## Research Article

**Open Access**

# SoEM: a novel PCR-free biodiversity assessment method based on small-organelles enriched metagenomics

## Jihoon Jo[1,2], Hyun-Gwan Lee[2,3], Kwang Young Kim[2,3] and Chungoo Park[1,2,*]

[1]*School of Biological Sciences and Technology, College of Natural Sciences, Chonnam National University, Gwangju 61186, Korea*
[2]*Marine Ecosystem Disturbing and Harmful Organisms (MEDHO) Research Center, Gwangju 61186, Korea*
[3]*Department of Oceanography, College of Natural Sciences, Chonnam National University, Gwangju 61186, Korea*

DNA metabarcoding is currently used for large-scale taxonomic identification to understand the community composition in various marine ecosystems. However, before being widely used in this emerging field, this experimental and analytic approach still has several technical challenges to overcome, such as polymerase chain reaction (PCR) bias, and lack of well-established metabarcoding markers, a task which is difficult but not impossible to achieve. In this study, we present an adapted PCR-free small-organelles enriched metagenomics (SoEM) method for marine biodiversity assessment. To avoid PCR bias and random artefacts, we extracted target DNA sequences without PCR amplification from marine environmental samples enriched with small organelles including mitochondria and plastids because their genome sequences provide a valuable source of molecular markers for phylogenetic analysis. To experimentally enrich small organelles, we performed subcellular fractionation using modified differential centrifugation for marine environmental DNA samples. To validate our SoEM method, two marine environmental samples from the coastal waters were tested the taxonomic capturing capacity against that of traditional DNA metabarcoding method. Results showed that, regardless of taxonomic levels, at least 3-fold greater numbers of taxa were identified in our SoEM method, compared to those identified by the conventional multi-locus DNA metabarcoding method. The SoEM method is thus effective and accurate for identifying taxonomic diversity and presents a useful alternative approach for evaluating biodiversity in the marine environment.

## INTRODUCTION

Biodiversity has been considered as the integration of biological variability across all scales, from genetic levels, through species and ecosystems, to landscapes (Walker 1992). Raising awareness and maintaining biodiversity is an important goal of conservation biology. However, a fundamental challenge to our understanding of biodiver-

sity is identifying the distribution and abundance of species across space and time (Losos 2010, Lee et al. 2017). Over the past few decades, identification and mapping of biodiversity based on morphological and microscopic approaches has been developed and is being widely used in a variety of studies (Roy and Foote 1997, McManus and Katz 2009). Expert taxonomic knowledge and expensive infrastructure are required to illustrate as many morphological details as possible in a high-throughput manner (Appeltans et al. 2012). However, many morphological characters that define species cannot be clearly distinguished using light or electron microscopy, and an unignorably large number of morphological patterns tend to vary substantially within populations (Scotland et al. 2003). Even if all observable characters for taxonomic identification are available, many ecological assessments by trained taxonomists are misleading due to misidentifications caused by the use of different criteria for species delimitation or different interpretations (Hebert et al. 2004, Fontaneto et al. 2009, Leasi and Todaro 2009).

DNA barcoding, the linkage of a specimen to a unique DNA sequence, has been applied to the taxonomic identification of species (Hebert et al. 2003*b*, Kress and Erickson 2008, Zhou et al. 2009). The DNA barcode consists of an easily sequenced marker fragment that can distinguish a species based on the principle that genetic variation between species exceeds the variation within them. For example, cytochrome oxidase subunit I (COI) (Hebert et al. 2003*b*), *tufA* (Kress et al. 2015), and internal transcribed spacer (ITS) (Eberhardt 2012) have been widely utilized for identifying animals, green algae, and fungi, respectively. While DNA barcodes link an individual to a taxon with similar genetic sequences, DNA metabarcoding or environmental DNA (eDNA) barcoding is an emerging approach to investigate the ecological community structure of mixed samples from a target environment. As next-generation sequencing (NGS) technologies have been developed and successfully exploited to obtain a vast amount of genome sequences at a lower cost (Schuster 2007, Ansorge 2009), DNA metabarcoding relying on the high-throughput sequencing approach (Taberlet et al. 2012) has been popularly utilized in ecological studies. In addition, because NGS-based DNA metabarcoding is undertaken using automated and standardized data processing procedures (e.g., MEGAN4, Mothur, QIIME, and UPARSE pipeline) (Schloss et al. 2009, Caporaso et al. 2010, Huson et al. 2011, Edgar 2013) with well-established and curated databases (e.g., NCBI taxonomy, Federhen 2012; IBOL, http://www.ibolproject.org/; CBOL, Pawlowski et al. 2012; and SILVA, Quast

et al. 2013), datasets even from different studies and laboratories can be comparable and will thus facilitate biomonitoring studies with large spatiotemporal resolution (Hajibabaei et al. 2011).

Despite the potential power of DNA metabarcoding for large-scale species identification, several limitations have been pointed out and should be considered (Taberlet et al. 2012). In particular, the main reason of these limitations is a step involving DNA amplification by polymerase chain reaction (PCR). First, the PCR method can introduce errors by degradation of the template DNA or during DNA amplification and sequencing (Cha and Thilly 1993). Second, standardized DNA barcoding markers (e.g., COI, *tufA*, or ITS) are not necessarily suitable as DNA metabarcoding markers (Elbrecht and Leese 2015). Because the target bulk environmental samples contain organisms from a wide range of lineages, a multi-marker metabarcoding approach is proposed to represent a more accurate method (Drummond et al. 2015). However, it is difficult to determine suitable metabarcoding primer sets for some taxonomic groups such as nematodes that have very divergent mitochondrial genomes (mitogenomes) and ribosomal sequences (Porazinska et al. 2009, Bik et al. 2013). Third, because many variable species have to be amplified in the same PCR experiment, the primers used for amplification must be highly versatile (Elbrecht and Leese 2015). This means that many different target molecules should be amplified with the same efficiency, without missing species containing target sequences that do match well with the primers. However, standardized DNA barcoding primers do not provide equally good matches to all target sequences derived from a bulk environmental sample and thus introduce PCR amplification bias resulting in taxonomic biases (Hebert et al. 2003*a*, Bellemain et al. 2010, Taberlet et al. 2012, Pawluczyk et al. 2015, Krehenwinkel et al. 2017).

There have been several attempts (e.g., metagenome skimming (Dodsworth 2015, Linard et al. 2015, Greshake et al. 2016), mitochondrial metagenomics (Tang et al. 2014, Crampton-Platt et al. 2015, 2016), a mitochondrial capture microarray (Liu et al. 2016), and mitochondrial enrichment by differential centrifugation (Zhou et al. 2013) to overcome these limitations of using PCR-dependent techniques in ecological research. For example, metagenome skimming, first coined by Straub et al. (2012), is the low-coverage shotgun metagenomics sequencing of bulk DNA from environmental samples. *De novo* assembly analyses are performed through low-coverage genomic sequencing to provide a high-copy of metabarcoding markers that have been used as phylo-

genetic markers (e.g., mitochondrial DNA, chloroplast DNA, and rRNA genes) (Andújar et al. 2015, Greshake et al. 2016, Hillmann et al. 2018). The resulting assemblies are assigned for taxonomic identification to any of the genomes in the mixtures. However, details of the recovery of metabarcoding markers from low-coverage genomic sequencing and the accuracy of taxonomic identification from genomic sequence assembly of species mixtures is still vague. Next, the mitochondrial metagenomics approach attempts to capture DNA sequences, especially mitogenomes for species-level diagnosis. Using the *de novo* genomic assembly method and high-throughput sequencing without DNA enrichment or amplification, the obtained mitogenomes are used for taxonomic identification by comparing sequences in the NCBI database (Tang et al. 2014). However, this method is expensive and a burden on computational resources. Recently, Liu et al. (2016) presented a novel PCR-free mitogenome enrichment pipeline for biodiversity analyses. Using a mitochondrial capture microarray designed with the mitogenome sequences from the 1,000 insect transcriptome evolution project (1KITE, http://www.1kite.org), they demonstrated higher mitochondrial capture enrichment efficiency compared to the PCR-based identification method. Such hybridization-based approaches are high throughput and relatively inexpensive but have some limitations including reliance on the existing knowledge of mitogenome sequences and cross-hybridization problems. Zhou et al. (2013) adopted a differential centrifugation method for mitochondrial enrichment to investigate the biodiversity of Arthropoda. Although this approach proved efficient for fully controlled environmental samples, it has not been applied to actual environmental samples to our knowledge.

In this study, we present and evaluate the performance of an adapted PCR-free metagenomics method (called small-organelles enriched metagenomics [SoEM]) for biodiversity assessment in actual environmental samples. Without PCR amplification, subcellular fractionation by modified differential centrifugation for complex environmental samples was performed to enrich small organelles including mitochondria and plastids. Using two real marine environmental samples, we tested the taxonomic capture capacity of our SoEM method against the conventional multi-locus DNA metabarcoding method and observed that our SoEM method can detect a broader range of biodiversity compared to the most widely used DNA metabarcoding method.

## MATERIALS AND METHODS

### Sample collection

To collect marine micro-eukaryotic environmental samples, coastal seawater from two geologically separated sampling sites in South Korea was pumped using a submersible-pump (with capacity of 69 L min$^{-1}$). These sites are Yamido, Gunsan-si (35°851′ N, 126°501′ E; site #1) and Baealdo, Gwangyang-si (34°958′ N, 127°763′ E; site #2) (Fig. 1A). The seawater pre-filtered through a 200-μm mesh was concentrated using a homemade-conical shaped filtration unit with a 1-μm mesh filter attached to a 100 mL bottle at the tip. Thus, the environmental samples whose were captured species by this filtration unit included ranging from 1 to 200 μm in size. The concentrated seawater containing the environmental samples was transferred from filtration unit to 50 mL conical tubes, and was finally centrifuged at 2,500 ×g for 20 min to remove seawater (Fig. 1B). About 100 L of seawater was concentrated to one pellet of environmental samples. Above procedures were promptly performed on the sampling site. The pelleted samples were frozen in liquid nitrogen, delivered to the laboratory, and stored at -80°C until DNA extraction.

### Small organelles DNA enrichment and extraction

To enrich small organelles including mitochondria and plastids (Fig. 2), we slightly modified the differential centrifugation method described in a previous study (Zhou et al. 2013). About 1 g of frozen filtered sample (corresponding to about 100 L of sea-water) was resuspended in 20 mL of ice-cold homogenizing buffer (250 mM sucrose, 30 mM Tris-HCl, 10 mM EDTA, and pH 7.5). The suspension was homogenized using the IKA-T10 homogenizer (IKA, Staufen, Germany) at the maximum RPM for five cycles of 15 s on and 15 s off on ice. To eliminate the nuclei, cellular debris, and sedimentary pollutants, the homogenate was centrifuged at 1,300 ×g for 10 min, at 4°C. Supernatants enriched in small organelles were transferred to a new-tube and recentrifuged at 4°C at 17,000 ×g for 30 min. Finally, DNA was extracted from the collected pellet using the Qiagen DNeasy Blood/Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol, and the purified DNA was stored at -80°C until needed.
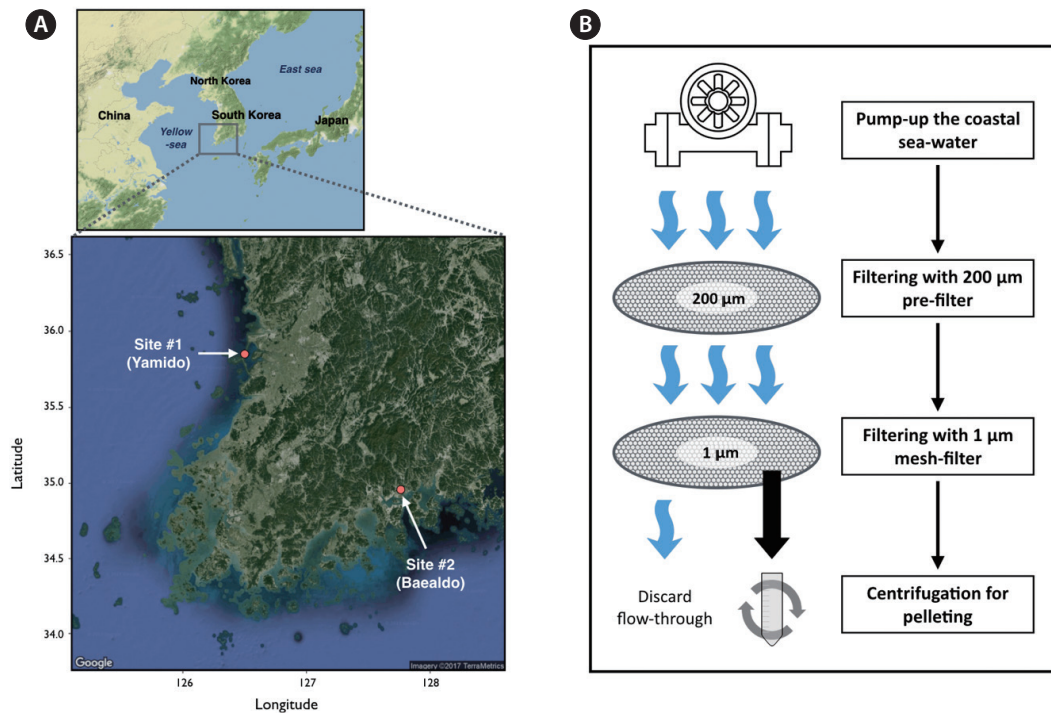
**Fig. 1.** Geological location of sampling sites and schematic workflow of the sampling procedure. (A) Geological location of two sampling sites in South Korea. Site #1 was located at Yamido, Gunsan-si (35°851′N, 126°501′E). Site #2 was located at Baealdo, Gwangyang-si (34°958′N, 127°763′E). (B) Schematic workflow of the sampling procedure. First, coastal sea-water was collected using a submersible-pump. The water intake was pre-filtered through a 200-μm filter and further filtration was performed with a 1-μm mesh filter. The collected environmental samples ranging from 1 to 200 μm were centrifuged at 2,500 ×g for 20 min.

## eDNA extraction

The eDNA for metabarcoding analysis was extracted using the CTAB protocol (Zhou et al. 1996) with a phenol-chloroform extraction method (Fig. 2). In short, about 1 g of frozen environmental sample was resuspended in 10 mL of 2% CTAB lysis buffer (2% cetyltrimethylammonium bromide, 100 mM Tris-HCl [pH 8.0], 20 mM EDTA, 1.4 M NaCl) and homogenized with the IKA-T10 homogenizer at the maximum RPM for five cycles of 15 s on and 15 s off on ice. The homogenate was treated with 0.5 mL of 20 mg mL$^{-1}$ proteinase K (Ambion; Thermo Fisher Scientific, Inc., Austin, TX, USA) at 60°C for 1 h, followed by phenol-chloroform extraction and ethanol precipitation of DNA. Finally, precipitated DNA was washed twice with 70% ethanol and was dissolved in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The eDNA was stored at -80°C until needed.

**Table 1.** Primary barcodes and primer sequences used for the metabarcoding analyses

| Clade | Barcode region | Primer name | Sequence information (5′ to 3′) |
|---|---|---|---|
| Animals | *COI* | LCO1490 | GGT CAA ATC ATA AAG ATA TTG G |
| | | HCO2198 | TAA ACT TCA GGG TGA CCA AAA AAT CA |
| Algae | *COI-5P* | GWSFn | TCA ACA AAY CAY AAA GAT ATY GG |
| | | GWSRx | ACT TCT GGR TGI CCR AAR AAY CA |
| Green algae | *tufA* | tufGF4 | GGN GCN GCN CAA ATG GAY GG |
| | | tufAR | CCT TCN CGA ATM GCR AAW CGC |
| Land plants | *rbcL* | CfD | CCR TTY ATG CGT TGG AGA GA |
| | | DPrbcL7 | AAR CAA CCT TGT GTA AGT CT |

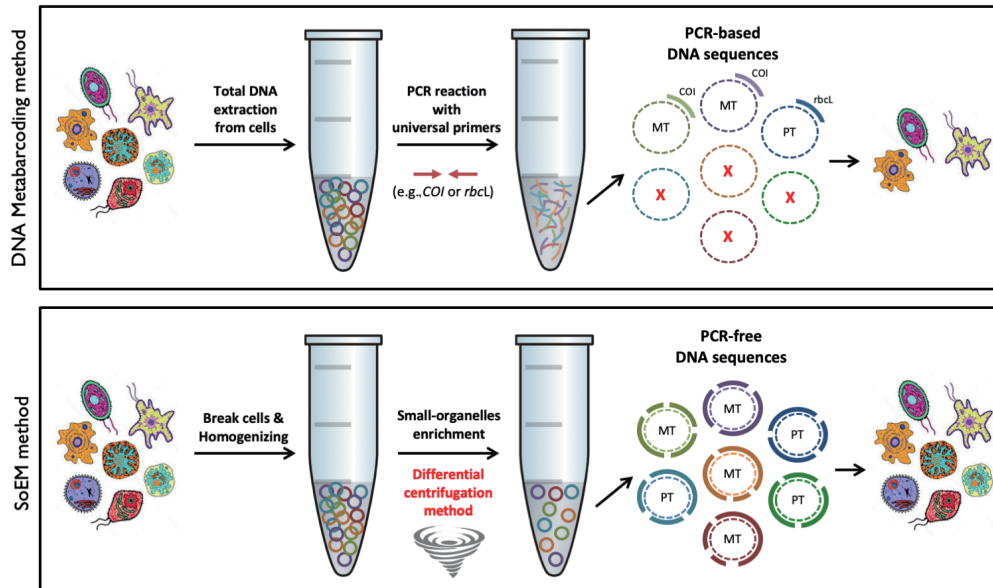Data from Saunders and McDevit (2012).

**Fig. 2.** Comparison of the experimental procedure of DNA metabarcoding and the small-organelles enriched metagenomics (SoEM) method. The upper panel represents the schematic workflow of the conventional DNA metabarcoding method. This method utilizes total genomic DNA extracted from environmental samples. Specific sequences are amplified by polymerase chain reaction (PCR) with universal primers of target barcoding markers. The resultant amplicons subjected to next-generation sequencing are mapped to databases for taxonomical assignment. The lower panel represents the schematic workflow of the SoEM method. Subcellular organelle enriched DNAs are extracted by differential centrifugation with a sucrose-containing buffer system using the weight-difference of subcellular organelles. The resultant sequences generated by next-generation sequencing are not biased to some barcoding markers but are subcellular enriched sequences with random DNA shearing. COI, cytochrome oxidase subunit I; MT, mitochondria; PT, plastid.

## DNA metabarcoding amplification

Because there is no generally accepted single universal DNA barcode for all organisms, we used four major barcode markers (*COI* for animal, *COI-5P* for algae, *tufA* for green algae, and *rbcL* for land plants) for DNA metabarcoding in this study (Table 1) (Saunders and McDevit 2012, Kress et al. 2015). The primer-sets were ligated with a pre-adaptor and a primer sequence for amplicon sequencing. The target marker fragments belonging to the barcoding loci were amplified in a 25-μL reaction containing 12.5 μL of KAPA HiFi Hotstart ReadyMix (KAPA Biosystems, Woburn, MA, USA), 5 μL of amplicon primer (1 μM), and 2.5 μL of diluted DNA template (about 5 ng μL$^{-1}$). The cycling conditions of PCR amplification were one cycle of 94°C for 3 min; 25 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s; and a final extension of 72°C for 5 min. The PCR products were purified with AMPure XP beads (Beckman-Coulter, Indianapolis, IN, USA). Index PCR was then performed for the purified PCR products in a 50-μL reaction containing 25 μL of KAPA HiFi Hotstart ReadyMix, 5 μL of NexteraXT Index, 10 μL of sterile deionized water, and 5 μL of PCR product from the pre-

vious amplification. Index PCR conditions included one cycle of 95°C for 3 min; 8 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s; and a final extension of 72°C for 5 min. The products were purified with AMPure XP beads, and the final purified products were stored at -20°C until required for library preparation. Additional details and primer sequences are summarized in Table 1. The PCR analysis for validation of the enrichment of small-organelles was performed with same condition described above.

## Library preparation and sequencing

To construct the sequencing libraries of SoEM assay using the TruSeq Nano DNA kit (Illumina, San Diego, CA, USA), initially about 200 ng of DNA was randomly fragmented for a 550 bp insertion size using a LE220 Focused-ultrasonicator (Covaris, Woburn, MA, USA). The fragmented DNA was blunt-ended and phosphorylated following end repair. The appropriate library size was selected using different ratios of sample purification beads. A single 'A' base was ligated to the 3′ end, and Illumina adapters were then ligated to the fragments. The ligated

fragments were PCR-amplified and cleaned-up. To verify the size of PCR-enriched fragments of both library types (amplicon and SoEM), the template size distribution was qualified using the Agilent Technologies 2100 Bioanalyzer and 2200 Tapestation (Agilent Technologies, Palo Alto, CA, USA). The final purified products were quantified by qPCR using KAPA Library Quantification kits (KAPA Biosystems) and the LabChip GX HT DNA High Sensitivity Kit (PerkinElmer, Waltham, MA, USA) for Illumina Sequencing platforms following the qPCR Quantification Protocol Guide. Finally, the qualified and quantified library was sequenced with paired-ends (2 × 301 bp) using the Illumina Miseq platform (Illumina). Purification of the PCR product and sequencing reactions were performed by Macrogen Inc. (Seoul, Korea). All the raw sequences obtained in this study were deposited to the NCBI SRA database under accession numbers ranging from SRR8185552 to SRR8185561, and to NCBI Bioproject under the accession number PRJNA505143.

## Bioinformatics analysis

To acquire high-quality sequence reads, all raw sequence data from both SoEM and DNA metabarcoding assays were preprocessed using Trimmomatic (v0.35) with a mean Phred score of 15 (Bolger et al. 2014). For preprocessed sequences from the SoEM assay, we applied two approaches to maximize the length of the amplified target sequences. First, the paired reads (i.e., forward and reverse sequences) were stitched together using PEAR (*P*aired-*E*nd re*A*d merge*R*) (v0.9.10) (Zhang et al. 2014) with minimum overlap size '-v 10' and minimum assembly length '-n 400' options. Second, three *de novo* metagenomic assemblers (SOAPdenovo2, IDBA-UD, and OMEGA) (Luo et al. 2012, Peng et al. 2012, Haider et al. 2014) were used to generate different sets

of contigs. The SOAPdenovo2 assembler was run under '–k 45, -D, -d, -R, -M 3' options, and screening of the '–K' option range from 71 to 91. IDBA-UD and Omega assemblers were run using default parameters and corrections of maximum read-length ('kMaxShortSeqeunce = 565') and minimum overlap length ('-l 180'), respectively. For the preprocessed sequences from the DNA metabarcoding assay, we removed the DNA barcoding primer set using Cutadapt v1.16 (Martin 2011) and the forward and reverse sequences from the final clean reads were concatenated using an in-house python script.

Next, to define an operational taxonomic unit (OTU), for merged sequences from the SoEM assay, a BLAST search was performed using each query sequence against the four DNA barcode reference databases (NT, non-redundant nucleotide database; PM, primary-marker sequences from GenBank; MT, mitochondria whole genome database; and PT, plastid whole-genome database) collected and built (Table 2), with an E-value < 1e-10 and aligned length ≥ 400 bp. For the BLAST result of each preprocessed sequence from the DNA metabarcoding assay, the following four steps were performed to eliminate chimeric concatenated reads: (1) accept only sequences belonging to the same target from forward and reverse reads; (2) accept only read pairs having the length of gaps and mismatches ≤5; (3) select one with the longest alignment length if there are many outputs passing the previous two steps; and (4) filtering out the sequences with an abnormal length of the assigned barcode region (i.e., <400 bp or >1,000 bp) because the lengths of the amplified target regions used in this study were estimated to be from 650 to 850 bp (Saunders and McDevit 2012, Kress et al. 2015). The remaining reads were then processed for final taxonomic identification.

To perform taxonomic assignment, we added a taxonomic ID to each OTU using the NCBI file gi_taxid_nucl.

**Table 2.** Four reference databases used in this study

| Database | Abbreviation | Download date (size) | Description |
|---|---|---|---|
| Non-redundant nucleotide database | NT | Oct 10, 2016 (128.3 Gb) | Non-redundant nucleotide database in the 'NCBI nucleotide' database (https://www.ncbi.nlm.nih.gov/nucleotide/) |
| Primary-marker database | PM | Nov 11, 2016 (2.6 Gb) | Manually downloaded sequences from the NCBI nucleotide database through key word search using '*COI*', '*COI-5P*', '*rbcL*', or '*tufA*' |
| Mitochondria whole-genome database | MT | Nov 11, 2016 (181 Mb) | All sequences of mitochondrial genomes registered in the NCBI organelle genome resource (https://www.ncbi.nlm.nih.gov/genome/organelle/) |
| Plastid whole-genome database | PT | Dec 1, 2016 (214 Mb) | All sequences of plastid genomes registered in the NCBI organelle genome resources (https://www.ncbi.nlm.nih.gov/genome/organelle/) |
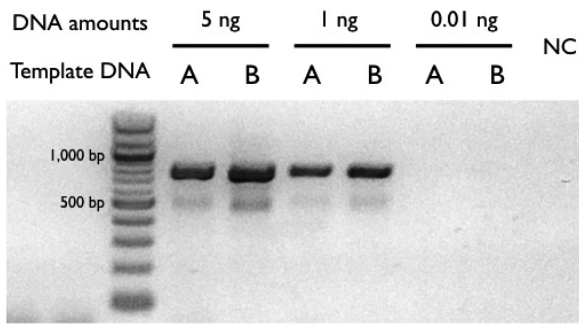
**Fig. 3.** Conventional polymerase chain reaction (PCR) products of the mitochondrial marker cytochrome oxidase subunit I. Two environmental DNA samples were investigated. (A) A PCR-based DNA sample used for the DNA metabarcoding method, and (B) a PCR-free DNA (namely, small-organelles enriched DNA) sample used in the small-organelles enriched metagenomics method. The same result was observed with both 5 ng and 1 ng of template DNA. However, PCR products of 0.01 ng template DNA were not amplified. NC, negative control.

dmp (ftp://ftp.ncbi.nih.gov/pub/taxonomy/gi_taxid_nucl.dmp.gz) and an in-house python script. Verification of species names, classification of taxonomic hierarchies, and conversion of scientific to common names was carried out using the 'taxize' package in the statistical program R (Chamberlain and Szöcs 2013). Unavailable taxonomic information at each level was represented as 'N/A' in further steps. Furthermore, in this study, we excluded all OTUs assigned to the super-kingdoms 'Bacteria,' 'Virus,' 'Archaea,' and 'Unclassified' because we assessed micro-eukaryotic biodiversity.

## RESULTS AND DISCUSSION

### Sampling and high throughput sequencing

To assess the newly established SoEM method, marine environmental samples from two geologically distinct sites were collected and prefiltered through 200 and 1 μm meshes (Fig. 1). We enriched the small organelles in the environmental samples using the SoEM method (see

Materials and Methods for details) (Fig. 2). To validate whether the small organelles were experimentally enriched, we performed PCR amplification of the *COI* mitochondrial DNA gene (Andújar et al. 2018) in samples before or after differential centrifugation. Our PCR results indicate that the SoEM method can be used effectively to enrich small organelles in environmental samples (Fig. 3). We then performed high-throughput sequencing using an Illumina MiSeq (v2, 301-cycle). After quality filtering, 35.6 and 35.1 million clean reads were generated for site #1 and site #2, respectively (Table 3). The cleaned paired-end reads were merged, and the resultant longer reads (6.92 and 6.65 million for site #1 and site #2, respectively) were used (Table 3). Because many sequence reads were excluded from the preceding analysis (the merging rates of between 37.90% and 38.86%) (Table 3) and to further boost the accuracy of taxonomical classification, we again performed a metagenome assembly using the cleaned paired-end reads to obtain longer contiguous sequences of the genomes of the organisms present in a given eDNA sample. Three different metagenome assemblers were run using multiple parameters (see Materials and Methods for details, Supplementary Table S1). As expected from previous studies (Kunin et al. 2008), longer stretches of contiguous sequences result in the formation of chimeric contigs that contain a mixture of sequences from multiple genomes (Supplementary Table S2); their results were thus excluded from further analysis. To compare whether our SoEM method could provide better performance of taxonomic classification compared to more conventional DNA metabarcoding, we additionally performed amplicon sequencing of four PCR-amplified markers (*COI*, *COI-5P*, *tufA*, and *rbcL*) using the Illumina Miseq platform. After trimming the adapters and removing PCR primers, 0.49 to 1.10 million clean reads were obtained from site #1 and site #2 (Table 4).

### The SoEM method can detect more biodiversity than classical DNA metabarcoding

In this study, we conducted taxonomic assignment us-

**Table 3.** Summary of sequencing statistics for samples used in the SoEM method

| Data | Total reads (raw data) | Total read base pairs (raw data) | Total reads (trimmed data) | PEAR-assembled read-pairs to single-end | Assembled reads (%) |
|---|---|---|---|---|---|
| Yamido (site #1) | 36,847,740 | 11,091,169,740 | 35,635,742 | 6,924,354 | 38.86 |
| Baealdo (site #2) | 36,503,228 | 10,987,471,628 | 35,101,464 | 6,651,439 | 37.90 |

Note: The read-length of all raw reads is 301 bp.
SoEM, small-organelles enriched metagenomics; PEAR, Paired-End reAd mergeR (see Materials and Methods).
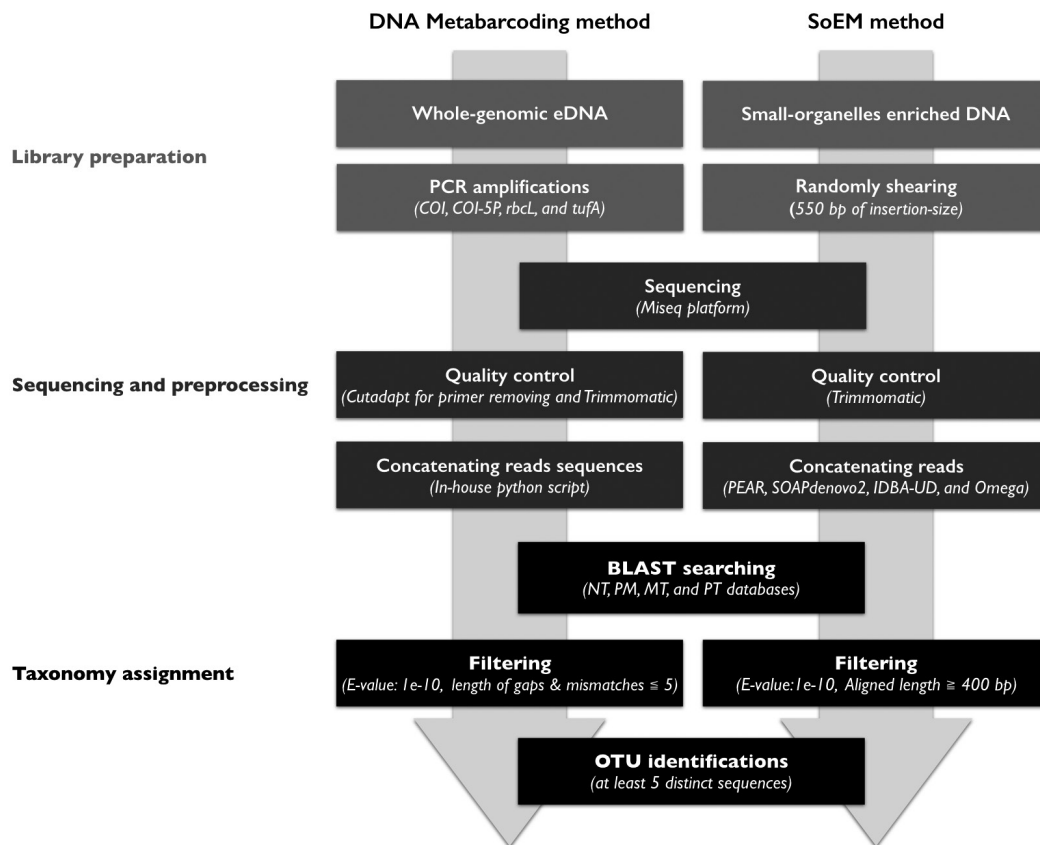
**Fig. 4.** Schematic workflow of the overall process of DNA metabarcoding and the small-organelles enriched metagenomics (SoEM) method. COI, cytochrome oxidase subunit I; NT, non-redundant nucleotide database; PM, primary-marker sequences from GenBank; MT, mitochondria whole genome database; PT, plastid whole-genome database; OTU, operational taxonomic unit.

ing all qualified merged reads (see Materials and Methods for details), which differs from other widely used taxonomic assignment approaches that rely on sequence clustering (Davenport and Tümmler 2013, Nilakanta et al. 2014, Porter and Hajibabaei 2018). These sequence reads were taxonomically assigned via BLAST search using four reference databases (NT, PM, MT, and PT). We then defined a taxonomic class if at least five distinct sequences, hereafter described as one OTU in this study, were classified as the same taxonomic classes. The de-

tailed operation procedures are described in Fig. 4.

For sampling site #1, a total of 430 SoEM OTUs were classified at genus level, whereas only 45 of the DNA metabarcoding OTUs were classified to this level (Fig. 5). Thus, almost 10-fold greater number of genus groups were identified using our SoEM method. When taxonomic assignment was restricted to a higher level, a 6.17-fold (185 vs. 30) and 3.11-fold (28 vs. 9) greater number of order and phylum groups were assigned, respectively. At the species level, the SoEM method identified 568 spe-

**Table 4.** Summary of sequencing statistics for samples used in the DNA metabarcoding method

| Data | Barcode region | Total read (raw data) | Total reads (trimmed) | Total average read length | Total average read length after trimming |
|---|---|---|---|---|---|
| Yamido (site #1) | *COI* | 550,186 | 528,630 | 440.11 | 389.70 |
| | *COI-5P* | 1,021,354 | 976,302 | 422.18 | 377.68 |
| | *tufA* | 1,005,994 | 968,436 | 332.46 | 297.48 |
| | *rbcL* | 1,026,754 | 973,032 | 404.19 | 355.03 |
| Baealdo (site #2) | COI | 1,172,700 | 1,088,284 | 406.00 | 357.61 |
| | COI-5P | 984,548 | 913,592 | 386.71 | 344.81 |
| | *tufA* | 1,009,662 | 953,572 | 399.75 | 368.48 |
| | *rbcL* | 527,250 | 485,476 | 389.01 | 350.25 |
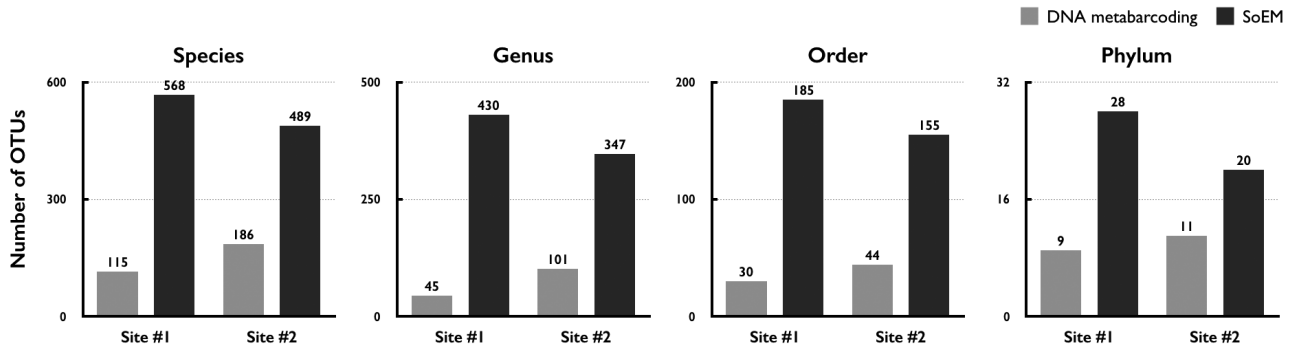
**Fig. 5.** Number of unique taxonomic groups (operational taxonomic units, OTUs) predicted by the DNA metabarcoding and small-organelles enriched metagenomics (SoEM) methods.
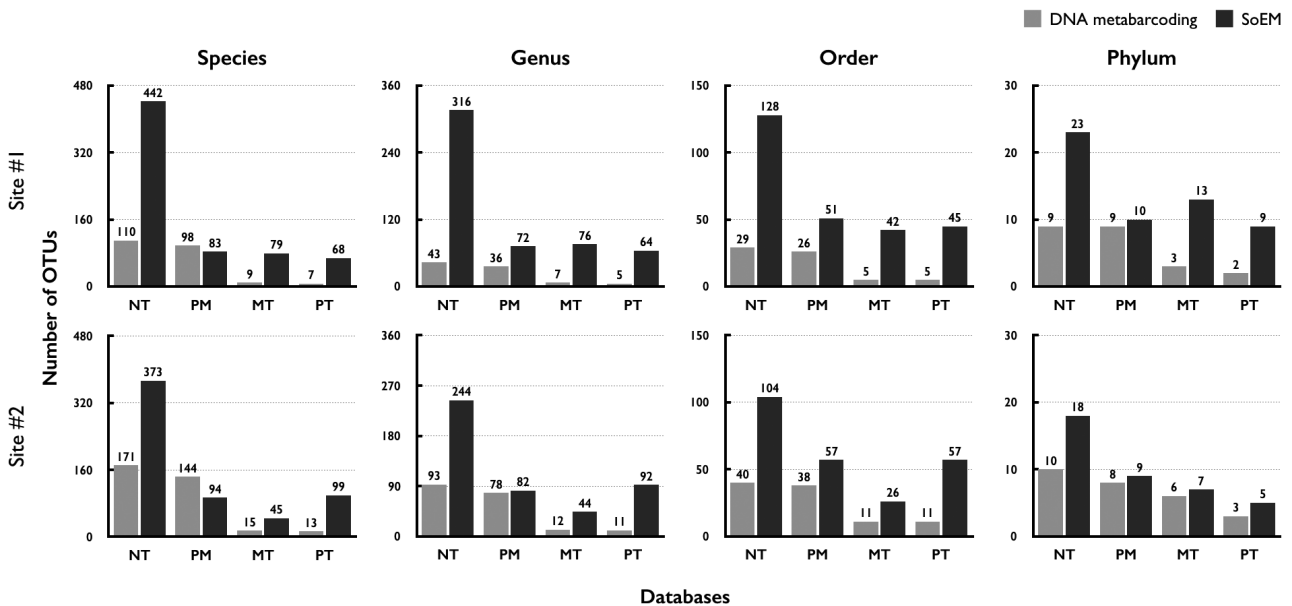


**Fig. 6.** The number of operational taxonomic units (OTUs) annotated by four distinct databases. SoEM, small-organelles enriched metagenomics; NT, non-redundant nucleotide database; PM, primary-marker sequences from GenBank; MT, mitochondria whole genome database; PT, plastid whole-genome database.

cies, compared to 115 species identified by DNA metabarcoding. This suggests that our SoEM method detects a broader range of biodiversity compared to the most widely used DNA metabarcoding method.

The SoEM method directly sequences small organelle genomic sequences, whereas DNA metabarcoding uses PCR primers to amplify a taxonomically informative marker gene from eDNA mixtures. This means that the quality of taxonomic identification relies on the selected reference databases. Here, we collected and used datasets composed of one all nucleotide collection database (NT), one marker-sequence database (PM), and two genome databases (MT and PT). Because these databases had different areas of interest in taxonomic classification, we repeated the same analysis on each database individually for taxonomic assignment. Except for the result

obtained when PM was used for species-level taxonomic classification (83 vs. 98 by SoEM and DNA metabarcoding, respectively), the same trends and slightly different magnitudes were observed (Fig. 6), indicating that the outperformance of the SoEM method was not due to the effect of selecting a particular database. Moreover, OTUs detected from both methods are mainly assigned to marine planktonic phyla such as Rotifera, Arthropoda, Nematoda, Mollusca, and Bacillariophyta. The information of taxon-assigned OTUs is presented in Supplementary Tables S3-S6.

Next, we determined how many taxonomic groups assigned by DNA metabarcoding method could also be identified by the SoEM method to examine the former to be superseded by the latter. At the species and genus levels, more than half of these taxonomic groups were not
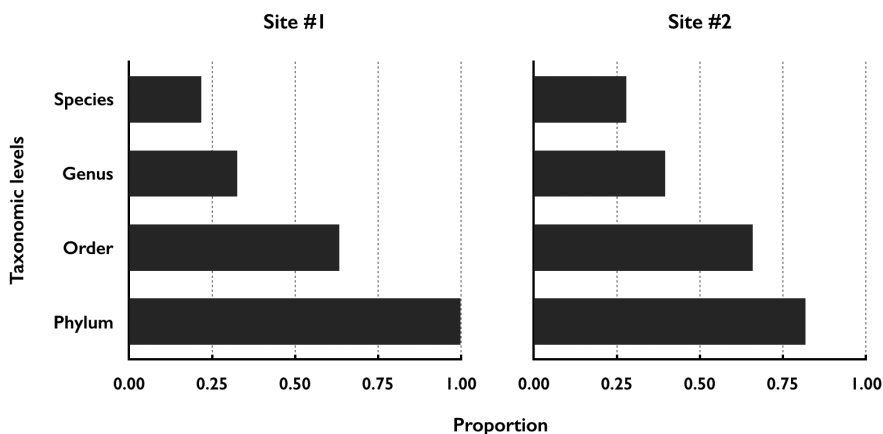
**Fig. 7.** The proportion of operational taxonomic units detected by the DNA metabarcoding method to those detected by the small-organelles enriched metagenomics method.
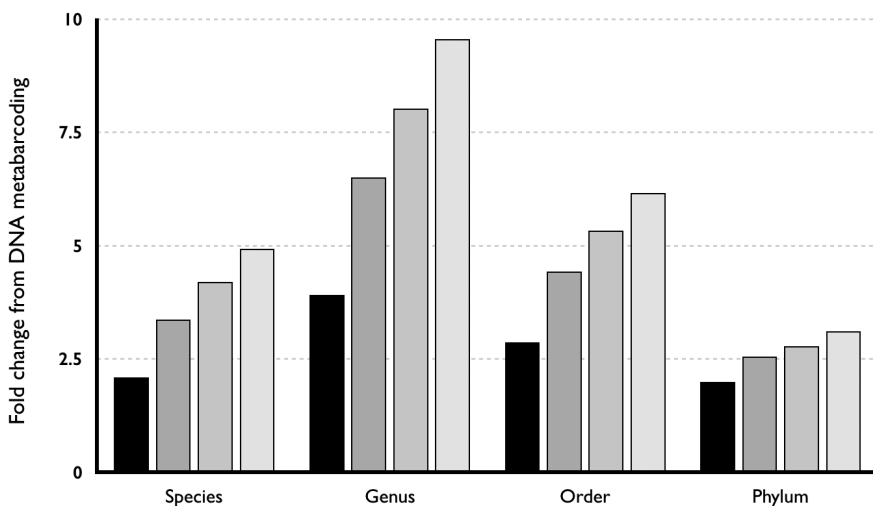


**Fig. 8.** Fold change differences in the number of taxa detected by DNA metabarcoding against those using different data sizes for the small-organelles enriched metagenomics (SoEM) method. Black color indicates 25% of the total raw sequence reads used for the SoEM method. Dark grey, grey, and light grey colors are their 50, 75, and 100%, respectively.

identified by the SoEM method, whereas this score was decreased in the higher taxonomic levels (Fig. 7). The reason for these results might be that DNA metabarcoding method does not provide an equally sensitive PCR assay for species-level detection to all target sequences derived from eDNA. Further studies are thus required to address this issue.

Our results may have arisen due differences in sequencing depth: we obtained an average 3.44 and 13.58 million-reads in the BLAST search for the DNA metabarcoding and SoEM method, respectively, which indicates about a 4-fold difference. To evaluate this hypothesis, we simulated different coverage depths by randomly sampling the raw sequence reads used for the SoEM method. Three different depths involving two higher coverages

(75% and 50% of total SoEM) and one similar coverage (25% of total SoEM) of sequence reads used for DNA metabarcoding method were presented. Following the same experimental approach, we observed that molecular DNA-based identification of taxonomic diversity depends on sequencing depth (Fig. 8). Nevertheless, when directly compared from a sample size perspective (25% of total SoEM), the numbers of detected taxa detected by the SoEM method were still higher (at least 2-fold) in all four taxonomic levels than those detected by DNA metabarcoding method. These findings indicate that our overall results, that the SoEM method detected a broader range of taxa than did DNA metabarcoding method, were not a by-product of the difference in sample size. To validate our method and experimental results, we repeated

the same analyses using data from sampling site #2 and observed no significant difference in results.

## Caveats

Despite the advantages to SoEM method over DNA metabarcoding method, there are several caveats in our method. First, morphological characteristics of the cells in the sea such as cell-size, shape, hardness, and tissue complexity were highly diverse. For this reason, the disruption efficiency of those cell walls and membranes was not evenly applied to all species in the homogenization process, and this disparity might affect the enrichment of small-organelles from rigid-cells such as diatoms (Kröger and Poulsen 2008). Thus, the experimental process for small-organelle enrichment with environmental samples has room for improvement. Second, the SoEM method required more DNA than the DNA metabarcoding method for successfully preparing the sequencing library. In fact, in our study, the SoEM method required about 200 ng of DNA per library, whereas about 12.5 ng of DNA was required for each library in the DNA metabarcoding method (see Materials and Methods). This difference in the demands for DNA between both methods arises from the fact that the SoEM library intrinsically requires a lot of DNA for whole genome shotgun-sequencing, and the amounts of DNA are inevitably reduced in the experiment of small-organelle enrichment. Thus, more efficient and optimized library preparation strategies are still required for further studies. Third, oceanic plankton communities highly interact with each other in response to changes in environmental conditions (Hays et al. 2005). Symbiotic relationships and prey-predator interactions between microbial eukaryotes and prokaryotes are well known and have been largely studied (Fenchel 1988, Decelle et al. 2012). We only focused on marine microbial eukaryotic ecology, and thus, numerous non-eukaryotic sequences were excluded in our analyses.

## CONCLUSION

In this paper, we introduce a newly established PCR-free metagenomics method for marine biodiversity assessment. The SoEM method and its taxonomic capturing capacity were investigated in two marine environmental samples. At least 3-fold higher numbers of taxa were identified in our SoEM method compared to those identified by the conventional multi-locus DNA metabarcoding method. These results are independent of different taxonomic levels (phylum, order, genus, and species). Consequently, our SoEM method is effective and accurate for identifying taxonomic diversity and presents a useful alternative approach to evaluate biodiversity in the marine environment. Although our proposed method is not optimized and still has room for improvement, we suggest that it will be a promising and alternative approach for phylogenetic diversity investigations from various environmental samples and believe that it will be an attractive method for many marine ecologists and conservation biologists.

## SUPPLEMENTARY MATERIAL

**Supplementary Table S1.** Statistics of metagenome assemblies (www.e-algae.org).

**Supplementary Table S2.** Number of chimeric-contigs in each assembler (www.e-algae.org).

**Supplementary Table S3.** List of taxon-assigned operational taxonomic units (OTUs) of small-organelles enriched metagenomics (SoEM) at site #1 (www.e-algae.org).

**Supplementary Table S4.** List of taxon-assigned operational taxonomic units (OTUs) of DNA metabarcoding at site #1 (www.e-algae.org).

**Supplementary Table S5.** List of taxon-assigned operational taxonomic units (OTUs) of small-organelles enriched metagenomics (SoEM) at site #2 (www.e-algae.org).

**Supplementary Table S6.** List of taxon-assigned operational taxonomic units (OTUs) of DNA metabarcoding at site #2 (www.e-algae.org).

## REFERENCES

Andújar, C., Arribas, P., Ruzicka, F., Crampton-Platt, A., Timmermans, M. J. T. N. & Vogler, A. P. 2015. Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. Mol. Ecol. 24:3603-3617.

Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P. & Emerson, B. C. 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. Mol. Ecol. 27:3968-3975.

Ansorge, W. J. 2009. Next-generation DNA sequencing techniques. N. Biotechnol. 25:195-203.

Appeltans, W., Ahyong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., Bamber, R., Barber, A., Bartsch, I., Berta, A., Błażewicz-Paszkowycz, M., Bock, P., Boxshall, G., Boyko, C. B., Brandão, S. N., Bray, R. A., Bruce, N. L., Cairns, S. D., Chan, T. -Y., Cheng, L., Collins, A. G., Cribb, T., Curini-Galletti, M., Dahdouh-Guebas, F., Davie, P. J. F., Dawson, M. N., De Clerck, O., Decock, W., De Grave, S., de Voogd, N. J., Domning, D. P., Emig, C. C., Erséus, C., Eschmeyer, W., Fauchald, K., Fautin, D. G., Feist, S. W., Fransen, C. H. J. M., Furuya, H., Garcia-Alvarez, O., Gerken, S., Gibson, D., Gittenberger, A., Gofas, S., Gómez-Daglio, L., Gordon, D. P., Guiry, M. D., Hernandez, F., Hoeksema, B. W., Hopcroft, R. R., Jaume, D., Kirk, P., Koedam, N., Koenemann, S., Kolb, J. B., Kristensen, R. M., Kroh, A., Lambert, G., Lazarus, D. B., Lemaitre, R., Longshaw, M., Lowry, J., Macpherson, E., Madin, L. P., Mah, C., Mapstone, G., McLaughlin, P. A., Mees, J., Meland, K., Messing, C. G., Mills, C. E., Molodtsova, T. N., Mooi, R., Neuhaus, B., Ng, P. K. L., Nielsen, C., Norenburg, J., Opresko, D. M., Osawa, M., Paulay, G., Perrin, W., Pilger, J. F., Poore, G. C. B., Pugh, P., Read, G. B., Reimer, J. D., Rius, M., Rocha, R. M., Saiz-Salinas, J. I., Scarabino, V., Schierwater, B., Schmidt-Rhaesa, A., Schnabel, K. E., Schotte, M., Schuchert, P., Schwabe, E., Segers, H., Self-Sullivan, C., Shenkar, N., Siegel, V., Sterrer, W., Stöhr, S., Swalla, B., Tasker, M. L., Thuesen, E. V., Timm, T., Todaro, M. A., Turon, X., Tyler, S., Uetz, P., van der Land, J., Vanhoorne, B., van Ofwegen, L. P., van Soest, R. W. M., Vanaverbeke, J., Walker-Smith, G., Walter, T. C., Warren, A., Williams, G. C., Wilson, S. P. & Costello, M. J. 2012. The magnitude of global marine species diversity. Curr. Biol. 22:2189-2202.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P. & Kauserud, H. 2010. ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases. BMC Microbiol. 10:189.

Bik, H. M., Fournier, D., Sung, W., Bergeron, R. D. & Thomas, W. K. 2013. Intra-genomic variation in the ribosomal repeats of nematodes. PLoS ONE 8:e78230.

Bolger, A. M., Lohse, M. & Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7:335-336.

Cha, R. S. & Thilly, W. G. 1993. Specificity, efficiency, and fidelity of PCR. PCR Methods Appl. 3:S18-S29.

Chamberlain, S. A. & Szöcs, E. 2013. taxize: taxonomic search and retrieval in R. F1000Res 2:191.

Crampton-Platt, A., Timmermans, M. J., Gimmel, M. L., Kutty, S. N., Cockerill, T. D., Vun Khen, C. & Vogler, A. P. 2015. Soup to tree: the phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. Mol. Biol. Evol. 32:2302-2316.

Crampton-Platt, A., Yu, D. W., Zhou, X. & Vogler, A. P. 2016. Mitochondrial metagenomics: letting the genes out of the bottle. Gigascience 5:15.

Davenport, C. F. & Tümmler, B. 2013. Advances in computational analysis of metagenome sequences. Environ. Microbiol. 15:1-5.

Decelle, J., Probert, I., Bittner, L., Desdevises, Y., Colin, S., de Vargas, C., Galí, M., Simó, R. & Not, F. 2012. An original mode of symbiosis in open ocean plankton. Proc. Natl. Acad. Sci. U. S. A. 109:18000-18005.

Dodsworth, S. 2015. Genome skimming for next-generation biodiversity analysis. Trends Plant Sci. 20:525-527.

Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C. M., Heled, J., Ross, H. A., Tooman, L., Grosser, S., Park, D., Demetras, N. J., Stevens, M. I., Russell, J. C., Anderson, S. H., Carter, A. & Nelson, N. 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. Gigascience 4:46.

Eberhardt, U. 2012. Methods for DNA barcoding of fungi. Methods Mol. Biol. 858:183-205.

Edgar, R. C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10:996-998.

Elbrecht, V. & Leese, F. 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass: sequence relationships with an innovative metabarcoding protocol. PLoS ONE 10:e0130324.

Federhen, S. 2012. The NCBI Taxonomy database. Nucleic Acids Res. 40(Database issue):D136-D143.

Fenchel, T. 1988. Marine plankton food chains. Annu. Rev. Ecol. Syst. 19:19-38.

Fontaneto, D., Kaya, M., Herniou, E. A. & Barraclough, T. G. 2009. Extreme levels of hidden diversity in microscopic animals (Rotifera) revealed by DNA taxonomy. Mol. Phylogenet. Evol. 53:182-189.

Greshake, B., Zehr, S., Dal Grande, F., Meiser, A., Schmitt, I. & Ebersberger, I. 2016. Potential and pitfalls of eukaryotic metagenome skimming: a test case for lichens. Mol. Ecol. Resour. 16:511-523.

Haider, B., Ahn, T. -H., Bushnell, B., Chai, J., Copeland, A. & Pan, C. 2014. Omega: an overlap-graph *de novo* assembler for metagenomics. Bioinformatics 30:2717-2722.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C. & Baird, D. J. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6:e17497.

Hays, G. C., Richardson, A. J. & Robinson, C. 2005. Climate change and marine plankton. Trends Ecol. Evol. 20:337-344.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. 2003*a*. Biological identifications through DNA barcodes. Proc. Biol. Sci. 270:313-321.

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. & Hallwachs, W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. U. S. A. 101:14812-14817.

Hebert, P. D. N., Ratnasingham, S. & deWaard, J. R. 2003*b*. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. Biol. Sci. 270(Suppl. 1):S96-S99.

Hillmann, B., Al-Ghalith, G. A., Shields-Cutler, R. R., Zhu, Q., Gohl, D. M., Beckman, K. B., Knight, R. & Knights, D. 2018. Evaluating the information content of shallow shotgun metagenomics. mSystems 3:e00069-18.

Huson, D. H., Mitra, S., Ruscheweyh, H. -J., Weber, N. & Schuster, S. C. 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res. 21:1552-1560.

Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B. & Gillespie, R. G. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Sci. Rep. 7:17668.

Kress, W. J. & Erickson, D. L. 2008. DNA barcodes: genes, genomics, and bioinformatics. Proc. Natl. Acad. Sci. U. S. A. 105:2761-2762.

Kress, W. J., Garcia-Robledo, C., Uriarte, M. & Erickson, D. L.

2015. DNA barcodes for ecology, evolution, and conservation. Trends Ecol. Evol. 30:25-35.

Kröger, N. & Poulsen, N. 2008. Diatoms-from cell wall biogenesis to nanotechnology. Annu. Rev. Genet. 42:83-107.

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K. & Hugenholtz, P. 2008. A bioinformatician's guide to metagenomics. Microbiol. Mol. Biol. Rev. 72:557-578.

Leasi, F. & Todaro, M. A. 2009. Meiofaunal cryptic species revealed by confocal microscopy: the case of *Xenotrichula intermedia* (Gastrotricha). Mar. Biol. 156:1335-1346.

Lee, H. -G., Kim, H. M., Min, J., Kim, K., Park, M. G., Jeong, H. J. & Kim, K. Y. 2017. An advanced tool, droplet digital PCR (ddPCR), for absolute quantification of the red-tide dinoflagellate, *Cochlodinium polykrikoides* Margalef (Dinophyceae). Algae 32:189-197.

Linard, B., Crampton-Platt, A., Gillett, C. P., Timmermans, M. J. & Vogler, A. P. 2015. Metagenome skimming of insect specimen pools: potential for comparative genomics. Genome Biol. Evol. 7:1474-1489.

Liu, S., Wang, X., Xie, L., Tan, M., Li, Z., Su, X., Zhang, H., Misof, B., Kjer, K. M., Tang, M., Niehuis, O., Jiang, H. & Zhou, X. 2016. Mitochondrial capture enriches mitoDNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. Mol. Ecol. Resour. 16:470-479.

Losos, J. B. 2010. Adaptive radiation, ecological opportunity, and evolutionary determinism. American Society of Naturalists E. O. Wilson award address. Am. Nat. 175:623-639.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S. -M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T. -W. & Wang, J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. GigaScience 1:18.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17:10-12.

McManus, G. B. & Katz, L. A. 2009. Molecular and morphological methods for identifying plankton: what makes a successful marriage? J. Plankton Res. 31:1119-1129.

Nilakanta, H., Drews, K. L., Firrell, S., Foulkes, M. A. & Jablonski, K. A. 2014. A review of software for analyzing molecular sequences. BMC Res. Notes 7:830.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S. S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A. M., Gile, G. H., Holzmann, M., Jahn, R., Jirků, M., Keeling, P. J., Kostka, M., Kudryavtsev, A., Lara, E., Lukeš, J., Mann, D. G., Mitchell, E. A. D., Nitsche, F., Romeralo, M., Saunders, G. W., Simpson, A. G. B., Smirnov, A. V., Spouge, J. L., Stern, R. F., Stoeck, T., Zim-

mermann, J., Schindel, D. & de Vargas, C. 2012. CBOL Protist Working Group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol. 10:e1001419.

Pawluczyk, M., Weiss, J., Links, M. G., Egaña Aranguren, M., Wilkinson, M. D. & Egea-Cortines, M. 2015. Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. Anal. Bioanal. Chem. 407:1841-1848.

Peng, Y., Leung, H. C., Yiu, S. M. & Chin, F. Y. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420-1428.

Porazinska, D. L., Giblin-Davis, R. M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., Powers, T. O., Tucker, A. E., Sung, W. & Thomas, W. K. 2009. Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. Mol. Ecol. Resour. 9:1439-1450.

Porter, T. M. & Hajibabaei, M. 2018. Scaling up: a guide to high-throughput genomic approaches for biodiversity analysis. Mol. Ecol. 27:313-338.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41(Database issue):D590-D596.

Roy, K. & Foote, M. 1997. Morphological approaches to measuring biodiversity. Trends Ecol. Evol. 12:277-281.

Saunders, G. W. & McDevit, D. C. 2012. Methods for DNA barcoding photosynthetic protists emphasizing the macroalgae and diatoms. Methods Mol. Biol. 858:207-222.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. & Weber, C. F. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75:7537-7541.

Schuster, S. C. 2007. Next-generation sequencing transforms today's biology. Nat. Methods 5:16-18.

Scotland, R. W., Olmstead, R. G. & Bennett, J. R. 2003. Phylogeny reconstruction: the role of morphology. Syst. Biol. 52:539-548.

Straub, S. C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C. & Liston, A. 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am. J. Bot. 99:349-364.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21:2045-2050.

Tang, M., Tan, M., Meng, G., Yang, S., Su, X., Liu, S., Song, W., Li, Y., Wu, Q., Zhang, A. & Zhou, X. 2014. Multiplex sequencing of pooled mitochondrial genomes: a crucial step toward biodiversity analysis using mito-metagenomics. Nucleic Acids Res. 42:e166.

Walker, B. H. 1992. Biodiversity and ecological redundancy. Conserv. Biol. 6:18-23.

Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614-620.

Zhou, J., Bruns, M. A. & Tiedje, J. M. 1996. DNA recovery from soils of diverse composition. Appl. Environ. Microbiol. 62:316-322.

Zhou, X., Adamowicz, S. J., Jacobus, L. M., Dewalt, R. E. & Hebert, P. D. N. 2009. Towards a comprehensive barcode library for arctic life: Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. Front. Zool. 6:30.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., Tang, M., Fu, R., Li, J. & Huang, Q. 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. Gigascience 2:4.