LIMNOLOGY and OCEANOGRAPHY: METHODS

ASLO

Limnol. Oceanogr.: Methods 21, 2023, 178–191

doi: 10.1002/lom3.10538

Small-organelle-enriched metagenomics: An improved method for environmental DNA-based identification of marine plankton

Soyeong Jin,[1] Hyun-Gwan Lee,[2] Chungoo Park,[1]* Kwang Young Kim [2]*

[1]School of Biological Sciences and Technology, College of Natural Sciences, Chonnam National University, Gwangju, Republic of Korea

[2]Department of Oceanography, College of Natural Sciences, Chonnam National University, Gwangju, Republic of Korea

Abstract

Understanding marine biodiversity is of paramount importance. Recently, we developed a polymerase chain reaction (PCR)-free small-organelle-enriched metagenomics (SoEM) method with enhanced performance in species identification compared to multimarker DNA metabarcoding. However, this method has a problem of applicability for field surveys, for example, the need to filter large volumes of seawater. Here, we propose an improved version of the original SoEM method (SoEM pioneer version [SoEM-pv]) for marine biodiversity assessment. There are three major contributions to this proposed SoEM method (SoEM v1.0): (1) presentation of PCR-free protocols for seawater environmental DNA analysis; (2) provision of widely applicable optimal DNA extraction protocols that can be applied to small volumes (approximately 2 liters) of coastal waters; and (3) description of easy-to-use bioinformatics pipelines for analysis of the SoEM data. To validate the accuracy and sensitivity of SoEM v1.0, we evaluated the performance of the taxonomic capture capacity from seawater samples and found that, given the sample volume of 2 liters, (1) SoEM v1.0 had an approximately 5- to 10-fold higher DNA extraction yield than the SoEM-pv; (2) SoEM v1.0 had the highest taxonomic diversity compared to the other methods, 1.6-fold higher than PCR-based multi-locus DNA metabarcoding and 1.2-fold higher than shotgun metagenomic sequencing; and (3) the number of species identified did not increase proportionally with increasing sample size from 2 to 20 liters. We suggest that SoEM v1.0 is a promising and feasible method for marine eukaryotic diversity studies and expect it to be welcomed by biological oceanographers and coastal resource managers.

Marine biodiversity is a fundamental feature of marine ecosystems and is essential for the sustainability of complex communities of interdependent and interacting organisms (Worden et al. 2015). Species in a community provide the biological foundation of ecosystem services that underpin human well-being, sustainable development, and healthy ecosystems (Cardinale et al. 2012; van der Plas 2019). Although the marine environment is widely recognized as a significant resource for human societies, marine biodiversity is severely threatened by multiple anthropogenic pressures, including habitat degradation and destruction (Crowe et al. 2013; Rogers et al. 2018), pollution (Magris et al. 2019), overfishing (Fernandes et al. 2017), invasive alien species (Ulman et al. 2019), and other various ecological or environmental issues related to climate change (McQuatters-Gollop et al. 2019). Despite the overall decline in global biodiversity (Díaz et al. 2019), species richness does not necessarily decline in locally disturbed areas if the extirpation of certain vulnerable species allows for the recruitment, establishment, and coexistence of more resilient species, according to the intermediate disturbance hypothesis (Roxburgh et al. 2004; Hillebrand et al. 2018).

Several institutions such as the US EPA's National Aquatic Resource Surveys, US Geological Survey, California Cooperative Oceanic Fisheries Investigations, Marine Biodiversity Observation Network (MBON), and Ocean Biogeographic Information System (OBIS) maintain databases of marine species in response to the unraveling and importance of species richness or taxon composition in the marine ecology. They have also responded

*Correspondence: chungoo@jnu.ac.kr; kykim@chonnam.ac.kr

Soyeong Jin and Hyun-Gwan Lee contributed equally to this study.

Author Contribution Statement: The study was designed by Kim, Lee, and Park. Lee and Jin conducted the field and laboratory experiments, and Jin and Park conducted the bioinformatics analysis. Kim and Park took the lead in writing the paper, but all authors contributed to it.

Additional Supporting Information may be found in the online version of this article.

with strategies for protecting, conserving, and restoring marine biodiversity. Notably, MBON and OBIS have worked closely with the Global Biodiversity Information Facility, which was established to develop services and technical solutions to integrate and coordinate biological observations from local to global scales and to build a community of practice to implement ecosystem-scale resource assessment and management (Canonico et al. 2019). In addition, intergovernmental organizations such as the Convention of Biological Diversity (CBD), Sustainable Development Goals (SDGs), Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, and International Union for Conservation of Nature also collect data on marine species and work toward coordinating and sharing information about marine biodiversity. In this context, the World Register of Marine Species is a recognized platform that lists over 240,000 marine species and adds about 2000 new species yearly (https://www.marinespecies.org at VLIZ, accessed: 30 January 2022).

Marine plankton communities are beneficial as indicators of climate change in marine ecosystems and form the food web base that supports many ecosystem functions, such as productivity, fish stock maintenance, and carbon storage (McQuatters-Gollop et al. 2019). Typically, biological samples for plankton diversity are obtained using plankton nets with various mesh sizes to sample the water column or water samplers; either Niskin bottles mounted on a conductivity-temperature-depth (CTD)-rosette system or a peristaltic pump. However, studying plankton diversity in the ocean has always been challenging because plankton range in size from microns to a few millimeters and include unicellular protists and small animals, requiring a variety of sampling strategies and techniques. Recently, DNA-based species identification approaches have emerged as an important method for characterizing marine biodiversity and have provided remarkable insights into community assembly and structure in the oceans (Stat et al. 2019). This method is a step toward less reliance on highly specialized taxonomic expertise and the ability to detect hidden or rare species (Ruppert et al. 2019). Besides traditional microscopic identification of plankton, DNA-based approaches have been widely employed (Suter et al. 2021). However, they are still under investigation due to a lack of reliability/reproducibility between methods.

With the rapid development of high-throughput sequencing technology and decreasing sequencing costs (Shendure et al. 2017), it has become possible to comprehensively analyze plankton communities using metabarcoding, defined as DNA barcoding for the simultaneous identification of a large number of species (Taberlet et al. 2012). For instance, Xue et al. (2018) investigated the temporal patterns of eukaryotic plankton communities following a cyanobacterial bloom using DNA barcoding approach. They found a robust temporal turnover of rare planktonic eukaryotes that were not detected by traditional morphology-based identification, coining the term "rare biosphere" (Sogin et al. 2006). Despite the convenience and effectiveness of DNA metabarcoding for large-scale species identification, its limitations have been recognized and extensively discussed (Taberlet et al. 2012). For example, PCR for DNA metabarcoding can introduce significant errors due to the degradation of template DNA or during DNA amplification and sequencing. Second, standardized taxonomic marker sequences (e.g., *COI*, *tufA*, *rbcL*, or *ITS*) are not necessarily suitable for DNA metabarcoding of bulk environmental samples (Porazinska et al. 2009; Bik et al. 2013). Such disproportional amplification of target sequences from different species, leading to so-called PCR bias, could lead to substantial distortions and influence our view of the plankton community structure.

Recently, researchers have made several attempts to avoid misinterpretation due to PCR-dependent artifacts. For example, the metagenome skimming approach captures the genomic diversity of low-coverage shotgun genome sequences in environmental genomics studies (Dodsworth 2015; Linard et al. 2015; Greshake et al. 2016). The shallow genomic sequences, thought to amplify high copy numbers of metabarcoding markers potentially, are de novo assembled, and the resulting assemblies are assigned to one of the genomes in the mixtures for taxonomic identification. However, further studies are limited to providing accurate taxonomic identification and covering the broadest range of species using low-coverage genomic sequences. Next, the mitochondrial metagenomics approach attempts to bioinformatically capture mitogenomic sequences from the shotgun sequencing of total DNA from specimen mixtures (Tang et al. 2014; Crampton-Platt et al. 2015; Crampton-Platt et al. 2016). Considering the varying DNA concentration, intraspecific variation, and chimeric sequences, the success of the mitochondrial metagenomics method relies on the formation of numerous mitogenome contigs. Another practical constraint is the relatively high cost of sequencing and the bioinformatic analysis burden for data acquisition. Most sequence reads are ultimately not used because they are not mitochondria-derived sequences. To further develop a cost-effective and efficient method for enriching mitochondrial genomes, Liu et al. (2016) presented a mitochondrial capture microarray pipeline for screening environmental samples prior to PCR-free shotgun sequencing and demonstrated higher mitochondrial capture enrichment efficiency than the PCR-based identification methods using a gene capture chip designed by the mitogenome sequences of the 1000-insect transcriptome evolution project (1KITE, http://www.1kite.org). Although many hurdles associated with decreasing sequencing efficiency and increasing operational costs have been overcome, cross-hybridization problems with microarrays remain to be solved. Zhou et al. (2013) employed a differential centrifugation method for experimentally enriched mitochondrial sequences without PCR amplification in bulk arthropod samples. Although the proposed method proved successful for molecular biodiversity analysis in well-controlled mock samples, its direct applicability to environmental community samples with a greater diversity of taxonomic groups should be carefully assessed and considered.

Recently, we demonstrated the necessity of improving the existing differential centrifugation-based mitochondrial enrichment approach and proposed modified differential centrifugation with optimized DNA extraction procedures in marine environment samples (Jo et al. 2019). The novel approach using a high-yield PCR-free small-organelle-enriched metagenomics (SoEM) method was introduced to eliminate possible PCR-related issues in metabarcoding and improve the efficacy of whole-genome metagenomic analysis for marine biodiversity assessment, targeting DNAs of small subcellular organelles (e.g., mitochondria and plastids). The DNA content of these organelles is widely used for taxonomic identification and molecular systematics. In experiments with various marine environment samples, the SoEM method was shown to identify at least threefold more taxa, regardless of taxonomic levels, than the conventional multilocus DNA metabarcoding method.

Despite the SoEM method attempting to overcome PCR inhibition and bias, there are still limitations to the direct application of this technique to marine samples. Because DNA sequences derived from small organelles represent only a tiny fraction of the total environmental DNA (eDNA) that can be extracted from the water sample, a large volume of environmental water samples is needed to obtain a sufficient amount of DNA for next-generation sequencing (NGS) analysis (Jo et al. 2019). The original version of the method required at least 1 g of organism (cell) pellets, equivalent to approximately 100 liters of mesotrophic coastal water. However, a larger sample volume is required for low-biomass pelagic communities in oligotrophic oceans. This results in longer sampling times and makes it impractical to continuously study the biodiversity of plankton communities over multiple stations during research vessel (RV) surveys.

This paper describes SoEM version 1.0 (SoEM v1.0), which has many significant improvements over the original method. Most importantly, SoEM v1.0 can be applied using only a few liters of mesotrophic coastal water. To this end, experimental improvements in the yield and quality of DNA extraction from highly diverse eukaryotic plankton communities must also be considered. By testing eDNA samples using different cell disruption and DNA extraction methods, SoEM v1.0 combines the ability to extract eDNA from small-volume water samples efficiently. In addition to these experimental efforts, bioinformatics studies of NGS-based metagenomics have also been conducted. The required NGS sequencing coverage and library size are specified in SoEM v1.0, and we can provide an easy-to-use application for analyzing and visualizing the SoEM v1.0 data. As a result, we have successfully developed an experimental and analytical method for eDNA based identification that is more advanced than the commonly used PCR-based metabarcoding and shotgun metagenomic sequencing methods.

Materials and procedures

Environments and eDNA sampling in the field

Coastal seawater was collected using a submersible pump (with a capacity of ~10 L min^{-1}) in Tongyeong Bay, Korea (34°51′35″N, 128°25′59″E) on 07 May 2021. During the 6-h sampling period, the surface water temperature ranged from 18°C to 22°C, and the salinity ranged from 25 to 30 psu, as measured by a probe (YSI Model 2030). Water temperature closely followed the diurnal air temperature pattern. The variation in salinity can be attributed to freshwater runoff from the surrounding coastal region. In Tongyeong Bay, the tidal range on the day of sampling was approximately 1.3 m, with the lowest tidal elevation being 0.7 m (12:31) and the highest tidal elevation being 2.0 m (18:41) above the lowest low water (LLW). At the time of the first sample, the tide was approximately 0.8 m above LLW (12:10). The ocean was relatively calm during three of the sampling times, with the last sample taken at approximately 1.9 m above LLW (18:10). To reduce the differences in plankton diversity between tidal levels, the vacuum filtration samples from the three sampling periods were randomly combined into a pooled sample.

For the original SoEM method (hereafter referred to as SoEM pioneer version or simply SoEM-pv), plankton was harvested using a 1-μm mesh filter (conical shaped) from 1 ton of prefiltered seawater passed through a 200-μm mesh and pelleted by centrifugation at 2500 × g for 20 min. For performing the PCR-based metabarcoding and shotgun metagenomic approaches and the SoEM v1.0 method, 1 liter of seawater was filtered on a 0.45-μm pore cellulose acetate (CA) membrane filter using an in situ filtration device under low vacuum (< 100 mm Hg). Seawater samples of 2, 5, 10, and 20 liters were obtained by combining several randomly selected filters to obtain a representative sample for each sample volume. Based on the overall efficiency of DNA extraction of samples obtained using different filter materials, that is, glass fiber (Advantech Inc.), polycarbonate (Advantech Inc.), and CA membrane filters (Advantech Inc.), the CA filter was the best for further enrichment of small organelles or subcellular organelles (data not shown). All samples were filtered through a CA filter or sieved through plankton netting to obtain the cell pellet in the field, transferred to the laboratory in liquid nitrogen, and stored at −80°C until DNA purification.

Enrichment of small organelles via differential centrifugation

The first crucial step is to enrich intact small organelles containing DNA from filtered or netted samples. To mechanically disrupt cells without damaging the membranes of subcellular components, cells were homogenized in ice-cold sucrose buffer (250 mM sucrose, 30 mM Tris–HCl, and 10 mM ethylene-diamine-tetra-acetic acid (EDTA); pH 7.5). For the SoEM-pv method, we followed the protocol described by Jo et al. (2019); however, for the SoEM v1.0 method, we used a new bead-beating protocol to extract DNA from the CA filtrate and cell pellet of plankton samples. Each collected sample was beaten with a mixture of 2- and 0.2-mm zirconia beads (Watson Co.) at 4000 rpm for 30 s. The process was performed in 10 mL of ice-cold homogenizing buffer (250 mM sucrose, 30 mM Tris–HCl, and 10 mM EDTA; pH 7.5) using FastPrep-24

(MP-Bio) and repeated four times with 1 min incubation on ice. The cell disruption steps were optimized based on beating cycles with different bead sizes and beating durations.

Next, we employed the differential centrifugation protocol for isolating and enriching small organelles, following the procedure used for insect community analysis (Tamura and Aotsuka 1988; Zhou et al. 2013). Briefly, the disrupted homogenates were centrifuged twice at 1300 × *g* for 10 min at 4°C to eliminate pelleted nuclei, cellular debris, and sedimentary contaminants. The supernatants containing the small organelles were pelleted by centrifugation at 17,000 × *g* for 30 min at 4°C. After removal of the supernatant, the final small organelle fractions were immediately frozen in liquid nitrogen and stored at −80°C until DNA purification.

Environmental DNA extraction

To select the most optimal DNA extraction procedure, we compared the efficiency of commercial DNA isolation kits for extracting DNA from fractionated small organelles after differential centrifugation in a sucrose buffer. In this study, three commercial methods were tested according to each kit's protocol guide: (1) commercial silica-based kits (Qiagen DNeasy Blood/Tissue Kit), (2) PowerSoil® DNA Isolation Kit (MO BIO Lab), and (3) Monarch Genomic DNA Prep Kit (NEB). Because the Monarch Genomic DNA Prep Kit produce at least two times higher amounts of DNA than the other kits in the comparative assessment (data not shown), this DNA extraction protocol was plugged into the SoEM v1.0 method. It was also used for PCR-based metabarcoding and conventional shotgun metagenomic analyses. The quality and concentration of the purified DNA samples were evaluated and measured using a Nanodrop-1000 (Thermo Scientific Co.) and a Qubit 3 fluorometer (Life Technologies).

PCR-based metabarcoding analysis

In this study, we used three representative marker genes, *18 S rRNA*, *COI*, and *rbcL*, for the eukaryotic, animal, and plant clades, respectively (Saunders and McDevit 2012; Kress et al. 2015). The hypervariable v4 region of the *18 S rRNA* gene was amplified using the TAReukFWD1 (5′-CCAGCASCYGCGGTAATTCC-3′) and TAReukREV3 (5′-ACTTTCGTTCTTGATYRA-3′) forward and reverse primers (Stoeck et al. 2010). The variable region of the mitochondrial *COI* gene was amplified using the mlCOIintF (5′-GGWACWGGWTGAACWGTWTAYCCYCC-3′) (Leray et al. 2013) and HCO2198 (5′-TAAACTTCAGGGTGACCAAAAAATCA-3′) forward and reverse primers (Folmer et al. 1994). The rbcL_F118_tak (5′-ACNTGGACNGTWGTDTGGAC-3′) and rbcL_R565_tak (5′-CKCATRAANGGYTGNGARTT-3′) forward and reverse primers were used to amplify the DNA fragment of the *rbcL* gene (Akita et al. 2020). For each primer set, 2 ng of gDNA was used for PCR amplification with 5× reaction buffer, 1 mM dNTP mix, 5 mM each of the forward and reverse PCR primers, and Herculase II fusion DNA polymerase (Agilent Technologies). Two rounds of PCR amplification were performed for each primer set. The cycling conditions of the first round of PCR amplification were 1 cycle at 95°C for 3 min, followed by 25 cycles at 95°C for 30 s, various annealing temperatures for 30 s, and 72°C for 30 s, with a final extension of 5 min at 72°C. An annealing temperature of 50.6°C, 40°C, and 55°C was used for *18 S rRNA*, *COI*, and *rbcL*, respectively. The first-round PCR products were purified using AMPure beads (Agencourt Bioscience), and 2 μL of the first-round PCR product was amplified using the NexteraXT Indexed Primer for final library construction containing the index. The cycling conditions for the second-round PCR amplification were 1 cycle at 95°C for 3 min, followed by 10 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s, with a final extension of 5 min at 72°C. The second-round PCR products were purified, and the final purified product was quantified by quantitative PCR (qPCR) according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using TapeStation D1000 ScreenTape (Agilent Technologies). Paired-end (2 × 301 bp) sequencing was performed by Macrogen Inc. using the Illumina MiSeq platform (Illumina Inc.). Raw sequencing data from the PCR-based metabarcoding study were deposited in the NCBI SRA database under the accession numbers SRR18489172, SRR18489180, and SRR18489181.

NGS for SoEM v1.0

To construct the sequencing libraries for SoEM v1.0 and conventional shotgun metagenomic sequencing using the TruSeq Nano DNA Library Prep kit (Illumina Inc.), a minimum of 100 ng of DNA was randomly fragmented for an insertion size of 350, 450, or 550 bp using a LE220 Focused-ultrasonicator (Covaris). The fragmented DNA was blunt-ended and phosphorylated following end repair. The appropriate library size was selected using different ratios of sample purification beads. These DNA fragments underwent the addition of a single "A" base and ligation of Illumina TruSeq indexing adapters. The purified libraries were quantified using qPCR according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using an Agilent Technologies 4200 TapeStation (Agilent Technologies). Paired-end sequencing (2 × 301 bp) was performed by Macrogen using the Illumina MiSeq platform (Illumina Inc.). All the raw sequencing data from the SoEM v1.0 and shotgun metagenomic sequencing were deposited in the NCBI SRA database under the accession numbers SRR18489173 to SRR18489179, SRR18489182, and SRR18489183 and under the NCBI Bioproject accession number PRJNA819962.

Meta-bioinformatics analysis

Raw sequence reads of the 5-, 10-, and 20-liter-based SoEM v1.0 and the shotgun metagenome were subsampled to equalize the average number of raw sequences of thsomeL-based

SoEM v1.0. Subsequently, these subsampled raw sequence reads were preprocessed using Trimmomatic (version 0.39) (Bolger et al. 2014) to remove the adapter sequences and bases with low sequencing quality (below an average Phred score of 33). PCR primer sequences were removed using Cutadapt (v3.4.) software (Kechin et al. 2017), allowing for a 10% mismatch in the primer sequence. To maximize the length of amplified target sequences, forward and reverse paired-end sequencing reads were aligned and joined using FLASH (*F*ast *L*ength *A*djustment of *SH*ort reads; v1.2.11) (Magoc and Salzberg 2011) with a minimum length of overlap for two short reads (—min-overlap = 10). Overlapping paired-end reads were discarded, and the resulting merged reads were used for further analyses.

A BLAST search (v2.8.1) (Camacho et al. 2009) was performed against the NCBI's non-redundant nucleotide sequence database, which includes approximately 71 million sequences (last updated in July 2021), to identify homologies of the merged sequences, with an *E*-value < 1e-10. We sorted the BLAST results in descending order of bit score and selected the one with the highest bit score. To obtain the taxonomic identification (taxID) of individual merged sequences, we assigned all taxonomic ranks from kingdom to species using the NCBI file gi_taxid_nucl.dmp (ftp://ftp.ncbi.nih.gov/pub/taxonomy/gi_taxid_nucl.dmp.gz) and in-house Python scripts. No available taxonomic information was represented as "N/A" at each level. All merged sequences assigned to the "Archaea," "Bacteria," and "Virus" superkingdoms were excluded. The case was manually curated if more than one merged sequence had the same highest bits core but with different or multiple taxIDs. If there were discrepancies in the taxIDs, the case was eliminated. We defined a taxonomic class if at least five distinct merged sequences (Supporting Information Fig. S1), hereafter described as an operational taxonomic unit (OTU), were classified into the same taxonomic class.

Total chlorophyll *a*

To assess fluctuations in phytoplankton biomass during the sampling period at the site, total chlorophyll *a* (Chl *a*) concentration was measured by high-performance liquid chromatography (HPLC) using a Waters Symmetry C8 column, equipped with diode array detection (DAD), through a pyridine salt gradient in methanol (Agilent HPLC1200 series). For pigment analysis, 1 liter of seawater was filtered under a low vacuum through a GF/F filter (Advantech Inc.), immediately frozen in liquid nitrogen, and extracted with 90% acetone at 4°C overnight after sonication. Chl *a* was identified by its retention time and optical absorption spectrum, compared with commercial pigment standards (DHI Lab), and quantified using external standard calibration curves of concentration vs. integrated peak area. All chemicals not specifically mentioned were of molecular or HPLC grade and were purchased from Sigma and Merck.

Assessment

To obtain sufficient DNA from small organelles such as mitochondria and plastids, large volumes of seawater must be filtered because their DNA constitutes only a small fraction of the total eDNA. To overcome this drawback in the existing SoEM-pv method, the procedures for extracting and purifying DNA from seawater samples were refined. Moreover, the effectiveness of different sample volumes and methods was evaluated (Fig. 1). After testing various combinations of assays for DNA yield and purity (*see* "Methods and Procedures" section for details), we developed a novel experimental protocol, SoEM v1.0, that incorporates the sophisticated and complex processes involved in collecting organisms using the 0.45-μm CA filter and combining bead-beating based cell lysis and the Monarch Genomic DNA Purification Kit. The SoEM v1.0 method also included bioinformatics and statistical analyses (Fig. 1B).

The coastal planktonic communities were sampled at one location during flood tide (*see* "Methods and Procedures" section for details) to assess how well the SoEM v1.0 method could identify the eukaryotic taxa represented in the eDNA communities. The eDNA fragments extracted by the four different methods were sequenced using NGS by Illumina MiSeq, followed by bioinformatics analyses (including preprocessing of large-scale NGS data, gathering of basic taxonomic information, and analyzing the OTUs for biodiversity profiling). The results are presented in Table 1.

Comparison of newly proposed SoEM v1.0 and original SoEM-pv methods

The amounts of DNA extracted using both SoEM v1.0 and SoEM-pv methods were quantified, and the DNA extraction yield was approximately 5–10 times higher in SoEM v1.0 than in SoEM-pv protocols (data not shown). However, the total amount of DNA obtained using the SoEM-pv protocol was insufficient for NGS-based metagenomic analyses, despite multiple attempts by three trained and experienced individuals following the same protocol. This lack of experimental reproducibility has led to the development of a new standard method for assessing plankton biodiversity, essentially the goal of the proposed method.

Biodiversity comparisons between SoEM v1.0 and PCR-based metabarcoding and shotgun metagenomic sequencing methods

We compared the results for taxonomic classification outcomes between SoEM v1.0 and metabarcoding to assess whether the new method is superior to the most widely used PCR-based metabarcoding method. In the preliminary study, two experimental sessions were conducted independently for the SoEM v1.0 method to examine its reproducibility and validity. More than three-quarters of the OTUs were assigned to the same species-level taxonomy, and this score increased

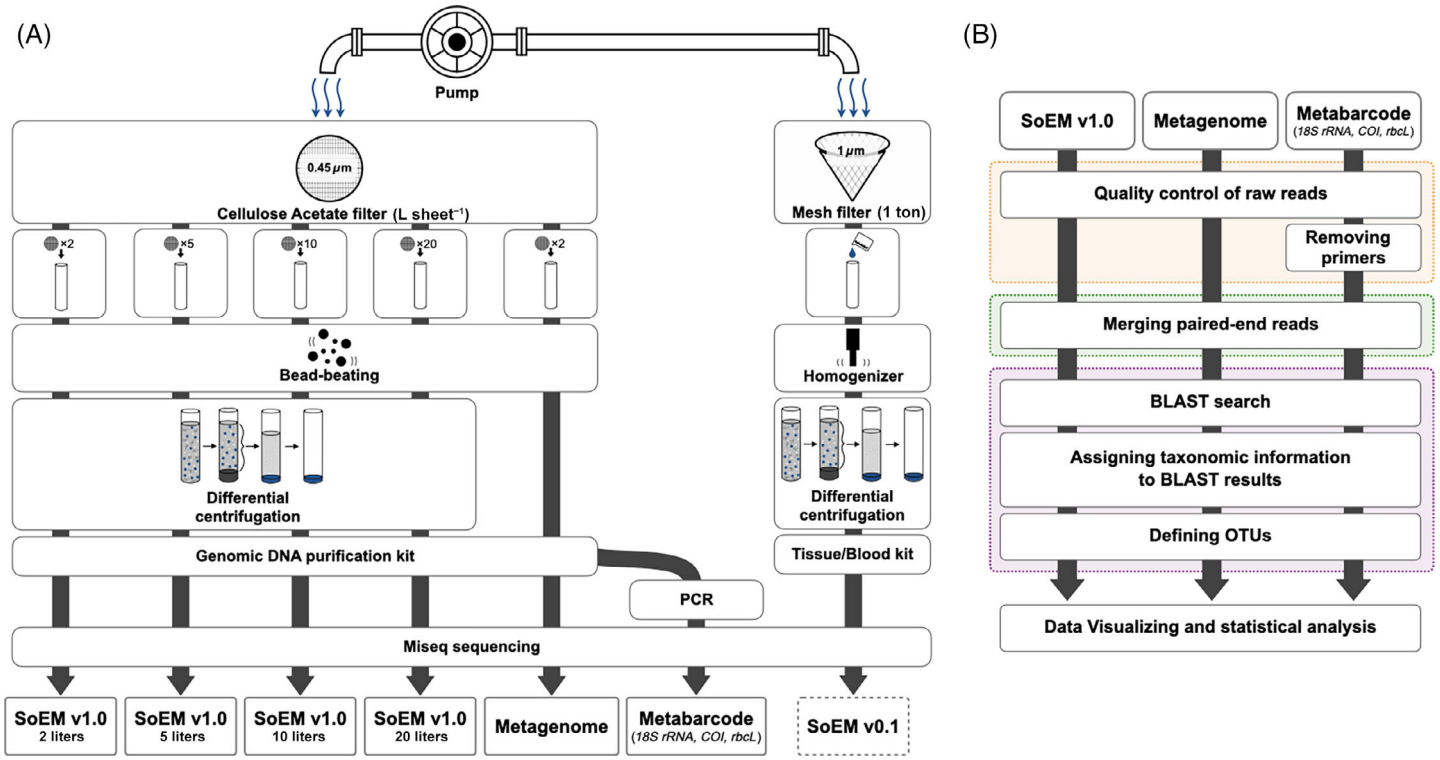

Fig. 1. End-to-end experimental (**A**) and bioinformatic (**B**) workflows for the SoEM-pv, SoEM v1.0, PCR-based metabarcoding, and shotgun metagenomic sequencing methods.

as expected at higher taxonomic levels (Supporting Information Fig. S2), indicating that the results were highly reproducible. While predicting all three marker genes (*18S rRNA*, *COI*, and *rbcL*) to be used for taxonomic profile determination, a total of 546 metabarcoding OTUs were classified at the species level. This value was mostly from *COI* genes (61.5%, 336 out of 546), which might be due to the continuous expansion of *COI* sequences through many studies. The SoEM v1.0 method identified 886 species, which was approximately 1.62 times more than that of the metabarcoding method (Fig. 2). When the taxonomic assignment was restricted to a higher level, 1.91-fold (647 vs. 338), 1.84-fold (441 vs. 240), and 1.73-fold (263 vs. 152) higher numbers of genera, families, and orders were assigned, respectively. We only identified an average of 12.9% of species-level OTUs in both methods, which was congruent with our previous study (Jo et al. 2019). This result

Table 1. Summary statistics for SoEM v1.0 (2, 5, 10, and 20 liters), PCR-based metabarcoding, and shotgun metagenomic sequencing methods.

Methods	Number of reads trimmed	Number of reads merged	Average length of merged reads (bp)	Number of reads merged (≥ 300 bp)	Number of OTUs
SoEM v1.0_R1$_{(2\ liters)}$	8,747,638	3,585,378	422.7	3,512,494	108,466
SoEM v1.0_R2$_{(2\ liters)}$	9,631,884	3,947,538	423.3	3,866,582	124,852
MB$_{18S\ rRNA}$	294,556	142,707	419.9	142,707	63,758
MB$_{COI}$	335,994	155,922	416.6	155,922	84,941
MB$_{rbcL}$	203,576	86,429	408.8	86,429	63,630
MG	9,534,418	3,713,157	428.1	4,250,392	159,996
SoEM v1.0$_{(5\ liters)}$	9,254,458	3,883,182	406.0	3,827,353	69,042
SoEM v1.0$_{(10\ liters)}$	8,931,538	3,666,262	409.7	3,615,215	55,748
SoEM v1.0$_{(20\ liters)}$	9,160,648	3,685,893	315.5	3,633,483	67,199

MB indicates the PCR-based metabarcoding method. MG indicates the shotgun metagenomic sequencing method.

may be because the metabarcoding method does not provide an equally sensitive PCR assay to all target sequences for species-level detection.

Instead of targeting a specific genomic locus for amplification, shotgun metagenomic sequencing, which has been referred to as the untargeted (shotgun) sequencing of all (meta) genomic DNAs (genomic), is an emerging methodology for direct genomic analysis of a complex community while avoiding the limitations of PCR-based methods (Tyson et al. 2004; Sharpton 2014). Shotgun-based metagenomic sequencing was also conducted on the same plankton community using the Illumina MiSeq platform. We compared its taxonomic classification data with the results obtained using the SoEM v1.0 method. At the species level, a total of 766 shotgun metagenomic OTUs were classified, which was more than that of the metabarcoding (546) but lesser than that of the SoEM v1.0 method (886). This differential pattern was retained across all taxonomic ranks (levels) (Fig. 2).

We further investigated whether the overall characteristics of taxonomic composition were consistent with the methods (Fig. 3). Using the top five taxonomic distributions that explained one-third of the entire taxonomic list, Metazoa, Viridiplantae, and Fungi were the major kingdoms detected by all three methods. At the phylum level, the top five taxonomic ranks detected by the SoEM v1.0 method were all detected by the shotgun metagenomic sequencing method, except for Streptophyta (green algae; the sixth taxonomic rank in the shotgun metagenomic method). Instead, Rhodophyta (red algae; the seventh taxonomic rank in the SoEM v1.0 method) was added to the top five lists of the shotgun metagenomic sequencing method. Compared to metabarcoding, three phyla (Arthropoda, Bacillariophyta, and Chlorophyta) were consistently included in the top five lists of the SoEM v1.0 method. However, Rhodophyta and Oomycota (water molds; the 11th taxonomic rank in the SoEM v1.0 method) were excluded. At the class level, only Hexanauplia (copepods; the 15th taxonomic rank in the shotgun metagenomic method) was not detected in the top five lists of the shotgun metagenomic method. Florideophyceae (red algae; the ninth taxonomic rank in the SoEM v1.0 method) was additionally detected by the shotgun metagenomic sequencing method. However, only Dinophyceae (dinoflagellates) was consistently included in the top five lists of the SoEM v1.0 and metabarcoding methods. From this taxonomic coverage and profiles, we suggest that the SoEM v1.0 method was superior to the metabarcoding method for taxonomic retrieval and was quantitatively comparable to the shotgun metagenomic sequencing method.

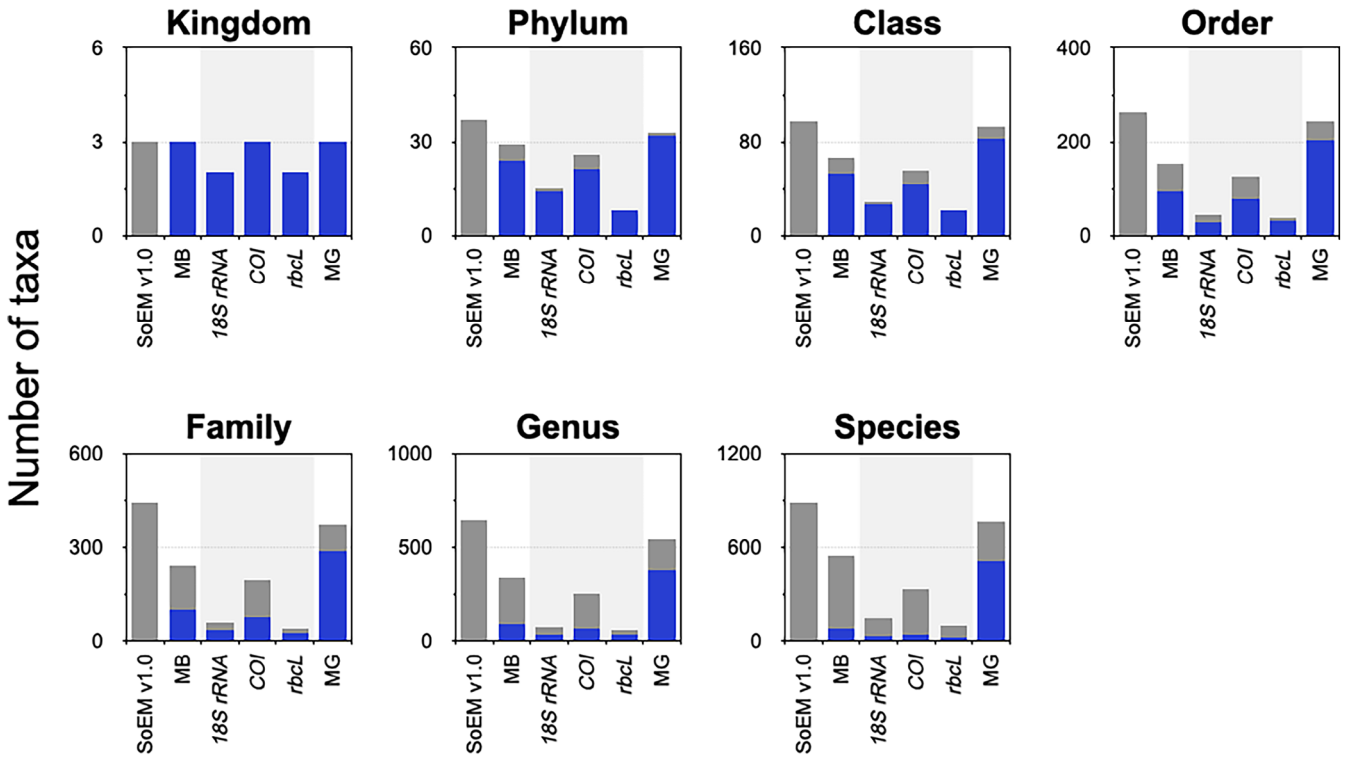

Fig. 2. Number of taxa detected using the SoEM v1.0, PCR-based metabarcoding, and shotgun metagenomic sequencing methods at each taxonomic level. Bars indicate the number of taxa detected by each method. Each vertical bar with blue color indicates the number of taxa commonly detected by SoEM v1.0 and each assessment method. MB stands for "PCR-based metabarcoding method," and the score is the sum of the number of taxa detected using three (*18 S rRNA*, *COI*, and *rbcL*) PCR-based metabarcoding methods. MG stands for "shotgun metagenomic sequencing method."

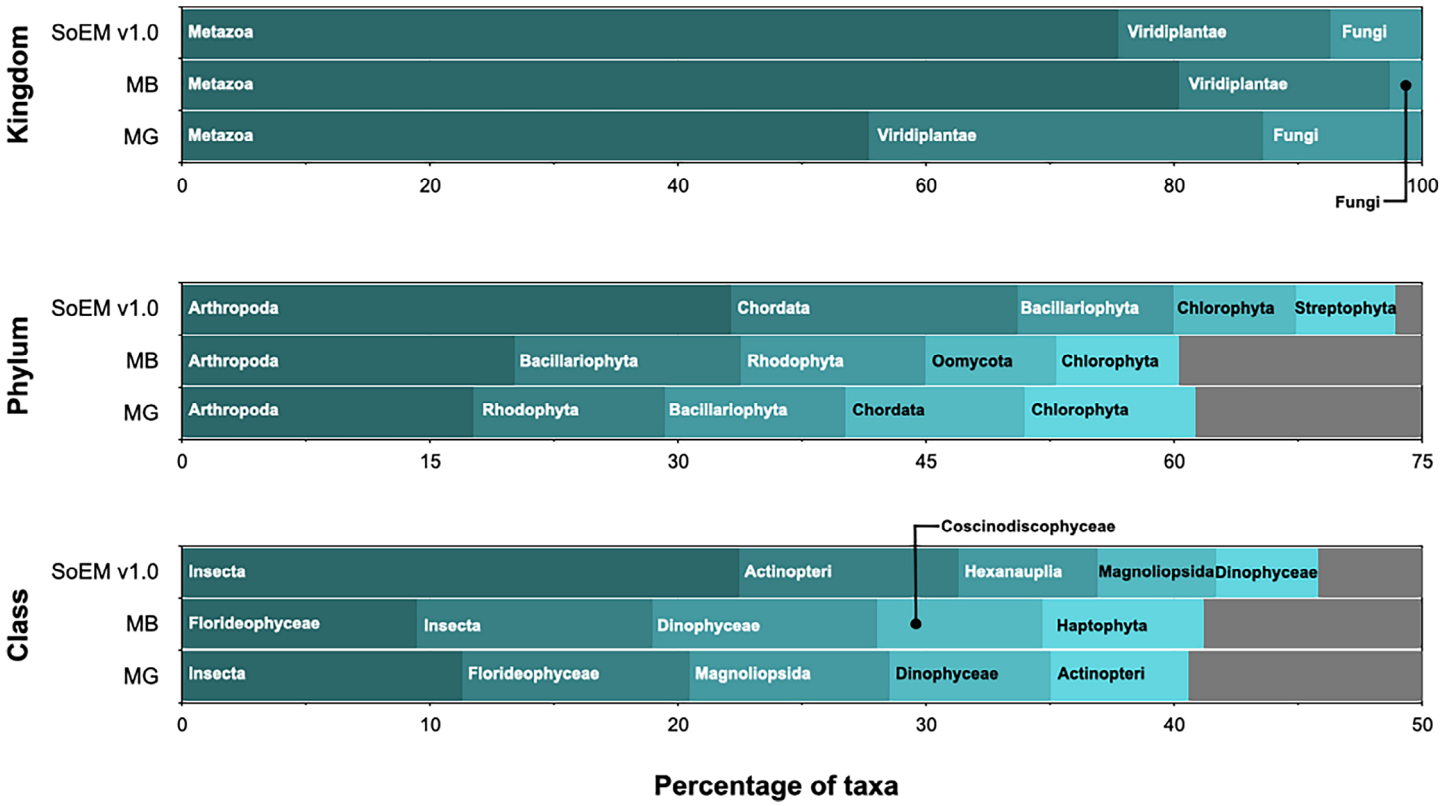

Fig. 3. Taxonomic distribution of plankton diversity detected by the SoEM v1.0, PCR-based metabarcoding, and shotgun metagenomic sequencing methods. The top five taxonomic distributions, which explain one-third of the total taxonomic list, are shown. MB and MG stand for "PCR-based metabarcoding method" and "shotgun metagenomic sequencing method," respectively.

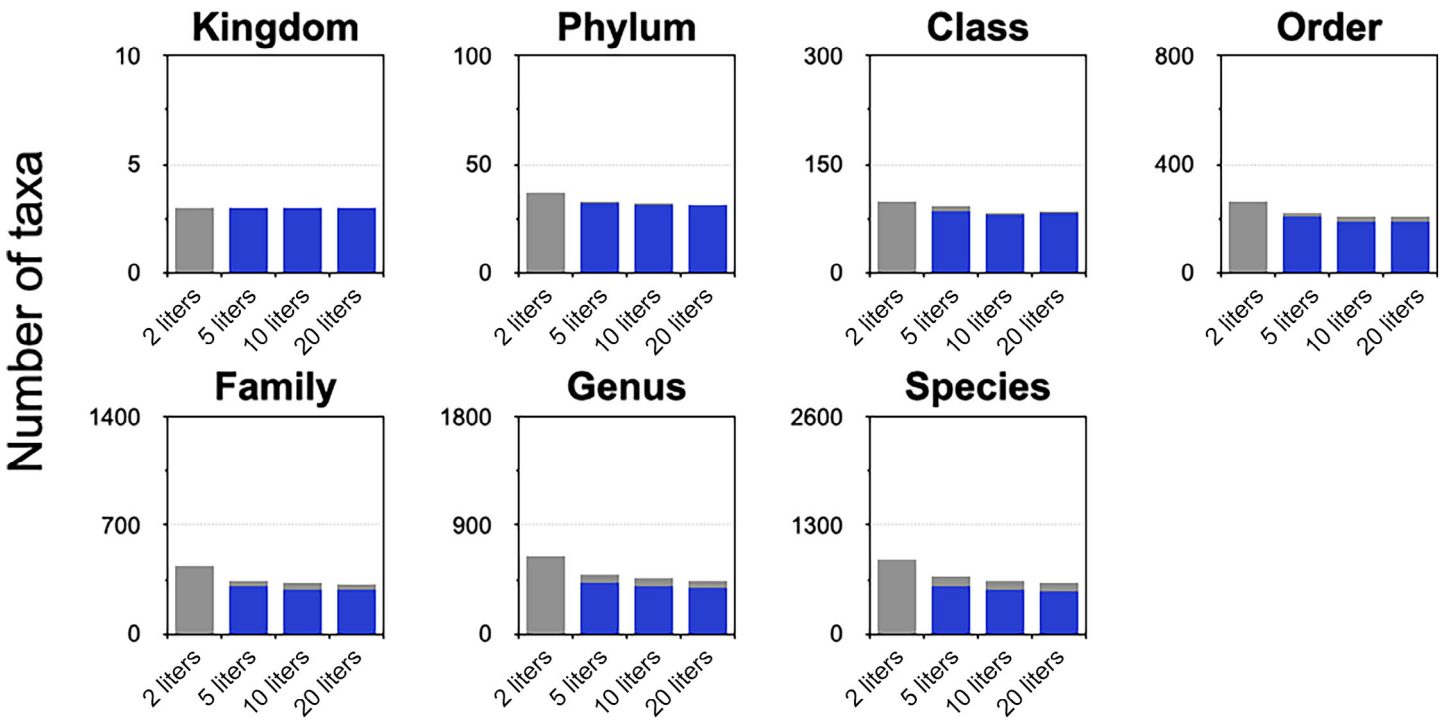

Fig. 4. Number of taxa detected in seawater samples of different volumes. Bars indicate the number of taxa detected in test samples. Each vertical bar with blue color indicates the number of taxa commonly detected by both the 2-liter-based SoEM v1.0 and the 5-liter-based (or 10- or 20-liter-based) SoEM v1.0 methods.

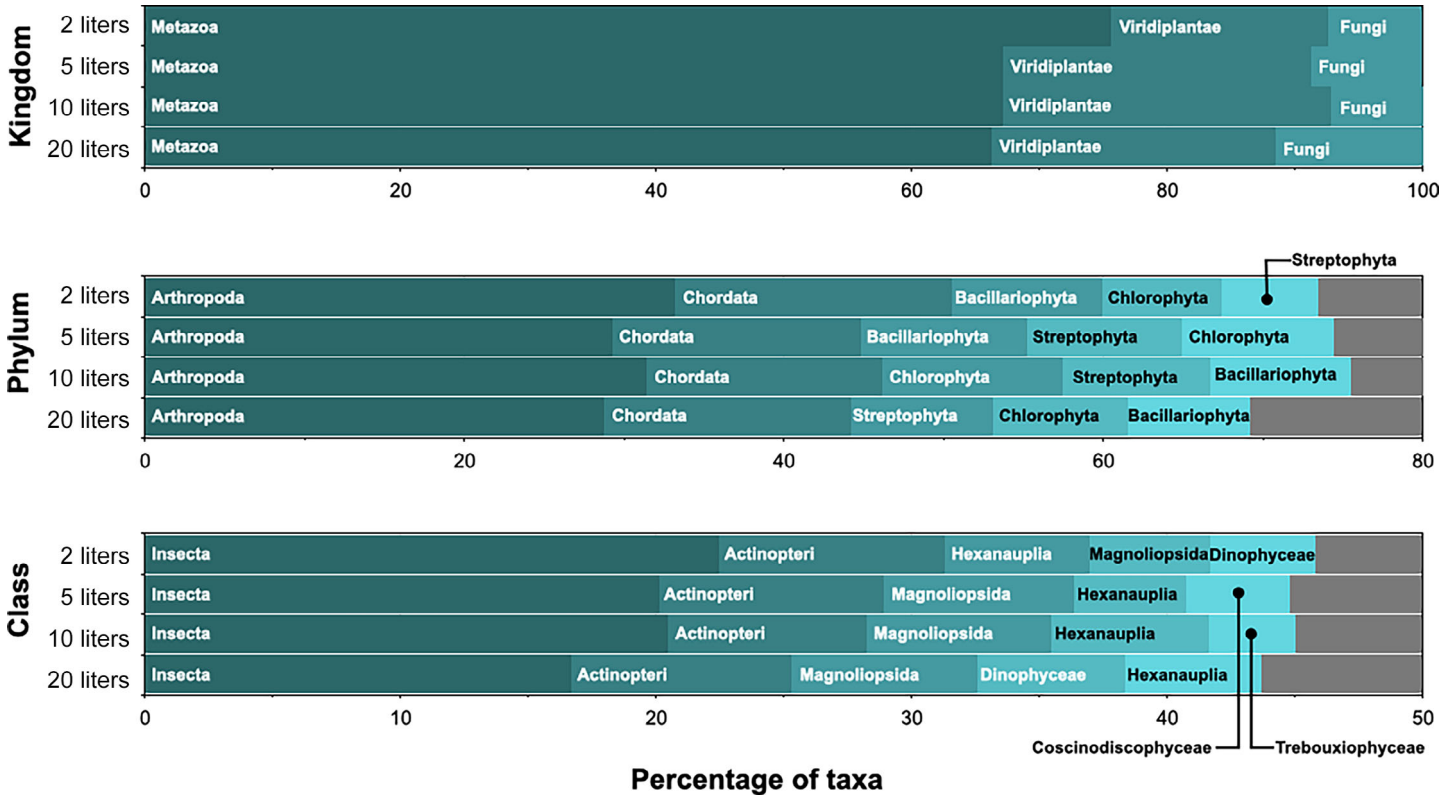

Fig. 5. Taxonomic distribution as per the 2-, 5-, 10-, and 20-liter-based SoEM v1.0 methods. The top five taxonomic distributions explain one-third of the total taxonomic list provided.

Estimate the necessary sample size for the SoEM v1.0 procedures

Regarding the implementation, 2 liters of seawater was used as the minimum sample volume for the SoEM v1.0 method. We questioned whether the experimental results were sensitive to the sample volume. If more seawater samples were available for the SoEM v1.0 test, the percentages of taxonomically assigned OTUs could be tweaked. In the present study, three different sample volumes (5, 10, and 20 liters) were prepared (Fig. 1A), and the results were compared with the data obtained from the 2-liter-based SoEM v1.0 analysis. Counterintuitively, the number of taxonomical OTUs detected did not increase proportionally with increasing sample volume and was nearly saturated at 2 liters (Fig. 4). Furthermore, the overall patterns of taxonomic composition detected using the 5-, 10-, and 20-liter-based SoEM v1.0 methods were almost concordant with those of the 2-liter-based SoEM v1.0 method (Fig. 5). These results suggest that our SoEM v1.0 protocol is much less sensitive to sample volume and that 2 liters are large enough to detect species at different taxonomic levels.

Discussion

Understanding spatiotemporal biodiversity patterns in marine ecosystems are crucial for the sustainable management of marine living resources, with a focus on global climate and environmental issues (Chavez et al. 2003; Miloslavich et al. 2018). As NGS technologies have been successfully developed and exploited to obtain a vast amount of DNA sequences at a lower cost, the NGS-based DNA metabarcoding approach is considered a potent tool to provide in-depth biodiversity data for various ecological and environmental applications (Djurhuus et al. 2020). However, two innate vulnerability issues for PCR-based DNA metabarcoding methods should be carefully addressed. The first is the difficulty of finding a DNA metabarcoding marker that works well across taxonomic groups, especially in diverse environmental samples containing complex eukaryotic assemblages. Although a multimarker metabarcoding approach has been proposed as a more practical alternative (Drummond et al. 2015; Zhang et al. 2018), it remains a challenge to determine suitable metabarcoding primer sets for studying marine eukaryotic communities. The second is PCR amplification bias, which leads to taxonomic confusion and difficulty in identification. Many different target DNA fragments derived from a bulk environmental sample should be amplified with the same efficiency, but standardized DNA barcoding primers cannot provide equally suitable matches to all target sequences (Pawluczyk et al. 2015; Krehenwinkel et al. 2017). These intrinsic technical limitations are the main weaknesses of this method.

Alternatively, the whole-genome shotgun metagenomic sequencing (wgSMS), in which the genomes of all organisms in biological samples are sequenced, has been proposed and applied to the eDNA-based taxonomic classification. Compared to DNA metabarcoding studies, wgSMS analysis provides a broader range of information with greater specificity and sensitivity for taxonomic classification. With the recent increase in the use of eukaryotic metagenomics (Parducci et al. 2019; Bovo et al. 2020), the wgSMS method has some practical issues, such as the increasing cost of sequencing and the complexity of the required bioinformatics analyses (e.g., marker gene analysis, metagenome binning, and sequence assembly). Usually, the cost of wgSMS is at least 5–10 times that of DNA metabarcoding sequencing. The wgSMS can produce vast amounts of ambiguous DNA sequences that are difficult to use for distinguishing taxonomic groups because of missing information in taxonomic databases.

The proposed improved SoEM v1.0 method was designed to circumvent the above-mentioned technical issues and resolve the practical issue of the large sample volume required in the original SoEM-pv method. In implementing the original SoEM-pv method, we focused on the technical feasibility of enriching small organelles from seawater samples and assessing DNA purity. Furthermore, the new advanced SoEM v1.0 method is applicable for analyzing small sample volumes (approximately 2 liters) of seawater. In the original SoEM-pv method, filtered seawater samples were lysed using a tissue homogenizer protocol, but we observed that many undisrupted cells were left untreated. Therefore, the SoEM-pv method requires processing at least several hundred liters of seawater. To improve the DNA extraction yield from a small volume of seawater and efficiently obtain a sufficient amount of DNA for NGS analysis, we adopted a bead-beating-based extraction protocol in the SoEM v1.0 method. The bead-beating-based lysis procedure, which combines different bead materials of various sizes, can efficiently disrupt marine eukaryotic organisms of varying sizes filtered (isolated) from pelagic seawater. Membrane filtration of samples obtained by the bead-beating protocol can also affect DNA extraction yields, allowing a small volume of seawater to obtain high quality and quantity of DNA suitable for NGS analysis. This approach can also reduce experimental runtime and thus inhibit intracellular DNA degradation.

Seawater samples were collected at different times during flood tide at a shallow bay site to compare the effectiveness of metabarcoding, shotgun metagenomics, and the SoEM v1.0 approaches in detecting plankton species. The Chl *a* concentration during the 6 h sampling period varied temporally from 3.61 to 8.28 μg L^{-1}. This variation may lead to differences in plankton assemblages at each sampling time associated with the local tidal advection of different water masses. As a cause of the change in species composition, our results included the effect sampling at different times of the day had on the variability of plankton assemblage (Figs. 4, 5). Despite the variation in Chl *a* concentration, an indicator of phytoplankton standing crop, the feasibility of the new sampling mode (accessible and affordable sample volumes) and the effectiveness of the new SoEM v1.0 protocol compared to other methods can be evaluated for the reliability or reproducibility of the method.

We also developed bioinformatics pipelines for SoEM v1.0 data analysis. Quality control of raw sequence reads, merging two overlapped paired-end reads, and a BLAST analysis against the NCBI nr database with taxonomic information was performed continuously. The collection of merged reads mapping to the same taxon was defined as that species of taxon observed in the natural water sample. This bioinformatics analysis can be applied to the analysis of DNA metabarcoding sequence data. However, it should be handled differently in shotgun metagenomic data. For example, two binning-based methods are used for shotgun metagenomic analysis. First, the supervised binning method uses already sequenced full genome sequences to label contig sequences with taxonomic information. In bacterial shotgun metagenomics, metrics based on *k-mer* frequencies are widely used to bin contigs; however, these *k-mer* frequencies cannot be used to differentiate among eukaryotic genomes. In addition, this supervised method is still challenging because most eukaryotic species have not been sequenced. Therefore, a large fraction of contig sequences cannot be mapped to reference genomes. Second, the unsupervised clustering method assigns contig sequences to groups according to similarities in their specific characteristics, such as total contig coverage and intra-genome GC content variation. Again, these genomic features can provide a powerful signal to group contigs in bacterial metagenomic data but not in eukaryotic data. To directly compare the results of these three distinct methods, we forcibly applied our bioinformatics pipelines to the shotgun analysis of metagenomic data. As a result, the total number of OTUs was higher in the shotgun metagenomic data (Table 1). However, the number of OTUs classified as species was higher in the SoEM v1.0 data (Fig. 2), indicating that the SoEM v1.0 method can more efficiently identify species from eDNA samples.

Because the longer the process, the better the performance of taxonomic assignments, the SoEM v1.0 method involves performing Illumina MiSeq shotgun metagenomic sequencing to obtain longer sequences from eDNA samples. According to the standard MiSeq platform, multiple dual-index libraries were prepared using the TruSeq DNA Prep Kit and sequenced on an Illumina MiSeq instrument using paired-end sequencing reactions (Katsuoka et al. 2014). For example, the Illumina protocol recommends using an average insert size of 550 bp or smaller for generating 2 × 300 bp paired-end reads. Therefore, the sequences with paired-end reads have about 50 bp of overlapping sequences in the middle. Indeed, we counted the number of paired-end reads that overlapped into single longer reads and found that only half of the total raw sequence reads with a library insert size of 550 bp had at least 10 bp of overlapping sequences (Fig. 6A), implying that many sequence

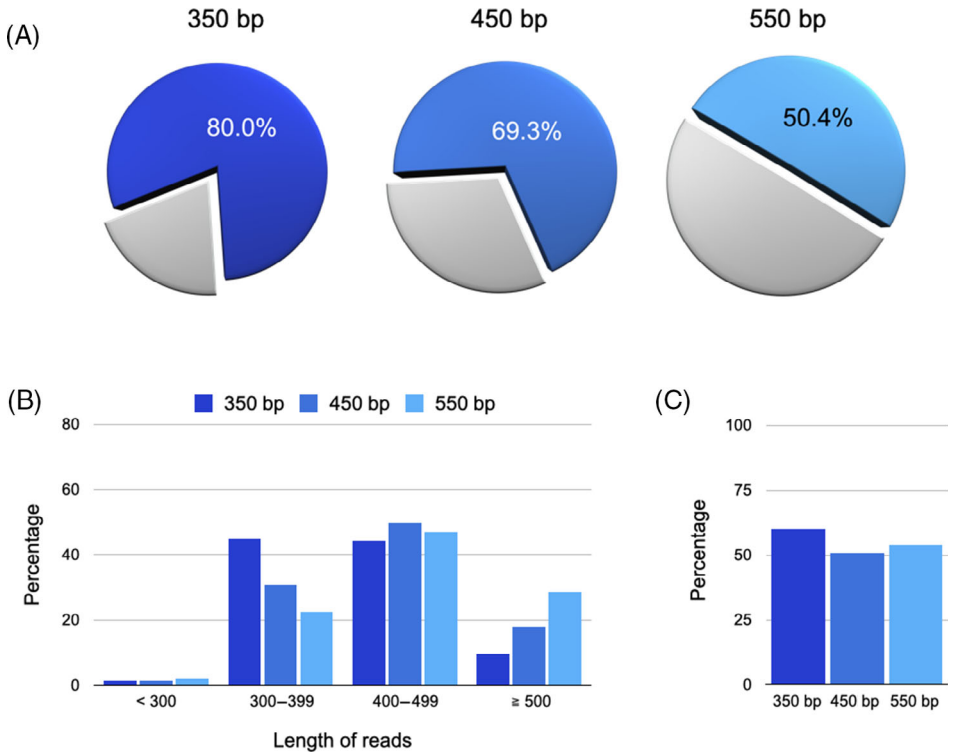

Fig. 6. Comparison of statistics in NGS tests of different library sizes in the SoEM v1.0 method. (**A**) Percentage of merged sequences. Blue indicates the percentage of merged sequences. The percentage of non-overlapping sequences is shown by gray. (**B**) Distribution of merged sequence lengths. (**C**) Percentage of OTUs identified by each library.

reads were excluded from downstream analysis. Alternatively, metagenomic assembly analysis can be performed to obtain longer contiguous sequences of the genomes of the organisms present in the eDNA samples. However, the de novo assembly of longer stretches of contiguous sequences could mainly result from the formation of chimeric contigs that contain a mixture of sequences from multiple genomes, as reported in our previous study (Jo et al. 2019). Thus, to obtain a higher proportion of overlapped (merged) sequences, two libraries with insertion sizes of 350 and 450 bp were constructed and sequenced with paired-end (2 × 301 bp) reads. As theoretically expected, 80% of the merged sequences were 350 bp, and between 50% and 80% of the merged sequences were 450 bp in size (Fig. 6A). All three library sizes generated merged sequences, mainly in the fragment size range of 400–499 bp (Fig. 6B). In the quantitative comparisons of different library sizes, the 350 bp library insert size produced the highest percentage of species-level OTUs (60%) (Fig. 6C). All these results suggest that a library insert size of 350 bp would be optimal for the SoEM v1.0 method. Again, all NGS experiments were run on an Illumina MiSeq with a library insert size of 350 bp.

The eDNA-based SoEM method can estimate the species richness in a given ecosystem with a high taxonomic resolution and monitor the temporal and spatial dynamics of plankton communities from the local to regional or global scale. In addition, this method can detect and anticipate anthropogenic impacts on plankton communities by assessing their functional trait diversity. Recognizing the power and limitations of the eDNA-based SoEM method will help ensure that the ongoing programs carried out at various monitoring sites will be useful for the conservation of marine ecosystems and the sustainable use of biodiversity resources. These can potentially contribute to achieving the targets of several international biodiversity commitments (e.g., the CBD Aichi Targets, SDG 14 by 2030).

There are still some caveats that need further attention for the SoEM v1.0 method to achieve better applicability. Independently of this study, we collected and analyzed several distinct seawater samples in the Korea Strait (Kim et al. 2019). Seawater was collected in volumes ranging from 2 to 100 liters for each sample, and the SoEM v1.0 protocol was used to extract the eDNA. However, some samples failed to yield the required amount of DNA for NGS analysis, probably due to a lack of detectable marine organisms in the sample volume. Other factors, such as the phytoplankton biomass, can be considered to figure out the best sample volume for the SoEM v1.0 method. Currently, we use a sample volume of 2 L when the Chl *a* concentration is between 3.61 and 8.28 $\mu g\ L^{-1}$. Another problem with eDNA-based biodiversity monitoring is the detection of dead eDNA. Since eDNA-based taxonomic identification relies on DNA itself, not on expressed gene products, it is challenging to differentiate between the

intracellular DNA from living cells and extracellular DNA shed by dead or living organisms. This theoretical possibility may increase the risk of false-positive results. Recent paleogenetic and forensic analyses (Cristescu 2019) are beginning to explore the use of environmental RNA (eRNA), a fragile molecule that degrades in vitro, for species identification; therefore, whether eRNA can capture biodiversity information should be further explored.

In this study, we investigated three categories of improvements to expand the usage of the SoEM method: (1) applying various sample volumes from mesotrophic coastal waters to capture targeted species or lineages of marine ecological and commercial interest; (2) implementing it for eDNA samples from different marine trophic levels and states; and (3) developing an easy-to-use web-based application for analyzing and visualizing the SoEM data.

References

Akita, S., H. Murasawa, M. Kondo, Y. Takano, Y. Kawakami, S. Nagai, and D. Fujita. 2020. DNA metabarcoding analysis of macroalgal seed banks on shell surface of the limpet *Niveotectura pallida*. Eur. J. Phycol. **55**: 467–477. doi:10.1080/09670262.2020.1750056

Bik, H. M., D. Fournier, W. Sung, R. D. Bergeron, and W. K. Thomas. 2013. Intra-genomic variation in the ribosomal repeats of nematodes. PLoS One **8**: e78230. doi:10.1371/journal.pone.0078230

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120. doi:10.1093/bioinformatics/btu170

Bovo, S., V. J. Utzeri, A. Ribani, R. Cabbri, and L. Fontanesi. 2020. Shotgun sequencing of honey DNA can describe honey bee derived environmental signatures and the honey bee hologenome complexity. Sci. Rep. **10**: 9279. doi:10.1038/s41598-020-66127-1

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: Architecture and applications. BMC Bioinform. **10**: 421. doi:10.1186/1471-2105-10-421

Canonico, G., and others. 2019. Global observational needs and resources for marine biodiversity. Front. Mar. Sci. **6**: 367. doi:10.3389/fmars.2019.00367

Cardinale, B. J., and others. 2012. Biodiversity loss and its impact on humanity. Nature **486**: 59–67. doi:10.1038/nature11148

Chavez, F. P., J. Ryan, S. E. lluch-Cota, and C. M. Ñiquen. 2003. From anchovies to sardines and back: Multidecadal change in the Pacific Ocean. Science **299**: 217–221. doi:10.1126/science.1075880

Crampton-Platt, A., M. J. T. N. Timmermans, M. L. Gimmel, S. N. Kutty, T. D. Cockerill, C. V. Khen, and A. P. Vogler. 2015. Soup to tree: The phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. Mol. Biol. Evol. **32**: 2302–2316. doi:10.1093/molbev/msv111

Crampton-Platt, A., D. W. Yu, X. Zhou, and A. P. Vogler. 2016. Mitochondrial metagenomics: Letting the genes out of the bottle. Gigascience **5**: 15. doi:10.1186/s13742-016-0120-y

Cristescu, M. E. 2019. Can environmental RNA revolutionize biodiversity science? Trends Ecol. Evol. **34**: 694–697. doi:10.1016/j.tree.2019.05.003

Crowe, T. P., and others. 2013. Large-scale variation in combined impacts of canopy loss and disturbance on community structure and ecosystem functioning. PLoS One **8**: e66238. doi:10.1371/journal.pone.0066238

Díaz, S., and others. 2019. Pervasive human-driven decline of life on Earthpoints to the need for transformative change. Science **366**: eaax3100. doi:10.1126/science.aax3100

Djurhuus, A., C. J. Closek, R. P. Kelly, K. J. Pitz, R. P. Michisaki, H. A. Starks, K. R. Walz, E. A. Andruszkiewicz, E. Olesin, K. Hubbard, and E. Montes. 2020. Environmental DNA reveals seasonal shifts and potential interactions in a marine community. Nat. Commun. **11**: 254. doi:10.1038/s41467-019-14105-1

Dodsworth, S. 2015. Genome skimming for next-generation biodiversity analysis. Trends Plant Sci. **20**: 525–527. doi:10.1016/j.tplants.2015.06.012

Drummond, A. J., and others. 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. Gigascience **4**: 46. doi:10.1186/s13742-015-0086-1

Fernandes, P. G., and others. 2017. Coherent assessments of Europe's marine fishes show regional divergence and megafauna loss. Nat. Ecol. Evol. **1**: 1–9. doi:10.1038/s41559-017-0170

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. **3**: 294–299.

Greshake, B., S. Zehr, F. Dal Grande, A. Meiser, I. Schmitt, and I. Ebersberger. 2016. Potential and pitfalls of eukaryotic metagenome skimming: A test case for lichens. Mol. Ecol. Resour. **16**: 511–523. doi:10.1111/1755-0998.12463

Hillebrand, H., and others. 2018. Biodiversity change is uncoupled from species richness trends: Consequences for conservation and monitoring. J. Appl. Ecol. **55**: 169–184. doi:10.1111/1365-2664.12959

Jo, J., H.-G. Lee, K. Y. Kim, and C. Park. 2019. SoEM: A novel PCR-free biodiversity assessment method based on small-organelles enriched metagenomics. Algae **34**: 57–70. doi:10.4490/algae.2019.34.2.26

Katsuoka, F., J. Yokozawa, K. Tsuda, S. Ito, X. Pan, M. Nagasaki, J. Yasuda, and M. Yamamoto. 2014. An efficient quantitation method of next-generation sequencing libraries by using MiSeq sequencer. Anal. Biochem. **466**: 27–29. doi:10.1016/j.ab.2014.08.015

Kechin, A., U. Boyarskikh, A. Kel, and M. Filipenko. 2017. cutPrimers: A new tool for accurate cutting of primers from

reads of targeted next generation sequencing. J. Comput. Biol. **24**: 1138–1143. doi:10.1089/cmb.2017.0096

Kim, H. M., J. Jo, C. Park, B. J. Choi, H. G. Lee, and K. Y. Kim. 2019. Epibionts associated with floating *Sargassum horneri* in the Korea Strait. Algae **34**: 303–313. doi:10.4490/algae.2019.34.12.10

Krehenwinkel, H., M. Wolf, J. Y. Lim, A. J. Rominger, W. B. Simison, and R. G. Gillespie. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Sci. Rep. **7**: 17668. doi:10.1038/s41598-017-17333-x

Kress, W. J., C. Garcia-Robledo, M. Uriarte, and D. L. Erickson. 2015. DNA barcodes for ecology, evolution, and conservation. Trends Ecol. Evol. **30**: 25–35. doi:10.1016/j.tree.2014.10.008

Leray, M., J. Y. Yang, C. P. Meyer, S. C. Mills, N. Agudelo, V. Ranwez, J. T. Boehm, and R. J. Machida. 2013. 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. **10**: 34. doi:10.1186/1742-9994-10-34

Linard, B., A. Crampton-Platt, C. P. Gillett, M. J. Timmermans, and A. P. Vogler. 2015. Metagenome skimming of insect specimen pools: Potential for comparative genomics. Genome Biol. Evol. **7**: 1474–1489. doi:10.1093/gbe/evv086

Liu, S., and others. 2016. Mitochondrial capture enriches Mito-DNA 100-fold, enabling PCR-free mitogenomics biodiversity analysis. Mol. Ecol. Resour. **16**: 470–479. doi:10.1111/1755-0998.12472

Magoc, T., and S. L. Salzberg. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics **27**: 2957–2963. doi:10.1093/bioinformatics/btr507

Magris, R. A., M. Marta-Almeida, J. A. F. Monteiro, and N. C. Ban. 2019. A modelling approach to assess the impact of land mining on marine biodiversity: Assessment in coastal catchments experiencing catastrophic events (SW Brazil). Sci. Total Environ. **659**: 828–840. doi:10.1016/j.scitotenv.2018.12.238

McQuatters-Gollop, A., and others. 2019. Plankton lifeforms as a biodiversity indicator for regional-scale assessment of pelagic habitats for policy. Ecol. Indic. **101**: 913–925. doi:10.1016/j.ecolind.2019.02.010

Miloslavich, P., N. J. Bax, S. E. Simmons, E. Klein, W. Appeltans, O. Aburto-Oropeza, M. Andersen Garcia, S. D. Batten, L. Benedetti-Cecchi, D. M. Checkley Jr, and S. Chiba. 2018. Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. Glob. Chang. Biol. **24**: 2416–2433. doi:10.1111/gcb.14108

Parducci, L., and others. 2019. Shotgun environmental DNA, pollen, and macrofossil analysis of Lateglacial Lake sediments from southern Sweden. Front. Ecol. Evol. **7**: 189. doi:10.3389/fevo.2019.00189

Pawluczyk, M., J. Weiss, M. G. Links, M. Egana Aranguren, M. D. Wilkinson, and M. Egea-Cortines. 2015. Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. Anal. Bioanal. Chem. **407**: 1841–1848. doi:10.1007/s00216-014-8435-y

Porazinska, D. L., and others. 2009. Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. Mol. Ecol. Resour. **9**: 1439–1450. doi:10.1111/j.1755-0998.2009.02611.x

Rogers, A., J. L. Blanchard, and P. J. Mumby. 2018. Fisheries productivity under progressive coral reef degradation. J. Appl. Ecol. **55**: 1041–1049. doi:10.1111/1365-2664.13051

Roxburgh, S. H., K. Shea, and B. Wilson. 2004. The intermediate disturbance hypothesis: Patch dynamics and mechanisms of species coexistence. Ecology **85**: 359–371. doi:10.1890/03-0266

Ruppert, K. M., R. J. Kline, and M. S. Rahman. 2019. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecol. Conserv. **17**: e00547. doi:10.1016/j.gecco.2019.e00547

Saunders, G. W., and D. C. McDevit. 2012. Methods for DNA barcoding photosynthetic protists emphasizing the macroalgae and diatoms. Methods Mol. Biol. **858**: 207–222. doi:10.1007/978-1-61779-591-6_10

Sharpton, T. J. 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci. **5**: 209. doi:10.3389/fpls.2014.00209

Shendure, J., S. Balasubramanian, G. M. Church, W. Gilbert, J. Rogers, J. A. Schloss, and R. H. Waterston. 2017. DNA sequencing at 40: Past, present and future. Nature **550**: 345–353. doi:10.1038/nature24286

Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc. Nat. Acad. Sci. U.S.A. **103**: 12115–12120. doi:10.1073/pnas.0605127103

Stoeck, T., D. Bass, M. Nebel, R. Christen, M. D. Jones, H. W. Breiner, and T. A. Richards. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol. Ecol. **19**: 21–31. doi:10.1111/j.1365-294X.2009.04480.x

Stat, M., J. John, J. D. DiBattista, S. J. Newman, M. Bunce, and E. S. Harvey. 2019. Combined use of eDNA metabarcoding and video surveillance for the assessment of fish biodiversity. Conserv. Biol. **33**: 196–205. doi:10.1111/cobi.13183

Suter, L., A. M. Polanowski, L. J. Clarke, J. A. Kitchener, and B. E. Deagle. 2021. Capturing open ocean biodiversity: Comparing environmental DNA metabarcoding to the continuous plankton recorder. Mol. Ecol. **30**: 3140–3157. doi:10.1111/mec.15587

Taberlet, P., E. Coissac, F. Pompanon, C. Brochmann, and E. Willerslev. 2012. Towards next-generation biodiversity

assessment using DNA metabarcoding. Mol. Ecol. **21**: 2045–2050. doi:10.1111/j.1365-294X.2012.05470.x

Tamura, K., and T. Aotsuka. 1988. Rapid isolation method of animal mitochondrial DNA by the alkaline lysis procedure. Biochem. Genet. **26**: 815–819. doi:10.1007/BF02395525

Tang, M., and others. 2014. Multiplex sequencing of pooled mitochondrial genomes—A crucial step toward biodiversity analysis using mito-metagenomics. Nucleic Acids Res. **42**: e166. doi:10.1093/nar/gku917

Tyson, G. W., and others. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428**: 37–43. doi:10.1038/nature02340

Ulman, A., J. Ferrario, A. Forcada, H. Seebens, C. Arvanitidis, A. Occhipinti-Ambrogi, A. Marchini, and N. O'Connor. 2019. Alien species spreading via biofouling on recreational vessels in the Mediterranean Sea. J. Appl. Ecol. **56**: 2620–2629. doi:10.1111/1365-2664.13502

van der Plas, F. 2019. Biodiversity and ecosystem functioning in naturally assembled communities. Biol. Rev. Camb. Philos. Soc. **94**: 1220–1245. doi:10.1111/brv.12499

Worden, A. Z., M. J. Follows, S. J. Giovannoni, S. Wilken, A. E. Zimmerman, and P. J. Keeling. 2015. Environmental science. Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. Science **347**: 1257594. doi:10.1126/science.1257594

Xue, Y., H. Chen, J. R. Yang, M. Liu, B. Huang, and J. Yang. 2018. Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. ISME J. **12**: 2263–2277. doi:10.1038/s41396-018-0159-0

Zhang, G. K., F. J. J. Chain, C. L. Abbott, and M. E. Cristescu. 2018. Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evol. Appl. **11**: 1901–1914. doi:10.1111/eva.12694

Zhou, X., and others. 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. Gigascience **2**: 2047-2217X-2042-2044. doi:10.1186/2047-217X-2-4

Acknowledgments

This research was supported by National Research Foundation (NRF) grants (NRF-2016R1A6A1A03012647, NRF-2020R1A2C3005053, NRF-2022M3I6A1085991) to K.Y.K. and C.P. funded by the Korean government (MSIT).

Submitted 28 December 2022
Accepted 30 January 2023

Associate editor: Steven W. Wilhelm