Identification of plant DNA Metabarcoding loci to increase taxonomic resolution in bat dietary analysis

Abigail Carruthers

St Johns College

Submitted for the degree of Master of Biology (MBiol)

2019

Supervised by Andreanna Welch

ABSTRACT

DNA metabarcoding combines DNA barcoding with Next-Generation sequencing, enabling the rapid identification of species in an environmental sample by comparing sequences against a database of known species. One use of this is dietary analysis by the identification of species present in faeces. The DNA in faeces is degraded, so finding a suitable barcode locus is challenging. In Cameroon, cacao is grown amongst shade trees, which, if managed correctly, can support a diversity of animal species. This biodiversity can provide ecosystem services that benefit cocoa yield, such as pest control by bats. Analysing bat diet can be used to show which shade tree species support these bats. A previous study analysing bat diet used the rbcL locus, but it provided very low taxonomic resolution. Therefore, this study aimed to find alternative metabarcoding loci to improve taxonomic resolution. The loci trnL and ITS1 were selected, and both provided higher taxonomic resolution than rbcL. trnL also had higher amplification success, suggesting it is the most suitable locus for bat dietary analysis. However, variation in the diet depending on the locus used showed that no locus is ideal. Nevertheless, testing these loci provided information about bat diet. This included similarities and differences between plant taxa detected in the faeces of insectivorous and frugivorous bats. Also, detecting cacao in insectivorous bat faeces suggests a possible link between bats and cacao pests. In general, this study identified new loci and demonstrated they are effective in analysing bat diet to identify important shade tree species.

CONTENTS

List of Tables and Figures

List of Appendices

Abbreviations

BOL	Barcode Of Life
BOLD	Barcode Of Life Database
CBOL	Consortium for the Barcode Of Life
DNA	Deoxyribonucleic acid
FOO	Frequency Of Occurrence
ITS	Internal transcribed spacer (barcode locus)
MOTU	Molecular Operational Taxonomic Unit
PCoA	Principle Coordinates Analysis
PCR	Polymerase Chain Reaction
RH	Indicates 49 samples of *Rhinolophus alcyone* and *Hipposideros ruber*
SDS	Sodium dodecyl sulphate (used for storage of faeces samples)

1 INTRODUCTION

1.1 DNA barcoding

1.1.1 What is DNA barcoding?

DNA barcoding consists of sequencing a short region of DNA, called a barcode locus, of a taxonomically identified specimen (Hebert *et al.*, 2003; Liu *et al.*, 2018). Locus sequences are often added to a database (DeSalle *et al.*, 2005; Hollingsworth *et al.,* 2011a). There are public global databases: BOLD (Barcode of Life Database), launched by the Consortium for the Barcode Of Life (CBOL), and Genbank (Ratnasingham and Hebert, 2007; Valentini *et al.*, 2008; Sayers *et al.*, 2019), which scientists all over the planet contribute to. Hence, DNA barcoding is a molecular method of recording the biodiversity of species across the planet.

Locus sequences in the barcode database must be distinct, so they can distinguish between species (Parmentier *et al.*, 2013). Therefore, barcode loci are selected for their high interspecific variation, but low intraspecific variation, so they are variable enough to distinguish between taxa, while being diagnostic of individual species (Luo *et al.*, 2011; Besnard *et al.*, 2014; Li *et al.*, 2015). Barcode loci must also have conserved flanking regions for primers to bind to so the loci can be amplified by PCR (Hollingsworth *et al.*, 2009). This enables loci to be amplified for sequencing, so the locus sequence of each individual species can be recorded.

Originally, the aim of CBOL was to use a single universal locus for barcoding all species in a kingdom (Ratnasingham and Hebert, 2007; CBOL Plant Working Group, 2009; Li *et al.*, 2015). These core loci should be standardised and universally agreed to enable barcode data to be shared (Hebert and Gregory, 2005; Hollingsworth *et al.,* 2011a; Liu *et al.*, 2018). For animals, the cytochrome c oxidase I mitochondrial gene (COI) is the core universal locus, as it is sufficiently variable to distinguish between most, although not all, species (Hebert *et al.*, 2003). For plants, no single barcode exists that can distinguish a large range of taxa (CBOL Plant Working Group, 2009; China Plant BOL Group *et al.*, 2011). Instead, several possible loci have been proposed, including chloroplast loci matK, rbcL and trnH-psbA, and the ribosomal internal transcribed spacer (ITS) (Cowan *et al.*, 2006; Newmaster *et al.*, 2008; CBOL Plant Working Group, 2009; Tan *et al.*, 2018). Two-locus combinations are often used for plants, for sufficient taxonomic coverage and resolution (CBOL Plant Working Group *et al.*, 2009; Li *et al.*, 2015). The CBOL Plant Working Group (2009) proposed a 2-locus barcode of rbcL and matK. rbcL has high universality and high sequence quality, but low discriminatory power, and matK has high discriminatory power but is less universal (CBOL Plant Working Group *et al.*, 2009), 2009). Core locus sequences are added to the database, which can then be used for species identification.

1.1.2 Barcoding for species identification

DNA barcode databases are frequently used for species identification (Hebert *et al.*, 2003; DeSalle *et al.*, 2005; Hebert and Gregory, 2005; Ratnasingham and Hebert, 2007). Barcode loci of unknown species are sequenced, and compared with locus sequences in the database, so they can be identified (Taberlet *et al.*, 2012b; Liu *et al.*, 2018; Tan *et al.*, 2018). This is useful for species that cannot be identified by morphology, or when only small plant fragments are available (Liu *et al.*, 2018). However, barcoding cannot always

identify specimens to species level. Major limitations in using barcodes for species identification are the availability of database sequences and if the barcode locus has sufficient variability between species (Parmentier *et al.*, 2013). The existence, and continued improvement of barcoding databases, as well as advancements in molecular genetic techniques, have led to the development of DNA metabarcoding.

1.2 DNA metabarcoding

1.2.1 What is DNA metabarcoding?

DNA metabarcoding combines traditional DNA barcoding with Next-Generation Sequencing (Taberlet *et al.*, 2012b; Clare *et al.*, 2016; Yang *et al.*, 2016). Next-Generation Sequencing is low cost and enables the sequencing of millions of DNA fragments in a single sequencing run (Shendure and Ji, 2008; Taberlet *et al.*, 2012b). Using Next-Generation sequencing, barcode loci from hundreds of species in a sample can be sequenced at once (Nichols *et al.,* 2016). Using metabarcoding, all species in a sample, such as soil, water or faeces, can be identified by comparison to a barcoding database (Taberlet *et al.*, 2012a; Bohmann *et al.*, 2014). Therefore, metabarcoding can be used to describe the species composition of environmental DNA (Deagle *et al.*, 2014).

Metabarcoding has a number of applications. Andersen *et al.*, (2012) demonstrated that the DNA present in soil accurately reflects the taxonomic richness and relative biomass of the individual species in that area. Therefore, identification of which species the DNA in the soil is from, by metabarcoding, can be used to estimate the biodiversity in an area. Metabarcoding can also be used to detect trophic interactions at the species level (Bohmann *et al.*, 2014; Clare, 2014a). For example, analysis of faeces, or rumen content or stomach content can be used to find out about the diet of a species (Clare *et al.*, 2009; Soininen *et al.*, 2013; Gebremedhin *et al.*, 2016; Yang *et al.*, 2016). Environmental samples are relatively easy to collect and analyse, by repeating this, the biodiversity of an area to be monitored over time (Bohmann *et al.*, 2014). Therefore, metabarcoding enables a deeper understanding of natural processes, such as the structure of ecosystems.

1.2.2 Plant metabarcoding challenges

There are several challenges in plant barcoding, that do not arise with animals. These problems therefore also occur in plant metabarcoding. The main problem in plant barcoding is the lack of a universal locus (CBOL Plant Working Group, 2009; Hollingsworth *et al.*, 2011a). Also, plant species can hybridise, so recognising distinct species is less clear cut. Discriminating between species or hybrids is very difficult or impossible, especially if their barcode locus sequences are very similar (Newmaster *et al.*, 2008; Coissac *et al.*, 2016). Similar species, especially those that have diverged recently, also have low sequence divergence, therefore low interspecific variability, making them hard to distinguish by barcoding (Dong *et al.*, 2012; Tan *et al.*, 2018). Therefore, this results in the species resolution achieved in plant metabarcoding studies being generally lower than that achieved in animal metabarcoding studies.

1.2.3 Basic process of metabarcoding

Metabarcoding has a similar basic process to species identification by traditional barcoding, but occurs on a much larger scale (Murray *et al.*, 2012; Yang *et al.*, 2016). The main steps involved are as follows. DNA is extracted from the environmental sample. Then, a barcode locus and primers are selected, and these primers are used to amplify the locus sequences of all species in the sample by PCR (Hibert *et al.*, 2013). A single pool,

containing locus sequences for all species in the sample that were amplified, is produced and sequenced. Sequencing produces a large dataset, which is analysed by bioinformatics, involving quality control and error removal, clustering into Molecular Operational Taxonomic Units (MOTUs), and finally taxonomic assignment (Kopylova *et al.*, 2016).

1.3 Key considerations in metabarcoding

1.3.1 Locus and primer choice

The initial and crucial decision in a metabarcoding study is the choice of barcoding locus and primers (Alberdi *et al.*, 2018; Omelchenko *et al.*, 2019; Pinol *et al.*, 2019). Loci are often chosen depending on database coverage (Alberdi *et al.*, 2018). It is important to select a locus for which there are database records that overlap with species analysed in a sample (Yang *et al.*, 2016; Alberdi *et al.*, 2018). There will always be taxa that are not in a database, but selecting a core barcode will maximise database coverage (Alberdi *et al.*, 2018).

Some barcoding and metabarcoding studies have compared the effectiveness of different loci, using several determining factors (e.g. CBOL Plant Working Group, 2009; Parmentier *et al.*, 2013; Yang *et al.*, 2016). Loci are selected based on their universality, production of high quality sequences and species discrimination (CBOL Plant Working Group, 2009; Bell *et al.*, 2016), although not all metabarcoding studies need species-level discrimination (Alberdi *et al.*, 2019). A single locus can be amplified by different primer pairs, so the most suitable primers must be selected to amplify a chosen barcode locus (Alberdi *et al.*, 2018).

For a locus to be considered 'universal', primers must be available that are able to bind and amplify the region in a large range of taxa (CBOL Plant Working Group, 2009). In metabarcoding, the composition of species in a sample is usually unknown, so a universal locus is important, to ensure sequences are obtained from all taxa in the sample (Pompanon *et al.*, 2012; Cheng *et al.*, 2016). However, 'universal' primers never perfectly match the targeted region in all species, so they may not be completely universal (Deagle *et al.*, 2014; Alberdi *et al.*, 2018). Certain primers may successfully amplify more the locus sequences of more species than others in a particular study (Cheng *et al.*, 2016). Also, some primers are designed to amplify the DNA of species from particular families. These can be used in addition to a 'universal' locus to increase the resolution of taxonomic assignments for taxa that have low variation in the 'universal' locus region (Hibert *et al.*, 2013; de Barba *et al.*, 2014; Liu *et al.*, 2018). Some loci may have primer binding regions that are divergent between taxa, and hence taxa specific, rather than universal, primers must be used (Cheng *et al.*, 2016; Omelchenko *et al.*, 2019). Therefore, the choice of primers is as important as the choice of locus, and locus choice can be influenced by primer availability.

1.3.2 Primer bias

The variation in the extent of mismatches between primers and primer binding regions between species leads to primer bias (Alberdi *et al.*, 2018). In some cases, the primer binding regions may be highly diverged, preventing primers from binding completely, so these taxa will be missing from the dataset (Omelchenko *et al.*, 2019). Taxa with fewer mismatches in their primer binding regions are preferentially amplified and sequenced over those with fewer mismatches (Lobo *et al.*, 2017). This results in some species in the sample being over-represented (Bellemain *et al.*, 2010; Gebremedhin *et al.*, 2016; Pinol *et*

al., 2019). Therefore, because of the potential of mismatches, using a single primer set is likely to always introduce bias. Using more than one locus can help minimise this bias, as well as increasing taxonomic resolution (Alberdi *et al.*, 2018).

1.3.3 Discrimination

Taxonomic discrimination (i.e. resolution) must also be considered when choosing a suitable locus. Species discrimination is a measure of how well a locus can resolve to species level (Liu *et al.*, 2018). Some loci are only able to resolve taxa to family or genus level, and not species level (Pompanon *et al.*, 2012). Loci with higher interspecific variability have better species discrimination (Luo *et al.*, 2011). The interspecific variability of a locus also varies between taxa. For example, the families Poaceae and Asteraceae have low variability in the trnL locus region (de Barba *et al.*, 2014). Consequently, species discrimination varies both depending on the barcode locus, and on the species analysed in the study (Li *et al.*, 2015; Liu *et al.*, 2018).

An effective barcode locus has broadly universal primers and good taxonomic discrimination. However, there is a trade-off between these factors. Loci with more universal primers tend to be in less variable regions, therefore have lower species discrimination (CBOL Plant Working Group, 2009; Pompanon *et al.*, 2012). Therefore, no locus is completely ideal for a barcoding study, but it is important to find the most effective locus, which may depend on the aim of the study. To minimise the effect of this trade-off, a combination of two loci can be used, to combine universality and taxonomic resolution. For example, matK and rbcL are often used together for standard plant barcoding, and trnL and ITS1 are sometimes used together in diet metabarcoding studies (CBOL Plant Working Group, 2009; de Barba *et al.*, 2014).

1.3.4 Quality filtering

Erroneous sequences (artefacts) are often produced as a result of errors in PCR and sequencing (Brown *et al.*, 2015). For example, chimeras, which are sequences of different species joined together, can be generated during PCR (Smyth *et al.*, 2010). Steps must be taken to remove erroneous sequences, to avoid false sequence diversity in analysis (Kunin *et al.*, 2010; Salinas-Ramos *et al.*, 2015). False sequences are assumed to be rare, so singletons or sequences of low copy number may be filtered out (Murray *et al.*, 2012; Alberdi *et al.*, 2018). Bioinformatics are also used for quality filtering (Nichols *et al.*, 2016). For example, chimeras can be detected using algorithms such as UCHIME (Edgar *et al.*, 2011). The stringency of both quality filtering and MOTU clustering (which occurs after quality filtering) can impact the measurement of diversity in the sample (Flynn *et al.*, 2015; Clare *et al.*, 2016). Quality filtered sequences are then clustered into MOTUs (Brown *et al.*, 2015).

1.3.5 MOTU clustering

One of the main challenges in metabarcoding is MOTU clustering (Clare *et al.*, 2016). Sequences are clustered into MOTUs using a similarity threshold (the 'clustering threshold') that is based on the interspecific divergence of the locus in the taxa in question. The aim of this is for each MOTU to represent an individual species, although this is rarely the case (Brown *et al.*, 2015; Clare *et al.*, 2016; Alberdi *et al.*, 2018). There is variability in the locus sequences of individuals of the same species, and even between loci copies in a single individual (Flynn *et al.*, 2015). Therefore, sequences are clustered

into with an aim of obtaining an comparable estimate for the number of species in the sample (Clare *et al.*, 2016).

There are three methods for MOTU clustering. '*De novo'* is the where sequences are clustered without the use of a reference database and 'closed-reference' involves the use of a database. 'Open-reference' is a combination of the two, where sequences that do not match the reference database are clustered by the '*de novo'* method (Kopylova *et al.*, 2016; Schloss, 2016). In this study, '*de novo'* clustering was used. The advantage of the '*de novo'* method is clustering can be carried out without the need for database coverage. This is useful in metabarcoding studies, where there is a large range of unknown taxa which may not all have matches to the database (Schloss, 2016).

The clustering threshold used has a large impact on the number of MOTUs produced. In metabarcoding, the number of species is unknown, and therefore it is not possible to determine which clustering threshold produces the most accurate number of MOTUs (Brown *et al.*, 2015). A commonly used MOTU clustering threshold is 97%, as the minimum divergence between species for most loci is 3% (Hebert *et al.*, 2003; Vetrovsky and Baldrian, 2013; Brown *et al.*, 2015; Alberdi *et al.*, 2018) However, this does vary in different taxa and loci (Brown *et al.*, 2015; Alberdi *et al.*, 2018). Therefore, finding the appropriate clustering threshold is a challenge, especially as metabarcoding studies involve samples containing a large range of unknown species, which will vary in the genetic divergence of their locus sequences.

A lower threshold gives a lower and more conservative estimate of the number of MOTUS (Salinas-Ramos *et al.*, 2015; Clare *et al.*, 2016). Studies may use this more conservative estimate to reduce the chance of the over-estimation of diversity caused by sequencing and PCR errors (Salinas-Ramos *et al.*, 2015). However, a lower estimate can result in under-splitting, where more than one species is represented by a single MOTU. A higher threshold, i.e. closer to 100%, can result in over-splitting and the production of artificial MOTUs, due to single species being represented by more than one MOTU (Brown *et al.*, 2015; Salinas-Ramos *et al.*, 2015).

Several studies have demonstrated the impact of varying the clustering threshold on the results of a metabarcoding study. Alberdi *et al.*, (2018), in a study of bat diet, showed that increasing the clustering threshold from 94% to 100% tripled the number of species. The filtering method used prior to MOTU clustering also impacts the results of a metabarcoding study. Clare *et al.* (2016), in a dietary overlap study, altered the filtering and clustering parameters, to give 176 different estimates for the number of MOTUs. However, while these estimates gave different measurements of dietary overlap, the general conclusion, of an overlap in the use of food resources by two species, remained the same. Flynn *et al.* (2015) used a mock zooplankton community to assess differences in filtering and clustering methods. The estimate of the number of MOTUs varied from 60, when the most stringent parameters were used, up to 5068, when the most relaxed parameters were used. However, this had a small impact on the number of species generated from the MOTUs, which ranged between 40 and 42.

While finding a single MOTU clustering threshold that produces the correct number of species is difficult or impossible, the studies discussed here have shown that varying MOTU numbers can have a minor effect on the results of an experiment.

An advantage of clustering sequences into MOTUs is that it is possible to roughly estimate the diversity of species in a sample without having a complete reference database (Blaxter *et al.*, 2005). Known and unknown taxa can be analysed together, or MOTUs can be left as unknowns, for statistical analyses to be carried out (Clare *et al.*, 2016). However, this is not particularly useful in that the data is not comparable with other studies (Evans *et al.*, 2016).

1.3.6 Reference database bias

MOTU sequences are matched to reference databases, such as Genbank (Sayers *et al.*, 2019), to identify the species they represent (Bohmann *et al.*, 2011; Mata *et al.*, 2018). A MOTU will not be identified to species level, or may be identified incorrectly, if its corresponding matching sequence is not in the database (Bohmann *et al.*, 2011; Richardson *et al.*, 2015a). Therefore, an incomplete database can limit taxonomic resolution in some species or families (Kesanakurti *et al.*, 2011). Also, only MOTUs that are resolved to a suitable taxonomic level will be used in the analysis of a study. Therefore, the results will be bias towards taxa that have been identified, and MOTUs that are not identified to any taxonomic level will be missing from the results (Bohmann *et al.*, 2011). Consequently, depending on database coverage, a metabarcoding study will not always give an accurate representation of the species composition of an environmental sample.

1.4 Plant metabarcoding applications

Metabarcoding, the identification of species in a sample by DNA sequencing, has proved to be useful in providing more precise details about biological and ecological processes.

Kartzinel *et al.*, (2015) used metabarcoding to analyse the faeces of large mammalian herbivores, to find out about dietary niche partitioning. Previous studies of dietary niche partitioning used stable isotope analysis, and showed how diet varied along a continuum between browsers and grazers, and also between ruminants and non-ruminants (Cerling *et al.*, 2003; Codron *et al.*, 2007). This study used the trnL-P6 locus due to its conserved primer sites and high interspecific variation. Metabarcoding also showed the broad diet differences shown using stable isotope analysis. However, it also provided much more precise information about the diet, showing how diet different between individual species, even those that were grazers of the same size and digestive physiology.

Raye *et al.*, (2011) used the trnL-P6 locus to study the diet of the alpine chamois. Using metabarcoding identified diet items to a high taxonomic resolution. This enabled the diets of individuals to be distinguished, and how they varied throughout the year.

Richardson *et al.*, (2015a) used metabarcoding in a novel way, to identify the pollen collected by bees in an agricultural landscape. The ITS2 locus was used, with the aim of genus level identification. In comparison to the more traditional method of microscopic identification of pollen, using metabarcoding revealed a much greater diversity of plant genera, and was especially useful at detecting pollen that is difficult to identify microscopically. However, metabarcoding identified some plant species that were unlikely to be present in the landscape, and it is possible that this was due to PCR or sequencing errors, or misidentifications due to an incomplete reference database. Despite these limitations, using metabarcoding still gave more precise information about the taxonomic identities of pollen collected by the bees.

Overall, these examples have shown how metabarcoding enables the taxa in environmental samples to be identified to much greater taxonomic resolution than previously, giving more precise details about ecological processes.

1.5 Dietary analysis

As demonstrated by Raye *et al.*, (2011) and Kartzinel *et al.*, (2015), DNA metabarcoding can be used to analyse diet, and provide greater taxonomic detail than other methods of dietary analysis. Understanding the diet of a species has many different uses. These include showing how the species fits into the food web (Valentini *et al.*, 2009), suggesting how resource availability might impact species abundance (Ambrose and DeNiro, 1986) and showing if an animal is an agricultural pest (Korine *et al.*, 1999). As well as DNA metabarcoding, there are a few other methods of diet analysis.

Direct observation is perhaps the most obvious method of dietary analysis, and it can be useful if the subject species and what they are feeding on are both easily identified. Perrin and Brereton-Stiles, (1999) and Kleynhans *et al.*, (2011) both used direct observation to study resource partitioning in large African savannah herbivores. It was an effective method in this case because the animals were feeding in an open space in the day time, so were easily observable. However, direct observation is time and labour intensive, and diet items are not always identifiable (Garnick *et al.*, 2018). Also, for small, secretive or nocturnal animals, direct observation is difficult or impossible, so other dietary analysis methods must be used (Ambrose and DeNiro, 1986; Korine *et al.*, 1999).

The stable isotope ratios of tissues or faeces correspond to diet composition. Carbon isotope ratios can distinguish between browsers and grazers. The ^{13}C /^{12}C ratio is more pronounced in C3, which are browsed plants (trees and shrubs) than C4 plants, which are grazed plants (grass) (Ambrose and DeNiro, 1986; Codron *et al.*, 2007). The ratio of ^{15}N to ^{14}N isotopes can be used to distinguish between herbivores and carnivores, as the concentration of ^{15}N increases up the trophic level (Ambrose and DeNiro, 1986). Stable isotope analysis is very useful for looking at energy flow through and between ecosystems. An advantage over metabarcoding is that stable isotope analysis can give long term diet information, by measuring isotope levels in bone collagen (Pompanon *et al.*, 2012). Metabarcoding can only show what an individual has eaten in recent days or hours. However, the taxonomic resolution achieved using stable isotope is very low (Pompanon *et al.*, 2012).

Microhistological analysis is the identification of food remains in stomach contents or faeces using a microscope (Forsyth and Davis, 2011). A reference collection is used to compare samples to for identification (Forsyth and Davis, 2011; Nichols *et al.,* 2016). Microhistology is very accurate and particularly useful, as it can give a much more reliable quantitative estimation of diet than metabarcoding can (Garnick *et al.,* 2018). However, it is labour intensive and requires taxonomic expertise (Soininen *et al.*, 2009; Nichols *et al.*, 2016). Diet items become unidentifiable with digestion, so bias can be towards more easily identifiable fragments, which are usually those that are less easily digested (Baamrane *et al.*, 2012).

Studies that have compared metabarcoding with other forms of dietary analysis, such as microhistology, have shown that metabarcoding gives much higher taxonomic resolution

and detects a larger diversity of species (Raye *et al.*, 2011; Baamrane *et al.*, 2012; Nichols *et al.*, 2016; Deagle *et al.*, 2019). Diet items that are too degraded for identification by microhistology can still be identified using DNA metabarcoding (Nichols *et al.,* 2016). A main advantage of metabarcoding is that hundreds of samples can be sequenced at once, allowing identification of species in the diets of hundreds of individuals at once (de Barba *et al.*, 2014; Kartzinel *et al.*, 2015; Lopes *et al.*, 2015; Galan *et al.*, 2018). Therefore metabarcoding is a useful form of dietary analysis when the priority is identifying to species in hundreds of samples at once, to a high taxonomic resolution.

However, metabarcoding does have some disadvantages as a method of dietary analysis. General metabarcoding limitations, such as the determining the correct MOTU clustering threshold, can have a large impact on the results (Stein *et al.*, 2014; Alberdi *et al.*, 2018). Also, quantitative estimation of diet is perhaps less reliable than if microhistological analysis was used (Elbrecht and Leese, 2015; Pinol *et al.*, 2019).

1.6 Key considerations in diet metabarcoding

1.6.1 Degradation of samples

The DNA in faeces and stomach contents is degraded, and this can cause problems in metabarcoding for dietary analysis (Deagle *et al.*, 2006; Pompanon *et al.*, 2012; Bohmann *et al.*, 2014). DNA becomes damaged during digestion, which means the majority of DNA fragments in faeces will be of short length (Deagle *et al.*, 2006). Therefore, longer loci have lower amplification success in degraded samples, as longer lengths of DNA are rarer (Pompanon *et al.*, 2012; Deiner *et al.*, 2017). Amplification success is a measure of the number or percentage, of samples that are successfully amplified using a particular set of primers. The fragment length that can be amplified in degraded samples varies between species, but on average, it is suggested that lengths of <250bp are suitable for amplifying degraded DNA (Deagle *et al.*, 2006; Meusnier *et al.*, 2008; Valentini *et al.,* 2008; Pompanon *et al.*, 2012; Nichols *et al.*, 2018). In general, there is an inverse relationship between amplification success and amplicon length (Deiner *et al.*, 2017).

However, longer fragments can be amplified in degraded samples. For example, Erickson *et al.*, (2017) used a 379bp fragment of rbcL to study the diet of white-tailed deer. However, it is likely that using a longer locus results in missing information from the species whose DNA is too degraded to amplify when using this locus (Deagle *et al.*, 2006).

Variation in amplicon length is also a source of bias. Shorter length amplicons are preferentially amplified and sequenced over longer amplicons and therefore over-represented (Wang *et al.*, 2015; Gebremedhin *et al.*, 2016; Deiner *et al.*, 2017).

This suggests that, in general, short loci will give a more accurate idea of the diet of a species. However, shorter loci tend to have lower interspecific variability, most simply because there are less possible variations in base pair sequence. This means shorter loci have lower taxonomic resolution (CBOL Plant Working Group, 2009; Pompanon *et al.*, 2012; Clare *et al.*, 2014; Deiner *et al.*, 2017). The choice of locus is therefore very important in diet metabarcoding. A locus must be short enough to be amplified in degraded samples, but also have high interspecific variation to distinguish between species. This is an additional complication to the trade-off between interspecific variability and universality that occurs in standard barcoding (Li *et al.*, 2015). A

metabarcoding locus needs universal primers, as the species composition of the sample analysed is often unknown and very diverse (Meusnier *et al.*, 2008).

High interspecific variability is not always related to length. Taberlet *et al.*, (2007) compared the taxonomic resolution of the 10-143bp trnL P6 and 91-98 bp rbcL. There is no obvious length variation between these loci, but trnL still had higher taxonomic resolution than rbcL, with 24% of sequences to species level, compared to 16% for rbcL. This suggests that trnL is a more variable locus than rbcL in general.

Traditional barcoding loci such as matK and rbcL are usually around 500bp or longer, therefore are unsuitable for degraded DNA (Valentini *et al.*, 2008; Murray *et al.*, 2012). So, other loci must be selected, or different primers used to amplify shorter fragments. Primer choice influences the amplicon length, and therefore can affect amplification success. So, primers could be selected to amplify a short region, suitable for degraded DNA, of a locus with high interspecific variation. For example, the chloroplast locus trnL has high interspecific variation, and primers g and h can be used to amplify a short region of it, the 10-143bp P6 loop (Taberlet *et al.*, 2007; Murray *et al.*, 2012). Murray *et al.*, (2012), in a study of ancient faeces DNA, successfully amplified this P6 loop, but had no success in amplifying the full length trnL locus.

Both locus length and position are important for species discrimination (Hajibabaei *et al.*, 2006). Although short loci have limited taxonomic resolution (Zeale *et al.*, 2010), some can still discriminate to species level to a certain extent (Taberlet *et al.*, 2007; Valentini *et al.*, 2009; Kartzinel *et al.*, 2015). For example, using trnL P6, Gebremedhin *et al.*, (2016) resolved 29.8% of sequences to species level, and Pegard *et al.*, (2009) resolved 24.4%. Also, although longer loci have low amplification success, they can still be used. They are unlikely to amplify in all samples, so short core locus with universal primers may be used in addition, to ensure all species are amplified (Meusnier *et al.*, 2008; Pompanon *et al.*, 2012). The locus ITS1, which can be up to 500bp (Wang *et al.*, 2015), has been used in addition to a core barcode, to increase taxonomic resolution. For example, Hibert *et al.*, (2013), in a study of tapir diet, designed a family specific ITS1 reverse primer for the Sapotaceae family, as it is a common family in the study site and the tapir diet. Baamrane *et al.*, (2012) used ITS1 specifically to increase taxonomic resolution in the Poaceae and Asteraceae families, as these have low sequence variation in the core barcode trnL. Therefore, there are many factors to consider when selecting loci for diet metabarcoding, and often no locus is ideal, due to trade-offs between amplification success and taxonomic resolution.

1.6.2 Quantitative diet estimates

When using metabarcoding for dietary analysis, ideally, the number of sequences corresponding to each diet item should reflect the actual abundance of that diet item that was consumed (Bohmann *et al.*, 2014; Pinol *et al.*, 2019). This relationship must remain after PCR amplification and through other steps in the metabarcoding process (Pompanon *et al.*, 2012), but there are several factors that prevent this.

The extent of mismatches in the primer binding regions of loci varies between species (Elbrecht and Leese, 2015). Some species are preferentially amplified over others, due to having less mismatches in their primer binding regions (Pinol *et al.,* 2019). There will be a higher number of sequences present for these preferentially amplified species, resulting in

them appearing to contribute to a higher proportion of the diet than is true (Pompanon *et al.*, 2012; Nichols *et al.*, 2016; Pinol *et al.*, 2019).

Pinol *et al.,* (2019) showed that in some primer pairs, there is a significant linear relationship between the abundance of species before and after PCR. However, this correlation is only reliable when the exact species composition is known. Therefore, metabarcoding may be quantitative depending on the primer pair used. A way of minimising primer bias is to select primers for which there is low variance in the number of mismatches between different species (Pinol *et al.*, 2019).

Length related bias also affects the ability to make a quantitative estimate of diet (Wang *et al.*, 2015; Cheng *et al.*, 2016), particularly if the length of a locus varies between species. Shorter loci are preferentially amplified over longer loci, so there will be more sequences amplified for species of shorter loci, resulting in them appearing more abundant than they actually are (Raye *et al.*, 2011; Pompanon *et al.*, 2012) Selecting primers with minimal length variation may help prevent this amplicon length bias (Pompanon *et al.*, 2012).

Variation in primer mismatches and locus length mean some sequences are preferentially amplified by PCR over others. Therefore, several studies have proposed avoiding PCR to avoid this amplification bias (Zhou *et al.*, 2013; Elbrecht and Leese, 2015; Pinol *et al.,* 2019). Instead, samples could be directly sequenced, or whole mitochondrial genome sequencing could be used (Bista *et al.*, 2018). However, this is not possible in diet metabarcoding, because samples are degraded, so long fragments cannot be sequenced, and often DNA is in low concentration and therefore needs to be amplified by PCR before it can be sequenced (Deagle *et al.*, 2006; Pinol *et al.,* 2019). There are also other sources of error in addition to amplification bias.

Metabarcoding of faeces adds the additional problem of variability in digestion rates, that prevents metabarcoding from being quantitative. If one diet item is less digested than another, it will be more intact in faeces, more easily detected, and therefore over-represented in the results (Deagle and Tollit, 2007; Pompanon *et al.*, 2012). Presumably, larger and harder diet items will remain more intact after digestion (Clare *et al.*, 2009), which suggests their DNA will be less degraded, easier to amplify, resulting in them being over represented.

Occasionally, quantitative data is used in metabarcoding, and has produced what appear to be reliable results. For example, Nichols *et al.,* (2016) found a significant correlation between the log proportion of DNA reads and the log proportion of macroscopic weight, suggesting metabarcoding could be quantitative. However, in most cases, it is suggested that metabarcoding cannot be used for quantitative assessment of species.

It is possible to estimate diet in a semi-quantitative way when metabarcoding the diet of a large number of individuals of the same species, by using frequency of occurrence (Raye *et al.*, 2011). Detecting a particular diet item in a large proportion of the individual samples would suggest it forms a large proportion of the diet of that particular species. This method uses presence-absence data. In most situations, it is preferable to use presence-absence data, as it is more reliable than sequence abundance data (Elbrecht and Leese, 2015).

1.7 Dietary analysis to construct food webs

Understanding how species interact is vital to find out about the structure and function of ecosystems (Evans *et al.*, 2016). Knowing about the diet of a species shows where it fits into the food web, which in turn may enable understanding of its ecological function, such as if it provides an ecosystem service such as pest control (Korine *et al.*, 1999; Pompanon *et al.*, 2012; Jedlicka *et al.*, 2017). Using molecular methods, such as DNA metabarcoding, to analyse species interactions, is easier and more precise than direct observation or morphological identification, particularly for interactions that are difficult to observe (Pompanon *et al.*, 2012; Evans *et al.*, 2016). In metabarcoding, locus sequences are pooled for a single sequencing run. However, loci can be tagged with unique sequences, so specific interaction data can be gathered, such as distinguishing the diet of different individuals (Evans *et al.*, 2016). Knowing the ecological functions of species and how they interact also means it is possible to identify keystone species, which interact with and support many other species, and are therefore vital for conservation (Jordan, 2009).

1.8 Ecosystem services

Maintaining biodiversity is important because it is the source of ecosystem services, such as pollination or pest control, which provide huge benefits to agricultural production (Fischer *et al.*, 2008, 2014; Cardinale *et al.*, 2012; Naeem *et al.*, 2012). Tropical forests are particularly biodiverse (Harvey *et al.*, 2006; Gardner *et al.*, 2009). However, the clearing of tropical forests due to conversion to agricultural land results in a loss of biodiversity and the ecosystem services it provides (Foley *et al.*, 2005; De Beenhouwer *et al.*, 2013). Therefore, there is a need to balance biodiversity and agricultural production, to maintain ecosystem services and conserve species (Fischer *et al.*, 2008).

Tropical crops, such as cacao, can be grown amongst a diversity of native trees, in a system of agroforestry (Sonwa *et al.*, 2007; Mbolo *et al.*, 2016). This tree diversity supports a wealth of animal diversity (Harvey *et al.*, 2006; De Beenhouwer *et al.*, 2013). For example, Tallamy and Shropshire, (2009) showed that native trees supported around four times the number of Lepidoptera species than introduced (non-native) plants. Using agroforests means biodiversity, and therefore the ecosystem services it provides, are retained (Steffan-Dewenter *et al.*, 2007; Sonwa *et al.*, 2017). Therefore, if managed to support biodiversity, agroforestry can be a sustainable method of agriculture.

High tree diversity supports bird and bat diversity (Waltert *et al.*, 2005; Maas *et al.*, 2015), and in turn birds and bats can benefit agriculture. For example, Maas *et al.*, (2013) excluded birds and bats from a patch of cacao agroforest in Indonesia, and demonstrated that pest control by birds and bats increased cacao yield by 31% and saved 730 USD per year. Van Bael *et al.,* (2007) showed that birds reduced arthropod abundance, decreasing plant damage in agroforests. In Cameroon Agroforests, shade tree cover can be useful in providing habitat for ants, and a decline in shade cover resulted in increased herbivory of cacao trees (Bisseleua *et al.*, 2009). Maintaining tree diversity supports the diversity of animals higher up the food chain. These animals can benefit agriculture by providing ecosystem services.

1.9 Study site and study species

1.9.1 Cacao agroforest

In southern Cameroon, cacao is a major source of income for many farmers (Gockowski and Dury, 1999; Jagoret *et al.*, 2011). It is traditionally grown in an agroforest system, which can be more sustainable than cacao monoculture practices in other countries, such as Côte d'Ivoire, where management is far more industrial and cacao trees are often grown under few or no shade trees (Gockowski and Dury, 1999; Sonwa *et al.*, 2007; Jagoret *et al.,* 2011). Some trees are typically cleared from Cameroonian agroforests to plant cacao, but a large variety, with different benefits, are retained. These include those that provide shade, and keep the soil cool and moist (Mbolo *et al.*, 2016). Shade trees also benefit cacao trees by protecting them from precipitation, supplying organic matter to the soil in the form of leaf litter, and also have been shown to promote production in older cacao trees (Ahenkorah *et al.*, 1974; Rice and Greenberg, 2000). Fruit and timber trees are also planted, providing some additional income to the farmers (Zapfack *et al.*, 2002; Sonwa *et al.*, 2007; Bisseleua *et al.*, 2009; Jagoret *et al.,* 2011; Mbolo *et al.*, 2016). These Cameroon agroforests have a complex and diverse structure, providing a habitat for wildlife and maintaining biodiversity, and therefore ecosystem services (Rice and Greenberg, 2000; Bisseleua *et al.*, 2009; De Beenhouwer *et al.,* 2013; Sonwa *et al.*, 2017). Therefore, although agroforestry may provide a lower yield than monoculture (Bisseleua *et al.*, 2009; Jagoret *et al.*, 2011), it does provide many other advantages which may offset the lower yield.

1.9.2 Bats

Common bat species in the agroforests include insectivorous species *Hipposideros ruber* and *Rhinolophus alcyone*, and frugivorous species *Epomops franqueti, Micropteropus pusillus* and *Rousettus aegyptiacus.* These bats roost in caves, buildings or trees (Decher and Fahr, 2005; Nkrumah *et al.*, 2016). For example, *H. ruber* mostly roosts in caves, but occasionally will roost in trees (Nkrumah *et al.*, 2016) and *R. aegyptiacus* roosts in large groups in caves or artificial structures such as mines or bunkers (Kwiecinski and Griffiths, 1999). Insectivorous bats select habitats depending on the availability of insect prey, and this in turn depends on the availability of plants that provide habitat and food for the insect prey (Nkrumah *et al.*, 2016).

1.9.3 Shade trees

Trees with useful products are common in the agroforests, including timber trees such as *Terminalia superba* and *Albizia adianthifolia* and fruit trees such as *Mangifera indica* (Mango) and *Dacryodes edulis* (African Plum). Tree diversity is high in cacao agroforests, but often, native forest trees are being replaced by exotic crop species (Sonwa *et al.*, 2007).

1.10 Rationale for this study

The goal of the wider project is to identify which shade tree species support high levels of biodiversity, and to find out if this biodiversity provides ecosystem services that benefit the cacao production. Analysing the faeces of frugivorous and insectivorous bats will show which shade tree species are required to support these bats.

Metabarcoding may provide direct and indirect formation about diet. Metabarcoding the faeces of frugivorous bats using plant primers will identify their diet directly. This is useful to show which tree species are important to support a diversity of frugivorous bat species. A possible theory is that metabarcoding the faeces of insectivorous bats using plant primers will indirectly identify plant species that were eaten by the bats' insect prey. However, it is not yet known whether it is possible to sequence DNA that has passed through both the insect and bat digestive systems. There could be other reasons for plant presence in insectivorous bat faeces. For example, some insectivorous bats roost in trees (Decher and Fahr, 2005; Nkrumah *et al.*, 2016), so may have unintentionally ingested plant fragments. However, if it is possible to detect plant species eaten by the insects, then detecting cacao sequences would potentially show if the bats are feeding on pests of the cacao.

To analyse bat diet, direct observation or microhistological analysis are not possible (Zeale *et al.*, 2010). Insectivorous bats feed aerially and in the dark, and chew up or avoid larger prey fragments, so they cannot be identified morphologically in faeces (Clare *et al.*, 2009; Zeale *et al.*, 2010; Bohmann *et al.*, 2014). Also, any plant material in the faeces of insectivorous bats is likely to be minimal and very degraded. In this study, high taxonomic resolution is required, in order to identify important tree species. Therefore, DNA metabarcoding is the only realistic method of dietary analysis in this study.

In a preliminary study, the rbcL locus was used to analyse the faeces samples of bats in the Cameroon cacao agroforest (A.J. Welch, unpublished). However, this locus gave very low taxonomic resolution, with just 2% of MOTUs resolved to species level. This was too low to gain useful information about important tree species supporting bat biodiversity. The aim of this study was to design and test the effectiveness of new metabarcoding loci, to improve taxonomic resolution. The goal was to design loci that could achieve species level resolution, although, as discussed, this is not always possible in diet metabarcoding studies.

2 MATERIALS AND METHODS

2.1 Study sites

Cameroon has a tropical humid climate, with the main dry season between December and February, a mini dry season in July and August, and wet seasons for the remainder of the year. Bat faeces samples were collected in sixteen cacao plantations in Southern Cameroon, consisting of eight traditional agroforests, six fruit tree and two full sun plantations. Data collection occurred during two field seasons, in August-September 2017 and in January-February 2018.

2.2 Sample Collection

Insectivorous and frugivorous bats were caught by passive mist netting using 20m by 20m nets, between 6pm and 12am. Individuals were identified, and their weight, sex and size were measured. They were then then released after they had produced faeces. Faeces were stored in Longmire buffer with 2% SDS (Sodium dodecyl sulphate). In this study, the diets of five insectivorous species and five frugivorous species were analysed (Appendix A). The most common were two insectivorous species, *Hipposideros ruber* and *Rhinolophus alcyone*, which were collected on almost every farm. Bark and leaf samples were collected from trees on farms in the Ayos landscape, and identified by botanist Hermann Taedoumg of Bioversity International. Plant samples were stored in silica.

2.3 Locus and Primer Selection

2.3.1 Literature review to identify candidate metabarcoding loci

A literature review of barcoding and metabarcoding studies was conducted to determine which barcode genes and primers would be suitable for this study. Google Scholar and Web of Science were used, with key search terms including: ‘Plant barcoding’, ‘faeces metabarcoding’ and ‘diet metabarcoding’. Some papers compared the effectiveness, universality and taxonomic resolution of barcode loci and primers, and others were empirical studies of dietary analysis by metabarcoding faeces or rumen content. For each barcode locus, the amplicon length, primers used, their universality, and the taxonomic resolution achieved was recorded (Appendix B). Candidate loci were chosen based on taxonomic resolution to species level, if they were effective on degraded DNA samples, and if there were universal primers available. Selected loci and primers are recorded in Table 1.

Table 1: Barcode loci and primers chosen from the literature review

Locus	Primer	Direction	Primer Sequence (5'-3')	Reference
trnL	trnL-c	F	CGAAATCGGTAGACGCTACG	1
trnL	trnL-d	R	GGGGATAGAGGGACTTGAAC	1
trnL	trnL-g	F	GGGCAATCCTGAGCCAA	1
trnL	trnL-h	R	CCATTGAGTCTCTGCACCTATC	1
trnH	psbA3f	F	GTTATGCATGAACGTAATGCTC	2
trnH	trnHf	R	CGCGCATGGTGGAATTCACAATCC	3
ITS1	ITS1	F	AGAAGTCGTAACAAGGTTTCCGTAGG	4
ITS1	5.8S1	R	AGAGCCDAGATATCCRTTGC	5
ITS1	5.8S2	R	ACGGGATTCTGCAATTCACAC	5
ITS1	ITS2	R	GCTRCGTTCTTCATCGATGC	6

Note: R: A/G, D: A/G/T
References: [1] Taberlet *et al.*, 2007; [2] Sang *et al.*, 1997; [3] Tate and Simpson, 2003 [4] China Plant BOL Group *et al.*, 2011 [5] This study [6] Stanford *et al.*, 2000

2.3.2 Identification of common shade tree families

A list of the seven most common tree families found in Cameroonian cacao forests was compiled, using information from several sources (Table 2). These included species identification carried out by botanist Hermann Taedoumg at the EBA001 site in the Ayos landscape and the top 18 shade tree species list (Sonwa *et al.*, 2017).

Table 2: The seven top plant families found in the Cameroon cacao agroforests

Anacardiaceae
Apocynaceae
Combretaceae
Moraceae
Phyllanthaceae
Fabaceae
Malvaceae

Locus sequences of species in the top seven families were aligned with the primers, to calculate the percentage divergence between species. Locus sequences of cacao (*Theobroma cacao*) were aligned with the primers to estimate locus lengths.

2.3.3 Designing primers for trnH-psbA and ITS

The existing primer pair for the trnH-psbA spacer (psbA3f/trnHf) amplifies a long region of 281-660bp (Sang *et al.*, 1997; Tate and Simpson, 2003; Kress *et al.*, 2005), and therefore produces fragments that are too long for degraded faecal samples. The reverse primers used to amplify ITS regions are often family specific (e.g. Baamrane *et al.*, 2012; Hibert *et al.*, 2013; de Barba *et al.*, 2014; Lopes *et al.*, 2015). Therefore, a reverse primer was designed specifically for the common families in the study site.

To design primers, barcode sequences of species belonging to seven common families were collected from the Genbank nucleotide database (Sayers *et al.*, 2019) and aligned to find conserved regions that would be suitable for primer design (Table 3). For each locus, the sequences were aligned in MAAFT online version 7 (Katoh *et al.,* 2017) and visualised in Sequencher version 5.4.6 (2017). However, trnH-psbA does not contain an internally conserved sequence suitable for primer design, so the original existing primers (trnHf and psbA3f) were tested. For the ITS segment, Primer3 4.1.0 (Untergasser *et al.*, 2012) was used to design reverse primers to bind to the 5.8S segment of the locus, to be used with the forward primer ITS1 (Figure 1) (China Plant BOL Group *et al.*, 2011).

Table 3: The species whose barcode locus sequences were aligned in primer design, along with their corresponding families, the size of the barcode locus and Genbank accession number (unique reference to Genbank record).

Family	Species name	Size of locus (bp)	Genbank Accession number
Anacardiaceae	*Mangifera indica*	1099	KJ833766
Anacardiaceae	*Rhus chinensis*	744	AB539919
Apocynaceae	*Sisyranthus compactus*	670	AJ310795
Combretaceae	*Conocarpus sericeus*	837	FJ381784
Combretaceae	*Terminalia arjuna*	695	FJ381783
Moraceae	*Morus bombycis*	691	AM042006
Moraceae	*Ficus boninsimae*	673	AB485917
Phyllanthaceae	*Phyllanthus emblica*	651	LC089029
Fabaceae	*Oxytropis pilosa*	676	AM401574
Malvaceae	*Napaea dioca*	682	AJ304940

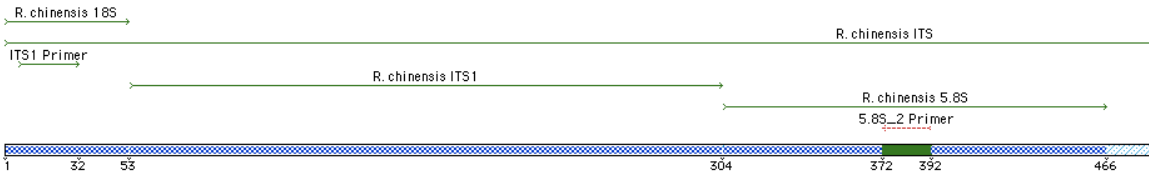

Figure 1: Part of the internal transcribed spacer region of *Rhus chinensis* as an example species, showing positions of the variable region ITS1 and conserved regions 18S and 5.8S. The positions where the chosen primers bind: ITS1 and 5.8S2 are also shown.

2.3.4 Assessment of reference taxonomic database

A search of the Genbank nucleotide database was carried out to estimate how many Cameroonian plant taxa it contains with records for the selected barcode regions. The search box entry "Locus name/country=Cameroon" was used, for the following loci names:

trnH-psbA, psbA-trnH, psbA; tRNA-Leu (trnL), trnL, trnL-trnF; 18S ITS1 5.8S, 18S ITS1

Data for neighbouring countries was also searched, but there were very few records. This search revealed around 200-300 records for each locus, which was sufficient for use in this study. Therefore, these barcode loci were selected to test (Table 1).

2.4 Laboratory work

2.4.1 Primer optimisation

Plant DNA extraction was carried out using the QIAGEN DNeasy Plant Mini kit, with a QIAGEN Tissuelyser II used for tissue disruption. Some modifications were made to the protocol involving the sample homogenisation steps (Appendix C).

The success of the DNA extractions was then tested by PCR amplification using the QIAGEN Multiplex PCR Kit. The PCR recipe contained of 5.3µL of water, 0.6µL at 10µM of each primer, 7.5µL multiplex and 1µL DNA. Universal rbcL primers that had previously been shown to work on Cameroonian plant DNA were used in multiplex: rbcLxpine 3/4, 5/6 and 7/8 (Table 4) (Jarrett *et al.*, in prep). PCR conditions for rbcL primers were an annealing temperature of 50°C and 30 seconds 72°C extension. The PCR products were visualised on a 2% agarose gel alongside a low molecular weight DNA ladder to determine success.

Table 4: rbcL loci and primer sequences used to test plant DNA extraction

Primer	Direction	Primer sequence (5'-3')
rbcLxpine3	F	CTAAATTGGGATTATCCGCT
rbcLxpine4	R	AGCGGTCTCTCCAACGCATA
rbcLxpine5	F	TTTCACTCAAGATTGGGTTTCT
rbcLxpine6	R	ATTTCCCCAAGGGTGTCCTA
rbcLxpine7	F	CTCCTGAATATGAAACCAAAGA
rbcLxpine8	R	GTAGCAGCGCCCTTTGTAAC

To test the effectiveness of the primers, PCR was carried out using extracted plant DNA. The annealing temperature and length of the 72°C extension were varied to find the optimal PCR conditions for each primer pair. Once they were shown to be effective, the primers were tested on DNA from faeces (see next section). Further optimisation of the PCR conditions was carried out, and optimal conditions for each primer pair were recorded (Table 5). After carrying out PCR, gel electrophoresis was used to visualise the success. PCR products were run on a 2% agarose gel alongside a low molecular weight DNA ladder. After testing all primer pairs on faeces DNA, it was concluded that trnL c/h and ITS1/5.8S2 had the highest amplification success rates. Therefore, these primer pairs were used with all faecal samples.

Table 5: Optimal conditions for each primer pair, found by PCR optimisation using both plant DNA and DNA extracted from faeces samples.

Primer Pair	Annealing temperature (°C)	Length of 72°C extension (Seconds)
trnL c/d	49	60
trnL g/h	55	30
trnL c/h	52	30
psbA3f/trnHf	60	60
ITS1/5.8S_1	50	30
ITS1/5.8S_2	49	30
ITS1/ITS2	50	30

2.4.2 Metabarcoding

Bat faecal DNA extraction was carried out using the QIAamp Fast DNA Stool Mini Kit. A modified version of the protocol was followed, involving more vigorous cell disruption and a longer lysis time (Appendix D).

Of the faeces samples that had amplified with PCR successfully, those with the brightest bands were prepared for sequencing. This consisted of 99 individual bat samples, 91 of which had trnL sequences and 83 of which had ITS1 sequences (Appendix A).

Triplicate PCRs (as described by Alberdi *et al.*, 2018) were carried out on each sample with trnL c/h and ITS1/5.8S primers, using the optimal PCR conditions found for each primer. The PCR recipe was contained 6.1µL of water, 0.2µL at 10µM of each primer, 7.5µL Multiplex and 2µL DNA. The PCR primers used for metabarcoding contained part of the illumina adapter as an overhang on their 3' end. The illumina overhang sequences are as follows:

Forward primers trnL-c and ITS1:
`TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[primer sequence]`

Reverse primers trnL-h and 5.8S2:
`GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[primer sequence]`

PCR success was visualised using a 2% agarose gel. Successful triplicate PCR products were then pooled to make one 15ul PCR product per sample.

DNA samples were cleaned using Sera-mag speed beads (Thermo Fisher Scientific), which were made into a solution using the protocol of Rohland and Reich, (2012) with 0.6x beads. The DNA concentration for each locus for each sample was measured using Qubit dsDNA High Sensitivity Kit (Thermo Fisher Scientific) and then the PCR products for each sample were pooled in equal concentrations.

Indexing PCR was carried out to extend the Illumina sequencing adapters to full length and add an individual specific pair of indices. PCR was carried out using KAPA Hifi Hotstart readymix PCR kit (Roche), with an initial cycle of 95°C for 3mins, then 98°C for 30secs, followed by 12 cycles of 98°C for 10secs, 63°C for 30secs, 72°C for 3 mins. The indexing PCR recipe contained 10μL water, 25μL Kapa Hifi Mastermix, 2.5μL of each primer and 10μL DNA. The indexed samples were cleaned using Sera-mag speed beads as before, and their concentration was measured using qubit. 2ul of each sample was run on a 2% agarose gel with a low Molecular Weight DNA ladder.

Libraries were validated on the Agilent 2200 TapeStation system, using High Sensitivity D1000 ScreenTape, for a subsample of eleven individuals, to estimate the average DNA fragment sizes. The qubit concentration and estimated fragment sizes were then used to calculate the nM concentration of each sample, and these values were used to dilute the samples so they could be pooled together in equal nM concentrations. The final sequencing pool was validated on the TapeStation, as above. The concentration of the pool was determined via qPCR using Kapa Universal Library Quantification Kit, and then the pool was sequenced on an Illumina MiSeq, with a 250bp paired end lane.

2.5 Bioinformatics

First, quality control of the data was carried out. Unless otherwise noted, default settings were used. CutAdapt was used to trim off any remaining Illumina adapter sequences. Reads were trimmed according to quality using Sickle (Joshi and Fass, 2011) and those below a length threshold (of 30bp) were removed. BayesHammer (Nikolenko *et al.*, 2013) was used for further error correction and the removal of highly erroneous sequences. Forward and reverse reads were merged using PEAR stand-alone (Zhang *et al.*, 2014). A custom python script, SortByLocus.py was used to split sequences for each sample according to locus, based on the first 40bp of the merged reads, which include the primer sequences. CutAdapt was used to trim off PCR primer sequences. Chimera detection was performed using DAMe in conjunction with usearch (Edgar *et al.*, 2011; Zepeda-Mendoza *et al.*, 2016). MOTUs were clustered using a de novo method with Sumaclust (Mercier *et al.*, 2013). Several possible clustering thresholds were tested to find the most appropriate (see results).

Following Alberdi *et al.*, (2018), taxonomy was assigned via a BLAST search of the Genbank NT database (Sayers *et al.*, 2019). The top 20 Genbank matches were retrieved and those with the highest bitscore were retained. The consensus taxonomy that agreed among these hits was determined, and taxonomy was assigned to the MOTUs based on percent identity: For matches ≥95%, order-level taxonomy was assigned, for ≥96.5%,

family-level and for ≥98%, genus and species level. MOTUs with less than five sequences were discarded.

2.6 Data analysis

Sequence data from *Rhinolophus alcyone* (27 samples) and *Hipposideros ruber* (22 samples) (insectivorous bats) from a previous experiment (A.J. Welch, unpublished) using the rbcL locus with rbcLxpine3 and rbcLxpine4 primers was also analysed. The 99 samples in this sequencing run included the same 49 *R. alcyone* and *H. ruber* samples. Each of the 49 samples contained sequences for at least two of the three loci. The remaining 50 samples in this run corresponded to bats of other species. Total numbers of bat species for each locus are listed in Appendix A. Analysis was carried out for the 49 *R. alcyone* and *H. ruber* bat sample using all three loci, and for the total 99 bat samples using just trnL and ITS1.

Only MOTUs that resolved to the phylum Streptophyta were used, to ensure that only plant sequences were included. Streptophyta includes embryophytes (land plants) and charophytes (some algae) (Bremer, 1985; Adl *et al.*, 2019). MOTUs that had no match in Genbank or were unique to negative control samples were also discarded from further analysis. The number of sequences and non-unique OTUs in the negative control sample were recorded.

Unless otherwise noted, taxonomic resolution was calculated as the percentage of MOTUs that were assigned species level. The frequency of occurrence (FOO) was calculated for each family, as the number of occurrences of a MOTU of that family in the diet.

Principle Co-ordinates analyses (PCoAs), and bar plots were constructed using Qiime v1.9 (Caporaso *et al.*, 2010). Samples were rarefied randomly to include 3000 sequences from each bat sample. This was done to ensure the largest number of samples could be included in the analysis. PCoAs were used to compare *R. alcyone* and *H. ruber* samples between each of the three loci, and insectivores and frugivores using trnL and ITS1. Alpha diversities of insectivorous and frugivorous bats were also calculated using Qiime, and a t-test was carried out to compare mean alpha diversities between insectivores and frugivores.

Statistical analyses were carried out using RStudio (RStudio Team, 2016) and the package dplyr. The number of MOTUs in present in the faeces of insectivorous and frugivorous bats, for both trnL and ITS1 sequence data, was analysed using a shapiro-wilk normality test (shapiro.test), followed by a Mann-Whitney U test (wilcox.test) to compare the number of MOTUs between insectivores and frugivores. The number of cacao sequences present in the faeces of insectivorous and frugivorous bats, for trnL sequence data only, was analysed using the same statistical tests. The package ggpubr was used to construct box plots (ggboxplot) to visualise the Mann-Whitney U test results.

3 RESULTS

3.1 Metabarcoding loci and amplification success

Chloroplast locus trnL, with c and h primers, and nuclear ribosomal locus ITS1, with ITS1 and 5.8S2 were selected to use in metabarcoding (Table 1). C and h primers had amplification success equal to g and h primers. ITS1 and 5.8S2 had the highest amplification success of all the ITS1 primers tested, when used to amplify the loci of shade tree samples. trnH-psbA had very low amplification success, so was discarded from further analysis. The chosen loci successfully amplified the majority of the shade tree samples they were tested on (Table 6).

Table 6: The 18 shade tree samples that were used to test the loci on, and whether they were successfully amplified

			Successfully amplified (Y/N)	
Tree number	**Family**	**Species**	**trnL ch**	**ITS1 and 5.8S2**
EBA001_4	Anacardiaceae	*Mangifera indica*	Y	Y
EBA001_5	Rutaceae	*Citrus sinensis*	Y	Y
EBA001_6	Rutaceae	*Citrus reticulata*	Y	Y
EBA001_7	Unidentified*	Unidentified	Y	Y
EBA001_8	Burseraceae	*Dacryodes edulis*	Y	Y
EBA001_9	Anacardiaceae	*Mangifera indica*	Y	Y
EBA001_10	Rutaceae	*Citrus reticulata*	Y	Y
EBA001_11	Fabaceae	*Albizia adianthifolia*	Y	Y
EBA001_12	Malvaceae	*Cola accuminata*	Y	N
EBA001_13	Fabaceae	*Albizia ferruginea*	Y	Y
EBA001_14	Burseraceae	*Dacryodes edulis*	Y	Y
EBA001_15	Moraceae	*Ficus exasperata*	Y	Y
EBA001_16	Burseraceae	*Dacryodes edulis*	Y	Y
EBA001_17	Unidentified*	Unidentified	N	N
EBA001_18	Moraceae	*Ficus exasperata*	Y	Y
EBA001_22	Malvaceae	*Sterculia tragacantha*	Y	N
EBA001_23	Anacardiaceae	*Mangifera indica*	Y	Y
EBA001_Tc	Malvaceae	*Theobroma cacao*	Y	Y

*Note: Some trees were identified but are still known to be common shade tree species in the site

Amplicon sizes of the loci are as follows: rbcL is 89bp, trnL is 169bp and ITS1 is 395bp. Percentage sequence divergences are 7.8% for rbcL, 21.8% for trnL and 43.5% for ITS1.

For trnL, DNA was extracted and loci amplified successfully from 100 out of 106 samples. For ITS1, amplification was attempted from more samples due to a lower amplification success of this locus. DNA was extracted and loci amplified successfully from 85 out of 115 samples.

Amplification success was calculated based on the 106 samples from which amplification had been attempted using both loci. Of these, 100 were successfully amplified using trnL (94% amplification success) and 83 were successfully amplified using ITS1 (78% amplification success). 99 samples were sent for sequencing, 91 of which had trnL sequences and 83 of which had ITS1 sequences (Appendix A).

3.2 DNA Sequencing and taxonomic assignment

Sequence data was obtained for all samples. The sequencing run returned a total of 13.1 million sequences, with an average of 57774 sequences per sample. 6.3% of the total sequences, including those corresponding to the CO1 loci included in the sequencing run, were removed during quality control steps. At the read merging stage, 0.32% of trnL reads and 13.0% of ITS1 reads failed to merge and were discarded. The remaining sequences were clustered into MOTUs. MOTU clustering thresholds were 96% for rbcL, 97% for trnL and 98% for ITS1. These were selected based on graphs which looked the effect of clustering threshold on the number of MOTUS generated (Figure 2).

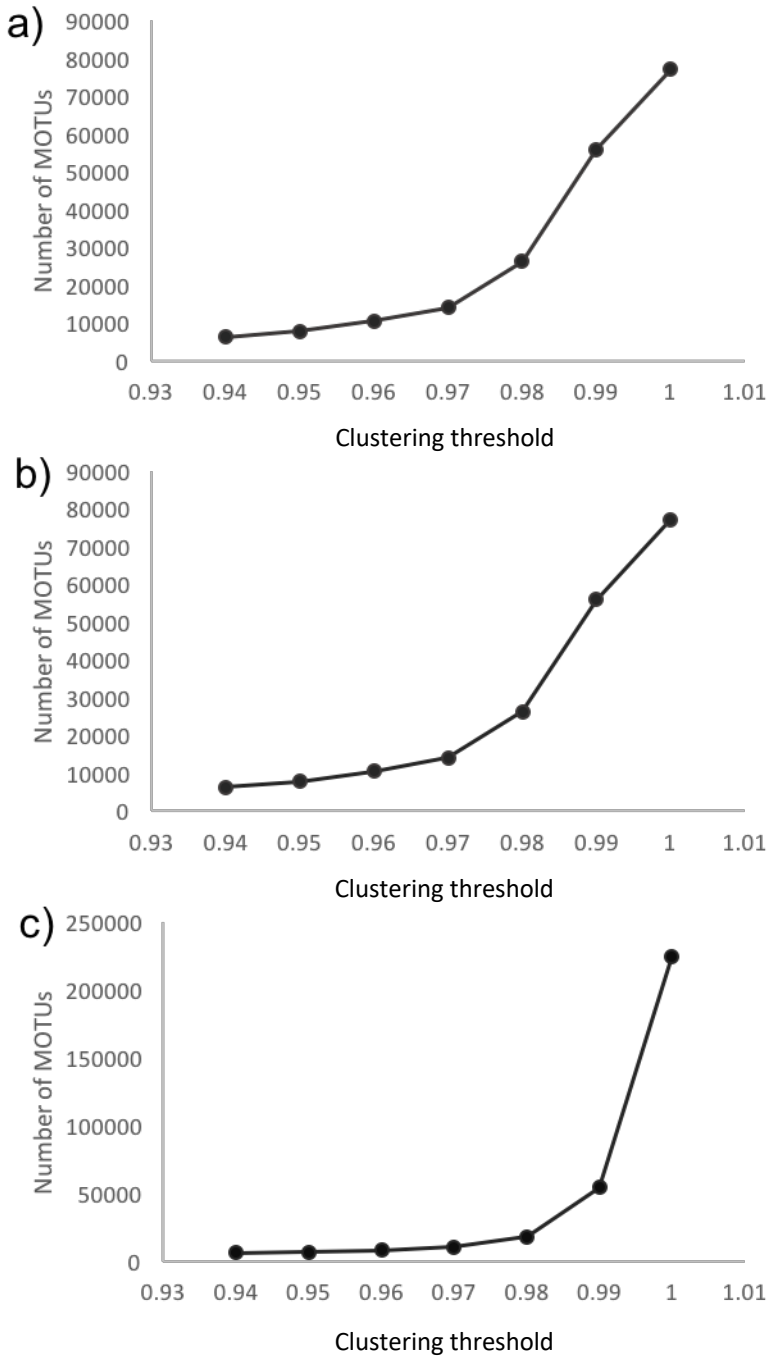

Figure 2: Graphs showing how MOTU clustering threshold determines the number of MOTUs produced. These graphs were used to determine the most appropriate MOTU clustering threshold, for a) rbcL, b) trnL, c) ITS1

Table 7 shows the number of sequences and MOTUs corresponding to each locus. Sequences and MOTUs are recorded both for all samples together, and for the 49 *R. alcyone* and *H. ruber* samples alone.

Table 7: Summary of numbers of sequences and MOTUs used in analyses, and the number of MOTUs that corresponded to non-plant taxa (these MOTUs were discarded before analysis). 'All samples' are the 99 samples used in this study (Appendix A), and the *R. alcyone* and *H. ruber* samples are 49 of these samples that were also analysed in the previous study using rbcL.

	Locus		
	rbcL	**trnL**	**ITS1**
Sequences in all samples	NA	1753826	736975
R. alcyone and *H. ruber* sequences	825940	972714	164902
Non-plant MOTUs	11	0	1918
MOTUs in all samples	NA	383	795
R. alcyone and *H. ruber* MOTUs	1583	267	199
Average MOTUs per individual (all samples)	NA	4	10
Average MOTUs per individual (*R. alcyone* and *H. ruber samples)*	33	6	6

3.3 Negative control samples

Negative control samples, from each triplicate PCR that used trnL or ITS1 primers, were sequenced. The negative control samples from rbcL PCRs were not sequenced. The overlap between negative control and bat sample (non-control) MOTUs is summarised in Figure 3.

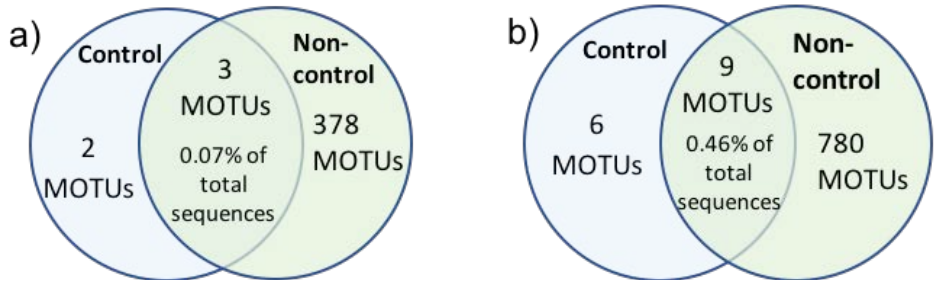

Figure 3: Venn diagrams showing the number of MOTUs found unique to the negative control sample, those shared between the negative control sample and non-control (bat faeces) samples, and those unique to the bat faeces samples, for a) trnL and b) ITS1. The percentage of total sequences corresponding to MOTUS shared between the negative control and bat faeces samples is also indicated.

Of the three shared MOTUs from the trnL samples, one resolved to species level, and for the nine shared MOTUs from the ITS1 samples, none resolved to species level. As the MOTUs that were found in both negative control samples and non-control samples were rare, and contained a very small proportion of sequences, they were not left in the analyses. MOTUs unique to the negative control samples were discarded.

3.4 Taxonomic resolution

Taxonomic resolution was measured as the percentage of MOTUs assigned to order, family, genus and species level for each locus (Table 8). All MOTUs used were resolved to phylum level, therefore taxonomic resolution for phylum level is 100%.

Table 8: Percentage taxonomic resolution for each locus. The table also includes taxonomic resolution calculated when using the 49 samples of *H. ruber* and *R. alcyone* that were compared between all three loci (labelled RH).

	trnL (%)	trnL RH* (%)	ITS1 (%)	ITS1 RH* (%)	rbcL RH* (%)
Order	79.9	80.1	50.6	52.3	59
Family	58.7	61.4	36.0	45.7	31.4
Genus	21.7	23.2	21.9	35.7	3.4
Species	14.1	15.7	11.4	21.1	1.8

*RH indicates taxonomic resolution calculated for the MOTUs belonging to the 49 *R. alcyone* and *H. ruber* samples

Taxonomic resolution was also calculated for insectivorous and frugivorous bats separately (Table 9 and Figure 4).

Table 9: Percentage taxonomic resolution of frugivore and insectivore bats, using trnL and ITS1 loci

	trnL Insectivores (%)	trnL Frugivores (%)	ITS1 Insectivores (%)	ITS1 Frugivores (%)
Order	80.8	81.3	53.4	51.9
Family	62.0	63.3	47.0	36.1
Genus	23.6	24.2	37.2	21.0
Species	15.2	14.6	21.9	9.8

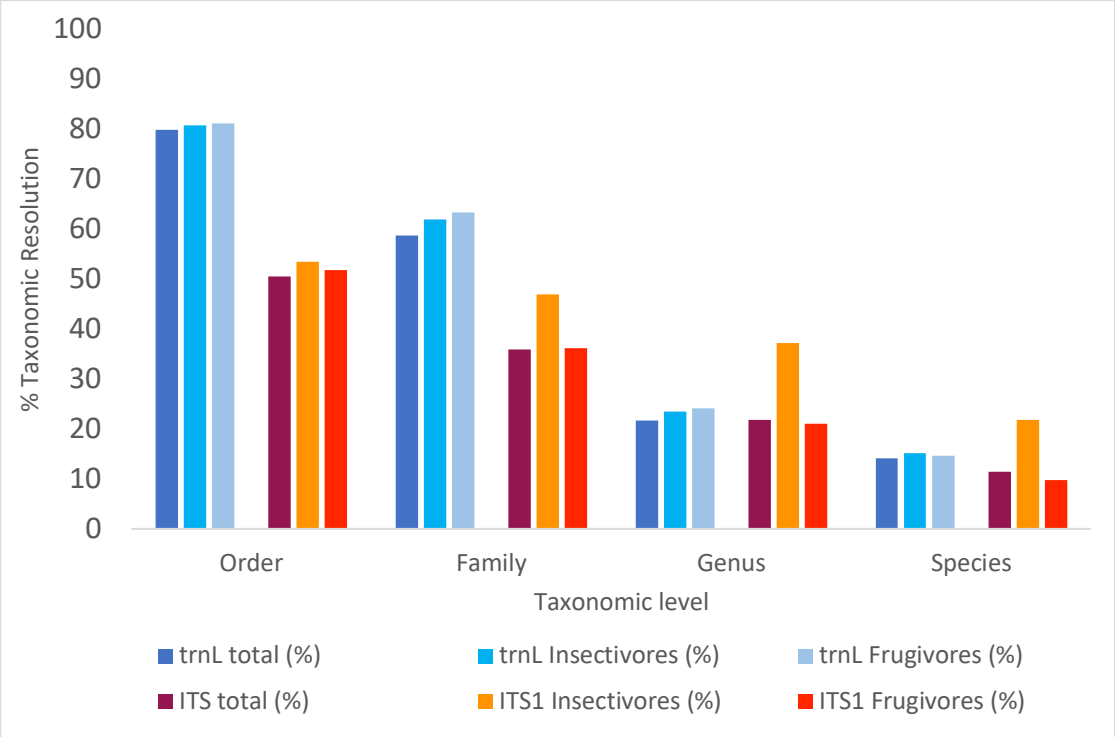

Figure 4: Bar chart comparing % taxonomic resolution achieved by trnL (Blue bars) and ITS1 (red/orange bars) when using total bat samples, insectivorous samples alone, and frugivorous samples alone.

3.5 Diet comparison between loci

Frequency of occurrence (FOO) was calculated, and used to show which were the most common families found in the *R. alcyone* and *H. ruber* samples, when using each locus. The top ten families, found when using each locus, were compared (Figure 5).

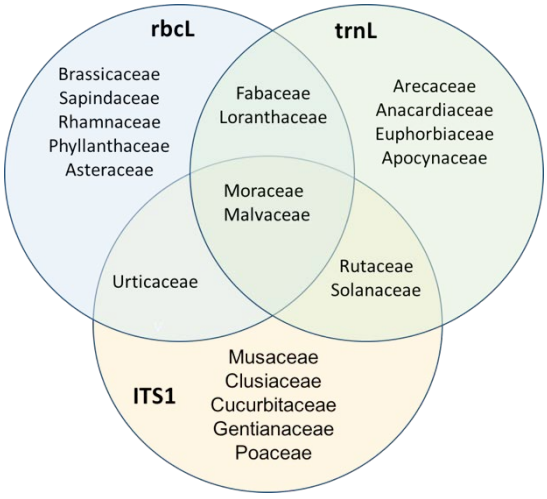

Figure 5: Venn diagram showing the top 10 families in the 49 H. ruber and R. alcyone bat faeces samples, when using each of the three loci. The abundance of families was calculated using frequency of occurrence (FOO)

Of the *R. alcyone* and *H. ruber* MOTUs that resolved to family level, there were 73 families recovered using rbcL, 58 for trnL and 29 for ITS1. Some of these families overlapped between loci and others were unique. Numbers of families are summarised in Figure 6.

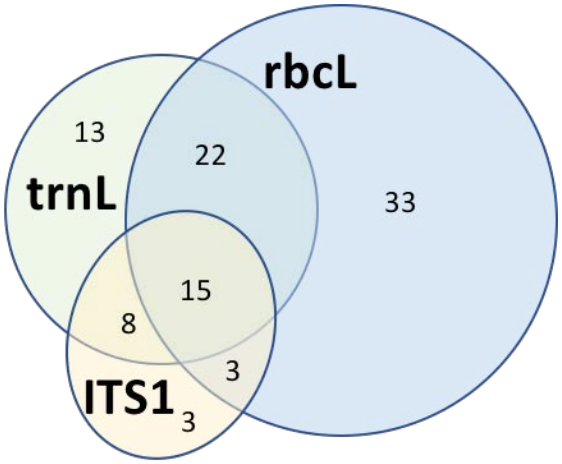

Figure 6: Venn diagram showing the total numbers of families identified in the H. ruber and R. alcyone families when using each of the three loci. Circle sizes are proportional to the number of families but are not to scale

Principle Coordinates analyses (PCoAs) were carried out to compare the species composition of the faeces of each *R. alcyone* and *H. ruber* individual (Figure 7). 3000 sequences were randomly sampled from each individual bat sample to use in the PCoA, and samples with fewer than 3000 sequences were discarded.

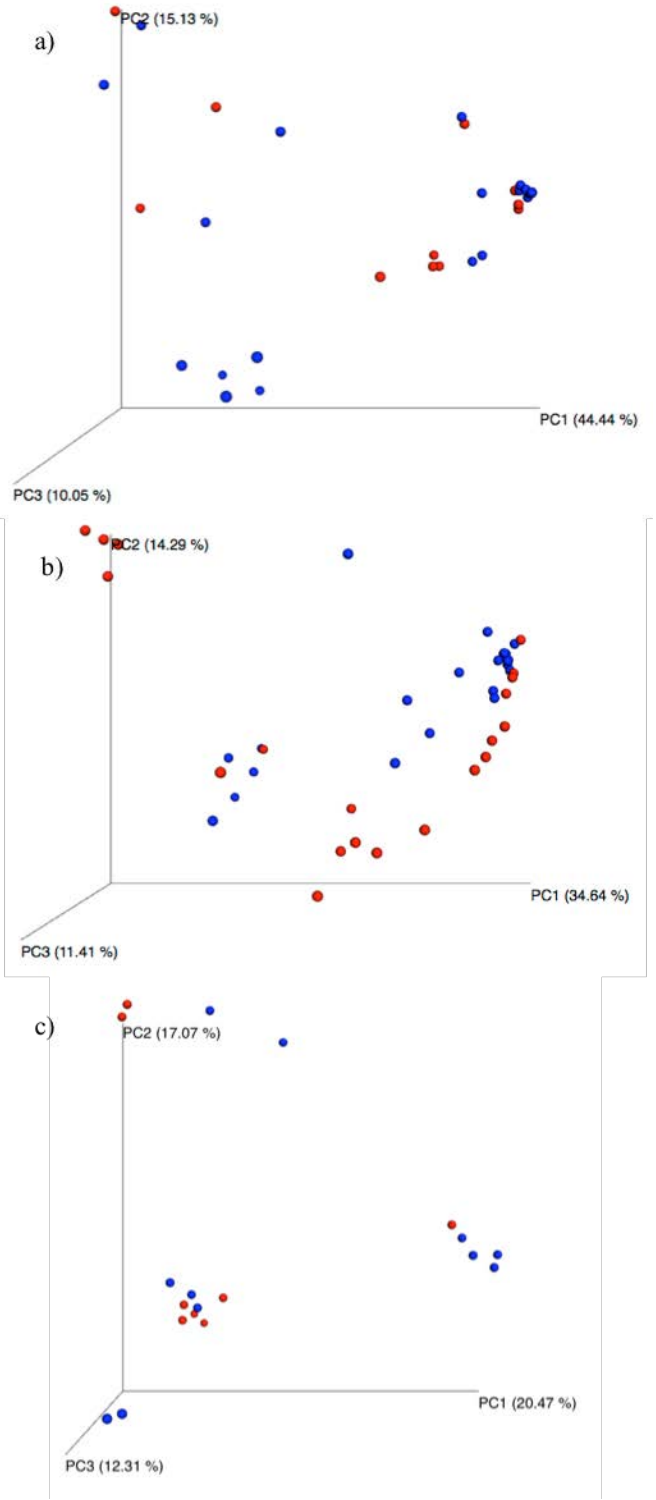

Figure 7: Results of Qiime PCoA (Principle Coordinates Analysis) showing detected taxa in the faeces of the *R. alcyone* (blue points) and *H. ruber* (red points) bats for a) rbcL: PC1 explains 44.44% of the variance, b) trnL: PC1 explains 34.64% of the variance and c) ITS1: PC1 explains 20.47 of the variance

3.6 Comparison of frugivorous and insectivorous bats

The taxa in the faeces samples of frugivorous and insectivorous bats were compared, using trnL and ITS1 MOTUs. The number of MOTUs in the faeces samples of individuals did not follow a normal distribution for neither trnL nor ITS1 data (Shapiro-Wilk. trnL: $W = 0.89205$, $p = 1.624 \times 10^{-6}$. ITS1: $W = 0.86635$, $p = 4.056 \times 10^{-7}$). Therefore, a non-parametric test was carried out.

Analysis using both trnL and ITS1 data suggested there is a significant difference between the number of MOTUs in the diet of frugivorous and insectivorous bats (Two-tailed Mann Whitney U. trnL: $W = 653.5$, $p = 0.002808$; ITS: $W = 1538.5$, $p = 6.288 \times 10^{-10}$)

Figure 8 a) shows that using trnL data, there are significantly more MOTUs in the diet of insectivores than frugivores, and figure 8 b) shows that when using ITS1 data, there are significantly more MOTUs in the diet of frugivores than insectivores.

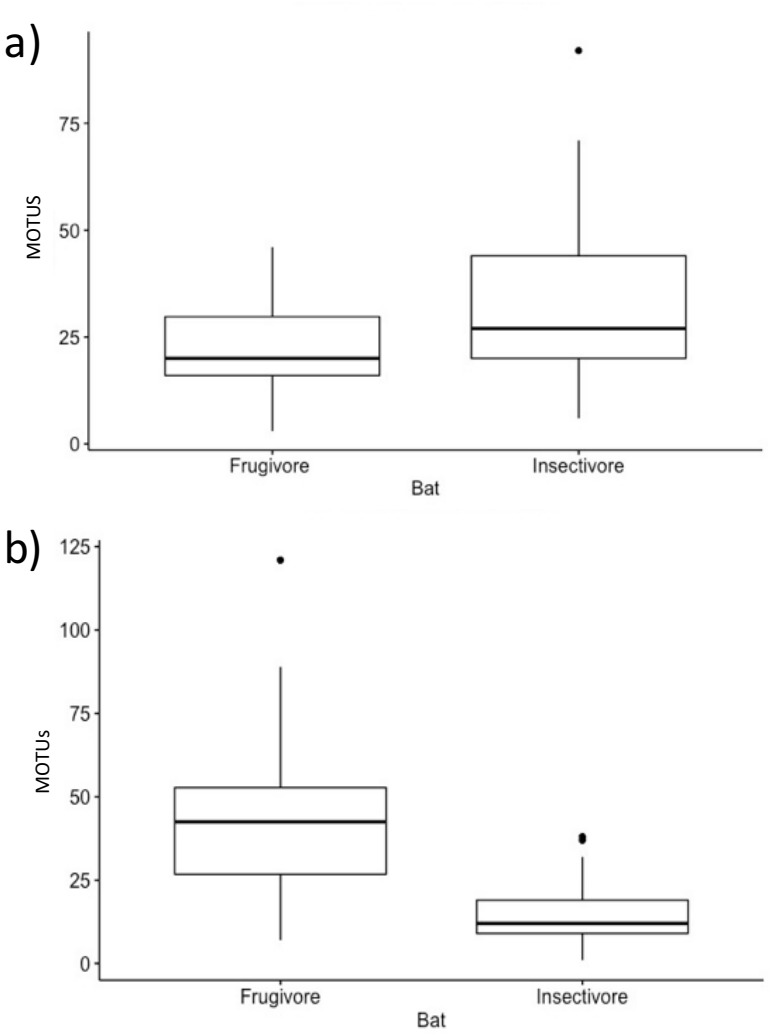

Figure 8: Boxplots comparing the number of MOTUs in the diet of frugivores and insectivores, when using a) trnL locus data, b) ITS1 locus data

Bar plots show how the species composition of bat faeces samples varies between insectivorous and frugivorous bats, and how this difference varied between the locus used (Figure 9).

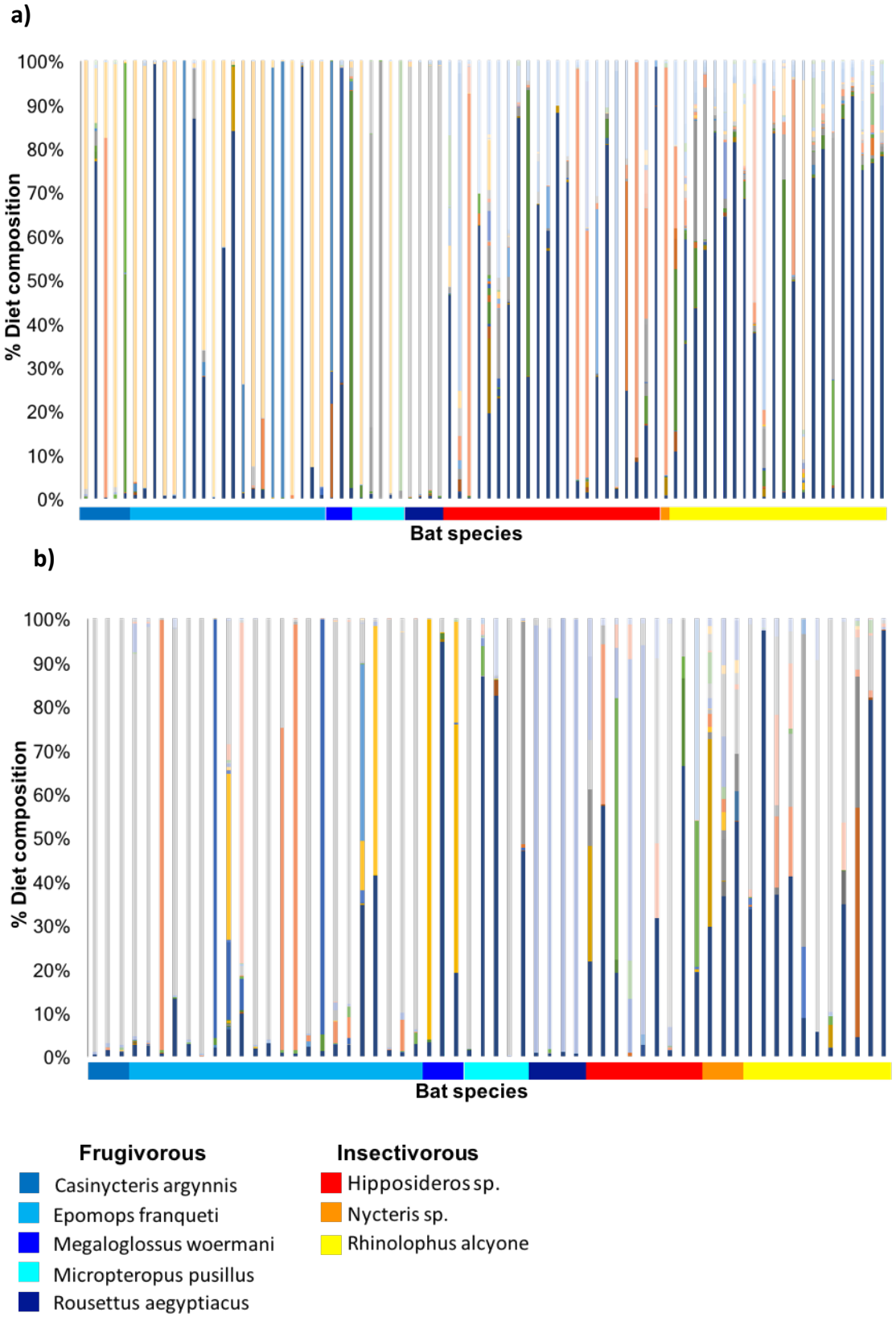

Figure 9: (Previous page): Qiime Taxa Summary bar plot showing the species composition of insectivorous and frugivorous bat faeces, found when using a) trnL and b) ITS1. Each individual bar represents an individual bat. The colours on the x axis represent the bat species, as indicated in the key. Colours within the bats indicate different diet item species. The diet item quantities are based on sequence abundance, standardised to total 100% for each bat species.

Alpha diversities for insectivores and frugivores, for both loci are summarised in Table 10.

Table 10: Mean alpha diversities and standard deviations for insectivores and frugivores, for trnL and ITS1

	trnL		**ITS1**	
	Insectivores	**Frugivores**	**Insectivores**	**Frugivores**
Mean	19.8	13.2	15	31.2
Std.dev	8.3	5.6	7	17.1

Boxplots are used to compare alpha diversities between frugivores and insectivores (Figure 10). For trnL, insectivores had a significantly higher alpha diversity than frugivores (t = 4.1, p = 0.001). For ITS1, frugivores had a significantly higher alpha diversity than insectivores (t = -4.25, p = 0.001).

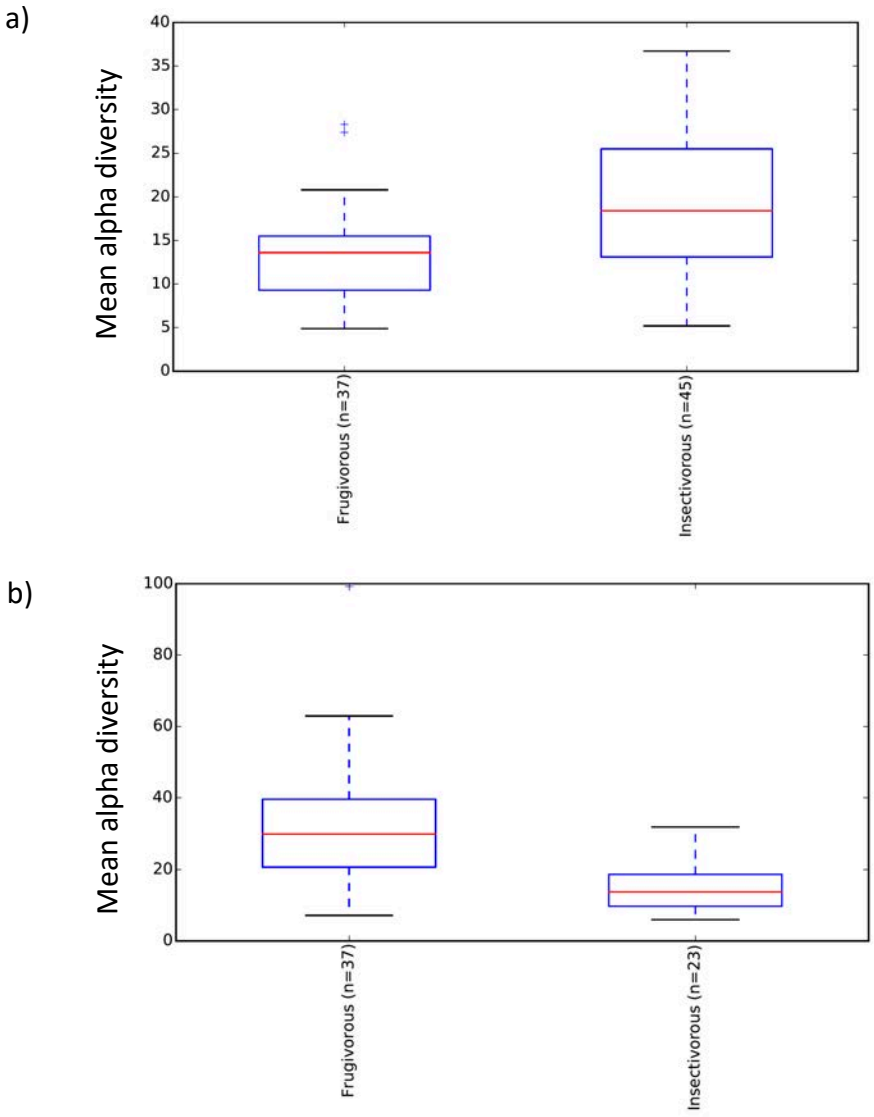

Figure 10: Boxplots showing mean alpha diversities for insectivores and frugivores for a) trnL and b) ITS1

Principle Coordinates analyses (PCoAs) were carried out to compare the composition of frugivore and insectivore bat faeces, using trnL and ITS1 data (Figure 11).

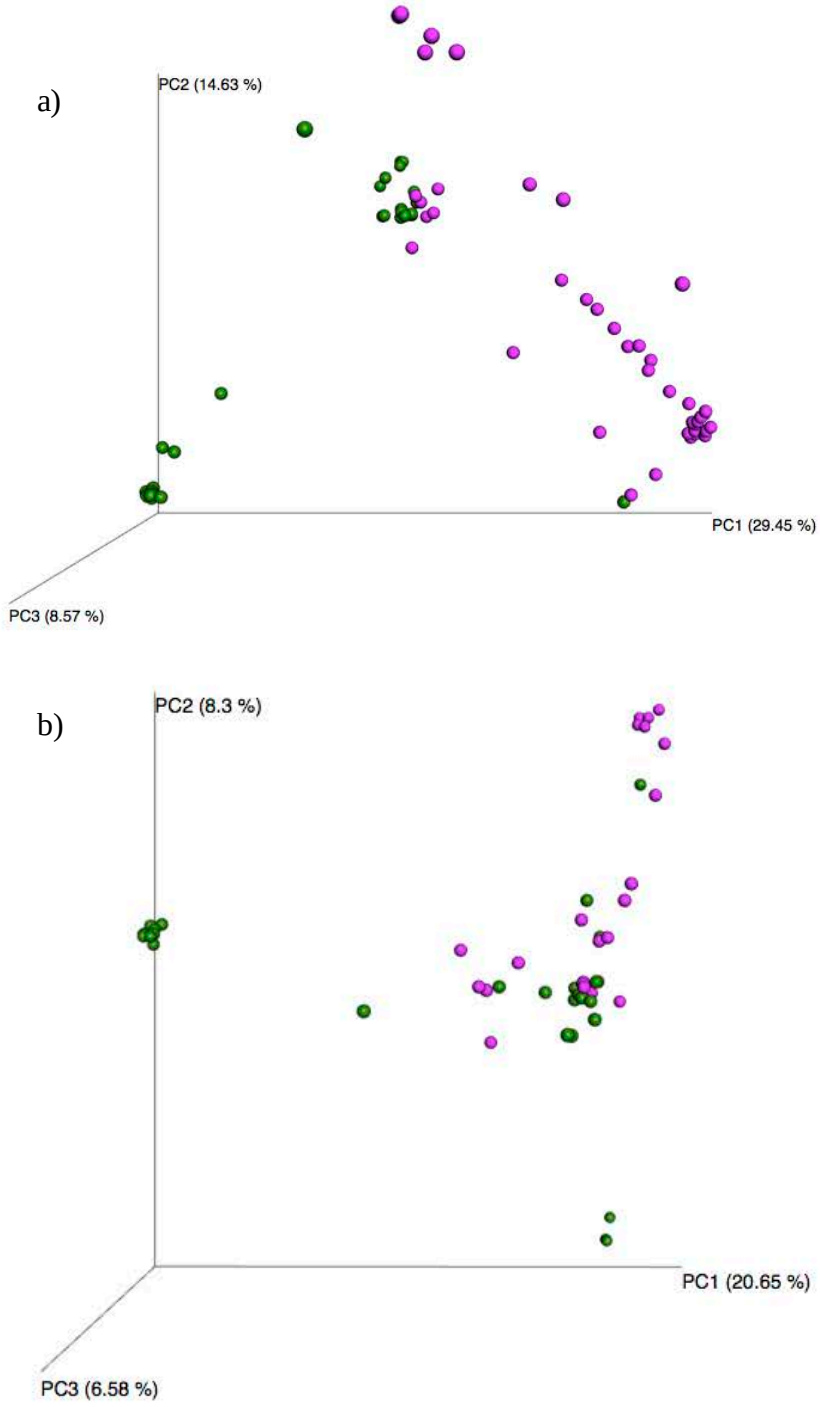

Figure 11: Results of Qiime PCoA showing detected taxa in the faeces of the Insectivorous (purple points) and Frugivorous (green points) bats for a) trnL, b) ITS1

3.7 Occurrence of cacao sequences in bat faeces samples

For trnL data, 42 out of the 49 insectivorous bats had at least five cacao sequences in their faeces, and twelve out of the 42 frugivorous bats had at least five cacao sequences in their faeces. For ITS1 data, two of the 43 insectivorous bats had cacao sequences in their faeces, and none of the 40 frugivorous bats had cacao sequences in their faeces, and

For trnL, the number of cacao sequences in the faeces samples of individuals did not follow a normal distribution (Shapiro-Wilk. $W = 0.37489$, $p = 2.20 \times 10^{-16}$). Therefore, a non-parametric test was carried out.

Analysis (using trnL data) suggested there is a significant difference between the number of cacao sequences in the faeces samples of frugivorous and insectivorous bats (Two-tailed Mann Whitney U. $W = 278.5$, $p = 1.864 \times 10^{-9}$). Figure 12 shows that there are significantly more cacao sequences in the faeces samples of insectivores than frugivores.

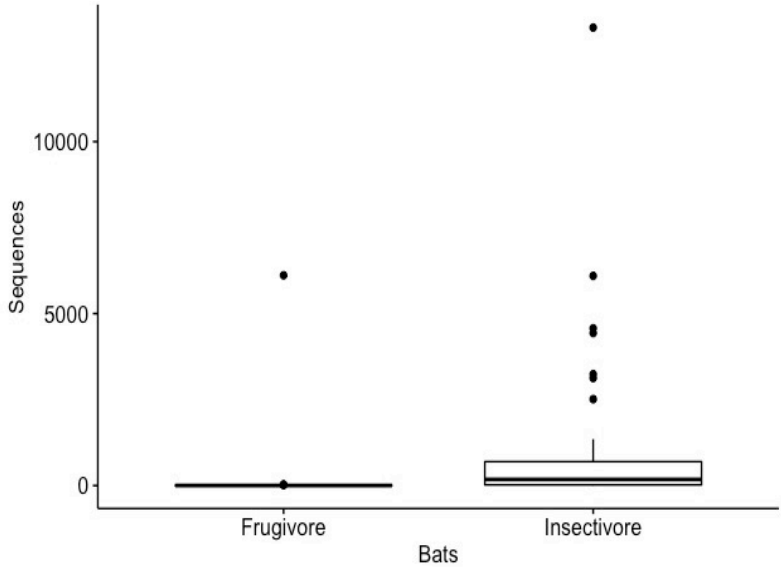

Figure 12: Boxplot comparing the number of cacao sequences in the faeces samples of insectivorous and frugivorous bats

4 DISCUSSION

The aim of this study was to find and test new metabarcoding loci to analyse bat diet, in order to achieve higher taxonomic resolution than that attained in the previous study (A. J. Welch, unpublished), which used the rbcL locus. Ideally, a locus would have high enough taxonomic resolution to discriminate between all species. However, the degraded nature of DNA in faeces and incomplete database coverage means this is unlikely. Therefore, it is important to find the most suitable locus, that will achieve the highest possible taxonomic resolution for this study. The chloroplast locus trnL and nuclear ribosome locus ITS1 were selected, and both compared with rbcL sequence data from the previous study. Both achieved higher taxonomic resolution than rbcL. Other factors, including amplification success and taxonomic coverage (i.e. universality), were also considered.

4.1 Optimal locus for metabarcoding

4.1.1 Low taxonomic resolution of rbcL

rbcL is a core locus in traditional barcoding, used due to its high universality and ease of amplification (CBOL Plant Working Group, 2009; de Vere *et al.*, 2012; Li *et al.*, 2015; Coissac *et al.*, 2016). However, rbcL has the lowest sequence divergence of all chloroplast loci (Kress *et al.*, 2005), which means it has very low species discrimination (Fazekas *et al.*, 2008; de Vere *et al.*, 2012). The rbcL locus used in traditional barcoding is usually around 600bp (de Vere *et al.*, 2012; Dong *et al.*, 2012). The previous rbcL study used a short fragment, of around 89bp, in order to obtain successful amplification in degraded samples. As shorter loci tend to have lower interspecific variability than longer loci, it is likely that the 89bp fragment has even lower discriminatory power than the 600bp fragment (Pegard *et al.*, 2009; Zeale *et al.*, 2010). This explains the low taxonomic resolution achieved using rbcL, showing the need to find alternative loci.

4.1.2 Selection of new loci

trnL

The trnL-P6 locus is commonly used in diet metabarcoding studies (e.g. Baamrane *et al.*, 2012; Hibert *et al.*, 2013; de Barba *et al.*, 2014; Gebremedhin *et al.*, 2016). This is a short fragment of the trnL locus, of around 10-143bp, amplified by g and h primers (Taberlet *et al.*, 2007). It is suitable for dietary analysis due to its short length, high amplification success and conserved primer regions, making it suitable for amplification of degraded samples in a wide range of taxa (Baamrane *et al.*, 2012; Murray *et al.*, 2012).

However, the short length of the trnL-P6 locus means it has lower taxonomic resolution than the full length trnL (Taberlet *et al.*, 2007; Murray *et al.*, 2012; Gebremedhin *et al.*, 2016). However, the full length trnL, amplified with c and d primers, is 254-767bp, is too long to amplify in degraded samples (Taberlet *et al.*, 2007). A compromise is to use c and h primers, which amplify a fragment of around 200bp (Taberlet *et al.*, 2007; Murray *et al.*, 2012). Yang *et al.*, (2016) used c and h primers to analyse the diet of herbivorous water birds, and achieved high interspecific divergence and some species-level resolution. When tested in this study, c and h primers were equally successful at amplifying the loci of shade tree samples as g and h primers were , so were chosen to analyse the bat faeces samples. Therefore, c and h primers were used in this study, with the aim of achieving higher taxonomic resolution than if the g and h primers were used.

ITS1

In barcoding, ITS has good discriminatory power, with higher interspecific divergence at low taxonomic levels than chloroplast loci such as rbcL and trnL (Kress *et al.*, 2005; China Plant BOL Group *et al.*, 2011; Hollingsworth *et al.*, 2011a; Li *et al.*, 2015). ITS can be amplified as a shorter fragment, of ITS1 or ITS2, short enough for amplification in degraded samples (Kress *et al.*, 2005; Li *et al.*, 2015). Wang *et al.*, (2015) proposed ITS1 as a better barcode region than ITS2, due to its higher universality and amplification success in many plant taxa. In diet metabarcoding, ITS1 is often used as supplementary to a core barcode locus (Hollingsworth *et al.*, 2011a). Baamrane *et al.*, (2012) and de Barba *et al.*, (2014) both used ITS1 alongside trnL, and identified 63% and 64% of sequences to species level, respectively. ITS1 is flanked by highly conserved regions 18S and 5.8S, so it can be amplified with universal primers that bind to these regions (Baldwin *et al.*, 1995; Alvarez and Wendel, 2003; Wang *et al.*, 2015). Therefore, ITS1 was used in this study. Some studies use family specific reverse primers for ITS1 (e.g. Baamrane *et al.*, 2012; de Barba *et al.*, 2014; Willerslev *et al.*, 2014; Kartzinel *et al.*, 2015). So, a reverse primer, 5.8S2, was designed for it in this study, based on species of families common in the study site.

4.1.3 Performance of trnL and ITS1 compared to rbcL

The aim of this study was to find loci that provided higher taxonomic resolution than rbcL. However, there is a trade-off between taxonomic resolution, amplification success and universality (CBOL Plant Working Group, 2009; Deiner *et al.*, 2017). Therefore, based on sequence data from the 27 *R. alcyone* and 22 *H. ruber* samples, all three of these factors were considered when comparing the three loci.

Taxonomic resolution

Both trnL and ITS1 achieved higher taxonomic resolution than rbcL, with 16% and 20% of MOTUs resolved to species level, compared to just 2% for rbcL. This higher taxonomic resolution is most likely due to the new loci having higher sequence divergence than rbcL. In this study, the sequence divergence of trnL was 21.8%, around three times that of rbcL (7.8%). ITS1 had an even higher sequence divergence, of 43.5%. Other barcoding studies have also demonstrated that both trnL (Taberlet *et al.*, 2007; Yang *et al.*, 2016) and ITS (Kress *et al.*, 2005; China Plant BOL Group *et al.*, 2011), have greater interspecific variability than rbcL. trnL (169bp) and ITS1 (395bp) are both longer than rbcL (89bp). Their higher interspecific variability could therefore be explained by their longer length, as longer loci tend to have higher interspecific variability than shorter loci (CBOL Plant Working Group, 2009; Clare *et al.*, 2014; Deiner *et al.*, 2017).

Out of the three loci, ITS1 had the highest taxonomic resolution to genus and species level, but lower taxonomic resolution at order and family level (Table 8). The reason for this is unclear, but resolution at this taxonomic level is less important for barcoding (Kress *et al.*, 2005), and for this study, species level resolution is more important.

Amplification success

Amplification success was similar for rbcL and trnL (95% and 94% respectively), but lower for ITS1 (74%). Other studies have also shown rbcL to have high amplification success, including in degraded samples (de Vere *et al.*, 2012). Another reason for its high amplification success is its short length. Shorter loci have higher amplification success (Hollingsworth *et al.*, 2011a). Also, they can amplify more loci of more species in

degraded samples, because a large proportion of DNA fragments in degraded samples are of short length (Deagle *et al.*, 2006).

Despite being nearly twice as long, the amplification success of trnL matched that of rbcL. This suggests that, at 169bp, trnL is short enough to successfully amplify degraded DNA samples in this study. As other studies suggest, loci of around 250bp or less are short enough to amplify in degraded samples (Deagle *et al.*, 2006; Meusnier *et al.*, 2008). ITS1 had lower amplification success, as also found by Chen *et al.*, (2010), probably due to its long length. Longer loci have low amplification success, particularly in samples containing degraded DNA (Pompanon *et al.*, 2012; Wang *et al.*, 2015). ITS, as a whole, has good species-level discrimination, but low amplification success (Gonzalez *et al.*, 2009; China Plant BOL Group *et al.*, 2011).

Universality

rbcL recovered five times as many MOTUs per individual than trnL or ITS1 did (Table 7). This means it amplified sequences in more species of each faeces sample than trnL or ITS1 did. Estimating what proportion of the total number species present in a sample is amplified is not possible without knowing the total number of species. However, comparing the number of MOTUs recovered per individual between different loci can show if one amplified more species than the other, and therefore shows which is the more universal locus. Therefore, rbcL appears to be a more universal locus than trnL and ITS1. trnL is a commonly used locus due to its highly conserved primer binding regions (Taberlet *et al.*, 2007). It may be that trnL does have high universality in this study, but rbcL, due to its short length, was able to recover very degraded samples, resulting in more MOTUs per individual than trnL or ITS1. Also, for ITS1, it is possible that the conserved regions 18S and 5.8S are divergent in some taxa (Omelchenko *et al.*, 2019). This would mean these taxa could not be amplified, which is a possible explanation for the low universality of ITS1 compared to rbcL.

4.1.4 Specific problems of ITS1

ITS1 achieved the highest taxonomic resolution in this study. However, it has some disadvantages, including sequence variation within species, length variation and amplification of non-plant taxa (Hollingsworth *et al.*, 2011a).

Sequence variation

Concerted evolution is where copies of a gene within a genome undergo gene conversion or crossing over, resulting in them becoming identical in sequence. However, incomplete concerted evolution of ITS1 can occur, and lead to the production of divergent copies within individuals (Alvarez and Wendel, 2003; Bailey *et al.*, 2003; Hollingsworth *et al.*, 2011a). This can potentially lead to incorrect identification in barcoding (Jeanson *et al.*, 2011).

Length variation

The length of ITS1 varies considerably between species, due to the presence of duplications and indels, and the presence of what appear to be regions with relaxed evolutionary constraint on length (Baldwin *et al.*, 1995; Gernandt *et al.,* 2001; Floyd *et al.*, 2002; Alvarez and Wendel, 2003). For example, Gernandt *et al.*, (2001) found insertions of around 180bp in ITS1 in some species. In this study, ITS1 agarose gels showed different band sizes, indicating length variation, whereas trnL gels did not. Wang *et al.*,

(2015) estimated its length could be up to 681bp in eudicots. Length variation is a problem in metabarcoding because shorter sequences are preferentially amplified over longer sequences (Wang *et al.*, 2015; Gebremedhin *et al.*, 2016; Deiner *et al.*, 2017). This means that species with exceptionally long ITS1 loci cannot be amplified in degraded samples. This could also explain the lower amplification success and low universality of ITS1 found in this study.

Amplification of non-plant taxa

In this study, the majority of MOTUs generated were assigned to non-plant taxa, most of which were fungi. ITS is a core locus for barcoding fungi (Schoch *et al.*, 2012), so fungal contamination is a common problem when using ITS for plant barcoding, even when barcoding fresh leaves directly (Zhang *et al.*, 1997). If fungal contamination occurs after faeces production, then it is likely that the fungi DNA is less degraded, therefore easier to amplify. If fungi are preferentially amplified by ITS primers, this may reduce plant amplification (Alvarez and Wendel, 2003). A suggestion to reduce contamination is to use plant-specific primers (Cullings and Vogler, 1998; Kress *et al.*, 2005), but this did not appear to make a difference in this study.

4.1.5 Comparison of trnL and ITS1 using all bat species analysed

Overall, rbcL has broad taxonomic coverage, rbcL and trnL have the high amplification success, and trnL and ITS1 have the high taxonomic resolution. The taxonomic resolution of ITS1 was only slightly higher than trnL. Therefore, it may be better to use trnL, in order to avoid the problems associated with using ITS1. Because the rbcL data used here was derived from a previous study, the comparison of rbcL, trnL and ITS1 was based on sequence data from just two species of insectivorous bat, 27 individuals of *R. alcyone* and 22 individuals of *H. ruber*. In this study, however, sequencing data from a total of 99 bat individuals, of five insectivorous and five frugivorous species, was also collected (Appendix A). Therefore, the taxonomic resolution and universality of trnL and ITS1 were compared again, using sequence data from all 99 samples.

Universality

ITS1 recovered twice as many MOTUs per individual than trnL, when using data from all samples together, suggesting it is more universal than trnL. A possible reason for this is that around 70% of ITS1 MOTUs were frugivorous, compared to trnL with just 30%, so adding the frugivorous samples increased the total number of MOTUs. It could be that plant DNA in insectivorous bat faeces is more degraded than plant DNA in frugivorous bat faeces. Therefore, as ITS1 is long, it cannot amplify the more degraded DNA in insectivorous bat faeces so successfully, resulting in it recovering fewer MOTUs from insectivore samples.

Taxonomic resolution

For analysis of all samples, trnL generally had a higher taxonomic resolution than ITS1. However, at genus and species level the taxonomic resolution of ITS1 and trnL were comparable, and ITS1 was higher for insectivores alone (Figure 4). Therefore, trnL appears to be the more effective for taxonomic resolution. If ITS1 is used, its low amplification success and variable length must also be considered.

ITS1 has been proposed as a core barcode due to its high interspecific divergence (China Plant BOL Group *et al.*, 2011). In contrast to this study, other studies have shown a large

improvement in resolution when using ITS over chloroplast loci (Kress *et al.*, 2005; Roy *et al.*, 2010; Muellner *et al.*, 2011). It was concluded that this high resolution of ITS outweighs its disadvantages (Hollingsworth, 2011b).

In general, the potential disadvantages of ITS appear to be less of a problem than was first thought (Hollingsworth 2011b; Song *et al.*, 2012). ITS is made up of the shorter fragments ITS1 and ITS2 (Yao *et al.*, 2010). ITS2 is shorter and amplifies more successfully than ITS, but still has high interspecific divergence (Chen *et al.*, 2010; China Plant BOL Group *et al.*, 2011; Jeanson *et al.*, 2011). This study used ITS1, which has similar advantages to ITS2 over ITS1, and also has higher divergence than ITS2 in some plant groups (Wang *et al.*, 2015).

Given findings from the literature that ITS1 has high taxonomic resolution, and the calculation of ITS1 sequence divergence in this study, which was around twice that of trnL, the results here showing ITS1 has lower taxonomic resolution than trnL are unexpected. This could be related to locus length or limited database coverage.

Shorter loci amplify more successfully than longer loci, particularly in degraded samples (Hollingsworth *et al.*, 2011a). As ITS1 has length variation, it is possible that in this study, only the shorter fragments were amplified. Shorter loci have lower taxonomic resolution (Deiner *et al.*, 2017), which could be why ITS1 had lower taxonomic resolution than expected.

Database coverage is a limiting factor in achieving taxonomic identification in metabarcoding studies (Pompanon *et al.*, 2012; Groom *et al.*, 2017). Given the low taxonomic resolution of ITS1 in comparison to trnL, despite its higher divergence, it is possible that ITS1 has less broad database coverage in the study site than trnL. It is not easy to estimate database coverage without knowing all plant species present. However, a broad estimate can be made by estimating database coverage for different countries. For trnL, there are around 420 Cameroon records, compared to 650 in the UK and 8600 in the USA. For ITS, there are 308 Cameroon plant records, 580 in the UK and 7400 in the USA. Therefore, trnL appears to have broader coverage than ITS1, but both are limited in Cameroon. This limited coverage in Cameroon suggests coverage for the study site in particular was also limited. In general, African rainforest trees have limited database coverage in Genbank (Parmentier *et al.*, 2013). Increasing the database or constructing a local database is not something that will happen instantly, so for the moment, trnL appears to be the most suitable barcode. ITS1 still has potential use in this study, possible for identification of particular plant groups (Gonzalez *et al.*, 2009). If database coverage is the main factor limiting the taxonomic resolution of ITS1, then it has the potential to provide higher taxonomic resolution in this study, should the database become more extensive.

Increasing database coverage could also increase taxonomic resolution for trnL. In a study of ibex diet, using a local reference database, Gebremedhin *et al.*, (2016) resolved 29.8% of sequences to species level, which is higher than this study, despite the use of a shorter locus. While there may be other factors affecting taxonomic resolution in this study, such as the study species, this suggests that higher taxonomic resolution could be achieved with a local database.

Another potential reason for low taxonomic resolution in any locus is that some taxonomic groups have low variability in a locus sequence, so cannot distinguish to species level,

even when there is species-level database coverage. For example, trnL has low variation in the families Asteraceae, Cyperaceae, Poaceae and Rosaceae (Raye *et al.*, 2011; de Barba *et al.*, 2014), and ITS has low variation in the family Phyllostachyae (Starr *et al.*, 2009). In this study, Moraceae was the most common family in the frugivore MOTUs recovered using ITS1, with around four times the FOO of any other family. Moraceae contained many MOTUs of the genus *Ficus.* In Genbank, *Ficus* sequences matched equally well with a large number of different species. This suggests that *Ficus* ITS1 loci are not divergent enough to distinguish between species. As the genus *Ficus* was so prominent, this may be part of the reason for the taxonomic resolution for ITS1 being particularly low in frugivores compared to insectivores (Table 9).

There are many other possible reasons why higher taxonomic resolution was not achieved in this study. While species resolution was relatively low, up to 60% of MOTUs (depending on locus) were resolved to family level. Identifying diet items to family level only may not be particularly useful, as the aim of this study is to identify shade trees to species level. However, it still gives an idea of which families the important shade tree species belong to. Also, given good knowledge of shade tree species present, it may be possible to predict the most likely plant diet species from knowing just the families.

4.2 Variation in diet between loci

So far, taxonomic resolution, amplification success and universality have been considered as factors that determine the most effective locus in this study. It is also important to assess how the taxa detected in the faeces varies depending on the locus used.

There was some variation and some overlap between families detected in the faeces samples using rbcL, trnL and ITS1 loci (Figures 5 and 6). No single locus can detect the total diet; instead each detects a different sample of taxa. Using each locus detected at least some unique families, suggesting that some families are more easily detected by using certain loci.

Amplification bias results in some taxa being preferentially amplified over others. Causes of amplification bias include primer bias, length related bias and reference database bias (Bohmann *et al.*, 2011; Deiner *et al.*, 2017; Lobo *et al.*, 2017; Alberdi *et al.*, 2018). The extent of how much the DNA is degraded in different taxa also varies (Deagle *et al.*, 2006). Amplification bias and the ability to detect degraded samples varies between loci. Therefore, in different loci, some taxa are preferentially amplified and over-represented, and others may be under-represented or not detected at all. This results in variation in the taxa detected using each locus.

Figure 5 shows the top ten families found in the bat faeces when using each of the three loci, calculated using Frequency Of Occurrence (FOO). Top families in the diet that are unique to a single locus are more likely to be present partly due to amplification bias, than those top families that overlap between loci, as amplification bias is reduced when using more than one locus (Alberdi *et al.*, 2018). This suggests that overlapping top families can be said to be common with more certainty than those that are unique. However, common families that are unique to a single locus may be very common, but just undetectable by the other loci. For example, the family Arecaceae had the highest FOO when using trnL, but was not detected using rbcL or ITS1.

Principle Coordinates Analyses were carried out to compare the diets of *R. alcyone* and *H. ruber* bat individuals between the three loci. For all three loci, there is variation and some clustering between individual bat diets (Figure 7). Clustering indicates that the bat individuals have similar diet composition. This could be due to these individuals having similar foraging areas or foraging habits, or living in the same geographical area. The clustering is more pronounced in rbcL and trnL than ITS1. This could be influenced by the numbers of individuals used in the analysis. The PCoAs only used data from individuals with 3000 sequences or more, and trnL and rbcL had more individuals than ITS1 with 3000 sequences or more. Alternatively, this could be because rbcL and trnL have more diet similarities between them than with ITS1. As shown in figure 6, there are more overlapping families between rbcL and trnL than with ITS1. In general, there was no clear distinction seen between *R. alcyone* and *H. ruber*, which suggests they have similar diets.

Using rbcL recovered a larger number of unique families than using trnL or ITS1. This is likely to be due to its higher universality, enabling it to detect a larger number of, and more degraded, taxa. Also, analysis involving rbcL was carried out on the same samples but separately, around six months before this study. This means a slightly different subsample of the faeces pellets were analysed, which may therefore contain slightly different taxa. Also, DNA amplification success of faeces samples has been shown to decrease over time, even as short as a few months (Franzten *et al.*, 1998; Murphy *et al.*, 2002). Therefore, it is possible that some species become too degraded in the time between the study using rbcL (A. J. Welch, unpublished) and this study.

4.3 Dietary analysis

In the process of assessing the effectiveness of different metabarcoding loci, useful information about bat diet was also gained. This included finding common families in the diet, as already discussed, comparing insectivore and frugivore diets and determining the occurrence of cacao in the diet.

4.3.1 Common families

All three loci had the families Malvaceae and Moraceae as two of their ten most common families (Figure 5). This suggests that these families make up a large proportion of the taxa in the faeces samples. Malvaceae includes cacao trees. Finding cacao sequences in the diet of insectivorous bats suggests that these bats could be feeding on the cacao pests. Moraceae includes fruit trees, such as figs and breadfruit, and high value timber (Sonwa *et al.*, 2007). Identifying species in the bat diet that have an economic value is useful as they will benefit cacao farms both economically and by supporting biodiversity (Rice and Greenberg, 2000). Having an economic value also gives farmers an incentive to plant these biodiversity-supporting tree species. Other families containing species of economic value include Arecaceae, which contains some fruit species, and Apocynaceae, which contains medicine and timber providing trees (Sonwa *et al.*, 2007).

4.3.2 Comparison of insectivorous and frugivorous bats

Plant DNA in frugivore bat faeces corresponds to the plants eaten by those bats, but plant DNA in insectivore faeces may correspond to insect diet, or be present for another reason. The differences in composition and abundance of taxa in the faeces of frugivorous and insectivorous bats was compared.

For trnL, the number of MOTUs detected per individual was significantly larger for insectivores than frugivores (Figure 8). Insectivores also had a significantly higher mean alpha diversity than frugivores (Table 10, Figure 10), as may be expected given the larger number of MOTUs. The trnL bar plot (Figure 9) shows more variation in insectivore diet compared to frugivore individuals, suggesting higher diversity in the insectivore diet. Assuming that the plant DNA present in the insectivorous bat faeces corresponds to the diet of the bats' insect prey, then these results might be expected. Adding together all the plants in the diet of each insect eaten by the bat would perhaps be greater than the plants in the diet of a frugivorous bat. Frugivorous bats may only feed on up to five fruits in one night (Thomas, 1984), but insectivorous bats could feed on seven or more species, and therefore would have a larger number of plant species in their faeces (Clare *et al.*, 2009).

However, using ITS1 gave contrasting results. There were significantly fewer MOTUs per individual for insectivores than frugivores bats (Figure 8). Insectivores also had a significantly lower mean alpha diversity than frugivores (Table 10, Figure 10) in contrast with trnL. DNA becomes degraded during digestion (Deagle *et al.*, 2006), therefore plant DNA that has passed through both the insect and bat digestive systems will be very degraded, so less easy to detect. It is possible that trnL is short enough to detect this degraded DNA, but ITS1 is not, and therefore fewer MOTUs were discovered, and the diet was less diverse for insectivores when using ITS1. There is a less obvious difference between insectivorous and frugivorous bats shown in the ITS1 bar plot (Figure 9). From the bar plot, the diet of frugivorous bats appears to be dominated by a single species, with other species of low abundance. However, this is not definitive, as the small sample size used in this analysis is unlikely to be representative of all frugivorous bats in Cameroon cacao agroforests.

The PCoAs comparing frugivorous and insectivorous bats both show variation between individuals (Figure 11). Each bat has a slightly different diet, which may be expected as both frugivorous and insectivorous species have diverse diets (Kwiecinski and Griffiths, 1999; Zeale *et al.*, 2010). There is also some clustering of frugivorous or insectivorous bats together, suggesting these individuals were from the same farm or shared similar foraging areas or foraging habits. trnL shows some division between frugivores and insectivores, although some individuals are clustered. ITS1 shows much greater overlap between frugivores and insectivores, but with some frugivorous bat outliers. The corresponding bar plots also suggested a clearer difference between insectivores and frugivores for trnL than ITS1. Clustering of insectivores and frugivores together suggests that the insect prey and frugivorous bats may share the same diet. Alternatively, there are both insectivorous and frugivorous species that roost in trees (Owen-ashley and Wilson, 1998; Decher and Fahr, 2005), so they could both ingest the same plant species, either intentionally or unintentionally.

Overall, the two loci appear to give very different results in terms of diet. Using a multi-locus barcode may help to give more consistent results. Also, this analysis was based on data from a small number of bat samples. Using a larger sample size would be more representative of the whole bat population, and may also help to give more consistent results.

4.3.3 Occurrence of cacao in bat diet

Malvaceae was found to be a commonly occurring family in the *R. alcyone* and *H. ruber* samples (Figure 5), and includes cacao trees. The occurrence of cacao in the diet of

insectivorous bats could indicate that they feed on insect pests of the cacao, assuming it is possible to detect plant DNA from the diet of these insects.

Using ITS1, just three insectivorous bat samples contained cacao sequences, and none were found in frugivorous bats. This could be because the cacao ITS1 sequence, at 395bp, is too long to be easily amplified, or becomes degraded if it passed through both insect and bat digestive systems. ITS1 has a large size range between species and it could be that cacao is a species with a particularly long ITS1 region compared to other species found in the bat diet in this study.

For trnL, there were significantly more cacao sequences in insectivore than frugivore samples (Figure 12). Frugivorous bats are very unlikely to eat cacao pods, as they are very large compared to the bats' small mouths (Luke L. Powell, Personal communication 23/4/2019). However, there were a few frugivorous bats that appeared to eat cacao, including a *Casinycteris argynnis* individual with over 4000 cacao sequences in the sample. Fruit bats do sometimes feed on leaves (Kwiecinski and Griffiths, 1999), so it is possible that some feed on cacao leaves.

Therefore, this has shown a link between insectivorous bats, which are potential cacao pest controllers, and cacao. This is a useful starting point in making a link between cacao, potential pest control bats and the shade trees that are support these bats, either by hosting their insect prey or providing roosts.

4.3.4 Plant DNA in insectivorous bat faeces

Plant DNA in insectivorous bat faeces may correspond to the plants eaten by their insect prey, but this is not definite. It is not possible to distinguish between primary and secondary predation using metabarcoding (Sheppard *et al.*, 2005). Therefore, while it is very unlikely that insectivorous bats ingest plant material directly, there may be other reasons for the present of plant DNA in their faeces. Plant DNA remains detectable after passing through insects (Valentini *et al.*, 2009), or bats (Aziz *et al.*, 2017), and insect DNA remains detectable after passing through bats (Clare *et al.*, 2009; Zeale *et al.*, 2010). However, it is not known whether plant DNA can be detected after passing through both the insect and bat digestive systems. Kalka and Kalko, (2006) demonstrated that the insectivorous bat *Micronycteris microtis* discarded plant material from the intestines of phytophagous insects before feeding, to avoid excess bulk and inhibitors. This suggests that even if plant DNA was detectable, bats may avoid consuming plants. However, there is still plant DNA present in the insectivorous bat faeces. Some insectivorous bats roost in trees (Decher and Fahr, 2005; Nurul-ain *et al.*, 2017), and therefore may unintentionally ingest plant fragments, which could explain the presence of plant DNA. Even if the plant DNA is present due to unintentional ingestion, or roosting in trees, the plant species present may still give an indication of which tree species are used by the insectivorous bats.

4.4 Methodological considerations

As demonstrated in this study, metabarcoding is a useful tool for dietary analysis. However, certain steps of the methods must be considered carefully, as they can have a large influence on the results.

4.4.1 MOTU clustering

The choice of MOTU clustering threshold can have a significant impact on the number of MOTUs generated, and therefore on the apparent number of species. In this study, clustering thresholds were selected that were most suitable for each locus by comparing how the clustering threshold impacted the number of MOTUs generated, as proposed by Alberdi *et al.*, (2018). This helps to reduce the likelihood of over- or under- splitting. In this study, some MOTUs matched the same species, which suggests over-splitting. However, as the faeces samples were very diverse, other taxa could be under-split. No single threshold is suitable for all species (Vetrovsky and Baldrian, 2013). The thresholds used in this study are similar to those used in other diet metabarcoding studies (Bohmann *et al.*, 2011; Alberdi *et al.*, 2018). It is not easy to find a more accurate threshold, unless enough is known about the diet to use group-specific thresholds (Brown *et al.*, 2015). Also, there is evidence to suggest that changes in clustering threshold can still result in the same general conclusions (Brown *et al.*, 2015; Clare *et al.*, 2016)

4.4.2 Error removal

Erroneous sequences (artefacts) are generated during PCR and sequencing (Brown *et al.*, 2015; Alberdi *et al.*, 2018). It is often assumed that very rare sequences are artefacts, and therefore rare sequences are excluded from analyses (Pompanon *et al.*, 2012; Alberdi *et al.*, 2018). However, discarding rare sequences risks also discarding real, but rare, taxa (de Barba *et al.*, 2014). In this study, PCRs were carried out in triplicate, and all PCR products were pooled. This additive replicative strategy aims to maximise the detection of diversity, but also increases the likelihood of rare sequences which could be artefacts (Alberdi *et al.*, 2018). Other studies (e.g. de Barba *et al.*, 2014) used restrictive strategies. Restrictive strategies involve only retaining sequences that are detected in more than one PCR replicate. The aim of this is to reduce the likelihood of erroneous sequences, which may be only found in single PCR replicates due to their rarity (Alberdi *et al.*, 2018). In this study, a more common error removal strategy was used, involving setting a minimum copy number (Alberdi *et al.*, 2018). In this study, MOTUs with less than five sequences were discarded. However, even the most stringent error removal strategies may not be completely effective, as false sequences may be much commoner than expected. This is a particular problem if PCR is used in later steps to add Illumina indices to libraries, as this can also increase the abundance of false sequences formed in earlier PCRs (Pompanon *et al.*, 2012; Alberdi *et al.*, 2019). It is important to find a balance between maximising biodiversity, but also minimising the occurrence of false sequences.

4.4.3 Quantitative diet estimates

A number of studies have shown that metabarcoding cannot give a reliable quantitative estimate of diet using sequence abundances, and presence-absence data should be used instead (Elbrecht and Leese, 2015; Pinol *et al.*, 2019). However, sequence abundance data is still frequently used, often in conjunction with presence-absence data (Raye *et al.*, 2011; Kartzinel *et al.*, 2015; Leray and Knowlton, 2015; Nichols *et al.*, 2016). In this study, presence-absence data was used to calculate FOO to compare the abundances of each family in the diet between loci. Sequence abundance data was used to compare the

abundance of cacao between insectivores and frugivores, and also for the Qiime PCoAs and bar plots. Sequence abundance data was required for these analyses in order to assess abundances of individual diet items within individuals. Both sequence abundance and presence-absence methods have disadvantages. Amplification bias and differences in degradation and digestion times of diet items result in some species having a larger or fewer number of sequences relative to their actual abundance (Bellemain *et al.*, 2010; Gebremedhin *et al.*, 2016; Pinol *et al.*, 2019). However, using presence-absence data means it is less easy to distinguish between rare and common taxa, and rare taxa can appear to be more common than they actually are (Alberdi *et al.*, 2019; Deagle *et al.*, 2019). Rare and erroneous sequences can also be disproportionately represented when presence-absence data is used (Deagle *et al.*, 2014). Neither method is perfect, but despite these limitations, both can provide reliable data to some extent.

4.5 Limitations

4.5.1 Sample size

A limited number of bats were sampled, from five frugivorous bat species and five insectivorous bat species (Appendix A). At least nine species of frugivorous bat and 20 species of insectivorous bat have been recorded in the Cameroon agroforest (Luke L. Powell, Personal Communication 05/12/2018), so this study did not sample the whole diversity. However, conclusions made about selecting loci are unlikely to differ if a larger sample size were to be used. Sampling a larger number of individuals and a larger number of species would give a more accurate representation of the bat diet. Analysing the contents of faeces only shows what was recently eaten by the bats (Deagle *et al.*, 2006; Alberdi *et al.*, 2019). Therefore, sampling over a time period would also help to give a more accurate picture of bat diet throughout the year.

The PCoA results indicated that the bats had a variety of different diets, which suggests they are generalists. Bats may be unlikely to make species level choices when foraging. Instead, food choice is more likely to depend on what is available in the time and space (Clare, 2014a). Therefore, identifying individual important shade tree species may require sampling of a large number of bat species and individuals over a longer time period, in order to find out which species are generally used most frequently and by most bat species.

4.5.2 Loss of diversity

Using metabarcoding, particularly with a single locus, does not recover the total diversity in a sample (Clarke *et al.*, 2014). Even when a sample is successfully amplified, not all species in that family are amplified, so will go undetected (Deagle *et al.*, 2014). Metabarcoding studies can only use sequences that were amplified. In this study, comparisons made about diet were based only on MOTUs that resolved to the taxonomic level in question, which in most cases was family. Only MOTUs that resolved to the phylum Streptophyta were used in analysis. For trnL, all MOTUs resolved to Streptophyta, but for ITS1, some resolved to other phyla (Table 7), and others did not resolve to phyla level at all. It is possible that some plant diversity was lost if it could not be distinguished from non-plant phyla. The QIIME analyses used to generate the PCoAs and bar plots used a randomly sampled 3000 sequences from each individual. This aimed to maximise the number of bat samples used, but resulted in a loss of sequence data from each sample.

4.5.3 Erroneous sequences

Despite steps to remove erroneous sequences, it appeared that some were still present in these results. In this study, ITS1 data contained some Porifera (sponge) sequences. These animals would definitely not be present in the study site, which suggests they are erroneous sequences. Therefore, it is possible that the plant sequence data used in analysis contains false sequences, but it is unlikely that this makes a large impact on the results.

4.5.4 Contamination

There were some contamination problems in the laboratory work. Some PCRs needed to be repeated due to contamination shown in the negative control lane of the gels. However, the absence of a band on the control lane does not necessarily mean there is no contamination, and therefore negative controls should be sequenced (Alberdi *et al.*, 2019). In this study, sequencing the negative control samples showed there was still some contamination. The MOTUs that were present in both negative control samples and bat samples were not removed because they made up a very low proportion of the total sequences, so were unlikely to have a significant impact on the results (Figure 3). Removing negative controls reduces the likelihood of retaining false diversity or artefactual sequences, but it also can result in the loss of true diversity (Alberdi *et al.*, 2019). Also, even removing all negative control MOTUs may not remove contamination completely. In this study, there were *Arabadopsis thaliana* sequences recovered in ITS1 MOTUs. It is very unlikely for *Arabadopsis* to be present in Cameroon (Hoffman, 2002; Brennan *et al.*, 2014), which suggests its presence is due to contamination from neighbouring plant laboratories that use *Arabadopsis* DNA. Contamination could be reduced by using a dedicated DNA extraction and PCR extraction hood, but these are costly.

4.6 Improvements and future directions

4.6.1 Local database

As database coverage is a major limiting factor in metabarcoding (Parmentier *et al.*, 2013), some studies construct local databases for their study area (e.g. de Barba *et al.*, 2014; Kartzinel *et al.*, 2015). Constructing a local database for the Cameroon agroforests is likely to give a major improvement in taxonomic resolution in studying bat diet. Also, using a local database alone, or restricting the database to plants only in the study area can help to increase discrimination as not all closely related species are present in the same area (Chase and Fay, 2009; Gonzalez *et al.*, 2009). This could also be used to avoid fungal sequences completely, which may avoid fungal contamination problems associated with ITS1.

4.6.2 Further loci optimisation

The ITS1 primer has some potential as a useful locus in this study, but it had low amplification success and amplified lots of fungal sequences. A possible improvement could be to re-design either or both of the primers. Cheng *et al.*, (2016) designed plant-specific ITS primers. These were designed with as many base sequence differences with fungi as possible, in order to reduce fungal contamination. Also, a mock community of known species could be constructed, to more accurately test the universality and extent of bias in loci. Several studies have used mock communities, and they are also often used to test if quantitative metabarcoding is reliable (Piñol *et al.*, 2015; Blanckenhorn *et al.*, 2016; Galan *et al.*, 2018).

As discussed, a using a combination of loci could improve dietary analysis studies, both by increasing taxonomic coverage and resolution and decreasing amplification bias. In this study, rbcL could be combined with trnL or rbcL. However, the extra cost and labour must be considered.

4.6.3 Multi-locus barcodes

A possible future direction to improve taxonomic coverage, resolution and reduce potential bias would be to use multi-locus barcodes (Chase and Fay, 2009; Alberdi *et al.*, 2018). As shown by the results of this study, locus choice influences the taxa identified in the diet, so sometimes more than one locus can be used (Alberdi *et al.*, 2018). The use of multi-locus barcodes is very common in plants (Kress and Erickson, 2007; Fazekas *et al.*, 2008). Often, a locus with broad coverage is combined with a locus with high species discrimination (Hollingsworth *et al.*, 2009). In this study, rbcL could be combined with trnL or ITS1 to achieve both high taxonomic resolution and broad coverage.

Using more than one locus increases the proportion of taxa detected (Fazekas *et al.*, 2008). As shown in Figures 5 and 6, more families were detected in the diet when combing the results from all three loci. A perhaps extreme example is the twelve-locus barcode developed by Arulandhu *et al.*, (2017) to identify protected species in processed food products. This multi-locus barcode was designed to detect both degraded DNA of both plants and animals. It identified a large range of samples to high resolution, including species that were present in very small quantities.

Using a multi-locus barcode would reduce the biases caused by using a single locus (Alberdi *et al.*, 2018). For example, Richardson *et al.*, (2015b) showed that using multilocus metabarcoding in pollen analysis improved the accuracy of quantitative measurements by decreasing the effect of primer bias. While quantitative analysis may not be possible in diet metabarcoding due to additional problems such as degradation and variation in digestion, this suggests that using multilocus barcodes would still help to reduce potential amplification bias. However, using a multi-locus barcode means there will be extra cost and time involved, and there is a problem of how to accurately combine results from more than one locus.

4.6.4 Bird diet

In the wider context, the result from this study could also be applied to analysing bird diet. Birds may also benefit cacao yield by pest control (Maas *et al.*, 2013). Therefore, these loci could be tested for their effectiveness in analysing bird samples. However, optimisation of the bird faeces DNA extraction protocol is required, as it has had limited success so far (A. J. Welch, Personal Communication 25/10/2018).

5 Conclusions

This study has compared the benefits and limitations of three metabarcoding loci for analysing the diet of insectivorous and frugivorous bats in a Cameroon cacao agroforest. The most suitable locus found was trnL, with c and h primers (Taberlet *et al.*, 2007), due to its relatively high taxonomic resolution and amplification success. Database coverage is potentially a major limiting factor to gaining high taxonomic resolution. Given broader database coverage, it is possible that ITS1 could provide higher taxonomic resolution than trnL. While the aim of this study was primarily optimisation of metabarcoding loci, through this some useful insights into the bat diet have been gained. In particular, this study has highlighted some common families in the bats' diet, shown differences and similarities between insectivore and frugivore diet, and shown a tentative link between insectivorous bats and cacao. Using a combination of loci is likely to increase taxonomic coverage and resolution. Overall, these new loci provide a starting point for further dietary analyses, in order to fully understand how farmers can manage shade tree species to maintain biodiversity, whilst also benefitting from bat pest control and other ecosystem services.

Acknowledgements

I would like to thank my supervisor Andreanna Welch for her help, support and patience throughout this project. Also, thank you to Luke Powell and the field team in Cameroon, and lastly to Rus Hoelzel and all the students in my lab.

REFERENCES

Adl, S. M., Bass, D., Lane, C. E., Lukeš, J., Schoch, C. L., Smirnov, A., Agatha, S., Berney, C., Brown, M. W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., del Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V. *et al.* (2019) 'Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes', *Journal of Eukaryotic Microbiology*, 66(1), pp. 4–119. doi: 10.1111/jeu.12691.

Ahenkorah, Y., Akrofi, G. S. and Adri, A. K. (1974) 'The end of the first cocoa shade and manurial experiment at the Cocoa Research Institute of Ghana', *Journal of Horticultural Science*, 49, pp. 43–51. doi: 10.1080/00221589.1974.11514550.

Alberdi, A., Aizpurua, O., Bohmann, K., Gopalakrishnan, S., Lynggaard, C., Nielsen, M. and Gilbert, M. T. P. (2019) 'Promises and pitfalls of using high-throughput sequencing for diet analysis', *Molecular Ecology Resources*, 19(2), pp. 327–348. doi: 10.1111/1755-0998.12960.

Alberdi, A., Aizpurua, O., Gilbert, M. T. P. and Bohmann, K. (2018) 'Scrutinizing key steps for reliable metabarcoding of environmental samples', *Methods in Ecology and Evolution*, 9, pp. 134–147. doi: 10.1111/2041-210X.12849.

Alvarez, I. and Wendel, J. F. (2003) 'Ribosomal ITS sequences and plant phylogenetic inference', *Molecular Phylogenetics and Evolution*, 29, pp. 417–434. doi: 10.1016/S1055-7903(03)00208-2.

Ambrose, S. H. and DeNiro, M. J. (1986) 'The isotopic ecology of East African mammals', *Oecologia*, 69, pp. 395–406. doi: 10.1007/BF00377062.

Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-madsen, H., Kjær, K. H., Orlando, L., Gilbert, M. T. P. and Willerslev, E. (2012) 'Meta-barcoding of " dirt " DNA from soil reflects vertebrate biodiversity', *Molecular Ecology*, 21(8), pp. 1966–1979. doi: 10.1111/j.1365-294X.2011.05261.x.

Arulandhu, A. J., Staats, M., Hagelaar, R., Voorhuijzen, M. M., Prins, T. W., Scholtens, I., Costessi, A., Duijsings, D., Gaspar, B., Teresa, M., Crespo, B., Holst-jensen, A., Birck, M., Burns, M., Haynes, E., Hochegger, R., Klingl, A., Lundberg, L., Natale, C. *et al.* (2017) 'Development and validation of a multi-locus DNA metabarcoding method to identify endangered species in complex samples', *GigaScience*, 6, pp. 1–18. doi: 10.1093/gigascience/gix080.

Aziz, S. A., Clements, G. R., Peng, L. Y., Campos-arceiz, A., Mcconkey, K. R., Forget, P. and Gan, H. M. (2017) 'Elucidating the diet of the island flying fox (Pteropus hypomelanus) in Peninsular Malaysia through Illumina Next-Generation Sequencing', *PeerJ*, 5, pp. 1–24. doi: 10.7717/peerj.3176.

Baamrane, M. A. A., Shehzad, W., Ouhammou, A., Abbad, A., Naimi, M., Coissac, E., Taberlet, P. and Znari, M. (2012) 'Assessment of the Food Habits of the Moroccan Dorcas Gazelle in M'Sabih Talaa, West Central Morocco, Using the trnL Approach', *PLoS ONE*, 7(4): e35643. doi: 10.1371/journal.pone.0035643.

Van Bael, S. A., Bichier, P. and Greenberg, R. (2007) 'Bird Predation on Insects Reduces Damage to the Foliage of Cocoa Trees (*Theobroma cacao*) in Western Panama', *Journal of Tropical Ecology*, 23(6), pp. 715–719. doi: 10.1017/S0266467407004440.

Bailey, C. D., Carr, T. G., Harris, S. A. and Hughes, C. E. (2003) 'Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes', *Molecular Phylogenetics and Evolution*, 29, pp. 435–455. doi: 10.1016/j.ympev.2003.08.021.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S. and Donoghue, M. J. (1995) 'The its Region of Nuclear Ribosomal DNA: A Valuable Source of Evidence on Angiosperm Phylogeny', *Annals of the Missouri Botanical Garden*, 82(2), pp. 247–277. doi: 10.2307/2399880.

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

De Beenhouwer, M., Aerts, R. and Honnay, O. (2013) 'A global meta-analysis of the biodiversity and ecosystem service benefits of coffee and cacao agroforestry', *Agriculture, Ecosystems and Environment*., 175, pp. 1–7. doi: 10.1016/j.agee.2013.05.003.

Bell, K. L., de Vere, N., Keller, A., Richardson, R. T., Gous, A., Burgess, K. S. and Brosi, B. J. (2016) 'Pollen DNA barcoding: current applications and future prospects', *Genome*, 59(9), pp. 629–640. doi: 10.1139/gen-2015-0200.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P. and Kauserud, H. (2010) 'ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases', *BMC Microbiology*, 10(189), pp. 1–9. doi: 10.1186/1471-2180-10-189.

Besnard, G., Christin, P.-A., Malé, P. G., Lhuillier, E., Lauzeral, C., Coissac, E. and Vorontsova, M. S. (2014) 'From museums to genomics: old herbarium specimens shed light on a C3 to C4 transition', *Journal of Experimental Biology*, 65(22), pp. 6711–6721. doi: 10.1093/jxb/eru395.

Bisseleua, D. H. B., Missoup, A. D. and Vidal, S. (2009) 'Biodiversity Conservation, Ecosystem Functioning, and Economic Incentives under Cocoa Agroforestry Intensification', *Conservation Biology*, 23(5), pp. 1176–1184. doi: 10.1111/j.1523-1739.2009.01220.x.

Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S., Christmas, M. and Creer, S. (2018) 'Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples', *Molecular Ecology Resources*, 18(5), pp. 1020–1034. doi: 10.1111/1755-0998.12888.

Blanckenhorn, W. U., Rohner, P. T., Bernasconi, M. V., Haugstetter, J. and Buser, A. (2016) 'Is qualitative and quantitative metabarcoding of dung fauna biodiversity feasible?', *Environmental Toxicology and Chemistry*, 35(8), pp. 1970–1977. doi:

10.1002/etc.3275.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R. and Abebe, E. (2005) 'Defining operational taxonomic units using DNA barcode data', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), pp. 1935–1943. doi: 10.1098/rstb.2005.1725.

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W. and Bruyn, M. De (2014) 'Environmental DNA for wildlife biology and biodiversity monitoring', *Trends in Ecology and Evolution*, 29(6), pp. 358–367. doi: 10.1016/j.tree.2014.04.003 358.

Bohmann, K., Monadjem, A., Noer, C. L., Rasmussen, M., Zeale, M. R. K., Clare, E., Jones, G., Willerslev, E. and Gilbert, M. T. P. (2011) 'Molecular Diet Analysis of Two African Free-Tailed Bats (Molossidae) Using High Throughput Sequencing', *PLoS ONE*, 6(6): e21441. doi: 10.1371/journal.pone.0021441.

Bremer, K. (1985) 'Summary of Green Plant Phylogeny and Classification', *Cladistics*, 1(4), pp. 369–385. doi: 10.1111/j.1096-0031.1985.tb00434.x.

Brennan, A. C., Méndez-Vigo, B., Haddioui, A., Martínez-Zapater, J. M., Picó, F. X. and Alonso-Blanco, C. (2014) 'The genetic structure of *Arabidopsis thaliana* in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe', *BMC Plant Biology*, 14(17), pp. 1–14. doi: 10.1186/1471-2229-14-17.

Brown, E. A., Chain, F. J. J., Crease, T. J., MacIsaac, H. J. and Cristescu, M. E. (2015) 'Divergence thresholds and divergent biodiversity estimates: Can metabarcoding reliably describe zooplankton communities?', *Ecology and Evolution*, 5(11), pp. 2234–2251. doi: 10.1002/ece3.1485.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M. *et al.* (2010) 'QIIME allows analysis of high-throughput community sequencing data', *Nature Methods,* 7, pp. 335-336. doi: 10.1038/nmeth.f.303.

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S. and Naeem, S. (2012) 'Biodiversity loss and its impact on humanity', *Nature*, 486, pp. 59–68. doi: 10.1038/nature11148.

CBOL Plant Working Group, Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M. W., Cowan, R. S., Erickson, D. L., Fazekasg, A. J., Graham, S. W., James, K. E., Kim, K.-J., Kress, W. J., Schneideri, H., van AlphenStahle, J., Barrett, S. C. H., van den Berg, C. *et al.* (2009) 'A DNA barcode for land plants', *PNAS*, 106(31), pp. 12794–12797. doi: 10.1073/pnas.0905845106.

Cerling, T. E., Harris, J. M. and Passey, B. H. (2003) 'Diets of East African Bovidae Based on Stable Isotope Analysis', *Journal of Mammalogy*, 84(2), pp. 456–470. doi:

10.1644/1545-1542(2003)084<0456:doeabb>2.0.co;2.

Chase, M. W. and Fay, M. F. (2009) 'Barcoding of Plants and Fungi', *Science*, 325(5941), pp. 682–683. doi: 10.1126/science.1176906.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L. and Zhu, Y. (2010) 'Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species', *PLoS ONE* 5(1): e8613. doi: 10.1371/journal.pone.0008613.

Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y. and Zhou, S. (2016) 'Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity', *Molecular Ecology Resources*, 16(1), pp. 138–149. doi: 10.1111/1755-0998.12438.

China Plant BOL Group, Li, D.-Z., Gao, L.-M., Li, H.-T., Wang, H., Ge, X.-J., Liu, J.-Q., Chen, Z.-D., Zhou, S.-L., Chen, S.-L., Yang, J.-B., Fu, C.-X., Zeng, C.-X., Yan, H.-F., Zhu, Y.-J., Sun, Y.-S., Chen, S.-Y., Zhao, L., Wang, K. *et al.* (2011) 'Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants', *PNAS*, 108(49), pp. 19641–19646. doi: 10.1073/pnas.1104551108.

Clare, E. L. (2014a) 'Molecular detection of trophic interactions: Emerging trends, distinct advantages, significant considerations and conservation applications', *Evolutionary Applications*, 7(9), pp. 1144–1157. doi: 10.1111/eva.12225.

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

Clare, E. L., Fraser, E. E., Braid, H. E., Fenton, M. B. and Hebert, P. D. N. (2009) 'Species on the menu of a generalist predator , the eastern red bat (Lasiurus borealis): using a molecular approach to detect arthropod prey', *Molecular Ecology*, 18(11), pp. 2532–2542. doi: 10.1111/j.1365-294X.2009.04184.x.

Clare, E. L., Symondson, W. O. C., Broders, H., Fabianek, F., Fraser, E. E., Mackenzie, A., Boughen, A., Hamilton, R., Willis, C. K. R., Martinez-Nuñez, F., Menzies, A. K., Norquay, K. J. O., Brigham, M., Poissant, J., Rintoul, J., Barclay, R. M. R. and Reimer, J. P. (2014b) 'The diet of Myotis lucifugus across Canada: Assessing foraging quality and diet variability', *Molecular Ecology*, 23(15), pp. 3618–3632. doi: 10.1111/mec.12542.

Clarke, L. J., Soubrier, J. and Weyrich, L. S. (2014) 'Environmental metabarcodes for insects : in silico PCR reveals potential for taxonomic bias', *Molecular Ecology Resources,* 14(6), pp. 1160–1170. doi: 10.1111/1755-0998.12265.

Codron, D., Codron, J., Lee-Thorp, J. A., Sponheimer, M., de Ruiter, D., Sealy, J., Grant, R. and Fourie, N. (2007) 'Diets of savanna ungulates from stable carbon isotope composition of faeces', *Journal of Zoology*, 273, pp. 21–29. doi: 10.1111/j.1469-7998.2007.00292.x.

Coissac, E., Hollingsworth, P. M., Lavergne, S. and Taberlet, P. (2016) 'From barcodes to genomes: Extending the concept of DNA barcoding', *Molecular Ecology*, 25(7), pp. 1423–1428. doi: 10.1111/mec.13549.

Cowan, R. S., Chase, M. W., Kress, W. J. and Savolainen, V. (2006) '300,000 Species to Identify: Problems, Progress, and Prospects in DNA Barcoding of Land Plants', *Taxon*, 55(3), pp. 611–616. doi: 10.2307/25065638.

Cullings, K. W. and Vogler, D. R. (1998) 'A 5.8S nuclear ribosomal RNA gene sequence database: applications to ecology and evolution', *Molecular Ecology*, 7(7), pp. 919–923. doi: 10.1046/j.1365-294x.1998.00409.x.

Deagle, B. E., Eveson, J. P. and Jarman, S. N. (2006) 'Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces', *Frontiers in Zoology*, 3(11), pp. 1–10. doi: 10.1186/1742-9994-3-11.

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(9), pp. 1–4. doi: 10.1098/rsbl.2014.0562.

Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarke, L. J., Vesterinen, E. J., Clare, E. L., Kartzinel, T. R. and Eveson, J. P. (2019) 'Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data?', *Molecular Ecology*, 28(2), pp. 391–406. doi: 10.1111/mec.14734.

Deagle, B. E. and Tollit, D. J. (2007) 'Quantitative analysis of prey DNA in pinniped faeces : potential to estimate diet composition ?', *Conservation Genetics,* 8(3) pp. 743–747. doi: 10.1007/s10592-006-9197-7.

Decher, J. and Fahr, J. (2005) 'Hipposideros cyclops', *Mammalian Species*, 763, pp. 1–7. doi: 10.2307/3504339.

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

DeSalle, R., Egan, M. G. and Siddall, M. (2005) 'The unholy trinity: Taxonomy, species delimitation and DNA barcoding', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), pp. 1905–1916. doi: 10.1098/rstb.2005.1722.

Dong, W., Liu, J., Yu, J., Wang, L. and Zhou, S. (2012) 'Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding', *PLoS ONE*, 7(4): e35071, doi: 10.1371/journal.pone.0035071.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. and Knight, R. (2011) 'UCHIME improves sensitivity and speed of chimera detection', *Bioinformatics*, 27(16), pp. 2194–2200. doi: 10.1093/bioinformatics/btr381.

Elbrecht, V. and Leese, F. (2015) 'Can DNA-Based Ecosystem Assessments Quantify

Species Abundance? Testing Primer Bias and Biomass — Sequence Relationships with an Innovative Metabarcoding Protocol', *PLoS ONE*, 10(7): e0130324. doi: 10.1371/journal.pone.0130324.

Erickson, D. L., Reed, E., Ramachandran, P., Bourg, N. A., Mcshea, W. J. and Ottesen, A. (2017) 'Interactions between White-Tailed Deer and Invasive Plants in North American Forests Reconstructing a herbivore's diet using a novel rbcL DNA mini-barcode for plants', *AoB Plants,* 9: plx015 doi: 10.1093/aobpla/plx015.

Evans, D. M., Kitson, J. J. N., Lunt, D. H., Straw, N. A. and Michael, J. O. (2016) 'Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems', *Functional Ecology*, 30, pp. 1904–1916. doi: 10.1111/1365-2435.12659.

Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., Percy, D. M., Hajibabaei, M. and Barrett, S. C. H. (2008) 'Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well', *PLoS ONE*, 3(7): e2802. doi: 10.1371/journal.pone.0002802.

Fischer, J., Abson, D. J., Butsic, V., Chappell, M. J., Ekroos, J., Hanspach, J., Kuemmerle, T., Smith, H. G. and Wehrden, H. Von (2014) 'Land Sparing Versus Land Sharing: Moving Forward', *Conservation Letters*, 7(3), pp. 149–157. doi: 10.1111/conl.12084.

Fischer, J., Brosi, B., Daily, G. C., Ehrlich, P. R., Goldman, R., Goldstein, J., Lindenmayer, D. B., Manning, A. D., Mooney, H. A., Pejchar, L., Ranganathan, J. and Tallis, H. (2008) 'Should agricultural policies encourage land Sparing or Wildlife-Friendly Farming?', *Frontiers in Ecology and the Environment*, 6(7), pp. 380–385. doi: 10.1890/070019.

Floyd, R., Abebe, E., Papert, A. and Blaxter, M. (2002) 'Molecular barcodes for soil nematode identification', *Molecular Ecology*, 11(4), pp. 839–850. doi: 10.1046/j.1365-294X.2002.01485.x.

Flynn, J. M., Brown, E. A., Chain, F. J. J., MacIsaac, H. J. and Cristescu, M. E. (2015) 'Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods', *Ecology and Evolution*, 5(11), pp. 2252–2266. doi: 10.1002/ece3.1497.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N. and Snyder, P. K. (2005) 'Global Consequences of Land Use', *Science*, 309(5734), pp. 570–575. doi: 10.1126/science.1111772.

Forsyth, D. M. and Davis, N. E. (2011) 'Diets of Non-Native Deer in Australia Estimated by Macroscopic Versus Microhistological Rumen Analysis', *The Journal of Wildlife Management*, 75(6), pp. 1488–1497. doi: 10.1002/jwmg.179.

Franzten, M. A. J., Silk, J. B., Ferguson, J. W. H., Wayne, R. K. and Kohn, M. H. (1998) 'Empirical evaluation of preservation methods for faecal DNA', *Molecular Ecology*,

7(10), pp. 1423–1428. doi: 10.1046/j.1365-294x.1998.00449.x.

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

Gardner, T. A., Barlow, J., Chazdon, R., Ewers, R. M., Harvey, C. A., Peres, C. A. and Sodhi, N. S. (2009) 'Prospects for tropical forest biodiversity in a human-modified world', *Ecology Letters*, 12(6), pp. 561–582. doi: 10.1111/j.1461-0248.2009.01294.x.

Garnick, S., Barboza, P. S. and Walker, J. W. (2018) 'Assessment of Animal-Based Methods Used for Estimating and Monitoring Rangeland Herbivore Diet Composition', *Rangeland Ecology & Management*, 71(4), pp. 449–457. doi: 10.1016/j.rama.2018.03.003.

Gebremedhin, B., Flagstad, O., Bekele, A., Chala, D., Bakkestuen, V., Boessenkool, S., Popp, M., Gussarova, G., Nemomissa, S., Brochmann, C., Stenseth, N. C. and Epp, L. S. (2016) 'DNA Metabarcoding Reveals Diet Overlap between the Endangered Walia Ibex and Domestic Goats - Implications for Conservation', *PLoS ONE*, 11(7): e0159133. doi: 10.1371/journal.pone.0159133.

Gernandt, D. S., Liston, A. and Pin, D. (2001) 'Variation in the nrDNA ITS of Pinus Subsection Cembroides : Implications for Molecular Systematic Studies of Pine Species Complexes', *Molecular Phylogenetics and Evolution,* 21(3), pp. 449–467. doi: 10.1006/mpev.2001.1026.

Gockowski, J. J. and Dury, S. (1999) 'The economics of cocoa-fruit agroforests in southern Cameroon', *International Symposium on Multi-Strata Agroforestry Systems with Perennial Crops*.

Gonzalez, M. A., Baraloto, C., Engel, J., Mori, S. A., Petronelli, P., Riera, B., Roger, A., Thebaud, C., Chave, J. (2009) 'Identification of Amazonian Trees with DNA Barcodes', *PLoS ONE,* 4(10): e7483. doi: 10.1371/journal.pone.0007483.

Groom, C., White, N. E., Mitchell, N., Roberts, J. D. and Mawson, P. (2017) 'Assessing the spatial ecology and resource use of a mobile and endangered species in an urbanized landscape using satellite telemetry and DNA faecal metabarcoding', *Ibis*, 159, pp. 390–405. doi: 10.1111/ibi.12442.

Hajibabaei, M., Smith, M. A., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B. and Hebert, P. D. N. (2006) 'A minimalist barcode can identify a specimen whose DNA is degraded', *Molecular Ecology Notes*, 6(4), pp. 959–964. doi: 10.1111/j.1471-8286.2006.01470.x.

Harvey, C. A., Gonzalez, J. and Somarriba, E. (2006) 'Dung beetle and terrestrial mammal diversity in forests, indigenous agroforestry systems and plantain monocultures in Talamanca, Costa Rica', *Biodiversity and Conservation*, 15(2), pp. 555–585. doi: 10.1007/s10531-005-2088-2.

Hebert, P. D. N., Cywinska, A., Ball, S. L. and DeWaard, J. R. (2003) 'Biological

Identifications through DNA barcodes', *Proceedings of the Royal Society B: Biological Sciences*, 270, pp. 313–321. doi: 10.1098/rspb.2002.2218.

Hebert, P. D. N. and Gregory, T. R. (2005) 'The Promise of DNA Barcoding for Taxonomy', *Systematic Biology*, 54(5), pp. 852–859. doi: 10.1080/10635150500354886.

Hibert, F., Taberlet, P., Chave, J., Scotti-saintagne, C., Sabatier, D. and Richard-Hansen, C. (2013) 'Unveiling the Diet of Elusive Rainforest Herbivores in Next Generation Sequencing Era? The Tapir as a Case Study', *PLoS ONE*, 8(4): e60799, pp. 1–11. doi: 10.1371/journal.pone.0060799.

Hoffman, M. H. (2002) 'Biogeography of *Arabidopsis thaliana* (L .) Heynh . (Brassicaceae)', *Journal of Biogeography*, 29(1), pp. 125–134. doi: 10.1046/j.1365-2699.2002.00647.x.

Hollingsworth, M. L., Clark, A. A., Forrest, L. L., Richardson, J., Pennington, R. T., Long, D. G., Cowan, R., Chase, M. W., Gaudeul, M. and Hollingswoth, P. M. (2009) 'Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants', *Molecular Ecology Resources*, 9(2), pp. 439–457. doi: 10.1111/j.1755-0998.2008.02439.x.

Hollingsworth, P. M., Graham, S. W. and Little, D. P. (2011a) 'Choosing and using a plant DNA barcode', *PLoS ONE*, 6(5): e19254. doi: 10.1371/journal.pone.0019254.

Hollingsworth, P. M. (2011b) 'Refining the DNA barcode for land plants', *PNAS,* 108(49), pp. 19451–19452. doi: 10.1073/pnas.1116812108.

Jagoret, P., Michel-Dounias, I. and Malézieux, E. (2011) 'Long-term dynamics of cocoa agroforests: A case study in central Cameroon', *Agroforestry Systems*, 81(3), pp. 267–278. doi: 10.1007/s10457-010-9368-x.

Jarrett, C., Powell, L. L., McDevitt, H., Helma, B. and Welch, A. J. (in prep) 'Bitter fruits of hard labour: Diet metabarcoding and telemetry reveal that urban songbirds travel further for lower-quality food.'

Jeanson, M. L., Labat, J.-N. and Little, D. P. (2011) 'DNA barcoding: a new tool for palm taxonomists?', *Annals of Botany*, 108, pp. 1445–1451. doi: 10.1093/aob/mcr158.

Jedlicka, J. A., Vo, A. E. and Almeida, R. P. P. (2017) 'Molecular scatology and high-throughput sequencing reveal predominately herbivorous insects in the diets of adult and nestling Western Bluebirds (*Sialia mexicana*) in California vineyards', *The Auk: Ornithological Advances*, 134, pp. 116–127. doi: 10.1642/AUK-16-103.1.

Jordan, F. (2009) 'Keystone species and food webs', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, pp. 1733–1741. doi: 10.1098/rstb.2008.0335.

Joshi, N. A. and Fass, J. N. (2011) 'Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software].'

Kalka, M. and Kalko, E. K. V (2006) 'Gleaning Bats as Underestimated Predators of

Herbivorous Insects: Diet of *Micronycteris microtis* (Phyllostomidae) in Panama', *Journal of Tropical Ecology*, 22(1), pp. 1–10. doi: 10.1017/S0266467405002920.

Kartzinel, T. R., Chen, P. A., Coverdale, T. C., Erickson, D. L., Kress, W. J., Kuzmina, M. L., Rubenstein, D. I., Wang, W. and Pringle, R. M. (2015) 'DNA metabarcoding illuminates dietary niche partitioning by African large herbivores', *PNAS*, 112(26), pp. 7873–8154. doi: 10.1073/pnas.1503283112.

Katoh, K., Rozewicki, J. and Yamada, K. D. (2017) 'MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization', *Briefings in Bioinformatics*, pp. 1–7. doi: 10.1093/bib/bbx108.

Kesanakurti, P. R., Fazekas, A. J., Burgess, K. S., Percy, D. M., Newmaster, S. G., Graham, S. W., Barrett, S. C. H., Hajibabaei, M. and Husband, B. C. (2011) 'Spatial patterns of plant diversity below-ground as revealed by DNA barcoding', *Molecular Ecology*, 20(6), pp. 1289–1302. doi: 10.1111/j.1365-294X.2010.04989.x.

Kleynhans, E. J., Jolles, A. E., Bos, M. R. E., Olff, H. and Boutin, S. (2011) 'Resource partitioning along multiple niche dimensions in differently sized African savanna grazers', *Oikos*, 120(4), pp. 591–600. doi: 10.1111/j.1600-0706.2010.18712.x.

Kopylova, E., Navas-Molina, J. A., Mercier, C., Xu, Z. Z., Mahe, F., He, Y., Zhou, H.-W., Rognes, T., Caporaso, J. G. and Knight, R. (2016) 'Open-Source Sequence Clustering Methods Improve the State Of the Art', *mSystems*, 1(1), pp. 1–16. doi: 10.1128/mSystems.00003-15.Editor.

Korine, C., Izhaki, I. and Arad, Z. (1999) 'Is the Egyptian fruit-bat *Rousettus aegyptiacus* a pest in Israel? An analysis of the bat's diet and implications for its conservation', *Biological Conservation*, 88, pp. 301–306. doi: 10.1016/S0006-3207(98)00126-8.

Kress, W. J. and Erickson, D. L. (2007) 'A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region', *PLoS ONE*, 2(6): e508. doi: 10.1371/journal.pone.0000508.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. and Janzen, D. H. (2005) 'Use of DNA barcodes to identify flowering plants', *PNAS*, 102(23), pp. 8369–8374. doi: 10.1073/pnas.0503123102.

Kunin, V., Engelbrektson, A., Ochman, H. and Hugenholtz, P. (2010) 'Wrinkles in the rare biosphere : pyrosequencing errors can lead to artificial inflation of diversity estimates', *Environmental Microbiology*, 12(1), pp. 118–123. doi: 10.1111/j.1462-2920.2009.02051.x.

Kwiecinski, G. G. and Griffiths, T. A. (1999) 'Rousettus egyptiacus', *Mammalian Species*, 611, pp. 1–9. doi: 10.2307/3504411.

Leray, M. and Knowlton, N. (2015) 'DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity', *PNAS*, 112(7), pp. 2076–2081. doi: 10.1073/pnas.1424997112.

Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y. and Chen, S. (2015) 'Plant DNA barcoding: from gene to genome', *Biological reviews of the Cambridge Philosophical Society*, 90(1), pp. 157–166. doi: 10.1111/brv.12104.

Liu, J., Milne, R. I., Möller, M., Zhu, G. F., Ye, L. J., Luo, Y. H., Yang, J. B., Wambulwa, M. C., Wang, C. N., Li, D. Z. and Gao, L. M. (2018) 'Integrating a comprehensive DNA barcode reference library with a global map of yews (Taxus L.) for forensic identification', *Molecular Ecology Resources*, 18(5), pp. 1–17. doi: 10.1111/1755-0998.12903.

Lobo, J., Shokralla, S., Costa, M. H., Hajibabaei, M. and Costa, F. O. (2017) 'DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities', *Scientific Reports*, 7, pp. 1–13. doi: 10.1038/s41598-017-15823-6.

Lopes, C. M., de Barba, M., Boyer, F., Mercier, C., da Silva Filho, P. J. A., Heidtmann, L. M., Galiano, D., Kubiak, B. B., Langone, P., Garcias, F. M., Gielly, L., Coissac, E., de Freitas, T. R. O. and Taberlet, P. (2015) 'DNA metabarcoding diet analysis for species with parapatric vs sympatric distribution: a case study on subterranean rodents', *Heredity*, 114(5), pp. 525–536. doi: 10.1038/hdy.2014.109.

Luo, A., Zhang, A., Ho, S. Y. W., Xu, W., Zhang, Y., Shi, W., Cameron, S. L. and Zhu, C. (2011) 'Potential efficacy of mitochondrial genes for animal DNA barcoding: A case study using eutherian mammals', *BMC Genomics*, 12(84), pp. 1–13. doi: 10.1186/1471-2164-12-84.

Maas, B., Clough, Y. and Tscharntke, T. (2013) 'Bats and birds increase crop yield in tropical agroforestry landscapes', *Ecology Letters*, 16(12), pp. 1480–1487. doi: 10.1111/ele.12194.

Maas, B., Tscharntke, T., Saleh, S., Putra, D. D. and Clough, Y. (2015) 'Avian species identity drives predation success in tropical cacao agroforestry', *Journal of Applied Ecology*, 52, pp. 735–743. doi: 10.1111/1365-2664.12409.

Mata, V. A., Rebelo, H., Amorim, F., McCracken, G. F., Jarman, S. and Beja, P. (2018) 'How much is enough? Effects of technical and biological replication on metabarcoding dietary analysis', *Molecular Ecology*, 28(2), pp. 165–175. doi: 10.1111/mec.14779.

Mbolo, M. M. A., Zekeng, J. C., Mala, W. A., Fobane, J. L., Chimi, C. D., Tangboulou, N., Nyako, C. M., Menyene, L. F. E. and Tamanjong, Y. V. (2016) 'The role of cocoa agroforestry systems in conserving forest tree diversity in the Central region of Cameroon', *Agroforestry Systems*, 90(4), pp. 577–590. doi: 10.1007/s10457-016-9945-8.

Mercier, C., Boyer, F., Bonin, A. and Coissac, E. (2013) *SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences*, *Programs and Abstracts of the SeqBio 2013 workshop. Abstract*. doi: 10.1109/RCIS.2013.6577673.

Meusnier, I., Singer, G. A. C., Landry, J. F., Hickey, D. A., Hebert, P. D. N. and Hajibabaei, M. (2008) 'A universal DNA mini-barcode for biodiversity analysis', *BMC Genomics*, 9(214), pp. 1–4. doi: 10.1186/1471-2164-9-214.

Muellner, A. N., Schaefer, H. and Lahaye, R. (2011) 'Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae)', *Molecular Ecology Resources*, 11(3), pp. 450–460. doi: 10.1111/j.1755-0998.2011.02984.x.

Murphy, M. A., Waits, L. P., Kendall, K. C., Wasser, S. K., Higbee, J. A. and Bogden, R. (2002) 'An evaluation of long-term preservation methods for brown bear (*Ursus arctos*) faecal DNA samples', *Conservation Genetics*, 3, pp. 435–440. doi: 10.1023/A:1020503330767.

Murray, D. C., Pearson, S. G., Fullagar, R., Chase, B. M., Houston, J., Atchison, J., White, N. E., Bellgard, M. I., Clarke, E., Macphail, M., Gilbert, M. T. P., Haile, J. and Bunce, M. (2012) 'High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens', *Quarternary Science Reviews*, 58, pp. 135–145. doi: 10.1016/j.quascirev.2012.10.021.

Naeem, S., Duffy, J. E. and Zavaleta, E. (2012) 'The Functions of Biological Diversity in an Age of Extinction', *Science*, 336, pp. 1401–1407. doi: 10.1126/science.1215855.

Newmaster, S. G., Fazekas, A. J., Steeves, R. A. D. and Janovec, J. (2008) 'Testing candidate plant barcode regions in the Myristicaceae', *Molecular Ecology Resources*, 8(3), pp. 480–490. doi: 10.1111/j.1471-8286.2007.02002.x.

Nichols, R. V, Åkesson, M. and Kjellander, P. (2016) 'Diet Assessment Based on Rumen Contents : A Comparison between DNA Metabarcoding and Macroscopy', *PLoS ONE*, 11(6): e0157977. doi: 10.1371/journal.pone.0157977.

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

Nikolenko, S. I., Korobeynikov, A. I. and Alekseyev, M. A. (2013) 'BayesHammer: Bayesian subclustering for error correction in single cell sequencing', *BMC Genomics*, 14(Suppl 1), pp. 1–11. doi: 10.1186/1471-2164-14-S1-S7.

Nkrumah, E. E., Vallo, P., Klose, S. M., Ripperger, S. P., Badu, E. K., Gloza-Rausch, F., Drosten, C., Kalko, E. K. V., Tschapka, M. and Oppong, S. K. (2016) 'Foraging Behavior and Habitat Selection of Noack's Round-Leaf Bat (*Hipposideros aff. ruber*) and Conservation Implications', *Tropical Conservation Science*, 9(4), pp. 1–11. doi: 10.1177/1940082916680428.

Nurul-ain, E., Rosli, H. and Kingston, T. (2017) 'Resource availability and roosting ecology shape reproductive phenology of rain forest insectivorous bats', *Biotropica*, 49(3), pp. 382–394. doi: 10.1111/btp.12430.

Omelchenko, D. O., Speranskaya, A. S., Ayginin, A. A., Khafizov, K., Krinitsina, A. A., Fedotova, A. V, Pozdyshev, D. V, Shtratnikova, V. Y., Kupriyanova, E. V, Shipulin, G. A. and Logacheva, M. D. (2019) 'Improved Protocols of ITS1-Based Metabarcoding and Their Application in the Analysis of Plant-Containing Products', *Genes*, 10(122), pp. 1–16. doi: 10.3390/genes10020122.

Owen-ashley, N. T. and Wilson, D. E. (1998) 'Micropteropus pusillus', *Mammalian Species*, 577, pp. 1–5. doi: 10.2307/0.577.1.

Parmentier, I., Duminil, J., Kuzmina, M., Philippe, M., Thomas, D. W., Kenfack, D., Chuyong, G. B., Cruaud, C. and Hardy, O. J. (2013) 'How Effective Are DNA Barcodes in the Identification of African Rainforest Trees?', *PLoS ONE*, 8(4): e54921. doi: 10.1371/journal.pone.0054921.

Pegard, A., Miquel, C., Valentini, A., Coissac, E., Bouvier, F., Francois, D., Taberlet, P., Engel, E. and Pompanon, F. (2009) 'Universal DNA-Based Methods for Assessing the Diet of Grazing Livestock and Wildlife from Feces', *Journal of Agricultural and Food Chemistry*, 57, pp. 5700–5706. doi: 10.1021/jf803680c.

Perrin, M. R. and Brereton-Stiles, R. (1999) 'Habitat use and feeding behaviour of the buffalo and the white rhinoceros in the Hluhluwe-Umfolozi Game Reserve', *South African Journal of Wildlife Research*, 29(3), pp. 72–80.

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

Pinol, J., Senar, M. A. and Symondson, W. O. C. (2019) 'The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative', *Molecular Ecology*, 28(2), pp. 407–419. doi: 10.1111/mec.14776.

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(8), pp. 1931–1950. doi: 10.1111/j.1365-294X.2011.05403.x.

Ratnasingham, S. and Hebert, P. D. N. (2007) 'BOLD: The Barcode of Life Data System', *Molecular Ecology Notes*, 7, pp. 355–364. doi: 10.1111/j.1471-8286.2006.01678.x.

Raye, G., Miquel, C., Coissac, E., Redjadj, C., Loison, A. and Taberlet, P. (2011) 'New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: chamois diet in autumn as a case study', *Ecological Research*, 26, pp. 265–276. doi: 10.1007/s11284-010-0780-5.

Rice, R. A. and Greenberg, R. (2000) 'Cacao Cultivation and the Conservation of Biological Diversity', *Ambio*, 29(3), pp. 167–173. doi: 10.1579/0044-7447-29.3.167.

Richardson, R. T., Lin, C.-H., Quijia, J. O., Riusech, N. S., Goodell, K. and Johnson, R. M. (2015b) 'Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding aproach', *Applications in Plant Sciences*, 3(11), pp. 1–9. doi: 10.3732/apps.1500043.

Richardson, R. T., Lin, C.-H., Sponsler, D. B., Quijia, J. O., Goodell, K. and Johnson, R. M. (2015a) 'Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem', *Applications in Plant Sciences*, 3(1), pp. 1–

6. doi: 10.3732/apps.1400066.

Rohland, N. and Reich, D. (2012) 'Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture', *Genome Research*, 22, pp. 939–946. doi: 10.1101/gr.128124.111.22.

Roy, S., Tyagi, A., Shukla, V., Kumar, A., Singh, U. M., Chaudhary, L. B., Datt, B., Bag, S. K., Singh, P. K., Nair, N. K., Husain, T. and Tuli, R. (2010) 'Universal Plant DNA Barcode Loci May Not Work in Complex Groups: A Case Study with Indian Berberis Species', *PLoS ONE*, 5(10): e13674. doi: 10.1371/journal.pone.0013674.

RStudio Team (2016) 'RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL'. Available at: http://www.rstudio.com/.

Salinas-Ramos, V. B., Montalvo, L. G. H., Leon-Regagnon, V., Arrizabalaga-Escudero, A. and Clare, E. L. (2015) 'Dietary overlap and seasonality in three species of mormoopid bats from a tropical dry forest', *Molecular Ecology*, 24(20), pp. 5296–5307. doi: 10.1111/mec.13386.

Sang, T., Crawford, D. J. and Stuessy, T. F. (1997) 'Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae)', *American Journal of Botany*, 84(9), pp. 1120–1136. doi: 10.2307/2446155.

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D. and Karsch-Mizrachi, I. (2019) 'GenBank', *Nucleic Acids Research*, 47, pp. D94–D99. doi: 10.1093/nar/gky989.

Schloss, P. D. (2016) 'Application of a Database-Independent Approach To Assess the Quality of Operational Taxonomic Unit Picking Methods', *mSystems*, 1(2), pp. 1–4. doi: 10.1128/mSystems.00027-16.Copyright.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W. and Fungal Barcoding Consortium (2012) 'Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi', *PNAS*, 109(16), pp. 6241–6246. doi: 10.1073/pnas.1117018109.

Sequencher version 5.4.6 (2017) 'DNA sequence analysis software'.

Shendure, J. and Ji, H. (2008) 'Next-generation DNA sequencing', *Nature Biotechnology*, 26(10), pp. 1135–1145. doi: 10.1038/nbt1486.

Sheppard, S. K., Bell, J., Sunderland, K. D., Fenlon, J., Skervin, D. and Symondson, W. O. C. (2005) 'Detection of secondary predation by PCR analyses of the gut contents of invertebrate generalist predators', *Molecular Ecology*, 14(14), pp. 4461–4468. doi: 10.1111/j.1365-294X.2005.02742.x.

Smyth, R. P., Schlub, T. E., Grimm, A., Venturi, V., Chopra, A., Mallal, S., Davenport, M. P. and Mak, J. (2010) 'Reducing chimera formation during PCR amplification to ensure accurate genotyping', *Gene*. Elsevier B.V., 469(1–2), pp. 45–51. doi: 10.1016/j.gene.2010.08.009.
Soininen, E. M., Valentini, A., Coissac, E., Miquel, C., Gielly, L., Brochmann, C.,

Brysting, A. K., Sønstebø, J. H., Ims, R. A., Yoccoz, N. G. and Taberlet, P. (2009) 'Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures', *Frontiers in Zoology*, 6(16), pp. 1–9. doi: 10.1186/1742-9994-6-16.

Soininen, E. M., Zinger, L., Gielly, L., Bellemain, E., Brathen, K. A., Brochmann, C., Epp, L. S., Gussarova, G., Hassel, K., Henden, J.-A., Killengreen, S. T., Rama, T., Stenøien, H. K., Yoccoz, N. G. and Ims, R. A. (2013) 'Shedding new light on the diet of Norwegian lemmings : DNA metabarcoding of stomach content', *Polar Biology*, 36, pp. 1069–1076. doi: 10.1007/s00300-013-1328-2.

Song, J., Shi, L., Li, D., Sun, Y., Niu, Y., Chen, Z., Luo, H., Pang, X., Sun, Z., Liu, C., Lv, A., Deng, Y., Larson-, Z., Wilkinson, M. and Chen, S. (2012) 'Extensive Pyrosequencing Reveals Frequent Intra-Genomic Variations of Internal Transcribed Spacer Regions of Nuclear Ribosomal DNA', *PLoS ONE*, 7(8): e43971. doi: 10.1371/journal.pone.0043971.

Sonwa, D. J., Nkongmeneck, B. A., Weise, S. F., Tchatat, M., Adesina, A. A. and Janssens, M. J. J. (2007) 'Diversity of plants in cocoa agroforests in the humid forest zone of Southern Cameroon', *Biodiversity and Conservation*, 16(8), pp. 2385–2400. doi: 10.1007/s10531-007-9187-1.

Sonwa, D. J., Weise, S. F., Nkongmeneck, B. A., Tchatat, M. and Janssens, M. J. J. (2017) 'Structure and composition of cocoa agroforests in the humid forest zone of Southern Cameroon', *Agroforestry Systems*, 91(3), pp. 451–470. doi: 10.1007/s10457-016-9942-y.

Starr, J. R., Naczi, R. F. C. and Chouinard, B. N. (2009) 'Plant DNA barcodes and species resolution in sedges (Carex, Cyperaceae)', *Molecular Ecology Resources*, 9(SUPPL. 1), pp. 151–163. doi: 10.1111/j.1755-0998.2009.02640.x.

Steffan-Dewenter, I., Kessler, M., Barkmann, J., Bos, M. M., Buchori, D., Erasmi, S., Faust, H., Gerold, G., Glenk, K., Gradstein, S. R., Guhardja, E., Harteveld, M., Hertel, D., Hohn, P., Kappash, M., Kohler, S., Leuschner, C., Maertens, M., Marggraf, R. *et al.* (2007) 'Tradeoffs between income, biodiversity, and ecosystem functioning during tropical rainforest conversion and agroforestry intensification', *PNAS*, 104(12), pp. 4973–4978. doi: 10.1073pnas.0608409104.

Stein, E. D., Martinez, M. C., Stiles, S., Miller, P. E. and Zakharov, E. V (2014) 'Is DNA Barcoding Actually Cheaper and Faster than Traditional Morphological Methods: Results from a Survey of Freshwater Bioassessment Efforts in the United States?', *PLoS ONE*, 9(4): e95525. pp. 1–10. doi: 10.1371/journal.pone.0095525.

Taberlet, P., Coissac, E., Hajibabaei, M. and Rieseberg, L. H. (2012a) 'Environmental DNA', *Molecular Ecology*, 21(8), pp. 1789–1793. doi: 10.1111/j.1365-294X.2012.05542.x.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev, E. (2012b) 'Towards next-generation biodiversity assessment using DNA metabarcoding', *Molecular Ecology*, 21(8), pp. 2045–2050. doi: 10.1111/j.1365-294X.2012.05470.x.
Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T.,

Corthier, G., Brochmann, C. and Willerslev, E. (2007) 'Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding', *Nucleic Acids Research*, 35(3). doi: 10.1093/nar/gkl938.

Tallamy, D. W. and Shropshire, K. J. (2009) 'Ranking Lepidopteran Use of Native Versus Introduced Plants', *Conservation Biology*, 23(4), pp. 941–947. doi: 10.1111/j.1523-1739.2009.01202.x.

Tan, S. L., Luo, Y. H., Hollingsworth, P. M., Burgess, K. S., Xu, K., Li, D. Z. and Gao, L. M. (2018) 'DNA barcoding herbaceous and woody plant species at a subalpine forest dynamics plot in Southwest China', *Ecology and Evolution*, 8(14), pp. 7195–7205. doi: 10.1002/ece3.4254.

Tate, J. A. and Simpson, B. A. (2003) 'Paraphyly of Tarasa (Malvaceae) and Diverse Origins of the Polyploid Species', *Systematic Biology*, 28(4), pp. 723–737. doi: 10.1043/02-64.1.

Thomas, D. W. (1984) 'Fruit Intake and Energy Budgets of Frugivorous Bats', *Physiological Zoology*, 57(4), pp. 457–467. doi: 10.1086/physzool.57.4.30163347.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. and Rozen, S. G. (2012) 'Primer3-new capabilities and interfaces', *Nucleic Acids Research*, 40(15), p. e115. doi: 10.1093/nar/gks596.

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

Valentini, A., Pompanon, F. and Taberlet, P. (2008) 'DNA barcoding for ecologists', *Trends in Ecology and Evolution*, 24(2), pp. 110–117. doi: 10.1016/j.tree.2008.09.011.

de Vere, N., Rich, T. C. G., Ford, C. R., Trinder, S. A., Long, C., Moore, C. W., Satterthwaite, D., Davies, H., Allainguillaume, J., Ronca, S., Tatarinova, T., Garbett, H., Walker, K. and Wilkinson, M. J. (2012) 'DNA barcoding the native flowering plants and conifers of Wales', *PLoS ONE*, 7(6): e37945, pp. 1–12. doi: 10.1371/journal.pone.0037945.

Vetrovsky, T. and Baldrian, P. (2013) 'The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses', *PLoS ONE*, 8(2): e57923. doi: 10.1371/journal.pone.0057923.

Waltert, M., Bobo, K. S., Sainge, N. M., Fermon, H. and Mühlenberg, M. (2005) 'From Forest to Farmland : Habitat Effects on Afrotropical Forest Bird Diversity', *Ecological Applications*, 15(4), pp. 1351–1366.

Wang, X., Liu, C., Huang, L., Bengtsson-palme, J., Chen, H., Zhang, J.-H., Cai, D. and Li, J.-Q. (2015) 'ITS1: a DNA barcode better than ITS2 in eukaryotes?', *Molecular Ecology Resources*, 15(3), pp. 573–586. doi: 10.1111/1755-0998.12325.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., Lorenzen, E. D., Alsos, I. G., Bellemain, E., Brysting, A. K. and Elven, R. (2014) 'Fifty thousand years of Arctic vegetation and megafaunal diet', *Nature,* 506, pp. 47-51, doi: 10.1038/nature12921.

Yang, Y., Zhan, A., Cao, L., Meng, F. and Xu, W. (2016) 'Selection of a marker gene to construct a reference library for wetland plants, and the application of metabarcoding to analyze the diet of wintering herbivorous waterbirds', *PeerJ*, 4, pp. 1–21. doi: 10.7717/peerj.2345.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P. and Chen, S. (2010) 'Use of ITS2 region as the universal DNA barcode for plants and animals', *PLoS ONE*, 5(10), p. e13102. doi: 10.1371/journal.pone.0013102.

Zapfack, L., Engwald, S., Sonke, B., Achoundong, G. and Madong, B. A. (2002) 'The impact of land conversion on plant biodiversity in the forest zone of Cameroon', *Biodiversity and Conservation*, 11, pp. 2047–2061. doi: 10.1023/A:1020861925294.

Zeale, M. R. K., Butlin, R. K., Barker, G. L. A., Lees, D. C. and Jones, G. (2010) 'Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces', *Molecular Ecology Resources*, 11(2), pp. 236–244. doi: 10.1111/j.1755-0998.2010.02920.x.

Zepeda-Mendoza, M. L., Bohmann, K., Carmona Baez, A. and Gilbert, M. T. P. (2016) 'DAMe: A toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses', *BMC Research Notes*. BioMed Central, 9(1), pp. 1–13. doi: 10.1186/s13104-016-2064-9.

Zhang, J., Kobert, K., Flouri, T. and Stamatakis, A. (2014) 'PEAR: A fast and accurate Illumina Paired-End reAd mergeR', *Bioinformatics*, 30(5), pp. 614–620. doi: 10.1093/bioinformatics/btt593.

Zhang, W., Wendel, J. F. and Clark, L. G. (1997) 'Bamboozled Again! Inadvertent Isolation of Fungal rDNA Sequences from Bamboos (Poaceae: Bambusoideae)', *Molecular Phylogenetics and Evolution*, 8(2), pp. 205–217. doi: 10.1006/mpev.1997.0422.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., Tang, M., Fu, R., Li, J. and Huang, Q. (2013) 'Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification', *GigaScience*, 2(4), pp. 1–12. doi: 10.1186/2047-217X-2-4.

APPENDICES

Appendix A: Bat species of the faeces samples used in analysis. The table also indicates the location and season of sample collection, and of which loci there is sequence data used in analysis

				Sequence data used		
Species	**Location**	**Season**	**Frugivore/Insectivore**	**rbcL**	**trnL**	**ITS1**
Hipposideros ruber	Konye	Dry	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Dry	Insectivore	Y	Y	Y
Hipposideros ruber	Ayos	Dry	Insectivore	Y	Y	Y
Hipposideros ruber	Bokito	Dry	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Dry	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Dry	Insectivore	Y		Y
Hipposideros ruber	Ayos	Dry	Insectivore	Y	Y	
Hipposideros ruber	Konye	Wet	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Wet	Insectivore	Y	Y	Y
Hipposideros ruber	Ayos	Wet	Insectivore	Y	Y	Y
Hipposideros ruber	Ayos	Wet	Insectivore	Y	Y	Y
Hipposideros ruber	Ayos	Wet	Insectivore	Y	Y	
Hipposideros ruber	Bokito	Wet	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Wet	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Wet	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Wet	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Wet	Insectivore	Y	Y	
Hipposideros ruber	Konye	Wet	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Wet	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Dry	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Dry	Insectivore	Y	Y	Y
Hipposideros ruber	Konye	Wet	Insectivore	Y	Y	Y
Rhinolophus alcyone	Konye	Dry	Insectivore	Y	Y	Y
Rhinolophus alcyone	Ayos	Dry	Insectivore	Y	Y	Y
Rhinolophus alcyone	Konye	Dry	Insectivore	Y		Y
Rhinolophus alcyone	Konye	Dry	Insectivore	Y	Y	Y
Rhinolophus alcyone	Konye	Dry	Insectivore	Y		Y
Rhinolophus alcyone	Konye	Dry	Insectivore	Y	Y	Y
Rhinolophus alcyone	Ayos	Wet	Insectivore	Y	Y	Y
Rhinolophus alcyone	Konye	Wet	Insectivore	Y	Y	
Rhinolophus alcyone	Ayos	Wet	Insectivore	Y	Y	Y

Rhinolophus alcyone	Ayos	Wet	Insectivore	Y	Y	Y
Rhinolophus alcyone	Ayos	Wet	Insectivore	Y	Y	Y
Rhinolophus alcyone	Ayos	Wet	Insectivore	Y	Y	Y
Rhinolophus alcyone	Konye	Wet	Insectivore	Y	Y	
Rhinolophus alcyone	Konye	Wet	Insectivore	Y	Y	Y
Rhinolophus alcyone	Konye	Dry	Insectivore	Y	Y	Y
Rhinolophus alcyone	Konye	Dry	Insectivore	Y	Y	Y
Rhinolophus alcyone	Konye	Dry	Insectivore	Y		Y
Rhinolophus alcyone	Konye	Dry	Insectivore	Y	Y	
Rhinolophus alcyone	Ayos	Dry	Insectivore	Y	Y	Y
Rhinolophus alcyone	Bokito	Wet	Insectivore	Y	Y	
Rhinolophus alcyone	Ayos	Wet	Insectivore	Y	Y	
Rhinolophus alcyone	Konye	Wet	Insectivore	Y	Y	Y
Rhinolophus alcyone	Ayos	Wet	Insectivore	Y	Y	
Rhinolophus alcyone	Ayos	Wet	Insectivore	Y	Y	
Rhinolophus alcyone	Ayos	Wet	Insectivore	Y	Y	
Rhinolophus alcyone	Konye	Wet	Insectivore	Y	Y	
Rhinolophus alcyone	Konye	Wet	Insectivore	Y	Y	
Hipposideros caffer	Ayos	Wet	Insectivore		Y	
Hipposideros caffer	Ayos	Wet	Insectivore		Y	Y
Hipposideros cyclops	Konye	Wet	Insectivore			Y
Hipposideros cyclops	Konye	Wet	Insectivore		Y	Y
Nycteris grandis	Konye	Dry	Insectivore		Y	Y
Nycteris hispida	Bokito	Wet	Insectivore			Y
Nycteris major	Bokito	Wet	Insectivore			Y
Nycteris sp.	Konye	Dry	Insectivore			Y
Casinycteris argynnis	Konye	Wet	Frugivore		Y	Y
Casinycteris argynnis	Konye	Wet	Frugivore		Y	Y
Casinycteris argynnis	Konye	Wet	Frugivore		Y	Y
Casinycteris argynnis	Konye	Wet	Frugivore		Y	Y
Casinycteris argynnis	Konye	Wet	Frugivore		Y	

Epomops franqueti	Konye	Wet	Frugivore		Y	Y
Epomops franqueti	Konye	Wet	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Konye	Wet	Frugivore		Y	Y
Epomops franqueti	Bokito	Wet	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Bokito	Dry	Frugivore		Y	Y
Epomops franqueti	Konye	Dry	Frugivore		Y	Y
Epomops franqueti	Konye	Dry	Frugivore		Y	Y
Epomops franqueti	Konye	Wet	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Bokito	Dry	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Konye	Dry	Frugivore		Y	
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Epomops franqueti	Ayos	Dry	Frugivore		Y	Y
Megaloglossus woermanni	Konye	Wet	Frugivore		Y	Y
Megaloglossus woermanni	Konye	Wet	Frugivore		Y	Y
Megaloglossus woermanni	Ayos	Dry	Frugivore		Y	Y
Micropteropus pusillus	Ayos	Dry	Frugivore		Y	Y
Micropteropus pusillus	Bokito	Dry	Frugivore		Y	Y
Micropteropus pusillus	Bokito	Dry	Frugivore		Y	Y
Micropteropus pusillus	Bokito	Dry	Frugivore		Y	Y
Micropteropus pusillus	Bokito	Dry	Frugivore		Y	Y
Micropteropus pusillus	Bokito	Dry	Frugivore		Y	Y
Rousettus aegyptiacus	Konye	Dry	Frugivore		Y	Y
Rousettus aegyptiacus	Konye	Dry	Frugivore		Y	Y
Rousettus aegyptiacus	Konye	Dry	Frugivore		Y	Y

Rousettus aegyptiacus	Konye	Dry	Frugivore		Y	Y

Appendix B: Summary of findings of the literature review carried out to identify candidate metabarcoding loci. Spaces are in the table due to some papers not providing all information

						Taxonomic resolution				
Publication	**Type of study**	**Locus**	**Length (bp)**	**Primers**	**Amplification success**	**Species**	**Genus**	**family**	**Advantages**	**Disadvantages**
Kress *et al.*, (2005)	Barcode loci comparison	ITS	560-709		88%				amplified in 2 smaller fragments (ITS1 and ITS2), high sequence divergence	
Kartzinel *et al.*, (2015)	Empirical diet metabarcoding	ITS		ITS1-F/ITS1Poa-R,ITS1Ast-R,ITS1Cyp-R Family specific primers					Short, conserved primer sites	
Yang *et al.*, (2016)	Barcode loci comparison	matK		matK-XF/matK-MALP	43%					Hard to amplify with available primers
Parmentier *et al.*, (2013)	Barcode loci comparison	matK		matK1RKimf/matK3FKim-r and matK390f/matK1326r	48%	75%				
Yang *et al.*, (2016)	Barcode loci comparison	rbcL		rbcLa-Fc/rbcLa-Rd	91%				Easy to amplify, sequence and align	Low resolution
Parmentier *et al.*, (2013)	Barcode loci comparison	rbcL		rbcLaF/rbcLaR and rbcL1F/rbcL724R	77%	72%	98%		easy to align	Not ID to species

Kress *et al.*, (2005)	Barcode loci comparison	rbcL	1428		95%					No species level discrimination
Yang *et al.*, (2016)	Barcode loci comparison	trnH-psbA		pasbA3_f/trnHf_05	71%					Low sequence quality
Kress *et al.*, (2005)	Barcode loci comparison	trnH-psbA	340-660		100%				High amplification success, good discrimination, high sequence divergence	
Parmentier *et al.*, (2013)	Barcode loci comparison	trnH-psbA		trnHf-05/psbA3f	71%	81%			good species identification	
Yang *et al.*, (2016)	Empirical diet metabarcoding	trnL		c/h	100%				High discriminating power, high sequence divergence	
Murray *et al.*, (2012)	Empirical aDNA metabarcoding	trnL		c/h	Failed					
Nichols *et al.,* (2016)	Empirical diet metabarcoding	trnL		g/h			90%			
Taberlet *et al.*, (2007)	Barcode loci comparison	trnL	254-767	c/d		67%	86%	100%	Conserved primers, robust	Low resolution
Murray *et al.*, (2012)	Empirical aDNA metabarcoding	trnL(UAA)-P6	90-120	g/h				28 families		
Kartzinel *et al.*, (2015)	Empirical diet metabarcoding	trnL(UAA)-P6		g/h		77%				

Taberlet *et al.*, (2007)	Barcode loci comparison	trnL(UAA)-P6	10-143	g/h		20%	41%	79%	Amplify degraded DNA, highly conserved primer regions	
Gebremedhin *et al.*, (2016)	Empirical diet metabarcoding	trnL(UAA)-P6				29%	67%	100%		
Soininen *et al.*, (2013)	Empirical diet metabarcoding	trnL(UAA)-P6		g/h and c/h						
Pegard *et al.*, (2009)	Empirical diet metabarcoding	trnL(UAA)-P6		g/h		24%	51%	91%		
Valentini *et al.,* (2008)	Empirical diet metabarcoding	trnL(UAA)-P6	20-85	g/h		50%				

Appendix C: Adaptation of protocol from QIAGEN DNeasy Plant Mini kit

1. Place the sample material (≤100 mg wet weight or ≤20 mg lyophilized tissue) into a 2ml safe-lock microcentrifuge tube, together with three 1/8-inch stainless steel ball bearings (Bearing Boys Ltd. 1/8-inch stainless steel 316 ball bearings, pack of 100).

2. Place the tubes into the TissueLyser Adapter Set 2 x 24, and fix into the clamps of the TissueLyser. Grind the samples for 2 min at 25 Hz.

3. Repeat step 2, reversing the position of the tubes within the adaptor set. Rotating the racks of tubes in this way ensures that all samples are thoroughly and equally disrupted.

4. Add 400µl Buffer AP1 and 4µl RNase A stock solution (100 mg/ml) to a maximum
of 100 mg (wet weight) or 20 mg (dried) disrupted plant tissue and vortex vigorously. No tissue clumps should be visible. Note: Do not mix Buffer AP1 and RNase A before use.

5. Incubate the mixture for 10 min at 65°C. Mix 2 or 3 times during incubation by inverting tube.

6. Add 130µl Buffer P3 to the lysate, mix, and incubate for 5 min on ice.

7. Centrifuge the lysate for 5 min at 20,000 x g (14,000 rpm).

8. Pipette the lysate into the QIAshredder Mini spin column (lilac) placed in a 2ml collection tube, and centrifuge for 2 min at 20,000 x g (14,000 rpm).

9. Transfer the flow-through fraction from step 9 into a new tube (not supplied) without disturbing the cell-debris pellet. Typically, 450µl of lysate is recovered. For some plant species less lysate is recovered. In this case, determine the volume for the next step.

10. Add 1.5 volumes of Buffer AW1 to the cleared lysate, and mix by pipetting. For example, to 450µl lysate, add 675µl Buffer AW1. Reduce the amount of Buffer AW1 accordingly if the volume of lysate is smaller. A precipitate may form after the addition of Buffer AW1, but this will not affect the DNeasy procedure. Note: It is important to pipet Buffer AW1 directly onto the cleared lysate and to mix immediately.

11. Pipette 650µl of the mixture from step 10, including any precipitate that may have formed, into the DNeasy Mini spin column placed in a 2ml collection tube (supplied). Centrifuge for 1 min at ≥6000 x g (corresponds to ≥8000 rpm for most microcentrifuges), and discard the flow-through. Reuse the collection tube in step 12.

12. Repeat step 11 with remaining sample. Discard flow-through and collection tube.

13. Place the DNeasy Mini spin column into a new 2 ml collection tube (supplied), add
    500µl Buffer AW2, and centrifuge for 1 min at ≥6000 x g (≥8000 rpm). Discard the flow-through and reuse the collection tube in step 14.

14. Add 500µl Buffer AW2 to the DNeasy Mini spin column, and centrifuge for 2 min at 20,000 x g (14,000 rpm) to dry the membrane. Discard flow-through and collection tube.
    Note: Following the centrifugation, remove the DNeasy Mini spin column from the
    collection tube carefully so the column does not come into contact with the flow-through, as this will result in carryover of ethanol.

15. Transfer the DNeasy Mini spin column to a 2 ml microcentrifuge tube (not supplied), and pipette 100µl Buffer AE directly onto the DNeasy membrane. Incubate for 5 min at room temperature (15–25°C), and then centrifuge for 1 min at ≥6000 x g (≥8000 rpm) to elute.

16. Repeat step 15 once.
    Note: More than 200 µl should not be eluted into a 1.5 ml microcentrifuge tube because the DNeasy Mini spin column will come into contact with the eluate.

Appendix D: Adaptation of protocol from QIAamp Fast DNA Stool Mini Kit

1. Weigh out 0.5g of 0.5mm white silica/zirconia beads into a 2ml microcentrifuge tube. Add at least 0.05g of faeces, using liquid if needed.

2. Add 1ml InhibitEX Buffer to each stool sample. Shake for 10m at frequency 20Hz on the QIAGEN Tissuelyser II.

3. Heat the suspension for 45min at 70°C on a heat block, upending/shaking the tubes several times during this time. Vortex for 15s.

4. Centrifuge sample for 5 min to pellet stool particles.

5. Pipette 25µl Proteinase K into a new 2ml microcentrifuge tube

6. Pipette 600µl supernatant from step 4 into the 2ml microcentrifuge tube containing Proteinase K.

7. Add 600µl Buffer AL and vortex for 15 s.

8. Give the tubes a quick spin in the microcentrifuge to get drops from the lid into the tube. Incubate at 70degrees for 10min.

9. Add 600µl of ethanol (96–100%) to the lysate, and mix by vortexing. Do a quick spin in the microcentrifuge.

10. Carefully apply 600µl lysate from step 9 to the QIAamp spin column. Close the cap and centrifuge for 1 min. Place the QIAamp spin column into a new 2ml collection tube, and discard the tube containing the filtrate. Repeat step 10 until all lysate is loaded.

11. Carefully open the QIAamp spin column and add 500µl Buffer AW1. Centrifuge for 1 min. Place the QIAamp spin column in a new 2 ml collection tube, and discard the collection tube containing the filtrate.

12. Heat ATE buffer to 37°C. Carefully open the QIAamp spin column and add 500µl Buffer AW2. Centrifuge for 3 min. Discard the collection tube containing the filtrate.
    Place the QIAamp spin column in a new 2 ml collection tube and discard the old collection tube with the filtrate. Centrifuge for 3 min.

13. Transfer the QIAamp spin column into a new, labelled 1.5 ml microcentrifuge tube and pipette 50µl Buffer ATE directly onto the QIAamp membrane. Incubate for 3 min at room temperature, then centrifuge for 1 min to elute DNA. Repeat step 14, using the same 50uL of buffer - so move the 50uL from bottom of the tube back into the column. This gives a higher concentration of DNA in the same volume.