Metabarcoding of marine zooplankton: prospects, progress and pitfalls

INTRODUCTION

Marine zooplankton are rapid-responders to environmental variation associated with regime shifts and climate change, which may cause significant and potentially accelerating losses in species diversity (Beaugrand et al., 2010; Möllmann and Diekmann, 2012). However, the systematic complexity of the zooplankton assemblage, with numerous cryptic and sibling species, and the lack of
diagnostic characters for immature (larval) stages constitute important impediments to our understanding of global-to-local patterns of biodiversity and biogeography. In recent years, high-throughput DNA sequencing (HTS) technologies have resulted in dramatic advances in practical, cost-effective molecular approaches to the analysis of environmental samples (Bourlat et al., 2013; Ji et al., 2013). In particular, metabarcoding (i.e. the large-scale taxonomic identification of a complex sample via analysis of one or few orthologous DNA regions, called barcodes) has the significant advantage of detecting the ‘hidden diversity’ of zooplankton assemblages, including mero-, holo- and ichthyoplankton (Lindeque et al., 2013). Metabarcoding of the pelagic assemblage is yielding new insights into marine biodiversity, as most marine species (including fish, macro-invertebrates etc.) are planktonic at some point in their life cycle.

The metabarcoding approach involves a variety of laboratory and data analysis steps upon which biodiversity estimates rely (Fig. 1). First, the DNA present in the sample (which can be intracellular or extracellular) is extracted from the whole sample; second, the barcode of choice is amplified using consensus or taxonomic group-specific PCR primers and appropriate amplification conditions; third, the PCR products are sequenced on an HTS platform (each nucleotide sequence is called a ‘read’); and finally, the obtained sequences are processed for quality control, grouped in molecular operational taxonomic units (MOTU), and compared with a reference database for taxonomic assignment. The metabarcoding approach, including these steps and their implications for biodiversity assessments of multicellular zooplankton, is the subject of this article.

**CHOICE OF BARCODE**

Metabarcoding studies of zooplankton assemblages have used a number of marker gene regions to characterize biodiversity patterns across different systematic levels and to address specific hypotheses. To date, the most frequently used gene regions are portions of the nuclear small-subunit ribosomal RNA gene (18S rRNA), which shows consistent patterns of divergence across invertebrate and vertebrate taxa, and discriminates genera, families and higher taxonomic groups (Mallatt et al., 2004). The V9 hypervariable region of 18S rRNA was developed as a standard marker of marine microbial eukaryotic diversity (Amaral-Zettler et al., 2009) and has also been used for analysis of zooplankton assemblages (Pearman et al., 2014; De Vargas et al., 2015; Pearman and Irigoien, 2015; Albaina et al., 2016). Several additional hypervariable regions of the 18S rRNA gene have been used for zooplankton metabarcoding studies, including V1–V2 (Lindeque et al., 2013); V4 (Sun et al., 2015) and V7–V9 (Hirai et al., 2015b). These studies allow examination of impacts of both different gene regions and also different sequence lengths, with the latter determined largely by the constraints of the HTS platform and associated protocols. Zooplankton metabarcoding studies have also used portions of the nuclear large-subunit 28S rRNA (Hirai et al., 2013, 2015a; Hirai and Tsuda, 2015). These nuclear rRNA regions have allowed accurate classification of novel sequences and reliable amplification with consensus primers, but due to the relatively conserved nature of this gene may underestimate the diversity of species in a community (Tang et al., 2012).

Analysis of species-level diversity and distribution of marine zooplankton using metabarcoding approaches will require development of reliable HTS protocols for more variable genes. The mitochondrial cytochrome oxidase I (COI) barcode region (Hebert et al., 2003) is one of the most commonly sequenced regions for analysis of species diversity among marine animals (Bucklin et al., 2011), including zooplankton (Bucklin et al., 2010a, b). A limited number of metabarcoding studies of zooplankton species biodiversity based on COI sequences has been published (Machida et al., 2009; Bourlat et al., 2013; Zaiko et al., 2015b). Another mitochondrial gene frequently used for identification and discrimination of zooplankton species is mitochondrial 16S rRNA (Lindeque et al., 1999, 2006; Goetze, 2010), which is considered by some to be a more reliable marker, especially for cnidarians (Zheng et al., 2014; Lindsay et al., 2015). These studies confirm the usefulness of both mitochondrial genes for accurate species identification and discrimination, but also clearly demonstrate that metabarcoding approaches for detection of species-level diversity face significant technical challenges, including the need for cocktails of group-specific primers (Bucklin et al., 2010b), and consequent inconsistent amplification success rates among the various taxonomic groups of marine zooplankton.

**REFERENCE BARCODE DATABASE FOR SPECIES IDENTIFICATION**

Metabarcoding-based species identification requires taxonomically complete and geographically comprehensive reference databases of DNA sequences for each species and for all gene regions. Absolutely essential is that reference specimens are accurately identified to species by a recognized expert taxonomist; inaccurate and incomplete identifications remain a persistent impediment to the use of metabarcoding for analysis of species-level zooplankton
biodiversity. A reference database can also serve as a valuable resource for researchers to confirm species identifications when morphological taxonomic expertise is limited and when training new taxonomists.

Reference databases are growing for several of the gene regions most frequently used for zooplankton metabarcoding. Most notable are the comprehensive SILVA databases (http://www.arb-silva.de/) with aligned small- (18S) and large-subunit (28S) rRNA sequences for all three domains of life (Quast et al., 2013). Also noteworthy is the archive of COI sequences for thousands of species of marine animals now available in public repositories, including the GenBank Barcode of Life section (http://www.ncbi.nlm.nih.gov/genbank/barcode) and the Barcode of Life Database.

Fig. 1. Schematic representation of the taxonomic analysis of a zooplankton sample using morphological identification (left arrows), barcoding (middle arrows) or metabarcoding (right arrows). Figure modified from Corell and Rodriguez-Ezpeleta (Corell and Rodriguez-Ezpeleta, 2014).
Clustered into MOTUs (Floyd et al. 2012). To make sense of the data, the sequences are usually large-scale biodiversity analyses of samples (Shokralla et al. 2010). DNA sequences in parallel, making them ideally suited for HTS technologies provide the ability to read millions of sequence reads that can be clustered into MOTUs and species richness (Lindeque et al., 2015). Analysis of zooplankton assemblages with metabarcoding has revealed previously hidden taxonomic richness, especially for hard-to-identify metazooplankton (e.g. bivalves, gastropods, polychaetes), rare species and parasites in comparison with morphological analysis (Lindeque et al., 2013). Recent metabarcoding studies, such as the TARA oceans expedition (de Vargas et al., 2015) with huge spatial-scale sampling, great depth of sequencing and comprehensive taxonomic analysis, revealed that MOTU diversity is likely to be much higher than described species of marine eukaryotic plankton, especially in the smaller organismal size fraction. As metabarcoding progresses, continued support for traditional morphological analysis will remain critically important to allow direct comparison between morphological and molecular approaches and to gain better understanding of how barcode choice, length and intra- and inter-specific variation influence the similarity threshold. Such studies will ensure that metagenetic assessment of diversity and species richness is both realistic and relevant.

COMPARISONS OF METABARCODING AND MORPHOLOGICAL MEASURES OF ZOOPLANKTON BIODIVERSITY

With the potential for newly emerging metabarcoding analyses to overtake, and perhaps be used as an alternative to morphological analysis in measuring zooplankton diversity, it is critical to compare and contrast the two approaches. While metabarcoding can provide a broad assessment of zooplankton diversity and taxon richness, we should not oversell this relatively new technique in its current state, nor be hasty in replacing morphological analysis, since both techniques have their costs and benefits.

Morphological analysis of zooplankton results in numerical abundance, with a possibility to convert to biomass. HTS technologies provide the ability to read millions of DNA sequences in parallel, making them ideally suited for large-scale biodiversity analyses of samples (Shokralla et al., 2012). To make sense of the data, the sequences are usually clustered into MOTUs (Floyd et al., 2002), based on a similarity threshold (Fonseca et al., 2010). When defining the similarity threshold, both barcode length and intra- and inter-specific degree of conservation need to be taken into account (Lindeque et al., 2013; Brown et al., 2015; Hirai et al., 2015a). Care must be taken in construction of the MOTUs to obtain the most relevant and realistic assessment of diversity and species richness. If the similarity threshold used to cluster sequences into MOTUs is too high, this may lead to an overestimation of taxon richness; however, if the threshold is too low, it is likely that the taxon richness will not align to species richness. Thus, a critical challenge in using metabarcoding to estimate biodiversity is to examine the relationship between MOTU number and species richness (Carugati et al., 2015).

Taxonomic identification of MOTUs can best be made by comparing a representative sequence from each MOTU against a nucleotide database. However, annotation must only be made against well-populated reference libraries based on correctly-identified specimens (see above). In summary, metabarcoding provides a number of sequence reads that can be clustered into MOTUs and with care and consideration can provide estimations of taxon richness that may approximate species richness.

While some pitfalls of metabarcoding, such as lack of identification of individual life stages (since DNA sequences will be identical for eggs, larvae, adults or other developmental stages of a given species), will be nearly impossible to overcome, metabarcoding does present promises above those of traditional morphological methods.

Metabarcoding can discriminate spatial and temporal patterns of variability in planktonic assemblages (Eiler et al., 2013; Massana et al., 2015). Analysis of zooplankton assemblages with metabarcoding has revealed previously hidden taxonomic richness, especially for hard-to-identify metazooplankton (e.g. bivalves, gastropods, polychaetes), rare species and parasites in comparison with morphological analysis (Lindeque et al., 2013). Recent metabarcoding studies, such as the TARA oceans expedition (de Vargas et al., 2015) with huge spatial-scale sampling, great depth of sequencing and comprehensive taxonomic analysis, revealed that MOTU diversity is likely to be much higher than described species of marine eukaryotic plankton, especially in the smaller organismal size fraction. As metabarcoding progresses, continued support for traditional morphological analysis will remain critically important to allow direct comparison between morphological and molecular approaches and to gain better understanding of how barcode choice, length and intra- and inter-specific variation influence the similarity threshold. Such studies will ensure that metagenetic assessment of diversity and species richness is both realistic and relevant.

CHALLENGES OF QUANTIFICATION OF TAXON ABUNDANCE OR BIOMASS USING METABARCODING

The sensitivity of metabarcoding analysis to detect and discriminate rare and cryptic species has been widely reported, including for zooplankton communities (Zhan and MacIsaac, 2015). Although presence/absence is critical for biodiversity monitoring, quantification of taxon abundance or relative abundance is needed for community characterization and for the assessment of many biological indices (Bourlat et al., 2013; Aylagas et al., 2014). Quantification of relative abundances of taxa above the species level has been shown to match morphological analyses for some zooplankton groups and samples (Lindeque et al., 2013; Hirai et al., 2015b), but metabarcoding analysis has not shown good agreement with species abundance data from morphological taxonomic analysis (Mohrbeck et al., 2015). In one study (Sun et al., 2015), a general trend was detected that low-abundance species usually corresponded to low-abundance sequence reads; however, the authors urged caution when using HTS-based approaches to make quantitative inferences.

The number of sequencing reads associated with a MOTU can approximate to biomass (Lindeque et al., 2013) and has been shown to correlate with dry weight of the taxon (Hirai et al., 2015b). It is likely that the correlation between biomass and sequencing reads is not linear and is affected by various biases introduced at different stages, e.g. DNA extraction, PCR amplification, DNA pooling and
bioinformatics sorting (Bik et al., 2012). Among possible approaches to address such biases are comparative analysis of RNA (which ensures detection only of living organisms) and approaches that do not require initial PCR amplification (Dowle et al., 2015). An important challenge for quantification using metabarcoding is that multicopy genes, such as ribosomal and mitochondrial genes, vary in copy number across different animal taxa (Prokopowich et al., 2003). It may eventually be possible to calibrate bias due to gene copy number variation (CNV) using low-diversity and/or mock samples, perhaps using quantitative PCR (qPCR) to determine gene copy numbers (Amend et al., 2010). Until these issues are resolved, metabarcoding will remain a semi-quantitative method for biodiversity analysis.

APPLICATIONS OF METABARCODING FOR ECO SYSTEM MONITORING AND MANAGEMENT

Potential applications of metabarcoding in marine monitoring include calculation of biotic indices based on taxonomic composition, characterization of trophic interactions and food web structure and detection of non-indigenous species (NIS). Although applications of metabarcoding for environmental monitoring seem quite straightforward and work in theory, the routine implementation of this approach still requires the development of standardized practices at each step of the procedure (Aylagas et al., 2014). Because some marine monitoring-related indices, e.g. AZTI’s Marine Biotic Index (Aylagas et al., 2014), rely on the presence of taxa that are either sensitive to or tolerant of pollution, the ability to detect all organisms present in the sample is crucial. Careful evaluation of the accuracy of the taxonomic composition inferred from metabarcoding is necessary before this method can be regularly applied to assess ecosystem status.

Another promising frontier for metabarcoding is the analysis of environmental DNA (eDNA) or free DNA molecules that are present outside of organisms (Bohmann et al., 2014). The use of eDNA to detect biodiversity shifts has been an active area of research (Lodge et al., 2012; Kelly et al., 2014; Kelly, 2016). Metabarcoding is also emerging as an invaluable tool in the examination of trophic relationships, through HTS analysis of gut contents and fecal material (Deagle et al., 2013; Albaina et al., 2016). Both these applications require specific examination of the impact of degraded DNA on the accuracy and reliability of the analyses, since amplicon size and copy number will impact metabarcoding uncertainty biases and will need to be addressed prior to use in ecological monitoring.

Metabarcoding may also substantially improve capabilities for accurate identification and early detection of introduced NIS (Mountfort et al., 2012; Kelly et al., 2014; Zaiko et al., 2015b). Such early-warning will provide managers with options to act before a harmful species can achieve high abundance (Robinson et al., 2011). Shipping, in particular transported ballast water, is considered to be one of the most important pathways of marine biological invasions worldwide (Molnar et al., 2008), yet traditional sampling does not always capture all organisms, especially at the early phase of invasion (Lehtiniemi et al., 2015). Previous studies have shown that estimates of taxon-specific DNA concentrations determined using qPCR correlate positively with abundance estimates of that taxon (Thomsen et al., 2012). Clearly, accurate detection and quantification are important for managers to determine the phase of the invasion and necessary approaches for eradication.

A promising use of metabarcoding for management of NIS concerns Ballast Water Management Convention compliance control. Metabarcoding can be used to verify positive results in control surveys (i.e. zero counts of organisms by microscopic analyses), because some species can be overlooked by conventional analysis. Metabarcoding, especially when used in combination with morphological analyses (Zaiko et al., 2015a), is a powerful new tool for NIS monitoring and management.

SUMMARY AND RECOMMENDATIONS

Metabarcoding is revolutionizing the analysis of marine biodiversity and has a significant advantage of detecting the ‘hidden diversity’ of marine zooplankton (Lindeque et al., 2013). Emerging results indicate that estimates of global zooplankton diversity will markedly increase with more accurate definition and higher resolution of time/space patterns made possible by metabarcoding. In addition, metabarcoding will allow for rapid detection and description of the impacts of climate change on biodiversity and biogeography. As HTS becomes more accessible and less expensive, the use of metabarcoding will expand into numerous applications in ocean monitoring and management, including calculation of biotic indices, trophic interactions and food web analysis and detection of introduced NIS. Among the challenges remaining for reliable and routine application of metabarcoding for analysis of zooplankton are evaluation and comparison of results using various barcode gene regions (as well as different primers and protocols); development of methodologies using more variable gene regions that can ensure identification, discrimination and detection of closely related, cryptic and rare species; impacts of degraded DNA (e.g. environmental DNA and DNA recovered from gut contents) and continued development of taxonomically comprehensive reference databases for all gene
regions. A particular need is to move metabarcoding applications from identification and detection of taxa to
t heir quantification in terms of abundance and/or biomass, which will require concerted effort to address
biases associated with gene CNV. Despite the remarkable promise of metabarcoding in yielding new understanding
and appreciation for global patterns of zooplankton diversity, it is critically important to maintain expertise and
capacity in morphological taxonomic identification of zooplankton to ensure that metabarcoding approaches can be validated. Such integrative morphological and molecular taxonomic approaches will provide the neces-
sary foundation and future of research, monitoring and management of the pelagic realm.

ACKNOWLEDGMENTS

This manuscript results from discussions of the International Council for the Exploration of the Sea (ICES) Working Group on Integrated Morphological and Molecular Taxonomy (WGIMT; see http://www.ices.dk/community/groups/Pages/WGIMTaspx and http://wgimt.net/). The authors gratefully acknowledge all members of WGIMT for their contribution to the development of the ideas presented here. This paper benefitted from collaboration with ICES WGITMO and WGBOSV. This is contribution number 748 of the Marine Research Division of AZTI.

FUNDING

We acknowledge support from the International Council for the Exploration of the Sea (ICES) for WGIMT, as well as conferences and symposia related to the Working Group’s mission. ICES Delegates from member countries provided financial support to some WGIMT members for travel to annual meetings. Additional support was provided by: NOAA Office of Ocean Exploration and Research (Grant No. NA15OAR00110209 to AB, University of Connecticut, USA); UK Natural Environment Research Council and Department for Environment, Food and Rural Affairs (Grant No. NE/L003279/1 to PKL, Plymouth Marine Laboratory, UK); DEVOTES (DEVelopment Of innovative Tools for understanding marine biodiversity and assessing Good Environmental Status) project funded by the European Union (EU) 7th Framework Program (Grant No. 308392 to NR-E, AZTI, Spain); Basque Government (Grant No. IT558-10 to AA, University of the Basque Country, UPV/EHU, Spain); and BONUS BIO-C3 project and BONUS (Art 185), funded by the EU and Finnish Academy (Grant No. 273660 to ML, Finnish Environment Institute, Finland).

REFERENCES


Sun, C., Zhao, Y., Li, H., Dong, Y., MacIsaac, H. J. and Zhan, A. (2015) Unreliable quantification of species abundance based on high-


