A flexible pipeline combining ~~bioinformatic~~clustering and correction tools for prokaryotic and eukaryotic metabarcoding

Short title:

A flexible metabarcoding pipeline based on read correction

Miriam I. Brandt[1], Blandine Trouche[2], Laure Quintric[3], Patrick Wincker[4,5], Julie Poulain[4,5], and Sophie Arnaud-Haond[1]

[1]MARBEC, Ifremer, Univ. Montpellier, IRD, CNRS, Sète, France

[2] Univ. Brest, CNRS, Ifremer, Laboratoire de Microbiologie des Environnements Extrêmes, Plouzané, France

[3]Ifremer, Cellule Bioinformatique, Brest, France

[4] Génomique Métabolique, Génoscope, Institut François Jacob, CEA, CNRS, Univ. Evry, Université Paris-Saclay, 91057 Evry, France

[5] Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/ Tara

Corresponding author: sarnaud@ifremer.fr, miriam.isabelle.brandt@gmail.com,

ABSTRACT

Environmental metabarcoding is an increasingly popular tool for studying biodiversity in marine and terrestrial biomes. ~~As~~With sequencing costs decreasing, multiple-marker metabarcoding~~ with multiple markers~~, spanning several branches of the tree of life, is becoming more accessible. However, bioinformatic ~~pipelines~~approaches need to ~~accommodate both micro- and macro biologists~~adjust to the diversity of taxonomic compartments targeted as well as to each barcode gene specificities. We built and tested a pipeline based on Illumina read correction with DADA2 allowing analysing ~~metabarcode~~metabarcoding data from prokaryotic (16S) and eukaryotic (18S, COI) life compartments. We implemented the option to cluster Amplicon Sequence Variants (ASVs) into Operational Taxonomic Units (OTUs) with swarm v2, a network-based clustering algorithm, and to further curate the ASVs/OTUs based on sequence similarity and co-occurrence rates using a recently developed algorithm, LULU. Finally, ~~a~~ flexible taxonomic assignment ~~of the Amplicon Sequence Variants (ASVs)~~ was ~~added~~implemented *via* ~~the~~ Ribosomal Database Project (RDP) Bayesian classifier ~~or by~~and BLAST. We validate this pipeline with ribosomal and mitochondrial markers using eukaryotic mock communities and 42 deep-sea sediment samples. The results show that ASVs, reflecting genetic diversity, may not be appropriate for alpha diversity estimation of organisms ~~defined by~~fitting the biological species concept. The results underline the advantages of clustering and LULU-curation for producing more reliable metazoan biodiversity inventories, and show that LULU is an effective tool for filtering metazoan molecular clusters, although the minimum identity threshold applied to co-occurring OTUs has to be increased for 18S. The comparison of BLAST and the RDP Classifier underlined the potential of the latter to deliver very good assignments, but highlighted the need for a concerted effort to build comprehensive, ~~yet specific databases adapted to the studied communities. The results underline the advantages of clustering and LULU-curation for producing metazoan biodiversity~~

~~inventories, and show that LULU is an effective tool for filtering metazoan molecular clusters while avoiding arbitrary relative abundance filters. Overall conservative estimates of diversity can be obtained using DADA2 and LULU correction algorithms alone, or in combination with the clustering algorithm swarm v2 (i.e. to obtain ASVs or OTUs), depending on the objective of the study.~~ ecosystem-specific, databases adapted to the studied communities.

Key words: Biodiversity, bioinformatics, environmental DNA, metabarcoding, mock communities, eukaryotes (18S and COI), prokaryotes (16S)

INTRODUCTION

High-throughput sequencing (HTS) technologies are revolutionizing the way we assess biodiversity. By producing millions of DNA sequences per sample, HTS now allows broad taxonomic biodiversity surveys through metabarcoding of bulk DNA from complex communities or from environmental DNA (eDNA) ~~DNA~~ directly extracted from soil, water, ~~or~~ and air samples~~, i.e. environmental DNA (eDNA)~~. First developed to unravel cryptic and uncultured prokaryotic diversity, metabarcoding methods have been extended to eukaryotes as powerful, non-invasive tools, allowing detection of a wide range of taxa in a rapid, cost-effective way using a variety of sample types (Valentini et al. 2009; Taberlet et al. 2012; Creer et al~~.,~~. 2016; Stat et al~~.,~~. 2017~~Creer et al., 2016; Stat et al., 2017; Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012; Valentini, Pompanon, & Taberlet, 2009~~). In the last decade, these tools have been used to describe past and present biodiversity in terrestrial (Ji et al~~.,~~. 2013; Yoccoz et al~~.,~~. 2012; Yu et al~~.,~~. 2012; Slon et al. 2017; Pansu et al. 2015)~~(Ji et al., 2013; Pansu et al., 2015; Slon et al., 2017; Yoccoz et al., 2012; Yu et al., 2012)~~, freshwater (Valentini et al. 2016; Deiner et al. 2016; Bista et al~~.,~~. 2015; ~~Deiner, Fronhofer, Mächler, Walser, & Altermatt, 2016;~~ Dejean et al~~.,~~. 2011; Evans et al~~., 2016; Valentini et al.,~~. 2016)~~(Bista et al., 2015; Deiner, Fronhofer, Mächler, Walser, & Altermatt, 2016; Dejean et al., 2011; Evans et al., 2016; Valentini et al., 2016)~~, and marine (~~Bik et al., 2012; Boussarie et al., 2018; De Vargas et al., 2015;~~ Fonseca et al~~.,~~. 2010; ~~Massana~~Sinniger et al~~., 2015~~. 2016; Pawlowski et al~~.,~~. 2011; Massana et al. 2015; De Vargas et al. 2015; Salazar et al~~., 2016; Sinniger et al.,~~. 2016; Boussarie et al. 2018; Bik et al. 2012) environments.

As every new technique brings on new challenges, a number of studies have put considerable effort into delineating critical aspects of metabarcoding protocols to ensure robust and reproducible results (see Fig.1 in Fonseca et al, 2018). Recent studies have addressed many issues regarding sampling methods (Dickie et al.~~,~~ 2018), contamination risks (Goldberg et al., 2016), DNA extraction protocols (Brannock ~~&~~ and Halanych, 2015; Deiner et al., 2015; Zinger et al., 2016), amplification biases and required PCR replication levels (Nichols et al. 2018; Alberdi~~,~~ et al.~~Aizpurua, Gilbert, & Bohmann,~~ 2017; Ficetola et al., 2015~~; Nichols et al., 2018~~). Similarly, computational pipelines, through which molecular data are transformed into ecological inventories of putative taxa, have also been in constant improvement. ~~Indeed,~~ PCR-generated errors and sequencing errors are major bioinformatic challenges for metabarcoding pipelines, as they can strongly bias biodiversity estimates (~~Bokulich et al., 2013;~~ Coissac~~, Riaz, & Puillandre,~~et al. 2012; Bokulich et al. 2013). A variety of tools have thus been developed for quality-filtering amplicon data ~~and removing~~to remove erroneous reads ~~to~~ and improve the reliability of Illumina-sequenced ~~metabarcode~~ metabarcoding inventories (Bokulich et al.~~,~~ 2013; Eren~~, Vineis, Morrison, & Sogin,~~ et al. 2013; Minoche~~, Dohm, & Himmelbauer,~~et al. 2011). Studies that evaluated bioinformatic ~~parameters~~ processing steps have generally found ~~these~~ that sequence quality-filtering ~~steps, as well as arbitrarily set~~parameters and clustering thresholds ~~are the parameters that~~ most strongly affect molecular biodiversity inventories, resulting in considerable variation during data analysis~~produced by metabarcoding~~ (Brannock ~~&~~ and Halanych~~,~~ 2015; ~~Brown, Chain, Crease, MacIsaac, & Cristescu, 2015;~~ Clare~~, Chain, Littlefair, & Cristescu~~ et al.~~,~~ 2016; Brownet al. 2015; Xiong & Zhan, 2018).

There were historically two reasons for clustering sequences into Operational Taxonomic Units (OTUs). The first was to limit the bias due to PCR and sequencing errors (and to some extent also intra-individual variability linked to the existence of pseudogenes) by clustering erroneous ~~(and non-target)~~ sequences with error free target sequences. The second was to delineate OTUs as clusters of homologous sequences (by grouping the alleles/haplotype at the same locus) that would best fit a “species level”, i.e. the Operational Taxonomic Units defined using a classical phenetic *proxy* (Sokal & Crovello, 1970). Recent bioinformatic algorithms alleviate the influence of errors

Code de champ modifié

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

and intraspecific variability in metabarcoding datasets. First, ~~Recent bioinformatic algorithms for the processing of metabarcode data have been developed to alleviate the influence of these two parameters. A~~amplicon-specific error correction methods, commonly used to correct sequences produced by pyrosequencing (Coissac et al., 2012), have now become available for Illumina-sequenced data. ~~Published~~ Introduced in 2016, DADA2 effectively corrects Illumina sequencing errors and has quickly become a widely used tool~~ for Illumina sequence correction~~, particularly in the microbial world, producing more accurate biodiversity inventories and resolving fine-scale genetic variation~~s~~ by defining Amplicon Sequence Variants (ASVs) ~~(Callahan et al., 2016; Nearing, Douglas, Comeau, & Langille~~et al.~~,~~ 2018).

~~Low abundance molecular clusters remain an issue in metabarcoding biodiversity inventories, as it is challenging to discriminate valid but rare clusters from spurious ones. Singleton removal (clusters with less than 1-2 total reads) is largely advocated in the metabarcoding community (Clare et al., 2016) to limit the inflation of diversity due to the occurrence of spurious sequences. However, this method is arbitrary and potentially hinders the detection of rare species (Frøslev et al., 2017).~~Second, LULU is a ~~newly~~ recently developed curation algorithm designed to filter out ~~remaining~~ spurious clusters originating from PCR and sequencing errors~~,~~ or from intra-individual variability (pseudogenes, heteroplasmy), ~~based on objective criteria. Spurious clusters are detected~~ based on their similarity and co-occurrence rate with more abundant clusters, allowing obtaining curated datasets while avoiding arbitrary abundance filters (Frøslev et al.~~,~~ 2017). The authors ~~demonstrated~~ validated their approach on metabarcoding of plants using ITS2 (nuclear ribosomal internal transcribed spacer region 2) and ~~comparing~~ evaluated it on several pipelines. Their results show that ASV definition with DADA2, subsequent clustering to address intraspecific variation, and final curation with LULU is the safest pathway for ~~obtaining~~ producing reliable and accurate metabarcoding data. The authors concluded that their validation on plants is relevant to other organism groups and other markers, while recommending future validation of LULU on mock communities as LULU's minimum match parameter may need to be adjusted to less variable marker genes.

~~There were historically two reasons for clustering sequences into Operational Taxonomic Units (OTUs). The first was to limit the bias due to PCR and sequencing errors (and to some extent also intra-individual variability linked to the existence of pseudogenes) by clustering erroneous (and non-target) sequences with error free target sequences. The second was to delineate OTUs as clusters of sequences that would best fit a "species level", i.e. the Operational Taxonomic Units defined using a classical phenetic *proxy* (Sokal & Crovello, 1970).~~

The ~~first issue~~impact of errors being ~~largely~~ strongly decreased ~~solved~~ by ~~the two~~ correction algorithms such as DADA2 and LULU, the relevance of ~~the second objective, i.e. the delineation~~ clustering sequences into~~of~~ OTUs, is now being ~~discussed~~debated. Indeed, after presenting their new algorithm on prokaryotic communities, the authors of DADA2 proposed that the reproducibility and comparability of ASVs across studies challenge the need for clustering sequences, as OTUs have the disadvantage of being study-specific and defined using arbitrary thresholds ~~(Callahan, McMurdie, & Holmes, 2017)~~ et al. 2017). However, clustering sequences may still be necessary in metazoan datasets, where very distinct levels of intraspecific polymorphism can exist in the same gene region among taxa due to both evolutionary and biological specificity (Bucklin et al. 2011; Phillips et al. 2019). ASV-based inventories will thus be biased in favour of taxa with high levels of intraspecific diversity, even though the latter are not necessarily the most abundant ones (Bazin et al. 2006). Such bias in biodiversity inventories based on ASVs is likely to be magnified in presence-absence metabarcode datasets, commonly used for metazoan communities (Ji et al.~~,~~ 2013) ~~2006). Such bias in biodiversity inventories based on ASVs is likely to be magnified in presence-absence~~. Similarly, imposing a "universal" clustering threshold on metabarcoding datasets is also introducing bias, penalizing groups with lower interspecific divergence, and overestimating species diversity in groups with higher interspecific divergence. However, this can be alleviated with tools such as swarm v2, a single-linkage clustering algorithm (Mahe et al. 2015). ~~, McMurdie, & Holmes, 2017).~~ Based on network theory, swarm v2 aggregates sequences iteratively and locally around seed sequences and determines coherent groups of sequences, independent of amplicon input order, allowing highly scalable and fine-scale clustering. Finally~~Nevertheless~~, it is widely recognized that homogeneous entities sharing a set of evolutionary and ecological properties, i.e. species ~~(de Queiroz, 2005; (~~(Mayr, 1942; de Queiroz, 2005), sometimes ~~proposed~~ referred to ~~be designed~~ as "ecotypes" for prokaryotes (Cohan~~,~~ 2001; Gevers et al.~~,~~ 2005), represent a fundamental category of biological organization that is the cornerstone of most ecological and evolutionary theories and empirical studies. ~~Keeping~~ Maintaining ASV information for feeding databases and cross-comparing studies is not incompatible with their clustering into OTUs, and this choice depends on the purpose of the study~~ (~~, i.e. providing a census of the extent and distribution of genetic polymorphism for a given gene, or a census of biodiversity to be used and manipulated in ecological or evolutionary studies~~)~~. ~~In fact, obtaining a biodiversity inventory of metazoan communities without clustering is likely to deliver a dataset hard to manipulate and interpret in a community ecology framework. In such datasets each haplotype of the target gene in a given species will represent an ASV, yet very distinct levels of intraspecific polymorphism can exist in the same gene region due to both evolutionary and biological specificity (Bucklin, Steinke, & Blanco-Bercial, 2011; Phillips, Gillis, & Hanner, 2019). For COI for example, this has been reported among species sampled in the same habitats (Plouviez et al., 2009). ASV-based inventories will thus be biased in favour of taxa with high levels of intraspecific diversity, even though the latter are not necessarily the most abundant ones (Bazin, Glémin, & Galtier, 2006).~~

~~Such bias in biodiversity inventories based on ASVs is likely to be magnified in presence-absence metabarcode datasets, commonly used for metazoan communities (Ji et al., 2013).~~

~~Clustering sequences while avoiding arbitrary clustering thresholds is possible with tools such as swarm v2, a single-linkage clustering algorithm (Mahe, Rognes, Quince, De Vargas, & Dunthorn, 2015). Based on network theory, this algorithm aggregates sequences iteratively and locally around seed sequences and determines coherent groups of sequences independent of amplicon input order, allowing highly scalable, fine-scale clustering.~~

Here we evaluate ~~the performance of~~ DADA2 and LULU, using them alone and in combination with swarm v2, to ~~test~~ assess the possi~~bilities offered by~~performance of these new tools ~~on~~ for metabarcoding of metazoan communities. Using both ~~revealed using both a~~ mitochondrial COI marker (Leray et al., 2013) and the ~~18S~~ V1V2 region of 18S ~~(Sinniger et al., 2016) small subunit~~ ribosomal RNA (~~SSU~~ rRNA) ~~barcode marker. For each of the markers~~(Sinniger et al., 2016), we evaluated the ~~effect of read correction (using DADA2)~~need for clustering~~, clustering (using Swarm v2)~~, and the effectiveness of LULU curation to select ~~the~~ pipeline parameters delivering the most accurate resolution ~~in~~ of two deep-sea mock communities. We then test the different bioinformatic tools on a deep-sea sediment dataset in order to select an optimal trade-off between inflating biodiversity estimates and loosing rare biodiversity. As a baseline for comparison and in the perspective of the joint study of metazoan and microbial taxa, we also analysed the 16S~~-~~ V4V5 rRNA barcode on these natural samples (Parada, ~~Needham,~~ et al.~~& Fuhrman~~, 2016).

Our objectives were to ~~(1) select the most appropriate tools allowing avoiding inflating biodiversity estimates while retaining rare biodiversity and (2)~~ discuss the use of ASV ~~and~~ vs OTU-centred datasets depending on taxonomic compartment ~~of interest~~ and ~~on~~ study objectives, and (2) determine the most adequate swarm-clustering and LULU curation thresholds that avoid inflating biodiversity estimates while retaining rare biodiversity.

.

1 MATERIALS AND METHODS

1.1 Preparation of samples

Mock communities

Genomic-DNA mass-balanced metazoan mock communities (5 ng/µL) were prepared using standardized 10 ng/µL DNA extracts of ten deep-sea specimens belonging to five taxonomic groups (Polychaeta, Crustacea, Anthozoa, Bivalvia, Gastropoda; Table S1). Specimen DNA was extracted using a CTAB extraction protocol, from muscle tissue or from whole polyps in the case of cnidarians. The mock communities differed in terms of ratios of total genomic DNA from each species, with increased dominance of three species and secondary species DNA input decreasing from 3% to 0.7%. We individually barcoded the species present in the mock communities: PCRs of both target genes were performed using the same primers as the ones used in metabarcoding (see below). The PCR reactions (25 µL final volume) contained 2 µL DNA template with 0.5 µM concentration of each primer, 1X *Phusion* Master Mix, and an additional 1 mM $MgCl_2$ for COI. PCR amplifications (98 °C for 30 s; 40 cycles of 10 s at 98 °C, 45 s at 48 °C (COI) or 57 °C (18S), 30 s at 72 °C; and 72 °C for 5 min) were cleaned up with ExoSAP (Thermo Fisher Scientific, Waltham, MA, USA) and sent to Eurofins (Eurofins Scientific, Luxembourg) for Sanger sequencing. The barcode sequences obtained for all mock specimens were added to the databases used for taxonomic assignments of metabarcoding datasets, and were submitted on Genbank under accession numbers MN826120-MN826130 and MN844176-MN844185.

Environmental DNA

Sediment cores were collected from thirteen deep-sea sites ranging from the Arctic to the Mediterranean during various cruises (Table S2). Sampling was carried out with a multicorer ~~(MUC)~~ or with a remotely operated vehicle~~ (ROV).~~. Three tube cores were taken at each sampling station (GPS coordinates in Table S2). The ~~sediment cores~~latter were sliced into depth layers~~,~~ ~~which~~ that were transferred into zip-lock bags, homogenised, and frozen at −80°C on board before being shipped on dry ice to the laboratory. The first layer (0-1 cm) was used ~~for~~in the present ~~analysis~~study. DNA extractions were performed using approximately 10 g of sediment with the PowerMax Soil DNA Isolation Kit (Qiagen, Hilden, Germany). To increase the DNA yield, the elution buffer was left on the spin filter membrane for 10 min at room temperature before centrifugation. The ~5 mL extract was then split into three parts, one of which was kept in screw-cap tubes for archiving purposes and stored at -80°C. ~~Negative extraction controls were included in each extraction run~~For the four field controls, the first solution of the kit was poured into the control zip-lock bag, before following the usual extraction steps. For the two negative extraction controls, a blank extraction (adding nothing to the bead tube) was performed alongside sample extractions.

1.2 Amplicon library construction and high-throughput sequencing

Two primer pairs were used to amplify the mitochondrial ~~Cytochrome c Oxidase subunit I~~ ~~(~~COI~~)~~ and the 18S~~ V1V2 small-subunit ribosomal RNA (SSU~~ V1-V2 rRNA~~)~~ barcode genes specifically targeting metazoans, and one pair of primer was used to amplify the prokaryote 16S~~-~~ ~~V4V5~~ V4-V5 region~~ (Table S 3).~~. PCR amplifications, library preparation, and sequencing were carried out at ~~Génoscope~~Genoscope (Evry, France) as part of the eDNAbyss project.

Eukaryotic 18S ~~V1V2~~ V1-V2 rRNA gene amplicon generation

Amplifications were performed with the *Phusion* High Fidelity PCR Master Mix with GC buffer (~~ThermoFisher~~Thermo Fisher Scientific, Waltham, MA, USA) and the SSUF04 (5’-GCTTGTCTCAAAGATTAAGCC-3’) and SSUR22*mod* (5’- CCTGCTGCCTTCCTTRGA-3’) primers (Sinniger et al. 2016~~, Table S 3).~~), preferentially targeting metazoans, the primary focus of this study. The PCR reactions (25 µL final volume) contained 2.5 ng or less of DNA template with 0.4 µM concentration of each primer, 3% of DMSO, and 1X *Phusion* Master Mix. PCR amplifications (98 °C for 30 s; 25 cycles of 10 s at 98 °C, 30 s at 45 °C, 30 s at 72 °C; and 72 °C for 10 min) of all samples were carried out in triplicate in order to smooth the intra-sample variance while obtaining sufficient amounts of amplicons for Illumina sequencing.

Eukaryotic COI gene amplicon generation

Metazoan COI barcodes were generated using the mlCOIintF (5’-GGWACWGGWTGAACWGTWTAYCCYCC-3’) and jgHCO2198 (5’-TAIACYTCIGGRTGICCRAARAAYCA-3’) primers (Leray et al. 2013~~, Table S 3~~). Triplicate PCR reactions (20 µl final volume) contained 2.5 ng or less of total DNA template with 0.5 µM final concentration of each primer, 3% of DMSO, 0.175 mM final concentration of dNTPs, and 1X Advantage 2 Polymerase Mix (Takara Bio, Kusatsu, Japan). Cycling conditions included a 10 min denaturation step followed by 16 cycles of 95 °C for 10 s, 30s at 62°C (−1°C per cycle), 68 °C for 60 s, followed by 15 cycles of 95 °C for 10 s, 30s at 46°C, 68 °C for 60 s and a final extension of 68 °C for 7 min.

Prokaryotic 16S rRNA gene amplicon generation

Mis en forme : Titre 3, Gauche, Interligne : simple

Prokaryotic barcodes were generated using 515F-Y (5'- GTGYCAGCMGCCGCGGTAA-3') and 926R (5'- CCGYCAATTYMTTTRAGTTT-3') 16S-V4V5 primers (Parada et al. 2016). Triplicate PCR mixtures were prepared as described above for 18S-V1V2, but cycling conditions included a 30 s denaturation step followed by 25 cycles of 98 °C for 10 s, 53 °C for 30 s, 72 °C for 30 s, and a final extension of 72 °C for 10 min.

Mis en forme : Normal (Web), Retrait : Première ligne : 1.27 cm, Interligne : Double

Amplicon library preparation

~~Prokaryotic barcodes were generated using 515F-Y and 926R 16S-V4V5 primers (Parada et al., 2016)~~PCR~~. Triplicate PCR mixtures were prepared as described above for 18S-V1V2, but cycling conditions included a 30 s denaturation step followed by 25 cycles of 98 °C for 10 s, 53 °C for 30 s, 72 °C for 30 s, and a final extension of 72 °C for 10 min.~~

~~In all cases, amplicon~~ triplicates were~~ then~~ pooled and PCR products purified using 1X AMPure XP beads (Beckman Coulter, Brea, CA, USA) clean up. Aliquots of purified amplicons were run on an Agilent Bioanalyzer using the DNA High Sensitivity LabChip kit (Agilent Technologies, Santa Clara, CA, USA) to check their lengths and quantified with a Qubit fluorimeter (Invitrogen, Carlsbad, CA, USA).

Mis en forme : Normal (Web), Retrait : Première ligne : 1.27 cm, Interligne : Double

~~*Amplicon library preparation*~~

One hundred ~~ng~~nanograms of ~~amplicons~~pooled amplicon triplicates were directly end-repaired, A-tailed and ligated to Illumina adapters on a Biomek FX Laboratory Automation Workstation (Beckman Coulter, Brea, CA, USA). Library amplification was performed using a Kapa Hifi HotStart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA, USA) with the same cycling conditions applied for all metagenomic libraries and purified using 1X AMPure XP beads.

Sequencing library quality control

~~Libraries~~Amplicon libraries were quantified by Quant-iT dsDNA HS assay kits using a Fluoroskan Ascent microplate fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and then by qPCR with the KAPA Library Quantification Kit for Illumina Libraries (Kapa Biosystems, Wilmington, MA, USA) on an MxPro instrument (Agilent Technologies, Santa Clara, CA, USA). Library profiles were assessed using a high-throughput microfluidic capillary electrophoresis system (LabChip GX, Perkin Elmer, Waltham, MA, USA).

Sequencing procedures

Library concentrations were normalized to 10 nM by addition of 10 mM Tris-Cl (pH 8.5) and applied to cluster generation according to the Illumina Cbot User Guide (Part # 15006165). Amplicon libraries are characterized by low diversity sequences at the beginning of the reads due to the presence of the primer sequence. Low-diversity libraries can interfere in correct cluster identification, resulting in a drastic loss of data output. Therefore, loading concentrations of libraries were decreased (8–9 pM instead of 12–14 pM for standard libraries) and PhiX DNA spike-in was increased (20% instead of 1%) in order to minimize the impacts on the run quality.

Libraries were sequenced on HiSeq2500 (System User Guide Part # 15035786) instruments (Illumina, San Diego, CA, USA) in a 250 bp paired-end mode.

1.3 Bioinformatic analyses

All bioinformatic analyses were performed using a Unix shell script on a home-based cluster (DATARMOR, Ifremer), available on Gitlab (https://gitlab.ifremer.fr/abyss-project/). The mock communities were analysed alongside the natural samples, and used to validate the metabarcoding pipeline in terms of detection of correct species and presence of false-positives. The details of the pipeline, along with specific parameters used for ~~both~~all three metabarcoding markers~~,~~ are listed in Table ~~S 4~~S3.

Reads preprocessing

Our multiplexing strategy relies on ligation of adapters to amplicon pools, meaning that contrary to libraries produced by double PCR, the reads in each paired sequencing run can be forward or reverse. DADA2 correction is based on error distribution differing between R1 and R2 reads. We thus developed a custom script (*abyss-preprocessing* in abyss-pipeline) allowing separating forward and reverse reads in each paired run and reformatting the outputs to be compatible with DADA2. Briefly, the script uses cutadapt v1.18 to ~~separate~~detect and remove primers, while separating forward and reverse reads in each paired sequence file~~, producing~~ to produce two pairs of sequence files per sample named R1F/R2R and R2F/R1R~~, while removing primers based on a maximum error rate (e 0.17 for 18S V1 and 0.27 for COI , O~~. Cutadapt parameters (Table S3) were set to require an overlap over the full length of the primer ~~1).).~~(default: 3 nt), with 2-4 nt mismatches allowed for ribosomal loci, and 7 nt mismatches allowed for COI (default: 10%). Each identified forward and reverse read is then renamed which the correct extension (/1 and /2 respectively), which is a requirement for DADA2 to recognize the pairs of reads. Each pair of renamed sequence files is then re-paired with BBMAP Repair v38.22 in order to remove singleton reads (non-paired reads). Optionally, sequence file names can also be renamed if necessary using a CSV correspondence file.

Mis en forme : Retrait : Première ligne : 0 cm

Read correction, amplicon cluster generation and taxonomic assignment

Pairs of Illumina reads were corrected with DADA2 v.1.10 ~~(Callahan et al., 2016) following the online tutorial for paired-end data (https://benjjneb.github.io/dada2/tutorial.html). Reads were filtered and trimmed with the *filterAndTrim* function and all reads containing ambiguous bases removed~~(Callahan et al. 2016) following the online tutorial for paired-end HiSeq data (https://benjjneb.github.io/dada2/bigdata_paired.html). Reads were filtered and trimmed with the *filterAndTrim* function and all reads containing ambiguous bases removed. The parameters were set based on tutorial recommendations and trimming lengths were adjusted based on sequence quality profiles, so that Q-scores remained above 30 (truncLen at 220 for 18S and 16S, 200 for COI, maxEE at 2, truncQ at 11, maxN at 0).

The error model was calculated for forward and reverse reads (R1F/R2R pairs and then R2F/R1R pairs) with *learnErrors* based on 100 million randomly chosen bases~~,~~ (default), and reads were dereplicated using *derepFastq*. After read correction with the *dada* function, forward and reverse reads were merged with a minimum overlap of 12 nucleotides, allowing no mismatches~~.~~ (default). The amplicons were then filtered by size. The size range was set to 330-390 bp for the 18S SSU rRNA marker gene, 300-326 bp for the COI marker gene, and 350-390 bp for the 16S rRNA marker gene.

Chimeras were removed with *removeBimeraDenovo* and ASVs were taxonomically assigned via the RDP naïve Bayesian classifier method, the default assignment method implemented in DADA2. ~~A second taxonomic assignment method was optionally implemented in~~

~~the pipeline, allowing assigning ASVs using BLAST+ (v2.6.0) based on minimum similarity and minimum coverage (-perc_identity 70 and –qcov_hsp 80). The Silva132 reference database was used for the 16S and 18S SSU rRNA marker genes (Quast et al., 2012), and MIDORI-UNIQUE (Machida, Leray, Ho, & Knowlton, 2017) was used for COI. The databases were downloaded from the DADA2 website (https://benjjneb.github.io/dada2/training.html) and from the FROGS website (http://genoweb.toulouse.inra.fr/frogs_databanks/assignation/). We individually barcoded the species present in the mock communities and added their barcode sequences to all the databases. Finally, to evaluate the effect on clustered data when OTUs are to be produced, ASV tables produced by DADA2 were clustered with swarm v2 (Mahe et al., 2015) at *d=4* for 18S, *d=6* for COI, and *d=1* for 16S in FROGS (http://frogs.toulouse.inra.fr/) (Escudié et al., 2018). Resulting OTUs were taxonomically assigned via~~A second taxonomic assignment method was optionally implemented in the pipeline, allowing assigning ASVs using BLAST+ (Basic Local Alignment Search Tool v2.6.0) based on minimum similarity and minimum coverage (-perc_identity 70 and –qcov_hsp 80). An initial test implementing BLASTn+ to assign taxonomy only to the COI dataset using a 96% percent identity threshold led to the exclusion of the majority of the clusters. Given observed inter-specific mitochondrial DNA divergence levels of up to 30% within a same polychaete genus (Zanol et al. 2010) or among some closely related deep-sea shrimp species (Shank et al. 1999), and considering our interest in the identities of multiple, largely unknown taxa in poorly characterized communities, more stringent BLAST thresholds were not implemented at this stage. The Silva132 reference database was used for the 16S and 18S SSU rRNA marker genes (Quast et al. 2012), and MIDORI-UNIQUE (Machida et al. 2017) was used for COI. The databases were downloaded from the DADA2 website (https://benjjneb.github.io/dada2/training.html) and from the FROGS website (http://genoweb.toulouse.inra.fr/frogs_databanks/assignation/). Finally, to evaluate the effect of clustering, ASV tables produced by DADA2 were clustered with swarm v2 (Mahe et al. 2015) at *d=1,3,4,5 and 11* for 18S and 16S, and *d=1,5,6,7, and 13* for COI in FROGS (http://frogs.toulouse.inra.fr/) (Escudié et al. 2018). Resulting OTUs were taxonomically assigned via RDP and BLAST+ using the databases stated above.

~~Molecular clusters were refined in R v.3.5.1 (R Core Team, 2018). A blank correction was made using the *decontam* package v.1.2.1 (Davis, Proctor, Holmes, Relman, & Callahan, 2018), removing all clusters that were more abundant in negative control samples than in other samples. ASV/OTU tables were refined taxonomically based on their RDP or BLAST taxonomy. For both assignment methods, unassigned clusters were removed. Non-target 18S and COI clusters (bacterial, non-metazoan) as well as all clusters with a terrestrial assignment (taxonomic groups known to be terrestrial-only, such as Insecta, Arachnida, Diplopoda, Amphibia, terrestrial mammals, Stylommatophora, Aves, Onychophora, Succineidae, Cyclophoridae, Diplommatinidae, Megalomastomatidae, Pupinidae, Veronicellidae) were removed. Samples were checked to ensure that a minimum of 10,000 metazoan reads were left after refining. Finally, an abundance renormalization was performed to remove spurious positive results due to random tag switching (Wangensteen & Turon, 2016).~~

Molecular clusters were refined in R v.3.5.1 (R Core Team 2018). A blank correction was made using the *decontam* package v.1.2.1 (Davis et al. 2018), removing all clusters that were prevalent (more frequent) in negative control samples. ASV/OTU tables were refined taxonomically based on their RDP or BLAST taxonomy. For both assignment methods, unassigned clusters were removed. Non-target 18S and COI clusters (bacterial, non-metazoan) as well as all clusters with a terrestrial assignment (taxonomic groups known to be terrestrial-only, such as Insecta, Arachnida, Diplopoda, Amphibia, terrestrial mammals, Stylommatophora, Aves, Onychophora, Succineidae, Cyclophoridae, Diplommatinidae, Megalomastomatidae, Pupinidae, Veronicellidae) were removed. Samples were checked to ensure that a minimum of 10,000 metazoan reads were left after refining. Finally, as tag-switching is always to be expected in multiplexed metabarcoding analyses (Schnell et al. 2015), an abundance renormalization was performed to remove spurious positive results due to reads assigned to the wrong sample (Wangensteen and Turon 2016, script from https://github.com/metabarpark/R_scripts_metabarpark).

To test LULU curation ~~(Frøslev et al., 2017)~~(Frøslev et al. 2017), refined 18S and COI ASVs/OTUs were curated with LULU v.0.1 following the online tutorial (https://github.com/tobiasgf/lulu). The LULU algorithm detects erroneous clusters by comparing their sequence similarities and co-occurrence rate with more abundant ("parent") clusters. LULU was tested with a minimum relative co-occurrence of 0.90~~ and~~, using a minimum similarity threshold (*minimum match*) ~~threshold of~~at 84% ~~and~~(default) and slightly higher at 90~~%.~~%, following recommendations of the authors for less variable loci than ITS.

The vast majority of prokaryotes usually show low levels (< 1% divergence) of intra genomic variability for the 16S SSU rRNA gene ~~(Acinas, Marcelino, Klepac-Ceraj, & Polz, 2004; Pei et al., 2010). Although we acknowledge that for a limited amount of cases, curation with LULU may still be useful to obtain a more rigorous census of biodiversity, this was not tested on the prokaryote communities used in this study. Indeed, parallelization not being currently available for LULU curation, the richness of those communities implied an unrealistic amount of calculation time, even on a powerful cluster (several weeks~~(Acinas et al. 2004; Pei et al. 2010). These low intragenomic divergence levels can be efficiently removed with swarm clustering at d=1. Although LULU curation may still be useful to merge redundant phylotypes in specific cases such as haplotype network analyses, this was not tested in this study. Indeed, parallelization not being currently available for LULU curation, the richness of prokaryote communities implied an unrealistic calculation time, even on a powerful cluster (e.g. LULU curation was at 20-40% after 4 days of calculation on our cluster).

~~**1.4 Statistical analyses**~~

~~Sequence tables were analysed using R with the packages phyloseq v1.22.3 (McMurdie & Holmes, 2013) following guidelines on online tutorials (http://joey711.github.io/phyloseq/tutorials-index.html), and vegan v2.5.2 (Oksanen et al., 2018). Each biodiversity inventory and its LULU curated version were merged into a single phyloseq object. The datasets were normalized by rarefaction to their common minimum sequencing depth, before analysis of the mock communities and the natural samples.~~

1.4 Statistical analyses

Sequence tables were analysed using R with the packages phyloseq v1.22.3 (McMurdie and Holmes 2013) following guidelines on online tutorials (http://joey711.github.io/phyloseq/tutorials-index.html), and vegan v2.5.2 (Oksanen et al. 2018). The datasets were normalized by rarefaction to their common minimum sequencing depth, before analysis of mock communities and natural samples.

To evaluate the functionality of the pipeline with the mock communities, taxonomically assigned metazoan clusters were considered as derived from one of the ten species used for the mock communities when the assignment delivered the corresponding species, genus, family, or class. Clusters not fitting the expected taxa were labelled as ‘Others’. ~~These~~Apart from PCR errors, these non-target clusters may ~~be spurious or reflect~~also originate from contamination by external DNA ~~or~~from associated microfauna, ~~such as commensals or parasites, which might have been present~~or gut content in the ~~extracted tissue~~case of whole polyps used for cnidarians.

Alpha diversity detected using each pipeline in the natural samples was evaluated with the number of observed target-taxa in the rarefied datasets via analyses of ~~deviance (ANODEV~~variance (ANOVA) on generalized linear models based on quasipoisson distribution models. Homogeneity of multivariate dispersions were verified with the *betapart* package v.1.5.1 ~~(Baselga & Orme, 2012)~~(Baselga and Orme 2012). Beta-diversity patterns were visualised via Principal Coordinates Analyses (PCoA), using Jaccard dissimilarities for metazoans and Bray-Curtis dissimilarities for prokaryotes. The effect of site and LULU curation~~, site and sediment core (nested within site)~~ on community composition was tested by means of PERMANOVA~~ on the rarefied incidence datasets. PERMANOVAs were calculated,~~ using the function ~~*adonis*~~*adoni*s2 (vegan), with ~~Jaccard~~the same dissimilarities as in PCoAs, and~~ 9999 permutations,~~ permuting ~~within sites for evaluating the Pipeline and Core effects.~~ 999 times.Finally, BLAST and RDP taxonomic ~~compositions in terms~~assignments of ~~cluster abundance~~the mock samples and the global dataset were compared ~~between pipelines and with results of a morphological inventory obtained from a first~~at the most adequate pipeline settings for each locus. BLAST-refined (minimum identity at 70%) and RDP-refined (minimum phylum bootstrap at 80%) datasets were compared on ASV-level ~~sorting in two sites~~for prokaryotes, and OTU-level for metazoans (swarm *d=3*, LULU at 84% for COI and 90% for 18S). As trials on MIDORI-UNIQUE resulted in very poor performance of RDP for COI (assignments belonging mostly to Insecta), the comparison was performed with MIDORI-UNIQUE subsampled to marine taxa only.

2 RESULTS

2.1 Alpha diversity in mock communities

A number of 2 million (18S) and 1.5 million (COI) raw reads were obtained from the two mock communities (Table S4). After refining, these numbers were decreased to 1.3 million for 18S and 0.7 million for COI.

Seven out of ten mock species were recovered in the 18S dataset and all species were detected in the COI dataset (Table 1), even with minimum relative DNA abundance levels as low as 0.7% (Mock 5). Taxonomically unresolved species were correctly assigned up to their common family or class level. Dominant species generally produced more reads in both the clustered and non-clustered datasets (Table S6).

When ASVs were clustered with swarm v2, this generally led to a slight loss of taxonomic resolution: *Chorocaris* sp. was not detected in Mock 5 for 18S at d > *1*, and the two bivalves *P. kilmeri* and *C. regab* were taxonomically misidentified for COI at d ≥ 1.

Clustering sequences with swarm v2 reduced the number of clusters produced per species, but some species still produced multiple OTUs even at *d* values as high as *d=11* for 18S (*A. arbuscula*, *Munidopsis* sp., and *E. norvegica*) and *d=13* for COI *D. dianthus*, *A. muricola*, *Chorocaris* sp., and *Paralepetopsis* sp.). Curating with LULU allowed reducing the number of clusters produced per species to nearly one for both loci, but the best results were obtained in datasets clustered at d > 1 for 18S and d ≥ 1 for COI. Moreover, LULU curation tended to decrease the number of non-target clusters ("Others") (Table 1). In the clustered COI dataset, curating with LULU at 84% *minimum match* resulted in the most accurate detection of community composition, and this for all *d* values tested. However, curating with LULU the 18S data (ASVs or OTUs) led to the loss of one shrimp species (*Chorocaris* sp) when the *minimum match* parameter was at 90% and an additional species was lost (the limpet *Paralepetopsis* sp.) when this parameter was at 84%.

LULU consistently merged the shrimp species *Chorocaris* sp with another shrimp species as the latter were always co-occurring in our mock samples.

Table 1. Number of ASVs/OTUs detected per species in the mock communities using different bioinformatic pipelines. White cells indicate an exact match with the number of OTUs expected, grey cells indicate a number of OTUs differing by ±3 from the number expected, and dark grey cells indicate a number of OTUs >3 from the one expected.

18S	**DADA2**	**DADA2 +LULU 90%**	**DADA2+ LULU 84%**		**DADA2+swarm d1/d3/d4/d5/d11**	**DADA2+swarm d1/d3/d4/d5/d11 + LULU 90%**	**DADA2+swarm d1/d3/d4/d5/d11 + LULU 84%**
Mock 3							
Alcyonacea;*A.arbuscula*	64	1	1	Alcyonacea;*A.arbuscula*	29/11/9/7/6	1/1/1/1/1	1/1/1/1/1
Caryophylliidae;*D.dianthus*	2	1	1	Caryophylliidae;*D.dianthus*	2/2/1/1/1	1/1/1/1/1	1/1/1/1/1
Alvinocaris muricola	2	1	1	*Alvinocaris muricola*	2/1/1/1/1	1/1/1/1/1	1/1/1/1/1
Chorocaris sp.	1	0	0	*Chorocaris* sp.	2/1/1/1/1	0/0/0/0/0	0/0/0/0/0
Munidopsis sp.	6	1	1	*Munidopsis* sp.	5/4/3/3/2	1/1/1/1/1	1/1/1/1/1
Gastropoda;*Paralepetopsis* sp.	1	1	0	Gastropoda;*Paralepetopsis* sp.	1/1/1/1/1	1/1/1/1/1	0/0/0/0/0
Vesicomyidae;*P. kilmeri/C. regab/V. gigas*	8	1	1	Bivalvia;*P. kilmeri/C. regab/V. gigas*	5/4/4/4/2	1/2/2/2/1	1/1/1/1/1
Polychaeta;*E.norvegica*	8	3	2	Polychaeta;*E.norvegica*	5/4/4/4/3	3/2/2/2/2	2/1/2/2/2
Others	3	3	2	Others	4/4/4/4/4	2/2/2/2/3	2/2/2/2/2
Mock 5							
Alcyonacea;*A.arbuscula*	54	1	1	Alcyonacea;*A.arbuscula*	28/11/9/7/6	1/1/1/1/1	1/1/1/1/1
Caryophylliidae;*D.dianthus*	1	1	1	Caryophylliidae;*D.dianthus*	1/1/1/1/1	1/1/1/1/1	1/1/1/1/1
Alvinocaris muricola	1	1	1	*Alvinocaris muricola*	1/1/1/1/1	1/1/1/1/1	1/1/1/1/1
Chorocaris sp.	1	0	0	*Chorocaris* sp.	1/0/0/0/0	0/0/0/0/0	0/0/0/0/0
Munidopsis sp.	4	1	1	*Munidopsis* sp.	4/3/3/3/2	1/1/1/1/1	1/1/1/1/1
Gastropoda;*Paralepetopsis* sp.	1	1	0	Gastropoda;*Paralepetopsis* sp.	1/1/1/1/1	1/1/1/1/1	0/0/0/0/0
Vesicomyidae;*P. kilmeri/C. regab/V. gigas*	5	1	1	Bivalvia;*P. kilmeri/C. regab/V. gigas*	5/3/3/3/2	1/1/1/1/1	1/1/1/1/1
Polychaeta;*E.norvegica*	11	3	2	Polychaeta;*E.norvegica*	5/4/4/4/3	3/2/2/2/1	2/1/2/2/2
Others	4	3	2	Others	3/4/4/4/2	4/2/2/2/1	4/2/2/2/3
COI	**DADA2**	**DADA2 +LULU 90%**	**DADA2+ LULU 84%**		**DADA2+swarm d1/d5/d6/d7/d13**	**DADA2+swarm d1/d5/d6/d7/d13 + LULU 90%**	**DADA2+swarm d1/d5/d6/d7/d13 + LULU 84%**
Mock 3							
Acanella arbuscula	1	1	1	*Acanella arbuscula*	1/1/1/1/1	1/1/1/1/1	1/1/1/1/1
Hexacorallia;*D.dianthus*	3	3	3	Hexacorallia;*D.dianthus*	3/4/4/4/3	3/3/3/3/3	3/3/3/3/3
Alvinocaris ;*A. muricola*	26	2	2	Alvinocaris;*A. muricola*	21/12/10/10/5	1/1/1/1/1	1/1/1/1/1
Chorocaris sp.	2	1	1	*Chorocaris* sp.	3/3/3/3/3	1/1/1/1/1	1/1/1/1/1
Munidopsis sp.	2	1	1	*Munidopsis* sp.	3/2/1/1/1	2/1/1/1/1	1/1/1/1/1
Gastropoda;*Paralepetopsis* sp.	8	2	3	Gastropoda;*Paralepetopsi* s sp.	3/3/3/3/2	2/2/2/2/2	2/2/2/2/2
Phreagena kilmeri	2	1	1	Bivalvia;*P. kilmeri*	2/3/3/3/3	2/2/2/2/2	2/2/2/2/2
Bivalvia;*C. regab*	2	1	1	Bivalvia;*C. regab*			
Vesicomya gigas	1	1	1	*Vesicomya gigas*	1/1/1/1/1	1/1/1/1/1	1/1/1/1/1
Polychaeta;*E.norvegica*	3	2	1	*Eunice norvegica*	2/1/1/1/1	2/1/1/1/1	1/1/1/1/1
Others	7	6	6	Others	3/3/3/3/4	4/5/5/5/5	5/5/5/5/5
Mock 5							
Acanella arbuscula	1	1	1	*Acanella arbuscula*	1/1/1/1/1	1/1/1/1/1	1/1/1/1/1
Hexacorallia;*D.dianthus*	3	3	3	Hexacorallia;*D.dianthus*	3/3/3/3/3	3/3/3/3/3	3/3/3/3/3
Alvinocaris ;*A. muricola*	26	2	2	Alvinocaris;*A. muricola*	21/12/10/10/5	1/1/1/1/1	1/1/1/1/1
Chorocaris sp.	1	1	1	*Chorocaris* sp.	2/2/2/2/2	1/1/1/1/1	1/1/1/1/1
Munidopsis sp.	2	1	1	*Munidopsis* sp.	2/2/1/1/1	1/1/1/1/1	1/1/1/1/1
Gastropoda;*Paralepetopsis* sp.	5	2	2	Gastropoda;*Paralepetopsis* sp.	3/2/2/2/2	2/2/2/2/2	2/2/2/2/2
Phreagena kilmeri	1	1	1	Bivalvia;*P. kilmeri*	2/2/2/2/2	2/2/2/2/2	2/2/2/2/2
Bivalvia;*C. regab*	2	1	1	Bivalvia;*C. regab*			
Vesicomya gigas	1	1	1	*Vesicomya gigas*	1/1/1/1/1	1/1/1/1/1	1/1/1/1/1
Polychaeta;*E.norvegica*	3	2	1	*Eunice norvegica*	2/2/2/2/2	1/1/1/1/1	1/1/1/1/1
Others	6	5	4	Others	2/2/2/2/2	1/2/2/2/2	1/1/1/1/1

.

2.2 Alpha-diversity patterns in natural samples

~~2.1~~ *High~~-~~-throughput* ~~DNA~~ *sequencing results*

A number of ~~45,828,979~~ 44 million (18S ~~reads, 34,639,914~~), 33 million (COI ~~reads~~) and 16~~,406,877~~ million (16S) reads were obtained from ~~six Illumina HiSeq runs of pooled amplicon libraries built from~~ 42 sediment samples, ~~2 mock communities (for 18S and COI), 6~~4 field controls, 2 extraction blanks, and 4-10 PCR ~~negative controls~~blanks (Table ~~1~~S4). Two sediment samples failed amplification for the COI marker gene (PCT_FA_CT2_0_1 and CHR_CT1_0_1). For metazoans, less reads were retained after bioinformatic processing in negative controls (36% ~~kept~~ for 18S, 47% for COI) ~~than in~~compared to true ~~or mock~~ samples (~60% ~~kept~~ for 18S, ~70~~-80~~% for COI), while the opposite was observed for 16S (74% of reads retained in control samples against 53% in true samples). ~~In total, 25,773,684 18S reads, 24,244,902 COI reads, and 9,446,242 16S reads remained after processing with DADA2.~~ Negative control samples (field, extraction, and PCR ~~blanks~~controls) contained 2,186,230 (~8%) 18S reads, 1,015,700 (~4%) COI reads, and 2,618,729 (28%) 16S reads. These reads were mostly originating from the ~~extraction~~field controls ~~(59~~for metazoans (48% for 18S, ~~65~~55% for COI~~,~~) and ~~72% for 16S). The corresponding clusters were removed from real samples if the number of reads in true samples was lower than in the negative~~extractions controls~~.~~ for 16S (50%).

After blank correction, data refining, and abundance renormalization, rarefaction curves showed that a plateau was achieved for all samples in both clustered and non-clustered datasets, suggesting an overall sequencing depth adequate to capture the diversity present (Fig. S1). The final 18S datasets (with and without clustering at selected *d* values) contained 8.9-9.6 million marine metazoan reads in 42 sediment samples (Table S4), and comprised 57,661 ASVs and

Mis en forme : Police :Italique

Mis en forme : Police :Italique

Mis en forme : Normal, Gauche

Mis en forme : Police :Italique

Mis en forme : Police :Italique

19,504-44,948 OTUs (Table S6). The final COI datasets contained 4.5-6.9 million marine metazoan reads in 40 sediment samples, and comprised 78,785 ASVs and 44,684-64,669 OTUs. The 16S datasets contained from 6.6 to 6.7 million prokaryotic reads in 42 sediment samples, producing 56,577 ASVs and 41,746-14,631 OTUs.

Mis en forme : Gauche

~~Table 1. Number of reads, ASVs, and OTUs obtained in samples after each pipeline step. Data refining was performed in R, based on BLAST assignments. Forward slashes separate ASV/OTU datasets (Dada2 without swarm clustering / Dada2 with swarm clustering).~~

Sample type	Number of samples	Raw reads	Quality-filtered reads	Merged reads	Reads before chimera removal	Non chimeric reads	% reads retained	Number of ASVs/OTUs before refining	Number of samples after refining	Number of target reads after refining	Number of target reads after renormalisation	Final number of target ASVs/OTUs	Number of target OTUs after LULU 84%	Number of target OTUs after LULU 90%
LOCUS														
18S-V1														
Control Sample	14	6 141 567	2 508 908	2 441 821	2 200 132	2 186 230	35,6	57,661 / 31,509	0	10,234,660 / 10,686,911	10,160,603 / 10,541,499	11,304 / 5,877	2,132 / 1,535	3, 639 / 2,889
Mock Sample	2	2 096 631	1 607 219	1 436 773	1 430 823	1 289 608	61,5		2					
True Sample	42	37 590 781	26 828 194	24 826 430	22 636 689	22 297 846	59,3		42					
COI														
Control Sample	16	2 146 476	1 053 997	1 024 547	1 015 821	1 015 700	47,3	78,785 / 52,216	0	7,601,973 / 5,179,905	7,552,406 / 5,129,293	21,663 / 8,249	11,987 / 4,849	17,265 / 7,251
Mock Sample	2	1 482 785	1 261 045	1 252 908	1 251 994	1 224 795	82,6		2					
True Sample	40	31 010 653	26 011 238	25 287 002	22 197 457	22 004 407	71,0		40					
16S - V4V5														
Control Sample	10	3,531,226	2,889,163	2,634,536	2,619,479	2,618,729	74.2	56,577 / 41,746	0	6,809,966 / 6,801,953	6,719,153 / 6,680,238	55,129 / 40,459	-	-
True Sample	42	12,875,651	9,307,729	7,122,154	7,114,195	6,827,513	53		42					

~~The 18S ASV dataset comprised 10,160,603 marine metazoan reads, with an average of 230,923 per sample (range of 42,119-721,972). When clustered with swarm v2, the final 18S dataset comprised 10,541,499 target reads, with an average of 239,579 per sample (range 45,259-721,753). The final COI ASV dataset comprised 7,552,406 marine metazoan reads, with an average of 179,819 per sample, (range of 54,585-438,324). When clustered with swarm v2, the final COI dataset comprised 5,129,293 target reads, with an average of 122,126 per sample (range of 31,228-349,805). The 16S ASV dataset comprised 6,719,153 prokaryotic reads, with an average of 159,979 per sample (range of 71,834 – 251,054). When clustered with swarm v2, the final 16S dataset comprised 6,680,238 prokaryotic reads, with an average of 159,253 per sample (range 71,601 - 250,032).~~

~~From the total 57,661 ASVs detected for 18S, 47,084 (82%) were assigned by BLAST to phylum level or lower. The assigned ASVs accounted for 97% of total 18S reads. BLAST detected 11,304 marine metazoan ASVs (Table 1). Samples contained 389 target ASVs on average, with a range of 88-881 per sample. LULU curation of 18S ASVs at 84% *minimum match* resulted in 2,132 clusters (134 per sample on average, range of 11-273), while 3,639 clusters remained after LULU curation at 90% *minimum match* (186 per sample on average, range of 14-402) (Table 1). From the total 31,509 18S OTUs obtained after clustering with swarm v2 (Mahe et al., 2015) at *d=4* (~1% divergence), 22,427 (71%) were assigned to phylum level or lower The assigned OTUs accounted for 93% of 18S reads. This resulted in 5,877 marine metazoan OTUs after data refining (286 metazoan clusters per sample on average, range of 29-698). The number of metazoan OTUs was reduced to 1,535 and 2,889 after LULU curation at 84% and 90% *minimum match* respectively (136 and 196 metazoan clusters per sample on average, range of 10-268 and 12-404 respectively).~~

~~The number of raw ASVs yielded by COI was higher: 78,785 from which 46,301 (59%) were assigned to phylum level or lower. The assigned ASVs accounted for 65% of total COI reads.~~

~~After data refining, BLAST identified 21,663 marine metazoan ASVs in the COI dataset (Table 1). Samples contained 914 ASVs on average, with a range of 56-1,955 per sample. LULU curation of COI ASVs at 84% *minimum match* resulted in 11,987 clusters (599 per sample on average, range of 22-1,210), while 17,265 clusters remained after LULU curation at 90% *minimum match* (787 per sample on average, range of 23-1,697). From the 52,216 COI OTUs obtained after clustering ASVs with swarm v2 at $d=6$ (~2% divergence), 21,924 (42%) were assigned to phylum level or lower. The assigned OTUs represented 52% of COI reads. After data refining, 8,249 marine metazoan COI OTUs remained in the dataset (470 per sample on average, range of 28-1,069). This number was reduced to 4,849 and 7,251 after LULU curation at 84% and 90% *minimum match* respectively (333 and 434 clusters per sample on average, range of 17-671 and 17-990 respectively).~~

~~From the total 56,577 ASVs detected for 16S, 55,804 (98.6%) were assigned by BLAST at phylum level or lower. The assigned ASVs accounted for 99.9% of total 16S reads, resulting in 55,129 final ASVs (Table 1). From the total 41,746 16S OTUs obtained after clustering with swarm v2 (Mahe et al., 2015) at $d=1$, 40,768 (97.7%) were assigned to phylum level or lower, resulting in 40,459 final OTUs.~~

~~Refining the ASV datasets based on RDP taxonomy resulted in decreased metazoan detection levels, but this was not the case for prokaryotes (Table S 5). For 18S, only 45% of ASVs could be assigned to phylum level or lower, resulting in 8,365 marine metazoan ASVs. For COI, although RDP assigned 76% of ASVS, only 2,526 target ASVs could be retrieved. We therefore reduced our COI database to only marine sequences. This resulted in 11% of assigned ASVs, but increased the number of target clusters to 8,466 (Table S 6).~~

~~2.2 Performance on mock samples~~

~~Assigning ASVs with BLAST allowed recovering 7 out of 10 mock species in the 18S dataset and all species in the COI dataset (Table 2), even with minimum relative DNA abundance levels as low as 0.7% (Mock 5).~~

~~When ASVs were clustered with swarm v2, this generally led to a slight loss of taxonomic resolution (*Chorocaris* sp. was not detected in Mock 3 for 18S and the two bivalves *P. kilmeri* and *C. regab* were taxonomically misidentified for COI). Taxonomically unresolved species were correctly assigned up to their common family or class level. Dominant species generally produced more reads in both the clustered and non-clustered datasets (Table S 7).~~

~~Clustering sequences with swarm v2 reduced the number of clusters produced per species, but some species still produced multiple (up to 10) OTUs (*A. arbuscula*, *Munidopsis* sp., and *E. norvegica* for 18S; *A. muricola*, *D. dianthus*, *Chorocaris* sp., and *Paralepetopsis* sp. for COI). Curating with LULU allowed reducing the number of clusters produced per species to nearly one, with and without clustering, and this for both loci. Moreover, LULU curation decreased the number of spurious clusters (“Others”), but this effect was more marked for 18S and at 84% *minimum match* (Table 2). However, curating with LULU the 18S data (ASVs or OTUs) led to the loss of one shrimp species (*Chorocaris* sp) when the *minimum match* parameter was at 90% and an additional species (the limpet *Paralepetopsis* sp.) when this parameter was at 84%. LULU consistently merged the shrimp species *Chorocaris* sp with another shrimp species as the latter were always co-occurring in our mock samples.~~

~~Table 2. Number of ASVs/OTUs detected per species in the mock communities using different bioinformatic pipelines. White cells indicate an exact match with the number of OTUs expected, grey cells indicate a number of OTUs differing by ±3 from the number expected, and dark grey cells indicate a number of OTUs >3 from the one expected.~~

18S	**DADA2**	**DADA2 +LULU 84%**	**DADA2+ LULU 90%**		**DADA2 +swarm**	**DADA2+swarm +LULU 84%**	**DADA2+swarm +LULU 90%**
Mock 3							
Alcyonacea;*A.arbuscula*	64	1	1	Alcyonacea;*A.arbuscula*	9	1	1
Caryophylliidae;*D.dianthus*	2	1	1	Caryophylliidae;*D.dianthus*	1	1	1
Alvinocaris muricola	2	1	1	*Alvinocaris muricola*	1	1	1
Chorocaris sp.	1	0	0	*Chorocaris* sp.	0	0	0
Munidopsis sp.	6	1	1	*Munidopsis* sp.	3	1	1
Gastropoda;*Paralepetopsis* sp.	1	0	1	Gastropoda;*Paralepetopsis* sp.	1	0	1
Vesicomyidae;*P. kilmeri/C. regab/V. gigas*	8	1	1	Bivalvia;*P. kilmeri/C. regab/V. gigas*	3	1	1
Polychaeta;*E.norvegica*	8	2	3	Polychaeta;*E.norvegica*	4	2	2
Others	3	2	3	Others	4	2	2
Mock 5							
Alcyonacea;*A.arbuscula*	54	1	1	Alcyonacea;*A.arbuscula*	9	1	1
Caryophylliidae;*D.dianthus*	1	1	1	Caryophylliidae;*D.dianthus*	1	1	1
Alvinocaris muricola	1	1	1	*Alvinocaris muricola*	1	1	1
Chorocaris sp.	1	0	0	*Chorocaris* sp.	1	0	0
Munidopsis sp.	4	1	1	*Munidopsis* sp.	3	1	1
Gastropoda;*Paralepetopsis* sp.	1	0	1	Gastropoda;*Paralepetopsis* sp.	1	0	1
Vesicomyidae;*P. kilmeri/C. regab/V. gigas*	5	1	1	Bivalvia;*P. kilmeri/C. regab/V. gigas*	4	1	2
Polychaeta;*E.norvegica*	11	2	3	Polychaeta;*E.norvegica*	4	2	2
Others	4	2	3	Others	4	2	2
COI	**DADA2**	**DADA2 +LULU 84%**	**DADA2+ LULU 90%**		**DADA2 +swarm**	**DADA2+swarm +LULU 84%**	**DADA2+swarm +LULU 90%**
Mock 3							
Acanella arbuscula	1	1	1	*Acanella arbuscula*	1	1	1
Hexacorallia;*D.dianthus*	3	3	3	Hexacorallia;*D.dianthus*	4	3	3
Alvinocaris ;A. muricola	26	2	2	Alvinocaris;*A. muricola*	10	1	1
Chorocaris sp.	2	1	1	*Chorocaris* sp.	3	1	1
Galatheidae;*Munidopsis* sp.	2	2	1	*Munidopsis* sp.	1	1	2
Gastropoda;*Paralepetopsis* sp.	8	3	3	Gastropoda;*Paralepetopsi* s sp.	3	2	2
Phreagena kilmeri	2	1	1	Bivalvia;*P. kilmeri*	3	2	2
Bivalvia;*C. regab*	2	1	1	Bivalvia;*C. regab*			
Vesicomya gigas	1	1	1	*Vesicomya gigas*	1	1	1
Polychaeta;*E.norvegica*	3	3	1	*Eunice norvegica*	1	1	1
Others	7	5	6	Others	3	4	5
Mock 5							
Acanella arbuscula	1	1	1	*Acanella arbuscula*	1	1	1
Hexacorallia;*D.dianthus*	3	3	3	Hexacorallia;*D.dianthus*	3	3	3
Alvinocaris ;A. muricola	26	2	2	Alvinocaris;*A. muricola*	9	1	1
Chorocaris sp.	1	1	1	*Chorocaris* sp.	2	1	1
Galatheidae;*Munidopsis* sp.	2	1	1	*Munidopsis* sp.	1	1	1
Gastropoda;*Paralepetopsis* sp.	7	2	2	Gastropoda;*Paralepetopsis* sp.	3	2	3
Phreagena kilmeri	1	1	1	Bivalvia;*P. kilmeri*	2	2	2
Bivalvia;*C. regab*	2	1	1	Bivalvia;*C. regab*			
Vesicomya gigas	1	1	1	*Vesicomya gigas*	1	1	1
Polychaeta;*E.norvegica*	2	2	3	*Eunice norvegica*	1	1	1
Others	5	6	5	Others	3	2	2

~~Assigning ASVs with the RDP Bayesian Classifier allowed recovering 4 out of 10 mock species in the 18S dataset (Fig S 2) and no species in the COI dataset using the full MIDORI database. The six incorrectly resolved species in the 18S dataset could only be resolved taxonomically up to their common class level (venerid bivalves and malacostracan crustaceans). For the COI dataset, using the full MIDORI database resulted in RDP assignments that never matched the expected taxon and were mostly assigned to arthropods (data not shown). When the database was reduced to marine-only taxa, all 10 species were detected (Fig S 2), although the dataset contained a considerable amount of spurious assignments (29 clusters assigned up to Arthropoda and Chordata). The latter were however always associated to a phylum bootstrap level < 98. As the taxonomic resolution using RDP was poorer in the mock communities using 18S, the remaining work was performed using BLAST assignments.~~

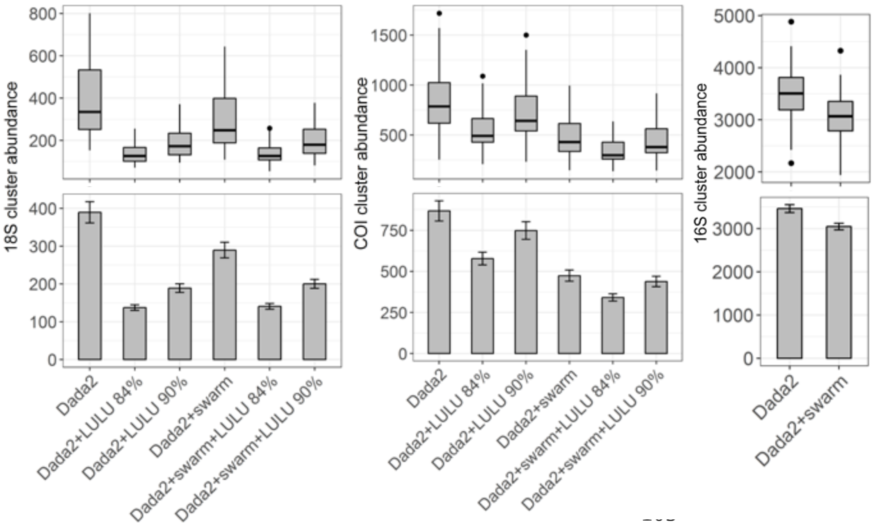
~~2.3 Alpha-diversity patterns between pipelines~~

~~*Eukaryotes*~~

Number of clusters among pipelines

The number of metazoan clusters detected in the deep-sea sediment samples varied significantly between bioinformatic pipelines chosen ~~(ANODEV: 18S, F(5,175)=599.91, p<0.001 and COI, F(5,195)=1,320.32, p<0.001, 16S, F(51,41)=2008.76, p<0.001, see Table S 8).~~ (, and also varied significantly among sites (Table 2). However, the pipeline effect was consistent across sites although mean cluster numbers detected per sample spanned a wide range in all loci (100-800 for 18S, 150-1,500 for COI datasets, and 1,500-5,000 for 16S, Fig. 1).

Expectedly, clustering ~~and LULU curation~~ significantly reduced the number of detected clusters per sample for all loci. ~~The reduction due~~Consistent to results observed in mock communities, clustering ~~was much more pronounced for metazoans, particularly~~at *d=1-13* resulted in comparable OTU numbers for COI, while significantly higher OTU numbers were obtained at $d=1$ than with $d>1$ for ~~16S data~~ribosomal loci (Fig. 1, Table 2). DADA2 detected on average ~~389 (SE=28) and~~ 863 (SE=61) metazoan ~~18S and~~ COI ASVs per sample~~ respectively, while~~ , and clustering ~~ASVs (at $d=4$ for 18S, $d=6$ for COI, and $d=1$ for 16S)~~ reduced ~~the~~this number ~~of metazoan OTUs detected~~ to ~~289 (SE=21) for 18S and 467 (SE=34) for COI.~~around 500, regardless the d-value. For ~~prokaryotes, the number of ASVs was on average 3,567 (SE=480) per sample~~ribosomal loci, clustering ~~decreased this mean~~at $d=3$-5 reduced OTU numbers of around 25-30% compared to ~~3,138 (SE=413) OTUs per sample.~~without clustering, while at $d=11$, cluster numbers were halved.



~~**Figure 1.** Number of clusters detected in sediment of 14 deep-sea sites with the Dada2 metabarcoding pipeline with or without LULU curation at 84% and 90% *minimum match* and swarm v2 clustering, using the 18S (left) and COI (centre) and 16S (right) marker genes. Cluster abundance was obtained after rarefaction to minimal sequencing depth. Boxplots represent medians with first and third quartiles. Barplots show means and standard errors.~~

Table 2. Effect of pipeline and site on the number of metazoan and prokaryote clusters. Results of the analysis of variance (ANOVA) of the rarefied cluster richness for the three genes studied. Pairwise comparisons were performed with Tukey's HSD tests. DS: Dada2+swarm; DSL: Dada2+swarm+LULU; d: swarm *d-value*. Significance codes: ***: p<0.001; **: p<0.01; *: p<0.05.

LOCUS	**F-value**	***p-value***	**Significant pairwise comparisons**
COI			
Pipeline	123.13	p<0.001	Dada2 > DS***; DS(d1) > DS(d13)***;
Site	356.37	p<0.001	Dada2 > DL***; DS > DSL 84%***; D(S)L 90% > D(S)L 84%***
Pipeline x Site	0.16	p>0.05	DL > DSL***; DL 90% > DS***
18S V1-V2			
Pipeline	129.16	p<0.001	Dada2 > DS(d>1)***; DS(d1) > DS(d>1)***; DS(d11) < DS(d1-5)***;
Site	154.52	p<0.001	Dada2 > DL***; DS > DSL 84%***; D(S)L 90% > D(S)L 84%***;
Pipeline x Site	0.49	p>0.05	DL 84% < DS***
16S V4-V5			
Pipeline	179.19	p<0.001	Dada2 > DS***;
Site	18.46	p<0.001	DS(d1) > DS(d>1)***; DS(d11) < DS(d1-5)***
Pipeline x Site	0.06	p>0.05	

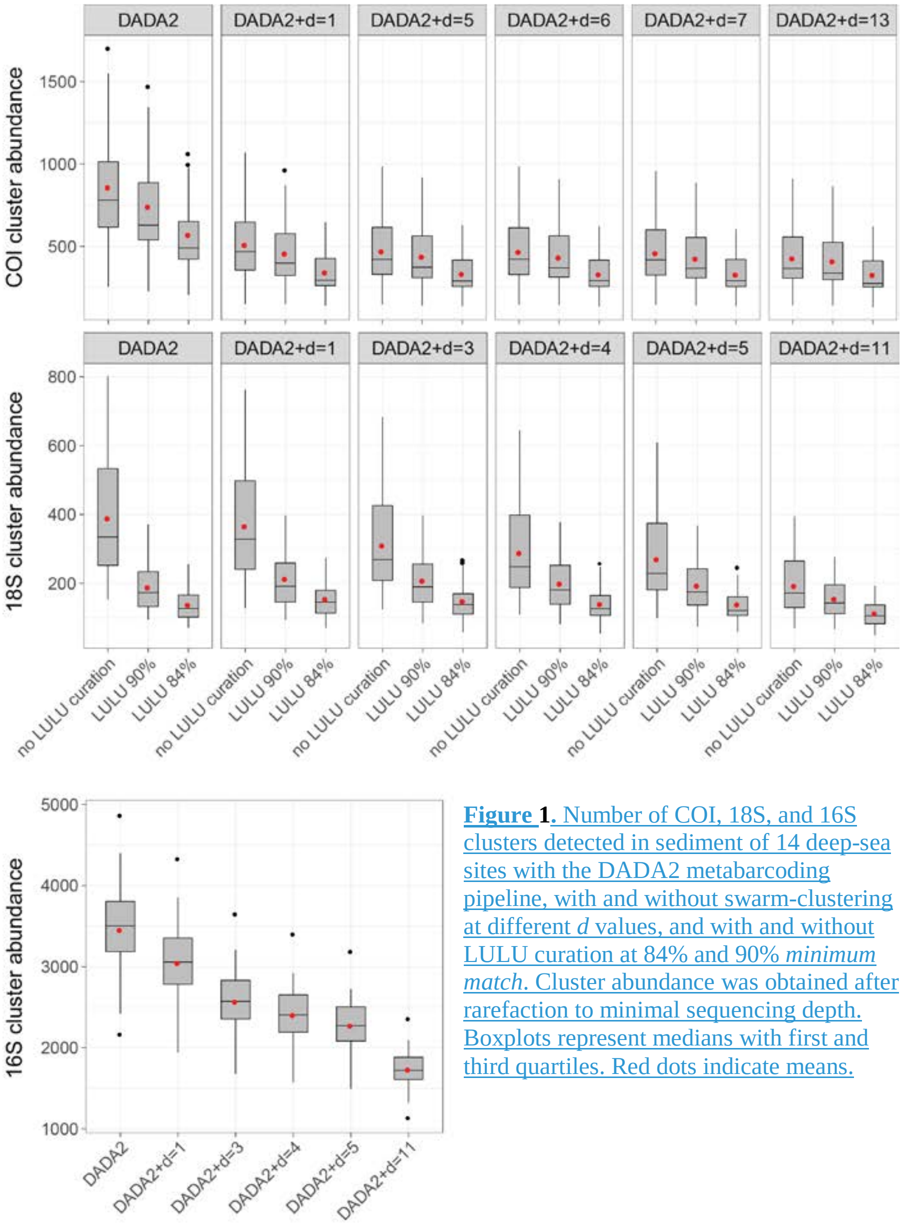


Figure 1. Number of COI, 18S, and 16S clusters detected in sediment of 14 deep-sea sites with the DADA2 metabarcoding pipeline, with and without swarm-clustering at different *d* values, and with and without LULU curation at 84% and 90% *minimum match*. Cluster abundance was obtained after rarefaction to minimal sequencing depth. Boxplots represent medians with first and third quartiles. Red dots indicate means.

LULU curation of metazoan ASVs significantly decreased the number of ~~metazoan~~ clusters detected ~~in~~at both ~~the ASV and~~tested *minimum match* values (Table 2). For OTU datasets~~.~~, the decrease was significant only when the *minimum match* parameter was at 84%. The effect of LULU curation was stronger at a lower *minimum match* ~~parameter. It~~ value for both loci, as LULU curation at 90% of ASVs or OTUs resulted in significantly more clusters than when the minimum match was at 84% (Table 2). The effect of LULU curation of was also more pronounced ~~in the ASV datasets and~~ for the 18S locus ~~(Fig. 1). At 90% minimum match,~~: LULU decreased by ~~51~~31-65% the number of 18S ~~and by 14% the number of COI ASVs, while this decrease was only of 31% for 18S OTUs and 7% for COI~~ ASVs/OTUs~~. When the *minimum match* parameter was at 84%, LULU decreased the number of detected metazoan clusters by 65% for 18S ASVs and~~ , compared to 7-33% for COI ~~ASVs, while in the clustered dataset this decrease was of 51% and 28% for 18S and COI OTUs respectively.~~. LULU curation of ASVs or OTUs ~~produced~~resulted in comparable ~~number of clusters~~cluster numbers in the 18S ~~dataset. At~~datasets, regardless the *d*-value used for clustering. For example, at 84% *minimum match*, LULU curation produced on average 137 ± 7 and 140 ± 8 clusters per sample after application on ASVs and OTUs (*d*=4) respectively. At 90%, these numbers were at 189 ± 11 and 200 ± 12 (Fig. 1). This was not the case for COI, where LULU curation of ASVs resulted in significantly more clusters (574 ± 38 at 84% and 742 ± 53 at 90%) than LULU curation of OTUs (334 ± 21 and 433 ± 31 for *d*=6).

~~The number of clusters detected also varied significantly among sites (ANODEV: 18S, F(11,175)=283.57, p<0.001 ; COI, F(13,195)=761.19, p<0.001; 16S, F(13,41)=507.37, p<0.01), and cores nested within sites (ANODEV: 18S, F(24,175)=32.21, p<0.001; COI, F(26,195)=72.91, p<0.001; 16S, F(28,41)=241.73, p<0.01). However, while the mean number of clusters detected per sample spanned a wide range in all loci (100-800 for 18S, 150-1,500 for COI datasets, and 2,000-5,000 for 16S), the pipeline effect was consistent across sites (Fig. S 3).~~

~~Taxonomic assignments and patterns~~ Looking at mean ASV and OTU numbers detected per phylum with each pipeline showed consistent effects of swarm clustering and LULU curation, but highlighted strong differences in the amount of intragenomic variation between taxonomic groups. For all loci investigated, some taxa displayed high ASV to OTU ratios, while others were hardly affected by clustering or LULU curation in terms of numbers of clusters detected (Fig S2).

~~2.4~~2.3 Patterns of beta-diversity between pipelines

~~Sequence identity varied strongly depending on phyla and marker gene (Fig. 2~~Community differences were visualized using PCoA ordinations (Jaccard and Bray-Curtis dissimilarities for metazoans and prokaryotes respectively) in clustered and non-clustered datasets (Fig. 2, Fig. S3). Expectedly, PERMANOVAs confirmed that sites differed significantly in terms of community structure, accounting from 45% to 89% of variation in data. Evaluating the effect of LULU curation (at 84% and 90%) for metazoans showed that LULU-curated data resolved similar ecological patterns than non-curated data, accounting from 0.5% (COI) to 1.3% (18S) of variation in data (Fig. 2).

Although ASV and OTU datasets detected similar levels of variation due to sites in PERMANOVAs, clustering levels affected the ecological patterns resolved by ordinations in rRNA loci (Fig 2). At low *d* values (*d=1-3*), ecological patterns were consistent to patterns observed in the ASV datasets, with samples segregating by site and depth. Increasing *d* values produced stronger segregation among sites, thus resulting in differentiation among ocean basins rather than depth. This change in resolution occurred with *d* values as low as *d=4* for 18S but was strongest at *d=11* for both rRNA loci (Fig. S3, Fig. 2).

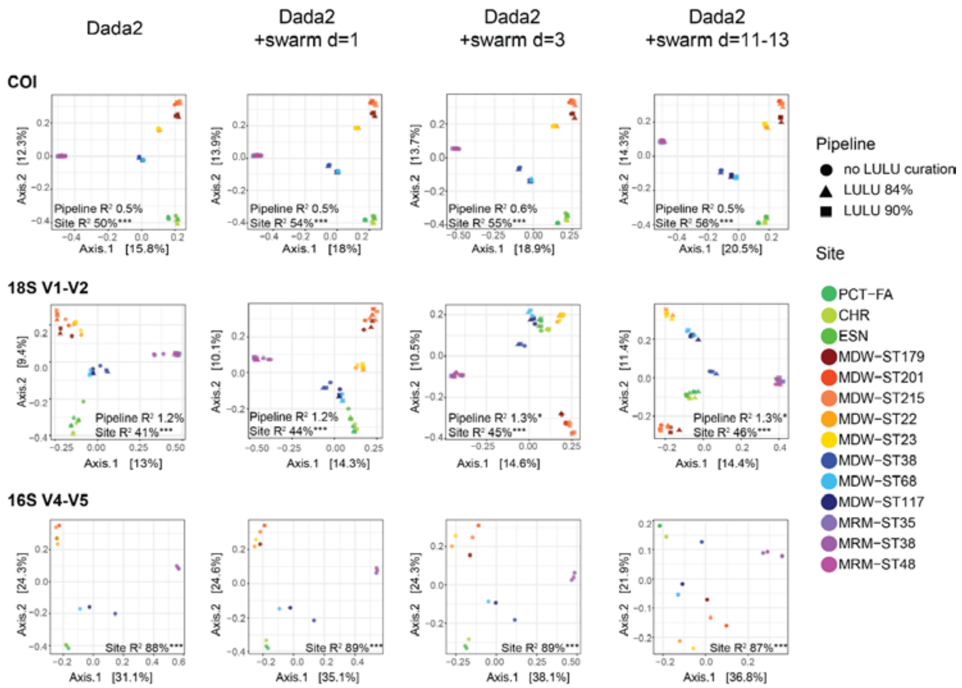


Figure 2. Beta-diversity patterns in ASV and OTU-centred datasets. PCoA ordinations showing community differentiation observed between sites and LULU *vs* not LULU curated samples, for the DADA2 metabarcoding pipeline with and without clustering. Metazoan datasets were clustered at *d=1-13* (COI) *d=1-11* (18S) and curated with LULU at two minimum match values. The prokaryote 16S dataset was clustered at *d=1-11*. R^2 values and associated p-values obtained in PERMANOVAs are shown in the ordination plots. Significance codes: ***: p<0.001; **: p<0.01; *: p<0.05. Colour codes: Green: Mediterranean < 1,000 m; Red-yellow: Mediterranean-Atlantic transition zone 300-1,000 m; Blue: North Atlantic < 1,000 m; Purple: Arctic < 1,000 m.

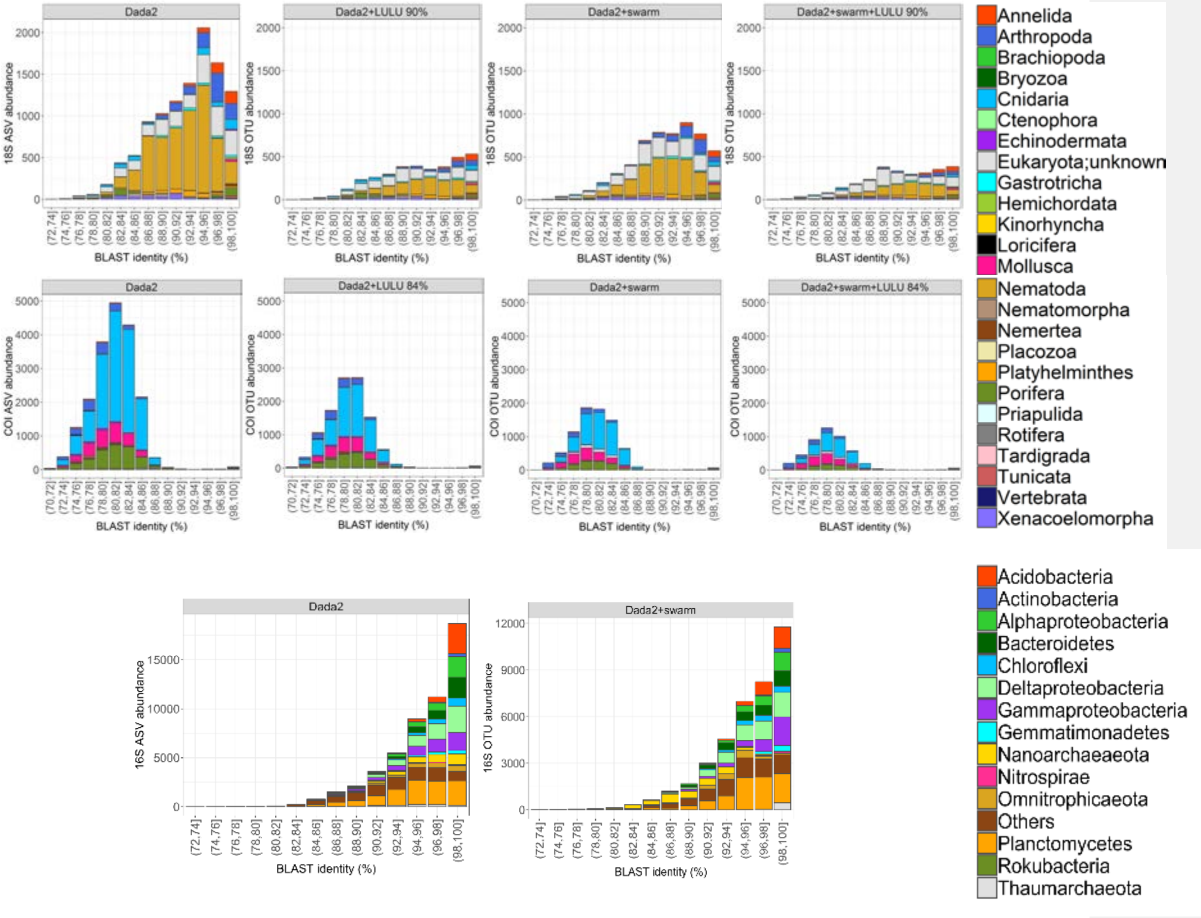
2.4 Taxonomic assignment quality

BLAST and RDP Bayesian Classifier assignments were compared in the mock communities and natural samples, on data clustered at *d=3* and curated with LULU at 84% for COI and 90% for 18S. For prokaryotes, assignment methods were compared on the ASV-level. BLAST and RDP assigned similar amounts of OTUs in the prokaryote dataset, but BLAST assigned 20-70% less OTUs in the metazoan datasets (Table S7). Assigning with BLAST at a minimum of 70% hit identity resulted in comparable results as described above. Eight of the ten species were recovered with COI and six species were recovered with 18S, while the vesicomyid bivalves were taxonomically unresolved with both loci (Fig. S4). Although most species produced one single OTU, between one and three species still resulted in 2-3 OTUs in each mock sample. Assigning the 18S dataset with RDP resulted in comparable taxonomic resolutions, although more species produced more than one OTU. Assigning the COI dataset with RDP using the MIDORI-UNIQUE database resulted in assignments of the mock samples that did not match the expected taxa and were mostly belonging to arthropods, a problem not observed with BLAST (data not shown). When the database was reduced to marine-only taxa, all 10 species were detected, and this at expected OTU abundances, once data was filtered for phylum bootstrap levels ≥ 80% (Fig S4). However, applying a phylum bootstrap minimum of 80% resulted in a strong decrease in the number of final target OTUs, particularly for COI where only 226 OTUs remained after filtering (Table S7). This reduced recovery with RDP after applying a minimum phylum bootstrap level was not observed in prokaryotes, where 51,000-55,000 ASVs were left after filtering with both assignment methods (Table S7).

BLAST hit identities of the overall datasets varied strongly depending on phyla and marker gene (Fig. 3). For 18S, most clusters had hit identities ≥ 90%. Poorly assigned clusters (hit identity < 90%) represented less than 20% of the dataset and were mostly assigned to

Mis en forme : Gauche

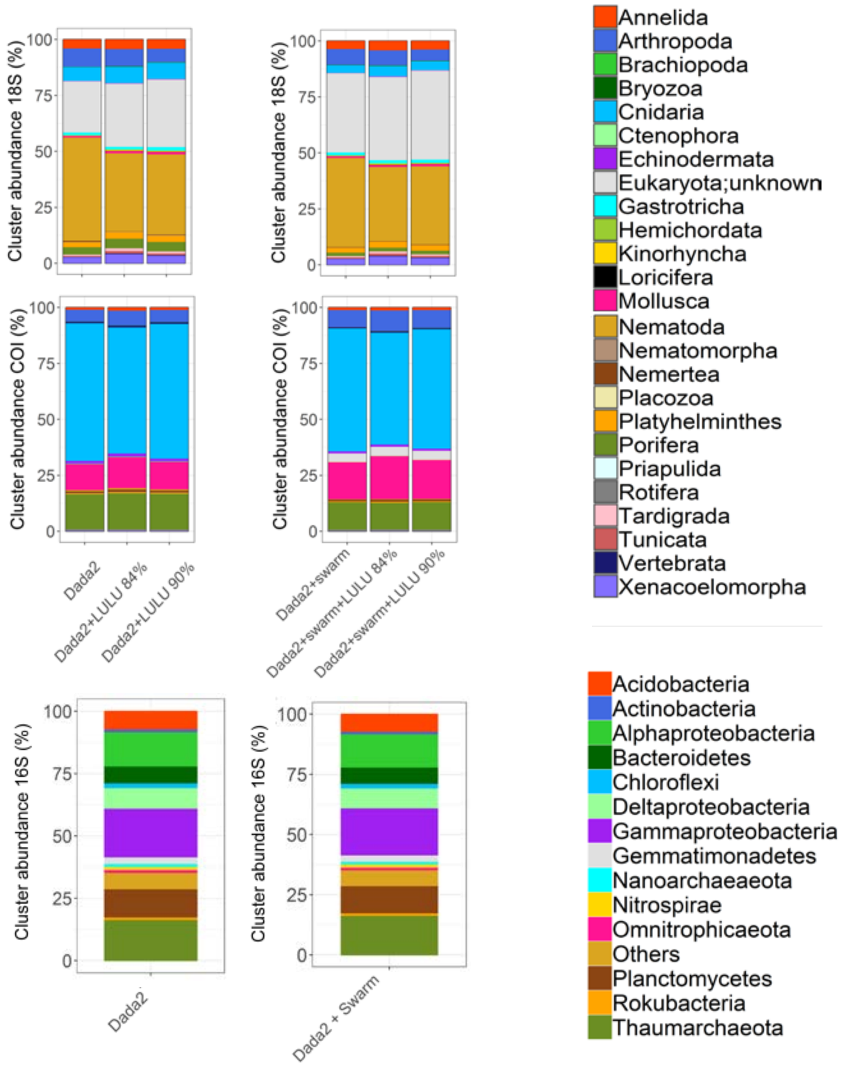
Nematoda, Cnidaria, Tardigrada, Porifera, and Xenacoelomorpha. For COI, nearly all clusters had similarities to sequences in databases lower than 90%. Overall, arthropods and echinoderms were detected at similar levels by both markers. The 18S barcode marker performed better in the detection of nematodes, annelids, platyhelminths, and xenacoelomorphs while COI mostly detected cnidarians, molluscs, and poriferans (Fig. ~~2~~3), highlighting the complementarity of these two loci. ~~Sequence~~BLAST hit identity was much higher for prokaryotes, with most clusters assigned ~~above~~with more than 90~~%.~~% similarity to sequences in databases. When datasets were filtered for RDP phylum bootstrap levels ≥ 80%, most assignments also had high genus bootstrap values for ribosomal loci. However, for COI, a considerable number of OTUs assigned to arthropods, cnidarians, molluscs, vertebrates, and poriferans still had genus bootstraps < 60%.



~~Figure 2. Taxonomic resolution in in metabarcoding datasets of 14 deep-sea sediment sites with four bioinformatic pipelines. Metazoan taxonomic assignment quality based on the 18S (top), COI (centre) and 16S (bottom) marker genes. BLAST hit identity of all metazoan clusters detected is given for four bioinformatic pipelines: DADA2, DADA2 curated with LULU at 84/90% *minimum match*, DADA2 clustered with swarm v2, and DADA2 clustered with swarm v2 and curated with LULU at 84/90% *minimum match*. BLAST hit identity for prokaryotes is given for two pipelines: DADA2 and DADA2 with swarm v2.~~

~~For metazoan loci, while clustering significantly decreased the number of OTUS detected, it increased the amount of clusters not assigned up to the phylum level in both loci (~10-20% increase, Fig. 2). In the 18S dataset, clustering led to the decrease in abundance of dominant taxa such as nematodes and non-dominant taxa such as cnidarians and poriferans (Fig. 2, Fig. 3). Similarly, for COI, clustering led to a decreased abundance of dominant taxa such as poriferans and cnidarians, while the number of clusters assigned to arthropods and molluscs increased (Fig. 2, Fig. 3). Changes were less marked for 16S data (Fig. 2), yet the number of some taxa clearly increased (i.e. Thaumarchaeota and Gammaproteonbacteria) whereas others decreased (i.e. Omnitrophicaeota).~~

~~For COI and 18S datasets, PERMANOVAs were performed to evaluate the effect of LULU curation at two *minimum match* thresholds. Multivariate analyses on clustered and non-clustered datasets showed significant differences in community structure between bioinformatic pipeline (i.e. with or without LULU), sites, and cores nested within sites (Table 3). LULU had a significant effect on taxonomic structure resolved, even though the percentage variation it explained was only around 1.3% for 18S and 0.5% for COI (R^2 values in Table 3), compared to 40-50% variation explained by sites, reflecting the predominant effect of biological signatures over bioinformatic processing in the resolution of community structure. Comparing the taxonomic composition resolved by all pipelines showed that LULU curation of ASVs or OTUS resulted in detected community compositions similar to non-curated datasets, although it increased the relative abundance of non-dominant taxa by decreasing the abundance of dominant phyla such as nematodes in 18S and cnidarians in COI (Fig. 3).~~



~~Figure 3. Patterns of relative cluster abundance resolved by different bioinformatic pipelines (ASV-centred on the left, OTU-centred on the right) in 14 deep-sea sites, using the 18S (top), COI (centre), and 16S (bottom) marker genes. LULU curation and clustering increase the abundance of non-dominant taxonomic groups in both metazoan loci, while this is not the case for prokaryotes.~~

~~Overall, community differences were visualized using PCoA ordinations of Jaccard distance matrices and showed that the different pipelines resolved the same ecological patterns, in which, consistently with the PERMANOVAs, the site effect was predominant (Fig. S 4).~~

~~**Table 3.** Effect of LULU curation on community structure detected in 14 deep-sea sites. Results of the permutational analysis of variance (PERMANOVA) of the rarefied OTU richness in clustered (Dada2+swarm+LULU) and non-clustered (Dada2+LULU) datasets, for the two genes studied. The tests were performed by permuting 9999 times using Jaccard distances. The pipeline and core effects were evaluated by permuting within sites.~~

	Dada2+swarm+LULU					**Dada2+LULU**					
LOCUS	***df***	**SS**	**Pseudo-F**	**P(>F)**	**R2**		***df***	**SS**	**Pseudo-F**	**P(>F)**	**R2**
18S-V1						**18S-V1**					
Pipeline	2	0.755	5.62	0.001	0.014	Pipeline	2	0.695	2.97	0.0001	0.012
Site	13	24.238	27.79	0.001	0.455	Site	13	23.658	15.57	0.0001	0.410
Site:Core	28	22.734	12.10	0.001	0.427	Site:Core	28	23.74	7.25	0.0001	0.412
Residuals	82	5.505			0.103	Residuals	82	9.584			0.166
Total	125	53.228			1.000	Total	125	57.677			1.000
COI						**COI**					
Pipeline	2	0.262	4.75	0.0001	0.005	Pipeline	2	0.244	2.68	0.0001	0.004
Site	13	29.555	82.47	0.0001	0.557	Site	13	27.525	46.61	0.0001	0.498
Site:Core	26	21.069	29.40	0.0001	0.397	Site:Core	26	24.984	20.31	0.0001	0.434
Residuals	78	2.15			0.041	Residuals	78	3.543			0.064
Total	119	53.036			1.000	Total	119	55.296			1.000

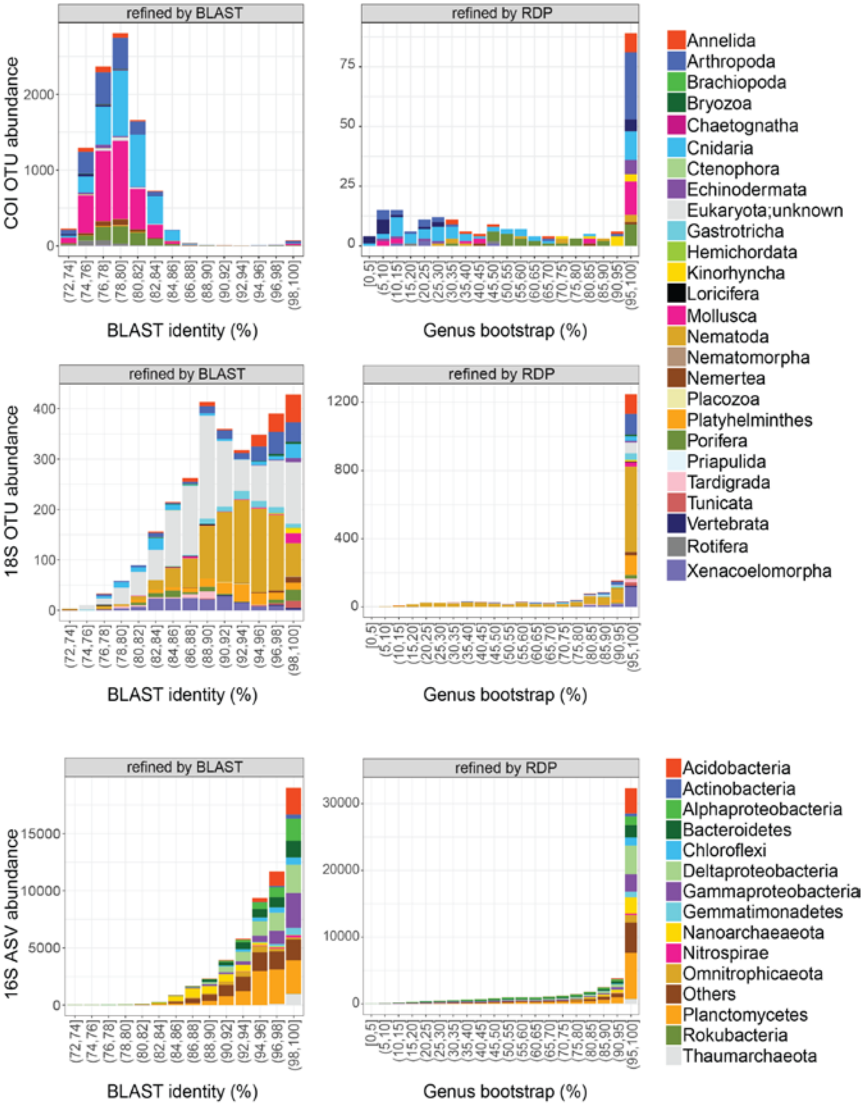


Figure 3. Taxonomic assignment quality of BLAST and RDP methods on metazoan and prokaryote metabarcoding datasets of 14 deep-sea sites. BLAST hit identity of all target clusters detected is given at hit identities > 70%. RDP-assigned data was filtered for phylum bootstraps ≥ 80%, and associated genus bootstraps are displayed. Taxonomic assignments were performed on the Silva132 database for 18S and 16S, and on the MIDORI-UNIQUE database, subsampled to marine taxa for COI.

3 DISCUSSION

3.1 ASVs ~~or~~and OTUs for ~~metazoans?~~genetic *vs* species diversity

The rise of HTS and the subsequent use of metabarcoding have revolutionized microbiology by unlocking the access to uncultivable microorganisms, which represent by far the great majority of prokaryotes (Klappenbach~~, Saxman, R., & Schmidt,~~ et al. 2001). The development and improvement of molecular and bioinformatic methods to perform inventories were historically primarily developed for 16S rRNA barcode loci, before being transferred to the eukaryotic kingdom based on the use of barcode markers such as 18S rRNA, ITS, or mitochondrial markers such as COI (~~Bellemain et al., 2010;~~ Valentini et al.~~,~~ 2009; Bellemain et al. 2010). Thus, most bioinformatics pipelines were initially developed accounting for intrinsic properties of prokaryotes and concepts inherent to microbiology (~~Boyer et al., 2016;~~ Caporaso et al.~~,~~ 2010; Schloss et al.~~,~~ 2009; Boyer et al. 2016), before being transferred to eukaryotes in general or metazoans in particular. Such application transfers ~~are not always straightforward, and~~ require adaptations to account~~ing~~ for differences in both concepts and basic biological features. One example is the question of the relevance of ~~the use of amplicon sequence variants (~~using ASVs~~)~~, advocated to replace OTUs "… *as the standard unit of marker-gene analysis and reporting*" (Callahan et al., 2017): an advice for microbiologists that may not apply to all cases, especially when working on metazoans.

First, metazoans are well known to exhibit variable and sometimes very high intraspecific polymorphism in 18S-V1 and above all in COI. Second, the results on the mock samples showed that single individuals prod~~d~~uced very different numbers of ASVs, indicating that ASV-centred datasets do not reflect actual species composition in metazoans. As this "demultiplication" will be highly variable across taxa (as seen in Fig. S2, and references such as Plouviez et al. 2009 and

Teixeira et al. 2013), the taxonomic compositions of samples based on ASVs will reflect genetic rather than species diversity.

Clustering ASVs into OTUs and/or curating with LULU alleviated the numerical inflation, but some species still produced more than one OTU, even at *d*-values as high as *d=11-13*. While clustering and LULU curation improved ~~COI~~numerical results in the mock communities~~ (where species always co-occurred),~~, they were associated with a decrease in taxonomic resolution, especially for 18S ~~data, as~~where some closely related species were merged~~, i.e.~~ with increasing clustering/filtering thresholds (i.e. the vesicomyid bivalves, the gastropod, and the shrimp species ~~(~~; Table ~~2~~1). When studying natural habitats, very likely to harbour closely related co-occurring species, both LULU curation and clustering are thus likely to lead to the loss of true species diversity, particularly for low-resolution markers such as 18S. Optimal results in the mock samples, i.e. ~~Optimal results in the mock samples, i.e.~~ delivering the best balance between the limitation of spurious clusters and the loss of true species diversity, were obtained with LULU curation at 90% for 18S and 84% for COI, highlighting the importance of adjusting bioinformatic correction tools to each barcode marker, a step for which mock communities are most adequate.

Mis en forme : Retrait : Première ligne : 0 cm

3.2 ASVs *vs* OTUs in natural communities: adapting pipeline parameters to marker properties

Life histories of organisms, together with intrinsic properties of marker genes, determine the level of intragenomic and intraspecific diversity. Intraspecific variation is a recognised problem in metabarcoding, known to generate spurious clusters (Brown et al. 2015), especially in the COI barcode marker. Indeed, this gene region has increased intragenomic variation due to its high evolutionary rate but also due to heteroplasmy and the abundance of pseudogenes, such as NUMTs, playing an important part of the supernumerary OTU richness in COI-metabarcoding (Bensasson et al. 2001; Song et al. 2008). Together with clustering, LULU curation at 84% proved effective in limiting the number of multiple clusters produced by single individuals, confirming its efficiency to correct for intragenomic diversity (Table 1).

~~3.2~~ The mock communities we used here did not contain several haplotypes of the same species (intraspecific variation), as is most often the case in environmental samples. ~~Application to real communities~~

~~The mock communities we used here did not contain several haplotypes of the same species (intraspecific variation), as is most often the case in environmental samples.~~ This prevents us from generalizing the comparable results ~~of LULU obtained with or without clustering to more complex communities. As distinct haplotypes do not always co-occur in nature,~~ obtained after LULU curation of ASVs ~~alone~~ and OTUs, and the apparently limited effect of clustering in the mock samples to communities that are more complex. However, LULU curation of ASVs is not suited to ~~correct~~account for natural haplotype diversity~~, and clustering ASVs may therefore~~: not all haplotypes co-occur and when they do so, they may vary in proportion and dominance relationships, making clustering more suited to account for natural haplotypic diversity. Thus, clustering ASVs will still be necessary to produce ~~datasets~~inventories of metazoan communities that reflect species rather than gene diversity.

As expected, ~~results~~evaluation of clustering and LULU curation on natural samples showed distinct ~~answers to this question for 18S and COI.~~results for 18S and COI. Indeed, concerted evolution, a common feature of SSU rRNA markers such as 16S (Hashimoto et al. 2003; Klappenbach et al. 2001) and 18S (Carranza et al. 1996), limits the amount of intragenomic polymorphism. In metazoans, a lower level of diversity is expected for the slower evolving 18S gene (Carranza et al. 1996), than for COI which exhibits faster evolutionary rates (Machida and Knowlton 2012; Machida et al. 2012). This is reflected in the lower ASV (DADA2) to OTU

(DADA2+swarm) ratios observed here for 18S (1.0-2.2.) compared to COI (2.0-2.7) data at clustering *d*-values comprised between one and seven (Table S6), underlining the different influence –and importance– of clustering on these loci, and the need for a versatile, marker by marker choice for clustering and curation parameters. When applying LULU to ASVs (DADA2) *versus* OTUs (DADA2+swarm) on 18S, similar ~~numbers of detected clusters were obtained (e.g. average of 137 ± 7 and 140 ± 8 clusters per sample after application at 84% on ASVs and OTUs respectively), again~~cluster numbers were obtained (Fig. 1), suggesting a limited added effect of clustering for this marker once DADA2 and LULU are applied~~ (Fig. 1).~~. This is in line with its slow evolutionary rate ~~(Carranza et al., 1996)~~(Carranza et al. 1996) leading to a limited number of haplotypes per species compared to COI. In contrast, ~~after~~for COI, LULU curation of the ~~COI~~ ASV dataset~~,~~ led to nearly twice the number of clusters ~~were obtained~~ (574 ± 38 at 84% and 742 ± 53 at 90%) compared to ~~the~~ LULU~~ curated OTU dataset (~~ curation of OTUs (at *d=6*: 334 ± 21 for 84% and 433 ± 31 for 90%). This confirms the ~~need for clustering on COI and the fact that LULU curation of ASVs is not sufficient to account for~~ higher intraspecific diversity ~~in natural samples for such a~~of COI, and the need to combine clustering with LULU curation to account for intraspecific diversity in natural samples, especially with highly polymorphic ~~marker.~~markers such as COI.

Finally, the reproductive mode and pace of selection in microbial populations may lead to locally lower levels of intraspecific variation than the one expected for metazoans. Prokaryotic alpha diversity was however also affected by the clustering of ASVs (Fig. 1), supporting the estimation of a 2.5-fold greater number of 16S rRNA variants than the actual number of bacterial “species” (Acinas et al. 2004). The significant decrease in the number of OTUs after clustering at *d=1* (Table 2, Fig. 1, decrease of ~25%) suggests the occurrence of very closely related 16S rRNA sequences, possibly belonging to the same ecotype/species. Such entities may still be important to delineate in studies aiming for example at identifying species associations (i.e. symbiotic relationships) across large distances and ecosystems, where drift or selection can lead to slightly different ASVs in space and time, with their function and association remaining stable.

Mis en forme : Retrait : Première ligne : 0 cm

3.3 Influence on beta diversity

~~prokaryotic alpha diversity was less affected by the clustering of ASVs (Table 1, Fig. 1), illustrating their lower intra-genomic variability (Pei et al., 2010) and the possibly lower diversity within ecotypes. Nevertheless, the differences suggest the occurrence of very closely related sequences of 16S rRNA, possibly belonging to the same ecotype/species. After focusing on alpha diversity estimates and the accuracy of inventories, the analysis of taxonomic structure showed that the non-clustered, clustered, and LULU-curated datasets resolved similar ecological patterns (Fig. S-4) and community compositions (Fig. 3), although differences in abundance were observed (Fig. 2). This is in accordance with other studies reporting severe impacts of bioinformatic parameters on alpha diversity while comparable patterns of beta diversity were observed, at least down to a minimum level of clustering stringency (Bokulich et al., 2013; Xiong & Zhan, 2018).~~

~~Clustering and LULU curation mainly led to the decrease of the number of clusters assigned to dominant taxa in both loci, i.e. nematodes for 18S, cnidarians and to lesser extent molluscs for COI. This is likely attributable to the low resolutive power of 18S, already acknowledged in general and for nematodes in particular (Derycke, Vanaverbeke, Rigaux, Backeljau, & Moens, 2010). Similarly the lack of resolution of COI for cnidarians has long been known (Hebert, Ratnasingham, & de Waard, 2003). Clustering also introduced more OTUs that could not be assigned at the phylum level with BLAST (Fig. 3), confirming the limitations of assigning taxonomy at the OTU level, as the representative sequence chosen for taxonomic assignment can lead to taxonomic ambiguity.~~

~~3.4 Assignment comparison~~

~~Finally, compared to BLAST assignment, lower taxonomic resolution was observed using the RDP Bayesian Classifier on the mock samples for 18S (Fig. S 2) and for COI when using the full MIDORI database. With this database, only five phyla were detected in the whole dataset: Arthropoda, Chordata, Mollusca, Nemertea, Porifera (data not shown). This is likely due to the size of the RDP training sets available for this study, and to the low coverage of deep-sea species in public databases. Small databases, taxonomically similar to the targeted communities, and with sequences of the same length as the amplified fragment of interest, are known to maximise accurate identification (Macheriotou et al., 2019). This limitation of databases, rather than the method itself, was confirmed by results using a reduced marine-only COI database. The latter (containing the barcodes of the mock species) resulted in accurate RDP assignments in the mock samples when the phylum bootstrap level was ≥ 98 (Fig. S 2), although the majority of clusters remained unassigned in the full dataset (89% compared to 45% with BLAST). The development of custom-built marine RDP training sets, without overrepresentation of terrestrial species, is therefore needed for this Bayesian assignment method to be effective on deep-sea datasets. With reduced trainings sets, removing clusters with phylum bootstrap level < 98 could be an efficient way to increase taxonomic quality of deep-sea metabarcoding datasets. At present, BLAST seems however the most efficient assignment method for deep-sea metabarcoding data, even though it has to be kept in mind that hit identities tend to be low, especially for COI, making it hard to work at taxonomic levels beyond phylum or class (Fig. 2).~~

After focusing on alpha diversity estimates, i.e. on the numerical accuracy of inventories, the analysis of community structures showed that the ~~non-clustered, clustered, and~~ LULU-curated datasets resolved similar ecological patterns as datasets not curated with LULU. However, clustering affected resolution of ecological patterns in ribosomal loci when *d* values were high, and this was not the case for COI, where similar patterns were resolved in all datasets (Fig. 2). This is in accordance with other studies reporting severe impacts of bioinformatic parameters on alpha diversity while comparable patterns of beta diversity ~~were~~ are observed in ASV and OTU datasets, at least down to a minimum level of clustering stringency (Xiong and Zhan 2018; Bokulich et al. 2013).

Clustering and LULU curation mainly led to the decrease of the number of clusters assigned to ~~dominant~~particular taxa in both loci, such as annelids, arthropods, nematodes, or platyhelminthes for 18S, and chordates, cnidarians, ~~and to lesser extent molluscs~~ echinoderms, or poriferans for COI (Fig. S2). The strong decrease in cluster numbers observed in these phyla suggests that the latter have greater intraspecific polymorphism, although the decrease could also be due to the merging of closely related species, as both markers have lower taxonomic resolution in particular taxa. This ~~is likely attributable to the low resolutive power of 18S, already~~ has been acknowledged for 18S in general, but in nematodes in particular (Derycke et al. 2010) ~~). Similarly the lack of resolution of COI for~~), and reported in cnidarians with COI (Hebert et al. 2003).

Overall, based on alpha and beta diversity results observed in mock communities and natural samples, applying LULU at 84% seems to efficiently curate metazoan COI datasets without significant loss of species, but clustering is required, at least at *d=1*, in order to address high intraspecific polymorphism. For 18S, LULU curation seems to require values above 84% (e.g. 90%) in order to avoid the loss of species, as seen in the mock communities. However, the low taxonomic resolution obtained with this marker suggests that clustering should be performed at low *d-values (d<4)* to address intraspecific polymorphism without affecting beta-diversity patterns. For prokaryotes, clustering 16S ASVs at *d=1* reduces the number of detected clusters by ~25%, which may help addressing intragenomic variation when needed.

3.4 Taxonomic resolution and assignment quality

The COI locus allowed the detection of all ten species present in the mock samples, compared to seven in the 18S dataset (Table 1). This locus also provided much more accurate assignments, most of them correct at the genus (and species) level, confirming that COI uncovers more metazoan species and offers a better taxonomic resolution than 18S (Tang et al. 2012; Clarke et al. 2017; Andújar et al. 2018). Our results also support approaches combining nuclear and mitochondrial markers to achieve more comprehensive biodiversity inventories (Cowart et al. 2015; Drummond et al. 2015; Zhan et al. 2014). Indeed, strong differences exist in amplification success among taxa (Bhadury et al. 2006; Carugati et al. 2015), exemplified by nematodes, which are well detected with 18S but not with COI (Bucklin et al. 2011). The high complementarity of COI and 18S in terms of targeted taxa (highlighted in Fig. S2), also supports the approach taken by Stefanni et al. (2018), as subsampling each gene dataset for its "best targeted phyla" and subsequently combining both seems to be a very efficient way to produce comprehensive and non-redundant biodiversity inventories.

Finally, compared to BLAST assignments, similar taxonomic resolution was observed using the RDP Bayesian Classifier on the mock samples for 18S (Fig. S4) ~~and for COI when using the full MIDORI database. With this database, only five phyla were detected in the whole dataset: Arthropoda, Chordata, Mollusca, Nemertea, Porifera (data not shown). This is likely due to the size of the RDP training sets available for this study, and to the low coverage of deep-sea species in public databases.~~and for COI when using the MIDORI-UNIQUE marine-only database. Poor performance of RDP using the full MIDORI database is likely due to the size of the database, and to its low coverage of deep-sea species. The problem of underrepresentation of deep-sea taxa is especially apparent with the BLAST assignments, which generally displayed low identities to sequences in databases, especially for COI (Fig. 3). Using minimum similarities of 80% for COI and 86% for 18S as cut-off values for metazoans has been shown to improve the taxonomic quality of metazoan metabarcoding datasets (Stefanni et al. 2018). However, phylogenies of marine invertebrates have found high levels of species divergence (up to ~30%), even within genera (Zanol et al. 2010). Consequently, studies on deep-sea taxa have found that some invertebrate species had COI sequences diverging more than 20% from any other species present in molecular databases (Shank et al. 1999; Herrera et al. 2015). At present, it thus seems difficult to work at taxonomic levels beyond phylum-level with deep-sea metabarcoding data when using large public databases. Small databases, taxonomically similar to the targeted communities, and with sequences of the same length as the amplified fragment of interest, are known to maximise accurate identification (Macheriotou et al. 2019). ~~This limitation of databases, rather than the method itself, was confirmed by results~~When using the reduced marine-only COI database, RDP provided the most accurate assignments in the mock samples when the phylum bootstrap level was ≥ 80 (Fig. S 4), although this filtering threshold drastically reduced the number of OTUs in the overall dataset (Table S7). The development of custom-built marine RDP training sets, without overrepresentation of terrestrial species, is therefore needed for this Bayesian assignment method to be effective on deep-sea datasets. With reduced ~~trainings sets~~ and more specific databases, removing clusters with phylum bootstrap-level < 80 should be an efficient way to increase taxonomic quality of deep-sea metabarcoding datasets. At present, if accurate taxonomic assignments are sought while using universal primers, we advocate assigning taxonomy in two steps: first, using BLAST and a large database including all phyla amplifiable by the primer set, extracting the clusters belonging to the groups of interest, then re-assigning taxonomy to these target taxa using RDP and a smaller, taxon-specific database.

CONCLUSIONS AND PERSPECTIVES

~~In this work based on~~Using mock communities and natural samples, we ~~propose a new pipeline using~~ evaluate several recent algorithms ~~allowing~~ and assess their capacity to improve the quality of molecular biodiversity inventories ~~based on metabarcoding data~~of metazoans and prokaryotes. Our r~~R~~esults ~~showed~~ support the fact that ASV data should be produced and communicated for reusability and reproducibility following the recommendations of Callahan et al. (2017). This is especially useful in large projects spanning wide geographic zones and time scales, as different ASV datasets can be easily merged *a posteriori*, and clustered if necessary afterwards. Nevertheless, clustering ASVs into OTUs will be required to obtain accurate species-level inventories, at least for metazoan communities, with a more severe influence of clustering observed on alpha diversity estimates than beta-diversity patterns. Considering 16S polymorphism observed in prokaryotic species (Acinas et al., 2004) and the possible geographic segregation of their populations, clustering may also be required in prokaryotic datasets, for example in studies screening for species associations (i.e. symbiotic~~ or parasitic relationships~~, ~~considering that~~ as symbionts may be prone to differential fixation through enhanced drift; Shapiro, Leducq, & Mallet, 2016).

~~Results~~Our results also demonstrated that LULU ~~curation is a good alternative to arbitrary relative abundance filters in~~effectively curates metazoan biodiversity inventories obtained through metabarcoding~~ pipelines~~. They also underline the need to adapt parameters for curation (e.g. LULU curation at 90% for 18S and 84% for COI) and clustering to each gene used and taxonomic compartment targeted, in order to identify an optimal balance between the correction for spurious clusters and the merging of closely related species.

Finally, ~~the results~~our findings also ~~show~~showed that accurate taxonomic assignments of deep-sea species can be obtained with the RDP Bayesian Classifier, but only with reduced databases containing ecosystem-specific sequences.

The pipeline is publicly available on Gitlab (https://gitlab.ifremer.fr/abyss-project/), and allows the use of sequence data obtained from libraries produced by double PCR or adaptor ligation methods, as well as having built-in options for using six commonly used metabarcoding primers.

ACKNOWLEDGEMENTS

This work is part of the "*Pourquoi Pas les Abysses?*" project, funded by Ifremer and the ~~project~~ eDNAbyss (AP2016 -228) project, funded by France Génomique (ANR-10-INBS-09) and Genoscope-CEA. We wish to thank Caroline Dussart for her contribution to the early developments of bioinformatic scripts and Patrick Durand for bioinformatic support, Cathy Liautard-Haag for her help with lab management, and Babett Guenther for help in barcoding. We also wish to express our gratitude to the participants and mission chiefs of the Essnaut (Marie-Anne Cambon Bonavita), MarMine (Eva Ramirez Llodra, project 247626/O30 (NRC-BI), PEACETIME (Cécile Guieu and Christian Tamburini), CanHROV (Marie Claire Fabri) and MEDWAVES (Covadonga Orejas) cruises. The MEDWAVES cruise was organised in the framework of the ATLAS Project, supported by the European Union 2020 Research and Innovation Programme under grant agreement number: 678760. We thank Stefaniya Kamenova, Tiago Pereira, and an anonymous referee for their comments on a previous version of this manuscript.

REFERENCES

Acinas, ~~S.~~Silvia G~~.,~~, Luisa A Marcelino, ~~L. A.,~~Vanja Klepac-Ceraj, ~~V., &~~and Martin F Polz~~, M. F. (~~. 2004~~).~~. ‘Divergence and Redundancy of 16S ~~rRNA~~RRNA Sequences in Genomes with Multiple ~~rrn Operons~~Rrn Operons’. *Journal of Bacteriology*~~,~~ 186 (9~~),~~): 2629–~~2635~~35. https://doi.org/10.1128/JB.186.9.2629-2635.2004.

Alberdi, ~~A.,~~Antton, Ostaizka Aizpurua, ~~O.,~~M. Thomas P. Gilbert, ~~M. T. P., &~~and Kristine Bohmann~~, K. (~~. 2017~~).~~. ‘Scrutinizing ~~key steps~~Key Steps for ~~reliable metabarcoding~~Reliable Metabarcoding of ~~environmental samples. (A.~~Environmental Samples’. Edited by Andrew Mahon~~, Ed.),~~. *Methods in Ecology and Evolution*, 2017. https://doi.org/10.1111/2041-210X.12849.

Andújar, Carmelo, Paula Arribas, Douglas W. Yu, Alfried P. Vogler, and Brent C. Emerson. 2018. ‘Why the COI Barcode Should Be the Community DNA Metabarcode for the Metazoa’. *Molecular Ecology* 27 (20): 3968–75. https://doi.org/10.1111/mec.14844.

Baselga, ~~A., &~~ Andrés, and C. David L. Orme~~, C. D. L. (~~. 2012~~). betapart : an~~. ‘Betapart : An R ~~package~~Package for the ~~study~~Study of ~~beta diversity.~~Beta Diversity’. *Methods in Ecology and Evolution*~~,~~ 3 (5~~),~~): 808–~~812~~12. https://doi.org/10.1111/j.2041-210X.2012.00224.x.

Bazin, ~~E.,~~Eric, Sylvain Glémin, ~~S., &~~and Nicolas Galtier~~, N. (~~. 2006~~).~~. ‘Population ~~size does not influence mitochondrial genetic diversity~~Size Does Not Influence Mitochondrial Genetic Diversity in ~~animals~~Animals’. *Science*~~,~~ 312 (5773~~),~~): 570–~~572~~72. https://doi.org/10.1126/science.1122033.

Bellemain, ~~E.,~~Eva, Tor Carlsen, ~~T.,~~Christian Brochmann, ~~C.,~~Eric Coissac, ~~E.,~~Pierre Taberlet, ~~P., &~~and Håvard Kauserud~~, H. (~~. 2010~~).~~. ‘ITS as an ~~environmental~~Environmental DNA ~~barcode~~Barcode for ~~fungi~~Fungi: An in ~~silico approach reveals potential~~Silico Approach

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Reveals Potential PCR ~~biases~~Biases’. *BMC Microbiology*~~,~~ 10~~,~~ (July): 189. https://doi.org/10.1186/1471-2180-10-189.

Mis en forme : Police :Non Italique

Bensasson, ~~D.,~~Douda, De Xing Zhang, ~~D. X.,~~Daniel L. Hartl, ~~D. L., &~~ and Godfrey M. Hewitt~~, G. M. (~~. 2001~~, June 1).~~ . ‘Mitochondrial ~~pseudogenes~~Pseudogenes: Evolution’s ~~misplaced witnesses.~~Misplaced Witnesses’. *Trends in Ecology and Evolution*. https://doi.org/10.1016/S0169-5347(01)02151-6.

Bhadury, P~~.,~~, M C Austen, ~~M. C.,~~D T Bilton, P J D~~. T.,~~ Lambshead, ~~P. J.~~A D~~.,~~ Rogers, ~~A. D., &~~and G R Smerdon~~, G. R. (~~. 2006~~).~~ . ‘Molecular ~~detection~~Detection of ~~marine nematodes~~Marine Nematodes from ~~environmental samples: overcoming eukaryotic interference.~~Environmental Samples: Overcoming Eukaryotic Interference’. *Aquatic Microbial Ecology*~~,~~ 44 (1~~),~~): 97–103. https://doi.org/Doi 10.3354/Ame044097.

Mis en forme : Police :Non Italique

Bik, ~~H.~~Holly M., Way Sung, ~~W.,~~Paul De Ley, ~~P.,~~James G Baldwin, ~~J. G.,~~Jyotsna Sharma, ~~J.,~~Axayácatl Rocha-Olivares, ~~A., & Thomas,~~ and W~~. K. (~~ Kelley Thomas. 2012~~).~~ . ‘Metagenetic ~~community analysis~~Community Analysis of ~~microbial eukaryotes illuminates biogeographic patterns~~Microbial Eukaryotes Illuminates Biogeographic Patterns in ~~deep-sea~~Deep-Sea and ~~shallow water sediments.~~Shallow Water Sediments.’ *Molecular Ecology*~~,~~ 21 (5~~),~~): 1048–~~1059~~59. https://doi.org/10.1111/j.1365-294X.2011.05297.x.

Mis en forme : Police :Non Italique

Bista, ~~I.,~~Iliana, G Carvalho, ~~G.,~~K Walsh, ~~K.,~~M Christmas, ~~M.,~~Mehrdad Hajibabaei, ~~M.,~~P Kille, ~~P., …~~D Lallias, and Simon Creer~~, S. (~~. 2015~~).~~ . ‘Monitoring ~~lake ecosystem health using metabarcoding~~Lake Ecosystem Health Using Metabarcoding of ~~environmental~~Environmental DNA: ~~temporal persistence~~Temporal Persistence and ~~ecological relevance~~Ecological Relevance’. *Genome*~~,~~ 58 (5~~),~~): 197.

Mis en forme : Police :Non Italique

Bokulich, ~~N.~~Nicholas A~~.,~~, Sathish Subramanian, ~~S.,~~Jeremiah J Faith, ~~J. J.,~~Dirk Gevers, ~~D.,~~Jeffrey I Gordon, ~~J. I.,~~Rob Knight, ~~R., …~~David A Mills, and J Gregory Caporaso~~, J. G. (~~. 2013~~).~~ .

‘Quality-~~filtering vastly improves diversity estimates~~Filtering Vastly Improves Diversity Estimates from Illumina ~~amplicon sequencing.~~Amplicon Sequencing’. *Nature Methods*~~,~~ 10 (1~~),~~): 57–59. https://doi.org/10.1038/nmeth.2276.

Boussarie, ~~G.,~~Germain, Judith Bakker, ~~J.,~~Owen S. Wangensteen, ~~O. S.,~~Stefano Mariani, ~~S.,~~Lucas Bonnin, ~~L.,~~Jean Baptiste Juhel, Jeremy J~~. B., … Mouillot, D. (~~. Kiszka, et al. 2018~~).~~. ‘Environmental DNA ~~illuminates~~Illuminates the ~~dark diversity~~Dark Diversity of ~~sharks~~Sharks’. *Science Advances*~~,~~ 4 (5~~),~~): eaap9661. https://doi.org/10.1126/sciadv.aap9661.

Boyer, F~~.,~~, C Mercier, ~~C.,~~A Bonin, ~~A.,~~Y Le Bras, ~~Y.,~~Pierre Taberlet, ~~P., &~~and Eric Coissac~~, E. (~~. 2016~~).~~. ‘OBITOOLS: ~~a~~A UNIX-~~inspired software package~~Inspired Software Package for DNA ~~metabarcoding~~Metabarcoding’. *Molecular Ecology Resources*~~,~~ 16 (1~~),~~): 176–~~182~~82. https://doi.org/10.1111/1755-0998.12428.

Brannock, P~~.~~ M~~., &~~, and K M Halanych~~, K. M. (~~. 2015~~).~~. ‘Meiofaunal ~~community analysis by high-throughput sequencing~~Community Analysis by High-Throughput Sequencing: Comparison of ~~extraction, quality filtering, and clustering methods.~~Extraction, Quality Filtering, and Clustering Methods’. *Marine Genomics*~~,~~ 23~~:~~: 67–75. https://doi.org/10.1016/j.margen.2015.05.007.

Brown, E~~.~~ A~~.,~~, F J J Chain, ~~F.~~T J~~. J.~~, Crease, ~~T.~~H J~~.,~~ MacIsaac, ~~H. J., &~~and M E Cristescu~~, M. E. (~~. 2015~~).~~. ‘Divergence ~~thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities?~~Thresholds and Divergent Biodiversity Estimates: Can Metabarcoding Reliably Describe Zooplankton Communities?’ *Ecology and Evolution*~~,~~ 5 (11~~),~~): 2234–~~2251~~51. https://doi.org/10.1002/ece3.1485.

Bucklin, ~~A.,~~Ann, Dirk Steinke, ~~D., &~~and Leocadio Blanco-Bercial~~, L. (~~. 2011~~).~~. ‘DNA Barcoding of Marine ~~Metazoa~~Metazoa’. *Annual Review of Marine Science*~~,~~ 3 (1~~),~~): 471–508. https://doi.org/10.1146/annurev-marine-120308-080950.

Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...

Callahan, ~~B.~~Benjamin J., Paul J. McMurdie, ~~P. J., &~~Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes~~, S. P. (2017).~~ . 2016. ‘DADA2: High-Resolution Sample Inference from Illumina Amplicon Data’. *Nature Methods* 13 (7): 581–83. https://doi.org/10.1038/nmeth.3869.

Callahan, Benjamin J., Paul J McMurdie, and Susan P Holmes. 2017. ‘Exact ~~sequence variants should replace operational taxonomic units~~Sequence Variants Should Replace Operational Taxonomic Units in ~~marker-gene data analysis.~~Marker-Gene Data Analysis’. *ISME Journal*~~,~~ 11 (12~~),~~): 2639–~~2643~~43. https://doi.org/10.1038/ismej.2017.119.

~~Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data.~~ ~~*Nature Methods, 13*~~~~(7), 581–583. https://doi.org/10.1038/nmeth.3869~~

Caporaso, J~~. G.,~~. Gregory, Justin Kuczynski, ~~J.,~~Jesse Stombaugh, ~~J.,~~Kyle Bittinger, ~~K.,~~Frederic D Bushman, ~~F. D.,~~Elizabeth K Costello, ~~E. K., … Knight, R. (~~Noah Fierer, et al. 2010~~).~~ . ‘QIIME ~~allows analysis~~Allows Analysis of ~~high-throughput community sequencing data.~~High-Throughput Community Sequencing Data’. *Nature Methods*~~,~~ 7 (5~~),~~): 335–~~336~~36. https://doi.org/10.1038/nmeth.f.303.

Carranza, ~~S.,~~Salvador, Gonzalo Giribet, ~~G.,~~Carles Ribera, ~~C.,~~Jaume Baguñà, ~~J., &~~and Marta Riutort~~, M. (~~. 1996~~).~~ . ‘Evidence ~~that two types~~That Two Types of 18S ~~rDNA coexist~~RDNA Coexist in the ~~genome~~Genome of Dugesia (Schmidtea) ~~mediterranea~~Mediterranea (Platyhelminthes, Turbellaria, Tricladida~~).~~)’. *Molecular Biology and Evolution*~~,~~ 13 (6~~),~~): 824–~~832~~32. https://doi.org/10.1093/oxfordjournals.molbev.a025643.

Carugati, ~~L.,~~Laura, Cinzia Corinaldesi, ~~C.,~~Antonio Dell’Anno, ~~A., &~~and Roberto Danovaro~~, R. (~~. 2015~~).~~ . ‘Metagenetic ~~tools~~Tools for the ~~census~~Census of ~~marine meiofaunal biodiversity~~Marine Meiofaunal Biodiversity: An ~~overview~~Overview’. *Marine Genomics*~~,~~ 24~~,~~

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique, Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Mis en forme : Espagnol (Espagne)

Mis en forme : Police :Non Italique

Mis en forme : Espagnol (Espagne)

Mis en forme : Police :Non Italique

(December): 11–20. https://doi.org/10.1016/j.margen.2015.04.010.

Clare, ~~E.~~Elizabeth L., Frédéric J.J. Chain, ~~F. J. J.,~~Joanne E. Littlefair, ~~J.~~and Melania E~~., &~~. Cristescu~~, M. E. (~~. 2016~~).~~. 'The ~~effects~~Effects of ~~parameter choice~~Parameter Choice on ~~defining molecular operational taxonomic units and resulting ecological analyses~~Defining Molecular Operational Taxonomic Units and Resulting Ecological Analyses of ~~metabarcoding data.~~Metabarcoding Data'. Edited by Kristy Deiner. *Genome*~~,~~ 59 (11~~),~~): 981–~~990~~90. https://doi.org/10.1139/gen-2015-0184.

Clarke, ~~L.~~Laurence J., Jason M. Beard, ~~J.~~Kerrie M~~.,~~. Swadling, ~~K. M., &~~and Bruce E. Deagle~~, B. E. (~~. 2017~~).~~. 'Effect of ~~marker choice and thermal cycling protocol~~Marker Choice and Thermal Cycling Protocol on ~~zooplankton~~Zooplankton DNA ~~metabarcoding studies~~Metabarcoding Studies'. *Ecology and Evolution*~~,~~ 7 (3~~),~~): 873–~~883~~83. https://doi.org/10.1002/ece3.2667.

Cohan, ~~F.~~Frederick M. ~~(~~2001~~).~~. 'Bacterial ~~species~~Species and ~~speciation~~Speciation'. Edited by M. Kane. *Systematic Biology*~~,~~ 50 (4~~),~~): 513–~~524~~24. https://doi.org/10.1080/10635150118398.

Coissac, ~~E.,~~Eric, Tiayyba Riaz, ~~T., &~~and Nicolas Puillandre~~, N. (~~. 2012~~).~~. 'Bioinformatic ~~challenges~~Challenges for DNA ~~metabarcoding~~Metabarcoding of ~~plants~~Plants and ~~animals~~Animals'. *Molecular Ecology*~~,~~ 21 (8~~),~~): 1834–~~1847~~47. https://doi.org/10.1111/j.1365-294X.2012.05550.x.

Cowart, ~~D.~~Dominique A~~.,~~., Miguel Pinheiro, ~~M.,~~Olivier Mouchel, ~~O.,~~Marion Maguer, ~~M.,~~Jacques Grall, ~~J.,~~Jacques Miné, ~~J., &~~and Sophie Arnaud~~-~~-Haond~~, S. (~~. 2015~~).~~. 'Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and Molecular Surveys of Seagrass ~~Communities~~Communities'. *PLoS One*~~,~~ 10 (2~~),~~): e0117562. https://doi.org/10.1371/journal.pone.0117562.

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Mis en forme : Français (France)

Mis en forme : Police :Non Italique

Creer, ~~S.,~~ Simon, Kristy Deiner, ~~K.,~~Serita Frey, ~~S.,~~Dorota Porazinska, ~~D.,~~Pierre Taberlet, ~~P.,~~W. Kelley Thomas, ~~W. K., … Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity.~~Caitlin Potter, and Holly M. Bik. 2016. 'The Ecologist's Field Guide to Sequence-Based Identification of Biodiversity'. Edited by Robert Freckleton. *Methods in Ecology and Evolution*~~,~~ 7 (9~~),~~): 1008–~~1018~~18. https://doi.org/10.1111/2041-210X.12574.

Davis, ~~N.~~Nicole M., Diana M. Proctor, ~~D. M.,~~Susan P. Holmes, ~~S. P.,~~David A. Relman, ~~D. A., &~~ and Benjamin J. Callahan~~, B. J. (~~ 2018~~).~~. 'Simple ~~statistical identification~~Statistical Identification and ~~removal~~Removal of ~~contaminant sequences~~Contaminant Sequences in ~~marker-gene~~Marker-Gene and ~~metagenomics data~~Metagenomics Data'. *Microbiome*~~,~~ 6 (1~~),~~): 226. https://doi.org/10.1186/s40168-018-0605-2.

~~de Queiroz, K. (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences*, *102*(Supplement 1), 6600–6607. https://doi.org/10.1073/pnas.0502030102~~

~~De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., … Velayoudon, D. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237). https://doi.org/10.1126/science.1261605~~

Deiner, ~~K.,~~Kristy, Emanuel A. Fronhofer, ~~E. A.,~~Elvira Mächler, ~~E.,~~Jean Claude Walser, ~~J. C., &~~and Florian Altermatt~~, F. (~~ 2016~~).~~. 'Environmental DNA ~~reveals that rivers are conveyer belts of biodiversity information.~~Reveals That Rivers Are Conveyer Belts of Biodiversity Information'. *Nature Communications*~~,~~ 7 (1~~),~~): 12544. https://doi.org/10.1038/ncomms12544.

Deiner, ~~K.,~~Kristy, Jean-Claude C Walser, ~~J. C. C.,~~E Machler, ~~E.,~~Florian Altermatt, ~~F.,~~Elvira Mächler, ~~E., &~~and Florian Altermatt~~, F. (~~ 2015~~).~~. 'Choice of ~~capture and extraction~~

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique, Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

~~methods affect detection of freshwater biodiversity~~Capture and Extraction Methods Affect Detection of Freshwater Biodiversity from ~~environmental DNA.~~Environmental DNA’. *Biological Conservation*~~,~~ 183~~,~~ (March): 53–63. https://doi.org/10.1016/j.biocon.2014.11.018.

Dejean, ~~T.,~~Tony, Alice Valentini, ~~A.,~~Antoine Duparc, ~~A.,~~Stephanie Pellier-Cuit, ~~S.,~~Francois Pompanon, ~~F.,~~Pierre Taberlet, ~~P., &~~and Claude Miaud~~, C. (~~. 2011~~).~~. ‘Persistence of Environmental DNA in Freshwater ~~Ecosystems~~Ecosystems’. *PLoS One*~~,~~ 6 (8). https://doi.org/10.1371/journal.pone.0023398.

Derycke, ~~S.,~~Sofie, Jan Vanaverbeke, ~~J.,~~Annelien Rigaux, ~~A.,~~Thierry Backeljau, ~~T., &~~and Tom Moens~~, T. (~~. 2010~~).~~. ‘Exploring the ~~use~~Use of ~~cytochrome oxidase~~Cytochrome Oxidase c ~~subunit~~Subunit 1 (COI) for DNA ~~barcoding~~Barcoding of ~~free-living marine nematodes.~~Free-Living Marine Nematodes’. Edited by Peter Roopnarine. *PLoS ONE*~~,~~ 5 (10~~),~~): e13716. https://doi.org/10.1371/journal.pone.0013716.

Dickie, ~~I.~~Ian A., Stephane Boyer, ~~S.,~~Hannah L. Buckley, ~~H. L.,~~Richard P. Duncan, ~~R.~~Paul P~~.,~~. Gardner, ~~P. P.,~~Ian D. Hogg, ~~I. D., … Weaver, L. (~~Robert J. Holdaway, et al. 2018~~).~~. ‘Towards ~~robust~~Robust and ~~repeatable sampling methods~~Repeatable Sampling Methods in ~~eDNA-based studies~~EDNA-Based Studies’. *Molecular Ecology Resources*~~, 18(5), 940–952.~~. Wiley/Blackwell (10.1111). https://doi.org/10.1111/1755-0998.12907.

Drummond, A~~.~~ J~~.,~~., R D Newcomb, T R~~. D.,~~ Buckley, ~~T. R.,~~D Xie, ~~D.,~~A Dopheide, ~~A., Potter,~~ B~~.~~ C~~.~~ M~~., … Nelson, N. (~~ Potter, J Heled, et al. 2015~~).~~. ‘Evaluating a ~~multigene environmental~~Multigene Environmental DNA ~~approach~~Approach for ~~biodiversity assessment.~~Biodiversity Assessment’. *Gigascience*~~,~~ 4. https://doi.org/ARTN 4610.1186/s13742-015-0086-1.

Eren, A~~. M.,~~ Murat, Joseph H Vineis, ~~J. H.,~~Hilary G Morrison, ~~H. G., &~~and Mitchell L Sogin~~, M.~~

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique, Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

~~L. (~~. 2013~~).~~ . ‘A Filtering Method to Generate High Quality Short Reads Using Illumina Paired-End ~~Technology~~Technology’. *PLoS ONE*~~,~~ 8 (6~~),~~): e66643. https://doi.org/10.1371/journal.pone.0066643.

Escudié, ~~F.,~~Frédéric, Lucas Auer, ~~L.,~~Maria Bernard, ~~M.,~~Mahendra Mariadassou, ~~M.,~~Laurent Cauquil, ~~L.,~~Katia Vidal, ~~K., … Berger, B. (~~Sarah Maman, et al. 2018~~).~~ . ‘FROGS: Find, Rapidly, OTUs with Galaxy ~~Solution.~~Solution’. Edited by Bonnie Berger. *Bioinformatics*~~,~~ 34 (8~~),~~): 1287–~~1294~~94. https://doi.org/10.1093/bioinformatics/btx791.

Evans, N~~.~~ T~~.,~~, B P Olds, ~~B. P.,~~M A Renshaw, ~~M. A.,~~C R Turner, ~~C. R.,~~Y Y Li, ~~Y. Y.,~~C L Jerde, ~~C. L., …~~A R Mahon, M E Pfrender, G A Lamberti, and D M Lodge~~, D. M. (~~. 2016~~).~~ . ‘Quantification of ~~mesocosm fish~~Mesocosm Fish and ~~amphibian species diversity~~Amphibian Species Diversity via ~~environmental~~Environmental DNA ~~metabarcoding~~Metabarcoding’. *Molecular Ecology Resources*~~,~~ 16 (1~~),~~): 29–41. https://doi.org/10.1111/1755-0998.12433.

Ficetola, ~~G. F.,~~Gentile Francesco, Johan Pansu, ~~J.,~~Aurélie Bonin, ~~A.,~~Eric Coissac, ~~E.,~~Charline Giguet-Covex, ~~C.,~~Marta De Barba, ~~M., … Taberlet, P. (~~Ludovic Gielly, et al. 2015~~).~~ . ‘Replication ~~levels, false presences~~Levels, False Presences and the ~~estimation~~Estimation of the ~~presence/absence~~Presence/Absence from ~~eDNA metabarcoding data.~~EDNA Metabarcoding Data’. *Molecular Ecology Resources*~~,~~ 15 (3~~),~~): 543–~~556~~56. https://doi.org/10.1111/1755-0998.12338.

Fonseca, ~~V.~~Vera G. ~~(~~2018~~).~~ . ‘Pitfalls in ~~relative abundance estimation using edna metabarcoding.~~Relative Abundance Estimation Using Edna Metabarcoding’. *Molecular Ecology Resources*~~,~~ 18 (5~~),~~): 923–~~926~~26. https://doi.org/10.1111/1755-0998.12902.

Fonseca, ~~V.~~Vera G., Gary R Carvalho, ~~G. R.,~~Way Sung, ~~W.,~~Harriet F Johnson, ~~H. F.,~~Deborah M Power, ~~D. M.,~~Simon P Neill, ~~S. P., … Creer, S. (~~Margaret Packer, et al. 2010~~).~~ . ‘Second-

Mis en forme : Français (France)

Mis en forme : Police :Non Italique, Français (France)

Mis en forme : Français (France)

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

~~generation environmental sequencing unmasks marine metazoan biodiversity.~~Generation Environmental Sequencing Unmasks Marine Metazoan Biodiversity'. *Nature Communications*~~,~~ 1. https://doi.org/9810.1038/ncomms1095.

Frøslev, ~~T. G.,~~Tobias Guldberg, Rasmus Kjøller, ~~R.,~~Hans Henrik Bruun, ~~H. H.,~~Rasmus Ejrnæs, ~~R.,~~Ane Kirstine Brunbjerg, ~~A. K.,~~Carlotta Pietroni, ~~C., &~~and Anders Johannes Hansen~~, A. J. (~~. 2017~~).~~. 'Algorithm for ~~post-clustering curation of DNA amplicon data yields reliable biodiversity estimates.~~Post-Clustering Curation of DNA Amplicon Data Yields Reliable Biodiversity Estimates'. *Nature Communications*~~,~~ 8 (1). https://doi.org/10.1038/s41467-017-01312-x.

Gevers, ~~D.,~~ Dirk, Frederick M. Cohan, ~~F. M.,~~Jeffrey G. Lawrence, ~~J.~~Brian G~~.,~~. Spratt, ~~B. G.,~~Tom Coenye, ~~T.,~~Edward J. Feil, ~~E. J., … Swings, J. (~~Erko Stackebrandt, et al. 2005~~).~~. 'Re-~~evaluating prokaryotic species.~~Evaluating Prokaryotic Species'. *Nature Reviews Microbiology*~~,~~ 3 (9~~),~~): 733–~~739~~39. https://doi.org/10.1038/nrmicro1236.

Goldberg, ~~C.~~Caren S., Cameron R. Turner, ~~C. R.,~~Kristy Deiner, ~~K.,~~Katy E. Klymus, ~~K. E.,~~Philip Francis Thomsen, ~~P. F.,~~Melanie A. Murphy, ~~M. A., … Taberlet, P. (~~Stephen F. Spear, et al. 2016~~).~~. 'Critical ~~considerations~~Considerations for the ~~application~~Application of ~~environmental~~Environmental DNA ~~methods~~Methods to ~~detect aquatic species~~Detect Aquatic Species'. Edited by M. Gilbert. *Methods in Ecology and Evolution*~~,~~ 7 (11~~),~~): 1299–1307. https://doi.org/10.1111/2041-210X.12595.

Hashimoto, ~~J.~~Joel G~~.,~~, Bradley S Stevenson, ~~B. S., &~~and Thomas M Schmidt~~, T. M. (~~. 2003~~).~~. 'Rates and ~~consequences~~Consequences of ~~recombination~~Recombination between ~~rRNA operons.~~RRNA Operons'. *Journal of Bacteriology*~~,~~ 185 (3~~),~~): 966–~~972~~72. https://doi.org/10.1128/JB.185.3.966-972.2003.

Hebert, ~~P.~~Paul D.~~-~~N., Sujeevan Ratnasingham, ~~S., &~~and Jeremy R. de Waard~~, J. R. (~~. 2003~~).~~.

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme ...

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme ...

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Mis en forme ...

Mis en forme : Police :Non Italique

‘Barcoding ~~animal life: cytochrome~~Animal Life: Cytochrome c ~~oxidase subunit 1 divergences~~Oxidase Subunit 1 Divergences among ~~closely related species.~~Closely Related Species’. *Proceedings of the Royal Society of London. Series B: Biological Sciences*~~,~~ 270 (suppl_1~~),~~): S96-9. https://doi.org/10.1098/rsbl.2003.0025.

Herrera, Santiago, Hiromi Watanabe, and Timothy M. Shank. 2015. ‘Evolutionary and Biogeographical Patterns of Barnacles from Deep-Sea Hydrothermal Vents’. *Molecular Ecology* 24 (3): 673–89. https://doi.org/10.1111/mec.13054.

Ji, ~~Y.,~~Yinqiu, Louise Ashton, ~~L.,~~Scott M Pedley, ~~S. M.,~~David P Edwards, ~~D. P.,~~Yong Tang, ~~Y.,~~Akihiro Nakamura, ~~A., … Yu, D. W. (~~Roger Kitching, et al. 2013~~).~~. ‘Reliable, ~~verifiable~~Verifiable and ~~efficient monitoring~~Efficient Monitoring of ~~biodiversity~~Biodiversity via ~~metabarcoding.~~Metabarcoding’. *Ecology Letters*~~,~~ 16 (10~~),~~): 1245–~~1257~~57. https://doi.org/10.1111/ele.12162.

Klappenbach, J~~.~~ A~~.,~~., Paul R. Saxman, ~~P.~~Cole James R., ~~R., C. J., &~~and Thomas M. Schmidt~~, T. M. (~~. 2001~~). rrndb: the~~. ‘Rrndb: The Ribosomal RNA Operon Copy Number ~~Database~~Database’. *Nucleic Acids Research*~~,~~ 29 (1~~),~~): 181–~~184~~84. https://doi.org/10.1093/nar/29.1.181.

Leray, ~~M.,~~Matthieu, J Y Yang, ~~J. Y.,~~C P Meyer, S C~~. P.,~~ Mills, ~~S. C.,~~N Agudelo, ~~N.,~~V Ranwez, ~~V., …~~J T Boehm, and Ryuji J. Machida~~, R. J. (~~. 2013~~). A new versatile primer set targeting.~~ ‘A New Versatile Primer Set Targeting a ~~short fragment~~Short Fragment of the ~~mitochondrial~~Mitochondrial COI ~~region~~Region for ~~metabarcoding metazoan diversity: application~~Metabarcoding Metazoan Diversity: Application for ~~characterizing coral reef fish gut contents.~~Characterizing Coral Reef Fish Gut Contents’. *Front Zool*~~,~~ 10~~,~~: 34. https://doi.org/10.1186/1742-9994-10-34.

Macheriotou, ~~L.,~~Lara, Katja Guilini, ~~K.,~~Tania Nara Bezerra, ~~T. N.,~~Bjorn Tytgat, ~~B.,~~Dinh Tu

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Nguyen, ~~D. T.,~~Thi Xuan Phuong Nguyen, ~~T. X., … Rigaux, A. (~~Febe Noppe, et al. 2019~~).~~ . ‘Metabarcoding ~~free-living marine nematodes using curated~~Free-Living Marine Nematodes Using Curated 18S and CO1 ~~reference sequence databases~~Reference Sequence Databases for ~~species-level taxonomic assignments.~~Species-Level Taxonomic Assignments’. *Ecology and Evolution*~~,~~ 9 (1~~),~~): 1–16. https://doi.org/10.1002/ece3.4814.

Machida, ~~R.~~Ryuji J., ~~Leray, M., Ho, S. L., &~~and Nancy Knowlton~~, N. (2017). Data~~. 2012. ‘PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences’. Edited by Jack Anthony Gilbert. *PLoS ONE* 7 (9): e46180. https://doi.org/10.1371/journal.pone.0046180.

Machida, Ryuji J., Matthew Kweskin, and Nancy Knowlton. 2012. ‘PCR Primers for Metazoan Mitochondrial 12S Ribosomal DNA Sequences’. *PLoS ONE* 7 (4). https://doi.org/10.1371/journal.pone.0035887.

Machida, Ryuji J., Matthieu Leray, Shian Lei Ho, and Nancy Knowlton. 2017. ‘Data Descriptor: Metazoan ~~mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples.~~Mitochondrial Gene Sequence Reference Datasets for Taxonomic Assignment of Environmental Samples’. *Scientific Data*~~,~~ 4. https://doi.org/10.1038/sdata.2017.27.

Mahe, F~~.~~., Torbjørn Rognes, ~~T.,~~C Quince, ~~C.,~~Colomban De Vargas, ~~C., &~~and M Dunthorn~~, M. (~~. 2015~~).~~ . ‘Swarm v2: ~~highly-scalable and high-resolution amplicon clustering.~~Highly-Scalable and High-Resolution Amplicon Clustering’. *PeerJ*~~,~~ 3. https://doi.org/Artn E142010.7717/Peerj.1420.

Massana, ~~R.,~~Ramón Ramon, Angélique Gobet, ~~A.,~~Stéphane Audic, ~~S.,~~David Bass, ~~D.,~~Lucie Bittner, ~~L.,~~Christophe Boutte, ~~C., … De Vargas, C. (~~Aurélie Chambouvet, et al. 2015~~).~~ . ‘Marine ~~protist diversity~~Protist Diversity in European ~~coastal waters~~Coastal Waters and ~~sediments~~Sediments as ~~revealed by high-throughput sequencing.~~Revealed by High-

Mis en forme : Police :Non Italique

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique, Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Throughput Sequencing'. *Environmental Microbiology*~~,~~ 17 (10~~),~~): 4035–~~4049~~49. https://doi.org/10.1111/1462-2920.12955.

Mayr, ~~E. (~~Ernst. 1942~~).~~. *Systematics and the ~~origin~~Origin of ~~species~~Species, from the ~~viewpoint~~Viewpoint of a ~~zoologist~~Zoologist*. New York, NY: Columbia University Press. ~~Retrieved from~~ http://www.hup.harvard.edu/catalog.php?isbn=9780674862500.

McMurdie, ~~P.~~Paul J., ~~&~~and Susan Holmes~~, S. (~~. 2013~~).~~. 'Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census ~~Data.~~Data'. Edited by Michael Watson. *PLoS ONE*~~,~~ 8 (4~~),~~): e61217. https://doi.org/10.1371/journal.pone.0061217.

Minoche, ~~A.~~André E~~.,~~., Juliane C Dohm, ~~J. C., &~~and Heinz Himmelbauer~~, H. (~~. 2011~~).~~. 'Evaluation of ~~genomic high-throughput sequencing data generated~~Genomic High-Throughput Sequencing Data Generated on Illumina HiSeq and Genome Analyzer ~~systems~~Systems'. *Genome Biology*~~,~~ 12 (11~~),~~): R112. https://doi.org/10.1186/gb-2011-12-11-r112.

Nearing, ~~J.~~Jacob T., Gavin M. Douglas, ~~G.~~André M~~.,~~., Comeau, ~~A. M., &~~ and Morgan G.I. Langille~~, M. G. I. (~~. 2018~~).~~. 'Denoising the Denoisers: ~~an independent evaluation~~An Independent Evaluation of ~~microbiome sequence error-correction approaches.~~Microbiome Sequence Error-Correction Approaches'. *PeerJ*~~,~~ 6~~,~~: e5364. https://doi.org/10.7717/peerj.5364.

Nichols, ~~R.~~Ruth V., Christopher Vollmers, ~~C.,~~Lee A. Newsom, ~~L. A.,~~Yue Wang, ~~Y.,~~Peter D. Heintzman, ~~P. D.,~~McKenna Leighton, ~~M., …~~Richard E. Green, and Beth Shapiro~~, B. (~~. 2018~~).~~. 'Minimizing ~~polymerase biases~~Polymerase Biases in ~~metabarcoding~~Metabarcoding'. *Molecular Ecology Resources*~~,~~ 18 (5~~),~~): 927–~~939~~39. https://doi.org/10.1111/1755-0998.12895.

Mis en forme : Police :Non Italique

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique, Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique, Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Oksanen, ~~J.,~~Jari, Michael Blanchet, Guillaume F. Friendly, ~~M.,~~Roeland Kindt, ~~R.,~~Pierre Legendre, ~~P.,~~Dan McGlinn, ~~D.,~~R. Peter Minchin, R.~~ P., … Wagner, H. (~~B. O'Hara, et al. 2018~~). vegan~~. 'Vegan: Community Ecology ~~Package. Retrieved from~~Package'. https://cran.r-project.org/package=vegan.

Pansu, ~~J.,~~Johan, Charline Giguet-Covex, ~~C.,~~Francesco Ficetola, ~~G. F.,~~Ludovic Gielly, ~~L.,~~Frederic Boyer, ~~F.,~~Eric Coissac, ~~E., …~~Isabelle Domaizon, Lucie Zinger, Jerome Poulenard, and Fabien Arnaud~~, F. (~~. 2015~~).~~ . 'Environmental DNA ~~metabarcoding~~Metabarcoding to ~~investigate historic changes in biodiversity.~~Investigate Historic Changes in Biodiversity'. *Genome*~~,~~ 58 (5~~),~~): 264.

Parada, A~~.~~ E~~.,~~, D M Needham, ~~D. M., &~~and J A Fuhrman~~, J. A. (~~. 2016~~).~~ . 'Every ~~base matters: assessing small subunit rRNA primers for marine microbiomes~~Base Matters: Assessing Small Subunit RRNA Primers for Marine Microbiomes with ~~mock communities, time series and global field samples.~~Mock Communities, Time Series and Global Field Samples'. *Environ Microbiol*~~,~~ 18 (5~~),~~): 1403–~~1414~~14. https://doi.org/10.1111/1462-2920.13023.

Pawlowski, ~~J.~~Jan W., Richard Christen, ~~R.,~~Beatrice Lecroq, ~~B.,~~Dipankar Bachar, ~~D.,~~Hamid Reza Shahbazkia, ~~H. R.,~~Linda Amaral-Zettler, ~~L., &~~and Laure Guillou~~, L. (~~. 2011~~).~~ . 'Eukaryotic Richness in the Abyss: Insights from Pyrotag ~~Sequencing~~Sequencing'. *PLoS One*~~,~~ 6 (4). https://doi.org/e1816910.1371/journal.pone.0018169.

Pei, ~~A.~~Anna Y., William E Oberdorf, Carlos W~~. E.,~~ Nossa, ~~C. W.,~~Ankush Agarwal, ~~A.,~~Pooja Chokshi, ~~P.,~~Erika A Gerz, ~~E. A., … Pei, Z. (~~Zhida Jin, et al. 2010~~).~~ . 'Diversity of 16S ~~rRNA genes~~RRNA Genes within ~~individual prokaryotic genomes.~~Individual Prokaryotic Genomes'. *Applied and Environmental Microbiology*~~,~~ 76 (12~~),~~): 3886–~~3897~~97. https://doi.org/10.1128/AEM.02953-09.

Phillips, ~~J.~~Jarrett D., Daniel J. Gillis, ~~D. J., &~~ and Robert H. Hanner~~, R. H. (~~. 2019~~, March 1).~~ .

Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...
Mis en forme ...

‘Incomplete ~~estimates~~Estimates of ~~genetic diversity~~Genetic Diversity within ~~species~~Species: Implications for DNA ~~barcoding~~Barcoding’. *Ecology and Evolution*. John Wiley & Sons, Ltd. https://doi.org/10.1002/ece3.4757.

Plouviez, S., T. M. Shank, ~~T. M.,~~B. Faure, ~~B.,~~C. Daguin-Thiebaut, ~~C.,~~F. Viard, F~~.,~~. H. Lallier, ~~F. H., &~~and D. Jollivet~~, D. (~~. 2009~~)~~. ‘Comparative ~~phylogeography~~Phylogeography among ~~hydrothermal vent species~~Hydrothermal Vent Species along the East Pacific Rise ~~reveals vicariant processes~~Reveals Vicariant Processes and ~~population expansion~~Population Expansion in the ~~South~~South’. *Molecular Ecology*~~,~~ 18 (18~~),~~): 3903–~~3917~~17. https://doi.org/10.1111/j.1365-294X.2009.04325.x.

Quast, ~~C.,~~Christian, Elmar Pruesse, ~~E.,~~Pelin Yilmaz, ~~P.,~~Jan Gerken, ~~J.,~~Timmy Schweer, ~~T.,~~Pablo Yarza, ~~P., …~~Jörg Peplies, and Frank Oliver Glöckner~~, F. O. (~~. 2012~~)~~. ‘The SILVA ~~ribosomal~~Ribosomal RNA ~~gene database project: improved data processing and web-based tools.~~Gene Database Project: Improved Data Processing and Web-Based Tools’. *Nucleic Acids Research*~~,~~ 41 (D1~~),~~): D590–~~D596~~96. https://doi.org/10.1093/nar/gks1219.

Queiroz, Kevin de. 2005. ‘Ernst Mayr and the Modern Concept of Species’. *Proceedings of the National Academy of Sciences* 102 (Supplement 1): 6600–6607. https://doi.org/10.1073/pnas.0502030102.

R Core Team. ~~(~~2018~~)~~. ‘R: A ~~language~~Language and ~~environment~~Environment for ~~statistical computing.~~Statistical Computing.’ R Foundation for Statistical Computing, Vienna, Austria.

Salazar, ~~G.,~~Guillem, Francisco M Cornejo-Castillo, ~~F. M.,~~Veronica Benitez-Barrios, ~~V.,~~Eugenio Fraile-Nuez, ~~E.,~~X Anton Alvarez-Salgado, ~~X. A.,~~Carlos M Duarte, ~~C. M., …~~Josep M Gasol, and Silvia G Acinas~~, S. G. (~~. 2016~~)~~. ‘Global ~~diversity~~Diversity and ~~biogeography~~Biogeography of ~~deep-sea pelagic prokaryotes~~Deep-Sea Pelagic Prokaryotes’. *Isme Journal*~~,~~ 10 (3~~),~~): 596–608. https://doi.org/10.1038/ismej.2015.137.

Mis en forme : Français (France)

Mis en forme : Français (France)

Mis en forme : Français (France)

Mis en forme : Français (France)

Mis en forme : Français (France)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Police :Non Italique

Schloss, ~~P.~~Patrick D., Sarah L. Westcott, ~~S. L.,~~Thomas Ryabin, ~~T.,~~Justine R. Hall, ~~J. R.,~~Martin Hartmann, ~~M.,~~Emily B. Hollister, ~~E. B., … Weber, C. F. (~~Ryan A. Lesniewski, et al. 2009~~).~~. ‘Introducing ~~mothur~~Mothur: Open-~~source, platform-independent, community-supported software for describing and comparing microbial communities.~~Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities’. *Applied and Environmental Microbiology*~~,~~ 75 (23~~),~~): 7537–~~7541~~41. https://doi.org/10.1128/AEM.01541-09.

Schnell, Ida Bærholm, Kristine Bohmann, and M. Thomas P. Gilbert. 2015. ‘Tag Jumps Illuminated - Reducing Sequence-to-Sample Misidentifications in Metabarcoding Studies’. *Molecular Ecology Resources* 15 (6): 1289–1303. https://doi.org/10.1111/1755-0998.12402.

Shank, Timothy M., Michael B. Black, Kenneth M. Halanych, Richard A. Lutz, and Robert C. Vrijenhoek. 1999. ‘Miocene Radiation of Deep-Sea Hydrothermal Vent Shrimp (Caridea: Bresiliidae): Evidence from Mitochondrial Cytochrome Oxidase Subunit I’. *Molecular Phylogenetics and Evolution* 13 (2): 244–54. https://doi.org/10.1006/mpev.1999.0642.

Shapiro, B. ~~J.,~~Jesse, Jean Baptiste Leducq, ~~J. B., &~~and James Mallet~~, J. (~~. 2016~~).~~. ‘What Is Speciation?’ Edited by Ivan Matic. *PLoS Genetics*~~,~~ 12 (3~~),~~): e1005860. https://doi.org/10.1371/journal.pgen.1005860.

Sinniger, ~~F.,~~Frederic, Jan W. Pawlowski, ~~J. W.,~~Saki Harii, ~~S.,~~Andrew J. Gooday, ~~A. J.,~~Hiroyuki Yamamoto, ~~H.,~~Pierre Chevaldonné, ~~P., … Creer, S. (~~Tomas Cedhagen, Gary Carvalho, and Simon Creer. 2016~~).~~. ‘Worldwide ~~analysis~~Analysis of ~~sedimentary~~Sedimentary DNA ~~reveals major gaps~~Reveals Major Gaps in ~~taxonomic knowledge of deep-sea benthos.~~Taxonomic Knowledge of Deep-Sea Benthos’. *Frontiers in Marine Science*~~,~~ 3 (June~~),~~): 92. https://doi.org/10.3389/FMARS.2016.00092.

Slon, ~~V.,~~Viviane, Charlotte Hopfe, ~~C.,~~Clemens L Weiß, ~~C. L.,~~Fabrizio Mafessoni, ~~F., de la~~Marco

Mis en forme : Police :Non Italique

Mis en forme : Français (France)

Mis en forme : Police :Non Italique

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique, Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

De La Rasilla, ~~M.,~~Carles Lalueza-Fox, ~~C., … Meyer, M. (~~Antonio Rosas, et al. 2017~~). ~~. 'Neandertal and Denisovan DNA from Pleistocene ~~sediments~~Sediments'. *Science* ~~(New York, N.Y.),~~ 356 (6338~~),~~): 605–~~608~~8. https://doi.org/10.1126/science.aam9695.

Sokal, ~~R. R., &~~ Robert R., and Theodore J. Crovello~~, T. J. (~~. 1970~~).~~. 'The Biological Species Concept : A Critical ~~Evaluation~~Evaluation'. *The American Naturalist*~~,~~ 104 (936~~),~~): 127–~~153~~53.

Song, ~~H.,~~Hojun, Jennifer E Buhay, ~~J. E.,~~Michael F Whiting, ~~M. F., &~~and Keith A Crandall~~, K. A. (~~. 2008~~).~~. 'Many ~~species~~Species in ~~one~~One: DNA ~~barcoding overestimates~~Barcoding Overestimates the ~~number~~Number of ~~species when nuclear mitochondrial pseudogenes are coamplified~~Species When Nuclear Mitochondrial Pseudogenes Are Coamplified'. *Proceedings of the National Academy of Sciences of the United States of America*~~,~~ 105 (36~~),~~): 13486–~~13491~~91. https://doi.org/10.1073/pnas.0803076105.

Stat, ~~M.,~~ Michael, Megan J. Huggett, ~~M. J.,~~Rachele Bernasconi, ~~R.,~~Joseph D. Dibattista, ~~J. D.,~~Tina E. Berry, ~~T. E.,~~Stephen J. Newman, ~~S. J., …~~Euan S. Harvey, and Michael Bunce~~, M. (~~. 2017~~).~~. 'Ecosystem ~~biomonitoring~~Biomonitoring with ~~eDNA: metabarcoding~~EDNA: Metabarcoding across the ~~tree~~Tree of ~~life~~Life in a ~~tropical marine environment.~~Tropical Marine Environment'. *Scientific Reports*~~,~~ 7. https://doi.org/10.1038/s41598-017-12501-5.

Stefanni, Sergio, David Stanković, Diego Borme, Alessandra de Olazabal, Tea Juretić, Alberto Pallavicini, and Valentina Tirelli. 2018. 'Multi-Marker Metabarcoding Approach to Study Mesozooplankton at Basin Scale'. *Scientific Reports* 8 (1): 12085. https://doi.org/10.1038/s41598-018-30157-7.

Taberlet, ~~P.,~~Pierre, Eric Coissac, ~~E.,~~Mehrdad Hajibabaei, ~~M., &~~and Loren H. Rieseberg~~. L. H. (~~. 2012~~).~~. 'Environmental ~~DNA~~DNA'. *Molecular Ecology*~~,~~ 21 (8~~),~~): 1789–~~1793~~93. https://doi.org/10.1111/j.1365-294X.2012.05542.x.

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique, Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique, Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Tang, ~~C.~~Cuong Q., Francesca Leasi, ~~F.,~~Ulrike Obertegger, ~~U.,~~Alexander Kieneke, ~~A.,~~Timothy G Barraclough, ~~T. G., &~~and Diego Fontaneto~~, D. (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity~~. 2012. 'The Widely Used Small Subunit 18S RDNA Molecule Greatly Underestimates True Diversity in ~~biodiversity surveys~~Biodiversity Surveys of the ~~meiofauna.~~Meiofauna.' *Proceedings of the National Academy of Sciences of the United States of America*~~,~~ 109 (40~~),~~): 16208–~~16221~~212. https://doi.org/10.1073/pnas.1209160109.

Teixeira, Sara, Karine Olu, Carole Decker, Regina L Cunha, Sandra Fuchs, Stéphane Hourdez, Ester A. Serrão, and Sophie Arnaud-Haond. 2013. 'High Connectivity across the Fragmented Chemosynthetic Ecosystems of the Deep Atlantic Equatorial Belt: Efficient Dispersal Mechanisms or Questionable Endemism?' *Molecular Ecology* 22 (18): 4663–80. https://doi.org/10.1111/mec.12419.

Valentini, ~~A.,~~Alice, François Pompanon, ~~F., &~~and Pierre Taberlet~~, P. (~~. 2009~~, February 1).~~. 'DNA ~~barcoding~~Barcoding for ~~ecologists~~Ecologists'. *Trends in Ecology and Evolution*. Elsevier Current Trends. https://doi.org/10.1016/j.tree.2008.09.011.

Valentini, ~~A.,~~Alice, Pierre Taberlet, ~~P.,~~Claude Miaud, ~~C.,~~Raphaël Raphael Civade, ~~R.,~~Jelger Herder, ~~J.,~~Philip Francis Thomsen, ~~P. F., … Dejean, T. (~~Eva Bellemain, et al. 2016~~).~~. 'Next-~~generation monitoring~~Generation Monitoring of ~~aquatic biodiversity using environmental~~Aquatic Biodiversity Using Environmental DNA ~~metabarcoding.~~Metabarcoding'. *Molecular Ecology*~~,~~ 25 (4~~),~~): 929–~~942~~42. https://doi.org/10.1111/mec.13428.

Vargas, Colomban De, Stéphane Audic, Nicolas Henry, Johan Decelle, Frédéric Mahé, Ramiro Logares, Enrique Lara, et al. 2015. 'Eukaryotic Plankton Diversity in the Sunlit Ocean'. *Science* 348 (6237). https://doi.org/10.1126/science.1261605.

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Mis en forme : Français (France)

Mis en forme : Police :Non Italique

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Mis en forme : Espagnol (Espagne)

Wangensteen, ~~O.~~Owen S., ~~&~~and Xavier Turon~~, X. (~~. 2016~~).~~. ‘Metabarcoding Techniques for Assessing Biodiversity of Marine Animal ~~Forests~~Forests’. In ~~S.~~ *Marine Animal Forests,* edited by S. Rossi, L. Bramanti, A. Gori, ~~&~~and C. Orejas Saco del Valle~~ (Eds.), *Marine Animal Forests* (pp.~~, 1–29~~)~~.. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17001-5_53-1.

Xiong, ~~W., &~~Wei, and Aibin Zhan~~, A. (~~. 2018~~).~~. ‘Testing ~~clustering strategies~~Clustering Strategies for ~~metabarcoding-based investigation~~Metabarcoding-Based Investigation of ~~community–environment interactions.~~Community–Environment Interactions’. *Molecular Ecology Resources*~~,~~ 18 (6~~),~~): 1326–~~1338~~38. https://doi.org/10.1111/1755-0998.12922.

Yoccoz, N~~.~~ G~~.,~~, K A Brathen, ~~K. A.,~~L Gielly, ~~L.,~~J Haile, ~~J.,~~M E Edwards, ~~M. E.,~~T Goslar, ~~T., … Taberlet, P. (~~H von Stedingk, et al. 2012~~).~~. ‘DNA from ~~soil mirrors plant taxonomic and growth form diversity.~~Soil Mirrors Plant Taxonomic and Growth Form Diversity’. *Molecular Ecology*~~,~~ 21 (15~~),~~): 3647–~~3655~~55. https://doi.org/10.1111/j.1365-294X.2012.05545.x.

Yu, ~~D.~~Douglas W~~.,~~, Yinqiu Ji, ~~Y.,~~Brent C Emerson, ~~B. C.,~~Xiaoyang Wang, ~~X.,~~Chengxi Ye, ~~C.,~~Chunyan Yang, ~~C., &~~and Zhaoli Ding~~, Z. (~~. 2012~~).~~. ‘Biodiversity ~~soup: metabarcoding~~Soup: Metabarcoding of ~~arthropods~~Arthropods for ~~rapid biodiversity assessment and biomonitoring.~~Rapid Biodiversity Assessment and Biomonitoring’. *Methods in Ecology and Evolution*~~,~~ 3 (4~~),~~): 613–~~623~~23. https://doi.org/10.1111/j.2041-210X.2012.00198.x.

Zanol, Joana, Kenneth M. Halanych, Torsten H. Struck, and Kristian Fauchald. 2010. ‘Phylogeny of the Bristle Worm Family Eunicidae (Eunicida, Annelida) and the Phylogenetic Utility of Noncongruent 16S, COI and 18S in Combined Analyses’. *Molecular Phylogenetics and Evolution* 55 (2): 660–76. https://doi.org/10.1016/j.ympev.2009.12.024.

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Anglais (Royaume-Uni)

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

Zhan, Aibin, Sarah A~~.,~~. Bailey, ~~S. A.,~~Daniel D. Heath, ~~D. D., &~~and Hugh J. Macisaac~~, H. J. (~~. 2014~~).~~. 'Performance ~~comparison~~Comparison of ~~genetic markers~~Genetic Markers for ~~high-throughput sequencing-based biodiversity assessment in complex communities.~~High-Throughput Sequencing-Based Biodiversity Assessment in Complex Communities'. *Molecular Ecology Resources*~~,~~, 14 (5~~),~~): 1049–~~1059~~59. https://doi.org/10.1111/1755-0998.12254.

Zinger, ~~L.,~~Lucie, Jérôme Chave, ~~J.,~~Eric Coissac, ~~E.,~~Amaia Iribar, ~~A.,~~Eliane Louisanna, ~~E.,~~Sophie Manzi, ~~S., … Taberlet, P. (~~Vincent Schilling, Heidy Schimann, Guilhem Sommeria-Klein, and Pierre Taberlet. 2016~~).~~. 'Extracellular DNA ~~extraction is~~Extraction Is a ~~fast, cheap~~Fast, Cheap and ~~reliable alternative~~Reliable Alternative for ~~multi-taxa surveys based~~Multi-Taxa Surveys Based on ~~soil DNA~~Soil DNA'. *Soil Biology and Biochemistry*~~,~~, 96~~,~~: 16–19. https://doi.org/10.1016/j.soilbio.2016.01.008.

Mis en forme : Police :Non Italique

Mis en forme : Police :Non Italique

DATA ACCESSIBILITY

The data for this work can be accessed in the European Nucleotide Archive (ENA) database (Study accession number will be given upon manuscript acceptance). The data set, including sequences, databases, as well as raw and refined ASV/OTU tables, has been deposited on ftp://ftp.ifremer.fr/ifremer/dataref/bioinfo/merlin/abyss/BioinformaticPipelineComparisons/. Bioinformatic scripts, config files, and R scripts are available on Gitlab (https://gitlab.ifremer.fr/abyss-project/).

AUTHOR CONTRIBUTIONS

MIB and SAH designed the study, MIB and JP carried out the laboratory and molecular work; MIB and BT performed the bioinformatic and statistical analyses. LQ assisted in the bioinformatic development and participated in the study design. MIB and SAH wrote the manuscript. All authors contributed to the final manuscript.