Investigating the effects of riparian timber extraction on trophic interactions of three Cyprynid Species from the the Kalabakan Basin, Northern Borneo Using DNA Metabarcoding

Phipps C. D.

September 5, 2014

Abstract

The accurate characterisation of trophic interactions is of importance to many fields of science, from single species conservation through to resolving community level food webs and assessing ecosystem function. with the advent of next generation DNA sequencing technologies, the study of trophic ecology experienced an important progression; the identification of resource species through the recovery of short, but informative 'barcode' regions of DNA from tissues present in gut contents or faeces. to date very few have focussed on freshwater systems, and to our knowledge none from the tropical streams of northern Borneo. the current study aims to characterise the diets of three locally abundant cyprinid species in order to identify the key resources they exploit and how timber extraction modifies resource use. Field work was carried out over three streams which had been subjected to varying logging intensity. PCR was conducted on stomach samples taken, employing two chloroplastic, rbcL and trnL markers and one mitochondrial marker, COI. 63, 19 and 2 taxa were detected, respectively. Taxonomic richness was found to be lower in heavily-logged sites, and species were found to respond differentially to logging intensity. The results of this study serve to emphasise the importance of riparian buffer zones in mitigating the detrimental effects of timber extraction on freshwater communities. The use of molecular techniques afforded a level of taxonomic resolution that would not have been possible through classical methods, however further research should be done in order to fully explore the dietary diversity these species exhibit.

1 Introduction

The accurate characterisation of trophic interactions is of importance to many fields of science, from single species conservation through to resolving community level food webs and assessing ecosystem function. Until relatively recently, attempts to discern a species' diet were carried out via direct behavioural observation, or through microscopic examination of gut contents or faeces. Although useful in lieu of more accurate methods, behavioural observation and morphological analysis possess a number of limitations, namely only large, easily observable species can be adequately studied and results can be skewed by the taxonomic expertise of the researcher identifying food items. A number of molecular approaches have also been employed in this field including serological tests for prey specific antibodies [Boreham and Ohiagu, 1978], protein electrophoretic approaches [Traugott, 2003], and stable isotope analysis [Jervis, 2005], all with mixed success and applicability. With the advent of readily available semi-automated DNA sequencing technologies, the study of trophic ecology experienced an important progression; the identification of resource species through the recovery of short, but informative 'barcode' regions of DNA from tissues present in gut contents or faeces. However, it was the development of next generation high throughput sequencing technologies, combined with a simultaneous expansion of publicly available genomic databases which has allowed the use of this technique to grow exponentially [Andrew et al., 2013]. To date, dietary barcoding (or metabarcoding where the identification of multiple species from a single sample is concerned) has been successfully carried out in an impressive array of taxa and ecological contexts, including both predatory and herbivorous invertebrates [Raso et al., 2014, Alcaide et al., 2009, Jurado-Rivera et al., 2009] fish [Corse et al., 2010, Barnett et al., 2010, Leray et al., 2013, Riemann et al., 2010], reptiles [Brown et al., 2012], birds [Oehm et al., 2011, Stech et al., 2011] and mammals [Shehzad et al., 2012, Baamrane et al., 2012, Alberdi et al., 2012, Quéméré et al., 2013, Kim et al., 2011]. Despite the large number of recent publications taking advantage of next generation sequencing technologies to characterise species' dietary habits, very few have focussed on freshwater systems [Corse et al., 2010], and to our knowledge none from the tropical streams of northern Borneo.

The freshwater systems of Borneo represent huge opportunities for the scientific study of the biodiversity, ecology and evolution of tropical fish communities. The Malay archipelago is considered a global hotspot for freshwater fish diversity, with icthyofaunal endemism and diversity especially high in Borneo given the size of the island [Kottelat and Whitten, 1996]. However it is likely even this represents an underestimation of true diversity, as with continued scientific research in the area, new species of fish have been consistently discovered on the island of Borneo for over 100 years [Boulenger, 1894, Seale, 1910, Herre, 1940, BANARESCU and Bianco, 1984, Lim, 1995, Britz et al., 2011]. It is worth note that many areas of Borneo currently remain understudied, and there exists a massive paucity of information regarding the functional ecology of fish species throughout the island. Simultaneously, the ecosystems of Borneo are currently experiencing the massive anthropogenic pressures of deforestation and conversion

to monoculture in the form of oil palm plantation. Borneo currently experiences the highest rate of deforestation in the world [Bradshaw et al., 2008], with nearly 80% of the land surface of Sabah and Sarawak affected by high-impact logging operations between 1990 and 2009 [Bryan et al., 2013]. Anthropogenic environmental degradation currently are exceeding our ability to study and inventory Bornean icthyofauna, and without a comprehensive knowledge of it's diversity and the ecological processes therein, we stand little chance of effectively conserving this hugely scientifically important group and therefore run the risk of losing much of this diversity before we fully understand it.

Degradation of riparian forests often represents a significant threat to riverine systems, and can have far reaching effects on stream communities through mechanisms such as the alteration of light and thermal regimes, increased sedimentation and the disruption of community structure [Pusey and Arthington, 2003]. However conclusions drawn from studies quantifying the impact of habitat alteration on south-east Asian stream communities are mixed. Martin-Smith [1998a] states that fish communities from undisturbed and logged forest exhibit some differences, but argues these differences cannot be unambiguously attributed to logging activity and that mesohabitats are more important to the persistence of fish species [Martin-Smith, 1998b]. In larger reaches of water, South-East Asian fish communities are known to stratify vertically in the water column, facilitating the partitioning of resources [Dudgeon, 1999]. This can mean habitat modification can differentially effect fish taxa, for example Iwata et al. [2003] observed a distinctive reduction in diversity across much of the benthic community in secondary forest reaches due to alterations to the depositional characteristics of these streams.

In order to fully understand the potential for a species to persist in a modified habitat one must first identify how the essential ecological interactions that a species experiences are in turn modified. One of the key aspects of this being the characterisation of diet. However despite the cyprinids of Borneo being one of the regions most diverse groups, as well as important ecologically, economically and exhibiting one of the highest levels of endemism [Sulaiman and Mayden, 2012], next to nothing is known about the feeding ecology of these species, save some broad field notes compiled by Inger and Chin [2002].

Taking this into consideration, the current study aims to characterise the diets of three locally abundant cyprinid species, *Nematabramis everetti*, *Rasbora sumatrana*, both surface strata species, and *Tor douronensis*, a benthic species, across varying degrees of logging intensity in order to identify; the key resources these species exploit in the headwater streams of the Kalabakan basin, how timber extraction modifies resource use, and whether these species are differentially vulnerable to habitat modification in a vertically compressed environment.

2 Methods

Study Site and Sample Collection

Field work was carried out in conjunction with, and based at the Stability of Altered Forest Ecosystems project [Ewers et al., 2011]. Located in South-Eastern Sabah, within the Kalabakan Forest Reserve (4°43'N, 117°35'E), the SAFE project area comprises 7,200 ha of previously logged lowland dipterocarp forest designated for conversion to oil palm plantation by the Malaysian government, and contains multiple stream catchments chosen to be similar in latitude, slope and elevation so as to minimise factors which could potentially confound the effects of land-use change. Samples were collected between the 9th-30th April 2014 from three streams which had been subjected to varying logging intensity, a stream located inside the Virgin Jungle Reserve (VJR) which had been subjected to some logging, but still retained most old-growth features, a stream located in a Logged Forest plot (LF) which had been logged twice, and a stream which had experienced multiple logging events (MLF). Fish were collected by carrying out 200m transects along the chosen streams over 3-5 consecutive days, throwing a 9 foot cast net approximately every 10m. Upon capture fish were identified using Inger and Chin [2002] and the SAFE freshwater fish list Hoek-Hui [2013]. All Individuals of the three focal species greater than 7cm were retained, and a latex catheter and syringe was inserted down the alimentary canal into the stomach of the individual, following which gastric lavage was carried out, flushing stomach contents with 5ml distilled water. All fish were then released unharmed at point of capture. Water was drained from the stomach contents using filter-paper and stomach contents stored in 95% ethanol at approximately 0°C for 1 week in the field, after which they were transported to laboratory conditions and stored at −80°C.A breakdown of species and locations from which stomach samples were collected can be found 1. An initial broad scale morphological assessment was carried out on all stomach samples in order to inform downstream analyses, from which gut contents were found to be composed of arthropod tissues, indicated by the presence of exoskeleton fragments, as well as terrestrial and aquatic plant tissues.

DNA Extraction and PCR

Total DNA extractions were initially performed using a CTAB-based extraction protocol in order to maximise DNA recovery from plant tissues. Total gut contents were initially homogenised and incubated for 1 h in extraction buffer (100 mM Tris HCL pH 8.0, 50 mM EDTA pH 8.0, 150 mM NaCL, CTAB 2%), extractions were then carried out using the protocol outlined by Karp et al. [1998] for plant DNA, with an overnight isopropanol precipitation, eluted in a final volume of 35μl TE buffer. In order to maximise recovery of non-plant DNA gut contents underwent a second DNA extraction using the DNA stool mini kit (Qiagen) following manufacturers instructions. Extracted samples from both protocols were then pooled on an equal volume basis (n = 6-15) according to species and location from which gut contents were collected. DNA pools were then purified using AMPure paramagnetic beads (Agincourt) and re-eluted in a final volume of 50μl. DNA extracts then underwent three seperate amplification experiments per pool.

For the identification of plant prey species, primer pairs h1aF and h2aR targeting the rbcL chloroplastic gene region [Poinar et al., 1998], and G and H, targeting the trnL P6 loop(UAA intron) [Taberlet et al., 2007] were selected and minor modifications made. Namely, the removal of two nucleotides from the 3' prime end of rbcL h1aF, and the addition of an A and T to the 5' prime end of rbcL h1aR, these modifications bring the primer melting temperatures within the range of the KAPA HiFi HotStart ReadyMix PCR kit (KapaBiosystems). Primers ZBJ-ArtFlc and ZBJ-ArtRlc targeting the COI mitochondrial gene region [Zeale et al., 2011] were selected to amplify invertebrate prey species. An ambiguity was added to the third nucleotide from the 3' prime end of ZBJ-ArtRlc which had the effect of increasing the BLAST hit score when tested against arthropod mitochondrial genomes from GenBank. All primers were selected due to their ability to amplify a wide range of target species, as well as short amplicon size and a pre-existing body of literature indicating them as suitable candidates for dietary studies (see Table 2 for full details of modified primer sequences). PCR protocols were optimised for a final reaction volume of 25μl, containing 12.5μl KAPA HiFi HotStart ReadyMix, 0.5μl forward and reverse primers (0.25μm) and 5μl DNA extract. PCR profiles consisted of an initial denaturation step of 95°C for 3 minutes, followed by 45 cycles of 95°C for 30 seconds, 30 seconds at primer specific annealing temperature, 72°C for 30 seconds and then 72°C for 5 minutes. All PCR experiments were run with positive and negative controls. Pineapple DNA extract was used as a positive control in the case of chloroplastic markers and Coleopteran DNA in the case of the COI marker. PCR grade water was used in negative controls

PCR products were run on a 1.5% agarose gel and subsequently excised at the full amplicon length and purified using the Qiagen gel extraction kit. Concentration of purified PCR products was then quantified using the Qubit 2.0 fluorometer (Invitrogen) and concentration was normalised before re-pooling at the species/location level. PCR products were then subjected to multiplex library preparation using Nextera XT reagents (Illumina), by first attaching dual index barcodes in a shortened second PCR stage (50μl

total reaction volume with 25μl KAPA Hifi ReadyMix, 5μl index primers, 5μl product from first PCR stage and 10μl PCR grade water) at 95°C for 3 minutes, followed by 8 cycles of 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds and then 72°C for 5 minutes. Libraries were then validated by bioanalyzer, concentrations quantified by qPCR, normalised and pooled before 151bp paired end reads were sequenced on a single lane of Illumina MiSeq using v3 reagent kit following manufacturers instructions.

Sequence analysis and taxonomic assignation

Raw sequence reads were de-multiplexed using CASSAVA v1.8.2 software (Illumina). Reads then underwent an initial quality filter using the NGS QC toolkit v2.3.2 [Patel and Jain, 2012] with a cutoff nucleotide quality score of 30, and a cutoff read length of high quality of 70%. Following filtering of low-quality reads, sequence analysis followed the methods outlined by Quéméré et al. [2013] and De Barba et al. [2014] using the OBITools (http://www.prabi.grenoble.fr/trac/OBITools). First, forward and reverse reads corresponding to a single molecule were alligned, and a single consensus read produced using the program *illuminapairedend* taking into account base call quality scores in the computation. Primer sequences from the three gene regions were then identified and removed using *ngsfilter*, retaining reads with a maximum of 2bp mismatch on primer sequences for further analysis. Libraries were then split according to gene region using *obisplit*. Identical sequences were clustered using *obiuniq*, retaining information on read counts. Sequences with a count higher than 1 were retained and then filtered by length, based on full length amplicon sizes produced by each of the three primer pairs using *obigrep*. Amplicons produced by ZBJ-ArtFlc and ZBJ-ArtRlc were filtered at 90-160bp, amplicons produced by h1aF and h2aR at 90-120bp, and those produced by G and H at 10-140 bp. Sequence variants were then identified using *obiclean* and tagged as 'head' 'internal' and 'singleton' (see Quéméré et al. [2013]). Only 'head' and 'singleton' sequences were retained for taxonomic assignment.

In order to assign taxa to sequences generated from gut contents without *a priori* knowledge of the species present in the study area sequence reference databases were built for each metabarcode marker by downloading the relevant gene regions from the NCBI nucleotide (nt) database. Namely, trnL (Viridiplanteae) for G and H, rbcL (Magnoliophyta) for h1aF and h2aR, and COI (Eukaryota) for ZBJ-ArtFlc and ZBJ-ArtRlc. Gut content sequences were then BLAST searched against reference databases using BLASTn with default settings as part of the standalone BLAST+ software [Camacho et al., 2009]. The popular metagenomic analysis software MEGAN5 [Huson et al., 2007] was employed for taxonomic binning based on sequence similarity. MEGAN employs a lowest common ancestor (LCA) algorithm, whereby reads are assigned to the LCA of hits generated from the BLAST output file. This can be seen as a somewhat conservative estimation of true biological diversity, however it does possess the benefit of reducing the risk of false positive taxonomic assignations. In the case of the current study, only those

BLAST hits with the highest bit-score were considered for taxonomic assignation (top-percent filter: 0.1%). Hits with a bit-score that fell below 50, or an expect value of greater than $1e^{-5}$ were also omitted. There is an important caveat to note when employing this method in that taxonomic assignations made are strictly putitative and represent the taxon, or lowest common ancestor of the taxa, present in the NCBI database that achieve the most significant allignment.

Statistical analyses

Due to a number of factors which can affect read frequencies generated from metabarcoding, such as PCR bias and primer specificity, dietary data was considered as presence-absence for statistical analysis and data generated from trnL and rbcL markers were analysed seperately. Presence-absence data were then used to calculate pairwise dissimilarity matrices for all species/stream combinations using the Bray-Curtis dissimilarity index [Bray and Curtis, 1957], where gut contents containing exactly the same species would generate a Bray-Curtis index of 0 and those sharing no species would generate an index of 1. A permutational multivariate analysis of variance (PERMANOVA) test was then conducted on the resulting dissimilarity matrix in order to establish whether interspecific differences or logging intensity explained the most variation in gut content composition. Non-metric mulitidimensional (NMDS) scaling was employed to visualise relative similarity or dissimilarity of gut contents, with metabarcodes similar in composition appearing closer together, and those more dissimilar further apart. NMDS ordination, PERMANOVA tests and indices calculations were computed using the R library *Vegan* [Oksanen et al., 2007].

3 Results

Next generation sequencing generated 15,114,221 reads, of these 6,627,586 (43.85%) passed Illumina demultiplexing filters. After the various quality control and filtering steps were applied, this corresponded to 539,139 full length barcodes of 5,522 strictly uniqe sequences (h1aF and h2aR: 194,488 full length and 1790 unique, G and H: 344,603 full length and 3,715 unique, and ZBJ-ArtFlc and ZBJ-ArtRlc: 48 full length and 17 unique). A total of 63 putitative taxa were identified from the primer pair h1aF and h2aR across all samples (Table 3), comprising of 16 species level assignations, 13 genus level, 4 subfamily level, 13 family level and 27 at higher taxonomic rankings. 19 putitative taxa were identified by Primer pairs G and H (Table 4) comprising of 3 species level assignations, 2 genus level, 8 family level and 6 at higher rankings. The mitochondrial primers ZBJ-ArtFlc and ZBJ-ArtRlc (Table 5) produced far fewer full length barcodes than the chloroplastic markers, and only two taxa were identified from the resulting amplicons, namely the staphylinid genus *Hydrosmecta*, and a pathgenic fungus of the genus *Pythium*. As a result of the relative paucity of amplicons produced by ZBJ-ArtFlc and ZBJ-ArtRlc these data were excluded from statistical analyses. Metabarcodes produced by h1aF and h2aR suggested Taxon richness was almost uniformly lower in logged sites than unlogged sites, save *Tor douronensis*, which exhibited higher taxon richness in unlogged sites, and *Rasbora sumatrana* which exhibited the highest taxon richness in the twice logged site (LF). However, this did not hold true for trnL metabarcodes, in which taxon richness was uniform across sites with the exception of *Rasbora* LF, which again exhibited higher taxon richness. Dissimilarity matrices generated from the two chloroplastic markers were shown to be non-correlative (Mantel test r = -0.06551, P = 0.621). Permutational multivariate analysis of variance of rbcL metabarcodes showed species to be a significant factor in stratifying gut contents (1000 permutations, F = 1.734, df = 2, P ¡ 0.05), explaining 46% of observed variance, however this did not hold true for trnL metabarcodes (1000 permutations, F = 2.479, df = 2, P = 0.0649) however similar levels of variance were explained (55%). Location was not found to be a significant factor in stratifying gut contents in either dataset (rbcl data: 1000 permutations, F = 0.659, df = 2, P = 0.8901, trnL data: 1000 permutations, F = 0.880, df = 2, P = 0.514) explaining just 24% and 30% of variance respectively. NMDS ordination of the two datasets showed some broad scale congruity, indicating differential responses to logging intensity between species. *Nematabramis* consistantly exhibitied the least variance over site treatments (Figure 1 and 2). *Rasbora* and *Tor* both exhibited greater variance than *Nematabramis* but in different directions and with different effect sizes. In both ordination plots there exists little overlap between species. a certain amount of caution should be exercised when interpreting results generated from these data as short-length barcodes commonly used in dietary studies lack taxonomic resolution, causing community data generated from these barcodes to be somewhat degenerate, this is exemplified by the fact that NMDS of trnL metabarcode data achieved a stress level of zero, whereas the longer length rbcL barcodes achieved a more normal stress level of 0.1.

4 Discussion

DNA analysis of cyprinid stomach contents presented in the current study indicates diets to be dominated by terrestrial plant tissue of angiosperm origin. Herbivory, including frugivory and granivory has been documented in a wide range of fish species across the Indomalayan Region [Corlett, 1998]. This is however, seemingly in contrast to previous studies investigating the trophic linkages of tropical headwater streams, which suggest a high proportion of the icthyofauna associated with these habitats predate terrestrial and aquatic invertebrate species, as well as other vertebrate species such as larval fish and amphibians [Martin-Smith and Laird, 1998, Ashraf et al., 2011, Mantel et al., 2004]. Terrestrial arthropod inputs into tropical streams has also been shown to exhibit strong seasonality, with Chan et al. [2008] documenting dry season inputs at approximately half that of the wet season. Considering samples used in the current study were collected throughout the month of April, this would be near the peak of low abundance of arthropod inputs. The importance of terrestrial inputs to the vertically compressed headwater streams of North Borneo was noted by Inger and Chin [2002] who, despite placing *Rasbora sumatrana* and *Nematabramis everetti* as second order predators, identified exogenous vascular plants as an important food source to these species. Combining the findings of previous studies with the current data supports a broad euryphagosity exhibited by the ichthyofauna of Bornean headstreams, and whilst terrestrial invertebrates may contribute more energetically to the diets of these fish, the seasonal and stochastic nature of this resource means the fruit, seeds and vascular tissues of plants provide a more stable food source. It is also worth note that the herbivory exhibited by these fish species represents an ecologically important interaction at the ecosystem level, facilitating nutrient transfer between terrestrial and aquatic ecosystems, and increasing the availability of particulate organic matter, carbon and nitrogen in the downstream food web.

Taxon richness of stomach samples from old growth forest was generally greater than that of repeatedly logged forest, however taxon richness alone does not mean much and in order to understand the mechanisms by which habitat disturbance modifies allochthnous resource use in headwater streams, it is neccesary to examine the taxonomic turnover between habitat types. Many of the taxa present only in metabarcodes from old growth forest represent important resource species, for example *Diospyros* (Ebanaceae), Laminaceae, and *Vitis* (Vitaceae) all contain fruiting species known to be utilised by terrestrial frugivores [Cannon et al., 2007, Bramley, 2009, Shaffiq et al., 2013], the seeds of some Zingiberaceae are utilised by ants [Pfeiffer et al., 2004], and species of the genus *Trigonia* contain large nectar filled flowers visited by a number of nectivorous species [Kato, 2005]. All of these resource species could be readily digested by fish should they happen to enter the aquatic system. This is not to say that sites experiencing greater logging intensity do not contain resource species that could be readily utilised by freshwater fish, but the higher proportions and diversity of these species in old growth forests, combined with a greater degree of canopy cover, contribute towards a more stable resource base which could sustain

fish populations in the absence of more energetically valuable resources such as terrestrial invertebrates.

Analysis of variance between metabarcode datasets indicated species to be a better predictor of diet than logging intensity alone, and visualisations of NMDS ordinations suggest that species respond differentially to habitat disturbance, as points do not tend towards, or away from any clustering pattern based on site alone. Considering the individual ecologies of these species, we can begin to explain the individual responses to terrestrial habitat disturbance. *Tor douronensis* was unique in that metabarcodes generated from this species exhibited higher taxonomic richness in the multiple-logged forest than old growth. *Tor*, whilst having been observed to readily consume terrestrial plant and invertibrate matter, has also been observed to feed on autochthnous food sources such as aquatic invertebrates and detritus [Nguyen, 2008]. The increase in terrestrial plant taxa observed, may indicate a move away from benthic feeding habits, as decreasing canopy cover can lead to the proliferation of filamentous green algae, which is both unpalatable and energetically unrewarding [Bunn et al., 1999]. Similarly The increased taxonomic richness of terrestrial plant matter in *Rasbora* in twice-logged habitat when compared to old growth may represent the decreasing abundance of invertebrate prey inputs in more open forest Chan et al. [2008] or less energetically rewarding plant matter.

There are some important limitations take into account when considering the results of the current study, perhaps most importantly is the conspicuous lack of aquatic plants or terrestrial invertebrates from the molecular data, as both of these groups were identified from initial morphological analysis of gut contents. The near absence of terrestrial invertebrates from is likely due to differential degradation rates between plant and animal DNA, which would be amplified by the presence of digestive enzymes from stomach samples. As such, the current data probably represents an under-representation of arthropod prey species. The mechanism behind the absence of algal species from metabarcode data is somewhat less clear, as previous work with the trnL UAA intron has reported successful amplification of these taxa [Taberlet et al., 1991]. It may be the case that the short-length barcodes generated by the P6 loop of the UAA intron were infact too short for BLAST to reliably identify. Another possible explanation would be a lack of molecular data of closely related taxa for this gene region in the NCBI database. Despite the wealth of amplicons generated, the lack of taxonomic resolution afforded by the G and H primer pair also hampered the assignment of taxa and the resulting statistical analyses of the data generated. this indicates that this particular marker should only be used in ecological studies when a comprehensive reference dataset of local species is available *a priori*. The incongruence of the data generated by the two chloroplastic markers is indicative of primer specificity, and serves to illustrate primers should be extensively tested *in scilico* before application to an ecological setting or generation of biodiversity estimates.

Nonetheless, the current study offers important insights into the ecology of *Rasbora sumatrana*, *Nematabramis everetti* and *Tor douronensis*, placing them as euryphagous omnivores which play a key role

in nutrient cycling between terrestrial and aquatic systems and contributing towards ecosystem functioning. furthermore, the results presented illustrate the role habitat disturbance can play in reducing the quality and stability of the resource base utilised by these species, with the distinct potential to disrupt the structure and functioning of these communities. This serves to emphasise the importance of riparian buffer zones in mitigating the detrimental effects of timber extraction on freshwater communities. The use of molecular techniques afforded a level of taxonomic resolution that would not have been possible through classical methods, however further research should be done in order to fully explore the dietary diversity these species exhibit, and to what extent they rely on the different food sources they exploit.

5 Acknowledgements

I greatfully Acknowledge SAFE project and Sime Darby foundation for funding this project. I extend great grattitude to Dr. Robert Ewers for offering me this opportunity and providing advice throughout. I also extend thanks to Prof Alfried Vogler who offered me much training and advice in molecular techniques. I also extend gratitude to Dr Vijay Kumar and Dr Chris Voo of Universiti Malaysia Sabah for allowing me to conduct labwork in-situ on the island of Borneo and offering guidance and advice throughout.

6 Tables and Figures

Table 1: Sample pools

	Species		
	Rasbora sumatrana	Nematabramis everetti	Tor douronensis
VJR	8	15	11
LF	6	0	0
MLF	7	7	9
Total	21	22	20

Table 2: Primer Sequences

Target.Gene	Target.Taxon	Name	Modified.Sequence	Fragment.Size..bp.	Annealing.Temperature..c.	Reference
rbcL	Plants	rbcL h1aF (F)	GGCAGCATTCCGAGTAACTCC	97	63.9	Poinar et al (1998)
-	-	rbcl h2aR (r)	ATCGTCCTTTGTAACGATCAAG	-	60.1	-
trnL	Plants	g (F)	GGGCAATCCTGAGCCAA	10 143	61.1	Taberlet et al (2006)
-	-	h (r)	CCATTGAGTCTCTGCACCTATC	-	61.4	-
COI	Arthropoda	ZBJ-ArtFlc (F)	AGATATTGGAACWTTATATTTTATTTTTGG	130	60.4	Zeale et al (2010)
-	-	ZBJ-ArtRlc (r)	GWACTAATCAATTWCCAAATCCWCC	-	61.1	-

Table 3: Presence-absence of putitative taxa identified by primer pair h1aF and h2aR

Taxa	Rs.VJR	Rs.LF	Rs.MLF	Ne.VJR	Ne.MLF	Td.VJR	Td.MLF
Mesangiospermae	X	X	X	X	X	X	X
Hedyosmum costaricense		X					
eudicotyledons		X					X
Pentapetalae	X	X	X	X			X
asterids	X	X	X	X		X	X
Aralia sp.		X					
Ilex						X	
Asteraceae							X
Diospyros	X						
Careya arborea							X
lamiids		X	X				
Loganiaceae		X					
Mostuea brunonis		X					
Uncaria	X	X	X			X	
Rubioideae							X
Solanales							X
Solandra grandiflora							X
Caryophyllales		X					
Drosera hamiltonii		X					
rosids	X	X	X	X	X	X	X
Zinowiewia australis			X				
Cucurbitales			X				
Fabaceae	X	X	X				
Mimosa	X			X			
Papilionoideae		X		X			
Dalbergieae			X			X	
Spatholobus						X	X
Amphithalea micrantha	X						
Echinosophora koreensis							X
Lithocarpus		X	X				X
Kiggelaria			X				
Wielandieae	X						
Trigonia	X						
Coussapoa villosa							X
malvids		X	X	X		X	
Malvales	X	X					
Dipterocarpoideae	X	X	X	X		X	X
Byttneria						X	
Thymelaeaceae							X
Sapindales		X					
Rutaceae	X	X	X	X	X	X	X
Neobyrnesia suberosa						X	
Sapindaceae	X						X
Nephelium lappaceum							X
Sapindus mukorossi							X
Leea	X	X		X		X	X
Urophysa henryi		X					
Araceae	X						
Petrosaviidae	X	X	X	X	X		
Asparagales		X	X				
Orchidaceae	X						
Arecaceae						X	
Raphia palm		X					
Eichhornia crassipes			X				
Eichhornia diversifolia		X					
Poales		X					
Poaceae	X	X		X	X		
Paspalum			X			X	
Zingiberaceae	X						
Pandanaceae			X				
Stemonaceae	X						
Magnoliidae	X	X					X
Magnoliales	X	X	X	X			X
Taxon richness	24	31	21	13	5	15	23

Table 4: Presence-absence of putitative taxa identified by primer pair G and H

Taxa	Rs.VJR	Rs.LF	Rs.MLF	Ne.VJR	Ne.MLF	Td.VJR	Td.MLF
Mesangiospermae	X	X					
Lamiaceae				X			
Convolvulaceae						X	X
Solanoideae	X	X	X	X	X		X
Papilionoideae		X					
Malpighiales		X					
Urticaceae		X	X	X	X		X
Dipterocarpoideae	X	X				X	X
Shorea		X					
Pterospermum heterophyllum							X
Melastomataceae		X	X	X	X		X
Anacardiaceae						X	
Rutaceae						X	
Storthocalyx sp.		X					
Vitaceae					X	X	X
Cayratia eurynema						X	
Vitis						X	
Liliopsida			X				
Myristicaceae	X	X					
Taxon richness	4	10	4	4	4	7	7

Table 5: Presence-absence of putitative taxa identified by primer pair ZBJ-ArtFlc and ZBJ-ArtRlc

Taxa	Rs.VJR	Rs.LF	Rs.MLF	Ne.VJR	Ne.MLF	Td.VJR	Td.MLF
Hydrosmecta	X						
Pythium			X		X		X

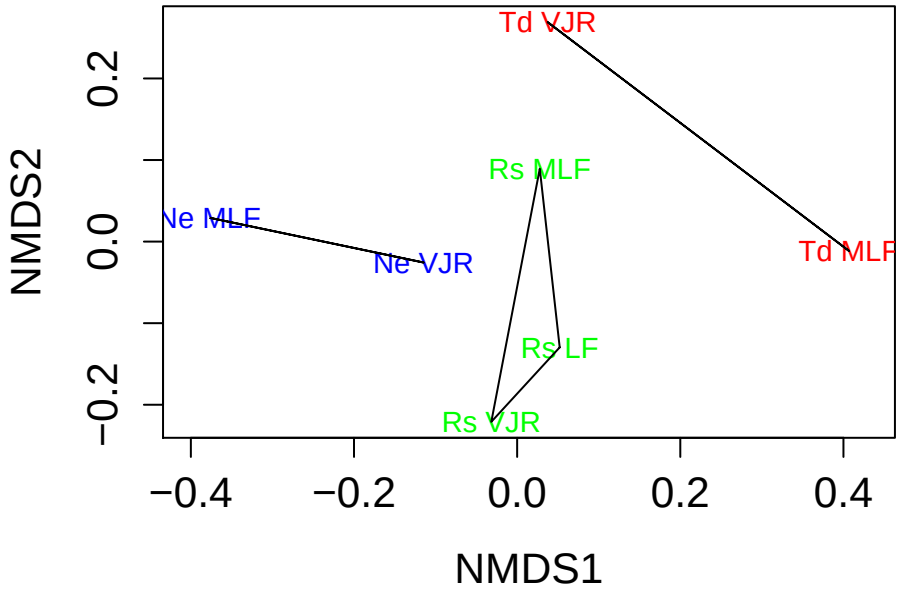

Figure 1: Non-metric multidimensional scaling of rbcL metabarcode data

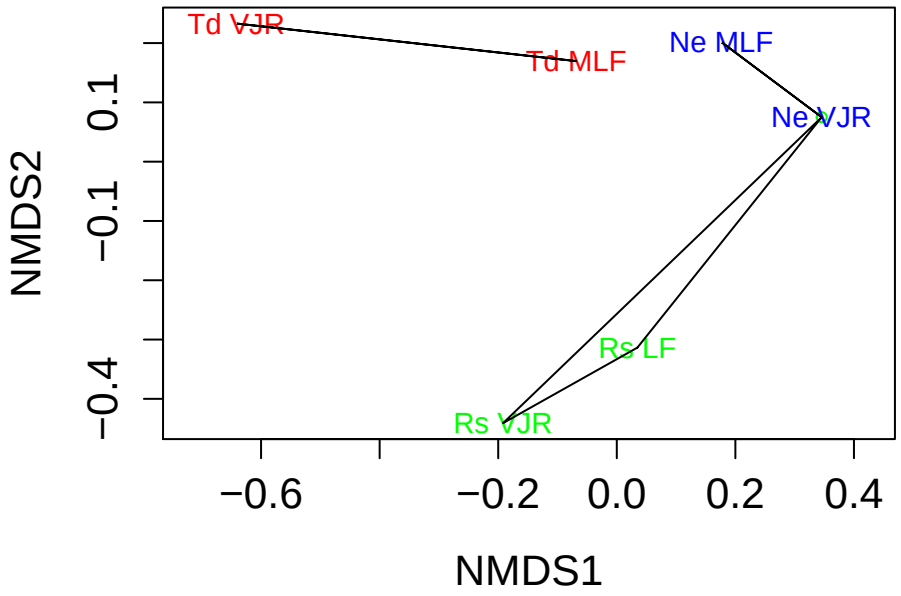

Figure 2: Non-metric multidimensional scaling of trnL metabarcode data

References

Antton Alberdi, Inazio Garin, Ostaizka Aizpurua, and Joxerra Aihartza. The foraging ecology of the mountain long-eared bat plecotus macrobullaris revealed with dna mini-barcodes. *PloS one*, 7(4): e35692, 2012.

Miguel Alcaide, Ciro Rico, Santiago Ruiz, Ramón Soriguer, Joaquín Muñoz, and Jordi Figuerola. Disentangling vector-borne transmission networks: a universal dna barcoding method to identify vertebrate hosts from arthropod bloodmeals. *PLoS One*, 4(9):e7092, 2009.

Rose L Andrew, Louis Bernatchez, Aurélie Bonin, C Alex Buerkle, Bryan C Carstens, Brent C Emerson, Dany Garant, Tatiana Giraud, Nolan C Kane, Sean M Rogers, et al. A road map for molecular ecology. *Molecular ecology*, 22(10):2605–2626, 2013.

MA Ashraf, MJ Maah, and I Yusoff. Bioaccumulation of heavy metals in fish species collected from former tin mining catchment. *International Journal of Environmental Research*, 6(1):209–218, 2011.

Moulay Abdeljalil Ait Baamrane, Wasim Shehzad, Ahmed Ouhammou, Abdelaziz Abbad, Mohamed Naimi, Eric Coissac, Pierre Taberlet, and Mohammed Znari. Assessment of the food habits of the moroccan dorcas gazelle in msabih talaa, west central morocco, using the trnl approach. *PloS one*, 7 (4):e35643, 2012.

Petru BANARESCU and PG Bianco. A contribution to fish fauna of kapuas river, kalimantan barat, indonesia borneo: Cyprinidae. *Cybium*, 8(1):59–70, 1984.

Adam Barnett, Kevin S Redd, Stewart D Frusher, John D Stevens, and Jayson M Semmens. Non-lethal method to obtain stomach samples from a large marine predator and the use of dna analysis to improve dietary information. *Journal of experimental marine biology and ecology*, 393(1):188–192, 2010.

Dennis A Benson, Ilene Karsch-Mizrachi, David J Lipman, James Ostell, and David L Wheeler. Genbank. *Nucleic acids research*, 36(suppl 1):D25–D30, 2008.

PFL Boreham and CE Ohiagu. The use of serology in evaluating invertebrate prey-predator relationships: a review. *Bulletin of entomological research*, 68(02):171–194, 1978.

GEORGE A Boulenger. Xxvi.descriptions of new freshwater fishes from borneo. *The Annals and Magazine of Natural History*, 13(75):245–251, 1894.

Corey JA Bradshaw, Navjot S Sodhi, and Barry W Brook. Tropical turmoil: a biodiversity tragedy in progress. *Frontiers in Ecology and the Environment*, 7(2):79–87, 2008.

Gemma LC Bramley. The genus callicarpa (lamiaceae) on borneo. *Botanical Journal of the Linnean Society*, 159(3):416–455, 2009.

J Roger Bray and John T Curtis. An ordination of the upland forest communities of southern wisconsin. *Ecological monographs*, 27(4):325–349, 1957.

Ralf Britz, Maurice Kottelat, and Tan Heok Hui. Fangfangia spinicleithralis, a new genus and species of miniature cyprinid fish from the peat swamp forests of borneo (teleostei: Cyprinidae). *Ichthyological Exploration of Freshwaters*, 22(4):327, 2011.

Barry W Brook, Corey JA Bradshaw, Lian Pin Koh, and Navjot S Sodhi. Momentum drives the crash: Mass extinction in the tropics1. *Biotropica*, 38(3):302–305, 2006.

David S Brown, Simon N Jarman, and William OC Symondson. Pyrosequencing of prey dna in reptile faeces: analysis of earthworm consumption by slow worms. *Molecular ecology resources*, 12(2):259–266, 2012.

Jane E Bryan, Philip L Shearman, Gregory P Asner, David E Knapp, Geraldine Aoro, and Barbara Lokes. Extreme differences in forest degradation in borneo: Comparing practices in sarawak, sabah, and brunei. *PloS one*, 8(7):e69679, 2013.

SE Bunn, PM Davies, and TD Mosisch. Ecosystem measures of river health and their response to riparian and catchment degradation. *Freshwater Biology*, 41(2):333–345, 1999.

Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. Blast+: architecture and applications. *BMC bioinformatics*, 10(1):421, 2009.

Charles H Cannon, Lisa M Curran, Andrew J Marshall, Mark Leighton, et al. Beyond mast-fruiting events: community asynchrony and individual dormancy dominate woody plant reproductive behavior across seven bornean forest types. *CURRENT SCIENCE-BANGALORE-*, 93(11):1558, 2007.

Michael A Chadwick. Stream ecology: Structure and function of running waters. *Freshwater Biology*, 53 (9):1914–1914, 2008.

Eric KW Chan, Yixin Zhang, and David Dudgeon. Arthropod'rain'into tropical streams: the importance of intact riparian forest and influences on fish diets. *Marine and Freshwater Research*, 59(8):653–660, 2008.

Robin L Chazdon, Carlos A Peres, Daisy Dent, Douglas Sheil, Ariel E Lugo, David Lamb, Nigel E Stork, and Scott E Miller. The potential for species conservation in tropical secondary forests. *Conservation Biology*, 23(6):1406–1417, 2009.

Richard T Corlett. Frugivory and seed dispersal by vertebrates in the oriental (indomalayan) region. *Biological Reviews of the Cambridge Philosophical Society*, 73(04):413–448, 1998.

Emmanuel Corse, Caroline Costedoat, Remi Chappaz, Nicolas Pech, JEAN-FRANÇOIS MARTIN, and Andre Gilles. A pcr-based method for diet analysis in freshwater organisms using 18s rdna barcoding on faeces. *Molecular ecology resources*, 10(1):96–108, 2010.

M De Barba, C Miquel, F Boyer, C Mercier, D Rioux, E Coissac, and P Taberlet. Dna metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular ecology resources*, 14(2):306–323, 2014.

David Dudgeon. *Tropical Asian streams: zoobenthos, ecology and conservation*, volume 1. Hong Kong University Press, 1999.

Robert M Ewers, Raphael K Didham, Lenore Fahrig, Gonçalo Ferraz, Andy Hector, Robert D Holt, Valerie Kapos, Glen Reynolds, Waidi Sinun, Jake L Snaddon, et al. A large-scale forest fragmentation experiment: the stability of altered forest ecosystems project. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1582):3292–3302, 2011.

AWCT Herre. Additions to the fish fauna of malaya and notes on rare or little known malayan and bornean fishes. *Bulletin of the Raffles Museum*, 16:27–61, 1940.

T. Hoek-Hui. Freshwater fish list with pictures. available online at: http://www.safeproject.net/projects/biodiversity/freshwater-fish-list-with-pictures/ (accessed on: 23/05/14)., 2013.

Daniel H Huson, Alexander F Auch, Ji Qi, and Stephan C Schuster. Megan analysis of metagenomic data. *Genome research*, 17(3):377–386, 2007.

Robert F Inger and Phui Kong Chin. fresh-water fishes of north borneo. 2002.

Tomoya Iwata, Shigeru Nakano, and Mikio Inoue. Impacts of past riparian deforestation on stream communities in a tropical rain forest in borneo. *Ecological Applications*, 13(2):461–473, 2003.

Mark A Jervis. *Insects as natural enemies*. Springer, 2005.

José A Jurado-Rivera, Alfried P Vogler, Chris AM Reid, Eduard Petitpierre, and Jesús Gómez-Zurita. Dna barcoding insect–host plant associations. *Proceedings of the Royal Society B: Biological Sciences*, 276(1657):639–648, 2009.

Angela Karp, Peter G Isaac, David S Ingram, et al. *Molecular tools for screening biodiversity: plants and animals*. Chapman & Hall Ltd, 1998.

Makoto Kato. Ecology of traplining bees and understory pollinators. In *Pollination Ecology and the Rain Forest*, pages 128–133. Springer, 2005.

Baek Jun Kim, Nam Sook Lee, and Sang Don Lee. Feeding diets of the korean water deer (hydropotes inermis argyropus) based on a 202 bp rbcl sequence analysis. *Conservation Genetics*, 12(3):851–856, 2011.

Maurice Kottelat and Tony Whitten. *Freshwater biodiversity in Asia: with special reference to fish*, volume 343. World Bank Publications, 1996.

MA Larkin, Gordon Blackshields, NP Brown, R Chenna, Paul A McGettigan, Hamish McWilliam, Franck Valentin, Iain M Wallace, Andreas Wilm, Rodrigo Lopez, et al. Clustal w and clustal x version 2.0. *Bioinformatics*, 23(21):2947–2948, 2007.

Matthieu Leray, Natalia Agudelo, Suzanne C Mills, and Christopher P Meyer. Effectiveness of annealing blocking primers versus restriction enzymes for characterization of generalist diets: unexpected prey revealed in the gut contents of two coral reef fish species. *PloS one*, 8(4):e58076, 2013.

Kelvin KP Lim. Rasbora kottelati, a new species of cyprinid fish from north-western borneo. *Raffles Bulletin of Zoology*, 43:65–74, 1995.

Sukhmani Kaur Mantel, Maria Salas, and David Dudgeon. Foodweb structure in a tropical asian forest stream. *Journal of the North American Benthological Society*, 23(4):728–755, 2004.

Keith M Martin-Smith. Effects of disturbance caused by selective timber extraction on fish communities in sabah, malaysia. *Environmental Biology of Fishes*, 53(2):155–167, 1998a.

KM Martin-Smith. Relationships between fishes and habitat in rainforest streams in sabah, malaysia. *Journal of Fish Biology*, 52(3):458–482, 1998b.

KM Martin-Smith and LM Laird. Depauperate freshwater fish communities in sabah: the role of barriers to movement and habitat quality. *Journal of Fish Biology*, 53(sA):331–344, 1998.

Isabelle Meusnier, Gregory AC Singer, Jean-François Landry, Donal A Hickey, Paul DN Hebert, and Mehrdad Hajibabaei. A universal dna mini-barcode for biodiversity analysis. *BMC genomics*, 9(1): 214, 2008.

Thuy TT Nguyen. Population structure in the highly fragmented range of tor douronensis (cyprinidae) in sarawak, malaysia revealed by microsatellite dna markers. *Freshwater biology*, 53(5):924–934, 2008.

Johannes Oehm, Anita Juen, Karin Nagiller, Sigrid Neuhauser, and Michael Traugott. Molecular scatology: how to improve prey dna detection success in avian faeces? *Molecular ecology resources*, 11(4): 620–628, 2011.

Jari Oksanen, Roeland Kindt, Pierre Legendre, Bob OHara, M Henry H Stevens, Maintainer Jari Oksanen, and MASS Suggests. The vegan package. *Community ecology package*, 2007.

Ravi K Patel and Mukesh Jain. Ngs qc toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*, 7(2):e30619, 2012.

Martin Pfeiffer, Jamili Nais, and K Eduard Linsenmair. Myrmecochory in the zingiberaceae: seed removal of globba franciscii and g. propinqua by ants (hymenoptera–formicidae) in rain forests on borneo. *Journal of tropical ecology*, 20(06):705–708, 2004.

Hendrik N Poinar, Michael Hofreiter, W Geoffrey Spaulding, Paul S Martin, B Artur Stankiewicz, Helen Bland, Richard P Evershed, Göran Possnert, and Svante Pääbo. Molecular coproscopy: dung and diet of the extinct ground sloth nothrotheriops shastensis. *Science*, 281(5375):402–406, 1998.

Bradley J Pusey and Angela H Arthington. Importance of the riparian zone to the conservation and management of freshwater fish: a review. *Marine and Freshwater Research*, 54(1):1–16, 2003.

Erwan Quéméré, Fabrice Hibert, Christian Miquel, Emeline Lhuillier, Emmanuel Rasolondraibe, Julie Champeau, Clément Rabarivola, Louis Nusbaumer, Cyrille Chatelain, Laurent Gautier, et al. A dna metabarcoding study of a primate dietary diversity and plasticity across its entire fragmented range. *PloS one*, 8(3):e58971, 2013.

Lorna Raso, Daniela Sint, Rebecca Mayer, Simon Plangg, Thomas Recheis, Silvia Brunner, Rüdiger Kaufmann, and Michael Traugott. Intraguild predation in pioneer predator communities of alpine glacier forelands. *Molecular ecology*, 2014.

Sujeevan Ratnasingham and Paul DN Hebert. Bold: The barcode of life data system (http://www. barcodinglife. org). *Molecular ecology notes*, 7(3):355–364, 2007.

Lasse Riemann, Hanna Alfredsson, Michael M Hansen, Thomas D Als, Torkel G Nielsen, Peter Munk, Kim Aarestrup, Gregory E Maes, Henrik Sparholt, Michael I Petersen, et al. Qualitative assessment of the diet of european eel larvae in the sargasso sea resolved by dna barcoding. *Biology letters*, 6(6): 819–822, 2010.

Alvln Seale. Fishes of borneo, with descriptions of four new species. *Philipp J Sci*, 5:263–289, 1910.

S Muhd Arif Shaffiq, B Japar Sidik, Z Muta Harah, and R Shiamala Devi. Marketable wild fruits of sarawak, borneo: Their mode of consumption, uses and sugar profiles. *Indian Journal of Traditional Knowledge*, 12(2):195–201, 2013.

Wasim Shehzad, Tiayyba Riaz, Muhammad A Nawaz, Christian Miquel, Carole Poillot, Safdar A Shah, Francois Pompanon, Eric Coissac, and Pierre Taberlet. Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (prionailurus bengalensis) in pakistan. *Molecular ecology*, 21(8):1951–1965, 2012.

M Stech, E Kolvoort, MJJE Loonen, K Vrieling, and JD Kruijer. Bryophyte dna sequences from faeces of an arctic herbivore, barnacle goose (branta leucopsis). *Molecular ecology resources*, 11(2):404–408, 2011.

Matthew J Struebig, Anthony Turner, Emily Giles, Felicia Lasmana, Simon Tollington, Henry Bernard, and Diana Bell. Quantifying the biodiversity value of repeatedly logged rainforests: gradient and comparative approaches from borneo. *Advances in Ecological Research*, 48:183–224, 2013.

Zohrah H Sulaiman and RL Mayden. Cypriniformes of borneo (actinopterygii, otophysi): An extraordinary fauna for integrated studies on diversity, systematics, evolution, ecology, and conservation. *Zootaxa*, 3586:359–376, 2012.

Pierre Taberlet, Ludovic Gielly, Guy Pautou, and Jean Bouvet. Universal primers for amplification of three non-coding regions of chloroplast dna. *Plant molecular biology*, 17(5):1105–1109, 1991.

Pierre Taberlet, Eric Coissac, François Pompanon, Ludovic Gielly, Christian Miquel, Alice Valentini, Thierry Vermat, Gerard Corthier, Christian Brochmann, and Eske Willerslev. Power and limitations of the chloroplast trnl (uaa) intron for plant dna barcoding. *Nucleic Acids Research*, 35(3):e14–e14, 2007.

Michael Traugott. The prey spectrum of larval and adult¡ i¿ cantharis¡/i¿ species in arable land: An electrophoretic approach. *Pedobiologia*, 47(2):161–169, 2003.

Alice Valentini, Christian Miquel, Muhammad Ali Nawaz, EVA Bellemain, Eric Coissac, FranCOis Pompanon, Ludovic Gielly, Corinne Cruaud, Giuseppe Nascetti, Patrick Wincker, et al. New perspectives in diet analysis based on dna barcoding and parallel pyrosequencing: the trnl approach. *Molecular Ecology Resources*, 9(1):51–60, 2009.

Diane Zarzoso-Lacoste, Emmanuel Corse, and Eric Vidal. Improving pcr detection of prey in molecular diet studies: importance of group-specific primer set selection and extraction protocol performances. *Molecular ecology resources*, 13(1):117–127, 2013.

Matt RK Zeale, Roger K Butlin, Gary LA Barker, David C Lees, and Gareth Jones. Taxon-specific pcr for dna barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, 11(2):236–244, 2011.