Post-bioinformatic methods to identify and reduce the prevalence of artefacts in metabarcoding data

Running title: Dietary metabarcoding artefacts

Authors: Lorna E. Drake[1], Jordan P. Cuff[1,2], Rebecca E. Young[1], Angela Marchbank[1], Elizabeth A. Chadwick[1] and William O.C. Symondson[1]

[1]School of Biosciences, Sir Martin Evans Building, Cardiff University, CF10 3AX

[2]Rothamsted Insect Survey, Rothamsted Research, West Common, Harpenden, Hertfordshire, AL5 2JQ, UK

Corresponding author: Lorna E. Drake School of Biosciences, Sir Martin Evans Building, Cardiff University, CF10 3AX. Email: Drake.Lorna@gmail.com

Abstract

Metabarcoding provides a powerful tool for investigating biodiversity and trophic interactions, but the high sensitivity of this methodology makes it vulnerable to errors, resulting in artefacts in the final data. Metabarcoding studies thus often utilise minimum sequence copy thresholds (MSCTs) to remove artefacts that remain in datasets; however, there is no consensus on best practice for the use of MSCTs. To mitigate erroneous reporting of results and inconsistencies, this study discusses and provides guidance for best-practice filtering of metabarcoding data for the ascertainment of conservative and accurate data. The most common MSCTs identified in the literature were applied to example datasets of Eurasian otter (*Lutra*

lutra) and cereal crop spider (Araneae: Linyphiidae and Lycosidae) diets. Changes in both the method and threshold value considerably affected the resultant data. Of the MSCTs tested, it was concluded that the optimal method for the examples given combined a sample-based threshold with removal of maximum taxon contamination, providing stringent filtering of artefacts whilst retaining target data. Choice of threshold value differed between datasets due to variation in artefact abundance and sequencing depth, thus studies should employ controls (mock communities, negative controls with no DNA and unused MID-tag combinations) to select threshold values appropriate for each individual study.

Keywords: contamination, diet, eDNA, environmental DNA, false positives

Introduction

Metabarcoding provides a powerful tool for ecological studies of biodiversity and trophic interactions (Deiner et al. 2017; Taberlet et al. 2018). By combining high throughput sequencing (HTS) with DNA barcoding, large volumes of high-resolution data can be generated from many samples simultaneously (Taberlet et al. 2018). As an accurate means of detecting and identifying not just common species, but also cryptic and rare species, metabarcoding has in many cases superseded traditional methods such as morphological analysis of prey remains in gut contents and faeces, and direct observation (Bowser et al. 2013; Roslin and Majaneva 2016; Elbrecht, Vamos, et al. 2017). The high sensitivity of metabarcoding does, however, render it vulnerable to error (Alberdi et al. 2018; Jusino et al. 2019), with differences in the treatment of samples producing distinct data, and thus conclusions, from the same samples (Alberdi et al. 2018; Alberdi et al. 2019). Better guidelines on best practice for data processing are thus required for metabarcoding studies as they become increasingly commonplace.

False positives, or 'artefacts', can be introduced at any stage of the metabarcoding process, from sample collection through to bioinformatic analysis (Alberdi et al. 2019; Jusino et al. 2019). These can occur through contamination from environmental or lab sources (Leonard et al. 2007; Siddall et al. 2009; Czurda et al. 2016), tag-jumping and sample mis-assignment (transfer of sample-specific tags between samples; Schnell et al. 2015) or PCR and sequencing errors (chimeras or mis-identified sequences; Shin et al. 2014; Bjornsgaard Aas et al. 2016). Artefacts may also be produced through errors in reference databases (such as GenBank and BOLD;

Valentini et al. 2009), resulting in sequences being assigned to the wrong taxon (Keskin et al. 2016; Rulik et al. 2017; Taberlet et al. 2018). Many of these artefacts can be limited through careful study design (e.g. pre- and post-PCR workstations; King et al. 2008; Murray et al. 2015) or the use of bioinformatics software to detect and remove erroneous sequences (e.g. UNOISE; Edgar 2016). However, it is likely that some artefacts will remain regardless of precautionary steps taken (Weyrich et al. 2019, Nakagawa et al. 2018), potentially inflating species richness (Schnell et al. 2015; Clare et al. 2016; Zinger et al. 2019) and distorting data interpretation.

Minimum sequence copy thresholds (MSCTs) are one adaptable method commonly used to reduce the prevalence of artefacts (e.g. Hänfling et al. 2016). The choice of threshold must be carefully considered as it can considerably impact the data; low thresholds will be unsuccessful at removing artefacts, leaving false positives in the resultant data, whereas high thresholds may remove too much data, resulting in false negatives (Hänfling et al. 2016). This is especially true for dietary studies in which DNA of the focal consumer can be present at much higher concentrations than that of the food items (i.e. prey) and is undegraded, often resulting in its greater degree of amplification, depending on the PCR primers used. The use of general primers that amplify the consumer will result in a lower proportion of each sample being assigned to food item DNA, whereas specific primers that avoid amplifying the consumer may reduce amplification of some food items over others due to primer bias (Piñol et al. 2014). This variation increases the risk of target sequences being excluded if inappropriate filtering thresholds are selected.

Experimental controls are valuable components for empirically assigning MSCT thresholds, as they provide a mechanism for estimating the proportion of artefacts within a dataset (Taberlet et al. 2018; Alberdi et al. 2019). Theoretically, negative controls (e.g. extraction blanks, PCR blanks and unused MID-tag (molecular identifier tag) combinations) should contain no DNA, and positive controls (e.g. mock communities) should only contain DNA from selected taxa. This is, however, rarely the case, and these unexpected reads facilitate effective determination of optimal thresholds for data clean-up. Reads in negative controls may be previously undetected contamination present in other samples (predominately identified using extraction and PCR blanks; Leonard et al. 2007; Czurda et al. 2016; Alberdi et al. 2019) or may occur due to tag-jumping or sequence mis-assignment (predominately identified using unused MID-tag combinations; Schnell et al. 2015). Such artefacts are impossible to identify with certainty without negative controls since they are mostly assigned to taxa that occur in high read abundances across many samples and are thus indistinguishable from environmental DNA (Carew et al. 2016; Jensen et al. 2016). Further artefacts are detected through the presence of positive control taxa in eDNA samples and eDNA taxa in positive controls, likely through tag-jumping, mis-assignment or sample cross-contamination. Unexpected reads in positive controls also allow low abundance artefacts from contaminants and PCR or sequencing errors, that may occur across eDNA samples too, to be identified. Control samples thus highlight artefacts prevalent throughout unfiltered data, with those identified through negative controls increasing the frequency of occurrence of taxa, those identified through positive controls inflating sample diversity, and both contributing to higher total read counts and, ultimately, false positives.

Application of MSCTs, and use of controls for assessing thresholds, remains ambiguous and non-standardised, with many studies employing entirely distinct methodologies and thresholds (e.g. Gebremedhin et al. 2016; Guardiola et al. 2016; McInnes et al. 2017). Here we compared common practices for removing artefacts from eDNA metabarcoding data using example datasets of Eurasian otter, *Lutra lutra* (Linnaeus, 1758), and cereal crop spider (Araneae: Linyphiidae and Lycosidae) dietary DNA. Samples were processed alongside experimental controls, allowing the practicality of controls for selecting filtering thresholds to be assessed. Through these examples, distinctions in the data outputs when using different techniques are highlighted, providing a basis for standardisation and outlining optimal solutions for the use of MSCTs on metabarcoding datasets. We hypothesised that; (i) data with MSCTs applied would still contain artefacts; (ii) the extent of artefact removal would differ depending upon the method of MSCT applied, with different MSCTs removing artefacts from different sources (e.g. artefacts in blanks vs. those in mock communities); (iii) low filtering thresholds would fail to remove many artefacts; (iv) high thresholds would remove too much data, resulting in the loss of target sequences and hence trophic relationships; (v) using multiple MSCTs simultaneously would remove more artefacts than MSCTs applied on their own; (vi) experimental controls would greatly benefit the choice of filtering method and threshold through identification of known target sequences and artefacts.

Methods

124 The methodologies discussed herein refer to techniques considered best practice for
125 environmental DNA metabarcoding, such as the use of negative controls (samples
126 included in DNA extraction and/or amplification steps), unused MID-tag combinations
127 (combinations of MID-tags that are not included in PCR amplification or sequencing),
128 positive controls (tissue extract DNA of a known species amplifiable by the selected
129 primers, but not expected to occur in eDNA samples) and mock communities
130 (mixtures of positive controls comprising DNA of several species). Theoretically,
131 blanks (i.e. negative controls and unused MID-tag combinations) should contain no
132 reads, and mock communities should contain reads only from selected taxa, with
133 these taxa only occurring within mock communities. Thus, unexpected reads in such
134 controls facilitate effective determination of optimal thresholds for data clean-up.
135 Therefore, inclusion of these controls throughout the metabarcoding process is
136 recommended for stringent data review, and is often necessary for the techniques
137 discussed.

138 To review existing artefact removal methodologies in use for DNA metabarcoding
139 data, the methods used in 154 studies conducting metabarcoding on eukaryotic DNA
140 for environmental monitoring or dietary analysis were tabulated (Table S1). Given the
141 focus of this study on the clean-up of dietary metabarcoding data, which presents
142 many unique challenges, each method was applied to four different datasets from
143 two dietary studies: a dietary study of the Eurasian otter, *Lutra lutra* (one COI and
144 one 16S dataset) and a dietary study of cereal crop money spiders (two COI
145 datasets).

146

Example dataset one: British otter diet

Faecal samples were collected during otter post-mortems by the Cardiff University Otter Project. Extracted faecal DNA was amplified using two metabarcoding primer pairs designed to amplify regions of the 16S rRNA and cytochrome *c* oxidase subunit I (COI) genes, each primer having ten-base-pair molecular identifier tags (MID tags) to facilitate post-bioinformatic sample identification. Extraction and PCR negative controls, unused MID tag combinations, repeat samples and mock communities were included alongside the focal eDNA samples. Mock communities comprised standardised mixtures of DNA of marine species not previously detected in the diet of Eurasian otters (Table S2; Supplementary Information 1). The resultant DNA libraries for each marker were sequenced on separate MiSeq V2 chips with 2x250bp paired-end reads. Greater detail regarding sample processing, amplification and sequencing is provided in Supplementary Information 2.

Example dataset two: cereal crop spider diet

Money spiders (*Bathyphantes*, *Erigone*, *Microlinyphia* and *Tenuiphantes*; Araneae: Linyphiidae) and wolf spiders (*Pardosa*; Araneae: Lycosidae) were visually located on transects through barley fields. Gut DNA, extracted from the whole spider abdomen, was amplified using two COI metabarcoding primer pairs. One primer pair was selected for broad amplification of all invertebrates present, including the predator, and the other designed to exclude spider DNA to avoid predator amplification, each primer having ten-base-pair MID tags to facilitate post-bioinformatic sample identification. Extraction and PCR negative controls, unused

MID tag combinations, repeat samples and mock communities were included alongside the focal eDNA samples. Mock communities comprised standardised mixtures of DNA of exotic species not previously recorded in Britain (Table S2; Supplementary Information 1). The resultant DNA libraries for each marker were sequenced on a MiSeq V3 chip with 2x300bp paired-end reads. Greater detail regarding sample processing, amplification and sequencing is provided in Supplementary Information 3.

Sequence analysis

Bioinformatic analyses were carried out using a custom pipeline. Sequences were first checked for truncation of MID-tags by determining the proportion of sequence files containing exactly 10bp before their respective primer. In all cases, the degree of truncation was deemed acceptable (≤10%).

FastP (Chen et al. 2016) was used to check the quality of reads, discard poor quality reads (<Q30, <125bp long or too many unqualified bases, denoted by 'N') and merge read pairs from MiSeq files (R1 and R2). Merged reads were assigned a sample ID based on the MID tags associated with each primer using the 'trim.seqs' function of Mothur (Schloss et al. 2009); this also removed the MID tag and primer sequences from the reads. Using the files created by Mothur, reads were demultiplexed to obtain one file per sample ID. Read headers were modified for each file to include the sample ID and reads were then concatenated back into one file. Sequences were denoised (removal of PCR and sequencing errors), clustered into zero-radius

operational taxonomic units (zOTUs) and an OTU table was created using the commands ‘fastx_uniques’, ‘unoise3’ and ‘otutab’ in Usearch (v. 11) (Edgar 2016; Edgar 2020). Taxonomic assignment for each zOTU was obtained using the ‘blastn’ command in BLAST+, using a threshold of 97% similarity and e-value of 0.00001, against a downloaded database of DNA barcoding sequences submitted to online databases (e.g. Genbank; National Center for Biotechnology Information 2008; Camacho et al. 2009).

Before assigning taxonomic identities to each zOTU, BLAST results were filtered using the ‘dplyr’ package in R [version 3.6.0] using R Studio [version 1.2.1335] (R Core Team 2019). This was used to retain only accession codes with the top BIT score for each zOTU. These data were then processed via MEGAN [version 6.12.3] (Huson et al. 2016) to assign taxonomic names to each zOTU. As erroneous entries on online databases can prevent species-level assignments, zOTUs for which the top BLAST hit (i.e. top BIT score) was not resolved to species-level were thus manually checked and assigned the most appropriate taxon. Taxonomic identity for each zOTU was added to the OTU table produced by Usearch and reads were aggregated by taxonomic identity for each sample in R using the ‘aggregate’ function with a sum base function. OTUs were allocated taxonomic identities to overcome issues such as over-splitting of taxonomic groups, and to facilitate ecological interpretation of the data, particularly regarding identification of artefacts (e.g. identifying marine species in non-coastal otters).

Minimum Sequence Copy Thresholds (MSCTs)

215 The seven most common MSCTs identified from the literature (Table 1) were tested
216 and their efficacy in cleaning all datasets compared. Filtering methods were enacted
217 in excel using IF formulae.

218

Table 1: Seven post-bioinformatic filtering methods often applied to eDNA metabarcoding datasets, selected from those identified in a review of 154 metabarcoding studies (Table S1). The 'method name', herein used to refer to these methods, is given alongside the description (how the methods are executed) and the aim of each.

Method name	Method Description	Method Aim
1. No filter	No OTU or sample filtering.	No clean-up/maximum preservation of data.
2. Singletons	Remove any read counts of one.	Remove extremely low frequency artefacts (e.g. sequencing artefacts).
3. <10	Remove any read counts that are less than ten.	Remove low frequency artefacts (e.g. sequencing artefacts, low-lying PCR contamination)
4. Max Contamination	Remove any read counts within each OTU that are lower than the highest read count within a negative/blank control for that OTU.	Remove contamination detected by the negative controls (e.g. extraction/PCR contamination, tag-jumping)
5. Total %	Remove any read counts less than a proportion of the total dataset read count for all reads.	Remove low frequency artefacts (e.g. sequencing artefacts, PCR contamination)
6. Sample %	Remove any read counts within a sample that are less than a proportion of the total sample read count for that sample.	Remove sample contamination (e.g. environmental, extraction or PCR contamination)
7. Taxon %	Remove read counts with an abundance less than a proportion of the total OTU read count for that OTU.	Remove cross contamination (e.g. cross contamination, tag-jumping)

If the read count (i.e. number of reads per sample per taxon) did not pass the designated threshold, then it was converted to zero (rather than subtracting the threshold, thus not altering the remaining read counts). For proportional methods (5-7, Table 1), a variety of thresholds were tested to explore how choice of threshold can affect data output. The range of thresholds tested were chosen based upon artefacts identified in control samples; we started with a low threshold and increased

231 the value until most of the identifiable artefacts were removed. We also explored the
232 effectiveness of using different MSCTs in pairwise combinations; this involved
233 simultaneously applying ‘Max Contamination’ with each proportional threshold
234 method (5-7), and ‘Sample %’ with ‘Taxon %’.

235 Basic statistics were calculated to assess the effectiveness of each filtering method;
236 total read count was used to assess the loss of reads across the whole dataset,
237 presence of singleton reads was used to assess removal of PCR and sequencing
238 errors, reads in blanks (negative controls and unused MID-tags) were used to assess
239 levels of contamination and tag-jumping, and mock communities were used to
240 assess presence of false positives within samples. Artefacts could also be identified
241 through taxa unexpectedly occurring in samples, such as taxa from dietary samples
242 in controls, marine taxa associated with otters that did not have access to marine
243 habitats, exotic taxa in British spider samples and mock community taxa in negative
244 controls, unused MID tags or dietary samples.

245 To visualise the results of each method, tables of reads were converted into heat
246 charts using the ‘ggplot2’ package (Warnes et al., 2012) in R. Frequency of
247 occurrence for each taxon across all MID-tag combinations was also calculated for
248 each filtering method and used to create heat charts. Relative frequencies were
249 calculated by dividing frequency of occurrence by the total number of MID-tag
250 combinations; these values then underwent non-metric multidimensional scaling
251 (NMDS) to visualise dissimilarity between the taxa present following application of
252 each MSCT. This was conducted using the ‘metaMDS’ function in the ‘vegan’
253 package (Oksanen et al. 2013) with two dimensions (stress <0.1) and a Bray-Curtis

254 dissimilarity calculation (Bray and Curtis 1957). Ellipses were created using the
255 'ordiellipse' function with the default 'sd' setting (standard deviation).

256

257 **Results**

258 *Sequencing output*

259 Sequencing yielded 17.6, 13.7, 11.2 and 11.0 million paired-end reads, for the otter
260 16S and COI, and spider general and exclusion datasets, respectively, which
261 decreased to 11.7, 7.9, 7.9 and 7.4 million, respectively, following bioinformatic
262 analysis. Comparison of post-bioinformatic clean-up methods produced the same
263 general patterns across the four datasets (otter 16S, otter COI, spider general COI
264 and spider exclusion COI). We therefore used the simplest dataset (otter 16S) to
265 graphically represent artefact removal (Figures 1-2; Table 2), with supplementary
266 information presenting the same data for the other datasets (otter COI, spider
267 general COI and spider exclusion COI; Figures S1-3; Tables S2-4), as well as graphs
268 depicting read counts per sample (Figures S4-7) and the spatial distribution of otter
269 faecal samples with marine taxa presences (Figures S8-9). The effectiveness of each
270 clean-up method across all datasets is also summarised in Table 3.

271

272 *No filter applied ('No Filter')*

273 The highest read counts and occurrence of artefacts were observed in data with no
274 MSCT applied. False positives in mock communities, reads in blanks, mock

community taxa present in blanks and eDNA samples, taxa from eDNA samples occurring in control samples, and obviously erroneously present taxa (e.g. marine taxa occurring in faecal samples from otters with no access to marine habitats) all occurred frequently across the datasets (Figure 1; Table 2). Artefacts appeared to be much more prevalent for taxa with high total read counts (e.g. mock community taxa, taxa commonly consumed by the predator and the focal predator itself). Many low abundance reads, including singletons, were also observed in the unfiltered data, possibly representing rare species but likely also sequencing errors.

Remove singleton reads ('Singletons')

Removing singleton reads resulted in data very similar to that of unfiltered data in all cases, with only few artefacts removed (Figure 1; Table 2).

Remove read counts less than 10 ('< 10')

Removing reads with an abundance less than 10 reduced the occurrence of artefacts in blanks, mock communities and the presence of mock community taxa in other samples. However, artefacts persisted in all controls and samples, producing data very similar to unfiltered data (Figure 1; Table 2).

Remove maximum taxon contamination ('Max Contamination')

Removing reads less than or equal to the maximum read count in blanks per taxon removed no reads from some taxa and high values from others (otter 16S: minimum read removal = 0, maximum = 8757, average = 394; otter COI: minimum = 0, maximum 23413, average = 117; spider amplification: minimum = 0, maximum = 5851, average = 136; spider exclusion: minimum = 0, maximum = 10764, average = 155). Taxa experiencing high levels of read removal were often those with high total read counts. This cleared all reads from blanks (Tables 2-3), all mock community taxa from eDNA samples and taxa with high read abundances in eDNA samples from controls (Figure 1). False positives were still present in mock communities though (Figure 1), as were singleton reads. This method also cleared several erroneously located taxa, such as marine species associated with inland otters, but not all (Figure 1; Table 2).

Proportion of total read count ('Total %')

This method removed artefacts present in blanks (Table 2), false positives in mock communities and erroneously located taxa (Figure 1; Table 2). Mock community taxa were cleared from blanks and eDNA samples to an extent, but some were still present even at high thresholds (Table 2). Taxa from dietary samples with high read abundances were not filtered efficaciously though, with many occurring in controls even at high thresholds. Thresholds tested across the datasets ranged between removing reads that contributed to less than 0.0001% and 0.02% of the total read count. The lowest thresholds only filtered out a proportion of the artefacts, whilst the highest thresholds filtered out all false positives within mock communities and almost

all reads in blanks (Figure 1; Table 2); however, the latter also removed target reads, shown by the loss of mock community taxa within mock communities. A lower threshold was therefore necessary to give a balance between false positives and false negatives. The optimal threshold was identified as 0.003%, 0.0008%, 0.0005% and 0.005% for otter 16S, otter COI, spider general amplification and spider exclusion, respectively, removing reads with abundances less than 79, 352, 39 and 236, respectively.

Proportion of read count per sample ('Sample %')

This method removed false positives from mock communities (Figure 1) and erroneously located taxa (Table 2). Low abundance taxa (e.g. foreign taxa occurring through sequencing errors) were less prevalent (Figure 1), as were singletons. Taxa with high total read abundances (e.g. mock community taxa and common taxa in dietary samples) and reads present in blanks were only filtered to an extent (Figure 1; Table 2), resulting in artefacts from both being prevalent in filtered data regardless of the threshold utilised. This method removed fewer reads from samples with low total read counts, therefore these samples were more likely to still contain artefacts. Thresholds tested across the datasets included removing reads that contributed less than 0.01% to 8% of a sample's reads. The highest thresholds were required to remove all false positives from mock communities. A much higher threshold was required for some datasets (e.g. otter 16S) when they contained taxa with greater relative read counts. The high thresholds required to clear mock communities of false positives also removed many target reads (highlighted by the loss of mock

community taxa), thus lower thresholds effectively balanced false positives and false negatives. The optimal threshold was identified as 1%, 0.3%, 0.38% and 1% for otter 16S, otter COI, spider general amplification and spider exclusion, respectively. These thresholds removed reads to a varying degree (otter 16S: minimum read removal for a sample = 0, maximum = 8757, average = 394; otter COI: minimum = 0, maximum = 23413, average = 117; spider general amplification: minimum = 1, maximum = 240, average = 80; spider exclusion: minimum = 1, maximum = 1704, average = 199).

Proportion of read count per taxon ('Taxon %')

This method filtered out reads in blanks (Figure 1; Table 2), as well as artefacts from taxa with high read abundances, clearing most of these from the datasets when using sufficient thresholds. A large proportion of reads were removed using this method (Figure 1; Table 2), especially from taxa with high total read counts. Taxa with low read counts had fewer reads removed, resulting in these containing more artefacts, highlighted by the prevalence of singleton reads and taxa identified as PCR or sequencing errors (e.g. foreign taxa; Figure 1). This method proved insufficient at removing false positives from eDNA samples, with false positives prevalent in mock communities regardless of the threshold used, and erroneously located taxa were only removed when using a high threshold (Figure 1; Table 2). Thresholds tested included removing reads that contributed to less than 0.1% - 3% of a taxon's reads. With low thresholds applied, many more artefacts were observed in blanks, but a threshold of 3% cleared most of these artefacts from the datasets in most cases. The highest thresholds removed a high proportion of reads, therefore lower thresholds

were selected to give a balance between clearing out artefacts and not losing too many reads; this was 0.5%, 0.8%, 0.5% and 1% for otter 16S, otter COI, spider general amplification and spider exclusion, respectively. These thresholds removed reads to different extents (otter 16S: minimum read removal for a taxa = 0, maximum = 26039, average = 553; otter COI: minimum = 0, maximum = 2040, average = 49; spider general amplification: minimum = 0, maximum = 306, average = 28; spider exclusion: minimum = 0, maximum = 1286, average = 76).

Combining methods

Many of the thresholds tested for MSCTs based on read counts ('Total %', 'Sample %' and 'Taxon %') did not clear all artefacts, particularly regarding clearance of blanks. Proportional methods were thus also combined with 'Max Contamination' to overcome this issue. 'Sample %' thresholds were also combined with 'Taxon %' thresholds given their complementary removal of artefacts. Combining methods removed more artefacts than using just one method. 'Total %' thresholds or 'Sample %' thresholds combined with 'Max Contamination' left very few artefacts in the data. These methods were highly complementary, with proportional thresholds clearing most false positives from mock communities and erroneously located taxa (Figure 1; Table 2), whilst the contamination threshold cleared reads in blanks and artefacts from taxa with high read counts (e.g. mock community taxa in non-mock community samples and faecal taxa in controls; Figure 1; Table 2). These combinations also cleared singletons and taxa suspected to be PCR or sequencing errors (Figure 1; Table 2). Combining these methods sometimes allowed lower thresholds to be used

387 concurrently for optimal results, but in other cases did not change the thresholds
388 required (otter 16S: optimal sample % = 0.5%, optimal total % = 0.002%; otter COI:
389 optimal sample % = 0.2%, optimal total % = 0.0008%; spider general amplification:
390 optimal sample % = 0.38%, optimal total % = 0.005%; spider exclusion: optimal
391 sample % = 0.39%, optimal total % = 0.005%).

392 'Taxon %' thresholds combined with 'Max Contamination' still contained many
393 artefacts; all reads in blanks and singletons were removed, but false positives were
394 still present in mock communities as were erroneously located taxa (although in
395 lower abundances compared to either filter alone; Figure 1; Table 2). This is likely
396 due to the similar action of both filters. Combining 'Taxon %' thresholds with 'Sample
397 %' thresholds removed more artefacts and performed similarly to MSCTs combining
398 'Sample %' thresholds with 'Max Contamination'. Combining these methods cleared
399 the majority of reads from blanks, all singleton reads, artefacts from taxa with high
400 read counts and most false positives in mock communities (Figure 1; Table 2);
401 however, there were still artefacts present in the negative controls and erroneously
402 located taxa were still present (Table 2). Combining these methods also removed
403 many overall reads. The optimal combination of thresholds changed between
404 datasets (otter 16S: sample = 0.5%, taxon = 0.3%; otter COI: sample = 0.2%, taxon =
405 0.3%; spider general amplification: sample = 0.5%, taxon = 0.3%; spider exclusion:
406 sample = 0.5%, taxon = 0.3%). Lowering the sample threshold introduced more false
407 positives to the data, whilst increasing the threshold removed target reads. Lowering
408 the taxon threshold retained more reads in blanks and artefacts from taxa with high
409 total read counts, whilst increasing the taxon threshold greatly decreased the total
410 read count, resulting in loss of target reads.

411

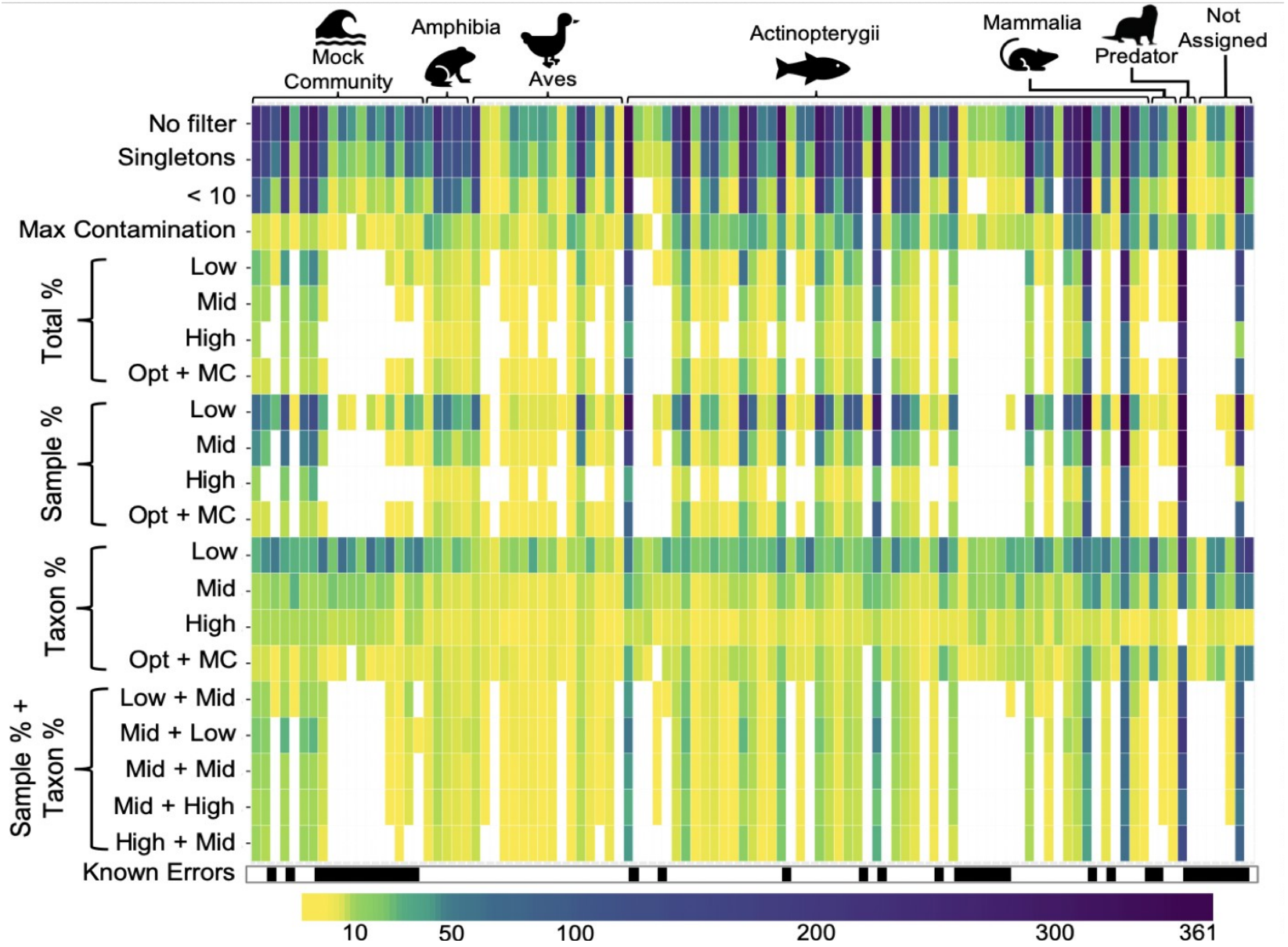

Mock Community
Amphibia
Aves
Actinopterygii
Mammalia
Predator
Not Assigned
No filter
Singletons
< 10
Max Contamination
Total %
Low
Mid
High
Opt + MC
Sample %
Low
Mid
High
Opt + MC
Taxon %
Low
Mid
High
Opt + MC
Sample % + Taxon %
Low + Mid
Mid + Low
Mid + Mid
Mid + High
High + Mid
Known Errors
10
50
100
200
300
361

Figure 1: Otter diet 16S counts. The number of presences of each taxon is displayed for each method (low count = yellow, high count = purple) along with the number of taxa in each dataset following clean-up. Differences in common taxa, mock communities, predator amplification and erroneous taxa can be observed. 'Low', 'Mid' and 'High' depict the context-dependent range of values utilised for proportional thresholds ('Total %', 'Sample %' and 'Taxon %'), with 'Opt + MC' denoting the threshold deemed 'optimal' combined with the 'Max Contamination' method (for specific values see Table S3). The same figure is available for three other datasets (otter COI, spider general COI and spider exclusion COI) in supplementary information (Figures S1-3).

Table 2: Performance of different minimum sequence copy thresholds on otter 16S data. 'Low', 'Mid' and 'High' depict the context-dependent range of values utilised for proportional thresholds ('Total %', 'Sample %' and 'Taxon %'), with 'Opt + MC' denoting the threshold deemed 'optimal' combined with the 'Max Contamination' method (for specific values see Table S3). Expected presences of marine taxa (~) were defined by the number of Eurasian otters (*Lutra lutra*) displaying reads for each marine taxon that was located along the coast or near an estuary. Similar tables were produced for three other datasets (otter COI, spider general COI and spider exclusion COI) and are presented in supplementary information (Tables S2-4).

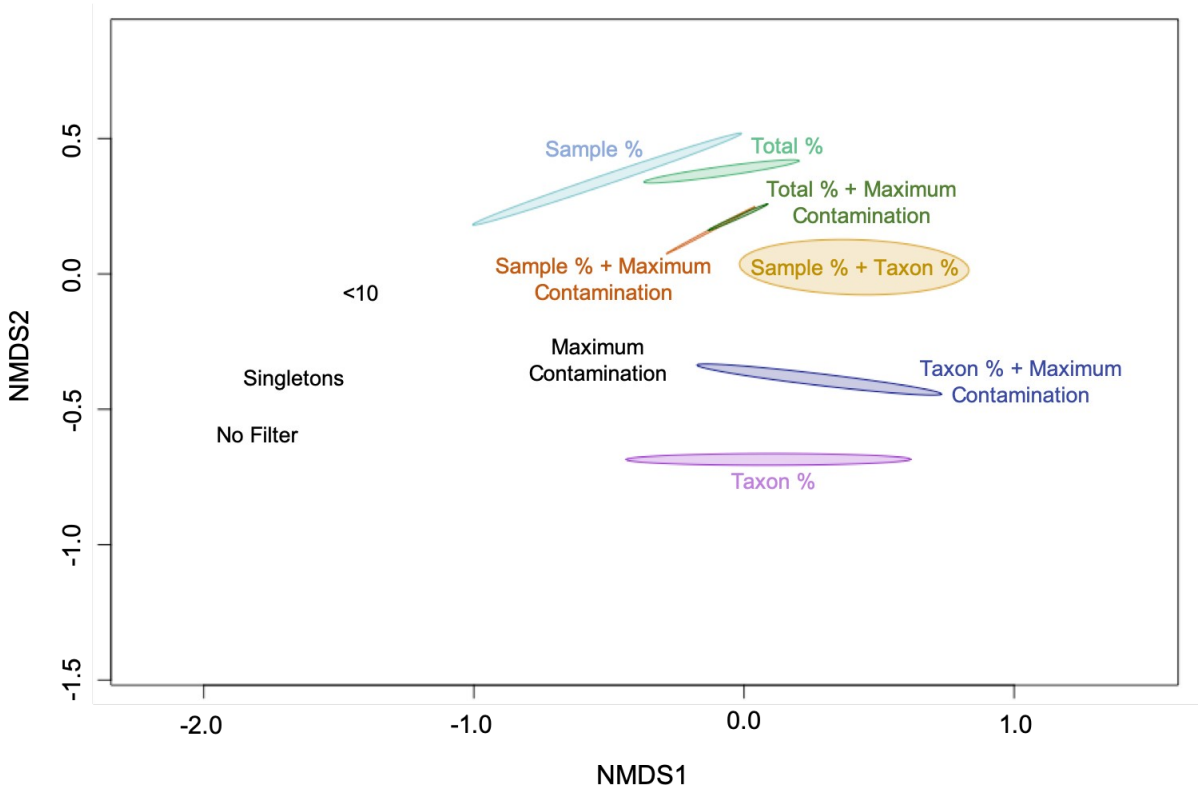
Minimum Sequence Copy Threshold	Total		Singletons	Blanks		Mock Communities			Marine taxa presences		
	Summed read count	Taxa	Number of presences	Summed read count	Average read count	Average false positive read count	Average false positive presences	Presences in eDNA samples/ blanks	*T. bubalis* (~1-3)	Pleuronectidae (~10-15)	*E. viperia* (~1)
No filter	11723871	105	2767	117460	1864	3121	38	295	166	324	37
Singletons	11721104	105	0	117032	1858	3113	30	259	84	291	28
< 10	11705943	99	0	114675	1820	3066	19	198	38	194	7
Maximum Contamination	10938496	102	63	0	0	314	4	0	36	14	7
Low Total %	11534535	71	0	96869	1538	2498	5	38	11	36	1
Mid Total %	11349821	60	0	78023	1238	2018	2	11	1	14	1
High Total %	10733900	46	0	35916	570	220	0	2	1	10	0
Opt Total % + MC	10874148	63	0	0	0	115	0	0	1	14	1
Low Sample %	11659268	89	218	116737	1853	3290	10	126	40	172	3
Mid Sample %	11478669	68	0	113804	1806	2113	2	51	6	38	1
High Sample %	10631707	46	0	86797	1378	0	0	21	1	8	1
Opt Sample % + MC	10875890	65	0	0	0	96	0	0	3	14	1
Low Taxon %	11031736	105	742	45985	730	376	13	21	36	27	22
Mid Taxon %	8669244	105	267	30812	489	163	8	2	19	12	5
High Taxon %	3660086	104	25	30645	486	99	5	1	1	7	1
Opt Taxon % + MC	8569029	102	0	0	0	96	2	0	19	12	5
Low Sample % + Mid Taxon %	10187214	72	2	30851	490	15	0	2	12	13	1
Mid Sample % + Low Taxon %	10959369	68	0	44471	706	140	0	19	4	16	1
Mid Sample % + Mid Taxon %	10177475	67	0	30434	483	124	0	2	4	13	1
Mid Sample % + High Taxon %	8647191	67	0	29865	474	124	0	2	4	12	1
High Sample % + Mid Taxon %	10155032	60	0	29886	474	0	0	2	2	13	1

436
437

Table 3: Success of different filtering methods in achieving the key objectives of post-bioinformatic data clean-up. Green, orange and red denote positive, neutral and negative outcomes, respectively. 'Low', 'Mid' and 'High' depict the value utilised for proportional thresholds ('Total %', 'Sample %' and 'Taxon %'), with 'Opt + MC' denoting the 'optimal' threshold combined with 'Max Contamination' methods (for specific values see Table S3).

NMDS analysis

Choice of MSCT method greatly affected the final composition of the data across all four datasets, as shown by NMDS (Figure 3). Application of ‘No Filter’, ‘Singletons’ and ‘< 10’ MSCTs produced similar outcomes, with the ‘< 10’ threshold also appearing to elicit similar effects to MSCTs based on ‘Total %’ and ‘Sample %’. ‘Sample %’ and ‘Total %’ thresholds were the most similar and gave results distinct from those of taxon MSCTs (‘Taxon %’ and ‘Maximum Contamination’). By combining taxon MSCTs with either ‘Sample %’ or ‘Total %’ thresholds, an intermediate result was obtained. All combinations of taxon filters with ‘Sample %’ or ‘Total %’ thresholds performed similarly to one another; however, with the otter 16S data those that combined ‘Sample %’ or ‘Total %’ with ‘Maximum Contamination’ were more dissimilar to taxon methods than combinations between ‘Sample %’ and ‘Taxon %’.



459

460 **Figure 2: Otter 16S non-metric multidimensional scaling of relative frequency of occurrence for each taxon following**
461 **application of different minimum sequence copy thresholds, including different methods and thresholds where possible.**
462 **Ellipse colours denote each method with None, Singletons, <10 and Maximum Contamination not having ellipses given**
463 **the lack of modifiable threshold. The same figure is available for three other datasets (otter COI, spider general COI and**
464 **spider exclusion COI) in supplementary information (Figures S7-9).**

Discussion

Here we have illustrated the efficacy of different filtering methods and thresholds for removal of artefacts from metabarcoding data, allowing us to identify an optimal method for artefact removal; utilising a threshold that removes a proportion of read counts per sample, combined with a threshold that removes reads with a count less than the maximum contamination identified per taxon ('Opt sample % + MC'; Table 3). For optimisation of thresholds, previous studies have disproportionately emphasised the importance of mock communities (e.g. Elbrecht and Leese 2017; Jusino et al. 2019); however, since the biases affecting true unknown mixtures of eDNA are almost impossible to experimentally replicate (Alberdi et al. 2018), data cannot be adequately filtered using only mock communities. By sequencing and analysing mock communities, blank samples and eDNA together, it was possible to fully assess which filters and thresholds were optimal in cleaning metabarcoding data of this nature.

Identifying artefacts

Despite all appropriate precautionary steps being taken to reduce contamination (e.g. screening negative controls, pre- and post-PCR workstations), and bioinformatic programmes used to remove erroneous sequences, artefacts were still observed in the unfiltered data. Such contamination is, however, largely unavoidable when using a method so broad-spectrum and sensitive (Alberdi et al. 2018; Jusino et al. 2019). Artefacts primarily manifested as unexpected reads in control samples, but also as erroneous taxa and misassigned reads. Erroneous taxa, usually existing in low read counts in the unfiltered data (De Barba et al. 2013; Ficetola et al. 2015) are,

in this case, taxa produced through PCR or sequencing errors that are ecologically highly unlikely to appear in their respective samples (e.g. foreign species), thus rendering them easy to identify and eliminate. Mis-assigned reads were more difficult to identify, primarily detected through mock community taxa occurring in eDNA samples and vice versa; however, some datasets also allow detection of mis-assignment between eDNA samples through the presence of, for example, marine taxa in land-locked sites (Figures S1-6). In such cases, reads were assumed to be derived from other samples through cross-contamination, tag-jumping or mis-assignment (Schnell et al. 2015; Alberdi et al. 2019). If easily identifiable, this can be fortuitous for threshold determination, but where samples share taxa that could theoretically co-occur, they will remain undetected.

Detection of artefacts is facilitated through the presence of unexpected reads in controls. Such reads in negative controls may occur due to low levels of contamination (e.g. from reagents or samples; Leonard et al. 2007; Czurda et al. 2016; Alberdi et al. 2019) that went undetected during screening of samples and may be present throughout only a few, or potentially all samples. Reads present in blanks may also occur due to tag-jumping or mis-assignment (Schnell et al. 2015), which are primarily identifiable through unused MID-tag combinations. These artefacts are hard to detect without blanks because they are frequently assigned to taxa that legitimately occur in high read abundances across many samples (Jensen et al. 2015; Carew et al. 2018), such as mock community taxa and common taxa in eDNA samples (e.g. commonly consumed taxa or the consumer itself). Further artefacts were detected through the presence of mock community taxa in eDNA samples and common eDNA taxa in mock communities; these were concluded to be primarily due to tag-jumping or mis-assignment rather than sample cross-

514 contamination because eDNA and mock community samples were processed
515 separately. Unexpected reads in mock communities also allowed low abundance
516 artefacts from contaminants and PCR or sequencing errors to be identified, which
517 may have occurred across the eDNA samples. Control samples showed artefacts
518 were prevalent throughout the unfiltered data, with those identified through blanks
519 increasing the frequency of occurrence of taxa, those identified through mock
520 communities inflating sample diversity and both contributing to higher total read
521 counts and, ultimately, false positives

522 The composition of mock communities is of great importance to the process of
523 identifying artefacts. If the mock communities are comprised of species that may
524 feasibly occur in the eDNA samples taken from the focal study system, the utility of
525 those controls is reduced. Although the mock communities in this study comprised
526 species considered highly unlikely to appear in the corresponding eDNA samples,
527 distinct problems were encountered for all datasets. For the otter dietary analysis,
528 the mock communities contained marine taxa unlikely to have been consumed by
529 otters, yet high read counts were observed in the COI mock communities for brill
530 (*Scophthalmus rhombus*), a species known to be consumed by otters and not
531 included in the mock community mixtures. The marine samples from which DNA was
532 extracted were collected as part of a larger marine surveying initiative and, whilst
533 care was taken by the practitioners responsible for the collection, cross-
534 contamination between species was possible. Since this taxon could legitimately
535 occur in both mock communities and eDNA samples, false presences are harder to
536 confirm, but its marine origin meant that in areas lacking access to marine prey by
537 otters, reads could still be identified as artefacts. The mock community mixtures
538 used for the spider dietary analysis included exotic species from Round Island,

539 Mauritius, collected as part of a separate study. These were selected for their
540 absence in Britain and taxonomic relevance to the expected prey species (also small
541 invertebrates). Given the poorly described entomological fauna of Round Island,
542 Mauritius, the identities of a minority of these species were not resolved in the
543 bioinformatics process, resulting in their designation as 'not assigned' and thus their
544 exclusion from the filtering process alongside other unassigned taxa.

545

546 *Performance of Minimum Sequence Copy Thresholds (MSCTs)*

547 Artefacts were removed to varying extents depending on the filtering method and
548 threshold utilised. Basic MSCTs commonly used in the literature, such as removing
549 singletons (e.g. Oliverio et al. 2018) or reads with an abundance less than 10 (e.g.
550 Gebremedhin et al. 2016), removed very few artefacts. This will, however, vary with
551 sequencing depth, with relatively greater depths increasing the likelihood of artefacts
552 having more than 10 occurrences (De Barba et al. 2014; Elbrecht & Leese 2015).
553 MSCTs removing reads with an abundance below a proportion of the total read
554 count performed better, reducing abundance of all detectable artefacts; however,
555 applying one threshold across all read counts potentially indiscriminately removes
556 target reads with low abundances and retains abundant artefacts. This bias can be
557 overcome by using MSCTs based on sample read counts, as the read count will
558 inevitably vary between samples despite best efforts to facilitate consistent sample
559 read depths (Deagle et al 2019). Sample MSCTs efficaciously removed artefacts
560 from within samples, with lowered levels of cross-contamination and erroneous taxa,
561 but did not clear artefacts from blanks, nor abundant taxa.

MSCTs that removed reads less than the maximum read count present in the blanks for each taxon ('Max Contamination'), and those which removed reads less than a given proportion of the total read count for that taxon ('Taxon %'), removed artefacts from blanks and abundant taxa, but not mock communities or erroneous taxa. Of these two methods, removal of maximum taxon contamination was more suitable as it removed all artefacts from negative controls and taxa with high read counts without removing too many reads overall. To achieve the same result using thresholds based on taxon read counts resulted in much greater read losses, increasing the likelihood of false negatives. Proportional taxon thresholds also showed a strong bias towards removing reads from abundant taxa. Whilst helping to remove artefacts produced through tag-jumping, this would potentially produce false negatives if taxa legitimately occurred in many samples. Comparing proportional taxon thresholds to others that cleared out similar amounts of artefacts revealed that proportional taxon thresholds produced the highest loss of reads, thus making this method more likely to lead to false negatives. Removal of maximum taxon contamination is logically superior given that the taxa for which the greatest number of reads will be removed will be based on those that are verifiably contaminating the blanks. Care must, however, be taken to ensure that the protocols followed to produce the blanks are sufficiently stringent but not unnecessarily conservative (e.g. negative control volumes included being based on the average volume pooled per plate, vs. the maximum volume pooled per plate), since this will cause this filtering method to produce many false negatives through overly strict removal of data.

Combining MSCTs

Combining different MSCTs improved the performance of all filters, leading to a greater reduction in artefact presence. The weakest combination used proportional taxon thresholds with removal of maximum taxon contamination ('Taxon %' with 'Maximum Contamination'); these analogous methods removed artefacts in similar ways (i.e. removal based on reads present across taxa, rather than across samples), with neither sufficiently mitigating artefacts within samples. Artefacts persisting in blanks, following application of total read count thresholds, were removed by combining this method with removal of maximum taxon contamination; however, this combination may introduce biases by not accounting for read depth variation between samples, thus providing overly conservative filtering to some samples and insufficient filtering to others. Taxon-based thresholds were complementary to sample-based thresholds, with one removing artefacts identified through blanks and abundant taxa and the other removing artefacts within samples, including erroneous taxa. Combining sample-based thresholds with removal of maximum taxon contamination performed better than combinations with proportional taxon thresholds, as a greater proportion of artefacts were removed with a lower total read loss, reducing the likelihood of false negatives. Due to its consistently improved performance over other MSCTs across all four metabarcoding datasets, we conclude that combining a sample-based threshold with removal of maximum taxon contamination is the optimal method for stringent filtering of metabarcoding data whilst retaining target data.

Choosing an appropriate threshold

In metabarcoding studies, removal of false positives tends to be prioritised over false negatives due to the assumption that reads prove taxon presence whilst a lack of reads does not prove absence because false negatives can occur due to experimental biases (e.g. sampling or primer bias; Oehm et al. 2011; Pinol et al. 2015). A trade-off exists whereby removal of false positives leads to an increase in false negatives (Zepeda-Mendosa et al. 2016; Alberdi et al. 2019), observed here when utilising high thresholds which removed many artefacts but also removed target reads, biasing results to taxa with high read abundance. Ultimately though, not all false positives are identifiable, meaning some artefacts may persist despite appropriate filtering removing all known artefacts. A balance can be achieved by which a high proportion of false positives are removed whilst retaining only very few false negatives that are easily disregarded (Clare et al. 2016; Hanfling et al. 2016; Zizka et al. 2019), thus better reflecting the true diversity within samples. The threshold at which this balance is achieved varies between studies depending on the sequencing depth and breadth of taxa. Appropriate thresholds should be chosen based on artefact removal from control samples. The aim of the study should, however, also be considered. Studies concerning commonly-detected taxa can employ more stringent filters that remove more artefacts at the expense of losing rare taxa that may not be of interest anyway (e.g. studies of major sources of nutrition to a predator). However, studies concerning rare taxa should consider refining their thresholds to optimally remove artefacts whilst retaining the greatest amount of sequencing data (e.g. surveys of species richness).

In this study, we chose to assess the effectiveness of different thresholds using taxa read counts as well as occurrences (count data converted to presence or absence). Occurrence data is often assumed to be a conservative method of assessing

metabarcoding data, as recovery biases (e.g. primer bias, starting amount of DNA) have a lower impact on such data (Deagle et al. 2019). Although occurrence data can inflate the importance of taxa that occur at low read counts (e.g. rare taxa or taxa consumed in small amounts; Deagle et al. 2019), and therefore also artefacts, we found it provided a simple and concise method for assessing artefact prevalence. Other methods, such as relative read abundance (RRA), may provide an alternative method for assessing abundance of artefacts and their impact on metabarcoding datasets by considering the proportion of reads each taxon contributes to a sample's total read count (this is analogous to the 'Sample %' MSCT). However, conversion of reads to RRA can produce misleading results due to biases such as differential digestion rates or primer amplifications (Pompanon et al. 2012; Clare 2014; Piñol et al. 2014; Thomas et al. 2014; Elbrecht and Leese 2015; Elbrecht et al. 2017; Alberdi et al. 2018), whilst the loss of read count data can potentially obscure interpretations of overall data loss. For these reasons we chose not to convert read count data into RRA in this study but instead use raw read counts to assess the use of different MSCTs, thus allowing both artefact abundance and overall loss of reads to be assessed and directly compared. Future developments may make RRA a useful tool for artefact detection and removal though, allowing identification of artefacts that are having a proportionally large impact on metabarcoding data.

Previous studies

A review of the relevant literature (154 DNA metabarcoding papers; Table S1) revealed a large proportion of eDNA studies did not employ MSCTs (29%) and those which did often used entirely distinct methodologies and thresholds, with no optimal

658 method apparent. Studies utilising one threshold across all read counts were
659 commonly used (18% of studies; Table S1), but often employed largely arbitrary
660 thresholds (e.g. removal of reads with an abundance of less than 10) that did not
661 consider the variation in artefact prevalence that can occur through differences in
662 sequencing depth (De Barba et al. 2014; Elbrecht and Leese 2015). Whilst some
663 studies circumvent this issue by using relative thresholds (2% of studies used
664 thresholds based on total read abundance, 18% sample read abundance and 9%
665 taxon abundance; Table S1), each of these methods is likely to have removed
666 artefacts to a different extent, introducing inconsistencies between datasets as a
667 consequence. This study shows how using different MSCTs can drastically affect
668 metabarcoding data, and in turn ecological interpretations of such data, therefore
669 highlighting the need for more stringent removal of artefact across metabarcoding
670 studies. Furthermore, the disparity in terminology and methodological descriptions
671 between studies identified in the literature search obviates confident inter-study
672 comparison and undermines an overall requirement for scientific transparency. By
673 comparing existing filtering methodologies, this study thus also provides effective
674 descriptions for such methods which can be applied to mitigate this disparity.

675

676 *Conclusions*

677 Here we have shown that artefacts persist in metabarcoding data even following
678 stringent lab and bioinformatic procedures. Although artefacts often occur in low
679 abundances, they can create a disproportionate representation of biodiversity and
680 produce misleading results, highlighting the need for read count filters. MSCTs
681 removed artefacts to differing extents, but combining sample-based thresholds with

682 removal of maximum taxon contamination provided an optimal outcome. Whilst the
683 optimal method was the same for all four datasets, thresholds applied differed due to
684 variation in sequencing depth and differential taxon amplification. The choice of
685 thresholds must thus depend on the individual study, taking into consideration the
686 sequencing depth, breadth of taxa amplified, artefact abundance and the
687 fundamental question under investigation. Control samples were crucial in assessing
688 filters and selecting appropriate thresholds, providing a means for assessing removal
689 of artefacts and target reads. We recommend that future metabarcoding studies
690 include mock communities and blanks, and, if possible, use taxa detected within
691 eDNA samples that can be used to identify artefacts in the resultant metabarcoding
692 data (e.g. marine taxa in inland samples) to facilitate identification of appropriate
693 thresholds. Given the broad variation in MSCTs applied to metabarcoding studies,
694 inconsistent results between these studies are inevitable. To mitigate erroneous
695 reporting of results and inconsistencies, effective guidance for best-practice filtering
696 of metabarcoding data for the ascertainment of conservative and accurate data
697 should be followed.

698

699 **Acknowledgements**

700 Analysis of dietary data from otter samples was conducted by LED and funded by
701 Knowledge Economy Skills Scholarship (KESS) and the Wildlife Trust of South and
702 West Wales who partnered this project. KESS is a pan-Wales higher level skills
703 initiative led by Bangor University on behalf of the HE sector in Wales and part
704 funded by the Welsh Government's European Social Fund (ESF) convergence
705 programme for West Wales and the Valleys. Analysis of dietary data from spider

706 samples was conducted by JPC and funded by the Biotechnology and Biological
707 Sciences Research Council through the South West Biosciences Doctoral Training
708 Partnership (grant BB/M009122/1). Sample collection of otter faeces was conducted
709 by Cardiff University Otter Project employees and placement students during
710 postmortems of otter carcasses. Samples for primer testing were provided by
711 Manuel Nicolaus from CEFAS (marine species), Shaun Leonard at Wild Trout Trust
712 (freshwater fish), Francois Edward and Monika Juergens at CEH (freshwater
713 invertebrates) and Derek Gow Consultancy (mammals). Thank you to Robert Reader
714 for allowing access to his farm for collection of terrestrial invertebrate samples and
715 thank you to Helen Hipperson and the NBAF team at University of Sheffield for their
716 initial training in bioinformatic analyses.

717

718 **Author contributions**

719 L.E.D, J.P.C, E.A.C and W.O.C.S conceived the ideas and oversaw the project;
720 L.E.D and J.P.C generated the data; A.M. carried out the sequencing and advised on
721 hypothetical implications for different data management strategies; L.E.D, J.P.C and
722 R.E.Y analysed the data; L.E.D led writing the manuscript. All authors commented
723 upon and contributed to the drafts and approved the final manuscript for publication.

724

725 **Data Accessibility**

726 The data relevant to this publication will be made publicly available via Dryad
727 following acceptance of the manuscript.

728

References

Abdelfattah, A., Li Destri Nicosia, M.G., Cacciola, S.O., Droby, S. and Schena, L. (2015). Metabarcoding analysis of fungal diversity in the phyllosphere and carposphere of olive (*Olea europaea*). PLoS ONE, 10(7), pp. 1–19. doi: 10.1371/journal.pone.0131069.

Aizpurua, O., Budinski, I., Georgiakakis, P., Gopalakrishnan, S., Ibañez, C., Mata, V., Rebelo, H., Russo, D., Szodoray-Parádi, F., Zhelyazkova, V., Zrncic, V., Gilbert, M.T.P. and Alberdi, A. (2018). Agriculture shapes the trophic niche of a bat preying on multiple pest arthropods across Europe: Evidence from DNA metabarcoding. Molecular Ecology, 27(3), pp. 815–825. doi: 10.1111/mec.14474.

Albaina, A., Aguirre, M., Abad, D., Santos, M. and Estonba, A. (2016). 18S rRNA V9 metabarcoding for diet characterization: A critical evaluation with two sympatric zooplanktivorous fish species. Ecology and Evolution, 6(6), pp. 1809–1824. doi: 10.1002/ece3.1986.

Alberdi, A., Aizpurua, O., Bohmann, K., Gopalakrishnan, S., Lynggaard, C., Nielsen, M. and Gilbert, M.T.P. (2019). Promises and pitfalls of using high-throughput sequencing for diet analysis. Molecular Ecology Resources, 19(2), pp. 327–348. doi: 10.1111/1755-0998.12960.

Alberdi, A., Aizpurua, O., Gilbert, M.T.P. and Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution, 9(1), pp. 134–147. doi: 10.1111/2041-210X.12849.

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W. and Iannone, R. (2020). rmarkdown: Dynamic Documents for R. R package version 2.3. Available at: https://github.com/rstudio/rmarkdown.

Alsos, I.G., Lammers, Y., Yoccoz, N.G., Jørgensen, T., Sjögren, P., Gielly, L. and Edwards, M.E. (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. PLoS ONE, 13(4), pp. 1–23. doi: 10.1371/journal.pone.0195403.

Andruszkiewicz, E.A., Starks, H.A., Chavez, F.P., Sassoubre, L.M., Block, B.A. and Boehm, A.B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. PLoS ONE, 12(4), pp. 1–20. doi: 10.1371/journal.pone.0176343.

Andújar, C., Arribas, P., Gray, C., Bruce, C., Woodward, G., Yu, D.W. and Vogler, A.P. (2018). Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. Molecular Ecology, 27(1), pp. 146–166. doi: 10.1111/mec.14410.

Anslan, S., Nilsson, R.H., Wurzbacher, C., Baldrian, P., Tedersoo, L. and Bahram, M. (2018). Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. MycoKeys, 39, pp. 29–40. doi: 10.3897/mycokeys.39.28109.

Arrizabalaga-Escudero, A., Clare, E.L., Salsamendi, E., Alberdi, A., Garin, I., Aihartza, J. and Goiti, U. (2018). Assessing niche partitioning of co-occurring sibling

bat species by DNA metabarcoding. Molecular Ecology, 27(5), pp. 1273–1283. doi: 10.1111/mec.14508.

Arulandhu, A.J., Staats, M., Hagelaar, R., Voorhuijzen, M.M., Prins, T.W., Scholtens, I., Costessi, A., Duijsings, D., Rechenmann, F., Gaspar, F.B., Barreto Crespo, M.T., Holst-Jensen, A., Birck, M., Burns, M., Haynes, E., Hochegger, R., Klingl, A., Lundberg, L., Natale, C., et al. (2017). Development and validation of a multi-locus DNA metabarcoding method to identify endangered species in complex samples. GigaScience, 6(10), pp. 1–18. doi: 10.1093/gigascience/gix080.

Aylagas, E., Borja, Á., Irigoien, X. and Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. Frontiers in Marine Science, 3(JUN), pp. 1–12. doi: 10.3389/fmars.2016.00096.

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E. and Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. Molecular Ecology Resources, 14(2), pp. 306–323. doi: 10.1111/1755-0998.12188.

Batovska, J., Lynch, S.E., Cogan, N.O.I., Brown, K., Darbro, J.M., Kho, E.A. and Blacket, M.J. (2018). Effective mosquito and arbovirus surveillance using metabarcoding. Molecular Ecology Resources, 18(1), pp. 32–40. doi: 10.1111/1755-0998.12682.

Beermann, A.J., Zizka, V.M.A., Elbrecht, V., Baranov, V. and Leese, F. (2018). DNA metabarcoding reveals the complex and hidden responses of chironomids to multiple stressors. Environmental Sciences Europe, 30(1). Available at: https://doi.org/10.1186/s12302-018-0157-x.

Bell, K.L., Burgess, K.S., Botsch, J.C., Dobbs, E.K., Read, T.D. and Brosi, B.J. (2019). Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. Molecular Ecology, 28(2), pp. 431–455. doi: 10.1111/mec.14840.

Bell, K.L., Fowler, J., Burgess, K.S., Dobbs, E.K., Gruenewald, D., Lawley, B., Morozumi, C. and Brosi, B.J. (2017). Applying Pollen DNA Metabarcoding to the Study of Plant–Pollinator Interactions. Applications in Plant Sciences, 5(6), p. 1600124. doi: 10.3732/apps.1600124.

Bell, K.L., Loeffler, V.M. and Brosi, B.J. (2017). An rbcL Reference Library to Aid in the Identification of Plant Species Mixtures by DNA Metabarcoding . Applications in Plant Sciences, 5(3), p. 1600110. doi: 10.3732/apps.1600110.

Beng, K.C., Tomlinson, K.W., Shen, X.H., Surget-Groba, Y., Hughes, A.C., Corlett, R.T. and Slik, J.W.F. (2016). The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. Scientific Reports, 6(April), pp. 1–13. Available at: http://dx.doi.org/10.1038/srep24965.

Berry, T.E., Osterrieder, S.K., Murray, D.C., Coghlan, M.L., Richardson, A.J., Grealy, A.K., Stat, M., Bejder, L. and Bunce, M. (2017). DNA metabarcoding for diet analysis and biodiversity: A case study using the endangered Australian sea lion (Neophoca cinerea). Ecology and Evolution, 7(14), pp. 5435–5453. doi: 10.1002/ece3.3123.

Bessey, C., Jarman, S.N., Stat, M., Rohner, C.A., Bunce, M., Koziol, A., Power, M., Rambahiniarison, J.M., Ponzo, A., Richardson, A.J. and Berry, O. (2019). DNA metabarcoding assays reveal a diverse prey assemblage for Mobula rays in the Bohol Sea, Philippines. Ecology and Evolution, 9(5), pp. 2459–2474. doi: 10.1002/ece3.4858.

Bjørnsgaard Aas, A., Davey, M.L. and Kauserud, H. (2017). ITS all right mama: investigating the formation of chimeric sequences in the ITS2 region by DNA metabarcoding analyses of fungal mock communities of different complexities. Molecular Ecology Resources, 17(4), pp. 730–741. doi: 10.1111/1755-0998.12622.

Blanckenhorn, W.U., Rohner, P.T., Bernasconi, M. V., Haugstetter, J. and Buser, A. (2016). Is qualitative and quantitative metabarcoding of dung fauna biodiversity feasible? Environmental Toxicology and Chemistry, 35(8), pp. 1970–1977. doi: 10.1002/etc.3275.

Bohmann, K., Monadjem, A., Noer, C., Rasmussen, M., Zeale, M.R.K., Clare, E., Jones, G., Willerslev, E. and Gilbert, M.T.P. (2011). Molecular diet analysis of two African free-tailed bats (molossidae) using high throughput sequencing. PLoS ONE, 6(6). doi: 10.1371/journal.pone.0021441.

Borrell, Y.J., Miralles, L., Do Huu, H., Mohammed-Geba, K. and Garcia-Vazquez, E. (2017). DNA in a bottle - Rapid metabarcoding survey for early alerts of invasive species in ports. PLoS ONE, 12(9), pp. 1–17. doi: 10.1371/journal.pone.0183347.

Bowser, A.K., Diamond, A.W. and Addison, J.A. (2013). From puffins to plankton: A DNA-based analysis of a seabird food chain in the Northern Gulf of Maine. PLoS ONE, 8(12), pp. 1–16. doi: 10.1371/journal.pone.0083152.

Boyer, S., Cruickshank, R.H. and Wratten, S.D. (2015). Faeces of generalist predators as 'biodiversity capsules': A new tool for biodiversity assessment in remote and inaccessible habitats. Food Webs, 3, pp. 1–6. Available at: http://dx.doi.org/10.1016/j.fooweb.2015.02.001.

Brandon-Mong, G.J., Gan, H.M., Sing, K.W., Lee, P.S., Lim, P.E. and Wilson, J.J. (2015). DNA metabarcoding of insects and allies: An evaluation of primers and pipelines. Bulletin of Entomological Research, 105(6), pp. 717–727. doi: 10.1017/S0007485315000681.

Brannock, P.M., Ortmann, A.C., Moss, A.G. and Halanych, K.M. (2016). Metabarcoding reveals environmental factors influencing spatio-temporal variation in pelagic micro-eukaryotes. Molecular ecology, 25(15), pp. 3593–3604. doi: 10.1111/mec.13709.

Brannock, P.M., Waits, D.S., Sharma, J. and Halanych, K.M. (2014). High-throughput sequencing characterizes intertidal meiofaunal communities in Northern

Gulf of Mexico (Dauphin Island and Mobile Bay, Alabama). Biological Bulletin, 227(2), pp. 161–174. doi: 10.1086/BBLv227n2p161.

Braukmann, T.W.A., Ivanova, N. V., Prosser, S.W.J., Elbrecht, V., Steinke, D., Ratnasingham, S., de Waard, J.R., Sones, J.E., Zakharov, E. V. and Hebert, P.D.N. (2019). Metabarcoding a diverse arthropod mock community. Molecular Ecology Resources, 19(3), pp. 711–727. doi: 10.1111/1755-0998.13008.

Bray, J.R. and Curtis, J.T. (1957). An ordination of the upland forest communities of southern Wisconsin. Ecological monographs, 27, pp. 325–349.

Brown, E.A., Chain, F.J.J., Crease, T.J., Macisaac, H.J. and Cristescu, M.E. (2015). Divergence thresholds and divergent biodiversity estimates: Can metabarcoding reliably describe zooplankton communities? Ecology and Evolution, 5(11), pp. 2234–2251. doi: 10.1002/ece3.1485.

Burgar, J.M., Murray, D.C., Craig, M.D., Haile, J., Houston, J., Stokes, V. and Bunce, M. (2014). Who's for dinner? High-throughput sequencing reveals bat dietary differentiation in a biodiversity hotspot where prey taxonomy is largely undescribed. Molecular Ecology, 23(15), pp. 3605–3617. doi: 10.1111/mec.12531.

Bylemans, J., Gleeson, D.M., Hardy, C.M. and Furlan, E. (2018). Toward an ecoregion scale evaluation of eDNA metabarcoding primers: A case study for the freshwater fish biodiversity of the Murray–Darling Basin (Australia). Ecology and Evolution, 8(17), pp. 8697–8712. doi: 10.1002/ece3.4387.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10, pp. 1–9. doi: 10.1186/1471-2105-10-421.

Carew, M.E., Coleman, R.A. and Hoffmann, A.A. (2018). Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? PeerJ, 2018(6). doi: 10.7717/peerj.4980.

Casey, J.M., Meyer, C.P., Morat, F., Brandl, S.J., Planes, S. and Parravicini, V. (2019). Reconstructing hyperdiverse food webs: Gut content metabarcoding as a tool to disentangle trophic interactions on coral reefs. Methods in Ecology and Evolution, 10(8), pp. 1157–1170. doi: 10.1111/2041-210X.13206.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17), pp. i884–i890. doi: 10.1093/bioinformatics/bty560.

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J.C., Bonin, A., Taberlet, P. and Pont, D. (2016). Spatial Representativeness of Environmental DNA Metabarcoding Signal for Fish Biodiversity Assessment in a Natural Freshwater System. PLoS ONE, 11(6), pp. 1–19. doi: 10.1371/journal.pone.0157366.

Clare, E.L., Chain, F.J.J., Littlefair, J.E. and Cristescu, M.E. (2016). The effects of parameter choice on defining molecular operational taxonomic units and resulting ecological analyses of metabarcoding data. Genome, 59(11), pp. 981–990. doi: 10.1139/gen-2015-0184.

892 Collins, R.A., Bakker, J., Wangensteen, O.S., Soto, A.Z., Corrigan, L., Sims, D.W.,
893 Genner, M.J. and Mariani, S. (2019). Non-specific amplification compromises
894 environmental DNA metabarcoding with COI. Methods in Ecology and Evolution,
895 10(11), pp. 1985–2001. doi: 10.1111/2041-210X.13276.

896 Cordier, T., Esling, P., Lejzerowicz, F., Visco, J., Ouadahi, A., Martins, C.,
897 Cedhagen, T. and Pawlowski, J. (2017). Predicting the Ecological Quality Status of
898 Marine Environments from eDNA Metabarcoding Data Using Supervised Machine
899 Learning. Environmental Science and Technology, 51(16), pp. 9118–9126. doi:
900 10.1021/acs.est.7b01518.

901 Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J. and Arnaud-
902 Haond, S. (2015). Metabarcoding is powerful yet still blind: A comparative analysis of
903 morphological and molecular surveys of seagrass communities. PLoS ONE, 10(2),
904 pp. 1–26. doi: 10.1371/journal.pone.0117562.

905 Creedy, T.J., Norman, H., Tang, C.Q., Qing Chin, K., Andujar, C., Arribas, P.,
906 O'Connor, R.S., Carvell, C., Notton, D.G. and Vogler, A.P. (2020). A validated
907 workflow for rapid taxonomic assignment and monitoring of a national fauna of bees
908 (Apiformes) using high throughput DNA barcoding. Molecular Ecology Resources,
909 20(1), pp. 40–53. doi: 10.1111/1755-0998.13056.

910 Czurda, S., Smelik, S., Preuner-Stix, S., Nogueira, F. and Lion, T. (2016).
911 Occurrence of fungal DNA contamination in PCR reagents: Approaches to control
912 and decontamination. Journal of Clinical Microbiology, 54(1), pp. 148–152. doi:
913 10.1128/JCM.02112-15.

914

915 Deagle, B.E., Chiaradia, A., McInnes, J. and Jarman, S.N. (2010). Pyrosequencing
916 faecal DNA to determine diet of little penguins: Is what goes in what comes out?
917 Conservation Genetics, 11(5), pp. 2039–2048. doi: 10.1007/s10592-010-0096-6.

918 Deagle, B.E., Clarke, L.J., Kitchener, J.A., Polanowski, A.M. and Davidson, A.T.
919 (2018). Genetic monitoring of open ocean biodiversity: An evaluation of DNA
920 metabarcoding for processing continuous plankton recorder samples. Molecular
921 Ecology Resources, 18(3), pp. 391–406. doi: 10.1111/1755-0998.12740.

922 Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. and Taberlet, P. (2014).
923 DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect
924 match. Biology Letters, 10(9), pp. 20140562–20140562. Available at:
925 http://rsbl.royalsocietypublishing.org/cgi/doi/10.1098/rsbl.2014.0562.

926 Deagle, B.E., Kirkwood, R. and Jarman, S.N. (2009). Analysis of Australian fur seal
927 diet by pyrosequencing prey DNA in faeces. Molecular Ecology, 18(9), pp. 2022–
928 2038. doi: 10.1111/j.1365-294X.2009.04158.x.

929 Deagle, B.E., Thomas, A.C., McInnes, J.C., Clarke, L.J., Vesterinen, E.J., Clare,
930 E.L., Kartzinel, T.R. and Eveson, J.P. (2019). Counting with DNA in metabarcoding
931 studies: How should we convert sequence reads to dietary data? Molecular Ecology,
932 28(2), pp. 391–406. doi: 10.1111/mec.14734.

933 Deagle, B.E., Thomas, A.C., Shaffer, A.K., Trites, A.W. and Jarman, S.N. (2013).
934 Quantifying sequence proportions in a DNA-based diet study using Ion Torrent
935 amplicon sequencing: Which counts count? Molecular Ecology Resources, 13(4), pp.
936 620–633. doi: 10.1111/1755-0998.12103.

937 Deagle, B.E., Tollit, D.J., Jarman, S.N., Hindell, M.A., Trites, A.W. and Gales, N.J.
938 (2005). Molecular scatology as a tool to study diet: Analysis of prey DNA in scats
939 from captive Steller sea lions. Molecular Ecology, 14(6), pp. 1831–1842. doi:
940 10.1111/j.1365-294X.2005.02531.x.

941 Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt,
942 F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E. and Bernatchez, L.
943 (2017). Environmental DNA metabarcoding: Transforming how we survey animal
944 and plant communities. Molecular Ecology, 26(21), pp. 5872–5895. doi:
945 10.1111/mec.14350.

946 Divoll, T.J., Brown, V.A., Kinne, J., McCracken, G.F. and O'Keefe, J.M. (2018).
947 Disparities in second-generation DNA metabarcoding results exposed with
948 accessible and repeatable workflows. Molecular Ecology Resources, 18(3), pp. 590–
949 601. doi: 10.1111/1755-0998.12770.

950 Djurhuus, A., Pitz, K., Sawaya, N.A., Rojas-Márquez, J., Michaud, B., Montes, E.,
951 Muller-Karger, F. and Breitbart, M. (2018). Evaluation of marine zooplankton
952 community structure through environmental DNA metabarcoding. Limnology and
953 Oceanography: Methods, 16(4), pp. 209–221. doi: 10.1002/lom3.10237.

954 Dunn, J.C., Stockdale, J.E., Moorhouse-Gann, R.J., McCubbin, A., Hipperson, H.,
955 Morris, A.J., Grice, P. V. and Symondson, W.O.C. (2018). The decline of the Turtle
956 Dove: Dietary associations with body condition and competition with other columbids
957 analysed using high-throughput sequencing. Molecular Ecology, 27(16), pp. 3386–
958 3407. doi: 10.1111/mec.14766.

959 Edgar, R. ([no date]). Usearch V11: Generating OTUs and ZOTUs. Available at:
960 https://www.drive5.com/usearch/manual/pipe_otus.html. [Accessed: 27 June 2020].

961 Edgar, R.C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS
962 amplicon sequencing. bioRxiv, . doi: https://doi.org/10.1101/003723.

963 Elbrecht, V. and Leese, F. (2015). Can DNA-based ecosystem assessments quantify
964 species abundance? Testing primer bias and biomass-sequence relationships with
965 an innovative metabarcoding protocol. PLoS ONE, 10(7), pp. 1–16. doi:
966 10.1371/journal.pone.0130324.

967 Elbrecht, V. and Leese, F. (2017). Validation and Development of COI
968 Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. Frontiers
969 in Environmental Science, 5(April), pp. 1–11. Available at:
970 http://journal.frontiersin.org/article/10.3389/fenvs.2017.00011/full.

971 Elbrecht, V., Vamos, E.E., Meissner, K., Aroviita, J. and Leese, F. (2017). Assessing
972 strengths and weaknesses of DNA metabarcoding-based macroinvertebrate

973 identification for routine stream monitoring. Methods in Ecology and Evolution, 8(10),
974 pp. 1265–1275. doi: 10.1111/2041-210X.12789.

975 Elbrecht, V., Vamos, E.E., Steinke, D. and Leese, F. (2018). Estimating intraspecific
976 genetic diversity from community DNA metabarcoding data. PeerJ, 2018(4), pp. 1–
977 13. doi: 10.7717/peerj.4644.

978 Emilson, C.E., Thompson, D.G., Venier, L.A., Porter, T.M., Swystun, T., Chartrand,
979 D., Capell, S. and Hajibabaei, M. (2017). DNA metabarcoding and morphological
980 macroinvertebrate metrics reveal the same changes in boreal watersheds across an
981 environmental gradient. Scientific Reports, 7(1), pp. 1–11. Available at:
982 http://dx.doi.org/10.1038/s41598-017-13157-x.

983 Evans, N.T., Olds, B.P., Renshaw, M.A., Turner, C.R., Li, Y., Jerde, C.L., Mahon,
984 A.R., Pfrender, M.E., Lamberti, G.A. and Lodge, D.M. (2016). Quantification of
985 mesocosm fish and amphibian species diversity via environmental DNA
986 metabarcoding. Molecular Ecology Resources, 16(1), pp. 29–41. doi: 10.1111/1755-
987 0998.12433.

988 Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M.,
989 Gielly, L., Lopes, C.M., Boyer, F., Pompanon, F., Rayé, G. and Taberlet, P. (2015).
990 Replication levels, false presences and the estimation of the presence/absence from
991 eDNA metabarcoding data. Molecular Ecology Resources, 15(3), pp. 543–556. doi:
992 10.1111/1755-0998.12338.

993 Fletcher, L.M., Zaiko, A., Atalah, J., Richter, I., Dufour, C.M., Pochon, X., Wood, S.A.
994 and Hopkins, G.A. (2017). Bilge water as a vector for the spread of marine pests: a
995 morphological, metabarcoding and experimental assessment. Biological Invasions,
996 19(10), pp. 2851–2867. doi: 10.1007/s10530-017-1489-y.

997 Folmer, O., Black, M., Hoeh, W., Lutz, R. and Vrijenhoek, R. (1994). DNA primers for
998 amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan
999 invertebrates. Molecular Marine Biology and Biotechnology, 3(5), pp. 294–299. doi:
1000 10.1371/journal.pone.0013102.

1001 Fonseca, V.G., Sinniger, F., Gaspar, J.M., Quince, C., Creer, S., Power, D.M., Peck,
1002 L.S. and Clark, M.S. (2017). Revealing higher than expected meiofaunal diversity in
1003 Antarctic sediments: A metabarcoding approach. Scientific Reports, 7(1), pp. 1–11.
1004 Available at: http://dx.doi.org/10.1038/s41598-017-06687-x.

1005 Frontalini, F., Greco, M., Di Bella, L., Lejzerowicz, F., Reo, E., Caruso, A.,
1006 Cosentino, C., Maccotta, A., Scopelliti, G., Nardelli, M.P., Losada, M.T., Armynot du
1007 Châtelet, E., Coccioni, R. and Pawlowski, J. (2018). Assessing the effect of mercury
1008 pollution on cultured benthic foraminifera community using morphological and eDNA
1009 metabarcoding approaches. Marine Pollution Bulletin, 129(2), pp. 512–524. Available
1010 at: http://dx.doi.org/10.1016/j.marpolbul.2017.10.022.

1011 Galal-Khallaf, A., Osman, A.G.M., Carleos, C.E., Garcia-Vazquez, E. and Borrell,
1012 Y.J. (2016). A case study for assessing fish traceability in Egyptian aquafeed
1013 formulations using pyrosequencing and metabarcoding. Fisheries Research, 174,
1014 pp. 143–150. Available at: http://dx.doi.org/10.1016/j.fishres.2015.09.009.

1015 Galan, M., Pons, J.B., Tournayre, O., Pierre, É., Leuchtmann, M., Pontier, D. and
1016 Charbonnel, N. (2018). Metabarcoding for the parallel identification of several
1017 hundred predators and their prey: Application to bat species diet analysis. Molecular
1018 Ecology Resources, 18(3), pp. 474–489. doi: 10.1111/1755-0998.12749.

1019 Gebremedhin, B., Flagstad, O., Bekele, A., Chala, D., Bakkestuen, V., Boessenkool,
1020 S., Popp, M., Gussarova, G., Schrøder-Nielsen, A., Nemomissa, S., Brochmann, C.,
1021 Stenseth, N.C. and Epp, L.S.E. (2016). DNA metabarcoding reveals diet overlap
1022 between the endangered walia ibex and domestic goats - Implications for
1023 conservation. PLoS ONE, 11(7). doi: 10.1371/journal.pone.0159133.

1024 Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M. and De Groot, G.A. (2015). Not all
1025 are free-living: High-throughput DNA metabarcoding reveals a diverse community of
1026 protists parasitizing soil metazoa. Molecular Ecology, 24(17), pp. 4556–4569. doi:
1027 10.1111/mec.13238.

1028 Gerwing, T.G., Kim, J.-H., Hamilton, D.J., Barbeau, M.A. and Addison, J.A. (2016).
1029 Diet reconstruction using next-generation sequencing increases the known
1030 ecosystem usage by a shorebird. The Auk, 133(2), pp. 168–177. doi: 10.1642/auk-
1031 15-176.1.

1032 Gosselin, E.N., Lonsinger, R.C. and Waits, L.P. (2017). Comparing morphological
1033 and molecular diet analyses and fecal DNA sampling protocols for a terrestrial
1034 carnivore. Wildlife Society Bulletin, 41(2), pp. 362–369. doi: 10.1002/wsb.749.

1035 Granquist, S.M., Esparza-Salas, R., Hauksson, E., Karlsson, O. and Angerbjörn, A.
1036 (2018). Fish consumption of harbour seals (Phoca vitulina) in north western Iceland
1037 assessed by DNA metabarcoding and morphological analysis. Polar Biology, 41(11),
1038 pp. 2199–2210. Available at: https://doi.org/10.1007/s00300-018-2354-x.

1039 Grealy, A.C., McDowell, M.C., Scofield, P., Murray, D.C., Fusco, D.A., Haile, J.,
1040 Prideaux, G.J. and Bunce, M. (2015). A critical evaluation of how ancient DNA bulk
1041 bone metabarcoding complements traditional morphological analysis of fossil
1042 assemblages. Quaternary Science Reviews, 128, pp. 37–47. doi:
1043 10.1016/j.quascirev.2015.09.014.

1044 Greiman, S.E., Cook, J.A., Tkach, V. V., Hoberg, E.P., Menning, D.M., Hope, A.G.,
1045 Sonsthagen, S.A. and Talbot, S.L. (2018). Museum metabarcoding: A novel method
1046 revealing gut helminth communities of small mammals across space and time.
1047 International Journal for Parasitology, 48(13), pp. 1061–1070. doi:
1048 10.1016/j.ijpara.2018.08.001.

1049 Grey, E.K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K.L.,
1050 Lacoursière-Roussel, A., Leong, S.C.Y., Li, Y., Olds, B., Pfrender, M.E., Prowse,
1051 T.A.A., Renshaw, M.A. and Lodge, D.M. (2018). Effects of sampling effort on
1052 biodiversity patterns estimated from environmental DNA metabarcoding surveys.
1053 Scientific Reports, 8(1), pp. 2–11. doi: 10.1038/s41598-018-27048-2.

1054 Guardiola, M., Uriz, M.J., Taberlet, P., Coissac, E., Wangensteen, O.S. and Turon,
1055 X. (2015). Deep-sea, deep-sequencing: Metabarcoding extracellular DNA from

1056 sediments of marine canyons. PLoS ONE, 10(10). doi:
1057 10.1371/journal.pone.0139633.

1058 Guardiola, M., Wangensteen, O.S., Taberlet, P., Coissac, E., Uriz, M.J. and Turon,
1059 X. (2016). Spatio-temporal monitoring of deep-sea communities using
1060 metabarcoding of sediment DNA and RNA. PeerJ, 2016(12), pp. 1–31. doi: 10.7717/
1061 peerj.2807.

1062 Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C. and Baird, D.J. (2011).
1063 Environmental barcoding: A next-generation sequencing approach for biomonitoring
1064 applications using river benthos. PLoS ONE, 6(4). doi:
1065 10.1371/journal.pone.0017497.

1066 Hänfling, B., Handley, L.L., Read, D.S., Hahn, C., Li, J., Nichols, P., Blackman, R.C.,
1067 Oliver, A. and Winfield, I.J. (2016). Environmental DNA metabarcoding of lake fish
1068 communities reflects long-term data from established survey methods. Molecular
1069 Ecology, 25(13), pp. 3101–3119. doi: 10.1111/mec.13660.

1070 Hardy, N., Berry, T., Kelaher, B.P., Goldsworthy, S.D., Bunce, M., Coleman, M.A.,
1071 Gillanders, B.M., Connell, S.D., Blewitt, M. and Figueira, W. (2017). Assessing the
1072 trophic ecology of top predators across a recolonisation frontier using DNA
1073 metabarcoding of diets. Marine Ecology Progress Series, 573(April), pp. 237–254.
1074 doi: 10.3354/meps12165.

1075 Harms-Tuohy, C.A., Schizas, N. V. and Appeldoorn, R.S. (2016). Use of DNA
1076 metabarcoding for stomach content analysis in the invasive lionfish pterois volitans in
1077 Puerto Rico. Marine Ecology Progress Series, 558, pp. 181–191. doi:
1078 10.3354/meps11738.

1079 Harper, L.R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H.C., Gough, K.C.,
1080 Lewis, E., Adams, I.P., Brotherton, P., Phillips, S. and Hänfling, B. (2018). Needle in
1081 a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection
1082 of the great crested newt (Triturus cristatus). Ecology and Evolution, 8(12), pp.
1083 6330–6341. doi: 10.1002/ece3.4013.

1084 Hart, M.L., Meyer, A., Johnson, P.J. and Ericsson, A.C. (2015). Comparative
1085 evaluation of DNA extraction methods from feces of multiple host species for
1086 downstream next-generation sequencing. PLoS ONE, 10(11), pp. 1–16. doi:
1087 10.1371/journal.pone.0143334.

1088 Hatzenbuhler, C., Kelly, J.R., Martinson, J., Okum, S. and Pilgrim, E. (2017).
1089 Sensitivity and accuracy of high-throughput metabarcoding methods for early
1090 detection of invasive fish species. Scientific Reports, 7(April 2016), pp. 1–10. doi:
1091 10.1038/srep46393.

1092 Hawlitschek, O., Fernández-González, A., Balmori-de la Puente, A. and Castresana,
1093 J. (2018). A pipeline for metabarcoding and diet analysis from fecal samples
1094 developed for a small semi-aquatic mammal. PLoS ONE, 13(8), pp. 1–19. doi:
1095 10.1371/journal.pone.0201763.

1096 Hibert, F., Taberlet, P., Chave, J., Scotti-Saintagne, C., Sabatier, D. and Richard-
1097 Hansen, C. (2013). Unveiling the Diet of Elusive Rainforest Herbivores in Next
1098 Generation Sequencing Era? The Tapir as a Case Study. PLoS ONE, 8(4). doi:
1099 10.1371/journal.pone.0060799.

1100 Hope, P.R., Bohmann, K., Gilbert, M.T.P., Zepeda-Mendoza, M.L., Razgour, O. and
1101 Jones, G. (2014). Second generation sequencing and morphological faecal analysis
1102 reveal unexpected foraging behaviour by Myotis nattereri (Chiroptera,
1103 Vespertilionidae) in winter. Frontiers in Zoology, 11(1), pp. 1–15. doi: 10.1186/1742-
1104 9994-11-39.

1105 Huson, D.H., Auch, A.F., Qi, J. and Schuster, S.C. (2007). MEGAN analysis of
1106 metagenomic data. Genome Research, 17(3), pp. 377–386. doi:
1107 10.1101/gr.5969107.

1108 Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh,
1109 H.J. and Tappu, R. (2016). MEGAN Community Edition - Interactive Exploration and
1110 Analysis of Large-Scale Microbiome Sequencing Data. PLoS Computational Biology,
1111 12(6), pp. 1–12. doi: 10.1371/journal.pcbi.1004957.

1112 Illumina ([no date]). Specifications for the MiSeq System. Available at:
1113 https://emea.illumina.com/systems/sequencing-platforms/miseq/specifications.html
1114 [Accessed: 27 June 2020].

1115 Iwanowicz, D.D., Vandergast, A.G., Cornman, R.S., Adams, C.R., Kohn, J.R.,
1116 Fisher, R.N. and Brehme, C.S. (2016). Metabarcoding of fecal samples to determine
1117 herbivore diets: A case study of the endangered Pacific pocket mouse. PLoS ONE,
1118 11(11). doi: 10.1371/journal.pone.0165366.

1119 Jakubavičiute, E., Bergström, U., Eklöf, J.S., Haenel, Q. and Bourlat, S.J. (2017).
1120 DNA metabarcoding reveals diverse diet of the three-spined stickleback in a coastal
1121 ecosystem. PLoS ONE, 12(10), pp. 1–16. doi: 10.1371/journal.pone.0186929.

1122 Jarman, S.N., McInnes, J.C., Faux, C., Polanowski, A.M., Marthick, J., Deagle, B.E.,
1123 Southwell, C. and Emmerson, L. (2013). Adélie penguin population diet monitoring
1124 by analysis of food DNA in scats. PLoS ONE, 8(12). doi:
1125 10.1371/journal.pone.0082227.

1126 Jeanniard-Du-Dot, T., Thomas, A.C., Cherel, Y., Trites, A.W. and Guinet, C. (2017).
1127 Combining hard-part and DNA analyses of scats with biologging and stable isotopes
1128 can reveal different diet compositions and feeding strategies within a fur seal
1129 population. Marine Ecology Progress Series, 584, pp. 1–16. doi:
1130 10.3354/meps12381.

1131 Jensen, R.H., Mollerup, S., Mourier, T., Hansen, T.A., Fridholm, H., Nielsen, L.P.,
1132 Willerslev, E., Hansen, A.J. and Vinner, L. (2015). Target-dependent enrichment of
1133 virions determines the reduction of high-throughput sequencing in virus discovery.
1134 PLoS ONE, 10(4), pp. 1–18. doi: 10.1371/journal.pone.0122636.

1135 Jusino, M.A., Banik, M.T., Palmer, J.M., Wray, A.K., Xiao, L., Pelton, E., Barber,
1136 J.R., Kawahara, A.Y., Gratton, C., Peery, M.Z. and Lindner, D.L. (2019). An

improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. Molecular Ecology Resources, 19(1), pp. 176–190. doi: 10.1111/1755-0998.12951.

Kartzinel, T.R., Chen, P.A., Coverdale, T.C., Erickson, D.L., Kress, W.J., Kuzmina, M.L., Rubenstein, D.I., Wang, W. and Pringle, R.M. (2015). DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. Proceedings of the National Academy of Sciences of the United States of America, 112(26), pp. 8019–8024. doi: 10.1073/pnas.1503283112.

Kaunisto, K.M., Roslin, T., Sääksjärvi, I.E. and Vesterinen, E.J. (2017). Pellets of proof: First glimpse of the dietary composition of adult odonates as revealed by metabarcoding of feces. Ecology and Evolution, 7(20), pp. 8588–8598. doi: 10.1002/ece3.3404.

Kelly, R.P., Port, J.A., Yamahara, K.M. and Crowder, L.B. (2014). Using environmental DNA to census marine fishes in a large mesocosm. PLoS ONE, 9(1). doi: 10.1371/journal.pone.0086175.

Keskin, E., Unal, E.M. and Atar, H.H. (2016). Detection of rare and invasive freshwater fish species using eDNA pyrosequencing: Lake Iznik ichthyofauna revised. Biochemical Systematics and Ecology, 67, pp. 29–36. Available at: http://dx.doi.org/10.1016/j.bse.2016.05.020.

King, R.A., Read, D.S., Traugott, M. and Symondson, W.O.C. (2008). Molecular analysis of predation: A review of best practice for DNA-based approaches. Molecular Ecology, 17(4), pp. 947–963. doi: 10.1111/j.1365-294X.2007.03613.x.

Kitson, J.J.N., Hahn, C., Sands, R.J., Straw, N.A., Evans, D.M. and Lunt, D.H. (2019). Detecting host–parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding. Molecular Ecology, 28(2), pp. 471–483. doi: 10.1111/mec.14518.

Klymus, K.E., Marshall, N.T. and Stepien, C.A. (2017). Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. PLoS ONE, 12(5), pp. 1–24. doi: 10.1371/journal.pone.0177643.

Klymus, K.E., Richter, C.A., Thompson, N. and Hinck, J.E. (2017). Metabarcoding of environmental DNA samples to explore the use of uranium mine containment ponds as a water source for wildlife. Diversity, 9(4), pp. 1–18. doi: 10.3390/d9040054.

Krehenwinkel, H., Wolf, M., Lim, J.Y., Rominger, A.J., Simison, W.B. and Gillespie, R.G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific Reports, 7(1), pp. 1–12. doi: 10.1038/s41598-017-17333-x.

Lacoursière-Roussel, A., Howland, K., Normandeau, E., Grey, E.K., Archambault, P., Deiner, K., Lodge, D.M., Hernandez, C., Leduc, N. and Bernatchez, L. (2018). EDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. International Journal of Business Innovation and Research, 17(3), pp. 7763–7777. doi: 10.1002/ece3.4213.

1178 Lanzén, A., Lekang, K., Jonassen, I., Thompson, E.M. and Troedsson, C. (2017).
1179 DNA extraction replicates improve diversity and compositional dissimilarity in
1180 metabarcoding of eukaryotes in marine sediments. PLoS ONE, 12(6), pp. 1–18. doi:
1181 10.1371/journal.pone.0179443.

1182 Laroche, O., Wood, S.A., Tremblay, L.A., Lear, G., Ellis, J.I. and Pochon, X. (2017).
1183 Metabarcoding monitoring analysis: The pros and cons of using co-extracted
1184 environmental DNA and RNA data to assess offshore oil production impacts on
1185 benthic communities. PeerJ, 2017(5). doi: 10.7717/peerj.3347.

1186 Leonard, J.A., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne,
1187 R.K. and Fleischer, R.C. (2007). Animal DNA in PCR reagents plagues ancient DNA
1188 research. Journal of Archaeological Science, 34(9), pp. 1361–1366. doi:
1189 10.1016/j.jas.2006.10.023.

1190 Leray, M. and Knowlton, N. (2015). DNA barcoding and metabarcoding of
1191 standardized samples reveal patterns of marine benthic diversity. Proceedings of the
1192 National Academy of Sciences of the United States of America, 112(7), pp. 2076–
1193 2081. doi: 10.1073/pnas.1424997112.

1194 Leray, M. and Knowlton, N. (2017). Random sampling causes the low reproducibility
1195 of rare eukaryotic OTUs in Illumina COI metabarcoding. PeerJ, 2017(3), pp. 1–27.
1196 doi: 10.7717/peerj.3006.

1197 Leray, M., Meyer, C.P. and Mills, S.C. (2015). Metabarcoding dietary analysis of
1198 coral dwelling predatory fish demonstrates the minor contribution of coralmutualists
1199 to their highly partitioned, generalist diet. PeerJ, 2015(6), pp. 1–19. doi:
1200 10.7717/peerj.1047.

1201 Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm,
1202 J.T. and Machida, R.J. (2013). A new versatile primer set targeting a short fragment
1203 of the mitochondrial COI region for metabarcoding metazoan diversity: Application
1204 for characterizing coral reef fish gut contents. Frontiers in Zoology, 10(1), pp. 1–14.
1205 doi: 10.1186/1742-9994-10-34.

1206 Li, W., Wang, M.M., Wang, X.G., Cheng, X.L., Guo, J.J., Bian, X.M. and Cai, L.
1207 (2016). Fungal communities in sediments of subtropical Chinese seas as estimated
1208 by DNA metabarcoding. Scientific Reports, 6, pp. 1–9. Available at: http://dx.doi.org/
1209 10.1038/srep26528.

1210 Lim, N.K.M., Tay, Y.C., Srivathsan, A., Tan, J.W.T., Kwik, J.T.B., Baloğlu, B., Meier,
1211 R. and Yeo, D.C.J. (2016). Next-generation freshwater bioassessment: eDNA
1212 metabarcoding with a conserved metazoan primer reveals species-rich and
1213 reservoir-specific communities. Royal Society Open Science, 3(11). doi:
1214 10.1098/rsos.160635.

1215 Lobo, J., Shokralla, S., Costa, M.H., Hajibabaei, M. and Costa, F.O. (2017). DNA
1216 metabarcoding for high-throughput monitoring of estuarine macrobenthic
1217 communities. Scientific Reports, 7(1), pp. 1–13. Available at:
1218 http://dx.doi.org/10.1038/s41598-017-15823-6.

1219 Lopes, C.M., De Barba, M., Boyer, F., Mercier, C., Da Silva Filho, P.J.S., Heidtmann,
1220 L.M., Galiano, D., Kubiak, B.B., Langone, P., Garcias, F.M., Gielly, L., Coissac, E.,
1221 De Freitas, T.R.O. and Taberlet, P. (2015). DNA metabarcoding diet analysis for
1222 species with parapatric vs sympatric distribution: A case study on subterranean
1223 rodents. Heredity, 114(5), pp. 525–536. doi: 10.1038/hdy.2014.109.

1224 Lopes, C.M., Sasso, T., Valentini, A., Dejean, T., Martins, M., Zamudio, K.R. and
1225 Haddad, C.F.B. (2017). eDNA metabarcoding: a promising method for anuran
1226 surveys in highly diverse tropical forests. Molecular Ecology Resources, 17(5), pp.
1227 904–914. doi: 10.1111/1755-0998.12643.

1228 Macher, J.N., Vivancos, A., Piggott, J.J., Centeno, F.C., Matthaei, C.D. and Leese,
1229 F. (2018). Comparison of environmental DNA and bulk-sample metabarcoding using
1230 highly degenerate cytochrome c oxidase I primers. Molecular Ecology Resources,
1231 18(6), pp. 1456–1468. doi: 10.1111/1755-0998.12940.

1232 Mata, V.A., Rebelo, H., Amorim, F., McCracken, G.F., Jarman, S. and Beja, P.
1233 (2019). How much is enough? Effects of technical and biological replication on
1234 metabarcoding dietary analysis. Molecular Ecology, 28(2), pp. 165–175. doi:
1235 10.1111/mec.14779.

1236 McInnes, J.C., Alderman, R., Deagle, B.E., Lea, M.A., Raymond, B. and Jarman,
1237 S.N. (2017). Optimised scat collection protocols for dietary DNA metabarcoding in
1238 vertebrates. Methods in Ecology and Evolution, 8(2), pp. 192–202. doi:
1239 10.1111/2041-210X.12677.

1240

1241 McInnes, J.C., Alderman, R., Lea, M.A., Raymond, B., Deagle, B.E., Phillips, R.A.,
1242 Stanworth, A., Thompson, D.R., Catry, P., Weimerskirch, H., Suazo, C.G., Gras, M.
1243 and Jarman, S.N. (2017). High occurrence of jellyfish predation by black-browed and
1244 Campbell albatross identified by DNA metabarcoding. Molecular Ecology, 26(18), pp.
1245 4831–4845. doi: 10.1111/mec.14245.

1246 Milhau, T., Valentini, A., Poulet, N., Roset, N., Jean, P., Gaboriaud, C. and Dejean,
1247 T. (2019). Seasonal dynamics of riverine fish communities using eDNA. Journal of
1248 Fish Biology, (October). doi: 10.1111/jfb.14190.

1249 Miller, K.E., Hopkins, K., Inward, D.J.G. and Vogler, A.P. (2016). Metabarcoding of
1250 fungal communities associated with bark beetles. Ecology and Evolution, 6(6), pp.
1251 1590–1600. doi: 10.1002/ece3.1925.

1252 Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T.,
1253 Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M. and Iwasaki, W. (2015). MiFish,
1254 a set of universal PCR primers for metabarcoding environmental DNA from fishes:
1255 Detection of more than 230 subtropical marine species. Royal Society Open
1256 Science, 2(7). doi: 10.1098/rsos.150088.

1257 Murray, D.C., Coghlan, M.L. and Bunce, M. (2015). From benchtop to desktop:
1258 Important considerations when designing amplicon sequencing workflows. PLoS
1259 ONE, 10(4), pp. 1–21. doi: 10.1371/journal.pone.0124671.

Nakagawa, H., Yamamoto, S., Sato, Y., Sado, T., Minamoto, T. and Miya, M. (2018). Comparing local- and regional-scale estimations of the diversity of stream fish using eDNA metabarcoding and conventional observation methods. Freshwater Biology, 63(6), pp. 569–580. doi: 10.1111/fwb.13094.

National Center for Biotechnology Information (2008). BLAST® Command Line Applications User Manual. Bethesda (MD), USA. Available at: www.ncbi.nlm.nih.gov/books/NBK279690/.

Nichols, R. V., Vollmers, C., Newsom, L.A., Wang, Y., Heintzman, P.D., Leighton, M., Green, R.E. and Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. Molecular Ecology Resources, 18(5), pp. 927–939. doi: 10.1111/1755-0998.12895.

Nichols, S.J., Kefford, B.J., Campbell, C.D., Bylemans, J., Chandler, E., Bray, J.P., Shackleton, M., Robinson, K.L., Carew, M.E. and Furlan, E.M. (2019). Towards routine DNA metabarcoding of macroinvertebrates using bulk samples for freshwater bioassessment: Effects of debris and storage conditions on the recovery of target taxa. Freshwater Biology, (September), pp. 1–14. doi: 10.1111/fwb.13443.

O'Donnell, J.L., Kelly, R.P., Lowell, N.C. and Port, J.A. (2016). Indexed PCR primers induce template- Specific bias in Large-Scale DNA sequencing studies. PLoS ONE, 11(3), pp. 1–11. doi: 10.1371/journal.pone.0148698.

Oehm, J., Juen, A., Nagiller, K., Neuhauser, S. and Traugott, M. (2011). Molecular scatology: How to improve prey DNA detection success in avian faeces? Molecular Ecology Resources, 11(4), pp. 620–628. doi: 10.1111/j.1755-0998.2011.03001.x.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. and Wagner, H. (2013). vegan: Community Ecology Package. R package version 2.0-7.

Oliverio, A.M., Gan, H., Wickings, K. and Fierer, N. (2018). A DNA metabarcoding approach to characterize soil arthropod communities. Soil Biology and Biochemistry, 125(March), pp. 37–43. Available at: https://doi.org/10.1016/j.soilbio.2018.06.026.

Piñol, J., Mir, G., Gomez-Polo, P. and Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular Ecology Resources, 15(4), pp. 819–830. doi: 10.1111/1755-0998.12355.

Piñol, J., San Andrés, V., Clare, E.L., Mir, G. and Symondson, W.O.C. (2014). A pragmatic approach to the analysis of diets of generalist predators: The use of next-generation sequencing with no blocking probes. Molecular Ecology Resources, 14(1), pp. 18–26. doi: 10.1111/1755-0998.12156.

Pochon, X., Zaiko, A., Fletcher, L.M., Laroche, O. and Wood, S.A. (2017). Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. PLoS ONE, 12(11), pp. 1–19. doi: 10.1371/journal.pone.0187636.

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., Roset, N., Schabuss, M., Zornig, H. and Dejean, T. (2018). Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. Scientific Reports, 8(1), pp. 1–13. doi: 10.1038/s41598-018-28424-8.

Pont, D., Valentini, A., Rocle, M., Maire, A., Delaigue, O., Jean, P. and Dejean, T. (2019). The future of fish-based ecological assessment of European rivers: from traditional EU Water Framework Directive compliant methods to eDNA metabarcoding-based approaches. Journal of Fish Biology, (October). doi: 10.1111/jfb.14176.

Pornon, A., Escaravage, N., Burrus, M., Holota, H., Khimoun, A., Mariette, J., Pellizzari, C., Iribar, A., Etienne, R., Taberlet, P., Vidal, M., Winterton, P., Zinger, L. and Andalo, C. (2016). Using metabarcoding to reveal and quantify plant-pollinator interactions. Scientific Reports, 6(February), pp. 1–12. Available at: http://dx.doi.org/10.1038/srep27282.

Prosser, S.W.J. and Hebert, P.D.N. (2017). Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. Food Chemistry, 214, pp. 183–191. Available at: http://dx.doi.org/10.1016/j.foodchem.2016.07.077.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, . Available at: https://www.r-project.org/.

Rennstam Rubbmark, O., Sint, D., Cupic, S. and Traugott, M. (2019). When to use next generation sequencing or diagnostic PCR in diet analyses. Molecular Ecology Resources, 19(2), pp. 388–399. doi: 10.1111/1755-0998.12974.

Rennstam Rubbmark, O., Sint, D., Horngacher, N. and Traugott, M. (2018). A broadly applicable COI primer pair and an efficient single-tube amplicon library preparation protocol for metabarcoding. Ecology and Evolution, 8(24), pp. 12335–12350. doi: 10.1002/ece3.4520.

Richardson, R.T., Lin, C.-H., Sponsler, D.B., Quijia, J.O., Goodell, K. and Johnson, R.M. (2015). Application of ITS2 Metabarcoding to Determine the Provenance of Pollen Collected by Honey Bees in an Agroecosystem. Applications in Plant Sciences, 3(1), p. 1400066. doi: 10.3732/apps.1400066.

Rivera, S.F., Vasselon, V., Jacquet, S., Bouchez, A., Ariztegui, D. and Rimet, F. (2018). Metabarcoding of lake benthic diatoms: from structure assemblages to ecological assessment. Hydrobiologia, 807(1), pp. 37–51. doi: 10.1007/s10750-017-3381-2.

Rocchi, S., Valot, B., Reboux, G. and Millon, L. (2017). DNA metabarcoding to assess indoor fungal communities: Electrostatic dust collectors and Illumina sequencing. Journal of Microbiological Methods, 139(January), pp. 107–112. doi: 10.1016/j.mimet.2017.05.014.

Rodgers, T.W., Xu, C.C.Y., Giacalone, J., Kapheim, K.M., Saltonstall, K., Vargas, M., Yu, D.W., Somervuo, P., McMillan, W.O. and Jansen, P.A. (2017). Carrion fly-

derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community. Molecular Ecology Resources, 17(6), pp. e133–e145. doi: 10.1111/1755-0998.12701.

Roslin, T. and Majaneva, S. (2016). The use of DNA barcodes in food web construction-terrestrial and aquatic ecologists unite! Genome, 59(9), pp. 603–628. doi: 10.1139/gen-2015-0229.

Rulik, B., Eberle, J., von der Mark, L., Thormann, J., Jung, M., Köhler, F., Apfel, W., Weigel, A., Kopetz, A., Köhler, J., Fritzlar, F., Hartmann, M., Hadulla, K., Schmidt, J., Hörren, T., Krebs, D., Theves, F., Eulitz, U., Skale, A., et al. (2017). Using taxonomic consistency with semi-automated data pre-processing for high quality DNA barcodes. Methods in Ecology and Evolution, 8(12), pp. 1878–1887. doi: 10.1111/2041-210X.12824.

Rytkönen, S., Vesterinen, E.J., Westerduin, C., Leviäkangas, T., Vatka, E., Mutanen, M., Välimäki, P., Hukkanen, M., Suokas, M. and Orell, M. (2019). From feces to data: A metabarcoding method for analyzing consumed and available prey in a bird-insect food web. Ecology and Evolution, 9(1), pp. 631–639. doi: 10.1002/ece3.4787.

Sato, H., Sogo, Y., Doi, H. and Yamanaka, H. (2017). Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. Scientific Reports, 7(1), pp. 1–12. Available at: http://dx.doi.org/10.1038/s41598-017-14978-6.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J. and Weber, C.F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology, 75(23), pp. 7537–7541. doi: 10.1128/AEM.01541-09.

Schnell, I.B., Bohmann, K. and Gilbert, M.T.P. (2015). Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. Molecular Ecology Resources, 15(6), pp. 1289–1303. doi: 10.1111/1755-0998.12402.

Schwarz, D., Spitzer, S.M., Thomas, A.C., Kohnert, C.M., Keates, T.R. and Acevedo-Gutiérrez, A. (2018). Large-scale molecular diet analysis in a generalist marine mammal reveals male preference for prey of conservation concern. Ecology and Evolution, 8(19), pp. 9889–9905. doi: 10.1002/ece3.4474.

Shaw, J.L.A., Clarke, L.J., Wedderburn, S.D., Barnes, T.C., Weyrich, L.S. and Cooper, A. (2016). Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. Biological Conservation, 197, pp. 131–138. Available at: http://dx.doi.org/10.1016/j.biocon.2016.03.010.

Shin, S., Lee, T.K., Han, J.M. and Park, J. (2014). Regional effects on chimera formation in 454 pyrosequenced amplicons from a mock community. Journal of Microbiology, 52(7), pp. 566–573. doi: 10.1007/s12275-014-3485-6.

Sickel, W., Ankenbrand, M.J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., Steffan-Dewenter, I. and Keller, A. (2015). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecology, 15(1), pp. 1–9. doi: 10.1186/s12898-015-0051-y.

Siddall, M.E., Fontanella, F.M., Watson, S.C., Kvist, S. and Erséus, C. (2009). Barcoding bamboozled by bacteria: Convergence to metazoan mitochondrial primer targets by marine microbes. Systematic Biology, 58(4), pp. 445–451. doi: 10.1093/sysbio/syp033.

Siegenthaler, A., Wangensteen, O.S., Soto, A.Z., Benvenuto, C., Corrigan, L. and Mariani, S. (2019). Metabarcoding of shrimp stomach content: Harnessing a natural sampler for fish biodiversity monitoring. Molecular Ecology Resources, 19(1), pp. 206–220. doi: 10.1111/1755-0998.12956.

da Silva, L.P., Mata, V.A., Lopes, P.B., Pereira, P., Jarman, S.N., Lopes, R.J. and Beja, P. (2019). Advancing the integration of multi-marker metabarcoding data in dietary analysis of trophic generalists. Molecular Ecology Resources, 19(6), pp. 1420–1432. doi: 10.1111/1755-0998.13060.

Smart, M.D., Cornman, R.S., Iwanowicz, D.D., McDermott-Kubeczko, M., Pettis, J.S., Spivak, M.S. and Otto, C.R.V. (2017). A comparison of honey bee-collected pollen from working agricultural lands using light microscopy and its metabarcoding. Environmental Entomology, 46(1), pp. 38–49. doi: 10.1093/ee/nvw159.

Soininen, E.M., Gauthier, G., Bilodeau, F., Berteaux, D., Gielly, L., Taberlet, P., Gussarova, G., Bellemain, E., Hassel, K., Stenøien, H.K., Epp, L., Schrøder-Nielsen, A., Brochmann, C. and Yoccoz, N.G. (2015). Highly overlapping winter diet in two sympatric lemming species revealed by DNA metabarcoding. PLoS ONE, 10(1), pp. 1–18. doi: 10.1371/journal.pone.0115335.

Srivathsan, A., Ang, A., Vogler, A.P. and Meier, R. (2016). Fecal metagenomics for the simultaneous assessment of diet, parasites, and population genetics of an understudied primate. Frontiers in Zoology, 13(1), pp. 1–13. Available at: http://dx.doi.org/10.1186/s12983-016-0150-4.

Srivathsan, A., Sha, J.C.M., Vogler, A.P. and Meier, R. (2015). Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (Pygathrix nemaeus). Molecular Ecology Resources, 15(2), pp. 250–261. doi: 10.1111/1755-0998.12302.

Stat, M., John, J., DiBattista, J.D., Newman, S.J., Bunce, M. and Harvey, E.S. (2019). Combined use of eDNA metabarcoding and video surveillance for the assessment of fish biodiversity. Conservation Biology, 33(1), pp. 196–205. doi: 10.1111/cobi.13183.

Stoeck, T., Kochems, R., Forster, D., Lejzerowicz, F. and Pawlowski, J. (2018). Metabarcoding of benthic ciliate communities shows high potential for environmental monitoring in salmon aquaculture. Ecological Indicators, 85(October 2017), pp. 153–164. Available at: https://doi.org/10.1016/j.ecolind.2017.10.041.

1422 Taberlet, P., Bonin, A., Coissac, E., Zinger, L. and Lucie, Z. (2018). Environmental
1423 DNA: For biodiversity research and monitoring. Oxford University Press.

1424 Theissinger, K., Kästel, A., Elbrecht, V., Makkonen, J., Michiels, S., Schmidt, S.,
1425 Allgeier, S., Leese, F. and Brühl, C. (2018). Using DNA metabarcoding for assessing
1426 chironomid diversity and community change in mosquito controlled temporary
1427 wetlands. Metabarcoding and Metagenomics, 2, p. e21060. doi:
1428 10.3897/mbmg.2.21060.

1429 Thermo Fisher Scientific ([no date]). Multiple Primer Analyzer. Available at:
1430 https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-
1431 biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-
1432 scientific-web-tools/multiple-primer-analyzer.html [Accessed: 27 July 2020].

1433 Thomas, A.C., Jarman, S.N., Haman, K.H., Trites, A.W. and Deagle, B.E. (2014).
1434 Improving accuracy of DNA diet estimates using food tissue control materials and an
1435 evaluation of proxies for digestion bias. Molecular Ecology, 23(15), pp. 3706–3718.
1436 doi: 10.1111/mec.12523.

1437 Toju, H. and Baba, Y.G. (2018). DNA metabarcoding of spiders, insects, and
1438 springtails for exploring potential linkage between above- and below-ground food
1439 webs. Zoological Letters, 4(1), pp. 1–12. doi: 10.1186/s40851-018-0088-9.

1440 Treonis, A.M., Unangst, S.K., Kepler, R.M., Buyer, J.S., Cavigelli, M.A., Mirsky, S.B.
1441 and Maul, J.E. (2018). Characterization of soil nematode communities in three
1442 cropping systems through morphological and DNA metabarcoding approaches.
1443 Scientific Reports, 8(1), pp. 1–12. Available at: http://dx.doi.org/10.1038/s41598-018-
1444 20366-5.

1445 Ushio, M., Murata, K., Sado, T., Nishiumi, I., Takeshita, M., Iwasaki, W. and Miya, M.
1446 (2018). Demonstration of the potential of environmental DNA as a tool for the
1447 detection of avian species. Scientific Reports, 8(1), pp. 1–10. doi: 10.1038/s41598-
1448 018-22817-5.

1449 Valentini, A., Miquel, C., Nawaz, M.A., Bellemain, E., Coissac, E., Pompanon, F.,
1450 Gielly, L., Cruaud, C., Nascetti, G., Wincker, P., Swenson, J.E. and Taberlet, P.
1451 (2009). New perspectives in diet analysis based on DNA barcoding and parallel
1452 pyrosequencing: The trnL approach. Molecular Ecology Resources, 9(1), pp. 51–60.
1453 doi: 10.1111/j.1755-0998.2008.02352.x.

1454 Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F.,
1455 Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet,
1456 N., Roset, N., Copp, G.H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.M., et al.
1457 (2016). Next-generation monitoring of aquatic biodiversity using environmental
1458 {DNA} metabarcoding. Molecular Ecology, 25, pp. 929–942.

1459 Vamos, E., Elbrecht, V. and Leese, F. (2017). Short COI markers for freshwater
1460 macroinvertebrate metabarcoding. Metabarcoding and Metagenomics, 1, p. e14625.
1461 doi: 10.3897/mbmg.1.14625.

1462 Vences, M., Lyra, M.L., Perl, R.G.B., Bletz, M.C., Stanković, D., Lopes, C.M., Jarek,
1463 M., Bhuju, S., Geffers, R., Haddad, C.F.B. and Steinfartz, S. (2016). Freshwater
1464 vertebrate metabarcoding on Illumina platforms using double-indexed primers of the
1465 mitochondrial 16S rRNA gene. Conservation Genetics Resources, 8(3), pp. 323–
1466 327. doi: 10.1007/s12686-016-0550-y.

1467 De Vere, N., Jones, L.E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., Hegarty,
1468 M.J., Creer, S. and Ford, C.R. (2017). Using DNA metabarcoding to investigate
1469 honey bee foraging reveals limited flower use despite high floral availability.
1470 Scientific Reports, 7(January), pp. 1–10. doi: 10.1038/srep42838.

1471 Vo, A.T.E. and Jedlicka, J.A. (2014). Protocols for metagenomic DNA extraction and
1472 Illumina amplicon library preparation for faecal and swab samples. Molecular
1473 Ecology Resources, 14(6), pp. 1183–1197. doi: 10.1111/1755-0998.12269.

1474 Wangensteen, O.S., Palacín, C., Guardiola, M. and Turon, X. (2018). DNA
1475 metabarcoding of littoral hardbottom communities: High diversity and database gaps
1476 revealed by two molecular markers. PeerJ, 2018(5), pp. 1–30. doi:
1477 10.7717/peerj.4705.

1478 Weyrich, L.S., Farrer, A.G., Eisenhofer, R., Arriola, L.A., Young, J., Selway, C.A.,
1479 Handsley-Davis, M., Adler, C.J., Breen, J. and Cooper, A. (2019). Laboratory
1480 contamination over time during low-biomass sample analysis. Molecular Ecology
1481 Resources, 19(4), pp. 982–996. doi: 10.1111/1755-0998.13011.

1482 Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York:
1483 Springer-Verlag.

1484 Wood, S.A., Pochon, X., Laroche, O., von Ammon, U., Adamson, J. and Zaiko, A.
1485 (2019). A comparison of droplet digital polymerase chain reaction (PCR), quantitative
1486 PCR and metabarcoding for species-specific detection in environmental DNA.
1487 Molecular Ecology Resources, 19(6), pp. 1407–1419. doi: 10.1111/1755-
1488 0998.13055.

1489 Xie, Y., Allaire, J. and Grolemund, G. (2018). R Markdown: The Definitive Guide.
1490 Boca Raton, Florida: Chapman and Hall/CRC. Available at:
1491 https://bookdown.org/yihui/rmarkdown.

1492 Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T.
1493 and Miya, M. (2017). Environmental DNA metabarcoding reveals local fish
1494 communities in a species-rich coastal sea. Scientific Reports, 7(December 2016),
1495 pp. 1–12. doi: 10.1038/srep40368.

1496 Yang, J., Zhang, X., Xie, Y., Song, C., Zhang, Y., Yu, H. and Burton, G.A. (2017).
1497 Zooplankton community profiling in a eutrophic freshwater ecosystem-lake tai basin
1498 by DNA metabarcoding. Scientific Reports, 7(1), pp. 1–11. Available at:
1499 http://dx.doi.org/10.1038/s41598-017-01808-y.

1500 Yoon, T.H., Kang, H.E., Kang, C.K., Lee, S.H., Ahn, D.H., Park, H. and Kim, H.W.
1501 (2016). Development of a cost-effective metabarcoding strategy for analysis of the
1502 marine phytoplankton community. PeerJ, 2016(6). doi: 10.7717/peerj.2115.

Zaiko, A., Samuiloviene, A., Ardura, A. and Garcia-Vazquez, E. (2015). Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. Marine Pollution Bulletin, 100(1), pp. 53–59. Available at: http://dx.doi.org/10.1016/j.marpolbul.2015.09.030.

Zepeda-Mendoza, M.L., Bohmann, K., Carmona Baez, A. and Gilbert, M.T.P. (2016). DAMe: A toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. BMC Research Notes, 9(1), pp. 1–13. doi: 10.1186/s13104-016-2064-9.

Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F., Chariton, A.A., Creer, S., Coissac, E., Deagle, B.E., De Barba, M., Dickie, I.A., Dumbrell, A.J., Ficetola, G.F., Fierer, N., Fumagalli, L., Gilbert, M.T.P., Jarman, S., Jumpponen, A., et al. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. Molecular Ecology, 28(8), pp. 1857–1862. doi: 10.1111/mec.15060.

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