DNA METABARCODING: THE FUTURE OF DIET TRACKING?

A review of the potential of DNA metabarcoding for diet tracking of insectivorous birds

Jaron Koelewijn
S2748169
Supervisor: C.Both
Faculty of Science and Engineering

Abstract
Diet tracking is at the core of ecological animal research. However, current methods are expensive, labor-intensive, and time-consuming. So, there is a need for new methods. DNA metabarcoding is an upcoming new method. Metabarcoding uses gut content or feces to look at the DNA of prey. Specifically, it looks at conserved regions in the DNA like, e.g., the Cytochrome Oxidase subunit 1 (COI) gene to identify species. Metabarcoding is potentially cheaper, less labor-intensive, and better for the researched animals, but the method also has potential biases. This review investigates if metabarcoding can be used for diet research in insectivorous birds. First, we address the general workflow of metabarcoding and highlight the potential biases. Then we look at how metabarcoding performs compared with traditional methods through the literature. It was found that metabarcoding generally performs better than traditional methods. Using metabarcoding, more taxa were identified and to a taxonomic level that is often impossible for traditional methods. However, we also found that metabarcoding can be improved. The reference databases are not yet covering all species and using single marker genes often led to under-identification compared to using multiple marker genes. Altogether we conclude that metabarcoding is a suitable method for diet research in insectivorous birds, provided that a researcher considers potential biases.

Table of Contents

Introduction

The world is going through unprecedented change due to man-made climate change. This impacts nature all around the world. The phenology of plants and animals around the world is especially affected. Phenology is the timing of seasonal activities of animals and plants, and they have been shown to occur progressively earlier since the 1960s (Walther et al., 2002). This has already caused significant problems for animals. Many animals have an optimal time window in which environmental conditions are most advantageous, and often their reproduction is timed to be synchronized with this period. This window is usually set by the phenology of other species (Visser & Gienapp, 2019). This would not be a problem if all species reacted equally to climate change. However, this is not the case. It has been shown that different trophic levels have a different sensitivity to climate change (Thackeray et al., 2016). Lower trophic levels react stronger than the higher levels. This can lead to a trophic mismatch; this occurs when there is a phenological desynchronization between trophic levels. An example from the wild is the pied flycatcher (*Ficedula hypoleuca).* This migratory bird typically tries to have young when the caterpillar peak is happening so that there is enough food for the young. However, caterpillars reacted to climate change while the pied flycatcher did less. This led to a mismatch that caused local population declines of 90% in the pied flycatcher populations with an early food peak (Both et al., 2006).

To fully understand the impact of climate change on migratory birds, we need a better understanding of their ecology. A good start for this is the question: What do migratory birds eat? Variants of this question have always been at the heart of animal ecological research. The diet can teach us about a lot of different topics like predator-prey relationships (Leray & Knowlton, 2015), niche partitioning (Kartzinel et al., 2015), and the flow of energy in whole food webs (McCann, 2007). However, diet research has several problems. The first one lies with humans: species are often not easily distinguished. Often techniques are used that rely on morphological characters. This requires experienced taxonomists, but there are too few experienced taxonomic experts, and the existing expertise is declining. (Pearson et al., 2011). The second problem lies more with the animals; Animals are variable during the year. If conditions are unfavorable, they can change their diet or even habitat. This leads to complex foraging patterns in time and space (McMeans et al., 2015). As a result, one needs to observe the diet in multiple places during extended periods of time. This is both costly and labor-intensive. Lastly, not all animals are easily tracked; some are elusive or incredibly rare. Altogether this has led to a search for new techniques.

One of the techniques that is becoming increasingly popular is DNA metabarcoding (hereafter metabarcoding). Metabarcoding aims to use a small piece of the genome to recognize all living taxa. Then a library would be constructed of this small piece of the genome so that one only needs to sequence this small part of the genome to recognize the species. For this to work, a DNA fragment needs to be found that has the following properties:

> "1. The DNA fragment must be nearly identical in specimens of the same species but different between individuals of different species,
> 2. The section must be standardised (the same section should be used in different taxonomic groups),
> 3. The marker must be robust, with conservative primer binding sites that allow it to be readily amplified and sequenced." (Fišer Pečnikar & Buzan, 2014)

Because of these needed properties, mitochondrial DNA makes a good target. Mitochondrial DNA has three benefits, recombination is rare in mitochondrial DNA, it has no introns, and it is always inherited from the mother, which reduces variation (Hebert et al., 2003). For animals, the mitochondrial gene encoding for cytochrome c oxidase subunit 1 (COI) has been proposed as the standardized gene (Kress et al., 2015). This gene is involved in the respiratory chain that catalyzes the reduction of oxygen to water. This step is crucial in cellular respiration, and as result, the gene is very slow to evolve (Fišer Pečnikar & Buzan, 2014). But the gene still allows for enough change to

recognize species from each other. COI sequencing allowed the discrimination of 98% of the animal species (Fišer Pečnikar & Buzan, 2014).
Metabarcoding can be used to identify environmental DNA (eDNA). Environmental DNA includes DNA coming from feces and stomach contents (Alberdi et al., 2019). This then can teach us what an animal has eaten. Using metabarcoding in diet research has several advantages. First, the resolution: metabarcoding can identify species down to taxonomic species levels. This usually is hard or even impossible to do with traditional methods (Alberdi et al., 2019). The second advantage is the workload: a large part of the work in metabarcoding is lab-based, and unlike traditional methods, there is no need to do all samples one by one. Multiple samples can be processed parallel during the lab work, which dramatically decreases the workload (Alberdi et al., 2019). Lastly, there is the benefit that metabarcoding can detect prey items that are either soft or liquid (Nielsen et al., 2018). These items are normally degraded, making them hard to detect with traditional methods. However, there are also downsides; it is hard to distinguish between DNA from prey and DNA from other sources. This can vary from secondary predations to DNA in water that is ingested together with aquatic prey (Nielsen et al., 2018). This problem is further increased because there is still debate about how we should process the data recovered from metabarcoding (Deagle et al., 2019). Lastly, there is the fact that metabarcoding can only tell you which species you have and not the size of the prey or in which life stages it is. Especially not knowing the life stage can be a handicap. For example, some insects change habitat depending on their life stage. Not knowing the life stage makes it impossible to know where the prey was found.

Altogether metabarcoding looks like a promising new technique for diet research, but it also has a few downsides. In this paper, we will examine if metabarcoding is a suitable method for diet research. We will do this with the question: How does DNA metabarcoding compare to traditional methods. We will first outline the general workflow for a metabarcoding study to answer this question. Then we will investigate how metabarcoding performs with mock samples and in field conditions. Lastly, we shortly provide some context to the debate of how metabarcoding should be analyzed. Altogether we hope to provide a framework that shows the strengths and weaknesses of metabarcoding for researchers considering this method.

The metabarcoding workflow

Study design

The first step of a metabarcoding diet study is the study design. A metabarcoding study should adhere to a robust ecological design, as with all ecological studies. That is to say; one should aim for a proper number of sites and replicates to have enough statistical power for the analysis. To further improve the analysis, one should also collect a good set of environmental parameters (Creer et al., 2016). If this groundwork is in order, one should also take multiple decisions into account that are specific to a metabarcoding study. One of the initial decisions should be taxonomic level because this will influence most of the choices downstream. Depending on the chosen taxonomic level, a marker region should be selected. This marker region should have a reference library compatible with the chosen taxonomic level. For animals, the COI gene is the most used gene for metabarcoding. The Barcode Of Life Database (BOLD) (Ratnasingham & Hebert, 2007) is an extensive library, but it does not yet include all species. As of writing, the database contains 242K animal species (*Bold Systems V4*, n.d.), and the coverage differs between geographical regions. So, it is necessary to check if the library contains the species that are expected in the diet. Otherwise, it might be necessary to build a library for the study by, e.g., using samples that are analyzed using traditional methods.
Next, the primers should be designed or selected, depending on the taxonomic level. Primer design is complicated by the fact that the COI gene is encoding a protein. Since mutations in the protein-encoding region often result in a protein that does not function, the protein-encoding part is highly conserved, but this is not the case for the primer binding site since changes here do not always influence the protein. So, it is challenging to locate a conserved area that can serve as a primer

region for all (study) species (Deagle et al., 2014). Altogether this makes it almost impossible to design primers that do not have a mismatch with at least some of the target species (Alberdi et al., 2018).
The mismatches with species can lead to a lower detection of these species. However, this does not have to be all bad. If a mismatch can be created for non-target species the PCR yields more DNA for the target species(Verkuil et al., 2022). Since both the marker region and the primer choice can affect downstream results it is wise to beforehand select them beforehand. This gives the opportunity to consider potential biases and decided if biases are relevant to the results.

Sample collection

There are two possible sample types possible for a diet metabarcoding study. The first is stomach contents; for this method, the stomach's content is collected post-feeding immediately. This has the significant advantage that the DNA of the prey should be less degraded than in other methods (Alberdi et al., 2019). However, this method has a significant downside. Extracting stomach contents is highly invasive for an animal. It requires capture, sedation, and in some cases, even killing the animal. These methods can significantly impact an animal and the study system. This can have consequences for the study itself, but these methods also face legal and ethical questions (Alberdi et al., 2019).
Consequently, feces is the more popular sample type for dietary analyses. In contrast to stomach content, feces can be collected non-invasively. It can either be collected directly from captured animals or the environment. The direct collection is preferable because there is less distortion by the environment. This is especially the case when animals defecate nearby (potential) prey items, e.g., arthropod eaters that eat arthropods that like to feed on feces. However, collection from the environment has a major benefit: you do not need to find the animal itself; this is a benefit when dealing with elusive, sensitive, or aggressive animals.
Regardless of the chosen sample type, it is essential to handle the sample afterward carefully. Any contamination can add foreign DNA to the sample, distorting the results. Furthermore, it is important to stop the degradation of DNA after collection. However, not all methods are suitable for preserving DNA; transformed alcohols and formalin should be avoided since these compounds denature nucleic acids. Denatured nucleic acids make the downstream analysis of the DNA impossible. Nonetheless, DNA is quite robust against other preservation forms; methods like drying, -20 °C freezing, and 100% ethanol are suitable (Creer et al., 2016). Although these are suitable methods, there is not much research about the effects of using different preservatives on metabarcoding results. So, it is highly recommended to use the same method for all samples (Alberdi et al., 2019).

Sample processing

After all the previous steps, there should now be a sample consisting of feces or stomach contents. Of this sample, only the DNA is needed for the metabarcoding. So, the next step is to isolate the DNA from the sample. Commonly this is done by using commercial DNA extraction kits. A plethora of kits are available, but analysis shows that different kits can result in different results (Dopheide et al., 2019; Verkuil et al., 2022). It is recommended to use only one type of extraction kit for all samples and report the type used when publishing.
With the isolated DNA, the PCR can be performed. Metabarcoding only relies on a small marker region in the DNA, so the PCR is performed with primers specific to the marker region. However, not all primers are identical; metabarcoding primers always have certain biases since it is virtually impossible to design primers that work the same for all taxa. In silico comparisons also reported large differences between primer sets (Piñol et al., 2019), proving that it is crucial to select proper primers beforehand. Additionally, it is important to use high-fidelity DNA polymerases because errors made by the DNA polymerases can lead to misidentification (Liu et al., 2020).
But primers are not the only thing that can influence the results of the PCR. There is some randomness involved in a standard PCR. To copy the DNA, a primer needs to meet the template. This

is a random process, and by pure chance, certain templates may be copied less. This will influence the results, and in the worst-case DNA is copied at such a low rate that it leads to a false negative. This problem is further increased because environmental samples are complex and can have low temple DNA (Ficetola et al., 2015). It is recommended to do at least 6-8 replicates. This should detect even the rarest taxa (Ficetola et al., 2015).
Lastly, there is the problem of chimeric sequences. Chimeric sequences are artifact sequences resulting from two or more sequences that have joined incorrectly. According to Shin et al., these are mainly formed due to sequence similarities and premature termination of DNA extension near the primer region (Shin et al., 2014). This can be a significant problem since metabarcoding always targets a similar (marker) region. However, there are algorithms that can detect chimeric sequences, although the chimeras formed in metabarcoding studies are particularly hard to detect (Alberdi et al., 2019).
After the PCR, there is now a sample with all copied and amplified marker region genes. The last step before analysis is reading out the nucleotide sequence of these genes. Typically this is done with a process called sequencing. Here the gene is copied using labeled nucleotides. The labeled nucleotides are then used to read out the nucleotide sequence. This method is used universally in studies that use DNA. However, two things need some more attention in a metabarcoding study. The first one is library preparation. In this process, the samples are given a tag used to recognize from which sample the DNA comes. This tag is attached with PCR using generic primers and is typically a 5'-nucleotide tag (Schnell et al., 2015). However, experiments show a phenomenon called tag-jumping during this process (Schnell et al., 2015). Tag-jumps happen when pieces of DNA are tagged with an incorrect tag. Later downstream, this DNA is then counted to the wrong sample. For example, Schnell et al. found bat diet in leech diet samples and vice versa when sequencing them together. Schnell et al. suggest that tag jumping is the result of blunt-ending of pools of tagged amplicons and chimera formation. So they advise avoiding methods that rely on blunt ending during the library build step and employing techniques to prevent or reduce chimera formation during the library index PCR (Schnell et al., 2015).

Data processing

After sequencing, the data now needs to be analyzed. This process is called bioinformatics. During this process, a so-called pipeline is constructed that converts the sequenced data into an OTU table (Operational Taxonomic Unit). The sequences of the OTUs are then matched with sequences of the reference database, to assign the correct taxonomic level (family, genus, species) to each OTU. Since this process typically falls outside the expertise of (field) biologists, it is recommended to consult a bioinformatics expert (Liu et al., 2020). Since bioinformatics has enough content for a whole paper, we cannot fully explain it, so we will only elaborate on the databases. Potential further reading on bioinformatics can be done in reviews on the matter (Deiner et al., 2017; Piper et al., 2019).

Databases are needed for the crucial last step of metabarcoding. Without them, the OTUs cannot be matched to taxons in the real world. However, it is important to consider the reliability of matches between OTUs and the database. This reliability depends on three things (Alberdi et al., 2019). The first one is the similarity between generated and reference sequences. This seems obvious at first glance: The better the match, the more reliable. This is the case, but this is also reliant on the second point of attention: the length of the sequence. Short sequences have a higher probability that they perfectly match a reference sequence than a longer sequence. However, this is not always desirable. If a short sequence matches multiple things, it is impossible to tell which one it should be. Lastly, an important point is the completeness of the database. This one might be the most important to think about beforehand. You can't find a match if your expected species are not in the database. But maybe more dangerous is a partial database. If you have a genus with three species A and B, but only A is in the database, you might high matches with A for B. This can lead to the conclusion that it is species A, while a complete database would have given a high match for A and B. In the latter case, only the genus identification was certain, while species stayed uncertain.

This is also why a good database can lead to a lower identification at lower levels of taxonomy.

FOO vs. RRA

After the whole DNA metabarcoding, there is now data to analyze. However, there is still debate what the best way to analyze metabarcoding data. Since this matters for the results of all studies, we will shortly explain the two schools of thought in this debate.
In essence, the debate revolves around whether we can use metabarcoding to quantify the abundance of species in the diet. The most common method does not quantify the abundance of species in the diet. Instead, it analyses the presence or absence of taxa. This is then often converted into Frequency Of Occurrence (FOO). Which is the percentage of samples that a taxon occurs in. Commonly there is a threshold set for the minimum amount of sequences needed. If it is above that, the taxa are present; otherwise, the taxa are absent. This method is considered safe and reliable because the result should be the same even when biological or technical biases distort the ratio between taxa. Because biological and technical biases are thought to be influential in metabarcoding, most studies state that using quantitative methods is not yet possible (Galan et al., 2018; Mcclenaghan et al., 2019; Swift et al., 2018; Vesterinen et al., 2016). This is often supported by references to papers like (Piñol et al., 2015; Thomas et al., 2014).
However, there is pushback against this idea. More recent papers point out that FOO also has its flaws. Deagle et al. point out that most molecular ecologists would agree that there should be a difference between a food taxon with 10.000 reads and one with only a few (Deagle et al., 2019). This difference will be discarded if you run an analysis with only presence/absence. As a result, there is little difference between a diet item that makes up 80% of the diet and an item that makes up 5% of the diet. Opponents of FOO argue that this leads to overestimating the importance of rare items. In turn, this can increase problems with contamination and secondary predation. Although these problems add only a little DNA to the sample, it can be enough to be detected with presence/absence (Deagle et al., 2019).

The alternative method of analysis would be relative read abundance (RRA). This method uses the percentages of DNA belonging to each prey species as a proxy for biomass consumed. This is beneficial because it retains the difference between a food item with only a few reads and one with many. Additionally, it lowers the problems with contamination and secondary predation, these things will add reads to the total, but the correct reads should still dominate the overall sample(Deagle et al., 2019). However, this method also has a significant drawback. It is especially vulnerable to biological or technical biases. If one of these biases leads to an over-or underestimation of one of the food items, it will also influence the other items in the sample since it is converted to percentages. Nonetheless, these biases can be mitigated, and there are already examples showing that RRA works and reflects data out of the field more accurately than FOO (Verkuil et al., 2022).

The debate if we should use FOO or RRA will continue for the foreseeable future. But more knowledge about the biases of metabarcoding and their causes might swing the favor of RRA. However, to do this, there is a need for more empirical studies to show the differences between FOO and RRA accurately. For the rest of this paper, it is important to remember that RRA and FOO studies are not the same and, as a result, are not always directly comparable to each other.

Field studies

Now the general principles of metabarcoding should be clear. However, the biggest question remains. Is this method suitable for diet research? To answer this question, we will look at multiple field studies that compare the performance of metabarcoding with either controlled conditions or compare it to traditional methods. Literature on this topic was searched on Web of Science.
We used the following criteria:

- The study uses metabarcoding to investigate diet

- The study validates metabarcoding in some way
- The study uses field samples
- The study subjects are insectivorous birds or bats.

Ideally, we would have only selected studies with insectivorous birds. However, since metabarcoding is a relatively new technique, there are not enough studies focusing on insectivorous birds. To still have enough literature, we also included bats. Metabarcoding research has been done in bats for a longer time, and they also forage on insects. This makes them a suitable proof of concept for metabarcoding an insectivorous diet. If a study also focussed on different topics like a nectivorous diet, we only deemed it suitable when the results of the insect component were presented separately. A overview of the used studies is given in supplement 1.

Known samples

To know if a method is suitable for diet research, we need to see if we measure what the animal eats. One way to do that is to feed an animal a known diet and test if the method picks up on everything that was part of the diet. Swift et al. used this method to test the effectiveness of metabarcoding in bats (Swift et al., 2018). Their study presented a population of zoo bats (*A. pallidus*) with eleven insect species and allowed them to feed ad libitum on preferred items. After having this diet for a week, the guano pellets of the bats were collected for three days. Then they used two markers, 16S and COI, to look if they could find the fed prey taxa in the diet. Using both markers together, they consistently found five of the diet items in the diet which was in line with dietary preferences observed by the zoo staff(Swift et al., 2018). Because this study used a known diet, they could create a mock sample based on the offered prey samples. They did this to control their results and found significant differences between the performance of the different markers. The 16S marker detected 8 out of the 11 species, while COI only identified 6. These results are not that good, and they think marker resolution and gaps in the database were to blame for the low identification. However, using both markers together led to identifying 10 out of the 11 species (91%) (Swift et al., 2018).

A similar experiment was done by Galan et al. (Galan et al. 2018). In this study, they tested mock samples and compared the results to the literature. In contrast to Swift et al., they used only COI as a marker. They made two different mock samples by mixing premade sequences with fixed percentages to test between FOO and RRA. In the first mock sample, they could detect 11 out of 12 species, and in the second mock sample 7 out of 7. However, they could not identify two species because they had no references in the used database. Three other species had multiple hits on species level because the sequences were too similar to closely related species(Galan et al., 2018). Furthermore, they found that they could not rely on the RRA. Their first mock sample had 12 species with an equal share, so all species should be 8.3%. However, species ranged between 0.4%-30.1%. The second mock sample had similar results: the frequencies should be 14.3% with seven species, but the values ranged from 1.5% to 27.3%. These results suggest that FOO works appropriately but that the PCR biases are too severe to analyze using RRA. However, this seems to be a somewhat preplanned risk. Galan et al. point out multiple sources of potential causes of PCR biases, with as most severe a mismatch of the primer for certain species. But their study does not seem to employ measures to lessen these biases. So while their current PCR protocol might lead to unusable RRA this might be improved by optimizing the PCR properly.

Taken together, these studies are promising for metabarcoding. Analysis of the mock communities shows that most sequences can be identified and down to species level. Achieving such a taxonomic resolution is typically difficult using traditional methods. Nevertheless, they also found that the technique is not yet perfect. Galan et al. showed that while FOO works well for the mock samples, their current PCR protocol was not ready for analysis using RRA. Additionally, both studies had problems identifying all their sequences due to gaps in the databases. However, using multiple markers improved the situation for Swift et al.

Morphological analysis

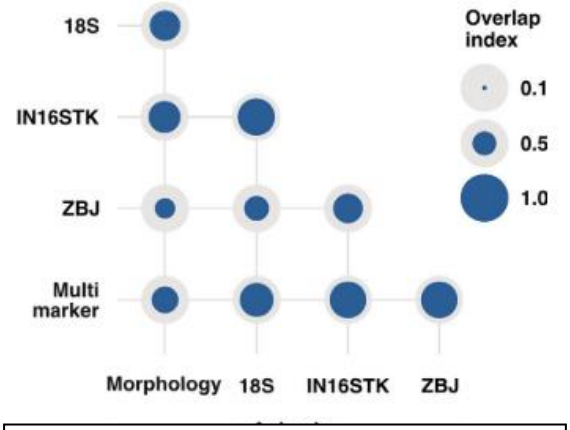
Galan et al. also looked at field samples to compare them to morphological data in the literature (Galan et al., 2018). In their study, Galan et al. captured bats between June and September 2015 from their summer roots on 18 sites in Western France. The bats were placed in a cotton holding bag during measurement. From this cotton bag, fecal pellets were collected and stored in microtubes. 357 samples were collected from 16 bat species. However, since the first intent was to analyze the fecal pellets using morphological analysis, the samples were stored at room temperature for 45-162 days. These conditions are not suitable for DNA; as mentioned earlier in this paper, the samples should be stored in ethanol or kept a -20 °C. But even with these conditions, they managed to get data out of 82% of the samples. The study investigated the dietary composition of 16 bat species using FOO. Their results were in line with the literature. That is to say; they found a high FOO for the orders reported by the literature.
Furthermore, metabarcoding was found to give a greater taxonomic resolution than the literature. The literature was based on morphological analysis; this allows for identification to order level or, at best, family level. Metabarcoding provided identification down to genus and species level (Galan et al., 2018).

Another study, done by da Silva et al., made a direct comparison (da Silva et al., 2019). They analyzed 115 feces of Black Wheatears (*Oenanthe leucura*) by doing both morphological identification and metabarcoding. They used three markers, one universal for eukaryotes (18S) and two specific for arthropods, namely, 16S and COI. They found that there were multiple significant differences between morphological methods and metabarcoding. As shown in table 1, they found 23 taxa out of 8 orders with morphological analysis, while all metabarcoding markers found more orders and more taxa. Secondly, the study found significant differences in composition among morphological and molecular methods. Lastly, they looked at the overlap between different methods. For this, they used the Czekanowski index. The formula of the Czekanowski index is:

Method	Number of samples with prey (N=115)	Orders	Taxa
Morphological examination	112	8	23
18S	94	21	91
16S	113	21	244
COI	108	18	231

Table 1. Performance of morphological examination and metabarcoding using three different markers(da Silva et al.,

$$CI = 1 - 0.5 \sum_{j=1}^{S} p_{x,i} - p_{y,i},$$



Czekanowski overlap for different methods. (da Silva et al., 2019).

Where both p's are the proportions of the ith diet item used by the prey and S as the total number of diet items, their results are shown in figure 1. They found an overlap of 0.435-0.673 between morphology and the metabarcoding markers. This means there is only between 45 and 67% overlap between the results. Da Silva et al. do not elaborate on why this is. However, looking at their results, it is clear that it results from higher detection by the markers. The only exception is the COI marker ZBJ. This marker fails to detect the Hymenoptera family Formicidae, while the other markers and morphological analysis detect this as an important diet item. However, ZBJ detects Lepidoptera significantly more than the other markers, and morphological analysis misses Lepidoptera entirely. This also leads to the striking results that the markers have only a slightly higher overlap between each other, 0.525-0.781. So Da Silva et al. also tested metabarcoding using all markers together. With this approach, there were 28 animal orders

detected in 112 samples. The overlap with morphology was 0.563, while the overlap with individual markers was 0.711-0.777 (da Silva et al., 2019). It is commendable that Da Silva et al. also tested this multimarker approach because these results show that this is the best approach. Using multiple overlapping markers looks redundant, but they show that it is complementing. This is especially clear with Lepidoptera and Formicidae. Lepidoptera was under-detected by markers other than ZBJ, but ZBJ missed the family Formicidae. The multimarker approach detected both.

Altogether these two studies show that metabarcoding performs better than morphology. Galan et al. found metabarcoding led to similar results for bats but with greater taxonomic resolution. While da Silva et al. showed that morphological examination detected far fewer taxa than metabarcoding. However, the study of Da Silva et al. and Swift et al. indicated that a single marker approach might be fallible. Both studies showed that markers failed to identify all species and, as result, missed parts of the diet. As shown by both studies, a multi-marker approach lessens this problem.

Camera traps

So far, we have looked at traditional methods that relied on the morphological examination of feces. However, this has a downside: the prey is already digested, so it is hard to detect soft-bodied insects. So, performing equal to or better than this method might not be enough. However, Verkuil et al. also showed that metabarcoding performs well compared to camera data(Verkuil et al., 2022). In their study, Verkuil et al. looked at the diet of the pied flycatcher. The study recorded the diet of nestlings using camera traps during 39 sessions on different nests for three years. From the resulting footage, the food item was identified down to the lowest taxonomical level possible. In addition, the relative size of the prey in relation to the adult's beak was scored. Lastly, the scaled biomass contribution of each taxon was calculated by multiplying the prey counts by a multiplication factor based on the prey size relative to the bill. This data was then compared to the metabarcoding data.
For the metabarcoding, the feces of 1-3 nestlings were collected. The feces were then mixed, resulting in 1 sample per nest sampled. The study used COI as marker gene, but instead of the common ZBJ primers, primers designed to match less with bird DNA and better with spiders were used. Another notable thing about this study is that it looked not only at FOO but also at whether RRA could be used as a quantitative estimate of the relative biomass contributed by each taxon. The study showed that metabarcoding performed better in identifying prey items than camera records. COI barcodes found 22 orders, and in contrast, camera records found 18. At the family level, the difference was larger. In the six most abundant orders, metabarcoding found 105 taxa, while camera records found 50 taxa. Furthermore, the study examined the correlation between the metabarcoding results and the scaled biomass detected on camera. It found that FOO had a correlation of 0.65 on order level and 0.61 on family level. More strikingly, the study found that RRA had a higher correlation, respectively 0.85 on order level and 0.74 on family level. This shows that RRA can be used as a quantitative estimate for relative biomass(Verkuil et al., 2022). This contrasts with the findings of Galan et al., who found that the PCR biases skewed the data of the RRA too much.

Environmental traps

So far, we have focussed on finding out what the animal has eaten. However, this is only one part of the puzzle. It requires knowledge of what an animal has available to understand the diet truly. Research into prey availability often samples the environment using traps. The prey in those traps is identified using time-consuming traditional methods. Using metabarcoding instead might be an improvement, so multiple studies looked into metabarcoding.

The first study was a proof-of-concept study by Rytkönen et al (Rytkönen et al., 2019). In this study, COI was used as a marker not only to describe the diet of four tit species but also to monitor the availability of potential invertebrate prey. To do this, feces of the birds and larvae (frass) were collected. Metabarcoding could identify 95% of OTUs to order level, 74% to family level, 66% to genus, and 60% to species level. The results showed that around three-quarters of the diet consisted

of Lepidoptera, which is consistent with the literature. However, compared to the literature, metabarcoding allowed further identification often to species level. Furthermore, the study showed that prey species could be identified using their frass. Although there was little overlap between the diet and the frass samples, this proves that metabarcoding can be used to determine a predator's diet and prey availability from their respective feces (Rytkönen et al., 2019).

Another study investigated barn swallow nestlings (*Hirundo rustica)* (Mcclenaghan et al., 2019). This study used COI metabarcoding to compare the diet with insects found in Malaise traps. The insects found in the Malaise traps were measured and identified by an expert, after which a selection of the insects was identified using metabarcoding. Using metabarcoding, 87.4 of the specimens were identified. Of these, 99.1% were identified to order level, 97.5% to family, 71.7% to genus, and 40.2% to species levels. This data was then used to create a custom database for the diet metabarcoding. For the diet, three fresh feces were collected and pooled every two days from day eight after hatching till fledging. From these feces, 88.8% of the OTUs were matched to a reference sequence; of these, 53.1% matched the custom database. Of the identified OTUs, 100% were identified to order level, 98.5% to family, 82.8% to genus level, and 14.5% to species level. The study's last step investigated prey selection, specifically if barn swallows preferred taxa of larger size. This was done by matching the diet OTUs with the custom database. Since this database also had length measurements, matches were used to assign both taxa and length. The study found that the diet had more prey of larger size (>3mm) than the environmental samples, leading to the conclusion that barn swallows preferred larger prey. The study mentions that metabarcoding cannot be used for quantitative methods due to biases, including the digestibility of prey items. Because of this, the study uses FOO. However, this leaves questions about the prey selection conclusion. First, why would differences in digestibility not influence FOO? Large prey items could digest slower, possibly leading to more detection, especially since samples are only taken every two days. Additionally, it is a known critique of FOO that it overestimates the importance of rare diet items. The environmental data shows that large insects are the rare category. Lastly, there is the question of how they compare. The study says that it uses FOO as qualitative measurement because quantitative is not possible with their metabarcoding setup. However, the abundance of insects in the environment is a quantitative measurement. Usually, these are not directly comparable, so it needs more explanation why they compare the frequency of occurrence of large prey items with the abundance in the environment.

Together these two studies show promise that metabarcoding can be used to describe birds' diet and their prey's availability. Both studies show that it is possible to use metabarcoding combined with a sampling method to identify prey availability in the environment. Nonetheless, the studies were not more than a proof of concept. Rytkönen et al. describe a low overlap between prey species found with the environmental samples and prey detected in the diet. At the same time, McClenaghan et al. could only identify 53.1% of the diet using sequences obtained from environmental samples.

Conclusion

This paper aimed to examine if metabarcoding was a suitable method to use in diet metabarcoding. We have shown that the basis of metabarcoding is solid; metabarcoding performed well in describing mock samples and a known diet in zoo bats(Swift et al., 2018). Additionally, the performance compared to traditional methods was good. Comparison with morphological analysis (both direct and with literature) showed that metabarcoding found what morphological analysis also found. However, in the case of Galan et al., metabarcoding gave a greater taxonomic resolution (Galan et al., 2018), while Da Silva et al. found both a greater taxonomic resolution and taxa that went undetected by morphological analysis (da Silva et al., 2019). Comparison with camera data gave the same picture; metabarcoding was comparable with camera data but detected more taxa (Verkuil et

al., 2022). Lastly, we looked at environmental traps. Both studies showed that it was possible to use metabarcoding to identify prey in their environment and diet.
Nonetheless, it also showed that comparing metabarcoding data with abundances is complicated, especially when using FOO. However, this might be improved by the use of RRA. This method was previously thought to be unusable due to biases during the metabarcoding process, but Verkuil et al. showed that it is possible to use RRA. But even with RRA, metabarcoding can only be used for relative measurements, not absolute ones.
Even though metabarcoding provided well, it was also shown that metabarcoding could miss certain taxa while they are important parts of the diet. Two problems mainly caused this. The first was lacking database coverage (Galan et al., 2018; Swift et al., 2018). This emphasizes the need to check if there is a suitable database when planning to use metabarcoding. Secondly, primer mismatches lead to underdetection of certain taxa. Altogether it is advisable to use multiple markers since the studies using multiple markers showed that these problems were mitigated in their design. If this is not possible, it is advisable to have some way of validation within the study to detect potential primer mismatches or gaps in the database.
All together, we can recommend metabarcoding as a suitable method for diet research. It is faster, cheaper, and offers a greater taxonomic resolution, provided that a researcher familiarizes themselves with the potential biases and pitfalls of metabarcoding beforehand.

Literature list

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

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

Bold Systems v4. (n.d.). Retrieved June 29, 2022, from https://boldsystems.org/

Both, C., Bouwhuis, S., Lessells, C. M., & Visser, M. E. (2006). Climate change and population declines in a long-distance migratory bird. *Nature*, *441*(1), 81–83. https://doi.org/10.1038/nature04539

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., Potter, C., & Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. In *Methods in Ecology and Evolution* (Vol. 7, Issue 9, pp. 1008–1018). John Wiley & Sons, Ltd. https://doi.org/10.1111/2041-210X.12574

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

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biology Letters*, *10*(9). https://doi.org/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., & Eveson, J. P. (2019). Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology*, *28*(2), 391–406. https://doi.org/10.1111/mec.14734

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

Dopheide, A., Xie, D., Buckley, T. R., Drummond, A. J., & Newcomb, R. D. (2019). Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. *Methods in Ecology and Evolution*, *10*(1), 120–133. https://doi.org/10.1111/2041-210X.13086

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

Fišer Pečnikar, Ž., & Buzan, E. V. (2014). 20 years since the introduction of DNA barcoding: From theory to application. *Journal of Applied Genetics*, *55*(1), 43–52. https://doi.org/10.1007/S13353-013-0180-Y/TABLES/1

Galan, M., Pons, J. B., Tournayre, O., Pierre, É., Leuchtmann, M., Pontier, D., & 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), 474–489. https://doi.org/10.1111/1755-0998.12749

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/RSPB.2002.2218

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

Kress, W. J., García-Robledo, C., Uriarte, M., & Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology and Evolution*, *30*(1), 25–35. https://doi.org/10.1016/j.tree.2014.10.008

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

Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burridge, C. P. (2020). A practical guide to DNA metabarcoding for entomological ecologists. In *Ecological Entomology* (Vol. 45, Issue 3, pp. 373–385). John Wiley & Sons, Ltd. https://doi.org/10.1111/een.12831

McCann, K. (2007). Protecting biostructure. *Nature 2007 446:7131*, *446*(7131), 29–29. https://doi.org/10.1038/446029A

Mcclenaghan, B., Nol, E., Kerr, K. C. R., Metabarcoding, D., Swallow, B., Mcclenaghan, B., Nol, E., & Kerr, K. C. R. (2019). DNA metabarcoding reveals the broad and flexible diet of a declining aerial insectivore. *The Auk*, *136*(1), 1–11. https://doi.org/10.1093/AUK/UKY003

McMeans, B. C., McCann, K. S., Humphries, M., Rooney, N., & Fisk, A. T. (2015). Food Web Structure in Temporally-Forced Ecosystems. *Trends in Ecology and Evolution*, *30*(11), 662–672. https://doi.org/10.1016/j.tree.2015.09.001

Nielsen, J. M., Clare, E. L., Hayden, B., Brett, M. T., & Kratina, P. (2018). Diet tracing in ecology: Method comparison and selection. In *Methods in Ecology and Evolution* (Vol. 9, Issue 2, pp. 278–291). John Wiley & Sons, Ltd. https://doi.org/10.1111/2041-210X.12869

Pearson, D. L., Hamilton, A. L., & Erwin, T. L. (2011). Recovery Plan for the Endangered Taxonomy Profession. *BioScience*, *61*(1), 58–63. https://doi.org/10.1525/bio.2011.61.1.11

Piñol, J., Mir, G., Gomez-Polo, P., & 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), 819–830. https://doi.org/10.1111/1755-0998.12355

Piñol, J., Senar, M. A., & 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), 407–419. https://doi.org/10.1111/mec.14776

Piper, A. M., Batovska, J., Cogan, N. O. I., Weiss, J., Cunningham, J. P., Rodoni, B. C., & Blacket, M. J. (2019). Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. In *GigaScience* (Vol. 8, Issue 8, pp. 1–22). Oxford Academic. https://doi.org/10.1093/gigascience/giz092

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*. https://doi.org/10.1111/j.1471-8286.2006.01678

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

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

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

Swift, J. F., Lance, R. F., Guan, X., Britzke, E. R., Lindsay, D. L., & Edwards, C. E. (2018). Multifaceted DNA metabarcoding: Validation of a noninvasive, next-generation approach to studying bat populations. *Evolutionary Applications*, *11*(7), 1120–1138. https://doi.org/10.1111/eva.12644

Thackeray, S. J., Henrys, P. A., Hemming, D., Bell, J. R., Botham, M. S., Burthe, S., Helaouet, P., Johns, D. G., Jones, I. D., Leech, D. I., MacKay, E. B., Massimino, D., Atkinson, S., Bacon, P. J., Brereton, T. M., Carvalho, L., Clutton-Brock, T. H., Duck, C., Edwards, M., … Wanless, S. (2016). Phenological sensitivity to climate across taxa and trophic levels. *Nature 2016 535:7611*, *535*(7611), 241–245. https://doi.org/10.1038/NATURE18608

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

Verkuil, Y. I., Nicolaus, M., Ubels, R., Dietz, M. W., Samplonius, J. M., Galema, A., Kiekebos, K., de Knijff, P., & Both, C. (2022). DNA metabarcoding quantifies the relative biomass of arthropod taxa in songbird diets: Validation with camera-recorded diets. *Ecology and Evolution*, *12*(5), e8881. https://doi.org/10.1002/ece3.8881

Vesterinen, E. J., Ruokolainen, L., Wahlberg, N., Peña, C., Roslin, T., Laine, V. N., Vasko, V., Sääksjärvi, I. E., Norrdahl, K., & Lilley, T. M. (2016). What you need is what you eat? Prey

selection by the bat Myotis daubentonii. *Molecular Ecology*, *25*(7), 1581–1594. https://doi.org/10.1111/mec.13564

Visser, M. E., & Gienapp, P. (2019). Evolutionary and demographic consequences of phenological mismatches. *Nature Ecology & Evolution 2019 3:6*, *3*(6), 879–885. https://doi.org/10.1038/S41559-019-0880-8

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J.-M., Hoegh-Guldberg, O., & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature 2002 416:6879*, *416*(6879), 389–395. https://doi.org/10.1038/416389A

Supplement 1

Reference	Database	FOO	RRA	Taxonomic level	Correlatie/result
Swift et al 2018	NCBI	0	0	species	Diet:Metabarcoding found diet items in line with observations by zoo staff. Mock: COI identified 8 out of 11 species, 16S 10/11 and combining both identified all species
Galan et al	BOLD	1	1	species	Metabarcoding identified 11 out of 12 species and 8/8 out of two mock samples. However the RRA was not comparable with starting percentages
					Results for metabarcoding were conguerent with literature based on morphological research. However metabarcoding provided greater taxonomic resolution
Da Silva et al 2019	BOLD, NCBI	1	0	Order and below	Metabarcoding identified more taxa then morphological analysis. Multimarker performed best overall. Overlap between morphological analysis and multi marker metabarcoding =0.563
Rytkonen et al 2018	BOLD	1	0	Order but mentione	Succes in identifying prey species from both predator feces and their own feces
Verkuil et al 2022	GenBank	1	1	Order, Family	Metabarcoding results have a high correlation with results from camera data. RRA has the highest correlation (R=0.85), while FOO has considerable lower correlation (R=0.65)
McClenaghan et al 2019	Custom	1	0	Order, Family	Succes in Identifying prey species from both predator feces and environomental samples.

Reference	Consumer	Prey type	Sample type	Geography	N	Aim	Marker	Mark length (bp)	Primers
Swift et al 2018	Bat	Arthropods	Guano	United states	42	Feed known diet to look if metabarcoding works	16S/COI	not mentioned	ZBJ-ArtF1c/ZBJ-ArtR2
Galan et al	Bat	Arthropods	Faecal pellets	France	336	Evaluate protocol with mock sample and do real s	COI	133/658	
Da Silva et al 2019	Black wheatears	Arthropods	Feces	Portugal	115	Comparing 3 markers with morphological analysis	18S, IN16STK (16S), ZBJ (COI)	not mentioned	COI= ZBJ 16S = IN16STK-1F_mod/ N16STK-1R_mod 18S = not specified
Rytkonen et al 2018	4 tit species	Arthropods	Nestling faeces	Finland	14	we test the metabarcoding methods in describing	COI	157	ZBJ-ArtF1c/ZBJ-ArtR2
Verkuil et al 2022	Pied flycatcher	Arthropods	Feces	Netherlands	63	Compare diet found with metabarcoding with cam	COI	not mentioned	LCO1490T/HCO1777T
McClenaghan et al 2019	Barn swallow	Arthropods	Nestling feces	Canada	271	Compare diet found with metabarcoding with Inse	COI	157	ZBJ-ArtF1c/ZBJ-ArtR2