DNA metabarcoding: A method in animal diet analysis studies

Oona Herzog

Bachelor thesis
Biology degree program
Oulu University
December 2019

Table of Contents

1. Introduction.. 1

2. Description of DNA metabarcoding 2

2.1 Sample collection 2

2.2 Sample preservation 4

2.3 DNA extraction 4

2.4 DNA barcode amplification 5

2.5 Sequencing 6

2.6 Data analysis... 6

3. Reference databases 7

3.1 BOLD 7

3.2 FinBOL 8

3.3 GenBank 8

3.4 Customized databases.. 9

4. Critical aspects of metabarcoding 9

4.1 Biological distortion factors 9

4.2 Technical distortion factors.. 11

5. Some other applications of metabarcoding... 14

5.1 A case study.. 15

6. Discussion and conclusion.. 17

References.. 18

1. Introduction

Understanding food webs in ecosystems is largely based on the knowledge of the diet of organisms. A better comprehension of biodiversity and the functioning of ecosystems is achieved when exploring the structure of those food webs in addition to temporal and spatial variations within them (Rytkönen et al., 2019). Furthermore, knowing predator diets and their responses to the variation of prey availability helps understanding food web structure along with population dynamics (McClenaghan et al., 2019).

Prey species identification is traditionally carried out with methods such as video documentations, direct observations or microscopy of fecal matter (Rytkönen et al., 2019). These traditional methods have been shown to be problematic especially with insectivores due to the difficulty of directly observing predation as well as undirect observations being challenging as insect preys are very variable and mostly digested in the guts (McClenaghan et al., 2019).

Non-invasive diet analyses for numerous species are nowadays much simplified with the help of molecular methods such as DNA barcoding and metabarcoding (McClenaghan et al., 2019). DNA barcoding is a method for species identification using variations in short sequences of the DNA (also called a DNA barcode) found in a standardized region of the genome (Kress et al., 2015; Waugh, 2007). Differentiation of species from one another is possible due to intraspecific variation in these sequences being inferior to interspecific variation (Waugh, 2007). DNA sequences which taxa have been confirmed, are used to build DNA barcode libraries, which allow the identification of unknown species by matching the DNA barcode of the unknown sample to the barcode library (Kress et al., 2015).

In the context of dietary analysis, prey DNA is sequenced from samples such as predator gut contents or feces, a reference database (i.e. DNA barcode library) is then used for species identification, which in turn can reveal the ecosystem's exact food webs (Rytkönen et al. 2019). This literature review is going to focus on the application of DNA metabarcoding in animal dietary studies. The objectives of this work are to describe the method of DNA metabarcoding as well as discussing the critical aspects of it and some other applications for this method.

2. Description of DNA metabarcoding

In general, DNA barcoding methods are based on four steps: 1. DNA is isolated from a sample. 2. The target DNA barcode region is amplified using polymerase chain reaction method (PCR). 3. The PCR product is sequenced. 4. The resulting sequences are compared to reference databases of known DNA barcodes to identify the corresponding species ("DNA Barcoding", 2019). The benefit of this method is the possibility for high taxonomic resolution identifications of many variable samples, demanding less time and expertise than morphological identification (McClenaghan et al., 2019).

DNA metabarcoding differs from DNA barcoding on the fact that DNA is sequenced from multiple species in a single sample, also called bulk samples (McClenaghan et al., 2019; McGee et al., 2019). The essential elements of DNA metabarcoding include the application of technologies such as high-throughput DNA sequencing; HTS (also known as next-generation sequencing; NGS), DNA barcodes, DNA barcode databases, as well as statistical analysis (Swift et al., 2018).

In the context of studies focusing on animal diet, the method of DNA metabarcoding is based on the following phases:

1. Collection of environmental samples. 2. Preservation of samples. 3. DNA extraction from the samples. 4. DNA barcode amplification using PCR. 5. Sequencing of the products. 6. Data analysis including: OTU-clustering and comparison to reference libraries (Creer et al., 2016).

2.1 Sample collection

When designing a dietary analysis study utilizing DNA metabarcoding, the biology and ecology of the studied system needs to be taken into consideration, as this system in question will frame processes such as the sampling technique, the type of sample and how the sample is preserved, processed and analyzed (Alberdi at al., 2019).

2.1.1 Type of sample-Environmental DNA

In a general sense, environmental DNA is defined as a mix of DNA, possibly degraded, originating from multiple different organisms. Environmental samples, for instance soil, water or air samples are usually used to extract eDNA (Alberdi et al., 2019). eDNA can also be extracted from mucus, skin, saliva, sperm, secretions, blood, urine, roots, leaves, pollen, fruit, larger organisms' rotting bodies or even from entire microorganisms (Ruppert et al., 2019). Feces and gut content are also considered as environmental samples, since they consist of a mix of DNA originating from various organisms, partly digested and deteriorated because of exposure to the environment (Alberdi et al., 2019).

Diet analyses have become faster and more economical with the help of HTS techniques that allow the identification of species using bulk mixture of prey samples (Rytkönen et al., 2019). In dietary studies, HTS is used on environmental DNA (eDNA) samples extracted from gut content or feces (Alberdi et al., 2019). Ruppert et al. (2019) point out that degradation of eDNA is one factor limiting the extent of studies using it. Environmental conditions, length of exposure to the outer environment or the media through which DNA travels highly affect the way trends of species and communities can be interpreted from such a sample. Frequently, only short segments of genetic material are left, this being particularly true in warm and tropical areas.

2.1.2 Sampling approaches

Alberdi et al. (2019) explain that stomach content is usually isolated through dissection, regurgitation, or digestion of the whole prey. Stomach content, when collected right after feeding, contains DNA that is less degraded than in feces. However, the collection process of such samples is invasive for the animal and is often limited by ethical and legal issues. In contrast, fecal samples are obtained from the environment or directly from the animal and are a popular sampling method as it is non-invasive. Deagle et al. (2019) raise an additional concern for the usage of stomach content for dietary analysis coming from the state of digestion of the recovered material. It might be in various state of digestion and in

consequence, this material cannot be used as well as fecal matter for assessing the content of the sample in a quantitative way.

King et al. (2008) recommend collecting fresh samples of feces, as they are more reliable. Old feces samples can result in false-negatives when it comes to prey that were actually consumed. The risk for contamination is also high when there is contact with a substrate, therefore, such samples should be avoided.

2.2 Sample preservation

Samples should be immediately preserved in order to prevent degradation from abiotic factors such as pH, temperature and light. Appropriate preservation methods are drying, freezing by -20 °C temperatures or the use of 100% ethanol (Ruppert et al., 2019). Storage of fecal matter can also include commercial kits like RNAlater, storage in dimethyl sulphomide (DMSO) salt solution, two-step storage with ethanol and silica or storage in commercial kits buffers (King et al., 2008). Regardless of the method used, preservation and storage of all samples in the exact same conditions is highly recommended to avoid technical distortions (Alberdi et al., 2019).

2.3 DNA extraction

The quality and quantity of the extracted DNA depends essentially on the DNA extraction method. eDNA extraction is generally accomplished using commercial kits providing fast and standardized solutions, such as PowerFecal/Soil DNA Kit or Qiagen QIAmp DNA Stool Mini Kit (Alberdi et al., 2019). The choice of the kit depends on the targeted taxa, as this might affect the efficiency of the kit (Ruppert et al., 2019). Efficiency might be altered because of the predator's intestinal characteristics or by the molecular composition of the prey. For instance, DNA extraction from bird and reptile feces or stomach content is less clear-cut than from mammal samples, due to bird and reptile feces containing uric acid (Alberdi et al., 2019).

Processing of eDNA should take place in a clean lab, where personnel and the equipment are correctly decontaminated in order to avoid contamination of samples (Ruppert et al., 2019). The use of sterile, DNA-free instruments and aseptic practices in the laboratory are an efficient way to prevent contamination. Possible contamination can be detected with the help of negative controls during DNA extraction (King et al., 2008).

2.4 DNA barcode amplification

In metabarcoding, PCR amplification targets only DNA barcodes (i.e. markers) which are taxonomically significant (Alberdi et al., 2019). These targeted sites need to be sufficiently conserved in order to reduce taxonomic bias (Clarke et al., 2014). In animal studies, the mitochondrial cytochrome C oxidase subunit I gene (COI) is employed as a DNA marker. However, other regions of the genome, such as nuclear 16S/18S ribosomal RNA genes or 12S mitochondrial DNA have been recognized to be, in some cases, more suitable for metabarcoding of specific taxa (Creer et al., 2016). Yet COI has far more available reference sequences (for over 500000 animal species) than any other gene regions, making it, at this moment, the better gene region for metabarcoding (Braukmann et al., 2019). The choice of DNA markers will be made depending on whether the need is to identify a precise species or a group of species (e.g. family, genus, or a whole order). Target-specific primers can then be designed from those DNA markers. Libraries of readily available primers targeting markers of particular prey taxa exist as well (King et al., 2008).

PCR is then used for the amplification of those targeted DNA regions which in turn produces numerous copies of them (Creer et al., 2016). Metabarcoding of eDNA requires short enough primers (under 300 base pair) in order to be able to amplify degraded samples such as gut and fecal samples. (Ruppert et al., 2019; King et al., 2008). The efficiency of PCR will depend on the type of primers used, the quality of the extracted DNA, PCR reagents and cycle conditions (King et al., 2008).

2.5 Sequencing

In the context of dietary studies, the use of high-throughput sequencing-based techniques has quickly increased in popularity (Alberdi et al., 2019). HTS is a modern sequencing technology making parallel DNA sequencing of millions of DNA fragments faster and cheaper (Creer et al., 2016). Many HTS platforms are nowadays available, however the best performing are Ion Torrent and Illumina sequencing (Ruppert et al., 2019; Rytkönen et al., 2019).

Before sequencing, PCR products (i.e. DNA sequences) must go through a process called library preparation or multiplexing (Alberdi et al., 2019; Ruppert et al., 2019). In this process, primers are tagged with short nucleotide sequences, also referred to as tags or indices, helping to relate the sequenced DNA to its PCR replicate and sample. These indices can be added through PCR or without PCR, where in this case, indices carrying sequencing adapters are attached to the sequences (Alberdi et al., 2019). HTS produces millions of reads, all describing the genetic code of the sequenced DNA (Ruppert et al., 2019).

2.6 Data analysis

For data analysis, the resulting sequences of HTS are first checked for errors, if such are found, these erroneous sequences are discarded. The remaining sequences are then classified according to their indices (Pompanon et al., 2012). The sequences are usually clustered into operational taxonomic units (OTUs), which are used for differentiating species or taxa using the similarities of genetic code (Ruppert et al., 2019).

Reliable taxonomic assignment of OTUs is a fundamental step in trophic interaction studies. This is achieved by matching the results of HTS to reference databases (Alberdi et al., 2019). The integrity and accuracy of the sequences found in the reference database will affect the quality of identification of taxa or species (Kress et al., 2015). Other factors, such as the length of the HTS product sequence and the similarity between the produced sequence and reference sequence influence the reliability of the taxonomic assignment. A short DNA sequence (e.g. 50 base pair or bp), will with a higher probability match perfectly a sequence in the reference database than a long sequence (e.g. 800bp) (Alberdi et al., 2019).

Classification algorithms, phylogeny, or species delimitation (allocation of sequences is based on the differences, rather than similarities, to known sequences) are often considered for species identification with OTUs as well (Ruppert et al., 2019).

3. Reference databases

Prey taxa identification is the most efficient when DNA sequences extracted from gut content and feces are compared to a reference DNA barcode library. A DNA barcode library is defined as a compilation of DNA barcode sequences of species with verified taxonomic origin (Kress et al., 2015). There are several public databases such as for example Barcode of Life Data Systems (BOLD) or GenBank (Kress et al., 2015; Waugh, 2007). Particularly in Finland can be found the Finnish Barcode of Life project (FinBOL), which is an extensive national reference database (Rytkönen et al., 2019). Customized databases may also be used in some cases when, for instance, a good collection of potential food DNA is not available in the database (Pompanon et al., 2012) or to be able to limit the scope to only relevant species found in a certain habitat (McClenaghan et al., 2019).

3.1 BOLD

BOLD (www.boldsystems.org) is a website managed by the Canadian Centre for DNA Barcoding which allows the uploading of sequences of the COI gene for species-level identification but also allows various types of data analysis for uploaded sequences (Waugh, 2007). With BOLD, animal, fungal and plant identification is possible using their identification engine which include among others, species level barcode records and public records (“Identification engine”, 2019).

BOLD has four different databases (“Databases”, 2019):

1. Public Data Portal that permits searching all 1.3M public records found in BOLD for data retrieval.

2. Bin Database, a database of Barcode Index Numbers (BINs) which provides a way to quickly validate and use barcode data in the case of lacking or unverified taxonomic data.
3. Publications database, which is a compilation of publications using barcode records.
4. Primer database, a record of primers utilized for generating barcode sequences.

3.2 FinBOL

FinBOL is one of the many national or regional projects creating DNA barcode databases ("What is DNA barcoding", 2019). The objective of FinBOL is to build a DNA barcode reference library including all multicellular organisms found in Finland. It has the ambition to enable automatic and unambiguous identification of any Finnish species for anyone. The data gained by FinBOL is eventually stored in BOLD and through that will be made public for species identification ("FinBOL", 2019). Therefore, the data is maintained by BOLD (Rytkönen et al., 2019).

3.3 GenBank

GenBank is a public nucleotide sequence database including sequences for about 260000 identified species. GenBank is managed by the National Center for Biotechnology Information (NCBI) in the USA (Benson et al., 2013).

Sequence data is originating from scientists, genome survey sequence, whole-genome shotgun and other high-throughput data generated in sequencing centers. In GenBank, sequence records are placed to a designated division depending on the source organism taxonomy or the sequencing technique used to generate data. There are 12 taxonomic divisions (e.g. bacteria, plant, primates, environmental samples) and 8 functional divisions (e.g. high-throughput genomic, genome survey sequences, transcriptome shotgun data) (Benson et al., 2013).

3.4 Customized databases

In cases of dietary assessment studies, when DNA sequences for possible diet component are missing in the database, the creation of a customized DNA barcode library might be necessary (Kress et al., 2015). In such cases, a set of DNA sequences is generated especially for the need of the study (Pompanon et al., 2012). This helps to avoid errors resulting from faulty taxonomic assignation due to missing data in the database (Kress et al., 2015).

Another application for the use of a customized reference database is found in a study from McClenaghan et al. (2019), where, in addition to using BOLD database, the researchers also used their own customized database in a dietary analysis study as they were interested only in prey items found in a certain habitat. They were then able to associate prey items in the diet with their availability in the habitat in question. Though the process of creating a customized reference database is time consuming (over 60 000 insect specimens were collected over two breeding seasons), it seems that utilizing it in conjunction with an international database such as BOLD can produce good results.

4. Critical aspects of metabarcoding

Although DNA metabarcoding is considered as one of the methods of choice to study biodiversity in environmental samples and appears as a clear-cut method, it also has its own share of challenges and limitations (Alberdi et al., 2018). These should be acknowledged in order to design reliable dietary studies, keeping out biological and technical distortion factors that might lead to the misrepresentation of diversity and taxonomic composition (Alberdi et al., 2019). The most important limitations and biases related to DNA metabarcoding are going to be described in the following sections.

4.1 Biological distortion factors

When designing a dietary study, Alberdi et al. (2019) recommend good knowledge of the biology of the studied species, in order to know which individuals to include in the study.

For instance, feeding behavior can vary depending on which stage of life cycle the individual is, as in some cases the individual might be in a stage where it does not feed. Digestion capability of the animal can also differ between species, lineages or developmental stages.

When considering feeding behavior of the predator, a source of distortion may arise from the fact that the predator might eat the whole prey, or only bits of it, which alters the quantity of prey DNA being finally digested (Alberdi et al., 2019). Another possible cause for bias in animal diet assessment is secondary predation, especially with omnivores, for instance when a predator feeds on another predator that has just ingested some food (De Barba et al., 2014).

Alberdi et al. (2019) raise the question of spatial and temporal scale, due to how common temporal and geographic variations are in trophic interactions. For example, depending on how general or specific the study aims to be, biased results can be avoided by collecting samples over a large spatial range or, in the contrary, on a local range. In the case of temporal scale, long-term dietary information can be obtained if the sampling process is repeated for a longer extent of time. This is due to the quality of fecal samples, as it has been shown that the amount of prey DNA is extremely low or almost inexistent when the prey has been consumed more than a few days before sample collection.

Environment is also considered as a biological distortion in eDNA samples, as it is known that temperature, pH, water and exposure to UV or other radiations accelerates considerably DNA degradation, thus decreasing PCR amplification success of fecal samples that were exposed to such conditions (Alberdi et al., 2019; McInnes et al., 2017). Exposure of feces to a certain type of substrate, such as dirt, reduces food DNA detectability and especially contamination from non-food DNA (e.g. insects, parasites and fungi) is common with such a substrate (McInnes et al., 2017).

De Sousa et al. (2019) explain that diverse environments such as aquatic or terrestrial environments bring their own challenges especially when it comes to sampling. For instance, access to feces of aquatic organisms is very restricted and sampling is mainly opportunistic, which makes samples of gut and stomach content highly valuable to study diet in aquatic

systems. However, prey items in such samples are often at different digestion states, which affects the amount of DNA that could be recovered.

4.2 Technical distortion factors

Sample collection being done, eDNA undergoes procedures including DNA extraction, PCR amplification, sequencing preparations, sequencing and bioinformatic procedures. All those steps are prone to technical distortion, from which arise the many challenges, limitations and biases related to DNA metabarcoding. Fortunately, procedural improvements can and are constantly being done, thus minimizing and controlling those biases (Alberdi et al., 2019).

4.2.1 Biases related to DNA extraction

Alberdi et al. (2019) discuss the challenges related to contamination. Sample contamination is a common issue faced anytime during the sample processing workflow. There are two types of contamination: external and cross-contamination. External contamination might occur in the field during sample collection or in the laboratory during sample processing. In the laboratory, cross-contamination between samples arises easily in the situation where large batches of sample are being processed, meaning that multiple tubes holding different samples are open at the same time with low spatial separation. Partly due to those contamination risks, it is crucial to include negative or positive controls throughout the whole workflow from sampling to final results. Reactions with no DNA in them are called negative controls. Positive controls include a single or a mixture of DNA extracts from known taxa (i.e. mock communities).

Alberdi et al. (2019) continues with several factors that may affect the relative quantity of DNA retrieved during DNA extraction. One of them is the characteristics of prey tissues, as for example, sclerotized animals (e.g. chitinous insects) cause more problem than soft-body species during DNA extraction. Bird and reptile feces are more problematic than mammal

feces, due to their uric acid content. Other enzymatic inhibitors, such as pectin and xylan can also cause difficulties if they are found in large amounts in the samples.

4.2.2 Biases related to PCR amplification

Good understanding of the role of predator-prey interactions in food webs is still challenging due to problems related to accurate and efficient detection of the intricate diversity of food consumed by the animal in the field (de Sousa et al., 2019). Such challenges are faced especially in dietary studies of generalists, herbivores and especially omnivores, where simultaneous identification of different groups of organisms is necessary (De Barba et al., 2014).

Amplification of predator DNA is a matter that needs to be taken into consideration in any dietary study (Alberdi et al., 2018). For example, amplification of the predator DNA can be problematic when primers target vertebrate prey, but the predator is also a vertebrate. To prevent the amplification of the predator DNA itself, De Barba et al. (2014), designed blocking oligonucleotides, also called blocking primers (McInnes et al., 2017), targeting mammalian sequences, which can also be used to block DNA amplification of human sequences in case of possible human contamination.

One fundamental decision in a metabarcoding study involves marker region and primer selection (Alberdi et al., 2018). COI is the most extensively used marker in dietary studies, although its use is not always unanimous. The question is whether to use a single primer targeting a single molecular marker (e.g. COI) or to combine multiple primer sets to amplify the marker in multiple taxonomic groups. Arising from this issue, other markers (such as 16S rRNA) are also used in dietary studies, which, in some cases, results in the recovery of a more extensive taxonomic range of prey items and less biased results than with COI marker (Alberdi et al., 2018; de Sousa et al., 2019). The combination of multiple primer sets is recommended since this method decreases taxonomic biases, in addition to taxonomic coverage being increased (Alberdi et al., 2018). As a result, this protocol provides a more exhaustive picture of diversity in the sample; however, it is more expensive and time-consuming (Alberdi et al., 2018; Ruppert et al., 2019).

An additional issue to consider when designing primers is PCR replicates (Ruppert et al., 2019). PCR replicates are repeated measurements of one same sample and are used to counter PCR stochasticity (Alberdi et al., 2019). Species detection is increased as well as the likelihood of false negatives is decreased when using several PCR replicates. The number of replicates needed can be estimated by considering detection probabilities, the objectives of the study, sequencing depth, primer choice and sequencing platform (Ruppert et al., 2019)

4.2.3 Biases related to sequencing

Alberdi et al. (2018) discuss that in order to provide an accurate picture of the diversity of environmental samples, a minimum sequencing depth (i.e. number of reads per library or number of reads per PCR replicates) is required. As sequencing depth increases, the detected diversity increases as well; nevertheless, exceedingly high levels of sequencing depth result in increased OTU dissimilarity between PCR replicates, which in consequence identifies falsely high diversity. Therefore, to improve diversity detection, primer choice, PCR optimization and inclusion of PCR replicates for every sample may be necessary aside from increasing sequencing depth.

4.2.5 Biases related to bioinformatic procedures

After sequencing, erroneous sequences should be removed. Several ways to do that exist, one of them being through mathematical approaches, when persistent error patterns are found (Alberdi et al., 2019). Clustering sequences into OTUs based on a similarity threshold is another way to be able to remove those erroneous sequences. Most metabarcoding studies use a threshold of 97% similarity. The risk is that such an approach generates either over- or underestimated species counts. No matter which approach is used to discard faulty sequences, if done too extremely, rare species might be discarded from the dataset, while if done too cautiously, errors might be represented as true diversity (Alberdi et al., 2018).

After the taxonomic assignment of sequences, the way the resulting data is interpreted can drastically influence the final results. The data can be interpreted either in a quantitative

way, meaning that proportions of different taxa in each sample is looked at, or in a qualitative way, in other words, looking at the presence or absence of taxa in a sample (Alberdi et al., 2019). De Barba et al. (2014) specify that most studies acknowledge that the data retrieved from HTS can be regarded as semi-quantitative and can therefore be used for comparative purposes. However, it is still difficult to accurately associate sequence counts information to the proportions of various food, especially in a complex dietary sample, such as of an omnivore. Multiple food items are found from such a sample, and their relative weight at sample level cannot be established, therefore, such data can only be considered as qualitative data.

Deagle et al. (2019) point out that in dietary analysis based on occurrence data (i.e. presence/absence), the importance of rare food taxa is mostly overestimated, leaving the food taxa eaten in abundant amounts underestimated. Relative read abundance (RRA) information may help to get a more accurate look at the diet at the population level, especially if individual fecal samples have several food taxa and this same food taxa are found in many samples. Both RRA and occurrence data approaches are more reliable in situations where the mean number of food taxa in samples is small. Morphological analysis may be useful to cross-validate RRA data. In addition, comparing results with other diet analysis methods, such stable isotope analyses, proved to be efficient to interpret sequence counts.

5. Some other applications of metabarcoding

A very appealing application of eDNA metabarcoding and probably one of the most researched one would be ecosystem and biodiversity monitoring (Ruppert et al., 2019). A more extensive understanding of global biodiversity is needed to better demonstrate the global consequences of a changing climate and human caused disruptions on earth’s ecosystems. DNA metabarcoding offers a cost-effective, easily implemented solution to tackle such challenges (McGee et al., 2019). Traditional methods for biodiversity monitoring are often destructive, costly and thorough taxonomic expertise is usually needed as well. The advantages of eDNA surveys are that they do not disturb the ecosystem, less effort is

demanded, they are sensitive for species detection and can be applied in regions where traditional surveys are impractical (Holman et al., 2019; Ruppert et al., 2019).

5.1 A case study

A study by Holman et al. (2019) applied DNA metabarcoding to detect introduced and resident marine species using eDNA samples of sediment and water. The focus of this study is on how different types of environmental samples might influence species detectability and how eDNA can be used for early detection of non-indigenous species (NIS).

In this study from Holman et al. (2019), eDNA metabarcoding of COI and 18S rRNA genes from sediment and water samples in artificial coastal sites throughout the United Kingdom were used to compare community composition between those samples. Four marinas were selected as sampling sites, those sites have been previously surveyed using rapid assessment (RA) surveying methods. Seawater and subtidal sediments sample were taken from 24 randomly selected sampling points. Seawater samples were collected by taking 50 ml of water from 10 cm below the surface using a sterile syringe and then filtered using a polyethersulfone filter. For the subtidal sediment samples, a sediment core of 600mm high and 60mm diameter was collected using a sediment corer. Subsamples of 10-20g of sediment from the top 2cm part of the core were taken with a sterile spatula, this in order to avoid collecting sediment from the sides of the core. DNA was extracted from those samples using DNA extraction kits and two sets of primers were chosen for metabarcoding. Those sets of primers were specifically chosen to be able to characterize extensively marine metazoic diversity. Sequencing was done using Illumina MiSeq instrument. Taxonomic assignment was done using MIDORI database for COI data and SILVA database for the 18S rRNA data. MIDORI is a database of unknown metazoan mitochondrial encoded gene sequences (Leray et al., 2018) and SILVA is a database of rRNA gene sequences from *Bacteria*, *Archaea* and *Eukaryota* domains (Quast et al., 2012). To identify NIS, only the taxonomic assignments of the COI gene were used, since the 18S region does not provide good resolution for species level taxonomic assignment (Holman et al., 2019).

Holman et al. (2019) found out in their study that in all sampled sites, sediment samples showed higher species richness than in water samples, demonstrating that the measured community composition is affected by the type of environmental sample. In addition, their eDNA metabarcoding data and the data from previous RA surveys correlated, indicating that various biodiversity assessment methods can complete each other. Moreover, throughout the four sites, 18 recently introduced NIS to the study region and 24 species documented in other regions as NIS were identified, showing that eDNA metabarcoding can be a powerful tool to detect NIS early. They were also able to confirm the eDNA detection of one NIS, the Asian date mussel (*Arcuatula senhousia*), with surveys targeted locally. When invasive, this species can have dramatic effects on benthic biodiversity. Finally, they advise to combine both traditional survey methods with eDNA methods, since NIS can be missed when solely eDNA surveys are used, as well as rare species can remain undetected using only traditional methods.

The authors raise also the issue of eDNA degradation in sediments and how it might affect their observations, although they processed only the uppermost part of the sampled core of the sediment, so to obtain a description of species composition of this present-day. Further research on eDNA deposition and degradation mechanisms in sea sediments is needed to be then able to temporally contextualize sediment samples (Holman et al., 2019).

The results of this study validate the efficiency of eDNA metabarcoding for biodiversity monitoring and underline that such molecular technologies applied in routine monitoring surveys have remarkable prospects (Holman et al., 2019). In a larger scale, biodiversity assessments unveiling global ecosystem changes can be greatly improved when DNA metabarcoding is implemented with ecological network analyses and machine learning algorithms, enabling automation of the process (Holman et al., 2019; McGee et al., 2019).

6. Discussion and conclusion

Our understanding of how ecosystems function is essentially based on the ability to describe trophic interactions and food webs. Food web dynamics are mostly explained with predator-prey interactions (de Sousa et al., 2019). Those interactions can be better understood when predator diets and their reaction to changes within available preys are known (McClenaghan et al., 2019). Many recent studies focusing on trophic interactions utilize metabarcoding as a method enabling the identification of prey remains in fecal or stomach contents (Deagle et al., 2019). Non-invasive diet analysis was probably one of the earliest applications of metabarcoding; however, beyond the purpose of characterizing animal diets, metabarcoding in this context has been used to research among other things: resource partitioning, dietary overlap and competition, habitat usage or species interaction (Ruppert et al., 2019).

Regardless of these remarkable promises, latest studies have brought to light diverse biological and technical factors that may distort taxonomic composition signal and diversity in metabarcoding dietary studies (Alberdi et al., 2019). One of the most important challenges being perhaps how to interpret count data in such studies (Deagle et al., 2019). In addition to encouraging researchers to investigate further these challenges, Deagle et al. (2019), suggest that using other methods of diet analysis to compare results of metabarcoding data, such as stable isotope analysis or traditional morphological analysis, has proven to be useful. Interestingly, another genetic approach applying eDNA methodology, quantitative PCR (qPCR), is a method that can be used to quantify the relative DNA abundance (and thus to a certain extent, the relative species abundance) and to increase detection sensitivity (Qu & Stewart, 2019). This approach has been recently used to detect specific species from eDNA, such as the endangered Gouldian finch, *Erythrura gouldiae* (Day et al., 2019).

To conclude, the data derived from DNA metabarcoding dietary studies has considerable utility in numerous conservation and ecological applications. Citing here only few of them: biological pest control or monitoring of diet shifts due to the modification of natural habitats under climate change or human activities development (De Barba et al., 2014). Finally, with proper scientific collaboration and coordination, the potential of DNA metabarcoding appears almost limitless (Ruppert et al., 2019).

References

Databases. (2019, November 11). Retrieved from http://www.boldsystems.org/index.php/databases

DNA barcoding. (2019, October 10). Retrieved from https://ibol.org/about/dna-barcoding/

FinBOL. (2019, November 03). Retrieved from https://en.finbol.org/finbol

Identification engine. (2019, November 03). Retrieved from http://www.boldsystems.org/index.php/IDS_OpenIdEngine

What is DNA barcoding. (2019, November 03). Retrieved from https://en.finbol.org/dna-viivakoodaus

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.

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.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. (2013). GenBank. *Nucleic Acid Research*, 41, D36-D42.

Braukmann, T.W., Ivanova, N.V., Prosser, S.W., Elbrecht, V., Steinke, D., Ratnasingham, S., de Waard, J.R., Sones, J.E., Zakharov, E.V. and Hebert, P.D. (2019). Metabarcoding a diverse arthropod mock community. *Molecular ecology resources*, 19(3), 711-727.

Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular ecology resources*, 14(6), 1160-1170.

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. *Methods in Ecology and Evolution*, 7(9), 1008-1018.

Day, K., Campbell, H., Fisher, A., Gibb, K., Hill, B., Rose, A., & Jarman, S. N. (2019). Development and validation of an environmental DNA test for the endangered Gouldian finch. *Endangered Species Research*, 40, 171-182.

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.

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

de Sousa, L. L., Silva, S. M., & Xavier, R. (2019). DNA metabarcoding in diet studies: Unveiling ecological aspects in aquatic and terrestrial ecosystems. *Environmental DNA*.

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2019). Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. *Scientific reports*, 9(1), 1-10.

King, R. A., Read, D. S., Traugott, M., & Symondson, W. O. C. (2008). INVITED REVIEW: Molecular analysis of predation: a review of best practice for DNA-based approaches. *Molecular ecology*, *17*(4), 947-963.

Kress, W. J., García-Robledo, C., Uriarte, M., & Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in ecology & evolution*, 30(1), 25-35.

Leray, M., Ho, S. L., Lin, I. J., & Machida, R. J. (2018). MIDORI server: a webserver for taxonomic assignment of unknown metazoan mitochondrial-encoded sequences using a curated database. *Bioinformatics*, 34(21), 3753-3754.

McClenaghan, B., Nol, E., & Kerr, K. C. (2019). DNA metabarcoding reveals the broad and flexible diet of a declining aerial insectivore. *The Auk: Ornithological Advances*, 136(1), uky003.

McGee, K. M., Robinson, C., & Hajibabaei, M. (2019). Gaps in DNA-Based Biomonitoring Across the Globe. *Frontiers in Ecology and Evolution*, 7, 337.

McInnes, J. C., Alderman, R., Deagle, B. E., Lea, M. A., Raymond, B., & Jarman, S. N. (2017). Optimised scat collection protocols for dietary DNA metabarcoding in vertebrates. *Methods in Ecology and Evolution*, 8(2), 192-202.

Pompanon, F., Deagle, B. E., Symondson, W. O., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: diet assessment using next generation sequencing. *Molecular ecology*, 21(8), 1931-1950.

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, e00547.

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.

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.

Qu, C., & Stewart, K. A. (2019). Evaluating monitoring options for conservation: comparing traditional and environmental DNA tools for a critically endangered mammal. *The Science of Nature*, 106(3-4), 9.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F.O., (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, 41(D1), D590-D596.

Waugh, J. (2007). DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays*, 29(2), 188-197.