



Research



Cite this article: Geraldi NR *et al.* 2024 Assessing patterns of metazoans in the global ocean using environmental DNA. *R. Soc. Open Sci.* **11**: 240724. <https://doi.org/10.1098/rsos.240724>

Received: 30 April 2024

Accepted: 16 July 2024

Subject Category:

Ecology, conservation, and global change biology

Subject Areas:

ecology, ecosystem, environmental science

Keywords:

deep sea, diversity, global abundance, metabarcoding, metagenome, metazoan

Author for correspondence:

Nathan R. Geraldi

e-mail: nathan.geraldi@kaust.edu.sa

[†]Present address: NatureMetrics, Guildford, UK.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7397762>.

Assessing patterns of metazoans in the global ocean using environmental DNA

Nathan R. Geraldi^{1,2,†}, Silvia G. Acinas³, Intikhab Alam², Josep M. Gasol^{3,4}, María Luz Fernández-de-Puelles⁵, Caterina R. Giner^{3,6}, Santiago Hernández León⁷, Ramiro Logares³, Ramon Massana³, Pablo Sánchez³, Vladimir Bajic², Takashi Gojobori² and Carlos M. Duarte^{1,2}

¹Red Sea Research Center, and ²Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

³Institut de Ciències del Mar, CSIC, Barcelona, Catalunya, Spain

⁴Center for Marine Ecosystem Research, Edith Cowan University, Joondalup, Western Australia, Australia

⁵Instituto Español de Oceanografía, Centro Oceanográfico de Baleares, Palma de Mallorca 07015, Spain

⁶Institute for the Oceans and Fisheries, University of British Columbia, UBC-AERL, Vancouver, Canada

⁷Instituto de Oceanografía y Cambio Global, IOCAG, Universidad de Las Palmas de Gran Canaria, Unidad Asociada ULPGC-CSIC, Campus de Taliarte, Telde, Gran Canaria, Canary Islands 35214, Spain

NRG, 0000-0002-2669-3867; JMG, 0000-0001-5238-2387

Documenting large-scale patterns of animals in the ocean and determining the drivers of these patterns is needed for conservation efforts given the unprecedented rates of change occurring within marine ecosystems. We used existing datasets from two global expeditions, *Tara Oceans* and *Malaspina*, that circumnavigated the oceans and sampled down to 4000 m to assess metazoans from environmental DNA (eDNA) extracted from seawater. We describe patterns of taxonomic richness within metazoan phyla and orders based on metabarcoding and infer the relative abundance of phyla using metagenome datasets, and relate these data to environmental variables. Arthropods had the greatest taxonomic richness of metazoan phyla at the surface, while cnidarians had the greatest richness in pelagic zones. Half of the marine metazoan eDNA from metagenome datasets was from arthropods, followed by cnidarians and nematodes. We found that mean surface temperature and primary

productivity were positively related to metazoan taxonomic richness. Our findings concur with existing knowledge that temperature and primary productivity are important drivers of taxonomic richness for specific taxa at the ocean's surface, but these correlations are less evident in the deep ocean. Massive sequencing of eDNA can improve understanding of animal distributions, particularly for the deep ocean where sampling is challenging.

1. Introduction

Anthropogenic impacts, such as climate change and overexploitation, are resulting in continued biodiversity loss on land and in the ocean [1,2]. Yet, we can barely understand the consequences of these changes in the ocean as we are far from achieving a comprehensive understanding of animal distribution and diversity in the marine realm, particularly in the open and deep ocean [3,4]. Enhanced knowledge of the open and deep ocean is important given they form one of the largest habitats in the biosphere [5,6]. Moreover, it is difficult to quantify fragile animals with conventional techniques, in particular, gelatinous organisms that forage, and diet studies suggest to be prevalent in the deep sea [7,8]. An enhanced understanding of marine animal distribution and environmental drivers across the ocean is needed to characterize ecosystem structure, and as a baseline for detecting and understanding change in marine communities as indicated by the Convention on Biological Diversity [9] and the United Nations 2030 Agenda for Sustainable Development [10].

Environmental DNA (eDNA) analyses—based on the detection of marker gene sequences (metabarcoding or metagenome sequencing)—can identify a wide array of taxa regardless of life stage, while traditional surveys often only measure animals of specific sizes and life stages. Although different definitions of eDNA exist, it can be defined as DNA extracted from an environmental sample, including intra- and extra-cellular DNA, to attain the most all-inclusive taxonomic data [11]. eDNA-based surveys have emerged as a powerful tool to accelerate the description of diversity in the marine environment by complementing traditional sampling methods because they bypass the need to capture or observe organisms [11–13].

Here, we utilized metabarcoding (PCR amplicons of the 18S rRNA gene) and metagenome (shotgun sequencing of whole community DNA) datasets from already published global expeditions (figure 1). These expeditions targeted microbial size fractions collected from seawater and publications characterized micro-eukaryotes in the photic zone from the *Tara Oceans* expedition [14,15] and in both photic and aphotic zones from the *Malaspina* expedition [16,17]. However, some of these studies reported the presence of metazoan DNA and we aimed to use these datasets to enhance the understanding of the patterns of taxonomic richness and the abundance of marine metazoans in the open and deep ocean based on eDNA. In particular, we used these datasets to (i) describe global patterns of taxonomic richness and the abundance of individual metazoan phyla in the ocean and (ii) identify environmental predictors of taxonomic richness and the abundance of all marine metazoans for the sun-lit and dark ocean.

To determine the predictors of oceanic metazoan richness (from metabarcoding datasets) and relative abundance (from metagenome datasets), we chose variables based on the following *a priori* hypotheses. First, taxon richness is expected to increase with temperature but to decrease with depth following multiple theories including metabolic-related hypotheses because temperature is positively related to physiological processes and mutation rates, which may lead to faster speciation rates [18,19], and 'the more-individuals hypothesis' based on the positive relationship between the number of individuals and available energy, which also suggests a positive relationship between primary productivity and richness [20,21]. These species richness-energy hypotheses (metabolic theory/kinetic energy) are particularly important for the deep sea [6]. In addition, 'the more-individuals hypothesis' inherently suggests a positive relationship between the abundance of taxa and available energy. Second, richness is expected to decrease as the distance to land increases because coastal margins may be a source of diversity to the open and deep ocean [6,22]. Third, greater environmental stability, proxied here as both the annual range in surface temperature and primary productivity, is hypothesized to be positively related to taxonomic richness [23] which could result from stable environments having species with more localized ranges resulting in higher species richness [24]. Finally, human activities can reduce animal abundance and diversity [25] and we used the ocean health index [26], which incorporates harvesting of organisms, global warming and transportation, as a proxy to assess the relationship

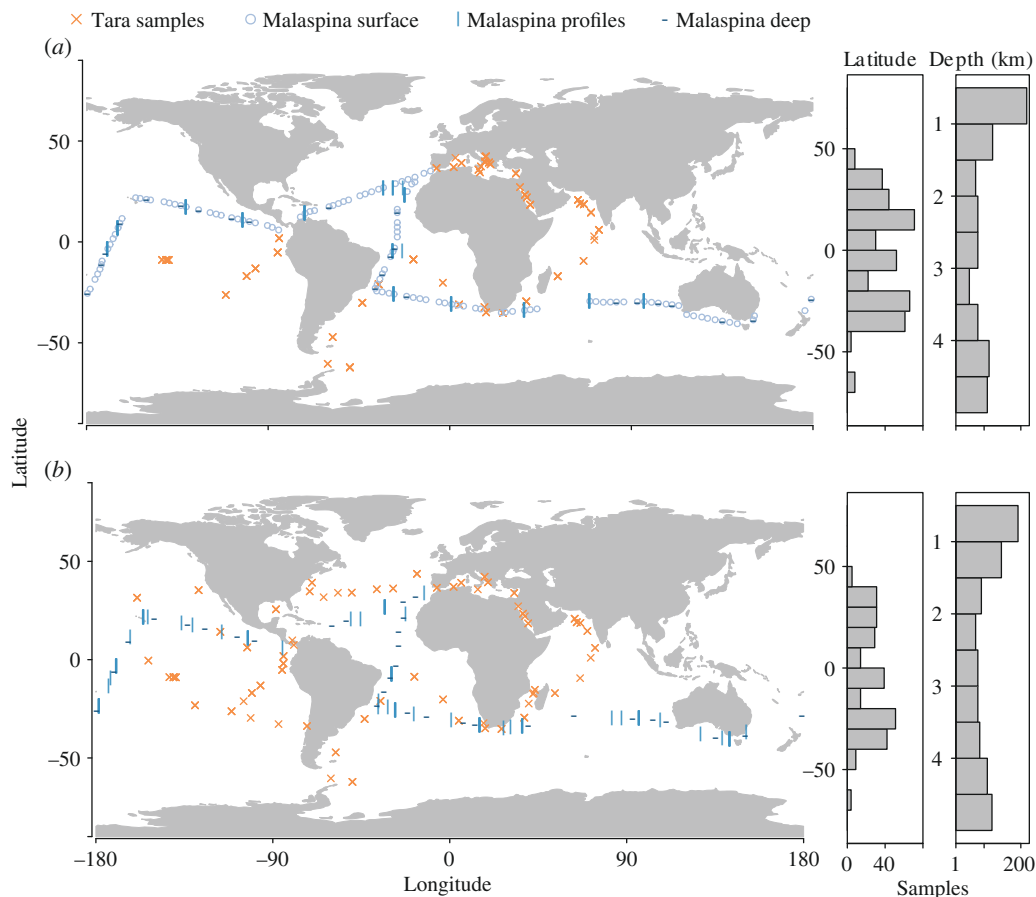


Figure 1. Sample locations for *Tara Oceans* and *Malaspina* global expeditions with sample distribution along the latitude and depth are shown in the right panels for metabarcode (a) and metagenome (b) datasets.

between anthropogenic impacts and metazoan eDNA. Thus, we assessed the patterns in taxonomic richness and the abundance of marine metazoans from eDNA and the associations of these patterns with environmental and human variables.

2. Material and methods

2.1. Method overview

This study used metabarcode and metagenome data. These two data types were collected along the survey tracks of both the *Tara Oceans* and the *Malaspina* expeditions (figure 1). Although it is generally accepted that metabarcoding is appropriate for assessing taxonomic richness, biases associated with PCR make metabarcoding's relationship with abundance less direct and context dependent [27,28]. However, metagenomes utilize shotgun sequencing and may have a better relationship between the abundance of DNA in the environment and post-sequencing data. As a result, we conducted an initial step of comparing results from the two genomic dataset types with morphology-based surveys conducted during the *Malaspina* cruise to determine which genomic data type, if either, relates to abundance and biomass determined by more traditional methods. In light of the results detailed below, we used metabarcode datasets to assess taxonomic richness and metagenome datasets to assess the abundance of marine metazoans. We used these datasets to describe basic patterns in the richness and abundance of the most prevalent metazoan phyla over the ocean's surface and in the deep ocean, while we used the cumulative taxonomic richness and abundance of all metazoans to analyse relationships with environmental variables. Each expedition had different collection, sequencing and sequence analysis protocols. Thus, we describe and analyse each of the genomic datasets from each expedition separately to make the multiple protocols and findings clear and concise. The expeditions were conducted before eDNA was targeted and thus did not include field or extraction controls that

are optimally used to detect contamination and consideration given the lack of these controls in the discussion. However, the expeditions did use common protocols to reduce contamination of samples in the laboratory and details of protocols are given in respective references. For example, PCR controls were always included for each PCR batch and samples were not sequenced if there was amplification in the control. The goal of these expeditions was to maximize the spatial coverage of the collected samples and replicates of samples were not sequenced.

2.2. Metabarcoding methods

The *Tara Oceans* sequences are open access and detailed methods are available elsewhere [14,29]. The *Tara Oceans* expedition was conducted from September 2009 to March 2012 and 241 samples from 45 locations were collected, primarily from two or three depths at each location (<400 m depth; figure 1). Sampling occurred day and night and there was no indication that this affected taxonomic richness (supplementary text and electronic supplementary material, figure S1). Before each surface sampling occasion, water was pumped for 10 min before collecting a sample. Before filling containers with samples, they were rinsed with diluted bleach (approx. 1–10%), freshwater twice and collected water twice. After each use, all equipment was rinsed with fresh water. Rinsing equipment with fresh water was sufficient to minimize contamination among samples based on field blanks in a plankton study using similar methods [30]. Water samples were filtered to collect four different size fractions, 0.8–5, 5–20, 20–180 and 180–2000 μm and the filters were flash frozen (other filter size fractions collected during *Tara Oceans* were not used because of inconstant sampling). However, to make results consistent with the other expedition only data from the 0.8–5 μm size fraction were used unless otherwise indicated (see electronic supplementary materials for details). DNA was extracted and a 130 bp section within the V9 region of the 18S rDNA gene (forward/reverse primer-pair 1389F 5'-TTGTACACACCGCCC-3' and 1510R 5'-CCTTCYGCAGGTTACCTAC-3' [31]) was PCR amplified using 5 ng of total DNA template. Illumina compatible adapters with the NEBNext DNASample Prep Master Mix were used to index samples [29]. Sequencing was performed using the Genome Analyser Ix systems (Illumina, San Diego, CA, USA). Sequences were merged using custom fastx software (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Sequences were retained if they had forward and reverse reads and less than 1% error based on the expected error in 50 bp sliding window. Chimeras were removed using the USEARCH program [32]. The V9 sequencing dataset was clustered using the 'Swarm' approach [33]. Unique operational taxonomic units (OTU) were removed if they did not occur in at least two samples and had at least three sequences, and taxonomy was assigned using a custom V9_PR2 reference database, primarily containing SILVA data with the ggsearch36 program [14]. These filtering and clustering parameters were used because they were used in analysing these data previously [14,15]. We used clustering methods for all amplicon data because this was the method used previously for this data [14,15], and because OTUs were similar to sequence variants when relating diversity to environmental gradients [34].

The *Malaspina* circumnavigation expedition was conducted from December 2010 to July 2011 (figure 1 [35]), and three separate datasets were collected: surface (119 samples from 119 stations taken at approximately 3 m depth, Logares *et al.* [36]), vertical profiles (86 samples from 13 stations [16]) and deep (26 samples from 26 stations taken at >2000 m [17]). Sampling for *Malaspina* began around 06.00 and finished by 09.00 local time. Niskin bottles were closed at the deepest sample (typically 4000 m depth for deep and profile datasets) and then sequentially shallower samples for the profile dataset. All equipment used during sampling was rinsed several times with MilliQ water and then several times with diluted HCl in between each sample and at the end of the day. Each station within the *Malaspina* vertical profiles usually included seven samples, one from the surface (approx. 3 m), one from the DCM (deep chlorophyll maximum), 2–3 from 200 to 1000 m and 2–3 samples from 1000 to 4000 m. Water was prefiltered through a 200 μm mesh and then sequentially filtered through 20, 3 and 0.2 μm filters (the intermediate filter was 0.8 μm instead of 3 μm for deep samples), which were then flash frozen.

DNA and RNA (RNA only from vertical profile samples) were simultaneously extracted using the Nucleospin RNA kit (Macherey-Nagel) with the NucleoSpin RNA/DNA Buffer Set (Macherey-Nagel) procedures. Samples included the smallest size fraction in surface samples and vertical profiles (0.2–3 μm) and the intermediate size fraction in deep samples (0.8–20 μm). RNA was transcribed to cDNA and the same procedures were followed for both DNA and cDNA. PCR amplification was conducted on a 380 bp section within the V4 region of the 18S rRNA gene (forward/reverse primer-pair TAREukFWD1 5'-CCAGCASCYGC GGTAATTCC-3' and TAREukREV3 5'-ACTTTCGTTCTTGATYRA-3'

[37]) using 5 ng of the total DNA template. Samples were indexed with Illumina sequencing adapters (i5 and i7) from the Illumina Nextera index kits. Amplicons for the surface and vertical profile samples were sequenced on an Illumina MiSeq platform (2 × 250 bp) and the PCR and sequencing were run at a specialized sequencing company (Research and Testing Lab, USA). Sequences were processed using a custom pipeline [38]. The paired-end reads were merged with PEAR [39], dereplicated and clustered with UPARSE at 99% similarity [40]. The deep samples differed from the other two sample sets and were sequenced on a 454 Life Sciences platform and clustered at 99% similarity using QIIME Uclust [41]. For all *Malaspina* data, taxonomy was assigned using BLAST search on the SILVA 18S database (v. 128 for surface and vertical profiles and 132 for deep samples). Taxonomy assignment to a reference sequence was based on the greatest per cent identity, the max score and then the total score.

Only the sequences (from both expeditions) with taxonomy assigned to Metazoa at a per cent identity match above 90% were retained for analysis. This is greater than past studies using global datasets (80% [14]). Ninety per cent was chosen as a balanced approach to minimize false positives in taxonomic assignment while reducing false negatives by identifying species or taxa not in the reference library, particularly for taxa from the open and deep ocean that are not well represented in reference libraries. All reads were summed for identical taxa (sequences with identical taxonomy) and taxonomic richness was the number of unique taxa for each sample. We present results at higher taxonomic levels (order or higher) given that the *Malaspina* pipeline did not use a common ancestor algorithm, as well as the rare chance that all three BLAST assignment scores were identical.

2.3. Metagenome methods

Samples for the metagenome datasets were collected following similar protocols to the metabarcoding datasets and were detailed for *Tara Oceans* [42–44] and the *Malaspina* expeditions [17,45]. There were 200 *Tara Oceans* metagenome samples available in this dataset from 67 locations and 138 unique sampling points, which included three size fraction categories 0.1–0.22, 0.22–3 (including 0.22–0.45, 0.22–1.6, 0.45–0.8 and 0.22–3 µm) and >0.22 µm. These three size categories included 19, 156 and 20 samples, respectively. There were 176 *Malaspina* metagenome samples included in this dataset with deep samples from 31 locations and profile samples from 35 locations and 101 unique location and depth sampling points, which included two size fractions for deep (0.2–0.8 and 0.8–20 µm) and for profile (0.2–3 and 3–20 µm) datasets. Libraries were created and sequenced using Illumina (GAIIx, HiSeq).

All metagenomes were re-annotated, standardized (among samples and calculated as sequences per million), and filtered using the Dragon Metagenomics Analysis Platform (DMAP). Gene abundance estimates from *Tara* samples were used without any modifications. However, for *Malaspina* datasets, we obtained normalized read counts considering the gene length and scaling of sequencing depth to one million reads, represented as fragments per kilobase of transcript per million mapped reads. DMAP assigns taxonomy to sequences from the expeditions using the UniProt database with high-throughput BLASTp using traverse lowest common ancestor from top hits. We searched the metagenome data for single-copy protein-encoding genes, or marker genes, that are taxonomically distinct. We conducted the BLASTp search with three different percent identity matches (PID), 50, 70 and 90. In addition, the search included three separate queries for different marker gene groups to assign sequences to metazoans. Specifically, three different searches of documented [46] enzyme encoding (ec) and mitochondrial domain (pf7) marker genes (search name followed by the exact genes in parentheses) included, ec3 (ec_id:1.6.5.3 OR ec_id:2.7.7.6 OR ec_id:6.3.2.2), ec4 (ec_id:1.6.5.3 OR ec_id:2.7.7.6 OR ec_id:2.7.11.1 OR ec_id:6.3.2.2) and pf7 (hmm_id:PF00189 OR hmm_id:PF01479 OR hmm_id:PF00411 OR hmm_id:PF01248 OR hmm_id:PF00181 OR hmm_id:PF01929 OR hmm_id:PF00252). Thus, nine distinct searches to assign taxonomy to sequences were performed for all metagenome datasets. We chose the query for all analyses (PID and marker gene combination) that had the greatest correlation coefficient (Spearman rank correlation) with morphology-based surveys (see electronic supplementary materials for survey methods). To determine the taxonomic breadth of the reference library, which can indicate biases in the results, the number of sequences for each phylum was obtained for each marker gene query from the UniProt database (<https://www.uniprot.org>; searched on 8 July 2019). These counts include all sequences as filtering out multiple occurrences of the same species and removing terrestrial species was not feasible with this database.

2.4. Statistical analysis

To determine what factors were related to changes in richness and relative abundance of metazoans based on eDNA, we analysed taxonomic richness using metabarcode datasets and abundance using metagenome datasets using either linear models (LM), linear mixed models (LMM) or spatial autoregressive lag and error models (SAR; see electronic supplementary materials for model choice and details). We conducted statistical analyses on the two response variables using eight different datasets: *Tara Oceans* metabarcode and metagenome datasets, *Malaspina* surface metabarcode dataset, *Malaspina* vertical profile DNA and RNA metabarcode datasets, *Malaspina* vertical profile metagenome dataset and *Malaspina* deep metabarcode and metagenome datasets. All analyses were conducted in R, v. 3.4.2 [47]. The datasets from each cruise were analysed separately because they used different methods and targeted different oceanic zones. For metabarcode datasets, the number of reads for each species within each sample was rarefied to the lowest number of reads within a sample for each dataset; 201 505 reads for *Tara Oceans*, 6461 for all *Malaspina* DNA and 21 825 for *Malaspina* vertical profile RNA (rarefy function from vegan package [48]). Rarefaction was chosen because it is a straightforward approach that can be more appropriate than normalization [49]. In addition, to reduce overestimating richness and to be conservative in our calculation of the number of species, reads from different OTUs assigned to the same species were summed and any reads were removed that did not include a species label because it was missing in the matching reference sequence. Finally, the analysis only included samples from similar size fractions (within approx. 0.2–5 μm) to make findings comparable among datasets and reflect findings from morphometric samples.

Predictor variables, chosen *a priori* to be included in statistical models, were sea surface temperature (SST) mean, SST range, surface primary productivity mean, surface primary productivity range, distance to land and the ocean health index. The mean and ranges of SST and primary productivity were extracted from Bio-ORACLE layers that are based on satellite data from 2000 to 2014 [50]. SST and primary productivity ranges were determined by calculating the difference between minimum and maximum temperature for each year of the dataset and then using the mean of these annual ranges. Distance to land was obtained from the global marine environment datasets [51]. The ocean health index is a standardized index of human influence based on many anthropogenic impacts and was used as an overarching proxy for anthropogenic impacts [26]. Details on why we chose surface variables and layer extraction, as well as details about additional analyses on the influence of latitude and the relationship between taxa richness and abundance, are provided in the electronic supplementary material.

3. Results

3.1. Richness of taxa from metabarcode datasets

Patterns of taxonomic richness within individual metazoan phyla were different at the surface and in the deep ocean. Arthropoda was the phylum with the highest taxonomic richness at the surface of the ocean, especially for the *Tara Oceans* dataset (figure 2a,b). The arthropod taxa were primarily from the order Calanoida. Arthropoda included taxa from six additional orders when limited to orders with a mean richness of greater than one per sample (figure 2f). Other phyla with high taxonomic richness included Chordata and Cnidaria. The Chordata taxa were primarily from Doliolida and Salpida orders, while the Cnidaria taxon was primarily from Siphonophorae (figure 2f,g). In comparison with the *Malaspina* surface data, the *Malaspina* profile data had lower richness within Arthropoda and Chordata, but similar richness of Cnidaria (figure 2c,h). *Malaspina* profile eDNA and eRNA had similar patterns of richness for both phyla and orders (figure 2c,d,h,i). The *Malaspina* deep data had greater taxonomic richness within Cnidaria compared with the other metabarcode datasets, which were primarily in Siphonophorae and Anthoathecata orders (figure 2e,j). There were changes in the richness of taxa within the major phyla as depth increased within the *Malaspina* profile dataset (figure 3a). The richness of arthropod taxon decreased with depth, while there was a peak in the richness of cnidarians within the mesopelagic. *Malaspina* datasets had fewer sequences per sample and a lower percentage assigned to metazoans (0.12% of approximately 97 000 sequences per sample) compared with the *Tara Oceans* (42% of approximately 1.7 million sequences per sample), and this probably resulted in the lower taxonomic richness and why we rarefied and analysed the datasets separately (see electronic supplementary materials for details of sequence depth and assignment). This pattern was counter to the *in*

silico results (the supplementary text and electronic supplementary material, figure S2) for which the primers used for *Malaspina* amplified more metazoans, but the percentage of metazoan sequences is dependent on the amplification of other organisms as well and the primers used for *Malaspina* probably amplified more microorganisms and at a greater frequency compared with the primers used for *Tara Oceans*.

3.2. Relative abundance of taxa from metagenomic datasets

Before we examined the patterns of relative abundance based on eDNA, we needed to determine which genomic protocols to use based on the correlation coefficients between the results of the different protocols and morphological surveys. We found that the most appropriate genomic protocol to use for eDNA-based abundance estimates was metagenomes with a per cent identity matching of >70%, the ec4 marker gene query and only samples from filtered size fractions between 0.2 and 3 μm (electronic supplementary material, figures S3–S6). The sequences identified with this marker gene query and per cent identity match had an $r^2 = 0.73$ with morphological abundance and an $r^2 = 0.45$ with morphological biomass (electronic supplementary material, figure S5). These relationships were similar to or better than the relationship between abundance and biomass of the morphological surveys ($r^2 = 0.48$) and much better than the relationship between metabarcoding sequences and either of the morphological surveys ($r^2 < 0.11$). Although the pf7 query had slightly higher r^2 values for biomass but not abundance compared with ec4 (electronic supplementary material, figure S3), this query resulted in assigning almost all sequences of the *Tara Oceans* dataset to Mollusca.

The metagenome datasets included 74 species assigned to sequences. This comprised 13 species of arthropods with the majority of sequences assigned to Branchiopoda (>90% of arthropod and 44% of metazoan sequences), 12 cnidarian species with the majority of sequences assigned to Anthozoa (>90% cnidarian and 16% metazoan sequences) and 10 species of nematodes with all sequences assigned to Chromadorea (27% of metazoan sequences).

Metagenomes indicated similar median relative abundance patterns among phyla at the surface, for depth profiles, and the deep ocean, as well as for different ocean basins and latitudinal zones, with a few exceptions (figure 4). Arthropods consistently made up 50% of the metazoan eDNA, regardless of dataset or location. Cnidarians were the second most abundant phylum and made up 10–20% of the metazoan sequences. One exception was the *Malaspina* deep samples for which cnidarians made up approximately 30% of the metazoan eDNA. These deep samples were slightly different from the profiles, which indicated that arthropods and cnidarians were more abundant at these depths than nematodes (figure 4*b,c*). Overall, nematodes were the third most abundant phyla and were most abundant in the *Malaspina* profile samples. All other metazoan phyla had eDNA that was undetected in many samples (had medians of zero) with the exceptions of Placozoa and Tardigrada within the *Tara Oceans* dataset, and Chordata and Mollusca within the *Malaspina* profile dataset. At the ocean's surface, arthropod eDNA (primarily zooplankton) was the most abundant of all metazoan phyla, followed by Cnidaria and Nematoda. However, these three phyla had similar abundances in the bathypelagic zone (figure 3*b*).

3.3. Environmental relationships with richness from metabarcode datasets

Overall, we found that temperature and primary production were related to the taxonomic richness of metazoans, however, there were differences among datasets in regard to specific significant predictor variables. The taxonomic richness of all metazoans from the *Tara Oceans* was positively related to the surface primary productivity mean and negatively related to the primary productivity range (electronic supplementary material, figure S7*a* and *b*), but there were also two interactions: one between mean primary productivity and ocean health index and the other between primary productivity range and distance to land (electronic supplementary material, table S1). The interaction between primary productivity and ocean health index was caused by the positive relationship between primary productivity and richness being most evident in areas of greater human impact. The negative relationship between primary productivity range and richness primarily occurred far from the land. The taxonomic richness of metazoans from the *Malaspina* surface dataset was also negatively related to the primary productivity range (electronic supplementary material, figure S7*e*), and positively related to the SST mean and range (electronic supplementary material, figure S7*c,d*). The *Malaspina* profile dataset indicated a positive relationship between richness and primary productivity for both

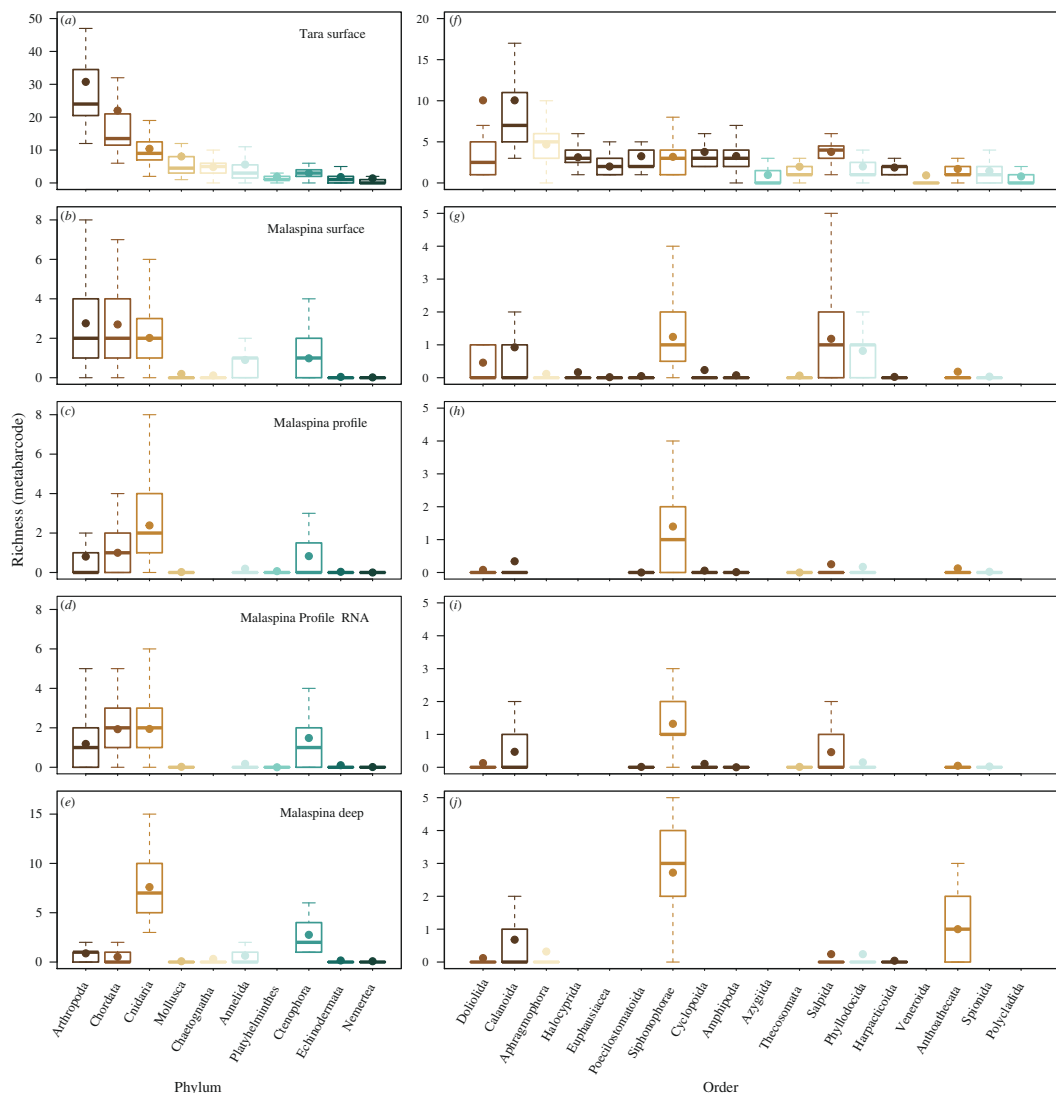


Figure 2. Taxonomic richness of metazoan phyla (a–e) and orders (f–j) with the greatest taxonomic richness from metabarcode data. Phyla or orders are organized from greatest to least rich (means). Plots show *Tara Ocean* samples (a and f), *Malaspina* surface samples (b and g), *Malaspina* profile samples for eDNA (c and h) and eRNA (d and i) and *Malaspina* deep samples (e and j). Only data from comparable filtered size fractions were included (0.2 – 0.8 and 0.2 – 3 μm size fractions for *Malaspina* datasets, and approx. 0.2 – 4.5 μm size fraction for *Tara Oceans*). The colour of the boxes in the second column indicates what phylum the order is in following the left column. Boxplots, based on per sample taxonomic richness, show the median with the upper and lower quartiles, while the whiskers extend to the extreme data point but no more than 1.5 times the respective quartile. The mean is indicated by the circle.

eDNA and eRNA (electronic supplementary material, figure S7f, g), while eRNA also indicated a negative relationship between richness and primary productivity range (electronic supplementary material, figure S7h). There were no significant relationships between taxonomic richness and predictor variables for the *Malaspina* deep dataset (electronic supplementary material, table S1), which included the auxiliary model that included the additional variables, *in situ* temperature and POC flux, at 2000 m. There were no significant relationships between the taxonomic richness and distance to land or ocean health index for any of the datasets. The explanatory power of these models differed among datasets (electronic supplementary material, table S1). The *Tara Oceans*, *Malaspina* surface, and *Malaspina* deep had r^2 values of 0.59, 0.47 and 0.44, respectively, while the *Malaspina* profile datasets had a low r^2 value (<0.15).

3.4. Environmental relationships with eDNA relative abundance from metagenome datasets

The relative abundance of metazoan eDNA within metagenome datasets was negatively related to depth and SST mean for the *Tara Oceans* metagenome dataset (electronic supplementary material,

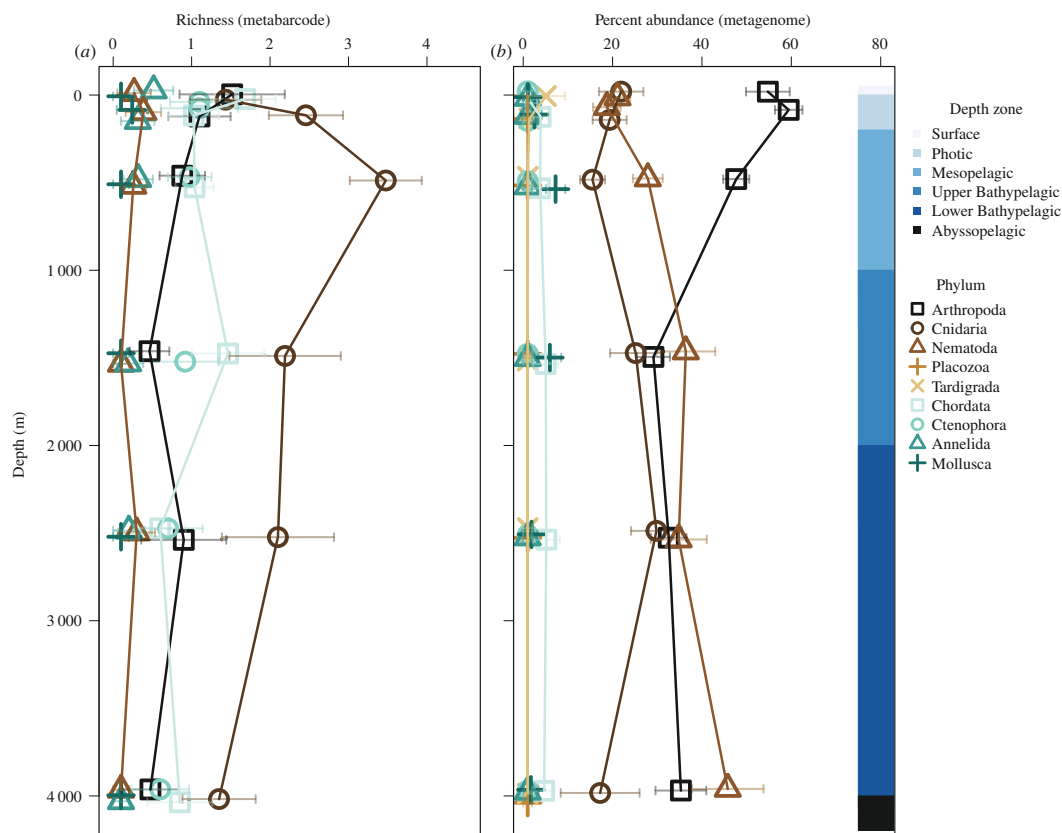


Figure 3. Taxonomic richness (a) and per cent abundance (b) of sequences within metazoan phyla from the ocean's surface to the abyssopelagic. Taxonomic richness was from metabarcoding data and abundance was from metagenome data, all from the *Malaspina* profile dataset (0.2–3 μm filtered size fractions). Phyla are indicated by shape and colour. The mean taxonomic richness or per cent abundance of samples within the respective categories is shown with symbols along with the standard error indicated by horizontal bars.

table S1 and figure S7*i,j*). These same patterns were evident in the *Malaspina* deep dataset (p -value < 0.05), but the *Malaspina* profile dataset had no significant relationships between eDNA relative abundance and predictor variables. The explanatory power of the models relating metazoan abundance to environmental variables was the highest for the *Malaspina* deep dataset ($r^2 = 0.75$), followed by the *Tara Oceans* ($r^2 = 0.36$) and *Malaspina* deep datasets ($r^2 = 0.16$; electronic supplementary material, table S1). Finally, we found no relationship between the richness and relative abundance of eDNA for any of the datasets with both metabarcoding and metagenome data (electronic supplementary material, figure S8).

4. Discussion

4.1. Global patterns of richness for metazoan phyla eDNA

Given the global coverage of these two expeditions, their combined information probably constitutes the most comprehensive dataset of metazoan eDNA retrieved from the open and deep ocean. These data offered a unique opportunity to enhance our understanding of the distribution and drivers of metazoans throughout the oceans based on eDNA. Patterns at the surface of the ocean indicate that Arthropoda was the most taxonomically rich phylum followed by Chordata and Cnidaria. In addition, the taxonomic richness of some gelatinous phyla increased with depth, including cnidarians, chordates and ctenophores. These basic patterns in the taxonomic richness of metazoan phyla enhance our knowledge of marine diversity from the ocean's surface to the lesser-studied deep ocean.

Biodiversity and abundance are patchy over the surface of the Earth, particularly, in the oceans, and determining the drivers of such patterns is an objective of ecology because it underpins our ability to understand and mitigate anthropogenic impacts on ecosystems [25]. Multiple theories of biodiversity

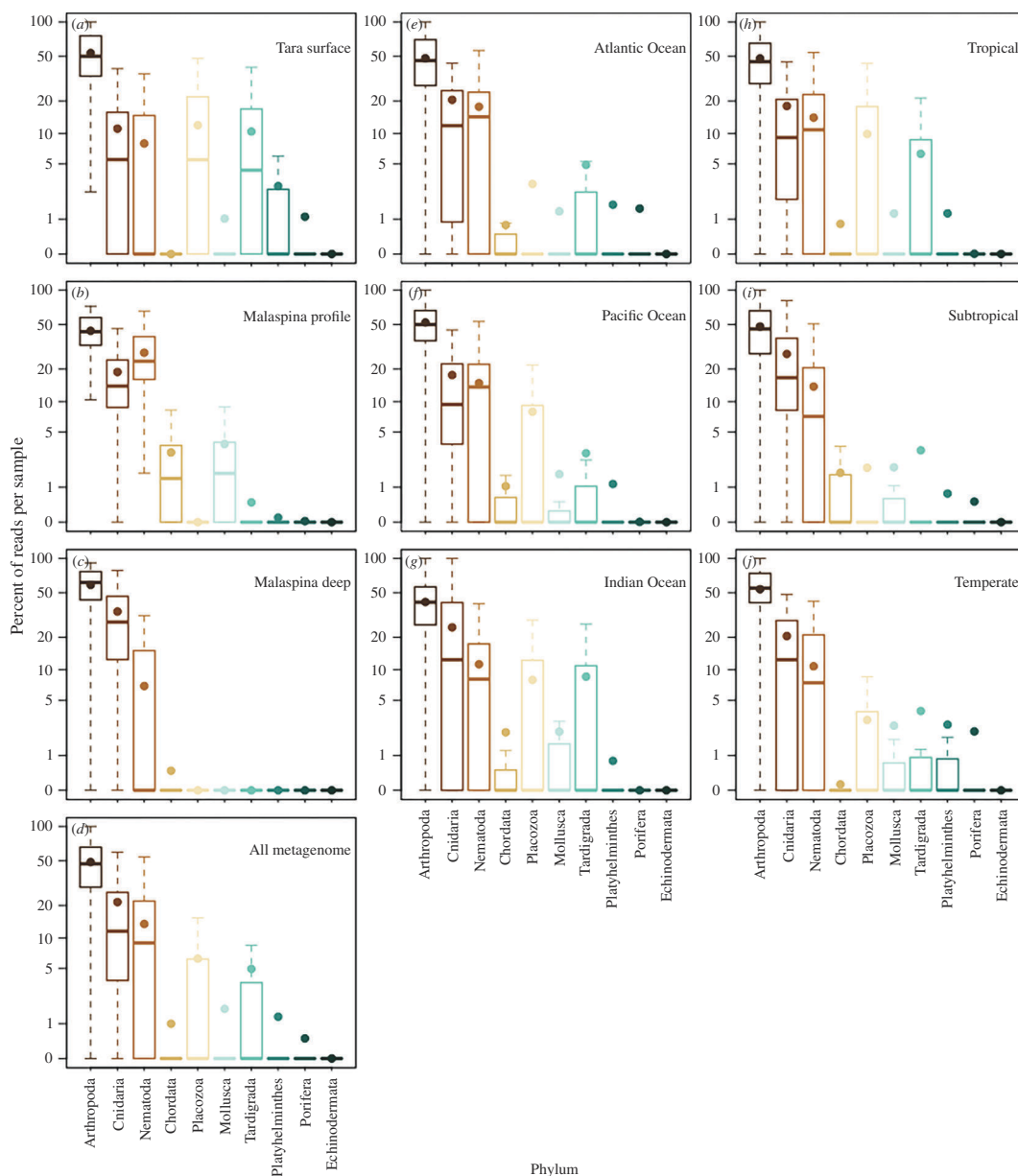


Figure 4. Mean abundance (expressed as the percentage of reads per sample of metazoan marker gene sequences from the metagenomic dataset) of the taxa identified by the two expeditions. Phyla shown are sorted based on highest medians. Plots show *Tara Ocean* samples (a), *Malaspina* profile samples (b) and *Malaspina* deep samples (c), all metagenome samples (d), all metagenome data from different ocean basins (e–g) and all metagenome data by latitudinal categories (h–j). Boxplots, based on per sample data across all samples, show the median with the upper and lower quartiles, while the whiskers extend to the extreme data point but no more than 1.5 times the respective quartile. The mean is indicated by the circle. Only data from comparable filtered size fractions were included (0.2–0.8 and 0.2–3 μm size fractions for *Malaspina* deep and profile datasets respectively and approx. 0.2–4.5 μm size fraction for *Tara Oceans*).

patterns focus on greater species richness being related to higher temperature and energy [20,21,52]. These hypotheses were supported by broad taxonomic assessments of global taxonomic richness, with temperature being the only variable that was positively related to species richness across multiple taxonomic groups [25,53] and a primary driver for taxonomic richness of copepods identified within *Tara Oceans* metabarcoding data [15]. However, the deep ocean may be an exception [53].

This assessment of metazoan taxonomic richness based on metabarcoding indicated that yearly mean SST and primary productivity were positively related to the taxonomic richness of metazoans, while greater ranges of primary productivity were negatively related to richness for the surface of the ocean (electronic supplementary material, table S1 and figure S7). The consistent findings for surface and profile samples from both expeditions, which had different sampling and processing methods,

give additional credence to our findings on oceanic patterns—particularly for the *Malaspina* dataset which had limited metazoan reads and low richness per sample (a mean of approx. 10 species per sample). In addition, the relationship between taxonomic richness and environmental variables was consistent between metabarcoding based on DNA and RNA. Although the relationship between eDNA and eRNA is complex, eRNA can be a more direct indicator of the presence of organisms, given that eRNA degrades quicker than eDNA in the environment [16,54]. These findings agree with past global assessments in that mean temperature is a primary driver of species richness [25,53] and that primary productivity is related to the diversity of copepods based on metabarcode data [15]. It is important to consider that our results are based on sampling locations at a single time point, which could influence our findings for areas with high ranges in the predictor variables, as this could also relate to changes in taxonomic richness.

Knowledge of the drivers of global metazoan taxonomic richness in the open ocean, from the surface to the deep sea, is minimal and resource pulses can both increase and decrease diversity [55]. In addition, a study on ophiuroids (brittle stars) throughout the oceans found that deep-sea species richness was positively related to high carbon export from surface primary production and proximity to continental margins, and negatively related to variation in surface primary productivity [6]. Although we found similar relationships with oceanic profiles, we found no support for a relationship between surface primary production and the taxonomic richness of metazoans in the deep ocean. Overall, our findings agree with global assessments that species richness and temperature are positively correlated at the surface of the ocean [25] and that temperature and primary production are positively related to taxonomic richness when samples are taken throughout the water column [6].

We also assessed the influence of latitude (electronic supplementary materials), along with the environmental variables already discussed, on metazoan taxonomic richness. Based on path analysis, SST mean had the strongest causal linkage with richness followed by latitude, and these variables also had a strong causal linkage, indicating that although these variables are correlated, richness is primarily driven by SST and then secondarily by latitude (electronic supplementary material, figure S8). In addition, there were no clear latitudinal trends for either taxonomic richness or the abundance of marine metazoans (electronic supplementary material S9), which is contrary to existing findings using disparate observations through time [53] and to *Tara Oceans* data showing peaks of copepod richness at low latitudes [15]. Perhaps these differences could be a result of single samples versus a culmination of many samples through time and that patterns in specific taxa may not persist in metazoans overall.

4.2. Global patterns of metagenomic sequence abundance

Quantification of the global abundance and biomass of organisms is needed for understanding large-scale bio-geochemical cycles and changes in ecosystem structure resulting from anthropogenic impacts. Therefore, finding novel methods to assess abundance can improve our estimates by complementing traditional survey techniques. With regard to eDNA, many studies have found a relationship between the abundance and biomass of organisms and the number of sequences, but there are multiple variables that may alter this relationship [27,56,57]. We found a significant relationship between the relative abundance of marker gene sequences within metagenome datasets and the abundance and biomass measured by morphological methods within metazoan phyla that were identified in both techniques (figure 2a,b). We did not find a relationship between biomass and metabarcoding data, which is different from some past studies [14,58], although this relationship is probably taxa dependent [59].

Our findings from metagenome datasets indicate that arthropods, largely represented by crustacean zooplankton, composed approximately 50% of the eDNA, followed by Cnidaria (approx. 11%) and Nematoda (approx. 8%), based on the median per cent of eDNA from marker gene sequences (figure 4d). A global estimate of ocean biomass found that arthropods (1.0 Gigatons of Carbon (Gt C)) and fish (0.7 Gt C) had the greatest biomass, followed by gelatinous cnidarians, mollusks and nematodes (0.04, 0.02 and 0.01 Gt C, respectively; [60]). The rank of these biomass estimates compares well with our relative sequence abundance, with the exception of more nematodes and fewer fish. The abundance of chordate DNA (mean of <1%) was orders of magnitude lower than would be expected based on the global biomass estimate (assuming fish biomass to be approximately 70% of arthropod biomass, which would equate to fish being 35% of all metazoan reads). The low relative abundance of fish eDNA is surprising given that the mesopelagic fish biomass may be orders of magnitude greater than previously thought [61]. However, only 0.3% of the DNA sequences were assigned to bony fish.

This was unlikely a result of few reference sequences, given that more than one-third of the chordate sequences within the reference database were fish. Another plausible explanation for this discrepancy could be that fish release less eDNA compared with other taxa such as jellyfish [62], resulting in an underestimation of fish abundance based on eDNA. A study using metabarcoding of a section of the CO1 gene also found a low detection of oceanic fishes [30]. An alternative explanation is that the biomass of other phyla is greater in comparison with the estimated fish biomass given the disproportionate sampling of the continental shelf compared with open and deep ocean surveys [5], and the difficulty in sampling soft-bodied and cryptic taxa. In addition, nematode eDNA was prevalent in metagenomes and this is discussed in the electronic supplementary material. Finally, our eDNA-based study suggests that cnidarians had approximately one-fourth of the eDNA of arthropods (based on medians; figure 3d), which would suggest that the global abundance of cnidarians may be greater than current existing estimates. These findings offer the first assessment of the relative abundance of animal phyla in the open ocean throughout the globe, based on a single assessment method.

Our ability to measure the abundance of gelatinous taxa may also be enhanced through genomic-based surveys. Delicate gelatinous organisms are notoriously difficult to sample, and the evidence that they have a major role in deep-sea food webs has been, thus far, derived indirectly from stomach content studies [7] or anecdotal observations from submersibles and cameras [8] in specific locations. Our results indicate that the abundance of cnidarian eDNA peaked in the bathypelagic zone (figure 3b). The greater abundance of gelatinous cnidarians in the deep ocean (bathypelagic zone) was also evident in the *Malaspina* deep samples, in which cnidarians made up almost one-third of metazoan sequences (figure 3b–c). This suggests that cnidarians are particularly prevalent in the deep sea, consistent with inferences derived from indirect sampling [7,8]. Past observations and findings from this study indicate that these gelatinous animals may be more abundant than previously thought in the open ocean and particularly in the deep sea.

The abundance of animals within metazoan phyla may also relate to environmental variables. This follows the ‘the more-individuals hypothesis’ which suggests that more energy (temperature or resources) will lead to more individuals, which in turn will result in more species [21], suggesting that taxonomic richness and abundance may have similar drivers. The abundance of metazoan eDNA was negatively related to depth and mean SST for the *Tara Oceans* samples and a similar trend was found for *Malaspina* deep samples (electronic supplementary material, table S1 and figure S7). Overall, these results agree with current assumptions that the biomass of animals decreases with depth but disagree because abundance did not increase with temperature [20,21]. In addition, comparing the richness and abundance of simultaneously sampled metabarcode and metagenome data indicates little or slightly negative relationship between abundance and taxonomic richness (electronic supplementary material, figure S8). The negative relationship between temperature and eDNA may result from greater preservation of eDNA in colder water [63], which could also affect the relationship between eDNA taxonomic richness and abundance.

4.3. Strengths and potential limitations of these datasets and eDNA surveys

Although using a single method enhances comparability among samples compared with using multiple methods, all sampling methods have imperfect detection [57] and a better understanding of biases when using eDNA to survey metazoans is needed. Results from eDNA studies are not based on counting or weighing individuals but on the DNA left from individuals and offer a ‘DNA view’ of the ocean. Sources of eDNA are being investigated and insights from different filtered size fractions can help elucidate eDNA origins (see electronic supplementary materials for a discussion on filtered size fractions).

There are multiple considerations when interpreting eDNA-based datasets. First, results from eDNA studies are limited by the coverage of the reference library used to assign DNA sequences to taxa. The metabarcode reference library did contain many taxa within almost all the major metazoan phyla with the exception of Nemertea, Gastrotricha and Ctenophora and the metagenome database did not have thaliaceans (salps and doliolids; electronic supplementary material, figure S2 and supplementary text for a detailed discussion on considerations). Improvements in reference databases will increase the diversity detected by eDNA-based surveys. A second consideration with eDNA-based studies is removing false positives while minimizing false negatives. This is particularly important for this study because these datasets were collected before field and extraction controls were common in molecular studies, although PCR controls were run for all datasets and samples were only sequenced if there was no indication of amplification in these controls (see electronic supplementary materials

for a detailed discussion on considerations and potential biases of eDNA-based surveys). It is possible that some of the taxa identified and used in our analysis were not present in samples and resulted from contamination. However, these taxa were probably a very small percentage of the total taxonomic richness and we chose our analysis to focus on broad taxonomic groups and the relationship between basic metrics (taxonomic richness) and environmental variables so that the effect of potential contamination would not affect our conclusions, which are extremely valuable given the unprecedented breadth of these data. Although we took many steps to minimize potential biases associated with genomic datasets so that taxonomic richness patterns based on eDNA reflected patterns of actual species richness, the findings should be interpreted knowing that they are based on diversity inferred from 18S rRNA sequences.

5. Conclusions

Ocean ecosystems are experiencing significant changes because of anthropogenic impacts [1,2], and determining marine taxa occurrence and environmental properties related to diversity and abundance are needed, especially for the deep ocean [3,4]. Surveys based on eDNA offer a complementary tool to traditional techniques to survey metazoans, particularly for species at low densities or with fragile body forms, as well as in locations difficult to sample such as the deep ocean. Taxonomic richness for all metazoans was positively related to mean SST and mean primary productivity, but negatively related to the range of primary productivity. However, the association with mean SST was not evident for profile samples and no relationship with any of the environmental variables was found for the deep ocean. In general, these findings based on eDNA agree with past findings focusing on specific groups of metazoans and offer a promising tool to survey ocean fauna.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. Sequences and data from the Tara Oceans are available at <http://taraoceans.sb731.roscoff.fr/EukDiv/>, at EBI under the project IDs PRJEB402 and PRJEB6610, and at PANGAEA. Sequences are available for Malaspina deep samples at the European Nucleotide Archive (<http://www.ebi.ac.uk/ena/data/view/PRJEB9943>), from Malaspina surface samples at the European Nucleotide Archive (<http://www.ebi.ac.uk/ena/data/view/PRJEB23913>), and for the Malaspina vertical profile samples at the European Nucleotide Archive (<http://www.ebi.ac.uk/ena/data/view/PRJEB23771>). R code used for analysis and figures are available at github [64] and data used in the R code are available on Zenodo [65].

Supplementary material is available online [66].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. N.R.G.: conceptualization, data curation, formal analysis, methodology, supervision, visualization, writing—original draft; S.G.A.: project administration, resources, writing—review and editing; I.A.: data curation, formal analysis, methodology, software, writing—review and editing; J.M.G.: investigation, resources, writing—review and editing; M.L.F.: data curation, resources, writing—review and editing; S.H.-L.: resources, writing—review and editing; R.L.: resources, writing—review and editing; R.M.: resources, writing—review and editing; P.S.: resources, writing—review and editing; V.B.: resources, writing—review and editing; T.G.: methodology, writing—review and editing; C.M.D.: conceptualization, funding acquisition, investigation, resources, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. No funding has been received for this article.

Acknowledgements. The Malaspina expedition samples were obtained thanks to a grant from Consolider-Ingenio. We thank the people participating in the Malaspina expeditions, in particular those obtaining and processing the DNA samples. Computation support was provided by the MarBits Platform at ICM-CSIC. We would like to thank the Tara Oceans scientists, particularly for making the data open access.

References

1. Diaz S *et al.* 2019 Pervasive human-driven decline of life on earth points to the need for transformative change. *Science* **366**, eaax3100. (doi:10.1126/science.aax3100)
2. 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)
3. Martin A *et al.* 2020 The oceans' twilight zone must be studied now, before it is too late. *Nature* **580**, 26–28. (doi:10.1038/d41586-020-00915-7)

4. Danovaro R *et al.* 2020 Ecological variables for developing a global deep-ocean monitoring and conservation strategy. *Nat. Ecol. Evol.* **4**, 181–192. (doi:10.1038/s41559-019-1091-z)
5. Webb TJ, Vanden Berghe E, O'Dor R. 2010 Biodiversity's big wet secret: the global distribution of marine biological records reveals chronic under-exploration of the deep pelagic ocean. *PLoS One* **5**, e10223. (doi:10.1371/journal.pone.0010223)
6. Woolley SNC, Tittensor DP, Dunstan PK, Guillera-Arroita G, Lahoz-Monfort JJ, Wintle BA, Worm B, O'Hara TD. 2016 Deep-sea diversity patterns are shaped by energy availability. *Nature* **533**, 393–396. (doi:10.1038/nature17937)
7. Hoving HJT, Haddock SHD. 2017 The giant deep-sea octopus haliphron atlanticus forages on gelatinous fauna. *Sci. Rep.* **7**, 44952. (doi:10.1038/srep44952)
8. Choy CA, Haddock SHD, Robison BH. 2017 Deep pelagic food web structure as revealed by *in situ* feeding observations. *Proc. Biol. Sci.* **284**, 20172116. (doi:10.1098/rspb.2017.2116)
9. UNEP, CBD and COP. 2010 The strategic plan for biodiversity 2011–2020 and the Aichi Biodiversity targets: living in harmony with nature. UNEP/CBD/COP/DEC/X/2. In *Decision adopted by the conference of the parties to the convention on biological diversity at its tenth meeting*. Nagoya.
10. UN General Assembly. 2015 Transforming our world: the 2030 Agenda for Sustainable Development, A/RES/70/1. <https://www.refworld.org/legal/resolution/unga/2015/en/111816>.
11. Taberlet P, Bonin A, Zinger L, Coissac E. 2018 *Environmental DNA: for biodiversity research and monitoring*. Oxford, New York: Oxford University Press. (doi:10.1093/oso/9780198767220.001.0001)
12. Deiner K *et al.* 2017 Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol. Ecol.* **26**, 5872–5895. (doi:10.1111/mec.14350)
13. Bik HM. 2014 Deciphering diversity and ecological function from marine metagenomes. *Biol. Bull.* **227**, 107–116. (doi:10.1086/BBLv227n2p107)
14. de Vargas C *et al.* 2015 Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605. (doi:10.1126/science.1261605)
15. Ibarbalz FM *et al.* 2019 Global trends in marine plankton diversity across kingdoms of life. *Cell* **179**, 1084–1097. (doi:10.1016/j.cell.2019.10.008)
16. Giner CR, Pernice MC, Balagué V, Duarte CM, Gasol JM, Logares R, Massana R. 2019 Marked changes in diversity and relative activity of picoeukaryotes with depth in the world ocean. *ISME J.* **14**, 1–13. (doi:10.1038/s41396-019-0506-9)
17. Pernice MC, Giner CR, Logares R, Perera-Bel J, Acinas SG, Duarte CM, Gasol JM, Massana R. 2016 Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J.* **10**, 945–958. (doi:10.1038/ismej.2015.170)
18. Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. 2004 Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789. (doi:10.1890/03-9000)
19. Brown JH. 2014 Why are there so many species in the tropics? *J. Biogeogr.* **41**, 8–22. (doi:10.1111/jbi.12228)
20. Srivastava DS, Lawton JH. 1998 Why more productive sites have more species: an experimental test of theory using tree-hole communities. *Am. Nat.* **152**, 510–529. (doi:10.1086/286187)
21. Storch D, Bohdalková E, Okie J. 2018 The more-individuals hypothesis revisited: the role of community abundance in species richness regulation and the productivity-diversity relationship. *Ecol. Lett.* **21**, 920–937. (doi:10.1111/ele.12941)
22. Holt RD. 1985 Population dynamics in two-patch environments: some anomalous consequences of an optimal habitat distribution. *Theor. Popul. Biol.* **28**, 181–208. (doi:10.1016/0040-5809(85)90027-9)
23. Fraser RH, Currie DJ. 1996 The species richness-energy hypothesis in a system where historical factors are thought to prevail: coral reefs. *Am. Nat.* **148**, 138–159. (doi:10.1086/285915)
24. Stevens GC. 1989 The latitudinal gradient in geographical range: how so many species coexist in the tropics. *Am. Nat.* **133**, 240–256. (doi:10.1086/284913)
25. Tittensor DP, Mora C, Jetz W, Lotze HK, Ricard D, Berghe EV, Worm B. 2010 Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098–1101. (doi:10.1038/nature09329)
26. Halpern BS *et al.* 2008 A global map of human impact on marine ecosystems. *Science* **319**, 948–952. (doi:10.1126/science.1149345)
27. Piñol J, Senar MA, Symondson WOC. 2019 The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Mol. Ecol.* 407–419. (doi:10.1111/mec.14776)
28. 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)
29. Alberti A *et al.* 2017 Viral to metazoan marine plankton nucleotide sequences from the tara oceans expedition. *Sci. Data* **4**, sdata201793. (doi:10.1038/sdata.2017.93)
30. Suter L, Polanowski AM, Clarke LJ, Kitchener JA, Deagle BE. 2020 Capturing open ocean biodiversity: comparing environmental DNA metabarcoding to the continuous plankton recorder. *Mol. Ecol.* **30**, 3140–3157. (doi:10.1111/mec.15587)
31. Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. 2009 A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One* **4**, e6372. (doi:10.1371/journal.pone.0006372)
32. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011 UCHIME improves sensitivity and speed of chimera detection. *Bioinforma. Oxf. Engl.* **27**, 2194–2200. (doi:10.1093/bioinformatics/btr381)
33. 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)
34. Clarke LJ, Beard JM, Swadling KM, Deagle BE. 2017 Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecol. Evol.* **7**, 873–883. (doi:10.1002/ece3.2667)

35. Duarte CM. 2015 Seafaring in the 21st century: the malaspina 2010 circumnavigation expedition. *Limnol. Oceanogr. Bull.* **24**, 11–14. (doi:10.1002/lob.10008)
36. Logares R, Deutschmann IM, Junger PC *et al.* 2020 Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* **8**, 55. (doi:10.1186/s40168-020-00827-8)
37. Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner HW, Richards TA. 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)
38. Logares R. 2017 Workflow for analysing miseq amplicons based on uparse. See <https://doi.org/10.5281/zenodo.259579>.
39. Zhang J, Kobert K, Flouri T, Stamatakis A. 2014 PEAR: a fast and accurate illumina paired-end read merger. *Bioinformatics* **30**, 614–620. (doi:10.1093/bioinformatics/btt593)
40. Edgar RC. 2013 UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998. (doi:10.1038/nmeth.2604)
41. Caporaso JG *et al.* 2010 QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336. (doi:10.1038/nmeth.f.303)
42. Pesant S *et al.* 2015 Open science resources for the discovery and analysis of tara oceans data. *Sci. Data* **2**, 150023. (doi:10.1038/sdata.2015.23)
43. Hingamp P *et al.* 2013 Exploring nucleo-cytoplasmic large DNA viruses in tara oceans microbial metagenomes. *ISME J.* **7**, 1678–1695. (doi:10.1038/ismej.2013.59)
44. Sunagawa S *et al.* 2015 Ocean plankton. structure and function of the global ocean microbiome. *Science* **348**, 1261359. (doi:10.1126/science.1261359)
45. Obiol A, Giner CR, Sánchez P, Duarte CM, Acinas SG, Massana R. 2020 A metagenomic assessment of microbial eukaryotic diversity in the global ocean. *Mol. Ecol. Resour.* **20**, 718–731. (doi:10.1111/1755-0998.13147)
46. Maier UG, Zauner S, Woehle C, Bolte K, Hempel F, Allen JF, Martin WF. 2013 Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. *Genome Biol. Evol.* **5**, 2318–2329. (doi:10.1093/gbe/evt181)
47. R Core Team. 2017 R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. See <https://www.R-project.org/>.
48. Oksanen J *et al.* 2018 *Vegan: Community Ecology Package*. See <https://CRAN.R-project.org/package=vegan>.
49. Cameron ES, Schmidt PJ, Tremblay BJM, Emelko MB, Müller KM. 2021 Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities. *Sci. Rep.* **11**, 22302. (doi:10.1038/s41598-021-01636-1)
50. Assis J, Tyberghein L, Bosch S, Verbruggen H, Serrão EA, De Clerck O. 2018 Bio-ORACLE v2.0: extending marine data layers for bioclimatic modelling. *Glob. Ecol. Biogeogr.* **27**, 277–284. (doi:10.1111/geb.12693)
51. Basher Z, Bowden DA, Costello MJ. 2018 Global marine environment datasets (GMED). *World Wide Web electronic publication Version 2.0*.
52. Clarke A, Gaston KJ. 2006 Climate, energy and diversity. *Proc. R. Soc. B Biol. Sci.* **273**, 2257–2266. (doi:10.1098/rspb.2006.3545)
53. Worm B, Tittensor DP. 2018 *A theory of global biodiversity*. Princeton University Press. (doi:10.2307/j.ctt1zkjz6q)
54. Coolen MJL, Shtereva G. 2009 Vertical distribution of metabolically active eukaryotes in the water column and sediments of the black sea. *FEMS Microbiol. Ecol.* **70**, 525–539. (doi:10.1111/j.1574-6941.2009.00756.x)
55. Holt RD. 2008 Theoretical perspectives on resource pulses. *Ecology* **89**, 671–681. (doi:10.1890/07-0348.1)
56. Bucklin A, Lindeque PK, Rodríguez-Ezpeleta N, Albaina A, Lehtiniemi M. 2016 Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *J. Plankton Res.* **38**, 393–400. (doi:10.1093/plankt/fbw023)
57. Shelton AO, O'Donnell JL, Samhouri JF, Lowell N, Williams GD, Kelly RP. 2016 A framework for inferring biological communities from environmental DNA. *Ecol. Appl.* **26**, 1645–1659. (doi:10.1890/15-1733.1)
58. Martinez MA *et al.* 2021 Epipelagic zooplankton diversity in the deep water region of the gulf of Mexico: a metabarcoding survey. *ICES J. Mar. Sci.* **78**, 3317–3332. (doi:10.1093/icesjms/fsab090)
59. Matthews SA, Goetze E, Ohman MD. 2021 Recommendations for interpreting zooplankton metabarcoding and integrating molecular methods with morphological analyses. *ICES J. Mar. Sci.* **78**, 3387–3396. (doi:10.1093/icesjms/fsab107)
60. Bar-On YM, Phillips R, Milo R. 2018 The biomass distribution on earth. *Proc. Natl Acad. Sci. USA* **115**, 6506–6511. (doi:10.1073/pnas.1711842115)
61. Irigoien X *et al.* 2014 Large mesopelagic fishes biomass and trophic efficiency in the open ocean. *Nat. Commun.* **5**, 3271. (doi:10.1038/ncomms4271)
62. Minamoto T, Fukuda M, Katsuhara KR, Fujiwara A, Hidaka S, Yamamoto S, Takahashi K, Masuda R. 2017 Environmental DNA reflects spatial and temporal jellyfish distribution. *PLoS One* **12**, e0173073. (doi:10.1371/journal.pone.0173073)
63. Salter I. 2018 Seasonal variability in the persistence of dissolved environmental DNA (eDNA) in a marine system: the role of microbial nutrient limitation. *PLoS One* **13**, e0192409. (doi:10.1371/journal.pone.0192409)
64. NathanRG. 2024 Global ocean genome analysis. GitHub. https://github.com/ngeraldi/Global_ocean_genome_analysis
65. Nathan RG. 2024 Assessing patterns of metazoans in the global ocean using environmental DNA. Zenodo. (doi:10.5281/zenodo.12735196)
66. Geraldi NR, Acinas S, Alam I, Gasol JM, Sanchez P, Fernandez ML *et al.* 2024. Data from: Assessing patterns of metazoans in the global ocean using environmental DNA. Figshare. (doi:10.6084/m9.figshare.c.7397762)