurate, Q30+raw reads for metabarcoding are fast becoming a reality [18]. It is thus quite foreseeable that the limiting factors of amplicon_sorter will resolve as nanopore read quality improves with time.

Nanopore metabarcoding costs and turnaround times

Various studies have compared sequencing costs between nanopore and Illumina for metabarcoding, and it is generally agreed upon that nanopore metabarcoding with the MinION is generally cheaper than Illumina MiSeq (28,29). We reduced reagent costs further by adopting a single-PCR tagging strategy, where each of our PCR primers were tagged on 5'-end with 13-bp tags [49]. This enabled us to pool multiple PCR replicates into just two pools for nanopore library preparation without further need to barcode them. The only downside was that it required a separate software (e.g., ONTbarcoder) rather than Guppy for sample demultiplexing. However, the single PCR-tagging saved us processing time because the tagging occurred during thermocycling rather than as an additional step in the library preparation process (thermocycling runs for the same length of time regardless whether tagging is performed). The general utility of tagged-PCR primers also means that it can be used for other DNA sequencing projects [50, 64, 81], and even for Illumina sequencing (like in this study).

Another attractive property of nanopore sequencing is its ability to sequence in real-time. Users can terminate the run when their sequencing needs have been met, wash the flow cell and even recycle it for future use. We were thus interested to know if there was a "sweet-spot" for MOTU richness obtained in relation to sequencing run time for metabarcoding sequencing, based on the observation that up to 90% of DNA barcodes were obtained within the first few hours [49]. Our preliminary

examination from subsampling nanopore reads with time was that both runs reached ~85% of the final MOTU count in under 12 h and 15 h for RUNs A and B respectively (Fig. 4 and Table S5), and sequencing beyond that did not lead to a substantial increase in the number of metazoan MOTUs recovered. We recognise that the relationship between run time and MOTUs recovered is not immediately clear for nanopore metabarcoding (vis-à-vis DNA barcoding). Metabarcoding is likely to be more sensitive to factors such as the number of samples pooled into one flow cell, flow cell health (different flow cells may start with different number of pores available for sequencing) and even pore occupancy (percentage of pores actively sequencing). More tests on the number of metabarcoding samples that can be comfortably multiplexed onto a MinION flow cell without compromising recovered MOTU diversity are needed. What was clear however, was that turnaround times were much faster; it took us three days to complete both nanopore runs (we ran RUN A and B consecutively), in contrast to outsourcing Illumina MiSeq sequencing, which would take 2–4 weeks at the very least. Researchers have even taken advantage of this quicker turnaround time in time-sensitive situations such as disease surveillance [94]. Even for zooplankton biomonitoring, where sampling intervals can be as often as every two weeks [95], a nanopore-based metabarcoding approach would enable a quicker generation of results that make proposed routine biomonitoring strategies like Song et al. [96] more operationally feasible.

Nanopore metabarcoding for community characterisation
From an operational perspective, we have demonstrated that nanopore-based metabarcoding is viable when benchmarked against Illumina sequencing. Our nanopore metabarcodes were virtually Illumina-like, even with (soon-to-be-obsolete) v9 library preparation kits and R10.3 MinION flow cells. This is only going to improve moving forward, and it is time to relinquish the perception that nanopore sequencing produces highly erroneous reads. Even though there were differences between sequencing platforms, we ultimately found that the same ecological conclusions were obtained regardless—that our zooplankton communities were structured by date, site and fraction, and using a different sequencer was not a significant factor in explaining zooplankton community dissimilarities. Even the relative abundance of MOTUs was fairly consistent across sequencing platforms (88% congruence) and both sequencers successfully recovered 10 metazoan phyla. This also means that future users can employ nanopore sequencing for community metabarcoding with the confidence that their results will be consistent with Illumina, with the potential to leverage the cost-effectiveness, portability and

real-time advantages that nanopore sequencing brings. For example, some studies have already incorporated in-situ nanopore metabarcoding on board marine vessels [23, 26], and we believe more will follow suit in future, especially in the field of plankton monitoring. We did observe however, that amplicon_sorter was less likely to recover rarer MOTUs in the community compared to Illumina. Hence, users who wish to detect rarer species with degenerate primer sets will have to go with conventional Illumina sequencing in order to increase the chances of detection. We do believe this drawback can be soon addressed given that the latest and most accurate R10.4.1 sequencing chemistry is already available, and there are an increasing number of promising reports regarding its use [18, 60, 92, 93]. Further benchmarking studies will be needed to investigate how these improvements impact metabarcoding.

Conclusions
DNA metabarcoding is a powerful technique that can be harnessed to generate numerous sequence reads in parallel for multi-species identification and much more. Presently, DNA metabarcoding is conducted using second generation sequencing mainstays like Illumina, and less so on third-generation sequencers like ONT's MinION sequencer. We surmised that this was likely due to the notoriously high error rates of nanopore reads, as well as the general lack of specialised programs that can process such erroneous reads. Existing nanopore metabarcoding workflows either incorporate complicated and time-consuming laboratory steps, or require custom reference databases, or additional polishing steps, which perhaps disincentives the use of nanopore sequencing for metabarcoding. However, recent improvements in nanopore read accuracy in conjunction with new bioinformatic pipelines have led us to posit that nanopore sequencing can now produce highly-accurate metabarcoding results that are consistent with conventional Illumina sequencing, and without the need to polish the sequences unlike in the past. We demonstrated this by metabarcoding 34 bulk zooplankton communities on two R10.3 MinION flow cells, and processed the reads with amplicon_sorter. Our results showed that: [1] nanopore metabarcodes are nearly Illumina-like in sequencing accuracy (99.6%) and are almost always indel-free (91.4%); [2] relative abundance of MOTUs were congruent (88%) across both platforms, and nanopore recovered the abundant MOTUs just as well as Illumina but struggled to capture the rarer taxa; and that [3] ecological conclusions were consistent across sequencing platforms when metabarcoding zooplankton communities despite some differences in species richness recovered. Reports of the newly released R10.4.1 sequencing chemistry already indicate vast improvements in the quality of nanopore sequences. We

are confident that our results will inspire greater assurance in the utility of nanopore technology for more, and perhaps even larger-scale, metabarcoding-related projects in the near future.

Abbreviations

MOTU	Molecular operational taxonomic unit
NGS	Next-generation sequencing
NMDS	Nonmetric multidimensional scaling
NUMT	Nuclear mitochondrial DNA
ONT	Oxford Nanopore Technologies
PCR	Polymerase chain reaction
PERMANOVA	Permutational multivariate analysis of variance
SFB	Short fragment buffer
SUP	Super accurate
WoRMS	World Register of Marine Species

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10767-4.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Author contributions
JJMC and DH conceived the project idea. MADM and ZJ led sample collections, assisted by WLN. WLN processed the samples and performed the wet laboratory processes together with YCAI. JJMC prepared the nanopore sequencing libraries, analysed the data and drafted the manuscript, with input from DH and YCAI. YCAI compiled the information for verification of taxonomic identities and geographic ranges. All authors reviewed the manuscript and approved the final draft for submission.

Data availability
The Illumina sequence reads, nanopore base-called fast5 files, and nanopore fastq reads have been uploaded onto NCBI Sequence Read Archive under BioProject PRJNA991449. Sample metadata, demultiplexing information, MOTU table, and taxonomic identifications can be found in Supplementary File S1.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Biological Sciences, National University of Singapore, 16 Science Drive 4, Singapore 117558, Singapore
[2]School of Marine and Environmental Affairs, University of Washington, 3707 Brooklyn Ave NE, Seattle, Washington 98105, USA
[3]Lee Kong Chian Natural History Museum, National University of Singapore, 2 Conservatory Drive, Singapore 117377, Singapore
[4]Tropical Marine Science Institute, National University of Singapore, 18 Kent Ridge Road, Singapore 119227, Singapore
[5]Centre for Nature-based Climate Solutions, National University of Singapore, 6 Science Drive 2, Singapore 117546, Singapore

References
1. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol. 2012;21(8):2045–50.
2. Hebert PD, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc Biol Sci. 2003;270(1512):313–21.
3. Glenn TC. Field guide to next-generation DNA sequencers. Mol Ecol Resour. 2011;11(5):759–69.
4. Ip YCA, Chang JJM, Huang D. Advancing and integrating Biomonitoring 2.0 with new molecular tools for marine biodiversity and ecosystem assessments. In: Hawkins SJ, Russell BD, Todd PA, editors. Oceanography and Marine Biology: an Annual Review. CRC; 2023. pp. 293–325.
5. Mikheyev AS, Tin MMY. A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour. 2014;14(6):1097–102.
6. Menegon M, Cantaloni C, Rodriguez-Prieto A, Centomo C, Abdelfattah A, Rossato M, et al. On site DNA barcoding by nanopore sequencing. PLoS ONE. 2017;12(10):e0184741.
7. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 2018;36(4):338–45.
8. Quick J, Ashton P, Calus S, Chatt C, Gossain S, Hawker J, et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. Genome Biol. 2015;16(1):1–14.
9. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12(8):733–5.
10. Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. Oxford Nanopore sequencing, hybrid error correction, and *de novo* assembly of a eukaryotic genome. Genome Res. 2015;25(11):1750–6.
11. Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol. 2019;37(7):783–92.
12. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 2015;7:99.
13. Davidov K, Iankelevich-Kounio E, Yakovenko I, Koucherov Y, Rubin-Blum M, Oren M. Identification of plastic-associated species in the Mediterranean Sea using DNA metabarcoding with Nanopore MinION. Sci Rep. 2020;10(1):17533.
14. del Socorro Toxqui Rodríguez M, Naya-Català F, Sitjà-Bobadilla A, Carla Piazzon M, Pérez-Sánchez J. Fish microbiomics: strengths and limitations of MinION sequencing of gilthead sea bream (*Sparus aurata*) intestinal microbiota. Aquaculture. 2023;569:739388.
15. Benítez-Páez A, Portune KJ, Sanz Y. Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION™ portable nanopore sequencer. Gigascience. 2016;5:4.

16. Calus ST, Ijaz UZ, Pinto AJ. NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. Gigascience. 2018;7(12):giy140.

17. Zhang T, Li H, Ma S, Cao J, Liao H, Huang Q, et al. The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. Appl Environ Microbiol. 2023;89(10):e0060523.

18. Stoeck T, Katzenmeier SN, Breiner HW, Rubel V. Nanopore duplex sequencing as an alternative to Illumina MiSeq sequencing for eDNA-based biomonitoring of coastal aquaculture impacts. Metabarcoding Metagenom. 2024;8:e121817.

19. Krehenwinkel H, Pomeranz A, Henderson JB, Kennedy SR, Lim JY, Swamy V et al. Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. Gigascience. 2019;8(5):giz006.

20. Baloğlu B, Chen Z, Elbrecht V, Braukmann T, MacDonald S, Steinke D. A workflow for accurate metabarcoding using nanopore MinION sequencing. Methods Ecol Evol. 2021;12(5):794–804.

21. Srivathsan A, Loh RK, Ong EJ, Lee L, Ang Y, Kutty SN et al. Network analysis with either Illumina or MinION reveals that detecting vertebrate species requires metabarcoding of iDNA from a diverse fly community. Mol Ecol. 2023;32(23):6418-35.

22. Semmouri I, De Schamphelaere KAC, Willemse S, Vandegehuchte MB, Janssen CR, Asselman J. Metabarcoding reveals hidden species and improves identification of marine zooplankton communities in the North Sea. ICES J Mar Sci. 2021;78(9):3411–27.

23. Carradec Q, Poulain J, Boissin E, Hume BCC, Voolstra CR, Ziegler M, et al. A framework for in situ molecular characterization of coral holobionts using nanopore sequencing. Sci Rep. 2020;10(1):15893.

24. Conti A, Casagrande Pierantoni D, Robert V, Corte L, Cardinali G. MinION sequencing of yeast mock communities to assess the effect of databases and ITS-LSU markers on the reliability of metabarcoding analysis. Microbiol Spectr. 2023;11(1):e0105222.

25. Munian K, Ramli FF, Othman N, Mahyudin NAA, Sariyati NH, Abdullah-Fauzi NAF, et al. Environmental DNA metabarcoding of freshwater fish in Malaysian tropical rivers using short-read nanopore sequencing as a potential biomonitoring tool. Mol Ecol Resour. 2024;24(4):e13936.

26. Truelove NK, Andruszkiewicz EA, Block BA. A rapid environmental DNA method for detecting white sharks in the open ocean. Methods Ecol Evol. 2019;10(8):1128–35.

27. Egeter B, Veríssimo J, Lopes-Lima M, Chaves C, Pinto J, Riccardi N, et al. Speeding up the detection of invasive bivalve species using environmental DNA: a Nanopore and Illumina sequencing comparison. Mol Ecol Resour. 2022;22(6):2232–47.

28. van der Reis AL, Beckley LE, Olivar MP, Jeffs AG. Nanopore short-read sequencing: a quick, cost-effective and accurate method for DNA metabarcoding. Environ DNA. 2023;5(2):282–96.

29. Huggins LG, Colella V, Young ND, Traub RJ. Metabarcoding using nanopore long-read sequencing for the unbiased characterization of apicomplexan haemoparasites. Mol Ecol Resour. 2024;24(2):e13878.

30. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat Biotechnol. 2015;33(3):296–300.

31. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantification. 2015;3:1–8.

32. Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL, et al. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. Sci Rep. 2018;8(1):10950.

33. Buchner D, Macher TH, Leese F. APSCALE: advanced pipeline for simple yet comprehensive analyses of DNA metabarcoding data. Bioinformatics. 2022;38(20):4817–9.

34. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–3.

35. Mousavi-Derazmahalleh M, Stott A, Lines R, Peverley G, Nester G, Simpson T, et al. eDNAFlow, an automated, reproducible and scalable workflow for analysis of environmental DNA sequences exploiting Nextflow and Singularity. Mol Ecol Resour. 2021;21(5):1697–704.

36. Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. Obitools: a unix-inspired software package for DNA metabarcoding. Mol Ecol Resour. 2016;16(1):176–82.

37. Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, et al. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Res. 2019;47(18):e103.

38. Maestri S, Cosentino E, Paterno M, Freitag H, Garces JM, Marcolungo L et al. A rapid and accurate MinION-based workflow for tracking species biodiversity in the field. Genes. 2019;10(6):468.

39. Sahlin K, Lim MCW, Prost S. NGSpeciesID: DNA barcode and ampli-con consensus generation from long-read sequencing data. Ecol Evol. 2021;11(3):1392–8.

40. Srivathsan A, Baloğlu B, Wang W, Tan WX, Bertrand D, Ng AHQ et al. A MinION™-based pipeline for fast and cost-effective DNA barcoding. Mol Ecol Resour. 2018;18(5):1035–49.

41. Srivathsan A, Hartop E, Puniamoorthy J, Lee WT, Kutty SN, Kurina O, et al. Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. BMC Biol. 2019;17:96.

42. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:e2584.

43. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–2.

44. Voorhuijzen-Harink MM, Hagelaar R, van Dijk JP, Prins TW, Kok EJ, Staats M. Toward on-site food authentication using nanopore sequencing. Food Chem X. 2019;2:100035.

45. Doorenspleet K, Jansen L, Oosterbroek S, Kamermans P, Bos O, Wurz E et al. The long and the short of it: Nanopore based eDNA metabarcoding of marine vertebrates works; sensitivity and specificity depend on amplicon lengths [Internet]. bioRxiv. 2023. p. 2021.11.26.470087. https://www.biorxiv.org/content/biorxiv/early/2023/07/11/2021.11.26.470087

46. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, et al. Evaluation of Oxford Nanopore's MinION sequencing device for Microbial whole genome sequencing applications. Sci Rep. 2018;8(1):10931.

47. Gunter HM, Youlten SE, Madala BS, Reis ALM, Stevanovski I, Wong T, et al. Library adaptors with integrated reference controls improve the accuracy and reliability of nanopore sequencing. Nat Commun. 2022;13(1):6437.

48. Vierstraete AR, Braeckman BP, Amplicon_sorter:. A tool for reference-free amplicon sorting based on sequence similarity and for building consensus sequences. Ecol Evol. 2022;12(3):e8603.

49. Srivathsan A, Lee L, Katoh K, Hartop E, Kutty SN, Wong J, et al. ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. BMC Biol. 2021;19(1):217.

50. Ip YCA, Chang JJM, Oh RM, Quek ZBR, Chan YKS, Bauman AG, et al. Seq' and ARMS shall find: DNA (meta)barcoding of Autonomous reef monitoring structures across the tree of life uncovers hidden cryptobiome of tropical urban coral reefs. Mol Ecol. 2023;32(23):6223–42.

51. Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, et al. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front Zool. 2013;10(1):1–14.

52. Lobo J, Costa PM, Teixeira MAL, Ferreira MSG, Costa MH, Costa FO. Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. BMC Ecol. 2013;13:34.

53. Ip YCA, Tay YC, Gan SX, Ang HP, Tun K, Chou LM, et al. From marine park to future genomic observatory? Enhancing marine biodiversity assessments using a biocode approach. Biodivers Data J. 2019;7:e46833.

54. Castro LR, Meyer RS, Shapiro B, Shirazi S, Cutler S, Lagos AM, et al. Metabarcoding meiofauna biodiversity assessment in four beaches of Northern Colombia: effects of sampling protocols and primer choice. Hydrobiologia. 2021;848(15):3407–26.

55. Leite BR, Vieira PE, Troncoso JS, Costa FO. Comparing species detection success between molecular markers in DNA metabarcoding of coastal macroinvertebrates. Metabarcoding Metagenomics. 2021;5:e70063.

56. Clarke LJ, Beard JM, Swadling KM, Deagle BE. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecol Evol. 2017;7(3):873–83.

57. Geller J, Meyer C, Parker M, Hawk H. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Mol Ecol Resour. 2013;13(5):851–61.

58. Chang JJM, Ip YCA, Bauman AG, Huang D. MinION-in-ARMS: Nanopore sequencing to expedite barcoding of specimen-rich macrofaunal samples from Autonomous Reef Monitoring Structures. Frontiers in Marine Science. 2020;7:448.

59. Yeo D, Srivathsan A, Meier R. Longer is not always better: optimizing barcode length for large-scale species discovery and identification. Syst Biol. 2020;69(5):999–1015.

60. Srivathsan A, Feng V, Suárez D, Emerson B, Meier R. ONTbarcoder 2.0: rapid species discovery and identification with real-time barcoding facilitated by Oxford Nanopore R10.4. Cladistics. 2024;40(2):192–203.

61. Sze Y, Miranda LN, Sin TM, Huang D. Characterising planktonic dinoflagellate diversity in Singapore using DNA metabarcoding. Metabarcoding Metagenomics. 2018;2:e25136.

62. Ip YCA, Tay YC, Chang JJM, Ang HP, Tun KPP, Chou LM, et al. Seeking life in sedimented waters: environmental DNA from diverse habitat types reveals ecologically significant species in a tropical marine environment. Environ DNA. 2021;3(3):654–68.

63. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina paired-end reAd mergeR. Bioinformatics. 2014;30(5):614–20.

64. Chang JJM, Ip YCA, Ng CSL, Huang D. Takeaways from mobile DNA barcoding with BentoLab and MinION. Genes. 2020;11(10):1121.

65. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. Bioinformatics. 2018;34(15):2666–9. NanoPack: visualizing and processing long-read sequencing data.

66. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate *de novo* genome assembly from long uncorrected reads. Genome Res. 2017;27(5):737–46.

67. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.

68. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.

69. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al. BLAST: architecture and applications. BMC Bioinformatics. 2009;10:421.

70. Srivathsan A, Sha JCM, Vogler AP, Meier R. Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (Pygathrix nemaeus). Mol Ecol Resour. 2015;15(2):250–61.

71. Lim LJW, Loh JBY, Lim AJS, Tan BYX, Ip YCA, Neo ML, et al. Diversity and distribution of intertidal marine species in Singapore. Raffles Bull Zool. 2020;68:396–403.

72. Wells FE, Tan KS, Todd PA, Jaafar Z, Yeo DCJ. A low number of introduced marine species in the tropics: a case study from Singapore. Manage Biol Invasions. 2019;10(1):23–45.

73. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2023 [cited 2023 Apr 28]. https://www.R-project.org/

74. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-; 2016. p. 213.

75. Oksanen J, Simpson GL, Guillaume Blanchet F, Roeland K. vegan: Community Ecology Package [Internet]. 2022. https://cran.r-project.org/web/packages/vegan/vegan.pdf

76. Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of. Methods Ecol Evol. 2016;7(12):1451–6.

77. Laporte M, Reny-Nolin E, Chouinard V, Hernandez C, Normandeau E, Bougas B, et al. Proper environmental DNA metabarcoding data transformation reveals temporal stability of fish communities in a dendritic river system. Environ DNA. 2021;3(5):1007–22.

78. Kolde R, pheatmap. Pretty Heatmaps. R package version 1.0. 12. 2019.

79. Kendall MG. A new measure of rank correlation. Biometrika. 1938;30(1/2):81–93.

80. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5(2):R12.

81. Chang JJM, Ip YCA, Cheng L, Kunning I, Mana RR, Wainwright BJ, et al. High-throughput sequencing for life-history sorting and for bridging reference sequences in Marine Gerromorpha (Insecta: Heteroptera). Insect Syst Divers. 2022;6(1):1.

82. Chan WWR, Chang JJM, Tan CZ, Ng JX, Ng MHC, Jaafar Z, Huang D. Eyeing DNA barcoding for species identification of fish larvae. J. Fish Biol. https://doi.org/10.1111/jfb.15920

83. Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLoS ONE. 2009;4(7):e6372.

84. Pearman JK, Irigoien X. Assessment of zooplankton community composition along a depth profile in the central Red Sea. PLoS ONE. 2015;10(7):e0133487.

85. Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A. Next generation sequencing reveals the hidden diversity of zooplankton assemblages. PLoS ONE. 2013;8(11):e81327.

86. Hirai J, Shimode S, Tsuda A. Evaluation of ITS2-28S as a molecular marker for identification of calanoid copepods in the subtropical western North Pacific. J Plankton Res. 2013;35(3):644–56.

87. Goetze E. Species discovery in marine planktonic invertebrates through global molecular screening. Mol Ecol. 2010;19(5):952–67.

88. Machida RJ, Hashiguchi Y, Nishida M, Nishida S. Zooplankton diversity analysis through single-gene sequencing of a community sample. BMC Genomics. 2009;10:438.

89. Zaiko A, Samuiloviene A, Ardura A, Garcia-Vazquez E. Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. Mar Pollut Bull. 2015;100(1):53–9.

90. Bourlat SJ, Borja A, Gilbert J, Taylor MI, Davies N, Weisberg SB, et al. Genomics in marine monitoring: new opportunities for assessing marine health status. Mar Pollut Bull. 2013;74(1):19–31.

91. Schroeder A, Stanković D, Pallavicini A, Gionechetti F, Pansera M, Camatti E. DNA metabarcoding and morphological analysis - Assessment of Zooplankton biodiversity in transitional waters. Mar Environ Res. 2020;160:104946.

92. Ni Y, Liu X, Simeneh ZM, Yang M, Li R. Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. Comput Struct Biotechnol J. 2023;21:2352–64.

93. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. Nat Methods. 2022;19(7):823–6.

94. Goenka SD, Gorzynski JE, Shafin K, Fisk DG, Pesout T, Jensen TD, et al. Accelerated identification of disease-causing variants with ultra-rapid nanopore genome sequencing. Nat Biotechnol. 2022;40(7):1035–41.

95. Mackas DL, Beaugrand G. Comparisons of zooplankton time series. J Mar Syst. 2010;79(3):286–304.

96. Song CU, Choi H, Jeon MS, Kim EJ, Jeong HG, Kim S, et al. Zooplankton diversity monitoring strategy for the urban coastal region using metabarcoding analysis. Sci Rep. 2021;11(1):24339.

Publisher's note