Worldwide Analysis of Sedimentary DNA Reveals Major Gaps in Taxonomic Knowledge of Deep-Sea Benthos

Sinniger, Frederic; Pawlowski, Jan; Harii, Saki; Gooday, Andrew J.; Yamamoto, Hiroyuki; Chevaldonne, Pierre; Cedhagen, Tomas; Carvalho, Gary; Creer, Simon

Frontiers in Marine Science

DOI: 10.3389/fmars.2016.00092

Published: 14/06/2016

Peer reviewed version

Cyswllt i'r cyhoeddiad / Link to publication

Dyfniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
https://doi.org/10.3389/fmars.2016.00092

Hawliau Cyffredinol / General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea benthos

Frederic Sinniger¹,²*, Jan Pawlowski³, Saki Harii¹, Andrew J. Gooday⁴, Hiroyuki Yamamoto², Pierre Chevaldonne⁵, Tomas Cedhagen⁶, Gary Carvalho⁷, Simon Creer⁷

¹Tropical Biosphere Research Center, University of the Ryukyus, Japan, ²R&D Center for Submarine Resources, Japan Agency for Marine-Earth Science and Technology, Japan, ³Department of Genetics and Evolution, University of Geneva, Switzerland, ⁴National Oceanography Centre, University of Southampton Waterfront Campus, United Kingdom, ⁵IMBE, CNRS, Aix Marseille Université, IRD, Avignon Université, France, ⁶Department of Biosciences, Aarhus University, Denmark, ⁷School of Biological Sciences, Bangor University, United Kingdom

Submitted to Journal:
Frontiers in Marine Science

Specialty Section:
Deep-Sea Environments and Ecology

ISSN:
2296-7745

Article type:
Original Research Article

Received on:
26 Mar 2016

Accepted on:
27 May 2016

Provisional PDF published on:
27 May 2016

Frontiers website link:
www.frontiersin.org

Citation:

Copyright statement:
© 2016 Sinniger, Pawlowski, Harii, Gooday, Yamamoto, Chevaldonné, Cedhagen, Carvalho and Creer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.
Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea benthos

Frédéric Sinniger 1,2*, Jan Pawlowski 3, Saki Harii 1, Andrew J. Gooday 4, Hiroyuki Yamamoto 2, Pierre Chevaldonné 5, Tomas Cedhagen 6, Gary Carvalho 7, Simon Creer 7

1 Tropical Biosphere Research Center, University of the Ryukyus, Motobu, Okinawa, Japan
2 R&D Center for Submarine Resources, Japan Agency for Marine-Earth Science and Technology, Yokosuka, Kanagawa, Japan
3 Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland
4 National Oceanography Centre, University of Southampton Waterfront Campus, European Way, Southampton, United Kingdom
5 IMBE, CNRS, Aix Marseille Université, IRD, Avignon Université, Station Marine d’Endoume, Marseille, France
6 Department of Biosciences, Aarhus University, Aarhus, Denmark
7 School of Biological Sciences, Bangor University, Bangor, Gwynedd, United Kingdom

*Correspondence:
Fredéric Sinniger
fredsinniger@hotmail.com

Keywords: environmental DNA, biodiversity, metabarcoding, meiobenthos, 18S rDNA

Abstract

Deep-sea sediments represent the largest but least known ecosystem on earth. With increasing anthropogenic pressure, it is now a matter of urgency to improve our understanding of deep-sea biodiversity. Traditional morpho-taxonomic studies suggest that the ocean floor hosts extraordinarily diverse benthic communities. However, due to both its remoteness and a lack of expert taxonomists, assessing deep-sea diversity is a very challenging task. Environmental DNA (eDNA) metabarcoding offers a powerful tool to complement morpho-taxonomic studies. Here we use eDNA to assess benthic metazoan diversity in 39 deep-sea sediment samples from bathyal and abyssal depths worldwide. The eDNA dataset was dominated by meiobenthic taxa and we identified all animal phyla commonly found in the deep-sea benthos; yet, the diversity within these phyla remains largely unknown. The large numbers of taxonomically unassigned molecular operational taxonomic units (OTUs) were not equally distributed among phyla, with nematodes and platyhelminthes being the most poorly characterized from a taxonomic perspective. While the data obtained here
reveal pronounced heterogeneity and vast amounts of unknown biodiversity in the
depth sea, they also expose the difficulties in exploiting metabarcoding datasets
resulting from the lack of taxonomic knowledge and appropriate reference databases.
Overall, our study demonstrates the promising potential of eDNA metabarcoding to
accelerate the assessment of deep-sea biodiversity for pure and applied deep-sea
environmental research but also emphasises the necessity to integrate such new
approaches with traditional morphology-based examination of deep-sea organisms.

Introduction

Prior to global industrialisation, the deep sea was protected from human influence by
its remoteness. However, the impacts of human activities have increased rapidly in
recent decades (Glover and Smith, 2003; Benn et al., 2010), mainly through waste
disposal (e.g. Thiel, 2003; Watters et al., 2010; Miyake et al., 2011; Ramirez-Llodra
et al., 2013) and the expansion of fishing and hydrocarbon extraction to bathyal
depths on continental margins (e.g. Koslow et al., 2000; Roberts, 2002; Clark, 2009).
The mining of metal-rich minerals in environments ranging from abyssal plains to
hydrothermal vents and seamounts is a serious prospect in the fairly near future (e.g.
Wedding et al., 2013; Fisher et al., 2014; Schlacher et al., 2014; Van Dover, 2014).
Such combinations of different direct anthropogenic stressors will likely exacerbate
multiple interacting stressors arising from climatic changes acting at a global scale
(Mora et al., 2013; Jones et al., 2014), creating major threats to the largest
environment on Earth (Ramirez-Llodra et al., 2011). Environmental stresses on whole
ecosystems led to a loss of biodiversity observed worldwide, with consequences to
ecosystem functioning (Worm et al., 2006; Hooper et al., 2012). It is, therefore,
essential to acquire baseline information on deep-sea diversity in order to establish
reference data reflecting near pristine or less impacted habitats. Such baseline studies
are crucial to the assessment of changes in deep-sea ecosystems resulting from the
increasing human activity.

The World Register of Marine Species (http://www.marinespecies.org) lists 23,708
metazoan species found in the depth sea (Fig. 1). Arthropods, chordates, mollusces,
anneids and echinoderms dominate this inventory of deep-sea species. Although such
richness is certainly an underestimate, there is no consensus on how many deep-sea
species exist (Miljutin et al., 2010), in part because of uncertainty concerning their
distribution patterns in a vast and chronically undersampled environment (McClain
and Hardy, 2010). The lack of publicly available molecular data, with only about one
fifth of the inventoried species (4918) being associated with such data in publicly
accessible databases (Fig. 1), coupled with the particularly challenging taxonomic
identification of meiofauna (Herman and Heip, 1988; Giere, 2008), illustrate clearly
the difficulty in assessing metazoan diversity in deep-sea sediments mainly inhabited
by small-sized animals (e.g., Thiel, 1975; 1983).

The development of DNA barcoding has substantially improved taxonomic
knowledge in some groups that are difficult to identify morphologically (Blaxter,
2004). Investigation of the molecular signatures of benthic fauna in environmental
samples was therefore the logical development of DNA barcoding approaches
(Markmann and Tautz, 2005). In recent years, en mass sequencing of environmentally
derived DNA has expanded rapidly with the availability of high-throughput
sequencing technologies, commonly referred to as metabarcoding (Taberlet et al.,
2012b). We consider environmental DNA (eDNA) to comprise not only DNA from
living species, including their eggs and larvae, but also DNA from fragments of dead organisms, gut contents and extracellular DNA (Taberlet et al., 2012a). Since marine sediments host a tremendous diversity of eukaryotic organisms, metabarcoding is particularly useful because of its potential to explore the biodiversity of all taxa in parallel (Bik et al. (2012b). Such an approach has revealed novel biodiversity in various coastal environments (e.g. Chariton et al., 2010; Fonseca et al., 2010; Bik et al., 2012a; Lallias et al., 2014; Pawlowski et al., 2014; Cowart et al., 2015). However, despite the dynamic expansion of eDNA studies, little metabarcoding information is available for benthic diversity at bathyal and abyssal depths (Pawlowski et al., 2011). High-throughput sequencing of deep-sea sedimentary eDNA has revealed a high level of previously unknown diversity among benthic foraminifera (Lecroq et al., 2011) and other deep-sea protists (Stock et al., 2013). Moreover, the capacity of deep-sea sediments to preserve DNA (Corinaldesi et al., 2011) has allowed inferences to be drawn about the past biodiversity of planktonic and benthic eukaryotes (Lejzerowicz et al., 2013). Metabarcoding has also been used to explore biogeographic patterns of microbial eukaryotes in the deep sea (Bik et al., 2012c; Guardiola et al., 2015). Thus, in many different ways, limited studies on eDNA clearly show the strong potential of metabarcoding in deep-sea biodiversity research where samples are scarce and expensive to collect.

Here, we use eDNA metabarcoding to identify gaps in our taxonomic knowledge of deep-sea biodiversity. We address this issue using en masse sampling of sediments from deep-sea environments distributed worldwide from upper bathyal (a few hundred meters depth) to abyssal (4 to 5 km below sea surface) depths. We use our global dataset to (a) test the potential of eDNA metabarcoding to assess deep-sea biodiversity and (b) identify the taxonomic breadth of hitherto unknown benthic diversity.

Methods

Sampling and DNA extractions

Sediments were collected during various cruises worldwide, mostly at abyssal and bathyal depths (see Table S1 and Fig. 2), and stored at -80°C. The sampling methods differed depending on the cruises (multicores, box cores, grab samples). Subsamples were obtained following a standardised method from the more or less undisturbed surficial sediments collected by the sampling gear. The samples consisted of fine mud, except for the Maud Rise samples that included a larger sand fraction. Only surface sediments (approximately within the first 3 cm) were processed. For each location, 4 DNA extractions were performed using MOBIO PowerMax extraction kits with less than 10 g of sediments (corresponding to a volume of between 5 and <10 ml depending on the nature of the sediments). In addition, for Northwest Pacific, Arctic, Southern Ocean and South Atlantic samples, 8 replicates of less than 1 g of sediments (roughly corresponding to 0.7 to 1 ml each) were extracted using MOBIO PowerSoil extraction kits.

PCR amplification and 454 sequencing

Preparation of amplicon libraries for 454 sequencing of the V1-V2 region of the nuclear small ribosomal subunit (18S) followed the protocols described previously (Fonseca et al., 2010). In order to maximise inclusion of metazoans in the primer mix,
primer R22 (Blaxter et al., 1998) was modified as follows: R22mod
5’CCTGCTGCCTTCCTTRGA3’, the primer F04 was left unmodified. Compared to
the original R22 primer and based on sequences of various phyla from GenBank,
R22mod was shortened to remove mismatches in some groups (cnidarians,
echinoderms, priapulids and kinorhynchs) and one ambiguity (R) was added to
accommodate the presence of a thymidine (T) instead of a cytidine (C) at this position
in brachiopods, bryozoan, kinorhynchs, rotifers and within several other groups. The
PCR amplifications were performed directly using the combined 454 adaptors, link,
MID (Molecular Identifier) tags and primers. MID tags were inserted only in the
forward primers as sequencing was made unidirectionally and 8bp MID tags were
used to distinguish between independent samples (Table S2). In order to reduce
chimera artefacts created during the PCR (Fonseca et al., 2012), we reduced the PCR
cycles to 23-25 cycles. The positive amplifications were identified
on a 1.5% agarose gel stained with ethidium bromide. Five amplifications were performed in parallel for
each core and extraction method. PCR products for each sample were pooled
on a 2 % agarose gel, and then excised bands were purified using the QIAquick PCR
purification kit (Qiagen). The purified products were quantified using a Bioanalyzer
(Agilent) and sequenced on a 454 Roche GSFLX sequencer on either quartets, or half
plates at the sequencing platform of Liverpool University.

**Sequence analyses**

Sequences were analysed using the QIIME 1.7 (Caporaso et al., 2010). Raw reads
were assigned to samples based on MID tags and checked for quality using
split_libraries.py. All sequences shorter than 200 bp were discarded, minimum quality
score was set at 25 and maximum homopolymer run was set at 6 bp. No mismatches
in primer or MID tags were tolerated. In order to reduce the potential bias introduced
by intragenomic variability and sequencing errors, OTUs were clustered at 97%
identities. Cluster seeds were selected as representative sequences for each OTUs
using pick_rep_set.py in QIIME 1.7. Sequences were aligned using align_seqs.py and
the aligned s108 SILVA database as template. Chimeric sequences were removed
using identify_chimeric_seqs.py and ChimeraSlayer in QIIME 1.7 and single
singletons (i.e. single sequences present in a single sample) were removed from the
dataset.

The 18S sequences were then compared, using the BLAST method with an E- value
threshold of 1e-100, against a reference database consisting of a “customised” version
of the Silva database s108 release formatted for use within the QIIME pipeline
(Caporaso et al., 2010), using assign_taxonomy.py in QIIME 1.7. The E-value
threshold was empirically identified following the observation that the default
parameter (E-value threshold of 0.001) provided an unrealistic number of identified
OTUs, which after verification were often only poorly related to the assigned taxa,
even at phylum level. Several OTUs were independently compared to the GenBank
database and the E-values below 1e-100 did not allow reliable identification at phylum
level in many cases. As a result, we chose to use a strict E-value threshold to limit
“folkloric” taxonomic assignments. The database was customised by correcting some
obvious misidentification (e.g., a copepod crustacean sequence labelled as octocoral
cnidarian) and adding recent deep-sea sequences obtained from public databases and
individual sequencing of deep-sea organisms. All “uncultured (marine) eukaryote” or
environmental samples identified above the phylum level were also removed from the

After taxonomic assignment, a phylogenetic tree was built with the unassigned sequences and the branch corresponding to metazoans was identified by independent blasts against the GenBank database. All the sequences forming the branch identified with confidence as metazoan were isolated and merged to the metazoan dataset with the following taxonomic assignment “Eukaryota;Metazoa; no blast hit”. However, because of the conservative approach chosen, and based on the limited phylogenetic resolution provided by the fragment analysed, the most basal metazoan OTUs (such as those assigned to sponges) may not have been included in the “metazoan branch”.

The decision to restrict the analyses to metazoans reflects the fact that the original primers were specific to metazoans and the observation that one major group of deep-sea protists, the foraminifera, was not found in our dataset, thus providing a biased estimation of deep-sea eukaryotic biodiversity.

Alpha diversity was measured using the simple “observed OTU” metric in order to estimate the depth of sequencing for the sediments analyses. OTU networks, linking the OTUs to the different biogeographic provinces in which they were found, were built using make_otu_network.py in QIIME 1.7 and drawn in Cytoscape 2.7.0 (Shannon et al., 2003) using the unweighted spring embedded layout. Beta diversity analyses were conducted using the unweighted UNIFRAC method (Lozupone and Knight 2005) implemented in QIIME 1.7. This method takes into account the phylogenetic information in the dataset. The unweighted approach allows the use of only qualitative data (i.e. presence/absence) and reduces bias from quantitative results. While this approach increases the importance of rare taxa, the qualitative bias potentially induced by the biomass (Bohmann et al., 2014; Hirai et al., 2015) was estimated too high considering the minimal amounts of sediments used. Unrarefied data allowed us to consider the total diversity recovered from the samples analysed. However, in order to make more objective comparisons between heterogeneous samples, Principal Coordinates Analyses of beta diversity was performed based on hundred rarefied datasets at 4488 metazoan reads per province. The same beta-diversity distances matrices were used to build UPGMA trees. Bootstrap support was calculated based on the 100 rarefied datasets. Biogeographic comparisons are based on the lower bathyal and abyssal provinces described in Watling et al. (2013).

According to the latter (Watling et al., 2013), the Mediterranean Sea and the North Atlantic both belong to the same bathyal province BY4. However, based on the differences in our sampling locations and data, as well as the suggestion made by Watling et al. (2013) that this province may require subdivision based on differences in environmental parameters such as temperature, we decided to treat both locations separately and refer to the bathyal Mediterranean as BY4 and the bathyal NW Atlantic as BY4b. Additionally, since Watling et al. (2013) did not divide upper bathyal regions (300-800 m depth) into provinces, we identified our upper bathyal regions by adding a “Z” to the name (i.e. BY1Z, BY9Z and BY11Z).

**Results**

**Taxonomic composition**

After quality checking, and removal of chimera and single singletons (i.e. single reads present in a single sample, rather than reads being single in each of a few samples),
the resulting dataset contained 530976 reads (from 3819 to 97456 reads per location),
38% of which could be confidently assigned to metazoa using a combination of
BLAST and the phylogenetic approach (Table 1). The remaining reads were assigned
to various other euakaryotic groups (e.g. Stramenopiles, Fungi). The distribution of
reads within the different metazoa groups was highly variable between locations
with a large dominance of annelids at nearly half of the sampling sites. Nematodes
were also sequenced abundantly in several samples followed by arthropods and a
large proportion of reads that could not be confidently assigned by BLAST methods
to any specific phylum (Fig 3).

OTU richness data showed less variation between locations compared to abundance
data (Fig 3). At a clustering threshold of 97%, metazoa OTUs represented between
8.4% (S Brazil Basin, abyssal) and 34.8% (Antarctic Peninsula) of the total
eukaryotic richness for each site. Metazoa diversity was largely dominated by
nematodes, which formed the most diverse group in all samples (from 25.4% to
48.9%). The next most diverse phyla were arthropods (mainly copepods) and annelids,
followed by platyhelmithes (Fig 3a). These four groups comprised nearly 88% of the
total number of assigned OTUs. The remaining OTUs were assigned to 19 other phyla.
Among the unassigned OTUs that could not be reliably recognised using BLAST with
the strict E-value threshold selected here, individual BLAST of some OTUs against
GenBank database suggests that they likely belong to the phyla Mesozoa and
Tardigrada. Combined, the taxa inferred to be present based on our data represented
almost all higher-level diversity of marine Metazoa (with the exception of a few
minor phyla such as Acanthocephala, Entoprocta and Phoronida).

A substantial diversity of orders and families was recovered within each of the three
major phyla (nematodes, annelids, arthropods), based on BLAST data. At order level,
the diversity was relatively equally distributed among nematodes, with less than a
quarter of OTUs belonging to the order Enoplida (Fig. 4a). The arthropods were
clearly dominated by harpacticoid copepods (68%) (Fig. 4b), while more than 50% of
annelids belonged to the infraclasse Scolecida or orders Spionida and Terebellida (Fig.
4c). Within orders, a wide diversity of families has been observed in the three major
phyla (Fig. 4a-b-c).

Phylogenetic analyses of metazoa OTUs (97% identity threshold) showed the
uneven distribution of unidentified OTUs (in red in fig. 5) throughout the resulting
tree. Although unassigned OTUs were found in almost all taxonomic groupings,
several clusters were composed mainly of unassigned OTUs. Our analyses confirmed
the impressive diversity of nematodes, representing almost half of the tree (Fig. 5),
although a significant part of this diversity may originate from intragenomic polymorphisms (Dell’Anno et al., 2015). Other monophyletic clusters were formed by
the superphylum Deuterostomia, the phylum Gastrotricha, the class Ostracoda and the
subclass Copepoda. The Copepoda comprised mainly harpacticoids as shown by a
comparison between OTUs assigned to copepods and sequences of harpacticoids
available in the database (Fig. S2).

Several clades were formed by OTUs belonging to different taxonomic groups (Fig.
5). The annelids grouped with the molluscs, the kinorhynchs grouped with mites and
potential tardigrades while the echinoderms clustered with hemichordates and
chordates forming a deuterostome clade. Within deuterostomes, detailed observation
showed that echinoderms, hemichordates, vertebrates and tunicates formed
independent sub-clusters with tunicates appearing clearly distinct at the base of this group (data not shown). Early metazoans (sponges, placozoans and cnidarians) also formed a monophyletic group branching between the deuterostomes and the clade comprising loriciferans, aceolomorphs and putative mesozoaons. Nemerteans were either located within the two clades containing annelids or formed an independent cluster nearby. Within this “nemertean only” cluster, it is interesting to note the presence of a few “platyhelminthes” OTUs assigned to the flatworm “Nematomorpha” sp.”. However, most likely the reference sequence from the public database originates from a misidentified specimen (sequence GenBank D85093). Except for this likely artefactual identification, all platyhelminthes clustered together in a monophyletic group. Interestingly, aside from the platyhelminthes cluster, which already includes a significant number of unassigned OTUs, another large clade is composed exclusively of unassigned OTUs. While no reliable BLAST identification could be obtained, these OTUs appear to be related to aceolomorphs. Although meio-benthic tunicates can be found, the presence of DNA from vertebrates (fish and cetacean) and most likely planktonic tunicates illustrates well the potential of eDNA to amplify not only organisms physically present in the sediments but also both indigenous and allochthonous extracellular DNA.

**Biogeographic patterns**

When considering all provinces regardless of the depth, out of the 1570 metazoan OTUs recovered, only 3 OTUs were shared among all 13 provinces (a harpacticoid copepod, a nematode and an unassigned OTU), 18 additional OTUs were found in 10 to 12 provinces (7 nematodes, 3 copepods and another undefined arthropod, 2 annelids, 1 hemichordate and 4 unassigned OTUs). No evidence was found for a higher proportion of predominantly planktonic groups, such as Ctenophora and Chaetognatha, among “cosmopolitan” OTUs, although some tunicates, hydrozoans and halocyprid ostracods observed in the abyssal provinces might have originated from the water column. Such findings correspond with the overall low representation of OTUs originating from the water column.

The spring embedded network visualisation of OTUs distributes the provinces on the network in order to minimize the differences in lengths of the edges connecting OTUs to the provinces (i.e. in a similar way as if the edges would be springs connecting balls corresponding to OTUs and provinces). Although easily saturated when including large amounts of OTUs, this method of visualising the distribution showed that the geographic distribution of OTUs appeared not random for several taxonomic groups. In annelids, the OTUs from abyssal provinces tended to cluster together, separated from the upper bathyal provinces by the lower bathyal provinces. Moreover, the two polar lower bathyal provinces grouped near the abyssal ones while the two polar upper bathyal provinces appeared more isolated (Fig. 3a).

Such patterns are not visible for all phyla, but the pattern observed in the Nematoda (Fig. 3b, based on OTUs shared by 6 provinces or more) tends to suggest relationships between abyssal provinces, although more appropriate sampling is required to explore biogeography in details. Compared to the patterns obtained with annelid OTUs, the upper bathyal Arctic also grouped close to the South Atlantic and Weddell Sea abyssal provinces. Moreover, the polar bathyal provinces clustered near or even within abyssal provinces in the nematodes data. The network made from arthropod MOTUs (Fig. 3c, including all OTUS) showed a different pattern with the Southern
Ocean provinces clustering together, as well as the Arctic and abyssal South Atlantic provinces. The single abyssal Pacific province (NW Pacific, on the edge of the Japan Trench) appears relatively isolated from the other abyssal provinces and the Mediterranean, NW Atlantic and two Andaman Sea provinces all appear quite isolated from each other and from the other provinces.

While networks facilitate the visualisation of the OTUs distributions, they are limited to taxonomic groups with limited numbers of OTUs (otherwise the network will saturate) and are directly affected by the sequencing depth (not rarefied). Principal coordinate analyses (PCoA) on rarefied dataset of all metazoan OTUs provided a more robust comparison of the different locations (Fig. 7). Unfortunately, the rarefaction threshold of 4488 metazoan reads per location did not allow the inclusion of several locations in the analyses (e.g. most abyssal locations including all the equatorial and South Atlantic locations, see table 1). Nevertheless, all the locations in Andaman Sea, the province with the most locations sampled, clustered relatively closely together. The two Mediterranean locations also grouped within this cluster. The situation was different for the lower bathyal Southern Ocean, for which the two sampled locations considered did not group together. One possible reason for the difference observed between the Lazarev Sea and the Maud Rise, may be related to the sediment characteristics, as Maud Rise sediments sampled were sandier than the fine muddy sediments from Lazarev Sea. Moreover, compared to the distances between the locations within Mediterranean Sea or within Andaman Sea, the Maud Rise was much more distant from the Lazarev Sea, increasing the possibility of different ecosystems being sampled (Giere, 2008).

**Discussion**

Most deep-sea diversity is unknown

Pioneering investigations in the 1960s (e.g., Hessler and Sanders, 1967; Sanders and Hessler, 1969), together with more recent studies (e.g., Snelgrove and Smith, 2002; Brandt et al., 2007; Rex and Ette, 2010) on bathyal and abyssal fauna, have challenged the long-held notion that the deep sea hosts a low diversity of metazoan organisms. Our results, based on the total DNA from the sediments (including organismal and extraorganismal DNA), reveal a large proportion of unassigned OTUs (Fig. 5). Although there is not necessarily a direct correspondence between DNA sequence data and morphological species diversity, these results do suggest that significant unknown diversity exists in deep-sea sediments at different taxonomic levels, supporting the idea of a highly diverse deep-sea fauna. Moreover, the irregular distribution of unassigned OTUs in the phylogenetic tree provides clear evidence that some taxonomic groups are particularly understudied. These less sequenced groups include cryptic and/or fragile organisms such as the acoelmorphs and loriciferans, which are rarely seen in deep-sea samples, as well as several groups of nematodes.

Based on the rate at which new taxonomic descriptions are being published, it has been proposed recently that most biodiversity on Earth might be described in the relatively near future (e.g. Appeltans et al., 2012; Costello et al., 2013). A distinction does need to be made between species that have been described taxonomically and those that have only been sequenced (Fig. 1). Nevertheless, by suggesting that important unknown genetic diversity exists within several deep-sea metazoan phyla (Fig. 4), our results tend to challenge these ambitious predictions and support the view
that a large part of the planet’s biodiversity remains to be discovered in the deep sea (e.g. Grassle and Maciolek, 1992; Poore and Wilson, 1993; Brandt et al., 2007; George et al., 2014). Many of these OTUs could not be assigned to any taxonomic group and some could therefore represent new higher taxa. Unfortunately, no DNA information is available for much of the known deep-sea metazoan diversity (Fig. 1) and their novelty is therefore impossible to confirm. These uncertainties should not undermine the potential of metabarcoding to better understand the diversity of poorly known communities. Indeed, the total information obtained from a large number of taxa in parallel provides a good estimator of environmental community diversity that has many practical applications for ecosystem assessment and monitoring (Chariton et al., 2010; Czernik et al., 2013; Stephenson et al., 2013; Chariton et al., 2014; Lallias et al., 2014; Pawlowski et al., 2014; Willerslev et al., 2014; Guardiola et al., 2015; Lejzerowicz et al., 2015; Pochon et al., 2015; Boschen et al., 2016).

Almost all marine benthic phyla were found in the sediments analysed and even with a conservative OTU clustering threshold of 97% and a limited number of samples, a wide diversity of OTUs was identified within the dominant phyla (Fig. 5, Fig. S2). The prevalence of nematodes and other meiofaunal groups is immediately apparent and confirms that meiofauna are an important component of the deep-sea benthic biodiversity. Such community composition is consistent with the slower rate of decline in the abundance and biomass of metazoan meiofauna from bathyal to abyssal regions compared to that of larger animals (macrofauna and megafauna) (e.g., Thiel, 1975; Rex et al., 2006; Rex and Etter, 2010). However, the absence of some macrobenthic taxa (e.g. peracarid crustaceans) and particularly of megabenthos such as decapod crustaceans, sea cucumbers and fish, which are very common in the deep sea, can be partially explained by the limited volume of analysed sediment samples (i.e. less than 10 ml per core). The megafaunal (vertebrate) sequences found in our data clearly originate from extracellular DNA and illustrate well the potential of environmental DNA to inform not only on the organisms physically present in the sample but also on DNA traces of large sized species.

In comparison with previous studies, we retrieved a lower proportion of metazoans than when the meiobenthos is isolated by decantation and sieving (45-1000µm size fraction) prior to DNA extraction (Creer et al., 2010; Fonseca et al., 2010; Bik et al., 2012a; Bik et al., 2012c; Fonseca et al., 2014). This finding is consistent with recent comparisons between different sampling sizes (Brannock and Halanych, 2015). However, in terms of metazoan diversity, our approach of analysing the DNA extracted directly from sediments does not appear to have retrieved significantly different patterns of diversity, as the dominant phyla are similar using both approaches. Brannock and Halanych (2015) recommend the use of elutriated samples (with meiofauna extracted from the sediments) to increase the amount of metazoan reads recovered. However, such approach requires larger volumes of sediments, which are not always available in deep-sea research. Extracting DNA from raw sediments has the advantage of including more extracellular DNA, but also the risk of including DNA from non-benthic organisms. Another study based on abyssal sedimentary DNA suggested that the DNA of planktonic species might account for more than 30% of all eDNA preserved in seafloor sediments (Pawlowski et al. 2011), although our new results do not reflect these findings. The length of the amplified DNA fragment likely explains the higher proportion of benthic diversity observed here. The DNA fragment sequenced in this study was significantly longer (approximately 450
bp) than the 150-bp-long V9 region used in Pawlowski et al. (2011). Comparison
between different genetic markers or data obtained with different primer pairs should
be considered with caution (Hadziavić et al., 2014). For example, different markers
will likely have different evolution rates that will additionally vary between
taxonomic groups, leading to confusing taxonomic interpretations of the results
obtained. However, overall, targeting larger DNA fragments will favour the
amplification of DNA from living organisms, or that of recently dead individuals
whose genomic content still persists in good condition in the environment.
Considering the logistic difficulties to sample in the deep sea, and the additional bias
induced during the meiofaunal isolation process (Bik et al., 2012c), sequencing the
total sedimentary eDNA represents a good compromise for exploring the biodiversity
of small-sized, deep-sea metazoans.

Biogeographic patterns

Broad spatial distributions are reported among small-sized eukaryotic taxa such as
rotalid foraminifera (Pawlowski et al., 2007; Gooday and Jorissen, 2012), nematodes
(Vanreusel et al., 2010; Zepplin et al., 2011) and harpacticoid copepods (Menzel et al.,
2011), as well as certain macrofaunal and megafaunal taxa (Sibuet, 1979; Allen,
2008). However, in some cases, detailed morphological and/or molecular re-
examination of putative cosmopolitan species resulted in the recognition of cryptic
species having much smaller distribution ranges (Moura et al., 2008; Brandão and
Yasuhara, 2013; Krapp-Schickel and De Broyer, 2014; Yasuhara et al., 2014). The
lower numbers of cosmopolitan taxa in our study originate either from higher than
expected biodiversity or undersampling of the vast ocean-floor environment.
Unfortunately, we cannot discriminate between these two hypotheses because the
samples on which this study is based were collected opportunistically, with different
numbers of samples and different sequencing depths at each location.

Deep-sea habitat heterogeneity at larger spatial scales is poorly understood but is
believed to play an important role in the maintenance of benthic biodiversity on the
ocean floor (e.g. Levin et al., 2001; Van Gaever et al., 2009; Vanreusel et al., 2010;
Durden et al., 2015). Our data suggest a considerable degree of taxonomic
differentiation, and hence biogeographic patterning, between the soft-sediment
benthic communities that are represented by our eDNA samples (Figs 6,7). For
example, the lower bathyal (1920-2160 m) polychaete data from the Lazarev Sea and
Maud Rise (BY9) and those derived from upper bathyal (290-500 m) samples taken
on the unusually deep Antarctic shelf (BY9Z) are strikingly different. This is
inconsistent with the extended bathymetric ranges often observed among species
living around the Antarctic continent (Brandt et al., 2007), but is not surprising given
the distinctive nature of benthic communities on the western side of the Antarctic
Peninsula (Scotia area; (De Broyer and Koubbi, 2014)). The composition of soft-
sediment communities in the lower and particularly the upper bathyal Andaman Sea
provinces will almost certainly be influenced by the oxygen minimum zone in this
region (Cedhagen et al., 2013). However, further biogeographic interpretation of our
data would be inappropriate given the fact that our samples were obtained
opportunistically from scattered locations. More extensive sampling, preferably
targeted in relation to environmental gradients (e.g. depth, productivity, bottom-water
oxygen levels), will be required in order to assess the full potential of eDNA
metabarcoding to explore the heterogeneity of deep-sea metazoan communities and their biogeographic patterns.

*Future challenges of deep-sea eDNA metabarcoding.*

From a molecular perspective, the main challenge for deep-sea metabarcoding studies is to find optimal molecular markers for metazoan species delimitation. Metazoan clades comprise highly diversified phyla with different rates of evolution (Johnson et al., 2014) and selecting a region of rRNA genes that would have a similar taxonomic resolution for all species is virtually impossible. For example, based on 18S rDNA data publicly available, different species within the deep-sea mollusc genus *Bathymodiolus* share between 99.2 and 100% identities, while species of the crustacean genus *Paramunida* share between only 91.5 and 99.4% identities and different genera within the cnidarian family Parazoanthidae share between 97.8 and 99.7% identities. Consequently both 97% and 99% identity clustering thresholds will merge OTUs representing very different taxonomic ranks depending on the taxa concerned. Even within organisms, a recent study of deep-sea nematodes by Dell’Anno et al. (2015) demonstrates how significant intragenomic polymorphism can impact the interpretation of metabarcoding data. Due to the arbitrary nature of species definitions and evolutionary differences between metazoan taxa, such issue will remain crucial despite efforts made to improve the taxonomic assignments of high-throughput sequencing data (Quince et al., 2011; Morgan et al., 2013).

The widely used COI gene has been proposed as an alternative metabarcoding marker to compensate for the lack of resolution of 18S rDNA at species, genus, or even higher taxonomic ranks in meio-benthos taxa (Tang et al., 2012). However, finding conserved COI priming sites in all metazoans is even more problematic than for 18S rDNA (Deagle et al., 2014). Moreover, the level of codon saturation provided by COI precludes us from identifying OTUs without an accurate and complete reference dataset. A test study conducted in parallel with this research has shown that less than 10% of the reads obtained for a standard COI fragment could be identified by BLAST (unpublished data). These results support published data on seagrass meadows where 93% of the COI OTUs recovered remained unassigned (Cowart et al., 2015).

Therefore, although the high-resolution power of COI for identifying species provides a significant advantage, the substantial inadequacy of available reference sequences is even more acute than for rRNA genes.

From a taxonomic perspective, the main challenge for deep-sea biodiversity research is to expand the reference database. Deep-sea diversity remains largely unknown (Costello et al., 2010; Danovaro et al., 2010) and even when identified using molecular taxonomic approaches, high-level assignment cannot be achieved in many cases. Moreover, a complete and reliable reference database is not only needed for taxonomic assignment but is also essential for post-sequencing processing of the data (Edgar et al., 2011; Quince et al., 2011; Fonseca et al., 2014). As shown in Figure 5, the level of taxonomic identification depends on the group. For example, the copepod clade, several groups of nematodes, and one group of annelids include only a modest proportion of unassigned OTUs. However, other groups are almost entirely composed of unassigned OTUs, suggesting the absence of reference sequences in the database. Overall, the amount and clustering of the unassigned OTUs observed here suggest the existence of largely uncharacterized taxonomic groups and highlights the potential extent of the unknown diversity in the deep sea.
DNA barcoding and morphology-based taxonomy have sometimes been perceived as antagonistic approaches (e.g. Ebach and Holdrege, 2005; Trewick, 2008; Boero, 2010). However, our results clearly emphasize the absolute necessity to increase taxonomic effort, including morphological analyses as proposed originally for DNA barcoding approaches (Hebert et al., 2003), in order to fully exploit the gigantic amounts of DNA data obtained by metabarcoding. On the one hand, morphological examination of the specimens that compose benthic communities is not always possible and even when possible is often extremely time consuming and usually requires the expertise of specialist taxonomists. Such limitations apply especially in the deep sea, where samples are difficult to obtain and often limited in size. On the other hand, as discussed above and in Dell’Anno et al. (2015), interpretation of metabarcoding data is limited by the reference database available. Therefore, rather than being in competition, the two approaches complement each other in providing a concerted framework that can be used to obtain the most accurate estimation of marine biodiversity on our planet for both pure and applied environmental research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

FS, JP, SC planned the experiments; FS conducted the experiments, JP, SC, HY, SH, PC, TC contributed to the experiments, FS, SC, JP, SH analysed the data, FS, JP, AG, SC wrote the manuscript; SC, GC, SH, PC critically reviewed the manuscript.

Funding

This study was funded through the Marie Curie international incoming fellowship project MARMEDIV (No 253251) and a Japan Society for Promotion of Science grant in aid for young researcher to FS (no 26870917). Part of this research was funded by the Japanese Cross-ministerial Strategic Innovation Promotion (SIP) Program for Development of New-generation Research Protocol for Submarine Resources and part of the sampling was supported by the Carlsberg foundation to TC (No 2010-01-0376), Swiss National Science Foundation to JP (grant 31003A-140766) and a British Antarctic Survey collaborative gearing scheme to SC (No 57). Mediterranean samples were obtained through funding by the “Deepsets” Responsive Mode Proposal (RMP) of the MarBEF NoE (EU Network of Excellence).

Acknowledgements

We thank the crews and participants to the various cruises that allowed us to obtain the deep-sea sediments, especially to Dr. P. Martinez-Arbizu, Dr. A. Brandt, Dr. T. Pérez, Dr. T. Maruyama and Dr. M. Tsuchiya. Thanks go also to the British Antarctic Survey, Dr. L. Peck and all the staff of the Rothera Antarctic station during the summer 2010-2011 who made the sampling possible.

References


systematic conservation planning in the deep sea. Proceedings of the Royal Society of London B: Biological Sciences 280, 20131684.


Legend to the figures

**Figure 1:** Proportions of deep-sea species with available DNA sequences on GenBank. The numbers are based on valid species described and registered in the World Register of Deep-Sea Species. The category “Other” groups Porifera, Bryozoan, Nematoda, Brachiopoda, Platyhelminthes, Sipuncula, Echiura, Chaetognatha, Nemertea, Ctenophora, Tardigrada, Hemichordata, Cephalorhyncha, Dicyemida, Acanthocephala, Entoprocta, Gastrotricha, Phoronida (listed by order of described species number). Data accessed 17.10.2014.

**Figure 2:** Map of the sampled regions. 1. Arctic (BY1 and BY1Z), 2. NW Atlantic (BY4b), 3. Mediterranean Sea (BY4), 4. Pernambuco Abyssal Plain (AB2), 5. Brazil Basin (AB3), 6. Argentinean Basin (AB5), 7. Southern Ocean (AB6 and BY9), 8. Antarctic Peninsula (BY9Z), 9. Andaman Sea (BY11), 10. NW Pacific (AB13).

**Figure 3:** Taxonomic composition. a) Species richness, contribution of OTUs from different phyla to the total metazoan species richness for each province. b) Quantitative distribution of the reads from each phyla to the total amount of metazoan reads for each province. The number of metazoan reads obtained for each location is indicated above each column. The last column represents the proportions based on all locations.

**Figure 4:** Order and family diversity. Proportions at order (inside chart) and family (outside chart) levels for a) nematodes, b) arthropods and c) annelids. Different shades of the same colour in the outside charts indicate different families within each order. Detailed legends on the family charts are available in Fig. S1.
Figure 5: Phylogenetic distance tree obtained from all metazoan OTUs. OTUs unassigned using BLAST are named in red. Major taxa identified through independent BLASTs are highlighted. Names within quotation marks indicate taxa corresponding to OTUs that could not be reliably identified by BLAST but that were subsequently identified using the complete GenBank database and phylogenetic distance with reference sequences. Branch lengths are representative of the genetic distances between sequences.

Figure 6: A) Annelid OTU network. Small dots represent OTUs and larger discs represent provinces. Lines connect OTUs to the provinces they were found. Lines are colored according to depth: blue = abyssal, red = lower bathyal, green = upper bathyal. B) Network of nematode OTUs shared by 6 provinces or more. C) Network of all arthropod OTUs.

Figure 7: Principal Coordinate Analyses (PCoA) plot of the beta diversity distances obtained from the unweighted UNIFRAC analyses on 100 independent resampling of 4488 metazoan reads per province (most abyssal samples did not reach this threshold and were discarded from the analyses). Symbol colour correspond to the biogeographic regions, overlapping ellipses (most often masked by the symbols) represent the interquartile range. Grey triangles = BY9 (Southern Ocean with Lazarev Sea and Maud Rise), pink pentagon = BY1 (Arctic), petrol blue triangles = BY4 (Mediterranea), green squares = BY11 (bathyal Andaman Sea), salmon squares = BY11Z (upper bathyal Andaman Sea), yellow diamond = BY9Z (upper bathyal Antarctic peninsula), blue hexagon = BY4b (NW Atlantic), purple circle = AB13 (Abyssal NW Pacific).

Supplementary material
Table S1: Sampling location and information.
Table S2: List and sequences of the MID tags used.

Figure S1: Order and family diversity (as in Fig. 4). Proportions at order (inside chart) and family (outside chart) levels for a) nematodes, b) arthropods and c) annelids.

Figure S2: Phylogenetic tree of copepods. Reference sequences obtained from GenBank are represented in red. OTUs obtained in this study are in black.
Table 1: Amounts and proportions of reads/OTUs for each location. Maximum and minimum values are indicated in green and red respectively.

<table>
<thead>
<tr>
<th>Location</th>
<th>Biogeographic region</th>
<th>Eukaryota</th>
<th>Metazoa</th>
<th>Metazoan proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Reads</td>
<td>OTUs</td>
<td>Reads</td>
</tr>
<tr>
<td>Andaman S1</td>
<td>BY11</td>
<td>35390</td>
<td>1886</td>
<td>19210</td>
</tr>
<tr>
<td>Andaman S3</td>
<td>BY11</td>
<td>41689</td>
<td>2278</td>
<td>18291</td>
</tr>
<tr>
<td>Andaman S4</td>
<td>BY11Z*</td>
<td>14583</td>
<td>1515</td>
<td>4936</td>
</tr>
<tr>
<td>Andaman S5</td>
<td>BY11Z*</td>
<td>16642</td>
<td>1267</td>
<td>7793</td>
</tr>
<tr>
<td>Mediterranea 890</td>
<td>BY4</td>
<td>26059</td>
<td>1866</td>
<td>8374</td>
</tr>
<tr>
<td>Mediterranea 950</td>
<td>BY4</td>
<td>10881</td>
<td>954</td>
<td>7002</td>
</tr>
<tr>
<td>NW Atlantic</td>
<td>BY4b*</td>
<td>42974</td>
<td>856</td>
<td>4488</td>
</tr>
<tr>
<td>Japan Trench</td>
<td>AB13</td>
<td>56946</td>
<td>1674</td>
<td>21550</td>
</tr>
<tr>
<td>Pernambuco</td>
<td>AB2</td>
<td>20564</td>
<td>1482</td>
<td>2324</td>
</tr>
<tr>
<td>Abyssal Plain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N Brazil Basin</td>
<td>AB3</td>
<td>7651</td>
<td>934</td>
<td>446</td>
</tr>
<tr>
<td>S Brazil Basin</td>
<td>AB3</td>
<td>42480</td>
<td>2383</td>
<td>2658</td>
</tr>
<tr>
<td>Argentinean Basin</td>
<td>AB5</td>
<td>6829</td>
<td>576</td>
<td>1412</td>
</tr>
<tr>
<td>Antarctic Peninsula</td>
<td>BY9Z*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weddell Sea</td>
<td>BY9</td>
<td>13295</td>
<td>979</td>
<td>2986</td>
</tr>
<tr>
<td>Lazarev Sea</td>
<td>BY9</td>
<td>11878</td>
<td>924</td>
<td>6543</td>
</tr>
<tr>
<td>Maud Rise</td>
<td>BY9</td>
<td>43974</td>
<td>2371</td>
<td>18600</td>
</tr>
<tr>
<td>Arctic lower bathyal</td>
<td>BY1</td>
<td>37866</td>
<td>2169</td>
<td>6191</td>
</tr>
<tr>
<td>Arctic upper bathyal</td>
<td>BY1Z*</td>
<td>3819</td>
<td>626</td>
<td>1522</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>530976</td>
<td>25462</td>
<td>201978</td>
</tr>
</tbody>
</table>

* indicates regions not listed in Watling et al. 2013, "Z" refers to upper bathyal depths, while BY4 was split to allow distinction between North Atlantic and Mediterranean Sea.