Title: Towards reproducible metabarcoding data - lessons from an international cross-laboratory experiment

Running title: Towards reproducible metabarcoding data

Anastasija Zaiko[1,2*], Paul Greenfield[3,4], Cathryn Abbott[5], Ulla von Ammon[1], Jaret Bilewitch[6], Michael Bunce[7], Melania E. Cristescu[8], Anthony Chariton[4], Eddy Dowle[9], Jonathan Geller[10], Alba Ardura Gutierrez[11], Mehrdad Hajibabaei[12], Emmet Haggard[10], Graeme J. Inglis[13], Shane Lavery[2,14], Aurelija Samuiloviene[15], Tiffany Simpson[16], Michael Stat[17], Sarah Stephenson[3], Judy Sutherland[6], Vibha Thakur[14], Kristen Westfall[5], Susanna A. Wood[1], Michael Wright[12], Guang Zhang[8], Xavier Pochon[1,2]

[1]Coastal and Freshwater Group, Cawthron Institute, Private Bag 2, Nelson 7042, New Zealand

[2]Institute of Marine Science, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

[3]Commonwealth Scientific and Industrial Research Organisation (CSIRO), GPO Box 1700, Canberra ACT 2601, Australia

[4]Environmental (e)DNA and Biomonitoring Lab, Department of Biological Sciences, Macquarie University, Sydney 2113, Australia

[5]Fisheries and Oceans Canada, Pacific Biological Station, 3190 Hammond Bay Rd, Nanaimo BC, V9T 6N7, Canada

[6]National Institute of Water & Atmospheric Research Ltd (NIWA), 301 Evans Bay Parade, Hataitai, Wellington 6021, New Zealand

[7]Environmental Protection Authority, Level 10/215 Lambton Quay, Wellington 6011, New Zealand

[8]McGill University, 1205 Dr. Penfield Ave, Montreal, Quebec, H3A 1B1, Canada

[9]School of Biomedical Sciences, University of Otago, PO Box 56, Dunedin 9054, New Zealand

[10]Moss Landing Marine Laboratories, San Jose State University, Moss Landing, California, 95039 USA

[11]Department of Functional Biology, University of Oviedo, 33006 Oviedo, Spain

[12]Canada University of Guelph, 50 Stone Rd E, Guelph, Canada

[13]National Institute of Water & Atmospheric Research Ltd (NIWA), PO Box 8602, Christchurch, New Zealand

[14]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

[15]Marine Research Institute, Klaipeda University, 92294 Klaipeda, Lithuania

[16]Curtin University, Kent St, Bentley WA 6102, Perth, Australia

[17]The University of Newcastle, University Dr, Callaghan NSW 2308, Newcastle, Australia

*Corresponding author: anastasija.zaiko@cawthron.org.nz

Abstract

Advances in high-throughput sequencing (HTS) are revolutionizing monitoring in marine environments by enabling rapid, accurate and holistic detection of species within complex biological samples. Research institutions worldwide increasingly employ HTS methods for biodiversity assessments. However, variance in laboratory procedures, analytical workflows and bioinformatic pipelines impede the transferability and comparability of results across research groups. An international experiment was conducted to assess the consistency of metabarcoding results derived from identical samples and primer sets using varying laboratory procedures. Homogenized biofouling samples collected from four coastal locations (Australia, Canada, New Zealand and the USA) were distributed to 12 independent laboratories. Participants were asked to follow one of two HTS library preparation workflows. While DNA extraction, primers and bioinformatic analyses were purposefully standardized to allow comparison, many other technical variables were allowed to vary among laboratories (e.g., amplification protocols, type of instrument used, etc.). Despite substantial variation observed in raw results, the primary signal in the data was consistent, with the samples grouping strongly by geographic origin for all datasets. Simple post-hoc data clean-up by removing low quality samples gave the best improvement in sample classification for nuclear 18S rRNA gene data, with an overall 92.81% correct group attribution. For mitochondrial COI gene data, the best classification result (95.58%) was achieved after correction for contamination errors. The identified critical methodological factors that introduced the greatest variability (preservation buffer, sample defrosting, template concentration, DNA polymerase, PCR enhancer) should be of great assistance in standardizing future comparative biodiversity studies using metabarcoding.

Keywords: reproducibility, high-throughput sequencing, 18S ribosomal rRNA (18S rRNA), mitochondrial Cytochrome c Oxidase subunit 1 (COI), metabarcoding, standardization.

Introduction

Recent advances in high-throughput sequencing (HTS) are revolutionizing ecology, offering unprecedented opportunities for new species discovery, monitoring ecological trends, diet analysis and environmental impact assessment (Aylagas, Borja, & Rodrigues-Ezpeleta, 2014; L. J. Clarke, Trebilco, Walters, Polanowski, & Deagle, 2020; Dowle, Pochon, Banks, Shearer, & Wood, 2016; Keeley, Wood, & Pochon, 2018; Valentini et al., 2016; Zhan et al., 2013). Metabarcoding enables community-wide biodiversity to be characterized from environmental DNA (eDNA) samples. It is based on the mass-sequencing of short, conserved DNA fragments (molecular markers or barcodes), which are then assigned taxonomy based on reference sequence databases, allowing the simultaneous characterization of a large range of taxa (Hajibabaei, Shokralla, Zhou, Singer, & Baird, 2011; Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012). Despite known limitations (e.g., incomplete reference sequence databases, uncertainties around sensitivity, specificity and detection probabilities), this approach is increasingly being advocated as a rapid, cost-effective and sensitive tool for environmental monitoring and assessment (Cordier et al., 2020; Lacoursière-Roussel et al., 2018; Wood et al., 2013). Metabarcoding can be used to detect all life-stages, including morphologically indistinguishable larval forms and rare or sparsely distributed populations, often with less time-effort than conventional approaches (Brown, Chain, Zhan, MacIsaac, & Cristescu, 2016; Comtet, Sandionigi, Viard, & Casiraghi, 2015; Zaiko, Samuiloviene, Ardura, & Garcia-Vazquez, 2015). In some situations, it offers greater potential for standardization than traditional morphological biodiversity assessment (Aylagas, Borja, Irigoien, & Rodríguez-Ezpeleta, 2016; Porter & Hajibabaei, 2018). However, metabarcoding is inherently complex, with many technical and analytical steps that inevitably vary among users, as well as multiple reputable but distinct metabarcoding workflows (Cristescu, 2014; Goldberg et al., 2016; Murray, Coghlan, & Bunce, 2015; Zaiko, Pochon, Garcia-Vazquez, Olenin, & Wood, 2018). Virtually all steps in the workflow vary among laboratories and individual studies, including sampling, preservation methods, DNA extraction, PCR amplification protocols, library preparation methods, sequencing platforms and bioinformatic pipelines (Bailet et al.,

2020; Bowers et al., 2021). This creates a significant challenge for the standardization of protocols and the implementation of a unified, cross-calibrated approach.

For sensitive applications such as biosecurity surveillance, a field devoted to monitoring environments for species with potentially harmful ecological and economic consequences (Molnar, Gamboa, Revenga, & Spalding, 2008; Simberloff et al., 2013), standards of quality assurance and control are of higher importance than in general biodiversity studies (Darling & Mahon, 2011; Lehtiniemi et al., 2015). Reporting unverified biosecurity risks from HTS data may lead to inadequate management responses and even potential legal entanglements for researchers (Darling, Pochon, Abbott, Inglis, & Zaiko, 2020). Despite recent calls for improved consistency in HTS-based research outputs (Jeunen et al., 2019; van der Loos & Nijland, 2020), laboratories working in this field often follow different analytical workflows, develop and apply in-house reference sequence databases, use customized bioinformatic pipelines and often fail to report methodologically relevant metadata consistently (Goldberg et al., 2016; Nicholson et al., 2020). This lack of harmonization may impede the reproducibility of research, transferability and comparability of results, and ultimately, the credibility of HTS-based biodiversity assessments (Zaiko et al., 2018). There are an increasing number of empirical studies addressing the effect of single components of the workflow, such as filtration, eDNA extraction methods, PCR thermocycling conditions, choice of polymerase or sequencing platform used (Aylagas et al., 2016; Braukmann et al., 2019; Djurhuus et al., 2017; Jeunen et al., 2019; Nichols et al., 2018). However, what is still largely unknown is the degree to which variations in common metabarcoding protocols distort community composition data and the impact of these potential biases on interpreting of metabarcoding-derived biodiversity information.

Acknowledging the need for a coordinated effort to accelerate the uptake of HTS-based tools for biodiversity surveys and conservation applications, an international cross-laboratory experiment was conducted to evaluate replicability of a metabarcoding protocol and explore how laboratory-based variance in sample handling and processing impacts biodiversity assessments. The overarching goal of the present study was to investigate the consistency and reproducibility of metabarcoding results when a set of identical

environmental samples are analyzed simultaneously by different laboratories (12 in this study) following a semi-standardized protocol and analytical workflow. The applied experimental design allowed us to identify discrepancies caused by laboratory-specific variation in technical steps and find the weakest links in the analytical pipelines that resulted in the greatest divergence in metabarcoding data. In our study we hypothesized that: (1) despite laboratory-to-laboratory variations (e.g., technical equipment, PCR conditions and reagents), consistent patterns of variability in biological community information would be derived for two markers (COI and 18S) when using a semi-standardized protocol; (2) the use of different HTS library build protocols (dual index primers *versus* fusion primers) would have no effect on the community composition data but might impact levels of contamination; (3) the greatest variability would be introduced in the PCR and clean-up steps, given that different reagents and equipment were used in each laboratory; and (4) if variability in community composition was observed between laboratories, this could be mitigated by applying post-hoc data correction.

Methods

An overview of the experimental set-up is presented in Figure 1, with additional details provided in the sections below.

Collection and processing of biofouling samples

Mature marine biofouling (~20 mL) was collected from upper 0.5-1 m depth at international locations adjacent to busy marine ports and marinas (Table 1). The samples were preserved with RNAlater™ (ThermoFisher Scientific, USA) or LifeGuard™ (Qiagen, Venlo, The Netherlands) by adding a volume of preservative solution equivalent to at least 150% of the biological sample volume (Table 1).

All samples were delivered to the Cawthron Institute, New Zealand (transported in ambient temperature within 2 days) and processed uniformly to ensure the same starting material was used by all

laboratories. Samples were homogenized by bead beating (1,500 rpm, 5 min; 1600 MiniG Spex SamplePrep NJ, USA) and split into 4 × 0.5 mL aliquots per laboratory (192 aliquots in total, Figure 1). Twelve well-established molecular laboratories in six countries participated in this study: New Zealand (Cawthron Institute, National Institute of Water and Atmospheric Research Ltd, University of Auckland and University of Otago); Australia (Curtin University and Macquarie University); USA (Moss Landing Marine Laboratory); Canada (Department of Fisheries and Oceans, McGill University and University of Guelph); Spain (University of Oviedo) and Lithuania (Klaipeda University).

Environmental DNA isolation and sequencing library preparation

Variation at the DNA extraction step was minimized by providing participants from each of the 12 laboratories with a MoBio/Qiagen PowerBiofilm DNA isolation kit (Qiagen, Venlo, The Netherlands) and a detailed manufacturer's protocol for DNA isolation, that all participants were asked to follow.

Participating laboratories were randomly allocated to a treatment group that followed one of two HTS library preparation workflows (dual index primers or fusion primers, see below) for amplification of both the 18S rRNA (Zhan et al. 2013; Uni18S: 5'-AGGGCAAKYCTGGTGCCAGC-3', Uni18SR: 5'-GRCGGTATCTRATCGYCTT-3') and COI (Leray et al. 2013; Geller et al. 2013; mlCOIintF: 5'-GGWACWGGWTGAACWGTWTAYCCYCC-3', jgHCO2198: 5'-TAIACYTCIGGRTGICCRAARAAYCA-3') genes. For analytical objectivity, the laboratories within each group were randomly assigned a letter code – A to F for the dual index and G to L fusion groups, respectively. All results are hereafter reported under those codes to avoid attributing analytical outcomes to a particular laboratory.

Protocols for PCRs and library preparation were also provided (Supporting Information, Appendices I and II); however, given that each laboratory uses different equipment and/or used their preferred reagents, some flexibility in equipment and reagents was allowed at the PCR and clean-up steps.

The participants recorded any deviations from the provided protocol, as well as specific details on sample handling and any issues encountered during the workflow (Table S1).

High-throughput sequencing

Libraries were sequenced on MiSeq™ Illumina instruments at two sequencing facilities: University of Guelph, Canada (libraries from laboratories A-F, dual index primers, n = 120) and Curtin University, Australia (libraries from laboratories G-L, fusion primers, n = 120). For the dual index library, 18S rRNA gene and COI amplicons were pooled for each sample and indexed using Illumina's Nextera XT v2 index adapters (FC-131-2001). Indexed samples were pooled and purified using AMPure XP magnetic bead kit (Beckman Coulter). The library concentration was quantified with Invitrogen PicoGreen dsDNA assay kit on a TBS-380 fluorometer (Turner Biosciences). Fragment length was assessed on an Agilent Bioanalyzer using Agilent's DNA 7500 kit. The library was diluted to 4nM and sequenced on an Illumina MiSeq using a v3 chemistry kit (2 × 300). Each run had a 10% PhiX spike-in. Samples were sequenced over two runs: Run 1 contained groups A, D and F. Run 2 contained groups B, C, and E.

For the fusion library, PCR products were quantified on arrival at the sequencing facility with Qubit (ThermoFisher Scientific, USA) and blended at equal molarity for three sequencing runs: Run 1 - 18S rRNA gene and COI from laboratories G and L; Run 2 - 18S rRNA gene and COI from laboratories H and K; Run 3 - 18S rRNA gene and COI from laboratories I and J. Amplicons from the laboratory J contained traces of green dye from the Taq buffer and were further cleaned using a QiaQuick Purification Kit (QIAGEN, Carlsbad, USA), and eluted in 30 µL elution buffer before quantification and library pooling. Final libraries were quantified again using Qubit. Molarity was calculated and diluted to 2 nM, except for Run 1, which had concentration of 1.3 nM. Libraries were denatured and blended with 20 pM PhiX. Sequencing was performed on the MiSeq (Illumina) system with a v2 chemistry kit (2 × 250).

Bioinformatics

Bioinformatics analysis of all the 18S rRNA and COI datasets was conducted at the Commonwealth Scientific and Industrial Research Organization (CSIRO, Australia), using the GHAP amplicon pipeline (https://doi.org/10.4225/08/59f98560eba25). This is a conventional amplicon clustering and classification pipeline built around tools from USearch (Edgar, Haas, Clemente, Quince, & Knight, 2011), combined with in-house scripted tools for demultiplexing and generating OTU tables. The amplicon reads were demultiplexed, split and trimmed as needed and the read pairs were then merged and de-replicated. The merged reads were trimmed/size-selected and UNOISE3 (Edgar, 2016) used to group almost identical sequences to generate zOTUs (Zero-radius Operational Taxonomic Units). Each zOTU sequence was then classified by BLASTing against a curated set of reference sequences. The 18S rRNA gene zOTUs were matched against curated sequences derived from the SILVA v128 SSU reference set (Quast et al., 2013). The COI zOTUs were matched against a set of COI reference sequences downloaded from GenBank (on 24-9-2019). The pipeline then mapped the merged reads back onto the zOTU sequences to get accurate read counts for each zOTU/sample pair, and generated zOTU abundance tables in both text and .biom (v1) formats, complete with taxonomic classifications and species assignments. The zOTU abundance tables were then summarized at all resolved taxonomic levels, combining the counts for identified taxa across all OTUs.

Both the dual index and fusion library sequencing was undertaken on pooled samples, so each data file contained both 18S rRNA gene and COI data. These files were split into 18S rRNA gene and COI files based on the primer sequences found at the start of each of the reads, and these primer regions were then trimmed from the reads. After splitting, there were found to be unequal numbers of 18S rRNA gene and COI reads in both datasets, with only 6-15% (fusion) and 30% (dual index) of the samples being assigned to the 18S rRNA gene.

The fusion sequencing protocol also added 6 or 7 bp barcode to the start of each read, with these barcodes removed from the reads during demultiplexing. The MiSeq runs used for the fusion libraries produced 260 bp reads (including barcodes and primers), and demultiplexing/trimming these resulted in

reads of 226-228 bp (18S rRNA) and 232-234 bp (COI). The dual index MiSeq runs produced 301 bp reads, including primer sequences, and used standard Illumina out-of-read barcodes. The number of bases removed by this post-sequencing trimming step varied from sample to sample, with higher levels of trimming observed for 18S rRNA reads (median 10.5 bp) than for COI (median 0.7 bp) reads. Samples from Lab C were more heavily trimmed than other samples (average trim for 18S rRNA was 34.5 bp, and 75.3 bp for COI). After splitting and primer-trimming, the reads going into the amplicon pipeline were 15-282 bp long (mean 246-280 bp) for 18S rRNA gene, and 9-275 bp (mean 200-275 bp) for COI. The inter-primer region being sequenced was ~410 bp for 18S rRNA gene, and ~310 bp for COI. Consequently, both the dual index and fusion library reads were long enough to have sufficient overlap for satisfactory pair-merging.

Statistical analyses

For each zOTU table (18S rRNA gene and COI) resulting from the bioinformatics pipeline, a permutational multivariate analysis of variance (PERMANOVA, Anderson 2001, 2017) was used to test how variation in HTS-derived biodiversity was partitioned in response to 'sample origin', 'library' (dual index *vs* fusion; both fixed factors) and 'laboratory' (a random factor nested in 'library'). Analysis was performed on Bray-Curtis similarity matrices of fourth root transformed datasets (zOTU reads data), applying 999 permutations of residuals under a reduced model with type III sum of squares applied. To assess the relative spread of data clouds in samples grouped by origin, a test for homogeneity of multivariate dispersions (PERMDISP) was applied for each dataset with 999 permutations. Individual sample distances to origin group centroids were retrieved for library type, 18S rRNA gene and COI separately. A distribution-free Wilcoxon signed-rank test was used to check whether 18S rRNA and COI OTU richness differ significantly between libraries, as well as whether distances to group centroids differ significantly between 18S rRNA and COI datasets.

To infer which factors affected the consistency of biodiversity information during sample handling and processing, a multiple regression model with automated stepwise selection function was

implemented in R Stats package version 3.5.3 (R-project, 2014). In the model, distance of individual samples to centroids of the sample origin group was treated as response variable, and the following sample handling and processing parameters were considered as predictor variables: preservation buffer (RNAlater/LifeGuard), storage time before DNA extraction (months), multiple (more than two times) DNA defrosting cycles due to PCR troubleshooting (yes/no), library type (fusion/dual index), amount of template used in PCR (µL), dilution of DNA template (yes/no), DNA polymerase used, addition of Bovine Serum Albumin (BSA) or other enhancers (yes/no), number of PCR cycles, contamination detected in control samples (yes/no). For each gene and library type we also performed correlation analysis using cor.test function implemented in R to test for association (Pearson product-moment correlation coefficient) between sequence Phred quality score (averaged per sample) and distance to origin group centroids.

For each dataset (18S rRNA gene and COI), a series of Canonical Analyses of Principal Coordinates (CAP) with 'sample geographic origin' as a grouping factor and 999 permutations were applied to assess the classification success (correct allocation of samples into a pre-defined group) on raw (uncorrected) HTS data and following post-hoc corrections: **Corr1**) samples with low sequence number (<1,000) excluded; **Corr2**) samples randomly rarefied to 10,000 sequence depth; **Corr3**) OTUs found in the negative controls removed from the corresponding subsets of data; **Corr4**) maximum read abundance of OTUs found in the negative controls subtracted from the corresponding subsets of data; **Corr5**) OTUs found in negative controls subtracted and samples with low sequence number (<1,000) excluded; **Corr6**) OTUs found in negative controls were subtracted and samples rarefied to 10,000 sequence depth. Clustering of samples according to their origin was visualized by CAP ordination plots for the raw dataset and corrected dataset with the best classification success. All multivariate statistical analyses were carried out using PRIMER 7 (v 7.0.13) with the PERMANOVA + add-on (K. R. Clarke & Gorley, 2015).

Following dataset corrections which resulted in the best 18S rRNA and COI sample classification according to the origin, the laboratory and library-driven variability in biodiversity data were explored

using Principal Coordinates Analysis (PCoA). For each sample origin, 18S rRNA and COI datasets, a two-dimensional visualization was performed using plot_ordination function on Bray-Curtis similarity matrices in phyloseq R package (McMurdie & Holmes, 2013).

Results

General observations

Although all 12 laboratories strictly adhered to the provided DNA extraction protocol, large variations in DNA concentration and quality were observed (Figure S1). However, different instruments were used to estimate DNA concentrations (Table S1) and some participants expressed concerns about accuracy of their measurements. The quantity and quality of extracted DNA is not a direct indication of how the targeted eukaryotic genes will amplify, as bulk samples are likely to comprise large proportions of bacterial DNA (not targeted in this study). Therefore, DNA concentration data were not used in subsequent statistical analyses.

At the PCR and clean-up steps, all laboratories applied variations to the suggested protocols. Seven laboratories reported difficulties in amplifying one or both marker genes and had to troubleshoot by adjusting cycling conditions, trying different reagents and/or template concentrations (Table S1). A variety of positive controls were used in library preparation by different laboratories, including DNA from a single species, or mock communities derived from combined DNA of multiple species (see Table S1).

The HTS from 240 amplicons (120 from the dual index and 120 from the fusion libraries) resulted in 15,104,359 and 41,241,053 pair-end sequence reads for dual index and fusion primers library, respectively. The total number of quality-filtered, de-noised (non-chimeric) sequences derived from 18S rRNA gene amplicons were 6,231,102 and 4,211,218 from dual index and fusion primers library, and 7,776,009 and 27,540,509 for COI amplicons from dual index and fusion primers library, respectively. Two samples from the 18S rRNA gene dataset failed to produce any high-quality sequence reads and

were therefore excluded from downstream analyses. The average sequence read Phred quality was relatively consistent for 18S rRNA and COI gene amplicons, for both library types (Figure S2). However, the sequence quality of the COI fusion library samples was, overall, slightly higher for both forward and reverse reads.

High variability in the number of filtered sequences and OTUs across samples and laboratories was observed (Figures 2 and 3). In the fusion library group, the number of good quality sequences yielded per sample varied from 1 to 50,101 for 18S rRNA gene amplicons and 347 to 556,775 for COI amplicons, respectively. In the dual index library, differences in sequence yields between genes were similar, with data from the18S rRNA gene ranging from 1 to 189,779 and COI from 6 to 219,694. The number of retrieved OTUs in the dual index library ranged from 1 to 465 and 4 to 1,622 for 18S rRNA gene and COI markers, respectively, and showed significant ($p<0.001$) moderate correlation with the number of sequence reads ($r=0.52$ and 0.68, 18S rRNA and COI samples, respectively). In the fusion library, OTU numbers correlated only with sequence reads for the 18S rRNA samples (0.46, $p<0.001$), ranging from 1 to 436 OTUs per sample. COI samples were characterized by an overall higher number of OTUs (79 to 2,363 per sample), and this did not correlate with the number of sequence reads ($r=0.18$, $p=0.08$). The 18S rRNA OTU richness did not differ significantly between libraries ($W = 4975$, $p = 0.22$), however overall higher COI OTU number was observed in the fusion library ($W = 2338$, $p < 0.001$). We observed no difference between the two markers ($W = 19429$, p-value = 0.2707) in relative mean spread of data clouds in samples grouped by origin between the two markers (average distance to group centroids were 53.89 ± SD 12.93 and 52.98 ± SD 15.14 for 18S rRNA and COI respectively).

Three laboratories reported contamination in their negative control samples, while sequencing data analysis revealed substantial contamination in DNA extraction blanks in six dual index datasets (Figure 4). Species used as positive controls were successfully retrieved from 17 out of 24 datasets (i.e. 70.8%). Failed positive controls affected three dual index and four fusion samples (Figure 4). None or negligible contamination was detected in the negative (no template) PCR controls (except for laboratory D, where >10,000 sequences of the same OTU found in extraction blanks were detected).

Biodiversity clustering according to sample origin – nuclear small subunit (18S rRNA gene) dataset

All factors considered in the PERMANOVA design (sample origin, library type and laboratory), as well as their interactions, had a significant effect on the partitioning variation in 18S rRNA gene datasets ($p<0.05$, Table S2). A test of homogeneity for the within-group multivariate dispersion (PERMDISP) revealed significant differences in the average distances to centroids among samples grouped by sample origin ($p = 0.001$, $F = 18.869$, PERMDISP).

The multiple linear regression followed by an automated stepwise selection of optimal model showed that 'preservation buffer' ($p<0.001$), type of DNA polymerase used in PCR reactions ($p<0.001$), addition of PCR enhancers ($p<0.001$) and 'multiple defrosting of DNA samples' ($p<0.001$) all had significant effects on the dispersion around the origin group centroid (Table S3). The best selected model explained 46% of variation in distance values. In dual index library samples, distances to group centroids showed weak ($r = -0.35$ and -0.34) but statistically significant ($p<0.001$) negative correlations with the sequence quality scores (forward and reverse reads respectively, Figure S3). No such correlations were detected in fusion library samples ($p>0.5$; $r = 0.11$ and -0.12, forward and reverse reads respectively, Figure S3).

Canonical analysis of principal coordinates (CAP) models derived from either the raw 18S rRNA dataset or datasets with post-hoc corrections were all characterized by high and significant canonical correlations (Table 2, Figure 4), and correctly allocated 88% of samples by their origin in the uncorrected

data and up to 93% after applying post-hoc corrections. This is considerably better than the 25% success expected by chance if samples were randomly allocated into four groups. Overall, the correct classification of samples by their geographic origin was improved most by removing samples with extremely low number of sequences (<1,000, Corr1, Table 2, Figure 4). Other corrections applied to the dataset had variable success in improving classification. The AUS samples were classified most accurately after rarefying sequencing data to equal depth across all samples (Corr2 and Corr6, Table 2). The NZ samples showed the lowest classification success overall, not exceeding 87.5% correct allocation (Corr1 and Corr6, Table 2).

Biodiversity clustering according to the geographic origin - mitochondrial Cytochrome c Oxidase subunit 1 dataset

All factors, including the interaction between origin and laboratory had significant effects ($p<0.01$, Table S4) on the partitioning variation in the COI datasets. Test of homogeneity for the within-group multivariate dispersion (PERMDISP) revealed significant differences in distances to centroids among the samples grouped by origin ($p = 0.01$, $F = 4.569$, PERMDISP).

Multiple linear regression followed by an automated stepwise selection of optimal models showed a significant effect of the amount of DNA template and the type of DNA polymerase used in PCR reactions ($p<0.001$), 'preservation buffer' ($p<0.001$), and addition of PCR enhancers ($p<0.001$) on the HTS data spread around group centroids (Table S5). The best selected model explained 58% of variation in distance values. Statistically significant ($p<0.001$) but moderate negative correlations between distances to group centroids and sequence quality scores were detected in dual index samples ($r = -0.64$ and -0.68, forward and reverse reads respectively, Figure S4). No such correlations were detected in fusion library samples ($p>0.5$; $r = 0.06$ and -0.16, forward and reverse reads respectively, Figure S4).

Canonical analysis of principal coordinates (CAP) analyses of COI datasets yielded high and significant canonical correlations (Table 3, Figure 5), and correctly allocated 88% of samples by their origin in the

raw datasets and up to 95% after applying post-hoc corrections (Table 3). The largest improvement in classification was after corrections for contamination effects (Corr3, Corr4 and Corr5, Table 3). New Zealand samples showed the lowest classification success, except for Corr3, when all OTUs found in negative controls were removed.

Variability in community biodiversity of the corrected datasets within the sample geographic origin groups

A more detailed exploration of biodiversity patterns following corrections that resulted in the most improved sample classification by origin (Corr1 and Corr3 for 18S rRNA and COI datasets, correspondingly) showed some variability within each location (Figures 6 and 7). It revealed variance between dual index and fusion library datasets, which was more apparent in 18S rRNA data. However, the first PCoA axis that separated the two library types explained only 19-24% and 10-24% of the total variance, in 18S rRNA and COI datasets, respectively.

Most samples clustered reasonably well by laboratory, however this clustering differed across datasets indicating stochasticity between sample origin groups. Some samples (e.g., fusion library samples from laboratories H and I) formed rather consistent outlying groups, which affected dispersion between fusion and dual index groups. This was particularly evident in the 18S rRNA dataset. Samples from these laboratories were characterized by comparatively high average sequence read quality scores (Lab H: 30.88 ± SD 2.32 and 34.84 ± SD 2.40; Lab I: 33.95 ± SD 1.27 and 35.07 ± SD 1.79, for 18S rRNA and COI sequences respectively), good recovery of target taxa in the positive control samples (except for 18S rRNA Lab H sample) and they did not encounter any contamination issues (Figure 4). In terms of the sample processing workflows, the only distinct difference in Labs H and I from other laboratories following fusion library protocol, was the addition of bovine serum albumin (BSA) at the PCR step (see Table S1).

Discussion

The increased use of environmental DNA (eDNA) metabarcoding across freshwater, marine, and terrestrial ecosystems is a result of advances in HTS technologies and their increasing affordability. Almost anyone and anywhere, including citizen scientists, can now collect a sample for DNA extraction and send it to a specialized laboratory to examine biodiversity in different biological matrices (Evans et al., 2016; Jeunen et al., 2019; Larson et al., 2020; Vandamme et al., 2016). This offers unprecedented opportunities for integrative analyses and comparisons of biodiversity patterns across wide temporal and spatial scales. However, as our findings clearly illustrate, there is a critical need for consistency and accuracy throughout the laboratory processing and data analysis workflows.

Metabarcoding (like any other biodiversity assessment technique) is prone to taxon identification errors, biases and contaminations that can result in false negatives and positives, and these can occur at any point in the sample processing and analytical steps (Doi et al., 2019; Ficetola et al., 2015). Improving knowledge of potential confounding factors and sources of variation is critical for increasing the credibility and reproducibility of metabarcoding approaches and for enabling large scale comparative studies and long-term monitoring (Baker, 2016). In the absence of standardized sample and analysis processing protocols for HTS metabarcoding, it is often assumed by default that reproducible data can be generated by closely following a credible workflow if it has been published in a peer-reviewed article. However, as demonstrated in the recent review by Nicholson et al. (2020), many studies do not report sufficient details to allow replication and/or appropriate data comparison between studies. It is also not usual practice to report troubleshooting incidents or small adjustments made to the protocols, which are common and evident in the present study. The importance of such variation on the outcome, therefore, remains largely unknown and is likely underestimated.

Our study revealed considerable variation in metabarcoding results from similar biodiversity analyses undertaken in different laboratories. However, the crucial questions addressed here were whether this methodological variation obscured the true biological signal, what factors introduced the greatest

variance, and whether the variance could be effectively reduced via appropriate data filtering. The raw data from all laboratories showed strong and consistent differentiation of biological samples from different geographical regions, for both the 18S rRNA and COI genes. We acknowledge, however, that the consistency in community composition clustering was assessed here at a broad geographical scale, and the effect of derived biases may be more significant when depicting more subtle (i.e. regional) biodiversity patterns.

Patterns in variability of biological community information derived from both markers and both library types

Our results showed no significant difference between 18S rRNA and COI data spread around geographic origin groups. This suggests that neither marker is more susceptible to differences in laboratory processing protocols and associated data biases. The 18S rRNA marker is more conserved than the COI, which typically results in comparatively lower species resolution and underestimation of species richness (Leray & Knowlton, 2016; Pearman et al., 2021; von Ammon et al., 2018). This may explain our observation of slightly stronger clustering of 18S rRNA data points according to the sample origin and substantial improvement of classification of 18S rRNA dataset following the removal of low-quality samples (i.e., those with only a small number of resulting sequence reads). This was in contrast to the COI data, which showed the best improvement in classification after removal of contaminating sequences. The reduced phylogenetic resolution in 18S rRNA data potentially makes it more resilient to contamination.

Despite a significant effect of library type on variance partitioning in samples detected by PERMANOVA (contributed only 2 and 13% to variation in 18S rRNA and COI samples, respectively) neither regression nor CAP analyses showed significant differentiation of samples between fusion and dual index libraries. However, an interesting interaction between library type and marker gene was observed in the numbers of sequences: the 18S rRNA yielded more sequence reads using dual index compared with the fusion method, and there was an opposite trend in the COI data.

Unlike dual index libraries, where amplicons are tagged with short nucleotide sequences using commercial kits at the additional PCR step (Bourlat, Haenel, Finnman, & Leray, 2016), in the fusion primers approach a target DNA fragment is simultaneously amplified and tagged with a long nucleotide tail (Zizka, Elbrecht, Macher, & Leese, 2019). The one-step tagging of amplicons from multiple samples with fusion primers was developed to improve cost- and time-efficiency of sample processing for metabarcoding (Elbrecht & Leese, 2015; Elbrecht & Steinke, 2018; Stat et al., 2017; Zizka et al., 2019). However, it may potentially introduce substantial biases into inferred biodiversity due to tag-specific mismatches with the PCR template and subsequent variation in primer-binding efficiency between taxonomic groups (O'Donnell, Kelly, Lowell, & Port, 2016). The biases of fusion primers reported however, refer to variation in sequence abundance across replicates and not in the detection of taxa, which is comparable between methods (O'Donnell et al., 2016; Zizka et al., 2019). Interestingly, in our study we found a higher richness of OTUs for the COI gene when using the fusion approach over the two-step approach, which contrasts to what one would expect if PCR inefficiencies resulted from longer primers. Regardless, the two library building methods produced similar community profiles. However, the fact that no effect of the average Phred scores on data dispersal was observed for both markers in the fusion library (Figs. S3 and S4), suggests that stochastic biases not directly related to the quality of sequences may be in play when applying the one-step PCR approach. It is also presumed that the dual indexing protocol is less susceptible to inhibitors, resulting in improved amplification of complex samples, compared to the one-step fusion approach (Zizka et al., 2019). While we did not expect varying levels of inhibiting substances between two markers (they were run on the same set of biofouling samples), longer (~410 bp) 18S rRNA amplicons might have been more prone to inhibition (McCord, Pionzio, & Thompson, 2014; Opel, Chung, & McCord, 2010). This likely explains overall lower number of sequences yielded for 18S rRNA with a higher number of sequences in the dual index library, as well as the best data consistency improvement following removal of 23 samples with extremely low sequence numbers (<1,000). The differential inhibition effect might also explain the outliers in the corrected 18S rRNA fusion dataset. These outliers came from samples where PCR was undertaken with addition of BSA, a protein known to

449 relieve amplification inhibition (Kreader, 1996). This observation suggests that the effect of BSA on
450 metabarcoding results and especially obtaining true diversity needs further exploration, preferably in
451 dedicated controlled experiments.

452 In contrast, the overall number of COI sequences was consistently higher than 18S rRNA
453 sequences under both dual index (~1.5 times higher on average) and, particularly, fusion (~25 times
454 higher on average) approaches. Because the 18S rRNA and COI gene fragments were pooled and
455 sequenced simultaneously, the most parsimonious explanation is that the shorter gene (~310 bp) COI
456 amplicons were subjected to preferential selection on the Illumina flow cell (Engelbrektson et al., 2010).
457 Future studies will be required to test this hypothesis further, as well as potential implications of gene
458 multiplexing on the derived biodiversity information.

459 Nevertheless, a clear advantage of the one-step fusion library approach (besides its cost/time-
460 efficiency) is that it is less prone to cross-contamination (Zizka et al., 2019). This was observed in our
461 study results, where substantial laboratory contamination was detected in the dual-index library samples,
462 with a significant effect on data clustering. This was especially evident in the COI dataset, where the best
463 classification results were achieved following corrections for contamination effects. The lowest
464 classification success was observed in the NZ COI dataset, and was also likely due to the dominance of
465 contaminating OTUs in the NZ sample from laboratory F (76% of all sequences were removed from their
466 dataset at the post-hoc correction step).

467 In this study we considered the two sequence library preparation methods most employed by
468 participating laboratories. Another common workflow not included here is the 'tagged PCR' approach,
469 which includes PCR amplifications with relatively short metabarcoding primers carrying 5' nucleotide
470 tags and subsequent ligation of adaptors (Binladen et al., 2007; Carøe & Bohmann, 2020). This method
471 has been recently touted as an effective alternative for metabarcoding library preparation, especially when
472 tag-jump free protocols are implemented to prevent false sequence assignment to samples and mitigate
473 chimera formation (Carøe & Bohmann, 2020). Additional comparative analysis is required to investigate
474 the robustness of this approach in the context of laboratory-induced variations.

The effect of sample handling and processing factors on variability in metabarcoding data

Analysis of the relative importance of considered random effects in introducing variability into the metabarcoding results showed that both 18S rRNA and COI datasets were most affected by the type of preservation buffer, type of DNA polymerase and the addition of PCR enhancers. Previous studies have shown that the choice of sample preservation method can have a substantial effect on the integrity of eDNA and may alter HTS-derived community structure to a greater or lesser degree (Gray, Pratte, & Kellogg, 2013; Lee, Adams, & Klassen, 2019; Tatangelo, Franzetti, Gandolfi, Bestetti, & Ambrosini, 2014). Therefore, although not always practical, immediate processing or cold storage (ideally at -20°C or lower temperatures) remains the gold standard for retaining utmost quality of genetic material in samples and reducing risks of preservation-related biases in biodiversity assessment (Lee et al., 2019; Renshaw, Olds, Jerde, McVeigh, & Lodge, 2015). However, on-site freezing is not always possible, and the use of DNA/RNA isolation buffers remains a popular alternative for samples collected in remote locations or when samples are to be shipped internationally. In the present study, we chose two commonly used commercial solutions, the RNAlater™ and LifeGuard™ isolation buffers for stabilizing the biofouling samples. The main criteria for this choice were i) the compatibility with the DNA extraction protocol, ii) the effective DNA stabilization capacity over a range of temperatures, and iii) no restrictive chemicals in relation to shipping. Since a number of studies have previously shown no substantial differences in DNA preservation efficacy between the two buffers (Gomez-Silvan et al., 2018) and they are referred to as equivalent in standardized genetic and genomic protocols (Duran & Cravo-Laureau, 2017; Hampton-Marcell, Frazier, Moormann, Owens, & Gilbert, 2017), we allowed sample providers to select either. The somewhat higher variance in metabarcoding data from the LifeGuard™ samples was largely determined by the presence of a few outliers. Taking into account the complexity of the sampled matrices (von Ammon et al., 2018), the observed effect of preservation solution could be related to sample-specific biases driven by the presence and concentration of inhibiting substances, or the prevalence of hard-shelled organisms reducing homogenization of the material. Therefore, we cannot confidently infer differential

501 performance of the two buffers for sample preservation based on the current results. These results
502 highlight the need to ensure that variables such as preservation buffer type, temperature and storage time
503 should be recorded and reported in the metabarcoding protocols and considered when analyzing data from
504 multiple sample sets, from different laboratories or collected over different timeframes.
505 DNA polymerases with proofreading activity (e.g., MyFi™ DNA Polymerase) may enhance the
506 formation of chimeric sequences compared to Taq DNA polymerases (e.g., AmpliTaq Gold™ DNA
507 Polymerase) due to the earlier occurrence of PCR saturation (Ahn, Kim, Song, & Weon, 2012; Judo,
508 Wedel, & Wilson, 1998). Among the six distinct DNA polymerases used in the present study, MyFi was
509 the only proofreading polymerase and it was used by five of the participating labs (C, G, H, K, and L).
510 We did not find any clear evidence showing that the observed variability related to 'polymerase' effect
511 was specifically linked to the use of a proofreading polymerase. Therefore, a more parsimonious
512 explanation is that the random effect was driven by somewhat differential performance of the five Taq
513 DNA polymerases used.
514 The PCR enhancer Bovine Serum Albumin (BSA) was added by four laboratories (A, E, H and I)
515 and appeared to have an impact on the derived metabarcoding outputs. As noted above, this is possibly
516 due to its ability to bind to inhibitory substances during PCR reactions, preventing inhibitory interactions
517 with DNA polymerase (Woide, Zink, & Thalhammer, 2010). Biofilms and biofouling material contain
518 many known PCR inhibitors, such as glycogen, polysaccharides and slat (Schrader, Schielke, Ellerbroek,
519 & Johne, 2012). Therefore, a certain level of inhibition was reasonably expected in our samples. The
520 results from the analyses suggest that the addition of BSA might have substantial effect on both variance
521 and composition of the detected communities (as evidenced by the outliers from corrected datasets of
522 Labs H and I). Because this study was not designed to fully investigate the impact of BSA and only a few
523 laboratories included enhancers, further in-depth studies are required to understand better the patterns of
524 BSA effects on derived biodiversity.
525 The 18S rRNA data also showed sensitivity to multiple freeze-thaw cycles of the samples. The
526 available (mostly anecdotal) evidence of the effect of multiple defrosting on DNA integrity, suggests that

even one cycle of freeze-thawing can reduce eDNA signal (Bowers et al., 2021). However, it is rarely the case that metabarcoding library construction is achieved without repetition of the PCR step (and thus multiple freeze-thawing of DNA material) for at least some samples (e.g., troubleshooting and adjusting amplification conditions for problematic samples). This parameter is difficult to control for, and it is rarely (if at all) reported in metabarcoding protocols. It would be practical, however, to keep track of and report freeze-thaw cycles of DNA samples, reduce their number where possible and, most importantly, divide DNA into multiple aliquots upon extraction.

The dispersion of the COI data from different laboratories was also affected by the amount of template DNA used in PCR (unlike 18S rRNA data). Low quantities of template DNA and stochasticity in early PCR cycles is known to affect reproducibility of metabarcoding results (Alberdi, Aizpurua, Gilbert, & Bohmann, 2018; Leray & Knowlton, 2017). This effect is particularly inherent for primers with reduced specificity (e.g., COI primers comprising a few degenerate bases to allow better matching the mutationally saturated target regions) applied to highly diverse environmental samples (e.g., marine samples), and can be reduced by performing multiple PCR technical replicates (Collins et al., 2019; Ficetola et al., 2015). Technical replication was advised in our suggested sample processing protocols, however, it was only followed by five laboratories: A, C (dual-index group) and G, H and L (fusion group). There is still lack of agreement in the literature whether technical replication is necessary for improved capture of "true" biodiversity (Marotz et al., 2019), especially considering the associated substantial increase in time and cost effort. Technical replication might be advised for studies where true (environmental) replication is restricted for some reasons, to mitigate the effect of within sample variability.

Effect of the post-hoc data correction on the observed variability in community composition

The datasets we dealt with in our analyses were somewhat atypical. We purposefully included all investigated sample sets, including those that amplified poorly, had low numbers of sequence reads, and

552 controls with some contamination, which would have usually been excluded in a typical study. These
553 "problematic" samples most likely would have been dropped either at the library preparation or
554 bioinformatic analysis step if standard quality control procedures were applied. In our study we aimed to
555 evaluate the widest possible range of issues encountered during metabarcoding analysis and test the effect
556 of different sources of variation (including extreme ones) on the final community data. The post-hoc
557 corrections applied to the datasets represent the usual quality checkpoints during sample processing and
558 initial data screening. For example, contaminated samples or those with low sequence yield would usually
559 be removed, as well as samples with extremely low sequence number or OTU diversity. These simple
560 corrections help to remove the noise from the dataset, introduced by the rare tail (i.e., singletons and
561 doubletons, that are in most cases amplification, sequencing or contamination artefacts), while conserving
562 and emphasizing the core patterns in the communities.

563

564 **Conclusions**

565 The results of this study suggest that, overall, community metabarcoding is relatively robust to
566 the random effects of laboratory-based variation within established sample processing protocols and
567 largely confirm our initial hypotheses: (1) Metabarcoding results from all laboratories provided consistent
568 patterns of discrimination among four community samples for both DNA markers (COI and 18S rRNA),
569 despite considerable variation; (2) The two different HTS library protocols (dual index primers *versus*
570 fusion primers) did not significantly affect the community comparisons, even though they appeared to
571 have some differential effects on numbers and quality of sequence reads from the two markers; (3) We
572 identified several factors that introduced the greatest variability (preservation buffer, sample defrosting,
573 template concentration, DNA polymerase, PCR enhancer [BSA]); (4) We confirmed that standard post-
574 hoc data filtering steps (e.g., excluding samples with low sequence number or eliminating contaminating
575 sequences) were very effective at removing noise in the metabarcoding data introduced by laboratory

variation. The main findings and further considerations resulting from our study are summarized in the provisional guidelines shown below (Text box 1).

Text box 1: Considerations for improved standardization and minimising effects of laboratory-based variation in metabarcoding community analyses.

Primer choice

18S rRNA: better detection of large-scale patterns, less sensitive to contamination
COI: better species resolution, higher sequence yield, less sensitive to freeze-thaw cycles

Library type

Dual index: higher sequence yield for the 18S rRNA gene, less susceptible to inhibitors,
Fusion: higher sequence yield for the COI gene, less prone to contamination

Sample handling and processing factors substantially affecting metabarcoding data variability

- Preservation buffer
- Type of DNA polymerase
- PCR enhancers

General recommendations

- Immediate processing or cold storage (ideally at -20°C or lower temperatures), whenever practical
- Preservation buffer type, temperature and storage time should be recorded and reported in the metabarcoding protocols and considered as covariates, whenever relevant
- Keep track of and report freeze-thaw cycles of DNA samples (avoid if possible)
- Divide DNA into multiple aliquots upon extraction
- Caution around using BSA and other PCR inhibitors, as these might affect variance and composition of investigated communities
- Technical replication at the PCR step is recommended to mitigate effects of within sample variability
- Consider appropriate data correction and de-noising for reducing effects from sequencing or contamination artefacts and better discerning core biological patterns
- Keep track and report all data corrections

Further considerations

- The effect of slightly different workflows may be more significant when discriminating assemblages with moderate overlap but some turnover (β diversity) in species assemblages. Therefore, bias rate imposed by methodological variation deserves further investigation at fine-scale levels for species diversity estimates.
- Due to large variation of sample types, primers, target taxa, differential laboratory access to reagents and many other factors, it is impractical to completely standardize PCR protocols across all metabarcoding studies. However, the maximum possible standardization is required in studies aimed at comparative biodiversity analyses.
- It is critical to clearly articulate methods in publications to enable a better understanding of the underlining causes of biological deviations or lack of reproducibility between studies.
- It is impossible to control the effects of all possible laboratory parameters in one study, therefore further comparative and cross-calibrations studies, at both large and small spatial scales, should be of great assistance for further identifying the key factors introducing biological variation and for improving standardization of metabarcoding protocols.

Acknowledgements

The experiment was facilitated by in-kind support from two biotech companies (Qiagen and Illumina) who donated free extraction and sequencing kits. This work is a result of the Quadrilateral Scientific Collaboration in Marine Biosecurity, a partnership funded by the New Zealand Ministry of Business, Innovation and Employment (Contract C01X1527) to foster cooperation in marine biosecurity research among New Zealand, Australia, Canada and the USA. Contribution of the Cawthron institute team was also supported by New Zealand Ministry of Business, Innovation and Employment funding (CAWX1904 – A toolbox to underpin and enable tomorrow's marine biosecurity system). GJI and JS were supported by the NZ Government's Strategic Science Investment Fund (SSIF) through NIWA Coasts & Oceans Programme 6 - 2019/20 SCI.

Author Contributions

AZ, PG, CA, UvA, MB, MEC, JG, AAG, MH, GJI, SL, AS, TS, MS, JS, SAW and XP conceived the study. AZ, CA, UvA, JG, MS, SAW and XP defined the design and collected the field samples. AZ, CA, UvA, JB, MB, MEC, AC, ED, JG, AAG, MH, EH, SL, AS, TS, MS, SS, JS, VT, KW, MW and GZ performed sample processing and library preparation. MB, MH, TS, MS and MW run the sequencing. PG run bioinformatic analyses on the datasets. AZ performed statistical analyses and produced the first draft of the manuscript. AZ, PG, CA, MEC, AC, GJI, SL, TS, SAW and XP revised the early versions of the manuscript and contributed to writing. All authors contributed to revision of later versions and final proof-reading.

References

Ahn, J. H., Kim, B. Y., Song, J., & Weon, H. Y. (2012). Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *J Microbiol, 50*(6), 1071-1074. doi:10.1007/s12275-012-2642-z

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *9*(1), 134-147. doi:10.1111/2041-210x.12849

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology, 26*, 32-46.

Anderson, M. J. (2017). Permutational Multivariate Analysis of Variance (PERMANOVA). In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, & J. L. Teugels (Eds.), *Wiley StatsRef: Statistics Reference Online* (pp. 1-15).

Aylagas, E., Borja, Á., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment. *Frontiers in Marine Science, 3*, Article 96. doi:10.3389/fmars.2016.00096

Aylagas, E., Borja, A., & Rodrigues-Ezpeleta, N. (2014). Environmental status assessment using DNA metabarcoding: towards a genetic based marine biotic index (gAMBI). *PLoS One, 9*(3), e90529. doi:90510.91371/journal.pone.0090529.

Bailet, B., Apothéloz-Perret-Gentil, L., Baričević, A., Chonova, T., Franc, A., Frigerio, J.-M., . . . Kahlert, M. (2020). Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. *Science of The Total Environment, 745*, 140948. doi:https://doi.org/10.1016/j.scitotenv.2020.140948

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*, 452-454.

Binladen, J., Gilbert, M. T., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., & Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One, 2*(2), e197. doi:10.1371/journal.pone.0000197

Bourlat, S. J., Haenel, Q., Finnman, J., & Leray, M. (2016). Preparation of Amplicon Libraries for Metabarcoding of Marine Eukaryotes Using Illumina MiSeq: The Dual-PCR Method. In S. J. Bourlat (Ed.), *Marine Genomics: Methods and Protocols* (pp. 197-207). New York, NY: Springer New York.

Bowers, H., Pochon, X., von Ammon, U., Gemmell, N. J., Stanton, J.-A., Jeunen, G.-J., . . . Zaiko, A. (2021). Towards optimization of eDNA/eRNA sampling technologies for marine biosecurity surveillance. *Water, 13*(8), 1113.

Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Elbrecht, V., Steinke, D., Ratnasingham, S., . . . Hebert, P. D. N. (2019). Metabarcoding a diverse arthropod mock community. *19*(3), 711-727. doi:10.1111/1755-0998.13008

Brown, E. A., Chain, F. J. J., Zhan, A., MacIsaac, H. J., & Cristescu, M. E. (2016). Early detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports. *Diversity and Distributions, 22*, 1045-1059.

Carøe, C., & Bohmann, K. (2020). Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples. *20*(6), 1620-1631. doi:https://doi.org/10.1111/1755-0998.13227

Clarke, K. R., & Gorley, R. N. (2015). *PRIMER v7: User Manual/Tutorial*. PRIMER-E, Plymouth, UK.

Clarke, L. J., Trebilco, R., Walters, A., Polanowski, A. M., & Deagle, B. E. (2020). DNA-based diet analysis of mesopelagic fish from the southern Kerguelen Axis. *Deep Sea Research Part II: Topical Studies in Oceanography, 174*. doi:https://doi.org/10.1016/j.dsr2.2018.09.001

Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., . . . Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. *10*(11), 1985-2001. doi:10.1111/2041-210x.13276

Comtet, T., Sandionigi, A., Viard, F., & Casiraghi, M. (2015). DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions, 17*, 905-822.

Cordier, T., Alonso-Sáez, L., Apothéloz-Perret-Gentil, L., Aylagas, E., Bohan, D. A., Bouchez, A., . . . Lanzén, A. (2020). Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. doi:10.1111/mec.15472

Cristescu, M. E. (2014). From barcoding single individuals metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution, 29*(10), 566-571.

Darling, J. A., & Mahon, A. R. (2011). From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental Research, 111*, 978-988.

Darling, J. A., Pochon, X., Abbott, C. L., Inglis, G. J., & Zaiko, A. (2020). The risks of using molecular biodiversity data for incidental detection of species of concern. *Diversity and Distributions*, in press.

Djurhuus, A., Port, J., Closek, C. J., Yamahara, K. M., Romero-Maraccini, O., Walz, K. R., . . . Chavez, F. P. (2017). Evaluation of Filtration and DNA Extraction Methods for Environmental DNA Biodiversity Assessments across Multiple Trophic Levels. *4*(314). doi:10.3389/fmars.2017.00314

Doi, H., Fukaya, K., Oka, S.-i., Sato, K., Kondoh, M., & Miya, M. (2019). Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model. *Scientific Reports, 9*(1), 3581. doi:10.1038/s41598-019-40233-1

Dowle, E., Pochon, X., Banks, J., Shearer, K., & Wood, S. A. (2016). Targeted gene enrichment and high throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. *Molecular Ecology Resources, 16*(5), 1240-1254.

Duran, R., & Cravo-Laureau, C. (2017). Protocols for Mudflat and Algal Mat In Situ Analysis. In T. J. McGenity, K. N. Timmis, & B. Nogales (Eds.), *Hydrocarbon and Lipid Microbiology Protocols: Field Studies* (pp. 305-317). Berlin, Heidelberg: Springer Berlin Heidelberg.

Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. 081257. doi:10.1101/081257 %J bioRxiv

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics, 27*, 2194-2200.

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass - sequence relationships with an innovative metabarcoding protocol. *PLoS One, 10*(7), e0130324. doi:10.1371/journal.pone.0130324

Elbrecht, V., & Steinke, D. (2018). Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. *Freshwater Biology, 64*(2), 380-387.

Engelbrektson, A., Kunin, V., Wrighton, K. C., Zvenigorodsky, N., Chen, F., Ochman, H., & Hugenholtz, P. (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *International Society for Microbial Ecology Journal, 4*(5), 642-647.

Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., . . . Lodge, D. M. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources, 16*(1), 29-41.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., . . . Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources, 15*(3), 543-556.

696 Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial
697 cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic
698 surveys. *Molecular Ecology, 13*(5), 851-861.
699 Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thosmsen, F. P., Murhpy, M. A., . . . Taberlet, P.
700 (2016). Critical considerations for the application of environmental DNA methods to detect
701 aquatic species. *Methods in Ecology and Evolution, 7*, 1299-1307.
702 Gomez-Silvan, C., Leung, M. H. Y., Grue, K. A., Kaur, R., Tong, X., Lee, P. K. H., & Andersen, G. L. (2018). A
703 comparison of methods used to unveil the genetic and metabolic pool in the built environment.
704 *Microbiome, 6*(1), 71. doi:10.1186/s40168-018-0453-0
705 Gray, M. A., Pratte, Z. A., & Kellogg, C. A. (2013). Comparison of DNA preservation methods for
706 environmental bacterial community samples. *FEMS Microbiol Ecol, 83*(2), 468-477.
707 doi:10.1111/1574-6941.12008
708 Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A., & Baird, D. J. (2011). Environmental barcoding: a
709 next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS*
710 *One, 6*(4), e17497.
711 Hampton-Marcell, J. T., Frazier, A., Moormann, S. M., Owens, S. M., & Gilbert, J. A. (2017). Preparation
712 and Analysis of Metatranscriptomic Libraries in Petroleum Hydrocarbon Microbe Systems. In T.
713 J. McGenity, K. N. Timmis, & B. Nogales (Eds.), *Hydrocarbon and Lipid Microbiology Protocols:*
714 *Genetic, Genomic and System Analyses of Communities* (pp. 51-67). Berlin, Heidelberg: Springer
715 Berlin Heidelberg.
716 Jeunen, G.-J., Knapp, M., Spencer, H. G., Taylor, H. R., Lamare, M. D., Stat, M., . . . Gemmell, N. J. (2019).
717 Species-level biodiversity assessment using marine environmental DNA metabarcoding requires
718 protocol optimization and standardization. *9*(3), 1323-1335. doi:10.1002/ece3.4843
719 Judo, M. S., Wedel, A. B., & Wilson, C. (1998). Stimulation and suppression of PCR-mediated
720 recombination. *Nucleic Acids Research, 26*(7), 1819-1825. doi:10.1093/nar/26.7.1819
721 Keeley, N. B., Wood, S. A., & Pochon, X. (2018). Development and preliminary validation of a multi-
722 trophic metabarcoding biotic index for monitoring benthic organic enrichment. *Ecological*
723 *Indicators, 85*, 1044-1057.
724 Kreader, C. A. (1996). Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32
725 protein. *62*(3), 1102-1106.
726 Lacoursière-Roussel, A., Howland, K., Normandeau, E., Grey, E. K., Archambault, P., Deiner, K., . . .
727 Bernatchez, L. (2018). eDNA metabarcoding as a new surveillance approach for coastal Arctic
728 biodiversity. *8*(16), 7763-7777. doi:10.1002/ece3.4213
729 Larson, E. R., Graham, B. M., Achury, R., Coon, J. J., Daniels, M. K., Gambrell, D. K., . . . Suarez, A. V.
730 (2020). From eDNA to citizen science: emerging tools for the early detection of invasive species.
731 *18*(4), 194-202. doi:10.1002/fee.2162
732 Lee, K. M., Adams, M., & Klassen, J. L. (2019). Evaluation of DESS as a storage medium for microbial
733 community analysis. *PeerJ, 7*, e6414. doi:10.7717/peerj.6414
734 Lehtiniemi, M., Ojaveer, H., David, M., Galil, B., Gollasch, S., McKenzie, C., . . . Pederson, J. (2015). Dose
735 of truth - Monitoring marine non-indigenous species to serve legislative requirements. *Marine*
736 *Policy, 54*, 26-35.
737 Leray, M., & Knowlton, N. (2016). Censusing marine eukaryotic diversity in the twenty-first century.
738 *Philos Trans R Soc Lond B Biol Sci, 371*(1702), 20150331. doi:doi:10.1098/rstb.2015.0331
739 Leray, M., & Knowlton, N. (2017). Random sampling causes the low reproducibility of rare eukaryotic
740 OTUs in Illumina COI metabarcoding. *PeerJ, 5*(e3006).
741 Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., . . . Machida, R. J. (2013). A new
742 versatile primer set targeting a short fragment of the mitochondrial COI region for

743 metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents.
744 *Frontiers in Zoology, 10*(1), 34. doi:10.1186/1742-9994-10-34
745 Marotz, C., Sharma, A., Humphrey, G., Gottel, N., Daum, C., Gilbert, J. A., . . . Knight, R. (2019). Triplicate
746 PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary. *67*(1), 29-32.
747 doi:10.2144/btn-2018-0192
748 McCord, B., Pionzio, A., & Thompson, R. (2014). *Analysis of the effect of a variety of PCR inhibitors on the*
749 *amplification of DNA using real time PCR, melt curves and STR analysis*. Retrieved from
750 https://nij.ojp.gov/library/publications/analysis-effect-variety-pcr-inhibitors-amplification-dna-
751 using-real-time-pcr
752 McMurdie, P. J., & Holmes, S. (2013). Phyloseq: an R package for reproducibleinteractive analysis and
753 graphics of microbiome census data. *PLoS One, 8*, e61217.
754 Molnar, J. L., Gamboa, R. L., Revenga, C., & Spalding, M. D. (2008). Assessing the global threat of invasive
755 species to marine biodiversity. *Frontiers in Ecology and the Environment, 6*(9), 485-492.
756 Murray, D. C., Coghlan, M. L., & Bunce, M. (2015). From Benchtop to Desktop: Important Considerations
757 when Designing Amplicon Sequencing Workflows. *PLoS One, 10*(4), e0124671.
758 doi:10.1371/journal.pone.0124671
759 Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., . . . Shapiro, B.
760 (2018). Minimizing polymerase biases in metabarcoding. *18*(5), 927-939. doi:10.1111/1755-
761 0998.12895
762 Nicholson, A., McIsaac, D., MacDonald, C., Gec, P., Mason, B. E., Rein, W., . . . Hanner, R. H. (2020). An
763 analysis of metadata reporting in freshwater environmental DNA research calls for the
764 development of best practice guidelines. *2*(3), 343-349. doi:10.1002/edn3.81
765 O'Donnell, J. L., Kelly, R. P., Lowell, N. C., & Port, J. A. (2016). Indexed PCR Primers Induce Template-
766 Specific Bias in Large-Scale DNA Sequencing Studies. *PLoS One, 11*(3), e0148698.
767 doi:10.1371/journal.pone.0148698
768 Opel, K. L., Chung, D., & McCord, B. R. (2010). A study of PCR inhibition mechanisms using real time PCR.
769 *J Forensic Sci, 55*(1), 25-33. doi:10.1111/j.1556-4029.2009.01245.x
770 Pearman, J. K., von Ammon, U., Laroche, O., Zaiko, A., Wood, S. A., Zubia, M., . . . Pochon, X. (2021).
771 Metabarcoding as a tool to enhance marine surveillance of nonindigenous species in tropical
772 harbors: A case study in Tahiti. *3*(1), 173-189. doi:https://doi.org/10.1002/edn3.154
773 Porter, T. M., & Hajibabaei, M. (2018). Scaling up: A guide to high-throughput genomic approaches for
774 biodiversity analysis. *Mol Ecol, 27*(2), 313-338. doi:10.1111/mec.14478
775 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glockner, F. O. (2013). The SILVA
776 ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic*
777 *Acids Res, 41*(Database issue), D590-596. doi:10.1093/nar/gks1219
778 R-project. (2014). R: A language and environment for statistical computing. Retrieved from
779 http://www.R-project.org
780 Renshaw, M. A., Olds, B. P., Jerde, C. L., McVeigh, M. M., & Lodge, D. M. (2015). The room temperature
781 preservation of filtered environmental DNA samples and assimilation into a phenol-chloroform-
782 isoamyl alcohol DNA extraction. *Mol Ecol Resour, 15*(1), 168-176. doi:10.1111/1755-0998.12281
783 Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors – occurrence, properties and
784 removal. *113*(5), 1014-1026. doi:https://doi.org/10.1111/j.1365-2672.2012.05384.x
785 Simberloff, D., Martin, J.-L., Genovesi, P., Maris, V., Wardle, D., Aronson, J., . . . Vilà, M. (2013). Impacts
786 of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution, 28*(1),
787 58-66. doi:10.1016/j.tree.2012.07.013
788 Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., . . . Bunce, M. (2017).
789 Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine
790 environment. *Scientific Reports, 7*, 12240.

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA: a special issue on DNA metabarcoding. *Molecular Ecology, 21*, 1789-1793.

Tatangelo, V., Franzetti, A., Gandolfi, I., Bestetti, G., & Ambrosini, R. (2014). Effect of preservation method on the assessment of bacterial community structure in soil and water samples. *FEMS Microbiol Lett, 356*(1), 32-38. doi:10.1111/1574-6968.12475

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., . . . Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol Ecol, 25*(4), 929-942. doi:10.1111/mec.13428

van der Loos, L. M., & Nijland, R. (2020). Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology*. doi:10.1111/mec.15592

Vandamme, S. G., Griffiths, A. M., Taylor, S. A., Di Muri, C., Hankard, E. A., Towne, J. A., . . . Mariani, S. (2016). Sushi barcoding in the UK: another kettle of fish. *PeerJ, 4*, e1891. doi:10.7717/peerj.1891

von Ammon, U., Wood, S. A., Laroche, O., Zaiko, A., Tait, L., Lavery, S., . . . Pochon, X. (2018). Combining morpho-taxonomy and metabarcoding enhances the detection of non-indigenous marine pests in biofouling communities. *Scientific Reports, 8*, 16290.

Woide, D., Zink, A., & Thalhammer, S. (2010). Technical Note: PCR Analysis of Minimum Target Amount of Ancient DNA. *American Journal of Physical Anthropology, 142*(2), 321-327.

Wood, S. A., Smith, K. F., Banks, J. C., Tremblay, L. A., Rhodes, L., Mountfort, D., . . . Pochon, X. (2013). Molecular genetic tools for environmental monitoring of New Zealand's aquatic habitats, past, present and the future. *New Zealand Journal of Marine and Freshwater Research, 47*(1), 90-119.

Zaiko, A., Pochon, X., Garcia-Vazquez, E., Olenin, S., & Wood, S. A. (2018). Advantages and limitations of environmental DNA/RNA tools for marine biosecurity: management and surveillance of non-indigenous species. *Frontiers in Marine Science, 5*(322).

Zaiko, A., Samuiloviene, A., Ardura, A., & Garcia-Vazquez, E. (2015). Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. *Marine Pollution Bulletin, 100*(1), 53-59.

Zhan, A., Hulak, M., Sylvester, F., Huang, X., Adebayo, A. A., Abbott, C., . . . MacIsaac, H. J. (2013). High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution, 4*(6), 558-565.

Zizka, V. M. A., Elbrecht, V., Macher, J.-N., & Leese, F. (2019). Assessing the influence of sample tagging and library preparation on DNA metabarcoding. *19*(4), 893-899. doi:https://doi.org/10.1111/1755-0998.13018

Data accessibility statement

The HTS data supporting the results presented in this manuscript were uploaded into CSIRO Data Access Portal and is accessible online at https://doi.org/10.25919/5j5x-0711.

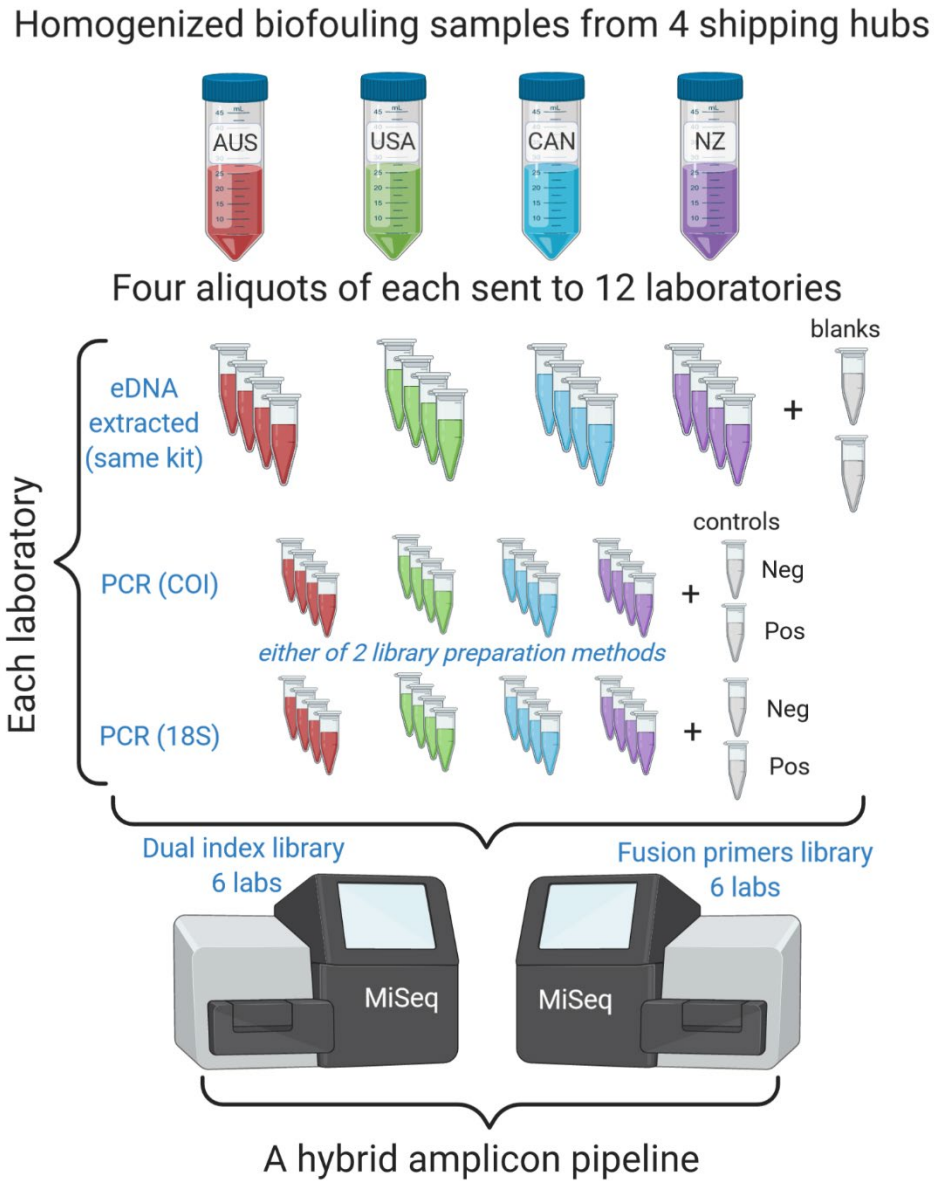


Figure 1. Conceptual scheme of the experimental workflow (created with BioRender.com). Countries from where the samples were sourced: AUS = Australia, USA = United States of America, CAN = Canada, NZ = New Zealand. 18S = 18S ribosomal rRNA gene, COI = mitochondrial Cytochrome c Oxidase subunit 1.

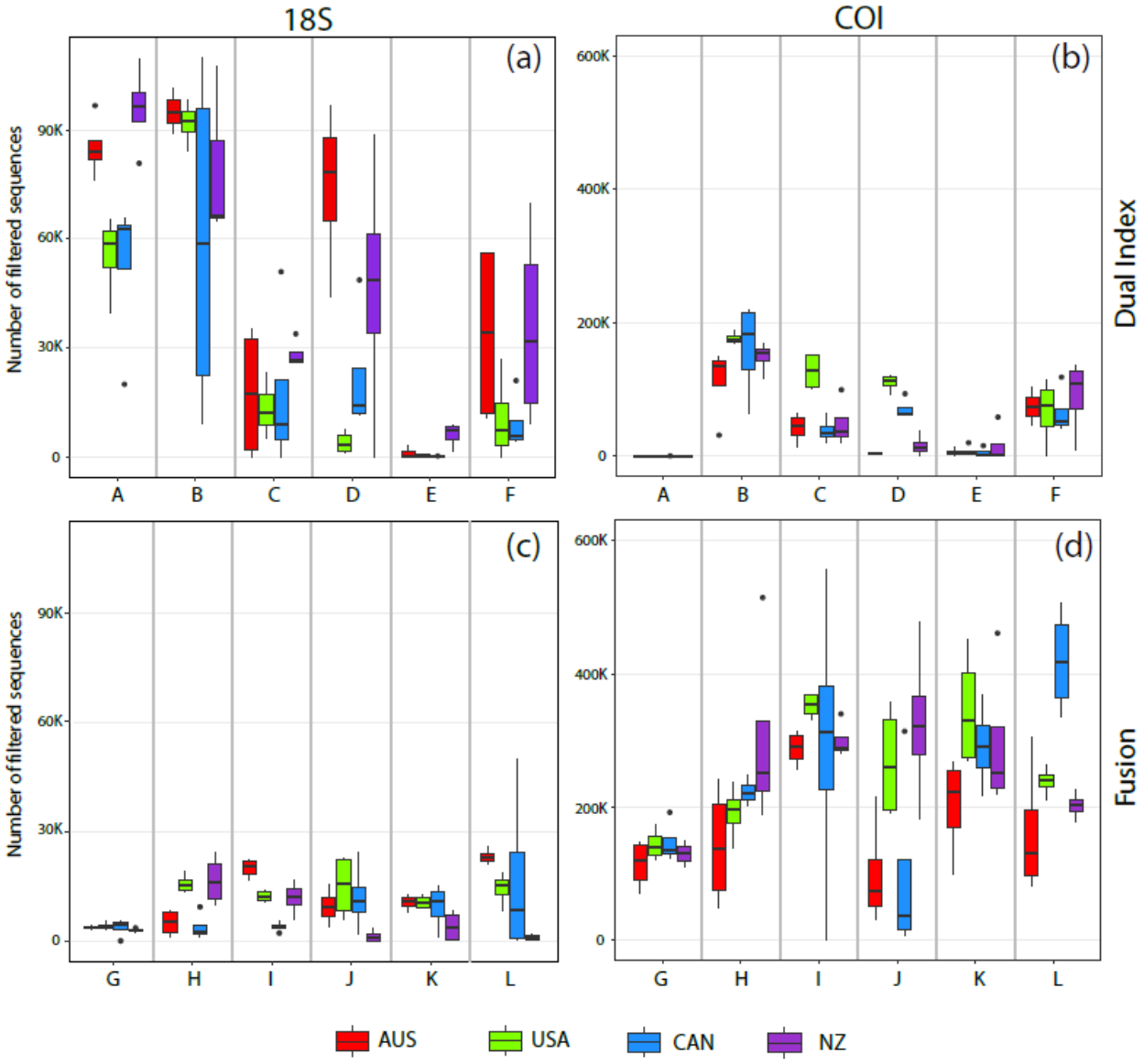


Figure 2. Number of filtered 18S rRNA gene (a, c) and Cytochrome c Oxidase subunit 1 (b, d) sequences obtained using the dual index (a, b) and fusion (c, d) methods. Each box plot represents the upper and lower quartiles (edges), median (horizontal line), the maximum value of the data that is within 1.5 times the interquartile range over the 75th percentile (the upper whisker), the minimum value of the data that is within 1.5 times the interquartile range under the 25th percentile (the lower whisker) and the outliers (black dots). The different colors represent the origin of the biofouling sample analyzed (AUS = Australia, USA = United States of America, CAN = Canada, NZ = New Zealand). Note the difference in the scale of y axis between markers.

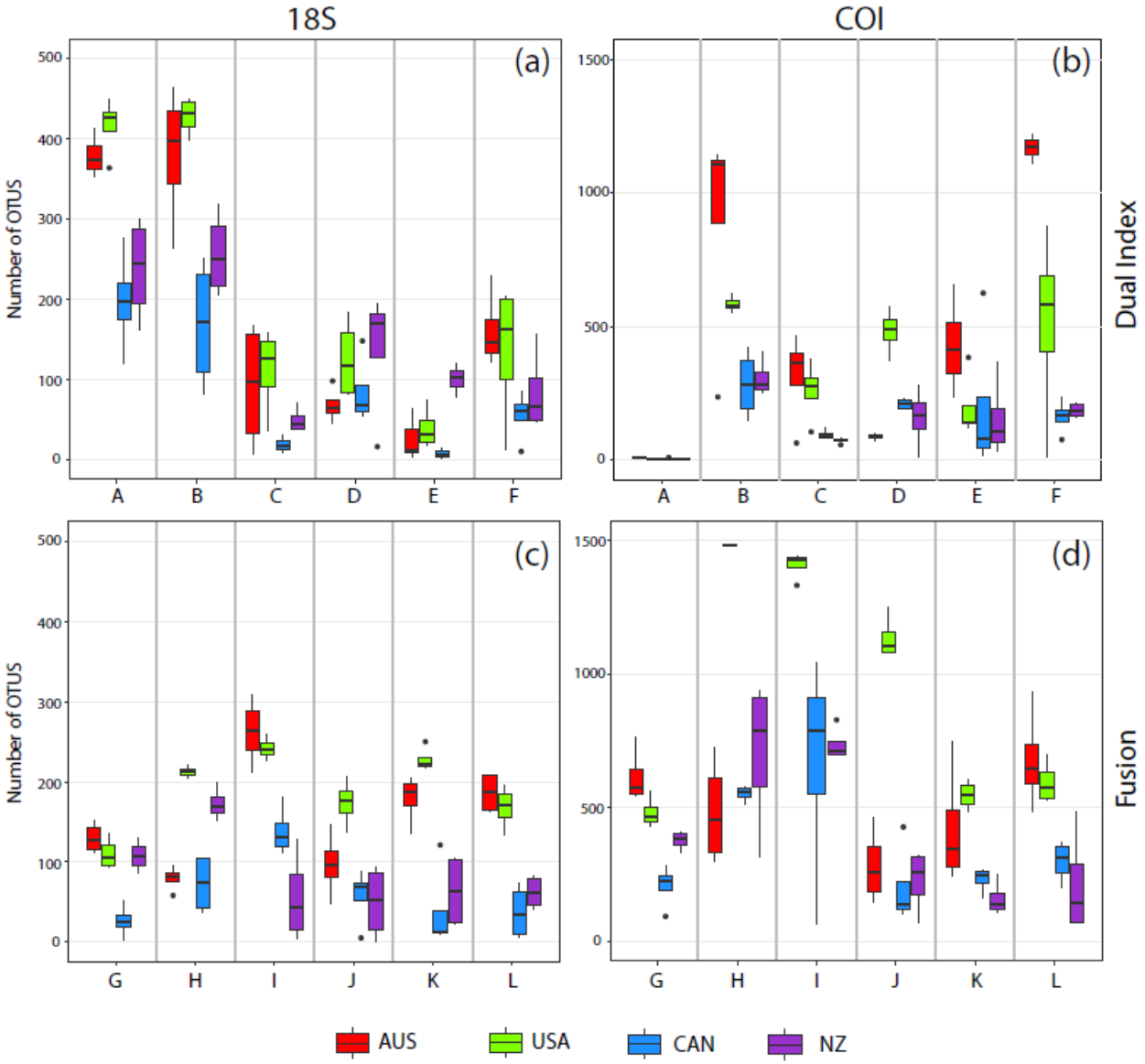


Figure 3. Number of Operational Taxonomic Units (OTUs) obtained for each 18S rRNA gene (a, c) and Cytochrome c Oxidase subunit 1 (b, d) amplicons using the dual index (a, b) and fusion (c, d) methods. Each box plot represents the upper and lower quartiles (edges), median (horizontal line), the maximum value of the data that is within 1.5 times the interquartile range over the 75th percentile (the upper whisker), the minimum value of the data that is within 1.5 times the interquartile range under the 25th percentile (the lower whisker) and the outliers (black dots). The different colors represent the origin of the biofouling sample analyzed (AUS = Australia, USA = United States of America, CAN = Canada, NZ = New Zealand). Note the difference in the scale of y axis between markers.

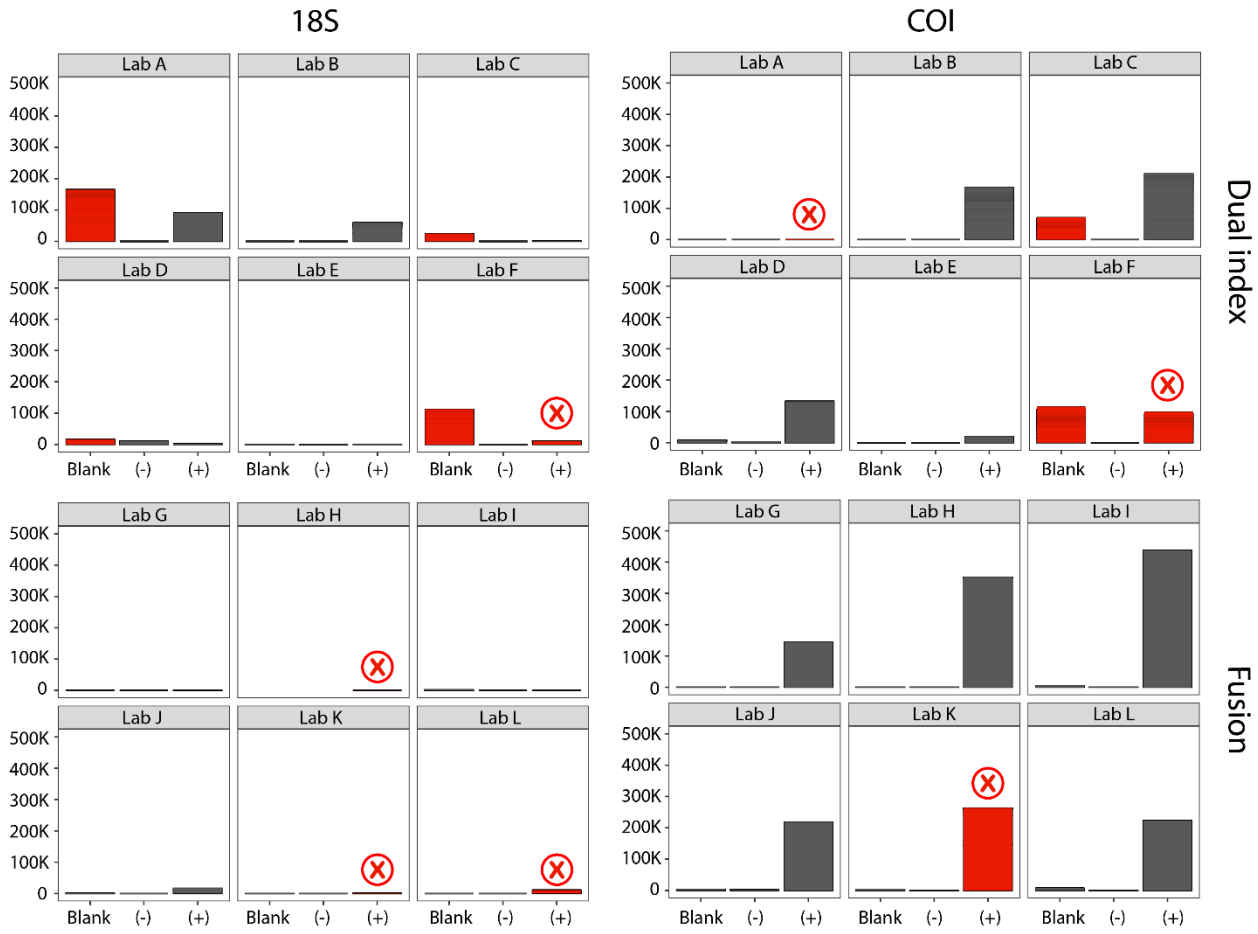
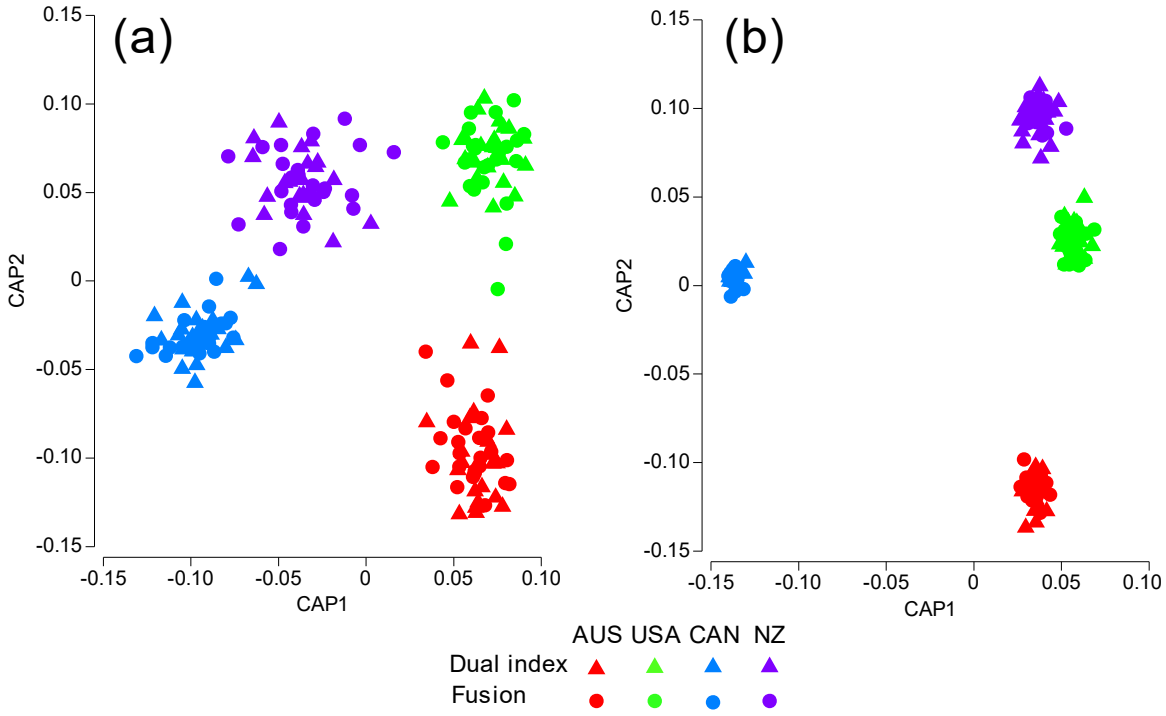


Figure 4. Number of filtered sequence reads recovered from the DNA extraction blanks (blank), negative PCR controls (-) and positive PCR controls (+). Red bars indicate samples with considerable contamination in the DNA extraction blanks; red crosses above red bars indicate positive PCR controls that failed to recover the target taxa. 18S = 18S rRNA gene, COI = Cytochrome c Oxidase subunit 1.

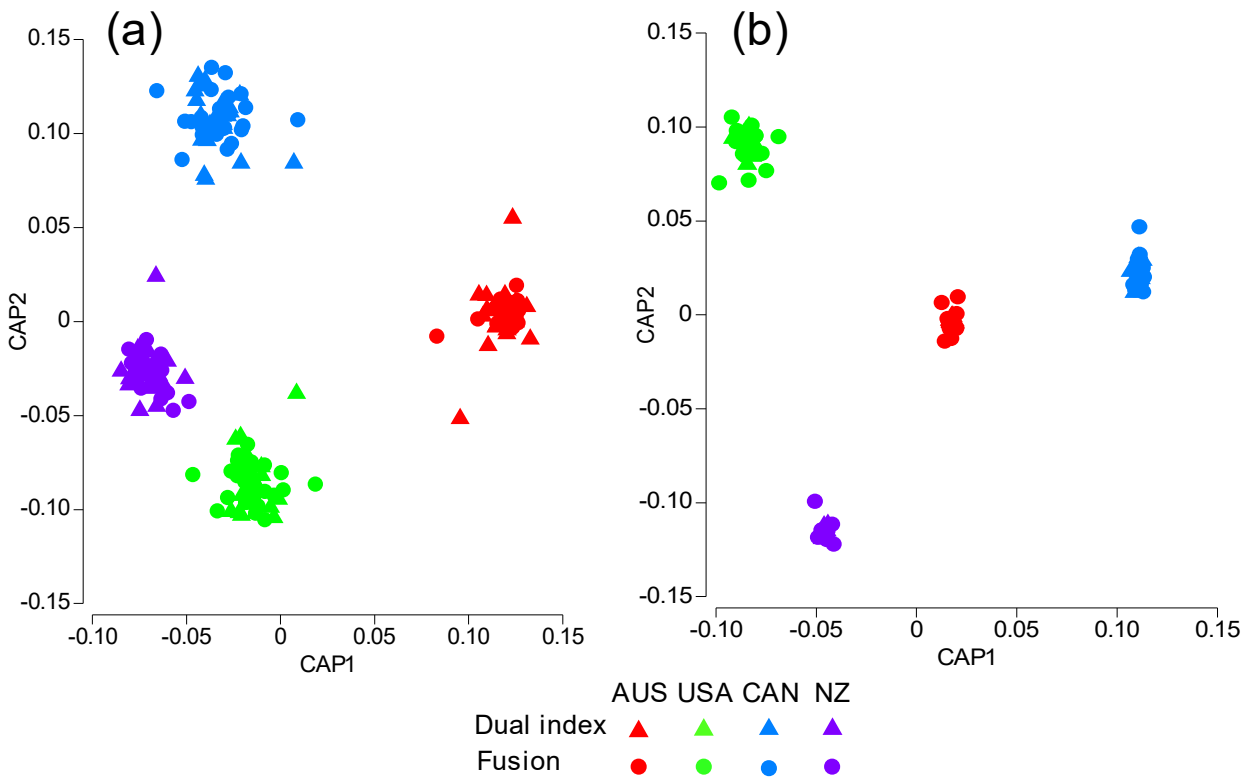


870

871 Figure 4. Canonical analysis of principal coordinates (CAP) plots based on Bray-Curtis dissimilarities of
872 fourth root transformed 18S rRNA gene read abundance data using sample origin as a grouping factor: a)
873 raw, uncorrected data; b) Corr1 data (samples with low sequence number (<1,000) excluded).

874

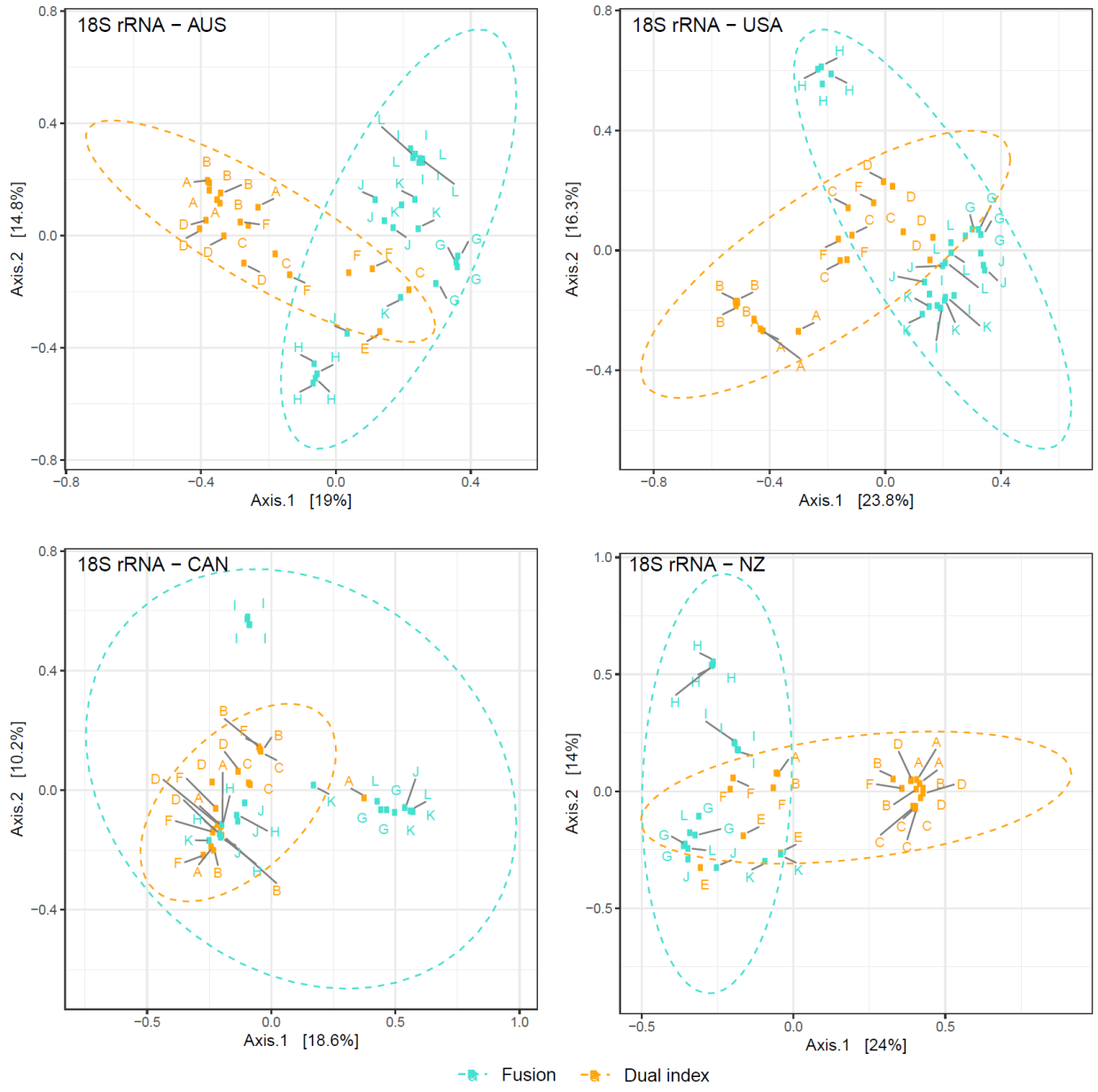
875



876

877 Figure 5. Canonical analysis of principal coordinates (CAP) plots based on Bray-Curtis dissimilarities of
878 fourth root transformed mitochondrial Cytochrome c Oxidase subunit 1 (COI) read abundance data using
879 sample origin as a grouping factor: a) raw, uncorrected data; b) Corr3 data (OTUs found in the negative
880 controls removed from the corresponding subset of data). Triangles represent samples from dual index
881 sequencing library, circles – samples from fusion sequencing library.

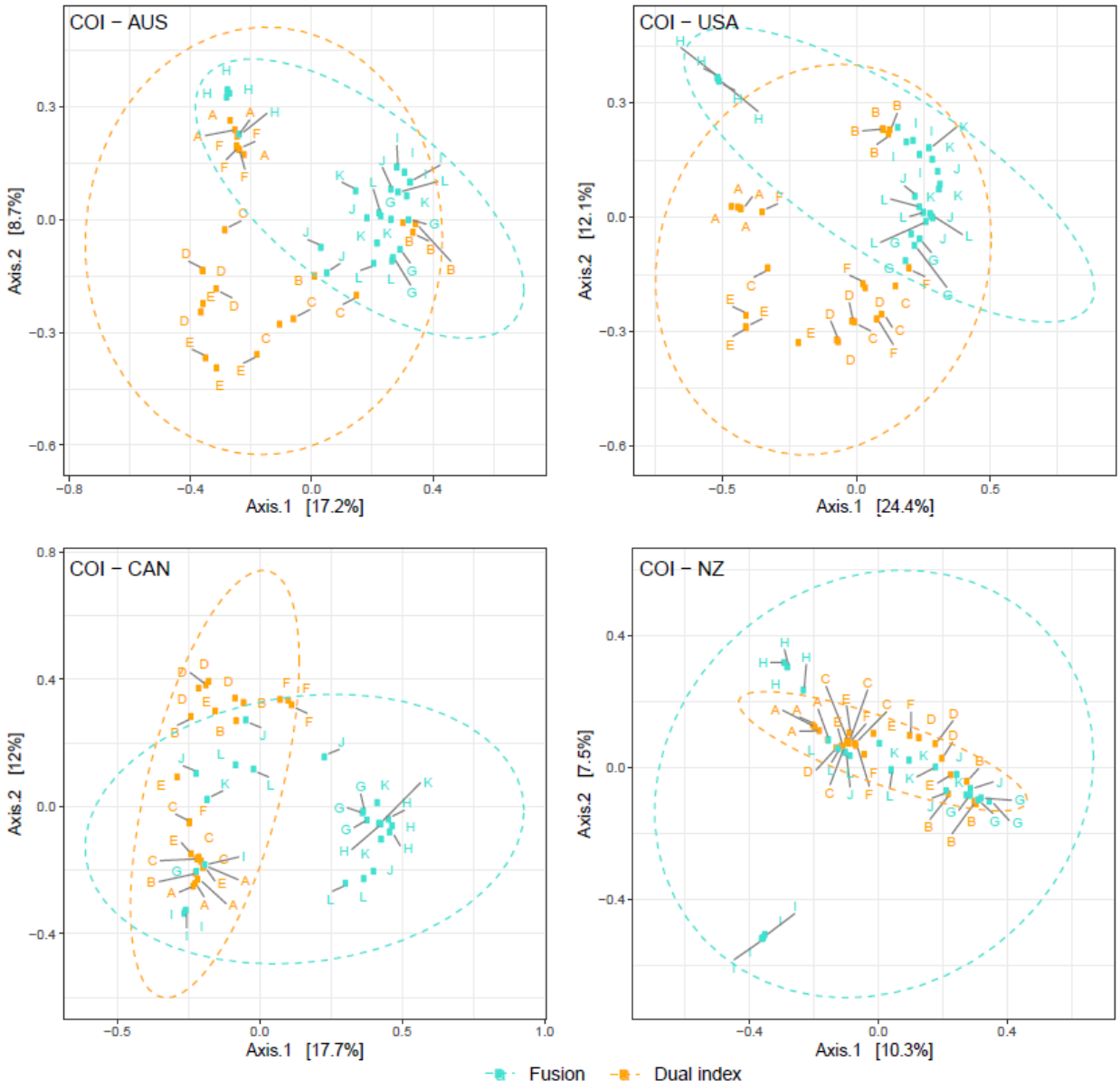
882



883

884 Figure 6. Two-dimensional Principal Coordinates Analysis (PCoA) visualizations of 18S rRNA
885 Operational Taxonomic Unit (OTU) diversity derived in the corrected datasets (Corr1) by different
886 laboratories (laboratory labels noted for each datapoint) for samples originating from Australia (AUS),
887 United States of America (USA), Canada (CAN) and New Zealand (NZ) samples.

888



889

890 Figure 7. Two-dimensional Principal Coordinates Analysis (PCoA) visualizations of mitochondrial
891 Cytochrome c Oxidase subunit 1 (COI) operational taxonomic unit (OTU) diversity derived in the
892 corrected datasets (Corr3) by different laboratories (laboratory labels noted for each datapoint) for
893 samples originating from Australia (AUS), United States of America (USA), Canada (CAN) and New
894 Zealand (NZ) samples.

895

896 Table 1. Summary information on the samples of marine biofouling (starting material).

Sample origin	Sample code	Date	Source	Preservation
Hillarys Boat Harbour, Perth, Australia	AUS	8 February 2017	Settlement plate (scraped biomass, homogenized and frozen before preservation)	RNAlater™
Monterey Bay Harbour, California, USA	USA	1 May 2017	Settlement plate biomass (scraped biomass)	RNAlater™
Victoria Harbour, British Columbia, Canada	CAN	9 May 2017	Marina ropes (scraped biomass)	LifeGuard™
Waitematā Harbour, Auckland, North Island, New Zealand	NZ	20 April 2017	Marina pontoon (scraped biomass)	LifeGuard™

897

898

Table 2. Summary of sample classification with raw and corrected 18S rRNA data from all laboratories. Canonical analysis of principal coordinates (CAP) analyses were carried out on fourth root transformed 18S rRNA gene environmental DNA read abundance data obtained from 12 laboratories: Raw – not corrected; Corr1 – samples with low sequence number (<1,000) excluded; Corr2 – samples randomly rarefied to 10,000 sequence depth; Corr3 – Operational Taxonomic Units (OTUs) found in the negative controls removed from the corresponding subset of data; Corr4 – maximum read abundance of OTUs found in the negative controls subtracted from the corresponding subset of data; Corr5 – OTUs found in negative controls subtracted and samples with low sequence number (<1,000) excluded; Corr6 – OTUs found in negative controls subtracted and samples rarefied to 10,000 sequence depth. Shading indicates the best classification result of samples by origin.

18S rRNA gene	**Raw**	**Corr1**	**Corr2**	**Corr3**	**Corr4**	**Corr5**	**Corr6**
No. of samples	190	167	105	190	190	166	101
No. of OTUs	2,701	2,700	2,463	2,646	2,654	2,648	2,409
No. of sequences	4,860,418	4,855,461	1,050,000	2,956,370	4,548,860	4,543,847	1,010,000
Canonical correlation	0.961	0.997	0.998	0.981	0.953	0.974	0.997
P value	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Correct classification %	88.42	92.81	91.43	87.89	87.89	90.36	91.09
AUS	89.36	90.91	100	87.23	91.49	90.7	96.43
USA	91.67	97.67	92.86	93.75	89.58	95.35	92.86
CAN	93.75	95	85.71	87.5	89.58	92.5	85.71
NZ	78.72	87.5	84	82.98	80.58	82.5	87.5

916 Table 3. Summary of sample classification with raw and corrected mitochondrial Cytochrome c Oxidase
917 subunit 1 (COI) data from all laboratories. Canonical analysis of principal coordinates (CAP) analyses
918 were carried out on fourth root transformed mitochondrial COI eDNA read abundance data obtained from
919 12 laboratories: Raw – not corrected; Corr1 – samples with low sequence number (<1,000) excluded;
920 Corr2 – samples randomly rarefied to 10,000 sequence depth; Corr3 – Operational Taxonomic Units
921 (OTUs) found in the negative controls removed from the corresponding subset of data; Corr4 – maximum
922 read abundance of OTUs found in the negative controls subtracted from the corresponding subset of data;
923 Corr5 – OTUs found in negative controls subtracted and samples with low sequence number (<1,000)
924 excluded; Corr6 – OTUs found in negative controls subtracted and samples rarefied to 10,000 sequence
925 depth. Shading indicates the best classification result of samples by origin.

COI Datasets	**Raw**	**Corr1**	**Corr2**	**Corr3**	**Corr4**	**Corr5**	**Corr6**
No. of samples	192	170	154	192	192	168	150
No. of OTUs	8,549	8,547	7,370	8,198	8,473	8,469	7,288
No. of sequences	27,866,837	27,865,889	1,540,000	13,049,764	27,286,419	27,284,006	1,500,000
Canonical correlation	0.98	0.993	0.994	0.999	0.996	0.997	0.993
p	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Correct classification %	88.02	94.71	92.86	95.58	94.35	95.24	92.67
AUS	93.75	97.73	97.3	97.78	100	97.62	97.06
USA	91.67	97.67	92.5	93.48	95.46	97.67	95
CAN	83.33	92.68	89.74	91.3	93.18	95.12	92.11
NZ	83.33	90.48	92.11	100	88.64	90.48	86.84

926

927