



CrossMark  
click for updates

**Cite this article:** Leray M, Knowlton N. 2016  
Censusing marine eukaryotic diversity in the  
twenty-first century. *Phil. Trans. R. Soc. B* **371**:  
20150331.  
<http://dx.doi.org/10.1098/rstb.2015.0331>

Accepted: 26 May 2016

One contribution of 16 to a theme issue  
'From DNA barcodes to biomes'.

**Subject Areas:**

ecology, genetics, taxonomy and systematics

**Keywords:**

meiofauna, protist, plankton,  
metabarcoding, 18S

**Author for correspondence:**

Matthieu Leray  
e-mail: [leray.upmc@gmail.com](mailto:leray.upmc@gmail.com)

# Censusing marine eukaryotic diversity in the twenty-first century

Matthieu Leray and Nancy Knowlton

National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA

ML, 0000-0002-7327-1878

The ocean constitutes one of the vastest and richest biomes on our planet. Most recent estimations, all based on indirect approaches, suggest that there are millions of marine eukaryotic species. Moreover, a large majority of these are small (less than 1 mm), cryptic and still unknown to science. However, this knowledge gap, caused by the lack of diagnostic morphological features in small organisms and the limited sampling of the global ocean, is currently being filled, thanks to new DNA-based approaches. The molecular technique of PCR amplification of homologous gene regions combined with high-throughput sequencing, routinely used to census unculturable prokaryotes, is now also being used to characterize whole communities of marine eukaryotes. Here, we review how this methodological advancement has helped to better quantify the magnitude and patterns of marine eukaryotic diversity, with an emphasis on taxonomic groups previously largely overlooked. We then discuss obstacles remaining to achieve a global understanding of marine eukaryotic diversity. In particular, we argue that 18S variable regions do not provide sufficient taxonomic resolution to census marine life, and suggest combining broad eukaryotic surveys targeting the 18S rRNA region with more taxon-focused analyses of hypervariable regions to improve our understanding of the diversity of species, the functional units of marine ecosystems.

This article is part of the themed issue 'From DNA barcodes to biomes'.

## 1. State of knowledge based on traditional taxonomy

Human activities are having an increasing effect on ocean biodiversity, although it remains the case that relatively few marine species are reported to have gone extinct [1]. However, impact assessments and monitoring initiatives are based almost exclusively on large and conspicuous species that represent a minor fraction of marine diversity. Small (less than 2 mm) and cryptic organisms, which play important ecological roles despite being inconspicuous, remain overlooked in biodiversity surveys [2]. This highlights a major limitation in our ability to monitor biological communities: How can we establish biological baselines, quantify changes in biodiversity over time and understand the consequences of community shifts on ecosystem services if most species are unknown to science or cannot be easily surveyed?

Estimates of the number of existing marine species, all based on indirect approaches (e.g. experts' opinion [3] and extrapolations from rates of species description [4], rates of discovery of higher taxa [5] or known areas or faunas [6]), range from 0.3 to more than 10 million species. According to the World Register of Marine Species (WORMS, [7]), there were 228 739 accepted eukaryotic marine species as of September 2015 (of which Animalia constituted 195 702 species, Plantae 9689 species, Chromista 21 403 species, Protozoa 589 species and Fungi 1356 species). This suggests that between 24 and 98% of all marine eukaryotic species remain to be described, with the proportion of unknown diversity varying greatly among groups. Taxonomic experts have estimated that fewer than 10% of species might be formally described in the most cryptic taxonomic groups (e.g. isopods, micro-gastropods, nematodes, copepods and some Chromista [3]). Even among well-known groups such as marine mammals, new species continue to be

discovered [8]. For many groups, the absence of diagnostic morphological characters (e.g. nematodes [9]), the lack of taxonomic expertise and the time required to describe or identify species [10] have been major impediments to obtaining a comprehensive understanding of marine diversity. When coupled with limited sampling, particularly in some regions of very high diversity and endemism [11,12], the challenge seems nearly insurmountable.

## 2. Another DNA revolution

Molecular barcodes, typically one or several small stretches of DNA, provide valuable characters to delineate species in all eukaryotic kingdoms of life. As such, standardized DNA regions were identified to supplement morphological identification: the mitochondrial cytochrome *c* oxidase 1 (COI) gene for animals [13], a combination of two chloroplast genes (matK + rbcL) for plants [14], the nuclear ribosomal internal transcribed spacer (ITS) for Fungi [15] and the universal 18S rRNA gene coupled with lineage-specific barcode genes in the highly diversified unicellular protists [16]. DNA barcoding has helped identify unrecognized taxa, and large databases of 'DNA identifiers' are now available in public databases (e.g. Barcode of Life Data Systems, BOLD [17]; the Protist Ribosomal Reference database, PR2 [18]). However, as communities typically comprise numerous small organisms with many rare species [19], Sanger sequence barcoding of individual organisms has proven inefficient for broad diversity surveys and impact assessments because of the time and money required.

In 2005, high-throughput sequencing (HTS) platforms became widely available, a technological revolution that now allows the detection of tens to hundreds of species simultaneously from whole-community samples in a matter of a few days. DNA is extracted from a collection of organisms (e.g. interstitial meiofauna, planktonic organisms or benthic sessile communities) or extracellular DNA (i.e. environmental DNA in water or sediments). Then a small fragment of a DNA marker gene is PCR amplified using general primers, yielding thousands of sequences per sample. DNA sequences are then sorted informatically, low-quality reads and contaminants are removed, and remaining sequences are clustered into molecular operational taxonomic units (OTUs). The resulting data can be compared with reference libraries of DNA barcodes to estimate richness and community composition [20].

This approach, also referred to as metabarcoding or metagenetics, was first applied to study bacterial and archaeal diversity [21], revealing up to 20 000 'species' per litre of seawater [22]. It is also now a cost- and time-effective alternative for eukaryote community profiling [23,24]. In the past 10 years, HTS has been used to study benthic meiofauna diversity in shallow [25–27], deep-sea [28–30] and estuarine [31–33] sediments, macro- and meiofaunal diversity in seagrass beds [34] and oyster reefs [35], as well as planktonic diversity across the globe [36–43], particularly the diversity of picoplanktonic size fractions (less than 3  $\mu\text{m}$ ) [44–46]. The metabarcoding revolution has especially benefited our understanding of microscopic eukaryotic diversity—the unicellular (e.g. protist) or small multicellular (e.g. metazoan less than 1 mm) species that belong to some of the most challenging groups for taxonomists. These tiny organisms are highly abundant in marine environments and have long been recognized as functionally important in both benthic [47] and pelagic systems (e.g. the microbial loop [2]). However, little was known about their

diversity patterns prior to using HTS. HTS has also been used to assess environmental impacts [48], understand trophic interactions [49–51] and track non-indigenous species [52].

## 3. What we have discovered in the ocean using metabarcoding

Studies using a metabarcoding approach have confirmed that diversity of eukaryotic organisms scales up with decreasing body size in both benthic and planktonic systems. Previously it was thought that intermediate-sized organisms held the most diversity [53–55], but earlier data were biased by our inability to distinguish the smallest species from each other using traditional methods. HTS-based studies directly comparing size-fractionated samples, either through sieving or filtration, highlight this new understanding. For example, Leray & Knowlton [35] found that two-thirds of the diversity on oyster reefs was smaller than 500  $\mu\text{m}$  and Logares *et al.* [43] and De Vargas *et al.* [41] found a peak of diversity in the nanoplankton (3–20  $\mu\text{m}$ ) and pico-nanoplankton (0.8–5  $\mu\text{m}$ ), respectively. The smallest size fractions comprise unicellular taxa (protists) that constitute much of the phylogenetic diversity in the domain Eukaryota. The key contributions of these micro-eukaryotes to the nutrient cycling, primary productivity and trophic activity of marine ecosystems have long been recognized [56]. However, their diversity remained one of the least-known prior to DNA metabarcoding because most of these species lack diagnostic morphological features and cannot be cultured. HTS of samples collected directly in the environment has unveiled a staggering number of taxa within most supergroups of protists (e.g. Alveolata [41,57–59], Stramenopila [60,61], Excavata [41], Haptophyta [62], Rhizaria [28,41,63–65], Opisthokonta [66]), increasing known diversity by several orders of magnitude. Moreover, novel clades, some of them considered new kingdoms, keep being discovered (i.e. Picobiliphyta [67], Rappemonada [68]; reviewed in [69]).

Metabarcoding has been a powerful tool for quantifying the relative diversity of various taxonomic groups of microscopic metazoans in benthic systems. Nematodes, for example, are known to be the most abundant metazoans in marine sediments, with up to 20 million individuals per square metre, a density one order of magnitude higher than that of any other taxon [70]. They had also been considered to be one of the most diverse groups, with global estimates converging to one million species [71], in stark contrast with the limited number of marine nematode species that have been described to date (7152, WORMS [7]). Most nematodes lack clear homologous morphological characters, and DNA barcodes are often the only way to delineate species [72,73]. Metabarcoding analyses of shallow and deep-water sediments have confirmed their high contribution to total meiofaunal diversity in marine sediments [25,26,29,30,33,34,74]. However, several of these studies also showed for the first time an equal or higher representation of Platyhelminthes and Annelida at some sites [25,26,29,30,34]. High-energy sediments in intertidal zones are preferred habitats for Platyhelminthes, where they may represent up to 95% of the total meiofaunal biomass [75]. However, because their soft-body structure is altered by preservatives, they can only be studied alive or with light histological techniques [76]. Hence, they have mostly been disregarded in traditional surveys. Interstitial species of Annelida

have been far less studied than their macrofaunal counterparts for similar reasons.

Similar analyses have broadened our understanding of the diversity of different pelagic groups. Because of their abundance, copepods are analogues of nematodes in pelagic systems. They represent important trophic links [77] and as such have been considered an important indicator of environmental changes. However, identifying copepods morphologically, a group with 11 181 described species and many more to describe [3], requires advanced taxonomic knowledge and time. Lower-level identifications are also problematic, with early life stages that are often misidentified or unidentified in monitoring programmes. As expected, the phylum Arthropoda is consistently the dominant metazoan group in metabarcoding analyses of plankton samples, with most OTUs belonging to copepods (i.e. 1009 out of 1554 in [37]; 5766 out of 7744 in [41]). Closely following Arthropoda, the phylum Chordata was particularly OTU-rich in the circum-global analysis by De Vargas *et al.* [41] with 6795 OTUs (40% of total). A great majority of these were assigned to the genus *Doliolum* (6387 OTUs) potentially as a result of the lack of reference sequences among ascidians; however, if this pattern is confirmed (i.e. not the result of sequencing artefacts, pseudogenes or intra-individual polymorphism), this would mean that there are over 92 times more taxa within Ascidiacea than morphologically described species [41]. In the same study, Hydrozoa represented 86% (587 out of 682 OTUs) of the diversity of Cnidaria, with almost half (46%) remaining unassigned at lower taxonomic levels.

In addition to revealing an astounding number of unknown taxa, rapid and cost-effective community profiling of microscopic communities using HTS has provided a better understanding of the ecological factors influencing distribution patterns. Studies of meiofauna have been particularly interesting in this regard. With nearly 60% of all animal phyla having interstitial representatives, meiofaunal communities comprise a diversity of life-history traits and occupy a wide range of ecological niches. With the power of metabarcoding, some have argued that they represent an optimal ecological unit for effective biomonitoring of benthic habitats [33]. Analyses conducted in estuaries demonstrated the response of meiofaunal communities to differences in salinity, sediment particle sizes, oxygen, water flow, nutrient, pH and turbidity [31,33] with groups such as Nematoda and Platyhelminthes responding differently to some abiotic factors [33]. Interstitial metazoans also proved very sensitive to exposure to toxic oil; following the Deepwater Horizon catastrophe in the Gulf of Mexico, there was a pronounced and prolonged reduction of meiofaunal OTU richness, with a shift from metazoan to fungus-dominated interstitial assemblages [48]. Benthic foraminiferans represent another indicator of environmental perturbations because of their sensitivity to abiotic conditions and their short lifespans [78], but traditional surveys used time-consuming morphological identification of shells. Metabarcoding successfully quantified changes in the composition of foraminiferal communities in response to a gradient of sediment oxygenation and nutrient enrichment caused by salmon farming [79,80]. Molecular data also helped measure the impact of fish farms on communities of macro-invertebrates and provided comparable estimates of biotic indices to those obtained using morphological data alone [81], further confirming the promise of these tools for assessing the health status of marine systems [82]. As marine ecosystems are also

increasingly threatened by invasive species, the taxonomic resolution of metabarcoding data may prove to be a powerful early-warning tool for the detection of non-indigenous benthic species at egg or larval stages [42,52,83].

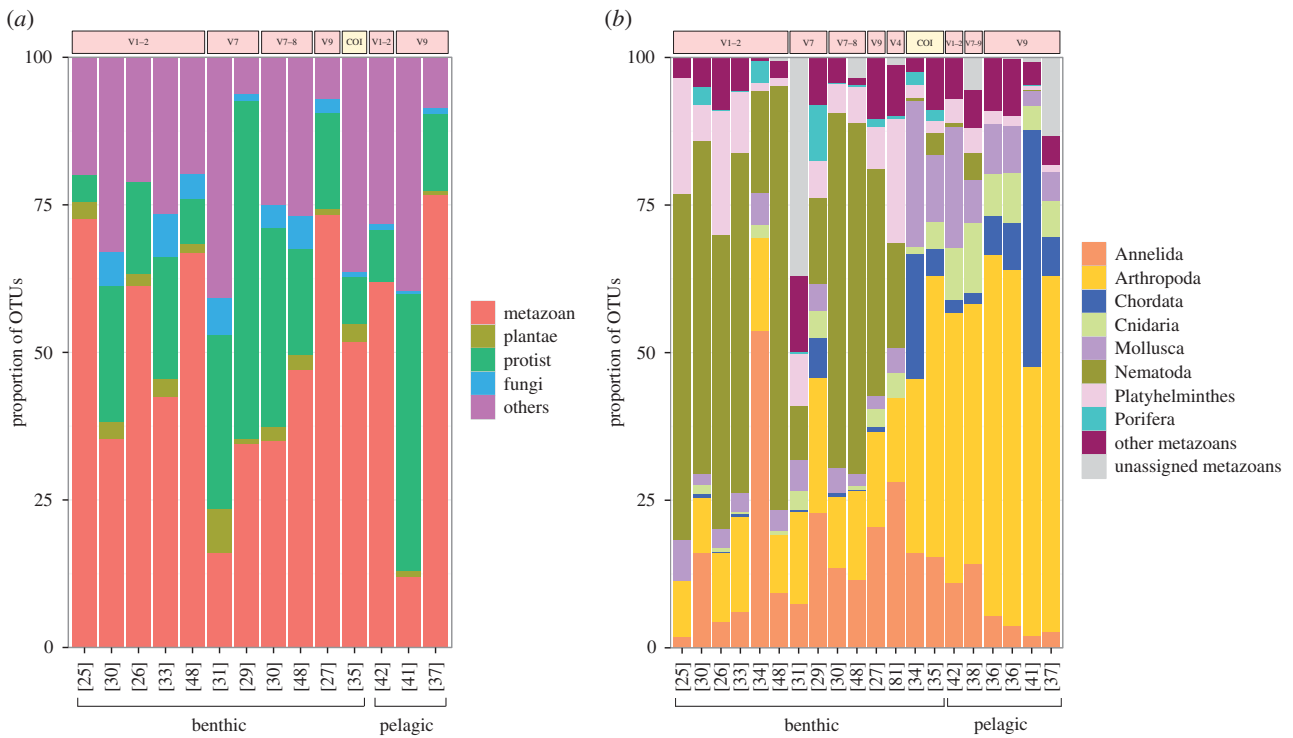
Finally, biogeographic patterns in microscopic groups have been unveiled with metabarcoding analyses that sidestep the taxonomic impediment. Available data have confirmed the previously proposed idea [84,85] that the most abundant taxa also often have wider distributions, providing strong evidence for a correlation between dispersal ability and other life-history traits. For example, Fonseca *et al.* [26] showed that the level of overlap in OTU composition of interstitial communities across a vast study area was highest among Nematoda and Platyhelminthes, the two most abundant phyla. Despite the presence of truly cosmopolitan taxa, groups of endemic OTUs were also identified as representative of biogeographic provinces [25,26,30]. Similarly, analysis of spatial patterns of planktonic diversity documented strong partitioning in plankton community composition that closely followed boundaries of water masses [40,41]. The first circumglobal metabarcoding analysis of ocean diversity [41] identified a significant correlation between geographical distance and community similarity at the scale of ocean basins for all size fractions of organisms living in the plankton. However, there were weaker levels of differentiation between communities of small-sized taxa than between communities of large sized taxa, suggesting increasing dispersal limitation with increasing body size.

## 4. Towards a global understanding of marine diversity?

Although trends are already emerging from individual studies, a global comparative dataset (i.e. Ocean Sampling Day [86]), integrating all kingdoms of life and all size classes, and bridging a range of marine ecosystems, would address outstanding questions in ecology beyond the most conspicuous species: do biodiversity patterns of unicellular eukaryotes and microscopic metazoans mirror in direction and magnitude the biodiversity patterns described for macrofauna (e.g. the latitudinal diversity gradient and the Coral Triangle biodiversity hotspot)? How do taxonomic, phylogenetic and functional diversity scale with geographical area, organism size and habitat diversity? How do patterns of commonness and rarity vary as a function of organism size and life-history traits? Does marine biodiversity vary predictably as a function of anthropogenic stress? Can we identify functional keystone groups that can be used as biological indicators? Answering these fundamental questions today is within reach but obstacles still remain (reviewed in [87]).

### (a) Methodological artefacts

The field of eukaryotic metabarcoding has been largely inspired by studies on prokaryotes for which analytical procedures have been extensively tested [88,89]. Studying eukaryotes presents its own technical challenges [87] that are now being addressed. Sources of variation in OTU detection and community profiling have been identified at each step of the metabarcoding analysis: sample collection [38], sample processing [90], PCR amplification (e.g. type of polymerase [91], target marker [34], primer set [30,36], primer selectivity



**Figure 1.** Proportion of OTUs among major groups of organisms (a) and among metazoan phyla (b) reported in studies characterizing benthic and planktonic communities using metabarcoding. Studies that strictly looked at protist diversity are not represented. The marker region targeted in each study (COI or 18S variable region) is indicated. Note that several studies targeted multiple markers. The category ‘Others’ in panel (a) comprises OTUs that were reported as unassigned, environmental sequences or prokaryotes. The category ‘other metazoans’ in (b) comprises the following phyla: Acanthocephala, Brachiopoda, Bryozoa, Cephaloryncha, Chaetognatha, Ctenophora, Cylindrozoa, Dicyemida, Echinodermata, Entoprocta, Gastrotricha, Gnathostomulida, Hemichordata, Micrognathozoa, Nemertea, Orthonectida, Phoronida, Placozoa, Rotifera, Sipuncula, Tardigrada and Xenacoelomorpha.

[35,92]), sequencing (due to random sampling [93]), denoising of raw reads [94], chimera detection [95–97] and sequence clustering (e.g. type of algorithm [90,98], threshold level [26,30,43,99]). For example, pervasive amplification biases of some primers have led some to suggest caution when using OTU relative abundance in ecological assessments [92]. Others have shown that presence–absence data are also sensitive to biases because random processes affect the reproducibility of rare OTUs [93]. In addition to methodological concerns, pseudogenes and intra-individual polymorphisms [100] can lead to inflated diversity metrics by increasing the number of predicted OTUs. The optimization of protocols and the use of controls (i.e. mock communities, technical replicates) can help increase the reliability of metabarcoding for a range of applications. Overall, there is growing evidence that methodological artefacts need to be carefully considered for the design and interpretation of studies using metabarcoding.

### (b) Lack of standardization of target markers

The use of different target fragments between independent studies represents the principal limitation to addressing regional scale and long-term questions in biodiversity because amplicon data obtained from non-homologous regions cannot be compared. Most published studies using broad-range primers ([25–38,41,42,48,81], omitting studies strictly looking at protists) target hypervariable regions of the 18S rRNA gene (94%), with the scientific community now slowly converging towards the use of the V1–2 and V9 (hyper)variable regions; these were amplified in 39% and 33% of studies, respectively (figure 1a,b). The prevalence

of studies targeting 18S regions largely stems from the versatility of PCR primers that represent convenient tools to screen the entire eukaryotic domain (figure 1a). By contrast, the mitochondrial COI gene, the standard barcoding gene for animals, has been targeted in only 11% of studies.

### (c) Lack of taxonomic resolution

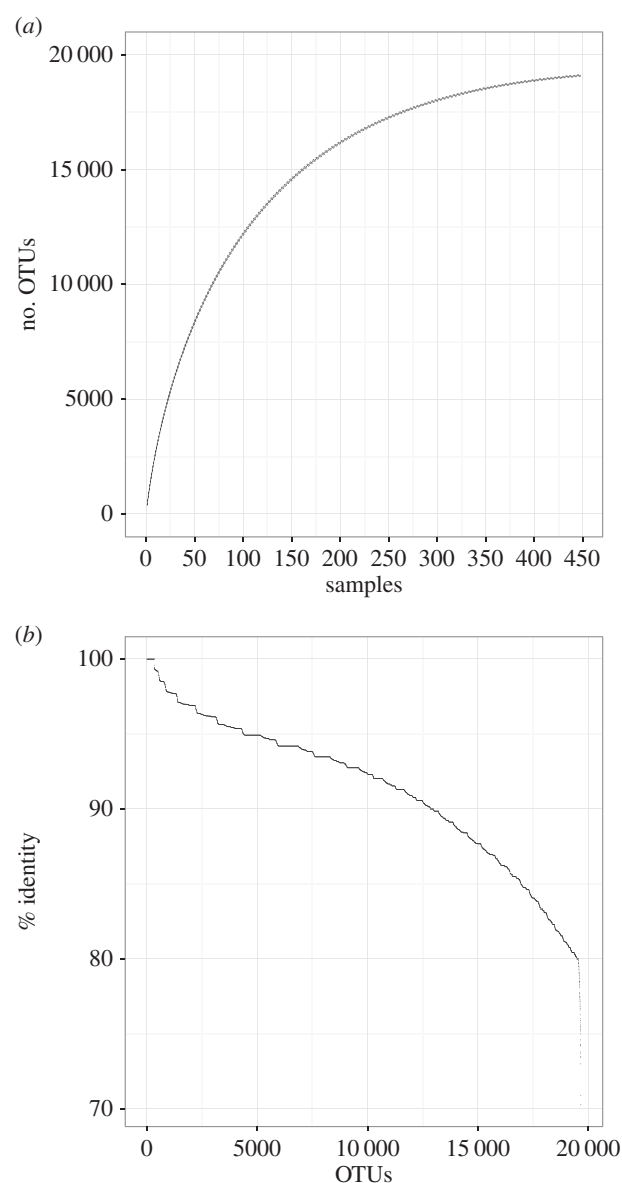
Primer versatility comes at the price of taxonomic resolution. There is now ample evidence that all 18S rRNA regions greatly underestimate the true number of metazoan species, the functional units of marine ecosystems. For example, Tang *et al.* [101] showed that 18S reduced estimates of the diversity of microscopic interstitial metazoans by a factor of 0.4 relative to morphology, whereas COI, which efficiently identified cryptic lineages, increased diversity estimates by a factor of 7.6. Similarly, Wu *et al.* [102] identified the V9 region as the most informative for taxonomic classification of copepods, but levels of nucleotide variation enabled clear differentiation only between genera and sometimes species within a few taxonomic groups. Mohrbeck *et al.* [103] came to the same conclusion after evaluating the effectiveness of the V1–V2 region for planktonic communities after identifying strictly identical sequences (100% similarity) shared by confamilial species. The performance of 18S for delineating protist species has been more challenging to evaluate because species boundaries have seldom been confirmed within an integrative taxonomic framework (i.e. morphology, ultrastructure and DNA taxonomy of cultured strains [104]). However, evidence from whole-genome sequences suggests similar limitations for biodiversity surveys [105].

**Table 1.** Total number of accepted metazoan species in the ocean according to World Register of Marine Species (WORMS) compared with the diversity of 18S V9 OTUs detected in metabarcoding analyses of plankton samples collected circumglobally.

	accepted marine species in WORMS	18S V9 OTUs in plankton <sup>a</sup>
Acanthocephala	468	0
Annelida	12 862	404 (48)
Arthropoda	57 871	9497 (205)
Brachiopoda	396	8 (4)
Bryozoa	6167	15 (5)
Cephalorhyncha	238	13
Chaetognatha	131	531 (7)
Chordata	22 248	6972 (32)
Cnidaria	11 407	842 (82)
Ctenophora	192	33 (9)
Cycliophora	2	0
Dicyemida	122	1
Echinodermata	7252	74 (19)
Entoprocta	180	12
Gastrotricha	497	2
Gnathostomulida	98	0
Hemichordata	130	12 (3)
Mollusca	45 219	560 (89)
Nematoda	7152	30 (6)
Nemertea	1359	81 (4)
Orthonectida	25	0
Phoronida	19	0
Placozoa	1	0
Platyhelminthes	12 230	148 (12)
Porifera	8476	25 (4)
Rotifera	187	7 (2)
Sipuncula	147	0
Tardigrada	193	2
Xenacoelomorpha	433	0
unassigned	n.a.	402
metazoan		
total	195 702	19 671 (531)

<sup>a</sup>De Vargas *et al.* [41] and Pearman *et al.* [36,37] for a total of 448 plankton samples and approximately 258 million reads. OTU-representative sequences of each study were clustered using SWARM [106] ( $d = 1$ ). Numbers between parentheses indicate OTUs with identity greater than 99% to the V9 PR2 database [18]. Note that some metazoans are unlikely to be routinely present in plankton samples as either adults or larvae.

To further illustrate the limitations of 18S for censusing metazoan marine diversity, we combined V9 datasets of the Tara Ocean circumglobal expedition [41] with plankton community profiles for surface water around coral reefs [36] and along a depth profile [37] in the Red Sea. Together, these studies account for most of the sampling and sequencing effort conducted in the ocean for metazoans, with 448



**Figure 2.** Metazoan diversity detected in the plankton based on 18S V9 sequencing. Sample-based rarefaction ((a), without replacement) and levels of sequence similarity to reference barcodes of the PR2 database (b) [18] were calculated using a dataset combining OTUs detected during the Tara Ocean circumglobal expedition [41] and OTUs collected in surface water around coral reefs [36] and along a depth profile [37] in the Red Sea. Together, these studies account for most of the sampling and sequencing effort conducted in the ocean for metazoans, with 448 samples and approximately 258 million metazoan sequences. OTU-representative sequences of each study were clustered using SWARM ( $d = 1$ ) [106].

samples and approximately 258 million metazoan sequences. After clustering OTU-representative sequences of respective studies using Swarm ( $d = 1$ ) [106], there were a total of 19 671 metazoan OTUs, which represents only a tiny fraction of what is known to live in the ocean (table 1). Yet, the rate of discovery of metazoan OTUs with the V9 region of the 18S rRNA gene is beginning to plateau after 448 samples (figure 2a). While this may be partly caused by the highly conservative quality control applied to the Tara dataset (e.g. removal of all unique reads present in a single sample) and the fact that some marine metazoans are unlikely to be regularly present in plankton samples, this pattern is also probably driven by low levels of variability of the V9 region.

## (d) Paucity of reference barcode data in public databases

While representative barcode sequences are available across most of the branches of the tree of life (however, see [16] for protists), the lack of coverage at lower taxonomic levels currently hinders interpretation of metabarcoding data. Most OTUs remain identified to taxonomic groups that have limited taxonomic and functional relevance in biodiversity inventories, ecological studies or monitoring initiatives. For example, very few metazoan OTUs in the 18S V9 circumglobal plankton dataset matched sequences of the Protist Ribosomal Reference (PR2 [18]) database with levels of sequence similarity typical of genus-level matches (2.7% with more than 99% similarity, figure 2b and table 1) (see [102] for threshold justification). This problem is also acute in studies targeting the hypervariable COI marker because COI tends to saturate at higher taxonomic levels. High levels of homoplasy between distant phylogenetic groups decrease the likelihood of confidently assigning OTUs in the absence of close representatives at lower taxonomic levels. This translated into a large number of OTUs classified as ‘unidentified’ in a recent benthic survey (28.3% [35]). Ongoing efforts by various working groups of the Consortium for the Barcode of Life (CBOL, www.barcodeoflife.org) to populate public barcode repositories and compile available data into curated databases (i.e. PR2 [18], Silva [107]) will help fill this gap.

## 5. Concluding remarks

The trade-off between amplification success and taxonomic resolution has long been recognized in PCR-based studies. Markers with highly conserved flanking regions underestimate the true number of species, whereas markers with more variable flanking regions usually provide better estimates of species richness. Shotgun metagenomics and metatranscriptomics remove the need to compromise, but they require a much higher sequencing depth per sample and have proved less effective at detecting rare taxa [108]. As a result, there is an urgent need to synergize and coordinate metabarcoding efforts because metabarcoding is likely to remain one of the prime methods for biodiversity monitoring and ecological studies for at least a few more years. For example, a strategy

in which samples would be analysed using a two-step approach combining gene regions that vary in taxonomic coverage and taxonomic resolution would better reflect marine diversity and enable future comparative studies. A first step could include profiling whole eukaryotic communities at a coarse taxonomic level by targeting 18S rRNA V1–V2 or V9 regions as a ‘pre-metabarcoding’ (cf. [16]), two regions for which highly versatile primers are available (see [109] for *in silico* tests of existing primer sets). A second step could target one or several more variable markers with broad-range lineage-specific primers (i.e. COI [50] and 12S [110] for metazoans, ITS for fungi [15]) to provide enhanced level of detail for groups that remain under-explored. Importantly, environmental vouchers (e.g. DNA and RNA extracts, remaining tissue homogenates) are as important for biodiversity studies as individual vouchers are for taxonomy. Therefore, we may also envision the creation of a network of museum-based repositories from which environmental vouchers could be loaned for complementary metabarcoding or genomics applications.

The ocean represents approximately 70% and more than 90% of the Earth’s surface and habitable volume, respectively. A holistic understanding of marine diversity will only become possible with coordinated efforts, further methodological developments, strict methodological standards and consistency of experimental designs [111]. Predicting how species respond to environmental change (e.g. range expansion and extinction) will only be possible if we include markers in metabarcoding analyses that are both consistent across taxa and hypervariable. Yet the need for biodiversity baselines cannot be postponed in this rapidly changing world, and today’s methods for metabarcoding are already very powerful. Thus, it is also important, as Voltaire noted, not to let the perfect become the enemy of the good.

**Authors’ contributions.** The authors contributed equally to this manuscript.

**Competing interests.** We have no competing interests.

**Funding.** The Sant Chair of the Smithsonian National Museum of Natural History funded this work.

**Acknowledgements.** We thank all authors who provided additional data and Ryuji Machida for his insightful comments on an early draft of the manuscript. We are also grateful to three anonymous reviewers who greatly helped improve this manuscript.

## References

1. McCauley DJ, Pinsky ML, Palumbi SR, Estes JA, Joyce FH, Warner RR. 2015 Marine defaunation: animal loss in the global ocean. *Science* **347**, 1255641. (doi:10.1126/science.1255641)
2. Azam F, Fenichel T, Field JG, Gray JC, Meyer-Reil LA, Thingstad F. 1983 The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10**, 257–264. (doi:10.3354/meps010257)
3. Appeltans W *et al.* 2012 The magnitude of global marine species diversity. *Curr. Biol.* **22**, 2189–2202. (doi:10.1016/j.cub.2012.09.036)
4. Costello MJ, Wilson S, Houlding B. 2012 Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Syst. Biol.* **61**, 871–883. (doi:10.1093/sysbio/syr080)
5. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011 How many species are there on Earth and in the Ocean? *PLoS Biol.* **9**, e1001127. (doi:10.1371/journal.pbio.1001127)
6. Bouchet P, Lozouet P, Maestrati P, Heros V. 2002 Assessing the magnitude of species richness in tropical marine environments: exceptionally high numbers of molluscs at a New Caledonia site. *Biol. J. Linn. Soc.* **75**, 421–436. (doi:10.1046/j.1095-8312.2002.00052.x)
7. Mees J *et al.* 2015 World Register of Marine Species. See <http://www.marinespecies.org> (accessed 17 October 2015).
8. Jefferson TA, Rosenbaum HC. 2014 Taxonomic revision of the humpback dolphins (*Sousa* spp.), and description of a new species from Australia. *Mar. Mammal Sci.* **30**, 1494–1541. (doi:10.1111/mms.12152)
9. Ley PDe. 2000 Lost in worm space: phylogeny and morphology as road maps to nematode diversity. *Nematology* **2**, 9–16. (doi:10.1163/156854100508854)
10. Wägele H, Klusmann-Kolb A, Kuhlmann M, Haszprunar G, Lindberg D, Koch A, Wägele JW. 2011 The taxonomist—an endangered race. A practical proposal for its survival. *Front. Zool.* **8**, 25. (doi:10.1186/1742-9994-8-25)
11. Hughes TP, Bellwood DR, Connolly SR. 2002 Biodiversity hotspots, centres of endemism, and the conservation of coral reefs. *Ecol. Lett.* **5**, 775–784. (doi:10.1046/j.1461-0248.2002.00383.x)

12. Fisher R, Radford BT, Knowlton N, Brainard RE, Michaelis FB, Caley MJ. 2011 Global mismatch between research effort and conservation needs of tropical coral reefs. *Conserv. Lett.* **4**, 64–72. (doi:10.1111/j.1755-263X.2010.00146.x)
13. Hebert PDN, Cywinska A, Ball SL, DeWaard JR. 2003 Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* **270**, 313–321. (doi:10.1098/rspb.2002.2218)
14. CBOL Plant Working Group. 2009 A DNA barcode for land plants. *Proc. Natl Acad. Sci. USA* **106**, 12 794–12 797. (doi:10.1073/pnas.0905845106)
15. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W. 2012 Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *Proc. Natl Acad. Sci. USA* **109**, 6241–6246. (doi:10.1073/pnas.1117018109)
16. Pawlowski J *et al.* 2012 CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* **10**, e1001419. (doi:10.1371/journal.pbio.1001419)
17. Ratnasingham S, Hebert PDN. 2007 BOLD: the Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). *Mol. Ecol. Notes* **7**, 355–364. (doi:10.1111/j.1471-8286.2007.01678.x)
18. Guillou L *et al.* 2013 The Protist Ribosomal Reference database (PR<sup>2</sup>): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, D597–D604. (doi:10.1093/nar/gks1160)
19. Albano PG, Sabelli B, Bouchet P. 2011 The challenge of small and rare species in marine biodiversity surveys: microgastropod diversity in a complex tropical coastal environment. *Biodivers. Conserv.* **20**, 3223–3237. (doi:10.1007/s10531-011-0117-x)
20. Hajibabaei M. 2012 The golden age of DNA metasystematics. *Trends Genet.* **28**, 535–537. (doi:10.1016/j.tig.2012.08.001)
21. Jonasson J, Olofsson M, Monstein H-J. 2002 Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16S rDNA fragments. *APMIS* **110**, 263–272. (doi:10.1034/j.1600-0463.2002.100309.x)
22. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006 Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc. Natl Acad. Sci. USA* **103**, 12 115–12 120. (doi:10.1073/pnas.0605127103)
23. Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ. 2011 Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* **6**, e17497. (doi:10.1371/journal.pone.0017497)
24. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012 Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050. (doi:10.1111/j.1365-294X.2012.05470.x)
25. Fonseca VG *et al.* 2010 Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat. Commun.* **1**, 98. (doi:10.1038/ncomms1095)
26. Fonseca VG *et al.* 2014 Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. *Glob. Ecol. Biogeogr.* **23**, 1293–1302. (doi:10.1111/geb.12223)
27. Brannock P, Waits D, Sharma J, Halanych K. 2014 High-throughput sequencing characterizes intertidal meiofaunal communities in northern Gulf of Mexico (Dauphin Island and Mobile Bay, Alabama). *Biol. Bull.* **227**, 161–174.
28. Pawlowski J, Christen R, Lecroq B, Bachar D, Shahbazkia HR, Amaral-Zettler L, Guillou L. 2011 Eukaryotic richness in the abyss: insights from pyrotag sequencing. *PLoS ONE* **6**, e18169. (doi:10.1371/journal.pone.0018169)
29. Guardiola M, Uriz MJ, Taberlet P, Coissac E, Wangensteen S, Turon X. 2015 Deep-sea, deep-sequencing: metabarcoding extracellular DNA from sediments of marine canyons. *PLoS ONE* **10**, e0139633. (doi:10.1371/journal.pone.0139633)
30. Bik HM, Sung W, De Ley P, Baldwin JG, Sharma J, Rocha-Olivares A, Thomas WK. 2012 Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. *Mol. Ecol.* **21**, 1048–1059. (doi:10.1111/j.1365-294X.2011.05297.x)
31. Chariton AA, Stephenson S, Morgan MJ, Steven ADL, Colloff MJ, Court LN, Hardy CM. 2015 Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Environ. Pollut.* **203**, 165–174. (doi:10.1016/j.envpol.2015.03.047)
32. Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM. 2010 Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Front. Ecol. Environ.* **8**, 233–238. (doi:10.1890/090115)
33. Lallias D *et al.* 2015 Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *ISME J.* **9**, 1208–1221. (doi:10.1038/ismej.2014.213)
34. Cowart DA, Pinheiro M, Mouchel O, Maguer M, Grall J, Miné J, Arnaud-Haond S. 2015 Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS ONE* **10**, e0117562. (doi:10.1371/journal.pone.0117562)
35. Leray M, Knowlton N. 2015 DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proc. Natl Acad. Sci. USA* **112**, 2076–2081. (doi:10.1073/pnas.1424997112)
36. Pearman JK, El-Sherbiny MM, Lanzen A, Al-Aidarous AM, Irigoien X. 2014 Zooplankton diversity across three Red Sea reefs using pyrosequencing. *Front. Mar. Sci.* **1**, 1–11. (doi:10.3389/fmars.2014.00027)
37. Pearman JK, Irigoien X. 2015 Assessment of zooplankton community composition along a depth profile in the central Red Sea. *PLoS ONE* **10**, e0133487. (doi:10.1371/journal.pone.0133487)
38. Hirai J, Yasuie M, Fujiwara A, Nakamura Y, Hamaoka S, Katakura S, Takano Y, Nagai S. 2015 Effects of plankton net characteristics on metagenetic community analysis of metazoan zooplankton in a coastal marine ecosystem. *J. Exp. Mar. Biol. Ecol.* **469**, 36–43. (doi:10.1016/j.jembe.2015.04.011)
39. Hirai J, Tsuda A. 2015 Metagenetic community analysis of epipelagic planktonic copepods in the tropical and subtropical Pacific. *Mar. Ecol. Prog. Ser.* **534**, 65–78. (doi:10.3354/meps11404)
40. Hirai J, Kuriyama M, Ichikawa T, Hidaka K, Tsuda A. 2015 A metagenetic approach for revealing community structure of marine planktonic copepods. *Mol. Ecol. Resour.* **15**, 68–80. (doi:10.1111/1755-0998.12294)
41. de Vargas C *et al.* 2015 Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605. (doi:10.1126/science.1261605)
42. Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A. 2013 Next generation sequencing reveals the hidden diversity of zooplankton assemblages. *PLoS ONE* **8**, e81327. (doi:10.1371/journal.pone.0081327)
43. Logares R *et al.* 2014 Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* **24**, 813–821. (doi:10.1016/j.cub.2014.02.050)
44. Massana R, Pernice M, Bunge JA, del Campo J. 2011 Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean. *ISME J.* **5**, 184–195. (doi:10.1038/ismej.2010.104)
45. Cheung MK, Au CH, Chu KH, Kwan HS, Wong CK. 2010 Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *ISME J.* **4**, 1053–1059. (doi:10.1038/ismej.2010.26)
46. Not F, del Campo J, Balagué V, de Vargas C, Massana R. 2009 New insights into the diversity of marine picoeukaryotes. *PLoS ONE* **4**, e7143. (doi:10.1371/journal.pone.0007143)
47. Kramer MJ, Bellwood O, Bellwood DR. 2013 The trophic importance of algal turfs for coral reef fishes: the crustacean link. *Coral Reefs* **32**, 575–583. (doi:10.1007/s00338-013-1009-1)
48. Bik HM, Halanach KM, Sharma J, Thomas WK. 2012 Dramatic shifts in benthic microbial eukaryote communities following the Deepwater Horizon oil spill. *PLoS ONE* **7**, e38550. (doi:10.1371/journal.pone.0038550)
49. Leray M, Meyer CP, Mills SC. 2015 Metabarcoding dietary analysis of coral dwelling predatory fish demonstrates the minor contribution of coral mutualists to their highly partitioned, generalist diet. *PeerJ* **3**, e1047. (doi:10.7717/peerj.1047)
50. Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ. 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)
51. Deagle BE, Chiaradia A, McInnes J, Jarman SN. 2010 Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out?

- Conserv. Genet.* **11**, 2039–2048. (doi:10.1007/s10592-010-0096-6)
52. Zaiko A, Martinez JL, Schmidt-Petersen J, Ribicic D, Samuiloviene A, Garcia-Vazquez E. 2015 Metabarcoding approach for the ballast water surveillance—an advantageous solution or an awkward challenge? *Mar. Pollut. Bull.* **92**, 25–34. (doi:10.1016/j.marpolbul.2015.01.008)
53. Warwick RM, Clarke KR. 1996 Relationships between body-size, species abundance and diversity in marine benthic assemblages: facts or artefacts? *J. Exp. Mar. Biol. Ecol.* **202**, 63–71. (doi:10.1016/0022-0981(96)00031-7)
54. Fenchel T. 1993 There are more small than large species? *Oikos* **68**, 375–378. (doi:10.2307/3544855)
55. May RM. 1988 How many species are there on Earth? *Science* **241**, 1441–1449. (doi:10.1126/science.241.4872.1441)
56. Caron DA, Countway PD, Jones AC, Kim DY, Schnetzer A. 2012 Marine protistan diversity. *Annu. Rev. Mar. Sci.* **4**, 467–493. (doi:10.1146/annurev-marine-120709-142802)
57. Santoferrara LF, Grattepanche J-D, Katz LA, Mcmanus GB. 2014 Pyrosequencing for assessing diversity of eukaryotic microbes: analysis of data on marine planktonic ciliates and comparison with traditional methods. *Environ. Microbiol.* **16**, 2752–2763. (doi:10.1111/1462-2920.12380)
58. Georges C, Monchy S, Genitsaris S, Christaki U. 2014 Protist community composition during early phytoplankton blooms in the naturally iron-fertilized Kerguelen area (Southern Ocean). *Biogeosci. Discuss.* **11**, 11 179–11 215. (doi:10.5194/bg-11-5847-2014)
59. Brown MV, Philip GK, Bunge JA, Smith MC, Bissett A, Lauro FM, Fuhrman JA, Donachie SP. 2009 Microbial community structure in the North Pacific ocean. *ISME J.* **3**, 1374–1386. (doi:10.5194/bg-11-5847-2014)
60. Yu L, Zhang W, Liu L, Yang J. 2015 Determining microeukaryotic plankton community around Xiamen Island, Southeast China, using Illumina MiSeq and PCR-DGGE techniques. *PLoS ONE* **10**, e0127721. (doi:10.1371/journal.pone.0127721)
61. Logares R, Audic S, Santini S, Pernice MC, de Vargas C, Massana R. 2012 Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J.* **6**, 1823–1833. (doi:10.1038/ismej.2012.36)
62. Bittner L *et al.* 2013 Diversity patterns of uncultured Haptophytes unravelled by pyrosequencing in Naples Bay. *Mol. Ecol.* **22**, 87–101. (doi:10.1111/mec.12108)
63. Lecroq B, Lejzerowicz F, Bachar D, Christen R, Esling P, Baerlocher L, Østerås M, Farinelli L, Pawlowski J. 2011 Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *Proc. Natl Acad. Sci. USA* **108**, 13 177–13 182. (doi:10.1073/pnas.1018426108)
64. Bass D, Cavalier-Smith T. 2004 Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *Int. J. Syst. Evol. Microbiol.* **54**, 2393–2404. (doi:10.1099/ijs.0.63229-0)
65. Bernhard JM, Edgcomb VP, Visscher PT, McIntyre-Wressnig A, Summons RE, Bouxsein ML, Louis L, Jeglinski M. 2013 Insights into foraminiferal influences on microfibrils of microbialites at Highborne Cay, Bahamas. *Proc. Natl Acad. Sci. USA* **110**, 9830–9834. (doi:10.1073/pnas.1221721110)
66. del Campo J, Ruiz-Trillo I. 2013 Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. *Mol. Biol. Evol.* **30**, 802–805. (doi:10.1093/molbev/mst006)
67. Not F, Valentin K, Romari K, Lovejoy C, Massana R, Töbe K, Vaulot D, Medlin LK. 2007 Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science* **315**, 253–255. (doi:10.1126/science.1136264)
68. Kim E, Harrison JW, Sudek S, Jones MDM, Wilcox HM, Richards TA, Worden AZ, Archibald JM. 2011 Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proc. Natl Acad. Sci. USA* **108**, 1496–1500. (doi:10.1073/pnas.1013337108)
69. Pawlowski J. 2013 The new micro-kingdoms of eukaryotes. *BMC Biol.* **11**, 40–41. (doi:10.1186/1741-7007-11-40)
70. Platt HM, Warwick RM. 1980 The significance of free-living nematodes to the littoral ecosystem. In *The Shore Environment-2 Ecosystems* (eds JH Price, DEG Irvine), pp. 729–759. London, UK: Academic Press.
71. Lamshead PJD, Boucher G. 2003 Marine nematode deep-sea biodiversity-hyperdiverse or hype? *J. Biogeogr.* **30**, 475–485. (doi:10.1046/j.1365-2699.2003.00843.x)
72. Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E. 2005 Defining operational taxonomic units using DNA barcode data. *Phil. Trans. R. Soc. B* **360**, 1935–1943. (doi:10.1098/rstb.2005.1725)
73. Floyd R, Abebe E, Papert A, Blaxter M. 2002 Molecular barcodes for soil nematode identification. *Mol. Ecol.* **11**, 839–850. (doi:10.1046/j.1365-294X.2002.01485.x)
74. Creer S *et al.* 2011 Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Mol. Ecol.* **19**, 4–20. (doi:10.1111/j.1365-294X.2009.04473.x)
75. Radziejewska T, Stańkowska-Radziun M. 1979 Intertidal meiofauna of Rekerchefjorden and Mablukta, West. Spitsbergen. *Sarsia* **64**, 253–258. (doi:10.1080/00364827.1979.10411385)
76. Martens PM, Schockaert ER. 1986 The importance of turbellarians in the marine meiobenthos: a review. *Hydrobiologia* **132**, 295–303. (doi:10.1007/BF00046263)
77. Beaugrand G, Brander KM, Alistair Lindley J, Souissi S, Reid PC. 2003 Plankton effect on cod recruitment in the North Sea. *Nature* **426**, 661–664. (doi:10.1038/nature02164)
78. Alve E. 1995 Benthic foraminiferal responses to estuarine pollution; a review. *J. Foraminifer. Res.* **25**, 190–203. (doi:10.2113/gsjfr.25.3.190)
79. Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA. 2014 Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Mol. Ecol. Resour.* **14**, 1129–1140. (doi:10.1111/1755-0998.12261)
80. Pochon X, Wood SA, Keeley NB, Lejzerowicz F, Esling P, Drew J, Pawlowski J. 2015 Accurate assessment of the impact of salmon farming on benthic sediment enrichment using foraminiferal metabarcoding. *Mar. Pollut. Bull.* **100**, 370–382. (doi:10.1016/j.marpolbul.2015.08.022)
81. Lejzerowicz F, Esling P, Pillet L, Wilding TA, Black KD, Pawlowski J. 2015 High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Sci. Rep.* **5**, 13932. (doi:10.1038/srep13932)
82. Aylagas E, Borja A, Rodríguez-Ezpeleta N. 2014 Environmental status assessment using DNA metabarcoding: towards a genetics based Marine Biotic Index (gAMBI). *PLoS ONE* **9**, e90529. (doi:10.1371/journal.pone.0090529)
83. Pochon X, Bott NJ, Smith KF, Wood SA. 2013 Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLoS ONE* **8**, e73935. (doi:10.1371/journal.pone.0073935)
84. Fenchel T, Finlay BJ. 2004 The ubiquity of small species: patterns of local and global diversity. *Bioscience* **54**, 777–784. (doi:10.1641/0006-3568(2004)054[0777:TUOSSP]2.0.CO;2)
85. Finlay BJ. 2002 Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–1063. (doi:10.1126/science.1070710)
86. Kopf A *et al.* 2015 The ocean sampling day consortium. *Gigascience* **4**, 27. (doi:10.1186/s13742-015-0066-5)
87. Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK. 2012 Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol. Evol.* **27**, 233–243. (doi:10.1016/j.tree.2011.11.010)
88. Huse SM, Welch DM, Morrison HG, Sogin ML. 2010 Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* **12**, 1889–1898. (doi:10.1111/j.1462-2920.2010.02193.x)
89. Sun Y, Cai Y, Huse SM, Knight R, Farmerie WG, Wang X, Mai V. 2012 A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. *Brief. Bioinform.* **13**, 107–121. (doi:10.1093/bib/bbr009)
90. Brannock PM, Halanach KM. 2015 Meiofaunal community analysis by high-throughput sequencing: comparison of extraction, quality filtering, and clustering methods. *Mar. Genomics* **23**, 67–75. (doi:10.1016/j.margen.2015.05.007)
91. Oliver AK, Brown SP, Callahan MA, Jumpponen A. 2015 Polymerase matters: non-proofreading enzymes inflate fungal community richness estimates by up to 15%. *Fungal Ecol.* **15**, 86–89. (doi:10.1016/j.funeco.2015.03.003)

92. 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. (doi:10.1371/journal.pone.0130324)
93. Zhan A, He S, Brown EA, Chain FJJ, Theriault TW, Abbott CL, Heath DD, Cristescu ME, Maclsaac HJ. 2014 Reproducibility of pyrosequencing data for biodiversity assessment in complex communities. *Methods Ecol. Evol.* **5**, 881–890. (doi:10.1111/2041-210X.12230)
94. Rossberg AG, Rogers T, McKane AJ. 2014 Current noise-removal methods can create false signals in ecogenomic data. *Proc. R. Soc. B* **281**, 20140191. (doi:10.1098/rspb.2014.0191)
95. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011 UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200. (doi:10.1093/bioinformatics/btr381)
96. Porazinska DL, Giblin-Davis RM, Sung W, Thomas WK. 2012 The nature and frequency of chimeras in eukaryotic metagenetic samples. *J. Nematol.* **44**, 18–25.
97. Fonseca VG, Nichols B, Lallias D, Quince C, Carvalho GR, Power DM, Creer S. 2012 Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Res.* **40**, e66. (doi:10.1093/nar/gks002)
98. Flynn JM, Brown EA, Chain FJJ, Maclsaac HJ, Cristescu ME. 2015 Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods. *Ecol. Evol.* **5**, 2252–2266. (doi:10.1002/ece3.1497)
99. Brown EA, Chain FJJ, Crease TJ, Maclsaac HJ, Cristescu ME. 2015 Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? *Ecol. Evol.* **5**, 2234–2251. (doi:10.1002/ece3.1485)
100. Decelle J, Romac S, Sasaki E, Not F, Mahé F. 2014 Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (radiolarians) assessed by high-throughput sequencing. *PLoS ONE* **9**, e104297. (doi:10.1371/journal.pone.0104297)
101. Tang CQ, Leasi F, Obertegger U, Kieneke A, Barraclough TG, Fontaneto D. 2012 The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proc. Natl Acad. Sci. USA* **109**, 16 208–16 212. (doi:10.1073/pnas.1209160109)
102. Wu S, Xiong J, Yu Y. 2015 Taxonomic resolutions based on 18S rRNA genes: a case study of subclass copepoda. *PLoS ONE* **10**, e0131498. (doi:10.1371/journal.pone.0131498)
103. Mohrbeck I, Raupach MJ, Martínez Arbizu P, Knebelberger T, Laakmann S. 2015 High-throughput sequencing—the key to rapid biodiversity assessment of marine metazoa? *PLoS ONE* **10**, e0140342. (doi:10.1371/journal.pone.0140342)
104. Hoef-Emden K, Küpper FC, Andersen RA. 2007 Meeting report: Sloan Foundation Workshop to resolve problems relating to the taxonomy of microorganisms and to culture collections arising from the barcoding initiatives; Portland ME, November 6–7, 2006. *Protist* **158**, 135–137. (doi:10.1016/j.protis.2007.02.001)
105. Piganeau G, Eyre-Walker A, Jancek S, Grimsley N, Moreau H. 2011 How and why DNA barcodes underestimate the diversity of microbial eukaryotes. *PLoS ONE* **6**, e16342. (doi:10.1371/journal.pone.0016342)
106. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. 2014 Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* **2**, e593. (doi:10.7717/peerj.593)
107. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013 The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596. (doi:10.1093/nar/gks1219)
108. Srivathsan A, Sha JCM, Vogler AP, Meier R. 2015 Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (*Pygathrix nemaeus*). *Mol. Ecol. Resour.* **15**, 250–261. (doi:10.1111/1755-0998.12302)
109. Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C. 2014 Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS ONE* **9**, e87624. (doi:10.1371/journal.pone.0087624)
110. Machida RJ, Kweskin M, Knowlton N. 2012 PCR primers for metazoan mitochondrial 12S ribosomal DNA sequences. *PLoS ONE* **7**, e35887. (doi:10.1371/journal.pone.0035887)
111. Duffy JE, Amaral-Zettler LA, Fautin DG, Paulay G, Ryneerson TA, Sosik HM, Stachowicz JJ. 2013 Envisioning a marine biodiversity observation network. *Bioscience* **63**, 350–361. (doi:10.1525/bio.2013.63.5.8)