SPECIAL ISSUE

CSIRO PUBLISHING

Marine and Freshwater Research
https://doi.org/10.1071/MF17209

How does marker choice affect your diet analysis: comparing genetic markers and digestion levels for diet metabarcoding of tropical-reef piscivores

Floriaan Devloo-Delva[A,B,C,I], *Roger Huerlimann*[A,D], *Gladys Chua*[A], *Jordan K. Matley*[E], *Michelle R. Heupel*[F], *Colin A. Simpfendorfer*[A] *and Gregory E. Maes*[A,D,G,H]

[A]Centre for Sustainable Tropical Fisheries and Aquaculture, College of Science and Engineering, James Cook University, 145 James Cook Drive, Townsville, Qld 4811, Australia.
[B]Oceans and Atmosphere, CSIRO, GPO Box 1538, Hobart, Tas 7000, Australia.
[C]School of Biological Sciences–Quantitative Marine Science, University of Tasmania, Private Bag 55, Hobart, Tas 7001, Australia.
[D]Comparative Genomics Centre, College of Science and Engineering, James Cook University, 145 James Cook Drive, Townsville, Qld 4811, Australia.
[E]Center for Marine and Environmental Studies, University of the Virgin Islands, 2 John Brewers Bay St Thomas, VI 00802, USA.
[F]Australian Institute of Marine Science, PMB 3, Townsville, Qld 4810, Australia.
[G]Laboratory of Biodiversity and Evolutionary Genomics, KU Leuven, Charles Deberiotstraat 32, BE-3000 Leuven, Belgium.
[H]Center for Human Genetics, UZ Leuven–Genomics Core, KU Leuven, Herestraat 49, BE-3000 Leuven, Belgium.
[I]Corresponding author. Email: floriaan.devloo.delva@hotmail.com

Abstract. Tropical reefs are highly diverse ecosystems, and reliable biomonitoring, through diet metabarcoding, is needed to understand present and future trophic relationships in this changing habitat. Several studies have assessed the reliability and effectiveness of single molecular markers; however, a cross-marker validation has rarely been performed. This study identified crucial properties for *12S* rDNA, *16S* rDNA and *COI* metabarcoding in tropical-reef piscivores (*Plectropomus* spp.). In addition, three new versatile primer sets for *16S* were designed *in silico* for metabarcoding of reef fish. Results showed that *COI* was overall better at recovering true diversity because of a well-supported database. Second, optimal *16S* amplicon sizes ranged between 160 and 440 base pairs for full diversity recovery, with increased species detection for the 270-base pairs region. Finally, blocking of predator-specific *COI* sequences was not equally effective in all host species, potentially introducing bias when diet compositions are directly compared. In conclusion, this novel study showed that marker success for prey identification is highly dependent on the reference database, taxonomic scope, DNA quality, amplicon length and sequencing platform. Results suggest that *COI*, complemented with *16S*, yields the best outcome for diet metabarcoding in reef piscivores. Findings in this paper are relevant to other piscivores and other metabarcoding applications.

Additional keywords: amplicon length, coral trout, gut contents, marine predators, metabarcoding performance, *16S* metabarcoding primers.

Received 9 July 2017, accepted 15 January 2018, published online 9 April 2018

Introduction

Tropical coral reefs are biodiversity hotspots, sustaining complex food webs by providing a wide array of prey for generalists, as well as specific trophic links and niches for specialists. However, reef ecosystems, such as the Great Barrier Reef (GBR), have recently become severely threatened by land-runoff, global warming, pollution and overfishing (Hughes *et al.* 2017). The disappearance of this valuable ecosystem can potentially have disastrous effects on its communities (Thompson *et al.* 2016). Reliable biomonitoring is needed to understand present and future trophic relationships in this changing habitat.

Dietary analyses are important for unravelling ecological processes, such as predator–prey interactions, resource use and community structuring (St John *et al.* 2001; Frisch *et al.* 2016*a*; Matley *et al.* 2017*a*); however, there are many problems with conventional methods. For example, visual identification has a low taxonomic resolution in diet analyses (Carreon-Martinez *et al.* 2011; Leray *et al.* 2012; Berry *et al.* 2015). Moreover, tropical water temperatures increase digestion rates, causing prey items to be less identifiable (Legler *et al.* 2010). One approach that can overcome various issues is molecular identification of gut contents.

Dietary DNA barcoding has been used in a wide range of organisms, such as bottlenose dolphins (Dunshea 2009), deepwater sharks (Dunn *et al.* 2010) and gilthead seabream (Avignon *et al.* 2017). However, DNA barcoding in generalist diets is still restricted by the completeness of the reference database (Dunn *et al.* 2010) and the presence of viable and distinguishable prey parts (Carreon-Martinez *et al.* 2011; Leray *et al.* 2012; Alonso *et al.* 2014). Regarding the latter, the application of high-throughput sequencing, in the form of metabarcoding, is preferable for more auditable, taxonomically comprehensive, reproducible studies (Ji *et al.* 2013; Zaiko *et al.* 2015*a*).

Metabarcoding has recently been applied to complex organisms, such as deepwater fishes (Thomsen *et al.* 2016), fur seals and penguins (Deagle *et al.* 2009, 2010). As a result of the development of standard laboratory and analytical protocols, diet compositions have been successfully determined for many species (reviewed in Pompanon *et al.* 2012). Moreover, Leray *et al.* (2013) developed new cytochrome oxidase I (*COI*) primers that target 313-base pair (bp) metazoan fragments to study diets of predatory reef fish, and were able to identify 51% of the prey to species level and 77% to a higher taxonomic level (Leray *et al.* 2015). Nevertheless, despite its broad application, metabarcoding is still prone to error and loss of information (Cruaud *et al.* 2017). To assess the reliability of metabarcoding, most research efforts have focused on its validation using standard sequence databases (Ji *et al.* 2013; Leray *et al.* 2013; Miya *et al.* 2015). However, many of these studies, with the exception of De Barba *et al.* (2014) and Berry *et al.* (2015), have analysed only single gene fragments.

The availability of reference DNA sequences and various molecular markers makes it difficult to reach an informed choice on the most suitable marker for a target species. Molecular markers should at least exhibit the following consistent properties for efficient use. First, the target gene should be effective at representing the true sample diversity, especially as marker choice strongly depends on the taxonomic focus (Deagle *et al.* 2014), along with the availability of a comprehensive and reliable sequence database for the target taxa (Berry *et al.* 2015; Cruaud *et al.* 2017). For example, in piscivorous diets both nuclear *18S* and *28S* genes would be less useful because of their slow evolutionary rate (i.e. lower level of variation) and, hence, under-estimation of diversity of closely related species. In contrast, mitochondrial DNA (mtDNA) regions show higher evolutionary rates and are, consequently, more suitable for species delineation (Machida and Tsuda 2010). For example, the commonly used *COI* gene is highly diverse among species and has a well-supported database (Lv *et al.* 2014), although it may not contain sufficiently conserved regions suitable for universal metabarcoding primers in highly diverse samples (Deagle *et al.* 2014). The alternative marker system is represented by ribosomal mtDNA genes (e.g. *12S* rDNA and *16S* rDNA), which have slower evolutionary rates than *COI* has (Machida and Tsuda 2010; Cawthorn *et al.* 2012), but provide more opportunity for optimal primer design. The *16S* rDNA marker is widely used for metazoans in marine systems (Alonso *et al.* 2014), whereas *12S* reference sequences are not always available. Second, because of DNA degradation, inherent to stomach contents, amplicon length should ideally be between 150 and 400 bp (Huber *et al.* 2009). There is an obvious trade-off between (1) high throughput sequencing platforms with a short fragment length (owing to maximum length capacity) yielding a restricted phylogenetic resolution (Kress *et al.* 2015) and (2) long amplicons on third-generation platforms (allowing longer reads), but with the risk of reduced amplification success and, hence, generating false negative results. Finally, diet analysis requires the reduction of the abundant (host) DNA amplification to allow reliable detection of rare species templates. A recent study found that, on average, 60% of *COI* and 63% of *16S* sequences belonged to host teleost species (Berry *et al.* 2015), prompting for the use of host-specific blocking primers (with a spacer C3CPG at the 3′ end to prevent elongation of host DNA) to increase the chance of detecting rare species (Vestheim and Jarman 2008). Such an approach has proven successful in several studies, ranging from penguins (Deagle *et al.* 2010) to reef fish (Leray *et al.* 2015).

The present study aimed at investigating various properties for optimal diet-metabarcoding analysis in tropical coral-reef piscivores, more specifically, a direct across-gene (*12S* rDNA, *16S* rDNA and *COI*) comparison of metabarcoding success in diet analysis. We compared (1) the performance of each marker gene on the basis of amplification and pipeline efficiency, amplification bias and relative accuracy in species recovery. Further, we investigated *in silico* (2) the pipeline efficiency and relative species identification accuracy of different *16S* rDNA amplicon lengths. Finally, we analysed (3) the total and host-specific amplification yield between *COI* and *COI* with host-blocking primers, and the relative accuracy of blocking components in closely related host species. Additionally, the effect of digestion levels on amplicon length and blocking performance was examined as a covariate in analyses, and three new versatile *16S* rDNA primer pairs have been developed *in silico*.

Materials and methods

Study species

Coral trout (*Plectropomus* spp.) are members of the Epinephelidae family. The genus contains several species with high economical, recreational and ecological value in reef ecosystems (Frisch *et al.* 2016*b*). It consists of seven species, of which the following three are commonly found in the GBR: *P. leopardus, P. maculatus* and *P. laevis*. Additionally, *P. laevis* undergoes a shift from footballer (FBT) to blue-spot (BST) colour phase that may be linked to behaviour, growth and maturity (Heupel *et al.* 2010). These closely related *Plectropomus* species occur sympatrically in tropical coral reefs (Matley *et al.* 2016, 2017*a*, 2017*b*). All species are likely to be polyphagous predators. For example, the diet of *P. leopardus* includes up to 422 prey species from 28 families, primarily fish (St John *et al.* 2001). This

Table 1. Primers used or designed in the study for metabarcoding purposes in tropical reef piscivores
Primer pairs in bold were used to compare genetic markers

Primer label	Sequence (5′ – 3′)	Reference
12S rDNA (170 bp)		
MiFish-U-F	GTCGGTAAAACTCGTGCCAGC	Miya *et al.* (2015)
MiFish-U-R	CATAGTGGGGTATCTAATCCCAGTTTG	Miya *et al.* (2015)
16S rDNA (160 bp)		
16S-R1-F	GTCYHGCCTGCCCHGTGA	This paper
16S-R1-R[A]	GCTCCAYAGGGTCTTCTCGTC	This paper
16S rDNA (270 bp)		
16S-R2-F[A]	GACGAGAAGACCCTRTGGAGC	This paper
16S-R2-R	CCAACATCGAGGTCGTAARCC	This paper
16S rDNA (440 bp)		
16S-R1-F	GTCYHGCCTGCCCHGTGA	This paper
16S-R2-R	CCAACATCGAGGTCGTAARCC	This paper
COI (313 bp)		
mlCOIintF	GGWACWGGWTGAACWGTWTAYCCYCC	Leray *et al.* (2013)
jgHCO2198	TAIACYTCIGGRTGICCRAARAAYCA	Geller *et al.* (2013)
Additional PCR components		
Illumina adaptor tail 1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG + forward primer	Illumina, USA
Illumina adaptor tail 2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG + reverse primer	Illumina, USA
P. laevis-B	CAAAGAATCAGAATAGRTGTTGGTAAAGG3	This paper
P. leopardus-B	CAAAGAATCAGMMTAAGTGYTGGTAGAG3	This paper
P. maculatus-B	CAAAGAATAAGAATAARTGCTGGTAGAG3	This paper

[A]Primers are tested only *in silico*.

generalist diet enabled testing of the methodology for a wide range of prey items.

Sample collection and DNA extraction

Specimens of *Plectropomus* spp. were collected along the GBR between August 2013 and February 2014 (Matley *et al.* 2017*a*, 2017*b*). Coral trout were obtained by speargun while SCUBA diving or by hooks baited with *Sardinella* spp. Once aboard, fish were sacrificed, stomachs removed and frozen whole immediately. The gut content was extracted in the laboratory and stored in ethanol at 4°C.

Eight samples, two per studied species or colour phase (i.e. *P. leopardus*, *P. maculatus*, *P. laevis* FBT and *P. laevis* BST), were selected for comparing three mitochondrial marker genes (*12S* rDNA, *16S* rDNA and *COI*). For each host species, a sample with high digestion (no visible residue) and one with low digestion (identifiable prey items) was chosen. Thirty-six samples, equally distributed across host species and digestion stages, were chosen to assess the use of host-specific blocking primers with the *COI* gene.

For DNA extraction, all ethanol was first discarded and large chunks were subsampled before removal to allow detection of rare DNA. Five millilitres of 2% cetyltrimethylammonium bromide (CTAB) extraction buffer (Tamari and Hinkley 2016) and 50 μL of proteinase K (20 mg mL^{-1}) were added. Samples were digested overnight at 65°C until all tissue was dissolved. A total of 700 μL of the homogenised solution was extracted following the CTAB procedure from Tamari and Hinkley (2016). DNA quantity and quality were subsequently assessed on a NanoDrop ND-1000 (NanoDrop Technologies Inc., Wilmington, DE, USA).

Library preparation and sequencing

Primer pairs for comparing the *12S*, *16S* and *COI* genes were selected based on amplicon size for Illumina MiSeq sequencing and efficiency and accuracy of taxon identification (see Leray *et al.* 2013; Brandon-Mong *et al.* 2015; Miya *et al.* 2015), except for *16S*, where a 440-bp versatile primer pair was designed to amplify tropical marine fish (Table 1).

Equal concentrations of DNA (20 ng μL^{-1}) were used for polymerase chain reactions (PCR). Amplicon PCR was accomplished in a 20-μL reaction volume with 1 μL of template DNA, 1× MyTaq reaction buffer (Bioline, London, UK), 0.4 μM tailed forward and reverse primer with 10% untailed primers (to initiate amplification) and 0.05 U μL^{-1} MyTaq DNA polymerase (Bioline). Additionally, for the *COI* blocking-primer treatments (hereafter called *COI*-B), 0.5 μL of blocking primer was added according to the identified host. PCR amplification was performed on a C1000 Thermo Cycler (BIO-RAD, Hercules, CA, USA). PCR conditions were set to initial denaturation of 60 s at 95°C, then 40 cycles of 30 s denaturation at 95°C, 30 s annealing at 50, 55 and 56°C for *12S*, *16S* and *COI* respectively, and an extension at 72°C for 30 s. Subsequently, the products were cleaned using the serapure beads protocol (Rohland and Reich 2012) on the Zephyr G3 Compact Liquid Handling Workstation (Caliper Life Sciences, Hopkinton, MA, USA). Cleaned products were barcoded using the 96-sample Nextera Index Kit (A-indices, Illumina, San Diego, CA, USA). In a final volume of 50 μL, we used 5 μL of clean PCR product, 1× MyTaq reaction buffer (Bioline), 5 μL of each indexing primer and 0.05 U μL^{-1} MyTaq DNA polymerase (Bioline). After amplification, PCR products were cleaned again by using the serapure bead protocol. The products were visualised at each

stage on a 1.5% gel, stained with 5000X GelGreen (Biotium, Fremont, CA, USA) to assess quality, before proceeding to the next step. Libraries were quantified with the Qubit dsDNA HS kit (Thermo Fisher Scientific, Waltham, MA, USA), normalised to 1 nM, pooled and sequenced on an Illumina MiSeq Sequencer using the v3 Reagent Kits (Illumina). Paired-end 300-bp sequencing was performed at the Molecular Facility of James Cook University, Townsville.

Metabarcoding pipeline

Sequence quality was checked with FastQC (ver. 0.11, S. Andrews, Babraham Bioinformatics, Cambridge, UK) and reads were filtered using Geneious (ver. 8.1.8, Biomatters, Auckland, New Zealand; Fig. S1, available as Supplementary material for this paper). Primer sequences were removed and bases with a Q-value lower than 20 were trimmed. Forward and reverse reads were paired and merged together at a minimum of 10-bp overlap with the fast length adjustment of short reads (FLASH) plugin implemented in Geneious (Magoč and Salzberg 2011). To investigate the influence of the assembly pipeline, reads were either merged before or after trimming, or not merged at all. Sequences within a certain size range (10% of the expected amplicon length) were selected. Consensus sequences (hereafter called contigs) were *de novo* assembled with high sensitivity (1% mismatch) to assemble more reads, using a greedy algorithm implemented in Geneious. Three local databases (*12S*, *16S* and *COI*) for all fish and one database for *COI* invertebrates were constructed with all non-redundant (nr) sequences from the NCBI GenBank database (see https://www.ncbi.nlm.nih.gov/genbank/, accessed January 2017). Finally, all contigs were blasted (using the Megablast algorithm) against their respective local databases. The completeness of the reference databases was checked and blast results that did not occur in all marker databases were excluded from gene-comparison analysis.

Additionally, the 440-bp *16S* rDNA reads were cut *in silico* after merging to six different test sizes (110, 160, 220, 270, 330 or 440 bp) to infer the accuracy of species identification relative to the fragment lengths.

In silico 16S-*primer and blocking-component design*

Group-specific primers were designed in the *16S* rDNA region. In total, 8488 sequences from 59 marine fish families (2761 species), commonly found around coral reefs, were obtained from GenBank (see https://www.ncbi.nlm.nih.gov/genbank/) and analysed for conserved regions after ClustalW alignment. Primer properties for optimal annealing were examined with Geneious, ver. 8.1.8. We designed three versatile primer sets for fragments of 160, 270 and 440 bp (Table 1).

Similarly, host-specific blocking primers were developed by aligning *COI* sequences from *P. leopardus* (12), *P. maculatus* (13) and *P. laevis* (9), and identifying a species-specific region adjacent to the jgHCO2198 primer. The addition of a spacer C3CPG prevented further elongation.

Data analysis

Per sample, all contigs with similar blast results were combined and the number of reads were summed up. The weighted averages (over the number of reads per contig) of pairwise identity and read lengths were calculated. Samples were grouped according to marker gene, *16S* fragment length or blocking component. Each group was compared based on several metrics that represent the quality of each blast result, including (1) read depth, (2) pairwise identity or (3) species detection accuracy. Blast results with a low number of reads and low identity are likely to be false positives because of sequencing errors or database incompleteness (Leray *et al.* 2015). Therefore, prey species with less than 0.1% of the total number of sequences per sample and species with a pairwise identity of less than 97% for *12S* (Miya *et al.* 2015), 99% for *16S* (Cawthorn *et al.* 2012; Alonso *et al.* 2014) and 98% for *COI* (Leray *et al.* 2013) were identified as poor blast results. Next, by comparing species unique to a group before and after blast-quality filtering, we can detect which prey species are false positives and which are considered a gain in diversity. This measure provides an estimate of relative accuracy of species detection for each marker. Blast-quality filtering removed blast results with less than 0.1% total sequences per sample, a low pairwise identity respective to the marker gene (see above) and an amplicon size outside a 10% range of the expected amplicon length. Two-way analyses of variance (ANOVAs) were performed in R (ver. 3.3.0, R Foundation for Statistical Computing, Vienna, Austria) to identify a significant difference among groups and metrics; a Tukey's HSD test was used to pinpoint the sources of differences *post hoc*, and $P < 0.05$ was considered statistically significant. Assumptions were tested by visual inspection of diagnostic plots, a Shapiro–Wilk test and a Levene's test. All graphs were constructed in R using the ggplot2 package (Wickham 2016).

Results

Sequencing results and barcoding database comparison

On average, we obtained 19 242 reads per sample, of which 30% were filtered out during the assembly pipeline (Fig. 1*a*). Because no stringent thresholds were set for contig assembly, we recovered an average of 525 contigs per sample across the three markers. When blasting these contigs against four custom databases, an average of 504 contigs retrieved a blast hit, leading to 21, 15, 25 and 31 potential prey items for *12S*, *16S*, *COI* and *COI*-B respectively, for all host species combined. This resulted in 37 unique species. On January 2017, the databases contained 19 621, 33 882 and 105 462 fish sequences for *12S*, *16S* and *COI*, and 1 199 452 invertebrate sequences for *COI*. Of the 37 unique species, several did not occur in the *12S* (12) and *16S* (10) databases, whereas all species were present in the *COI* database. Therefore, these species were excluded for gene comparison, leaving us with a total of 24 unique prey species for marker comparison. After quality filtering, this was reduced to seven prey species (not including the host) for eight samples.

Digestion stage had a significant effect on the species richness for all markers, where advanced digestion resulted in a low number of prey species detected ($F_{1,56} = 8.367$, $P = 0.005$; Fig. S2, available as Supplementary material for this paper). Further, no significant differences in species richness between genetic markers could be found. However, there was an effect of

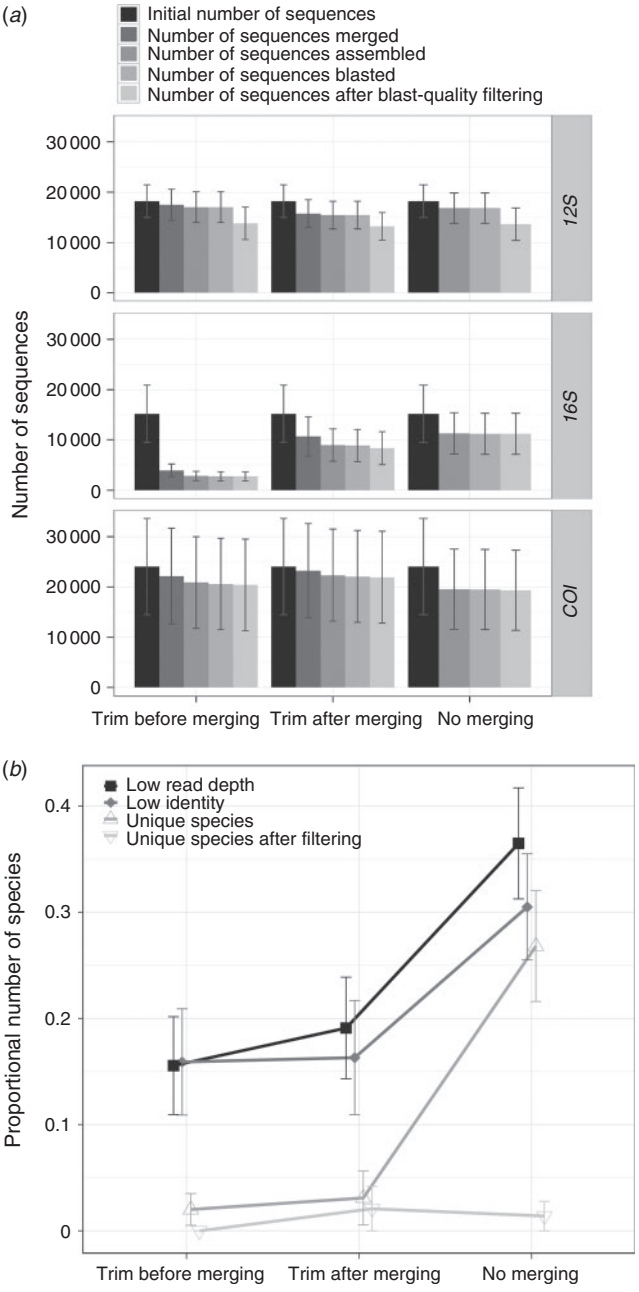

Fig. 1. Pipeline comparison for (*a*) sequencing and pipeline efficiency of *12S*, *16S* and *COI* markers and (*b*) blast-quality metrics for all markers combined. The proportion of each metric is calculated by dividing the number of species per group over the total number of species. Low read-depth species are those with a read depth < 0.1% of the total sequences per sample. Low-identity species are those with a pairwise identity <97, 99 or 98% for *12S*, *16S* or *COI* respectively. Error bars represent standard errors.

blast-quality filtering ($F_{1,56} = 4.532$, $P = 0.037$), going from an overall mean of 3.5 species per sample before to two species per sample after filtering. These two species usually, although not always, comprised one host and one prey. Additionally, we want to note that the *COI* and *COI*-B sequences yielded one invertebrate species of the crab (Brachyura) infraorder.

Metabarcoding performance of three marker genes

Pipeline

The merging step of the assembly pipeline was identified as crucial for retaining DNA sequences and detecting more species (Fig. 1*a*). The *16S* rDNA gene lost a non-significant (ns) amount of sequences if trimming was performed before merging ($F_{4,107} = 2.442$, $P = 0.051$). This decrease could be attributed to the merging step ($F_{2,109} = 3.962$, $P = 0.022$); the same, although less explicit, was noticeable for *COI*-B (results not shown). The *12S* pipelines showed a small, non-significant effect of blast-quality filtering (Tukey's test, $P = 0.293$). Generally, *COI* yielded and retained more sequences ($F_{2,320} = 16.957$, $P < 0.001$).

We found that blast quality decreased if only forward reads were used for species identification ($F_{2,276} = 17.764$, $P < 0.001$; Fig. 1*b*). Typically, non-merged reads returned 13% more low-quality species per sample. More specifically, without merging, more blast results had few reads ($F_{2,69} = 5.272$, $P = 0.007$) and lower, but non-significant, pairwise identity ($F_{2,69} = 2.63$, $P = 0.079$). Moreover, forwards reads yielded 23% more unique species per sample ($F_{2,69} = 16.350$, $P < 0.001$) that were filtered out afterwards. A Tukey's HSD test showed that merging before or after trimming low-quality bases did not have any significant effect on species recovery.

Amplification and pipeline bias

We checked the amplification bias for species unique to a gene and examined whether the loss of *16S* sequences (reported in previous section) introduced any bias towards certain species (Fig. 2). Overall, *16S* had fewer sequences per blast hit than *12S* and *COI* did , and was unable to detect most of the species identified by the other genes (only 6 of 24). *COI* returned the most prey species compared with *12S* and *16S*. Some species were detected only by *12S* or *COI*, but only at low read depth. Overall, 9 of the 24 species were retained after blast-quality filtering; of those, two were host species and *Sardinella lemuru* does not naturally occur in the GBR (http://www.fishbase.org, accessed May 2017). Two of the potential prey species were not detected by *COI* (only *12S*); however, one potential prey species (*Pterocaesio digramma*) was removed from the analysis because of its absence in the *16S* database. It is also noteworthy that *P. maculatus* was not recovered by any of the markers (although present in all databases) and that *12S* could not distinguish *P. laevis* from *P. leopardus* sequences.

Blast quality

All three mitochondrial marker genes yielded a high blast-quality level, returning an average of less than 15% low-quality species per sample (Fig. 3). Overall differences among marker genes were not significant ($F_{2,84} = 2.764$, $P = 0.069$). The relative accuracy of *12S* rDNA was most variable. For example, on average 20% of the species found per sample had a low identity, compared with *16S* (10%) and *COI* (9%). Next, *12S* contained 26% unique species per sample, which decreased non-significantly to 9% after filtering. The large difference in unique species before and after filtering indicated that *12S* returned many misidentified species because

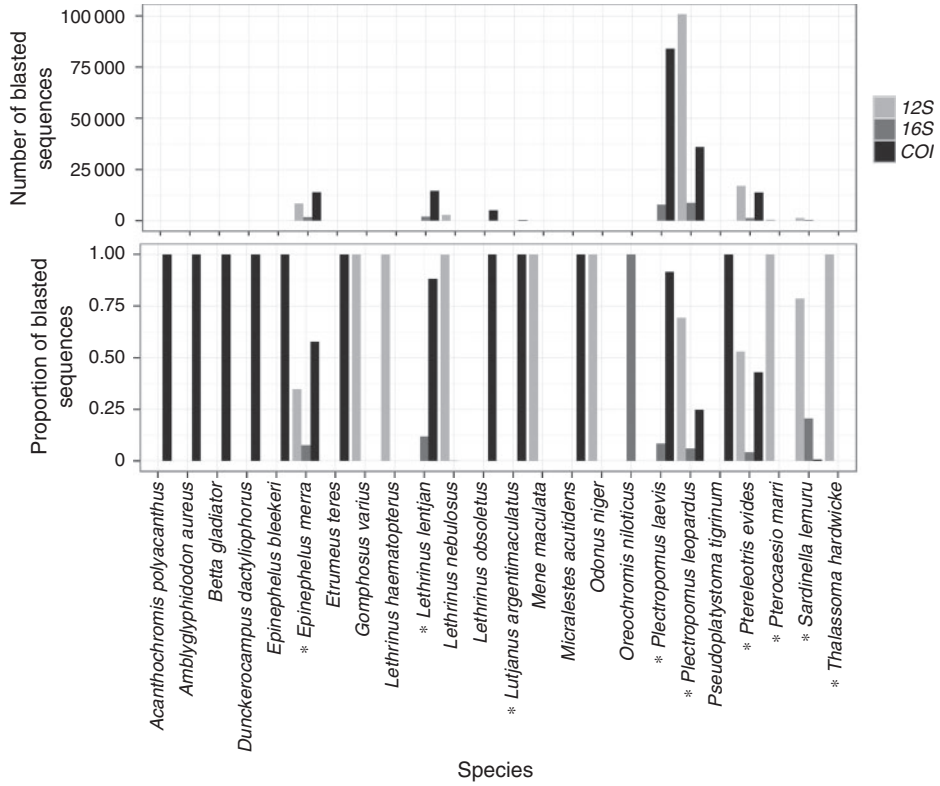

Fig. 2. Amplification and pipeline bias from eight samples, before blast-quality filtering. Species that were retained after blast-quality filtering are indicated with an asterisk (*). Top: absolute number of sequences per blast hit for each marker gene. Bottom: proportion of the number of blasted sequences per predator or prey species per gene over the total number of sequences per blast hit.

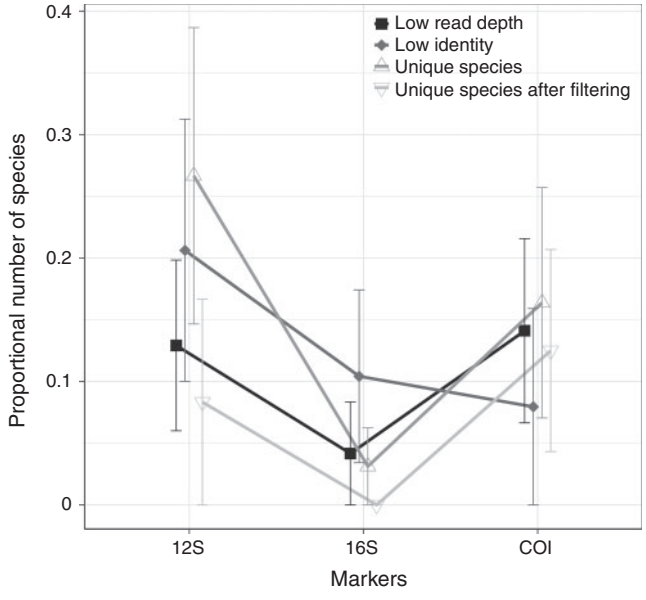

Fig. 3. Blast-quality metrics of marker-gene comparison. The proportion of each metric is calculated by dividing the number of species per group over the total number of species. Low read-depth species are those with a read depth <0.1% of the total sequences per sample. Low-identity species are those with a pairwise identity <97, 99 or 98% for *12S*, *16S* or *COI* respectively. Error bars represent standard errors.

of missing sequences in the database. Both *16S* and *COI* had a constant difference (3%) before and after filtering. Lastly, the many unique species after filtering in *COI* could suggest that the additional unique species were not misidentifications, but rather a gain in diversity because of better performance.

16S *amplicon-length species resolution*

Highly digested samples yielded and retained fewer sequences after blasting ($F_{1,36} = 18.971$, $P < 0.001$; Fig. 4*a*). Overall, no significant difference in the number of blasted sequences was found among lengths within digestion stages. The number of sequences decreased, non-significantly, from short to long fragments; this might have been caused by short sequences finding many random matches and, hence, more blasted sequences.

All fragment lengths were accurate at species recovery (Fig. 4*b*). The ratios of low-read and low-identity species were constant, but an amplicon length smaller than 160 bp led to an over-representation of unique species ($F_{15,168} = 4.20$, $P < 0.001$). The difference between unique species per sample before and after blast-quality filtering suggested that 110-bp (20%) and 160-bp (20%) fragments initially included numerous misidentifications. Amplicons of 220, 270 and 330 bp showed an equal ratio, but 270-bp fragments had a higher amount of unique species.

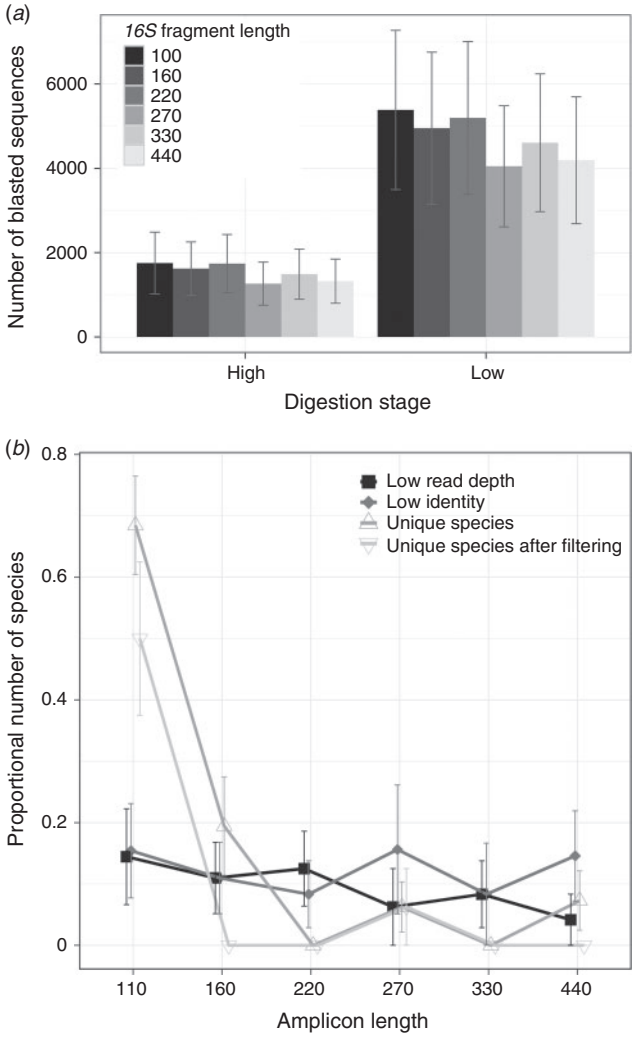

Fig. 4. *In silico 16S* rDNA-fragment analysis, comparing (*a*) pipeline efficiency among different amplicon lengths for samples with high and low digestive stages and (*b*) blast-quality metrics of *16S* rDNA amplicon lengths. The proportion of each metric is calculated by dividing the number of species per group over the total number of species. Low read-depth species are those with a read depth of <0.1% of the total sequences per sample. Low-identity species are those with a pairwise identity <99%. Error bars represent standard errors.

Host-blocking primers efficiency

Sequencing results from 36 samples identified a significant decrease of amplification with blocking primers for *P. laevis* samples ($F_{3,56} = 3.820$, $P = 0.015$; Fig. 5*a*), whereas the number of sequences remained equal for *P. leopardus* and *P. maculatus* samples ($F_{1,20} = 0.592$, $P = 0.451$ and $F_{1,14} = 2.312$, $P = 0.151$). Furthermore, Fig. 5*a* shows an equal number of sequences between digestion levels; however, Fig. S3 (available as Supplementary material for this paper) indicates that less digested samples had fewer host sequences ($F_{1,56} = 232.080$, $P < 0.001$), regardless of the use of blocking primers. This observation showed that although the number of sequences was similar, highly digested samples mostly consisted of the host, and blocking primers were not capable of removing all host sequences.

When applying the blocking primers, the number of low reads ($F_{1,70} = 17.000$, $P < 0.001$) and unique species ($F_{1,70} = 63.170$, $P < 0.001$) decreased, whereas the number of species with a low identity remained constant (ns; Fig. 5*b*). The difference before and after filtering was small for both *COI* (Tukey's test, $P = 0.159$) and *COI*-B (Tukey's test, $P = 0.906$), indicating that the recovered unique species might not be false positives. Remarkably, and to the contrary of our expectations, the ratio of the number of host *v.* prey species increased significantly when using host-blocking primers ($F_{1,70} = 8.437$; $P = 0.005$).

Discussion

We present the first diet metabarcoding study to directly compare marker-gene properties within the same tropical piscivore samples. We expected *COI* to perform better because of its higher evolutionary rate (Machida and Tsuda 2010) and well-supported database (Lv *et al.* 2014). Results showed that the high number of retained *COI* sequences could indicate its ability to identify fish species, as argued by Leray *et al.* (2015). In contrast, similar to Miya *et al.* (2015), *12S* exhibited several misidentifications. Moreover, false negatives were visible for the 440-bp *16S* fragment if trimming occurred before merging. When only forward reads were used, many incorrect blast results were found because of low-quality bases at the 3′ end, whereas the high- and low-quality bases supplement each other when paired reads are merged. Overall, more sequences and species were retained if trimming happened after merging.

In total, 7 of the 24 blast hits are potential prey species based on samples from eight different individuals. These species are from a broad taxonomic range (i.e. seven different families). All species, except for *Sardinella lemuru*, occur naturally on the GBR. Nevertheless, the genus is commonly used as bait when fishing for *Plectropomus* spp. Overall, we observed little PCR-amplification bias of *COI* relative to the other markers, which contrasts the results of Deagle *et al.* (2014) and Zaiko *et al.* (2015*b*). Instead, we retrieved *COI* sequences from many possible prey species. Similar results were reported by Leray *et al.* (2013) and Aylagas *et al.* (2016) who attributed the increased performance to the shorter target region. Hence, we recommend *COI* as a better gene to distinguish a broad range of species, as well as closely related species. Additionally, the versatile *COI* primers also allowed for detection of invertebrates, which is of importance for future biodiversity and diet-composition studies (Leray *et al.* 2013; Brandon-Mong *et al.* 2015; Aylagas *et al.* 2016). However, this feature is not unique to the *COI* gene; *12S* and *16S* rDNA could have similar properties.

Another argument in favour of *COI* is the database completeness. Potential food items can be missed by incomplete databases (Kress *et al.* 2015); *12S* and *16S* databases showed a substantial lack of tropical species barcodes, leading to the poor results found in the present study. The use of *COI* primers and database proved successful for diet of coral-reef fishes by Leray *et al.* (2012). However, we support Zaiko *et al.* (2015*b*) in arguing for a reference database that is taxonomically verified in

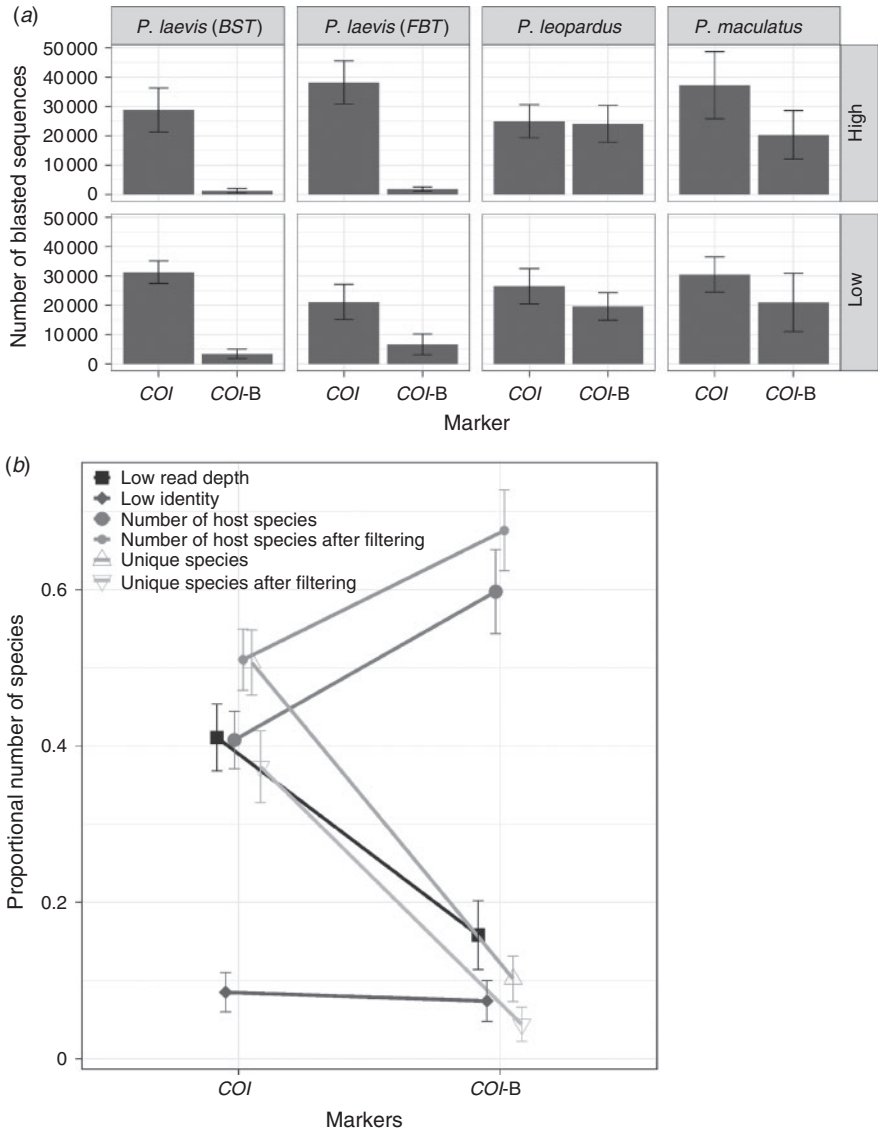

Fig. 5. (*a*) Sequencing and pipeline efficiency of host-blocking components for *COI*, contrasted against host species and digestion stages. (*b*) Blast-quality metrics of host-blocking comparison. The proportion of each metric is calculated by dividing the number of species per group over the total number of species. Low read-depth species are those with a read depth <0.1% of the total sequences per sample. Low-identity species are those with a pairwise identity <98%. Error bars represent standard errors.

order to prevent inaccurate or false assignment because of misidentified reference sequences.

COI could still be susceptible to false negatives in diet studies of other piscivorous animals, because of primer–template mismatches; consequently, Galal-Khallaf *et al.* (2016) and Cruaud *et al.* (2017) suggested that the marker should be complemented with a secondary gene. Our results showed that *16S* may be a good addition, provided an expansion of current vertebrate *16S* databases. Owing to its slower evolutionary rate, *16S* usually exhibits more conserved regions at higher taxonomic levels with a resolution similar to *COI* (Deagle *et al.* 2014). Combined with findings of Alonso *et al.* (2014) on fish diet of the Cory's shearwater, we suggest that with optimised *COI* primers and a better-curated database for *16S*, the most accurate fish diversity can be unveiled.

Amplicon size could also be the driver for the results favouring *COI*; therefore, we analysed different *16S* lengths *in silico*. Based on the results of Huber *et al.* (2009), we assumed that an intermediate length would be optimal for identifying the correct species, without any PCR and sequencing bias. We found that amplicons between 160 and 440 bp had enough variable sites for good blast qualities and reliable blast results. More specifically, our results showed that a fragment length of 270 bp yielded additional correct species. Although amplification success and cross-amplification across a broader range of species are yet to be tested for the *16S* rDNA fragment, it should

be considered to complement *COI* in metabarcoding studies (Alonso *et al.* 2014).

Whereas metabarcoding gene markers may benefit from host blocking-primer combinations by inhibiting abundant host DNA to amplify rare DNA sequences (Vestheim and Jarman 2008; Leray *et al.* 2015), our results did not show a significant increase in diversity between *COI* and *COI*-B. Host blocking was effective only in *P. laevis* samples. The lack of host inhibition for *P. leopardus* and *P. maculatus* individuals might arise from insufficient intraspecific variability knowledge (~12 sequences per species were used for primer design). Using blocking primers of different blocking quality could introduce bias if diets between these host species are compared. Overall, *COI*-B was not able to completely inhibit host templates from being amplified; in some conditions (high digestion with over-representation of host DNA), we even saw an increase in host species per sample. The increase is most likely due to unsuccessful amplification of rare sequences (diet DNA traces) caused by complex PCR reactions (Dunshea 2009), leading to a lower diversity in each sample. Therefore, given that designing blocking primers (1) is expensive and (2) requires prior knowledge of the host species, and that (3) digestion stage seemed equally important to decrease host DNA, we advise to invest in deeper sequencing of additional low-digestion samples and PCR replicates.

Findings of this paper can generally be extended to other piscivores and other metabarcoding applications. Given that *Plectropomus* spp. are common tropical predatory reef fish with a diverse diet (see results and St John *et al.* 2001), a wide range of additional tropical piscivores can be analysed in a similar way, such as coral dwelling fish (Leray *et al.* 2015), reef sharks (Frisch *et al.* 2016*a*) and dolphins (Dunshea 2009). For tropical systems that lack a thorough database, we suggest building a complete baseline of samples with known stomach content (positive controls and mock community), in addition to analysing sufficient replicates to avoid biased results. Moreover, the discussed metabarcoding markers may be useful in other metabarcoding applications. We believe that after additional *in silico* and *in vitro* verification, the *COI* primers in combination with the 270-bp *16S* primers can be applied to non-invasive diet assessments from faecal samples (e.g. Deagle *et al.* 2009, 2010) or biodiversity monitoring studies using environmental DNA from water to detect the presence–absence of fish species, expecting equal resolution and accuracy (e.g. Thomsen *et al.* 2016).

Conclusions

This study is the first to have performed an across-gene comparison for metabarcoding success in diet analysis of tropical generalist piscivores. It showed that marker success is highly dependent on the reference database, taxonomic scope, DNA quality, fragment length and sequencing platform. For further study of *Plectropomus* spp., we showed that *COI* is a reliable marker for identifying diet diversity. We also suggest ascertaining the efficiency of blocking primers for closely related species, because they could introduce strong amplification bias due to unknown intra-species diversity. We propose that the existing *COI* amplicons, complemented with short *16S* rDNA fragments, are sufficient to identify true biodiversity in diet studies to determine trophic roles of reef piscivores.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

The authors thank the following for field assistance: S. Moore, J. Eurich, J. Smart, P. Tenzing, V. Messmer, E. Lédée, F. de Faria, P. Yates, M. Espinoza and A. Tobin. This research pertained to permits issued by the Great Barrier Reef Marine Park Authority (G12/35236.1 and G14/36624.1) and Queensland Department of Primary Industries and Fisheries (144482). All research was conducted under JCU Animal Ethics Permit A1933. This project was funded by the Australian Government's National Environmental Research Program (Tropical Ecosystems Hub Project 6.1) and supported by an Australian Research Council Future Fellowship (#FT100101004). The authors sincerely appreciate the valuable comments of two anonymous reviewers to improve this manuscript.

References

Alonso, H., Granadeiro, J. P., Waap, S., Xavier, J., Symondson, W. O., Ramos, J. A., and Catry, P. (2014). A holistic ecological analysis of the diet of Cory's shearwaters using prey morphological characters and DNA barcoding. *Molecular Ecology* **23**, 3719–3733. doi:10.1111/MEC.12785

Avignon, S., Tastard, E., Weston, S., Duhamel, G., and Denis, F. (2017). Morphological identification and DNA barcoding used for diet analysis of gilthead seabream (*Sparus aurata*) in its expanding northerly range. *Aquatic Living Resources* **30**, 1. doi:10.1051/ALR/2016034

Aylagas, E., Borja, A., Irigoien, X., and Rodriguez-Ezpeleta, N. (2016). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Frontiers of Materials Science* **3**, 96. doi:10.3389/FMARS.2016.00096

Berry, O., Bulman, C., Bunce, M., Coghlan, M., Murray, D. C., and Ward, R. D. (2015). Comparison of morphological and DNA metabarcoding analyses of diets in exploited marine fishes. *Marine Ecology Progress Series* **540**, 167–181. doi:10.3354/MEPS11524

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**, 717–727. doi:10.1017/S0007485315000681

Carreon-Martinez, L., Johnson, T., Ludsin, S., and Heath, D. (2011). Utilization of stomach content DNA to determine diet diversity in piscivorous fishes. *Journal of Fish Biology* **78**, 1170–1182. doi:10.1111/J.1095-8649.2011.02925.X

Cawthorn, D.-M., Steinman, H. A., and Witthuhn, R. C. (2012). Evaluation of the *16S* and *12S* rRNA genes as universal markers for the identification of commercial fish species in South Africa. *Gene* **491**, 40–48. doi:10.1016/J.GENE.2011.09.009

Cruaud, P., Rasplus, J.-Y., Rodriguez, L. J., and Cruaud, A. (2017). High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Scientific Reports* **7**, 41948. doi:10.1038/SREP41948

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**, 306–323. doi:10.1111/1755-0998.12188

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

Deagle, B. E., Chiaradia, A., McInnes, J., and Jarman, S. N. (2010). Pyrosequencing faecal DNA to determine diet of little penguins: is

what goes in what comes out? *Conservation Genetics* **11**, 2039–2048. doi:10.1007/S10592-010-0096-6

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., and Taberlet, P. (2014). DNA metabarcoding and the cytochrome *c* oxidase subunit I marker: not a perfect match. *Biology Letters* **10**, 20140562. doi:10.1098/RSBL.2014.0562

Dunn, M. R., Szabo, A., McVeagh, M. S., and Smith, P. J. (2010). The diet of deepwater sharks and the benefits of using DNA identification of prey. *Deep-sea Research. Part I, Oceanographic Research Papers* **57**, 923–930. doi:10.1016/J.DSR.2010.02.006

Dunshea, G. (2009). DNA-based diet analysis for any predator. *PLoS One* **4**, e5252. doi:10.1371/JOURNAL.PONE.0005252

Frisch, A. J., Ireland, M., Rizzari, J. R., Lönnstedt, O. M., Magnenat, K. A., Mirbach, C. E., and Hobbs, J.-P. A. (2016*a*). Reassessing the trophic role of reef sharks as apex predators on coral reefs. *Coral Reefs* **35**, 459–472. doi:10.1007/S00338-016-1415-2

Frisch, A. J., Ireland, M., Rizzari, J. R., Lonnstedt, O. M., Magnenat, K. A., Mirbach, C. E., and Hobbs, J. P. A. (2016*b*). Reassessing the trophic role of reef sharks as apex predators on coral reefs. *Coral Reefs* **35**, 459–472. doi:10.1007/S00338-016-1415-2

Galal-Khallaf, A., Osman, A. G., Carleos, C. E., Garcia-Vazquez, E., and Borrell, Y. J. (2016). A case study for assessing fish traceability in Egyptian aquafeed formulations using pyrosequencing and metabarcoding. *Fisheries Research* **174**, 143–150. doi:10.1016/J.FISHRES.2015.09.009

Geller, J., Meyer, C., Parker, M., and Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome *c* oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources* **13**, 851–861. doi:10.1111/1755-0998.12138

Heupel, M. R., Williams, A. J., Welch, D. J., Davies, C. R., Adams, S., Carlos, G., and Mapstone, B. D. (2010). Demography of a large exploited grouper, *Plectropomus laevis*: implications for fisheries management. *Marine and Freshwater Research* **61**, 184–195. doi:10.1071/MF09056

Huber, J. A., Morrison, H. G., Huse, S. M., Neal, P. R., Sogin, M. L., and Mark Welch, D. B. (2009). Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environmental Microbiology* **11**, 1292–1302. doi:10.1111/J.1462-2920.2008.01857.X

Hughes, T. P., Kerry, J. T., Alvarez-Noriega, M., Alvarez-Romero, J. G., Anderson, K. D., Baird, A. H., Babcock, R. C., Beger, M., Bellwood, D. R., Berkelmans, R., Bridge, T. C., Butler, I. R., Byrne, M., Cantin, N. E., Comeau, S., Connolly, S. R., Cumming, G. S., Dalton, S. J., Diaz-Pulido, G., Eakin, C. M., Figueira, W. F., Gilmour, J. P., Harrison, H. B., Heron, S. F., Hoey, A. S., Hobbs, J. P. A., Hoogenboom, M. O., Kennedy, E. V., Kuo, C. Y., Lough, J. M., Lowe, R. J., Liu, G., Culloch, M. T. M., Malcolm, H. A., Mcwilliam, M. J., Pandolfi, J. M., Pears, R. J., Pratchett, M. S., Schoepf, V., Simpson, T., Skirving, W. J., Sommer, B., Torda, G., Wachenfeld, D. R., Willis, B. L., and Wilson, S. K. (2017). Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377. doi:10.1038/NATURE21707

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., and Edwards, F. A. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* **16**, 1245–1257. doi:10.1111/ELE.12162

Kress, W. J., García-Robledo, C., Uriarte, M., and Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution* **30**, 25–35. doi:10.1016/J.TREE.2014.10.008

Legler, N. D., Johnson, T. B., Heath, D. D., and Ludsin, S. A. (2010). Water temperature and prey size effects on the rate of digestion of larval and early juvenile fish. *Transactions of the American Fisheries Society* **139**, 868–875. doi:10.1577/T09-212.1

Leray, M., Boehm, J., Mills, S. C., and Meyer, C. (2012). Moorea BIOCODE barcode library as a tool for understanding predator–prey interactions: insights into the diet of common predatory coral reef fishes. *Coral Reefs* **31**, 383–388. doi:10.1007/S00338-011-0845-0

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

Leray, M., Meyer, C. P., and Mills, S. C. (2015). Metabarcoding dietary analysis of coral dwelling predatory fish demonstrates the minor contribution of coral mutualists to their highly partitioned, generalist diet. *PeerJ* **3**, e1047. doi:10.7717/PEERJ.1047

Lv, J., Wu, S., Zhang, Y., Chen, Y., Feng, C., Yuan, X., Jia, G., Deng, J., Wang, C., and Wang, Q. (2014). Assessment of four DNA fragments (*COI*, *16S* rDNA, *ITS2*, *12S* rDNA) for species identification of the Ixodida (Acari: Ixodida). *Parasites & Vectors* **7**, 93. doi:10.1186/1756-3305-7-93

Machida, R. J., and Tsuda, A. (2010). Dissimilarity of species and forms of planktonic *Neocalanus* copepods using mitochondrial *COI*, *12S*, nuclear *ITS*, and *28S* gene sequences. *PLoS One* **5**, e10278. doi:10.1371/JOURNAL.PONE.0010278

Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963. doi:10.1093/BIOINFORMATICS/BTR507

Matley, J., Tobin, A., Lédée, E., Heupel, M., and Simpfendorfer, C. (2016). Contrasting patterns of vertical and horizontal space use of two exploited and sympatric coral reef fish. *Marine Biology* **163**, 253. doi:10.1007/S00227-016-3023-7

Matley, J. K., Heupel, M. R., Fisk, A. T., Simpfendorfer, C. A., and Tobin, A. J. (2017*a*). Measuring niche overlap between co-occurring *Plectropomus* spp. using acoustic telemetry and stable isotopes. *Marine and Freshwater Research* **68**, 1468–1478. doi:10.1071/MF16120

Matley, J. K., Tobin, A. J., Simpfendorfer, C. A., Fisk, A. T., and Heupel, M. R. (2017*b*). Trophic niche and spatio-temporal changes in the feeding ecology of two sympatric species of coral trout (*Plectropomus leopardus* and *P. laevis*). *Marine Ecology Progress Series* **563**, 197–210. doi:10.3354/MEPS11971

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., and Araki, H. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science* **2**, 150088. doi:10.1098/RSOS.150088

Pompanon, F., Deagle, B. E., Symondson, W. O., Brown, D. S., Jarman, S. N., and Taberlet, P. (2012). Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology* **21**, 1931–1950. doi:10.1111/J.1365-294X.2011.05403.X

Rohland, N., and Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* **22**, 939–946. doi:10.1101/GR.128124.111

St John, J., Russ, G. R., Brown, I. W., and Squire, L. C. (2001). The diet of the large coral reef serranid *Plectropomus leopordus* in two fishing zones on the Great Barrier Reef, Australia. *Fishery Bulletin* **99**, 180–192.

Tamari, F., and Hinkley, C. S. (2016). Extraction of DNA from plant tissue: review and protocols. 'Sample Preparation Techniques for Soil, Plant, and Animal Samples'. (Ed. M. Mićić.) pp. 245–263. (Humana Press: New York, NY, USA.)

Thompson, A., Costello, P., Davidson, J., Logan, M., Coleman, G., Gunn, K., and Schaffelke, B. (2016). Marine monitoring program. Annual report for inshore coral reef monitoring: 2014 to 2015. Report for the Great Barrier Reef Marine Park Authority. Australian Institute of Marine Science, Townsville, Qld, Australia.

Thomsen, P. F., Moller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jorgensen, O. A., and Willerslev, E. (2016). Environmental DNA from seawater

samples correlate with trawl catches of subarctic, deepwater fishes. *PLoS One* **11**, e0165252. doi:10.1371/JOURNAL.PONE.0165252

Vestheim, H., and Jarman, S. N. (2008). Blocking primers to enhance PCR amplification of rare sequences in mixed samples–a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology* **5**, 12. doi:10.1186/1742-9994-5-12

Wickham, H. (2016). 'ggplot2: Elegant Graphics for Data Analysis', 2 edn. (Springer International Publishing: New York, NY, USA.)

Zaiko, A., Martinez, J. L., Schmidt-Petersen, J., Ribicic, D., Samuiloviene, A., and Garcia-Vazquez, E. (2015*a*). Metabarcoding approach for the ballast water surveillance: an advantageous solution or an awkward challenge? *Marine Pollution Bulletin* **92**, 25–34. doi:10.1016/J.MARPOLBUL.2015.01.008

Zaiko, A., Martinez, J. L., Ardura, A., Clusa, L., Borrell, Y. J., Samuiloviene, A., Roca, A., and Garcia-Vazquez, E. (2015*b*). Detecting nuisance species using NGST: methodology shortcomings and possible application in ballast water monitoring. *Marine Environmental Research* **112**, 64–72. doi:10.1016/J.MARENVRES.2015.07.002