Stockholm University

Department of Zoology

Systematics and Evolution

No. 2017:3

Animal metabarcoding: Methods and Applications

Author:
Daniel Marquina

Supervisor:
Fredrik Ronquist
Co-supervisors:
Johannes Bergsten
Niclas Gyllenstrand

Stockholm University

Arbetets titel:

Animal metabarcoding: Methods and Applications

Författarens namn: **Daniel Marquina**

Handledare: **Fredrik Ronquist, Johannes Bergsten, Niclas Gyllenstrand**

Forskningsämne: **Zoologisk Systematik**

Granskningskommitté: **Fredrik Ronquist, Johannes Bergsten, Niclas Gyllenstrand, Anders Andersson, Birgitta Tullberg**

Animal metabarcoding: Methods and Applications

Daniel Marquina Hernández

Background

'Environmental DNA' (eDNA) refers to DNA that can be extracted from samples without any kind of sorting or isolation of the organisms present in the sample (Taberlet *et al.* 2012). Although the term eDNA *sensu stricto* is limited to the free genetic material found in environmental samples (water, soil, faeces), sometimes eDNA is also used to refer to the DNA extracted from a homogenized sample of organisms from a general collecting method (pitfall traps, malaise traps, plankton net, etc), the extracellular DNA present in the preservative ethanol of such a sample, or even the DNA of prey present in a spider's cobweb (Hajibabaei *et al.* 2012, Gibson *et al.* 2015, Xu *et al.* 2015). The definition is not clear, and, depending on the scope, diet analysis can be considered within the field of eDNA if the purpose is not to assess what an organism is eating, but to know the species present in a certain habitat through their micropredators (Schnell *et al.* 2012). This reflects nothing but the advance in developing techniques of, non-invasive, non-destructive eDNA sampling.

It was microbiologists who first started using eDNA, in the early 2000s. The appearance of massive parallel sequencing techniques (MPS) – also called Next Generation Sequencing (NGS) or High Throughput Sequencing (HTS) – made it possible to analyze the genetics of uncultivable microorganisms. What was afterwards called 'metagenomics' originally consisted in sequencing the whole genomes of the bacteria present in soil and water samples (Liles *et al.* 2003, Venter *et al.* 2004, Tringe *et al.* 2005), but the potential of this technique was later applied to other fields,

such as medicine and bacterially mediated processes of human interest (Qin *et al.* 2010, Lücker *et al*. 2010).

Although there are some examples of real metagenomic studies focused on eukaryotic organisms (i.e. sequencing of the whole mitochondrial genomes of a pooled sample of extracted DNA from different species; this will be discussed later), the vast majority of such studies are based on 'DNA barcoding'. DNA barcoding consists of identification of species through sequencing of a single genetic marker, and comparison of that sequence with a reference library (Hebert *et al.* 2003). This means that each species would be represented by a unique DNA sequence or 'DNA barcode', shared between all specimens of the species (under a similarity threshold that can vary depending on the taxonomic group).

There are two main approaches to barcoding of environmental DNA: species-specific and multispecific or multitaxa (Valentini *et al.* 2015, Furlan *et al.* 2015). The first aims to detect a single species in the sample, typically using PCR amplification with species-specific primers, while the latter aims to detect either every organism present in the sample (Drummond *et al.* 2015, Brannock & Halanych 2015) or, more commonly, to detect the presence of every species of a certain taxonomic group (Dell'Anno *et al.* 2015, Valentini *et al.* 2015, Port *et al.* 2015, Hajibabaei *et al.* 2011). The multitaxa barcoding of eDNA samples is known as 'metabarcoding', a quite recent term (Pompanon *et al.* 2011, Riaz *et al.* 2011, Taberlet *et al.* 2012).

Barcoding theory and criticism

Hebert *et al.* (2003), analyzing a large set of animal taxa, showed that 98 % of congeneric species pairs (two species of the same genus) exhibit more than 2 %

sequence divergence[1] (average of 11.3 %) in the 658 bp long fragment at the 5' end of the cytochrome oxidase subunit I (hereafter mentioned as COI) tipically amplified by primers designed by Folmer *et al.* (1994), and that individuals of the same species showed a sequence divergence that rarely exceeds 2 %. Thus, their results indicated that the distance between sequences of different species is much higher than the distance between sequences belonging to individuals of the same species (Fig. 1). On average, the genetic distance between species appears to be an order of magnitude (10 times) higher than the distance within species (Hebert *et al.* 2004, Barret & Hebert 2005). Plotting the frequency of the genetic distances between individuals of the same or different species, one would expect to observe a region with very low frequencies between the intraspecific and interspecific comparisons, a phenomenon that has been called the 'barcoding gap' (Meyer & Paulay 2005) (Fig. 1). With DNA barcoding, then, two individuals can be considered to belong to the same species, or set apart in two different species, depending on whether the genetic distance between them falls into the intraspecific or interspecific region of distance values.

Some authors, however, doubt the adequacy of the use of the barcoding gap as an infallible method for species delimitation, and even consider its existence as an artifact of deficient taxon sampling within the genus. This is the case of Wiemers & Fiedler (2007), who found that there was an upper limit to the intraspecific genetic divergence in butterflies of the family Lycaenidae, but not a lower limit in the congeneric distances, existing an overlap between the two in some genera. Nowadays, DNA barcoding has been widely accepted although some corrections or considerations have been added to the initial method. Bergsten *et al*. (2012) found that the barcoding gap was narrower or even inexistent between species far a part from

[1] Measured as genetic distance: $mismatching\ positions \times 100/total\ positions$.

each other than for geographically close heterospecifics; Meier *et al*. (2008) recommend the use the minimum interspecific and the maximum intraspecific divergences as the thresholds for establishing the limits of the barcoding gap, rather than their average, since with a more comprehensive taxon sample both measures would be re separate than they are in reality. S it has been said before, considering these and other taxon-specific considerations, DNA barcoding is today a useful and trusted tool in molecular systematics.

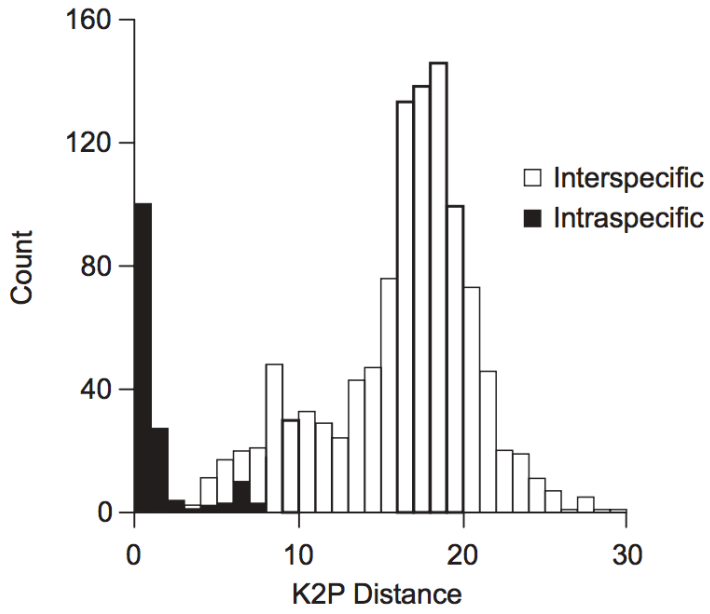

Fig. 1 Histogram of intraspecific and interspecific (congeneric) genetic divergence across 203 arachnid species. Divergences were calculated using Kimuras's two parameter (K2P) model. From Barret & Hebert (2005).

Field and laboratory protocols

As mentioned before, there are two main sources of eDNA, which are on the one hand actual tissue of the individuals present in the sample, and on the other hand medium in which the organisms were recently present and left traces of their DNA. These media are usually water, either freshwater or marine, and soil.

DNA from tissues. Examples of the first case can be found in many of the studies of the team of Mehrdad Hajibabaei at the University of Guelph (Canada). Insect samples are typically collected using a Malaise trap, a light trap or a net dragged through an aquatic (freshwater) transect. The sample is then mechanically homogenized, incubated with proteinase, and DNA is extracted using a commercial kit (Hajibabaei *et al.* 2011, Gibson *et al.* 2015). Sometimes, a pre-sorting of the sample is conducted in order to record the taxonomic composition, and only then homogenized (Shokralla *et al.* 2014, 2015, Gibson *et al.* 2014). In some of these examples, the sample is actually an artificial pool created by extracting small pieces of tissue (e.g. a leg) from each individual specimen and then pooling them before the extraction procedure. A different kind of pre-processing is exemplified by the study of Dell'Anno *et al*. (2015), which assessed the biodiversity of benthic nematodes from deep-sea samples by metabarcoding. The sample-processing consisted in freezing the sediment sample, sieving it through different sizes mesh net and separating the nematodes from all other meiofaunal organisms. Then, all nematodes from each sample were pooled together and total DNA was extracted in a similar procedure as in case of the above-mentioned insects.

A variant of the tissue-extracted eDNA involves genetic analysis of the intestinal content of organisms, or iDNA. Of course, the iDNA is affected by a certain degree of degradation by natural processes occurring the digestive tract, which challenges the laboratory protocol. However, the results can be quite valuable, both for biodiversity assessment, but especially for studying food webs (see special issue of *Molecular Ecology*; Symondson & Harwood 2014). Studies using this approach have been carried out with intestinal content from seals (Thomas *et al.* 2014, 2015), shrews and lizards (Brown *et al.* 2014), sea birds (Alonso *et al.* 2014), spiders (Welch

et al. 2014), and other invertebrates. Blood feeders are also a great source of iDNA. The iDNA can be used to study the specific prey of the blood feeders, but the blood feeders can also be used merely as 'sample collectors' to study biodiversity more broadly. Examples of the first case are found in mosquitoes (Townzen *et al.* 2008) or ticks (Gariepy *et al.* 2012), while leeches and hematophagous flies have been used to assess mammal biodiversity in a certain habitat in Vietnam and Côte d'Ivoire and Madagascar, respectively (Schnell *et al.* 2012, Calvignac-Spencer *et al.* 2013). Diet studies have also been carried out obtaining the DNA from the faeces of predators, bats being the most studied subjects (Clare *et al.* 2014, Emrich *et al.* 2014, Krüger *et al.* 2014). Excrements are also a source of DNA for other types of studies (e.g. population genetics of elusive mammals (Zhu *et al.* 2011)).

Shokralla *et al.* (2010) developed a method for extracting and amplifying DNA present in the ethanol that acts as a preservative in sample storage in order to keep the specimens intact for morphological examination. The method consisted of letting the ethanol evaporate at 56˚C and resuspend the DNA in molecular biology grade water. By doing this, they were able to amplify a short fragment of 130 bp of COI as well as the standard 658 bp barcoding region. Furthermore, they successfully applied the method to caterpillars found in small bottles of mescal (Mexican high-grading alcoholic beverage) (Fig. 2). Ethanol-DNA has been proven to be a good tool for metabarcoding analyses (Hajibabaei *et al.* 2012).

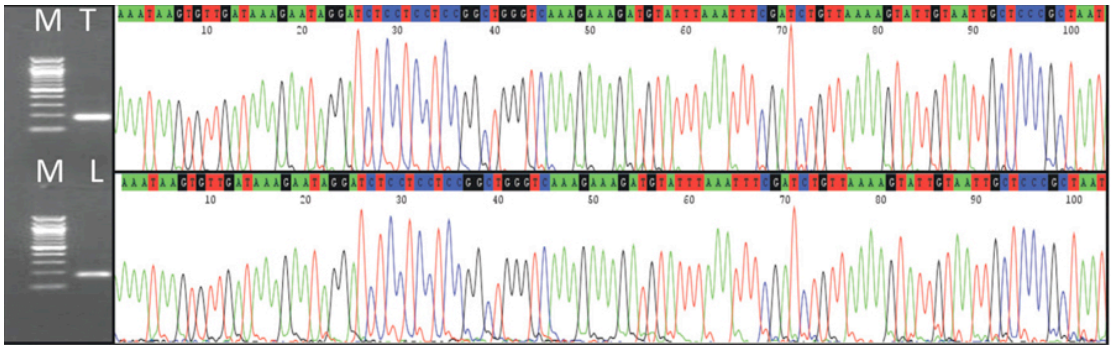

Fig. 2 DNA leaking from specimens in preservative liquids can be directly used for PCR and sequencing. (Left) Gels showing the 130 base–long amplicon from the cytochrome c oxidase 1 (CO1) gene from DNA extracted from caterpillar tissue specimen (T), compared with the same amplicon obtained directly from mescal liquid (L). M, molecular size marker. (Right) Corresponding sequencing electropherogram of the PCR amplicon shown at left.
From Shokralla *et al.* (2010).

DNA from water and soil. With regard to the second case, free DNA obtained directly from the environment, the most common is DNA present in water – either freshwater or seawater. Water is collected in large amounts (2-3 liters per sample) and taken to the laboratory. Valentini *et al.* (2016) tested a new methodology that involves filtering up to 100 L of water at the sampling site, but it is an exception in this type of studies. After collecting, the transport and handling to and in the laboratory must be carried with extreme caution to avoid DNA contamination. Since there is no physical proof in the sample of the species present at the sampling site, it is extremely difficult to control for post-sampling contamination that can distort and invalidate the results. Thus, the laboratory conditions are very strict and similar to those applied to handle ancient DNA (positive air pressure, UV treatment, full-body suits, etc.). Once in the laboratory, water samples are filtrated through standard, or specific commercial filter papers, in which the DNA is retained and afterwards extracted using a commercial kit (Furlan *et al.* 2015, Port *et al.* 2015). DNA can also be extracted from the water sample without filtering, using a density-based

centrifugation protocol, in which the DNA fraction can be obtained for further processing (Tréguier *et al.* 2014, Valentini *et al.* 2016).

Analysis of eDNA from water samples has been applied to the detection and monitoring of several animals such as crayfish, marine mammals, amphibians, fishes, and other vertebrates, in fresh and seawater, either species-specific or whole group detection (Tréguier *et al.* 2014, Foote *et al.* 2012, Biggs *et al.* 2015, Valentini *et al.* 2016, Furlan *et al.* 2015, Port *et al.* 2015).

Turner *et al.* (2015) studied eDNA from a hybrid of two introduced species of bigheaded Asian carp, *Hypophthalmichthys* spp.. They showed that the eDNA concentration of these species was up to 1800 higher in the river/pond sediments than in the water, and that the eDNA lasted without degradation (at least not enough to avoid amplification and identification) five times longer in the sediment than in the water. This can be explained by the sinking of the fish's faeces and other DNA-bearing particles that don't stay suspended in the water for long (Turner *et al.* 2014) (Fig. 3).

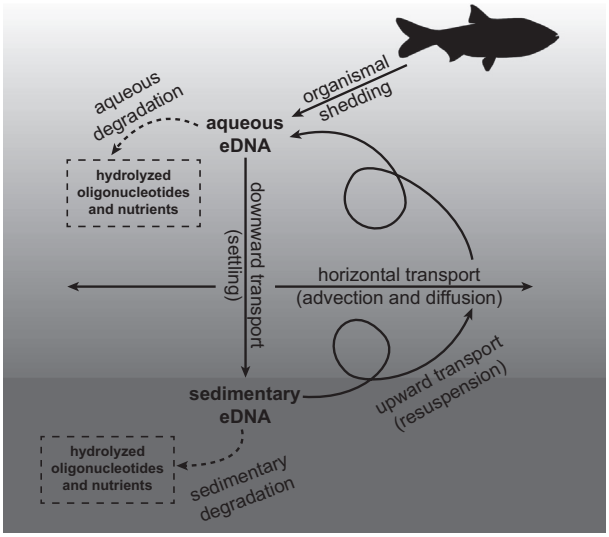

Fig. 3 Conceptual diagram of the processes affecting eDNA released into the water column by aquatic macrofauna. Because sedimentary eDNA persists longer than aqueous eDNA resuspension of sediments could influence the temporal resolution of inferences about organism presence made from aqueous eDNA. Horizontal transport of resuspended sediments could also influence the spatial resolution of inferences from aqueous eDNA. From Turner *et al.* (2015).

Finally, soil eDNA can be considered both tissue-based and medium-based eDNA: macrofauna leaves traces of DNA like fish would do in the water, while smaller animals can be caught whole in the extraction process. Soil eDNA has been used as an indicator of vertebrate biodiversity in zoological parks, to assess the Pleistocene Siberian fauna from permafrost samples, and the biodiversity of animals and other groups in Hauturu-O-Toi, a small island in New Zealand (Andersen *et al.* 2011, Epp *et al.* 2012, Drummond *et al.* 2015).

Sequencing and analyzing protocols

With the DNA extracted, it is time for sequencing and analyzing the data. There are two main approaches in this step, although one of them is much more widely adopted and the other needs further development. These approaches are amplicon-based and PCR-free, respectively.

PCR-free approach. PCR-free approach implies that there is no step consisting in amplifying the target DNA sequence prior to sequencing. Although at first looks counterproductive (the larger the amount of DNA, the easier is to sequence it), it appears as a solution to avoid certain flaws that go along with PCR-based or amplicon-based approaches (see below). The PCR-free approach is yet very poorly explored, and to date, only two methodologies have been proposed, and these have not been tested with totally satisfactory results. The first of them consists in shotgun sequencing of the mitochondrial genome of the whole mixed sample, and extracting the COI gene bioinformaticaly (Zhou *et al*. 2013). It presents a mayor flaw: it is (so far) impossible to obtain by shotgun sequencing the mitochondrial genome without sequencing the whole nuclear genome (that outnumbers the mitochondrial genome in

2000:1 bp). Thus, it generates a great amount of genetic information from the nucleus and other mitochondrial markers that is not essential for species identification. To overcome this problem, work has been done to optimize a step previous to the sequencing of mitochondrial enrichment. Using target-capture probes designed for several regions of the mitochondria, Liu *et al*. (2015) were able to increase 100 fold the proportion of mitochondrial versus nuclear DNA, and shotgun sequence almost the whole mitochondrial genome of 49 samples. Also using target capture, but with COI-specific probes, Shokralla *et al*. (2016, preprint) obtained COI sequences belonging to different insect orders from bulk samples (the previous attempts were made using mock communities samples), whit an effectiveness similar to metabarcoding process.

The big advantage of the PCR-free approach is that it avoids the PCR bias in detection of species associated with the amplicon approach (see below). However, right now its low cost-effectiveness makes it unattractive for researchers.

Amplicon-based approach. In an amplicon-based pipeline, there is a PCR step similar to the one used in the traditional Sanger-sequencing of individual specimens, typically employed in the generation of DNA barcode reference libraries, but instead of being sequenced one by one, all amplicons are sequenced at once in HTS platforms At least, that is the idea, but the choice of adequate markers and primers is much more difficult in multitaxa barcoding than in single, Sanger sequencing, DNA barcoding.

The ideal marker for metabarcoding should fulfill several requirements that, in most cases, present a trade-off between each other. These markers should have a high mutation rate, so they are variable enough to provide taxonomic resolution (i.e. to be

similar between individuals from the same species, but clearly different between individuals from closely related species), but be preceded and followed by conserved regions to which PCR primers can be attached. These regions should ideally be conserved only within the target group, so that no other groups are amplified. Finally, the ideal marker should display such features in a short sequence length of at most 500-600 bp. A short marker sequence is important for the recovery of information in case of DNA degradation, which is common in eDNA samples (Clarke *et al.* 2014, Yu *et al.* 2012). Short markers are also easier to sequence using MPS techniques. As an additional requirement, it is an advantage if the marker is already widely studied, so that publicly accessible reference databases allowing species identification are available. On the other hand, MPS techniques simplify the production of new reference libraries *de novo* given access to suitable DNA archives of the target biome.

Since the origins of DNA barcoding (single-specimen), the most used marker for metazoans has been a 658 bp long sequence at the 5' end of the cytochrome oxidase subunit I (COI), traditionally amplified with the primers designed by Folmer *et al.* (1994). For plants, fragments of the large subunit of ribulose 1,5-bisphosphate carboxylase gene (rbcL) and the maturase K gene (matK) are established as barcoding markers (Newmaster *et al.* 2006, Chase *et al.* 2007), while in fungi the internal transcribed spacer of the nuclear ribosome (ITS) is used (Schoch *et al.* 2012). For bacteria the 16S rRNA of the small ribosomal subunit (16S) (Andersson *et al.* 2008) and for unicellular eukaryotes the small nuclear ribosomal subunit (SSU-18S) are used as molecular barcodes (Pawlowski *et al.* 2012).

Multitaxa PCR requires the use of 'universal' primers, either universal for all metazoans or for the specific target group in the sample. Hajibabaei *et al.* (2006) and Meusnier *et al.* (2008) proved that much shorter fragments within the barcoding

region, from 135 bp, are able to provide equal capacity of discrimination. These fragments, of variable length and position depending on the taxonomic group, are known as mini-barcodes (Meusnier *et al.* 2008). Brandon-Mong *et al.* (2015) reviewed 6 universal primer pairs from 6 different studies, amplifying sequences within the barcoding region ranging from 130 to 550 bp. Once the primers are chosen, and the PCR reaction conducted, sequencing is a very straightforward step. In the beginning, metabarcoding studies used 454 pyrosequencing (Shokralla *et al.* 2012, Yu *et al.* 2012, Hajibabaei *et al.* 2012), and also more recent ones (Shokralla *et al.* 2014), but pyrosequencing quickly gave way to the use of Illumina platforms, either HiSeq or MiSeq sequencers (www.illumina.com). The Illumina techniques produce shorter reads, but they are much less expensive per base pair than those resulting from pyrosequencing.

Bioinformatics processing of MPS amplicon data

Illumina platforms generate huge amounts of reads. An example is the study of Gibson *et al.* (2015), producing more than 11.64 million reads of 250 bp. Primer sequences have to be removed ('trimmed' is the specific term) and then the reads should be merged to create paired-end reads of approximately 500 bp, that will constitute the barcode, and quality filtered to remove sequencing artifacts. This way, they obtain 4.37 million reads. Such reads are the PCR product, which was exponentially amplified from the original sample. This means that sequences from the same individual are represented many times in those 4.37 million sequences, and this redundant information must be eliminated. This step is known as 'de-replication', and consists of merging sequences above a high similarity threshold, usually set at 99-100 % similarity. In some cases, singletons (sequences retrieved in only one or two

replicas) are also removed, as they are likely to represent contaminants. Although most sequencing artifacts are eliminated in the quality filtering, a specific step is often introduced to get rid of chimeras (artifacts product of the pairing of two sequences that actually don't belong together, due to a fortuitous similarity between some regions of the mentioned sequences) that may have formed during sequencing since they may escape the quality filtering. Finally, the remaining sequences are clustered under a variable threshold of similarity that is decided by researchers, usually 97-98 %, and each cluster is considered as a separate MOTU (Molecular Operational Taxonomic Unit, barcoding's particular concept of 'species'). If the aim of the study is to assess name the species present in the sample, the MOTUs have to be compared with a reference database of sequences of the same marker from reliably identified specimens, e.g. using GenBank or BOLD. If the aim is only to generate biodiversity metrics, species identities may not be strictly necessary.

The above protocol only results in a qualitative analysis of the taxonomic composition of the sample. Often, it is also of interest to study the quantitative composition. If so, one goes back to the original reads, computing the proportion of the reads that match each of the MOTUs.

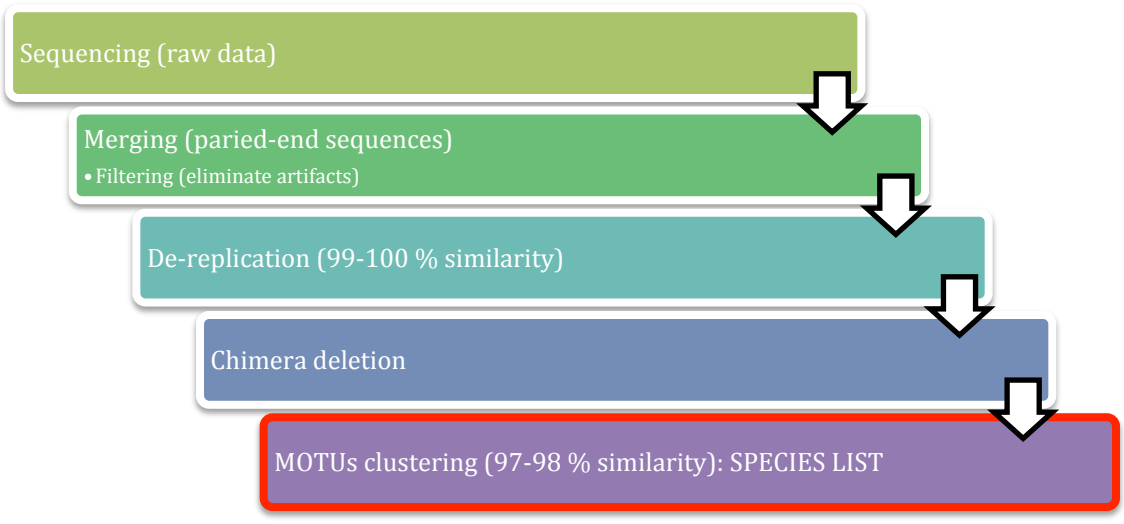

Fig. 4 Simplified bioinformatics pipeline for analyzing metabarcoding data.

There are several softwares available on the web to perform these steps, and separate programs can be used for each step. However, some software packages have been released that contain programs to perform the whole process. One of them is the USEARCH package (Edgar 2010) that includes the programs UPARSE (Edgar 2013) to de-replicate reads, UCHIME (Edgar *et al.* 2011) to eliminate chimeras and UCLUST (Edgar 2010) for MOTU clustering with variable similarity thresholds. More recently, Boyer *et al.* (2015) developed a package that contains all the necessary programs for analyzing NGS data in a metabarcoding context (de-replication, trimming of primers, chimera deletion, MOTU clustering and identification). The package, called OBITools (http://metabarcoding.org/obitools), can be freely downloaded with a practical tutorial available and it is linked to the ecoPrimers and ecoPCR softwares (Riaz *et al.* 2011).

Problems with amplicon-based approaches

To understand the flaws that accompanies amplicon-based approaches, it is important to define some concepts, that might have appeared before in the text, but that will be better defined here, related to barcoding, barcodes and primers.

The first of these concepts is *specificity*; specificity refers to the degree of complementarity between a primer sequence and its target sequence. The more the base pair matches between both sequences, the higher the specificity of the primer. This implies that primers can exist that don't match exactly the target sequence, and they are still able to produce amplification in a PCR reaction, although this capacity diminishes with the increase in the number of mismatches until a to in which there is no amplification success. Two target sequences can be amplified then by the same primer pair, even if one has the highest specificity (all base pairs matches between

primer and template) and the second has lower specificity. The problem appears when these two target sequences are in the same sample, since the template with higher specificity will bind with more affinity to the primer and have an easier and more abundant amplification than the other. Given that PCR reaction amplifies the template sequences exponentially (2^n, where n is the number of cycles), a deviation in the amplification success can increase the proportion of the best-binding templates and decrease the proportion of the worst-binding ones, so that in extreme cases the result will be no recovery of the latter. That is called *amplification bias*. Degeneracy is one of the most used solutions to overcome the problem of mismatches in primer binding. The idea is to allow different nucleotides in certain position(s) of the primer, where the mismatch is expected, so that each of the unique template sequences will have an exact (or close) match to at least one of the primers in the primer mixture.

Given that a degenerate primer is needed, the optimal situation would be simple degeneracy (only two possible nucleotides) in just one primer position. Unfortunately, that is not always sufficient for complete or near-complete coverage of a broad taxonomic group. In many occasions, more than one degenerate position is needed, and with more than two possible nucleotides at one or more of the degenerate sites. Valentini *et al.* (2015) designed a pair of primers for the subclass Teleostei (bony fishes) without any degeneracy, and for the superorder Batrachia (amphibians excluding Gymnophiona) with just one position with simple degeneracy. In constrast, primers designed by Geller *et al.* (2013) and Leray *et al.* (2013) for a much broader taxonomic group – all marine Metazoa – had up to one third of the primer positions degenerated. Also, with degenerate primers, amplification bias may result from different primer-template pairs showing different binding affinity depending on chemical factors, such as their GC content.

As mentioned above, it has been considered that the whole barcoding region of COI, and even smaller sequences down to 130-160 bp, is the best marker for barcoding, since it shows enough variation to discriminate between closely related species and there is plenty of reference data to compare with. However, there is another requirement that is in direct trade-off with sequence high variability: the presence of conserved regions to which the primers can be attached.

COI is a protein-coding gene, which means that the nucleotide sequence is translated to a protein that has 3D structure, location (embedded in the mitochondrion's inner membrane), and function: to oxidate the cytochrome c protein and generate a redox imbalance that can be used as energy to produce an ATP molecule. The structural and functional constraints result in an aminoacid sequence that is strongly conserved in certain regions, while other domains are susceptible to vary. Considering this, COI appears to fill both requirements. However, even if the amino acid sequence is strongly conserved, the corresponding DNA sequence can vary because of the redundancy of the genetic code: the third position in the codon is typically irrelevant for the produced amino acid, resulting in the third position being less constrained by selection (Deagle *et al.* 2014) and therefore more variable than the first and second codon positions. This makes it more difficult to obtain a universal primer that exactly matches sequences out of a particular target group. Taking this into account, different research teams have designed primers with the material they had available. This has on one hand lead to the existence of several barcode and mini-barcode 'universal' primers (Brandon-Mong *et al.* 2015), and more than 400 primer pairs for different animal groups (CBOL database, www.boldsystems.org), allowing one to potentially target a particular group of interest. On the other hand, but due to the same phenomenon (higher rates of change in third positions of each codon), a

primer designed for a certain taxonomic group could amplify non-targeted templates in the sample, since the saturation in change may produce homoplasy in the sequence, and thus, perfect match with a primer intended for a different group.

This bias in 'universal' primers has been documented by several studies: Hajibabaei *et al.* (2011) reported biases with Lep-F1/Lep-R1 (Hebert *et al.* 2004) (also Brandon-Mong *et al.* 2015), Yu *et al.* (2012) showed that Folmer primers (Folmer *et al.* 1994) fail to amplify many species of Hymenoptera (wasps, bees and ants), and Clarke *et al.* (2014) showed that several primer pairs present amplification bias resulting in overrepresentation of Diptera (flies and mosquitoes) and Lepidoptera (moths and butterflies) sequences. Increasing sequencing depth would not solve this problem, and could possibly generate chimeric reads (Deagle *et al.* 2014).

Finally, there is a last solution, which still needs further testing. It is known as multiplex PCR, and consists in a PCR reaction with several different primer pairs that amplify the same region of the marker, but they are specific of small taxonomic groups within the sample. Tested by Gibson *et al.* (2014), they were able to recover 91 % of the species present in an artificial bulk sample of known composition using 11 primer pairs. However, this methodology has to be improved with regard to cost-effectiveness.

In the *field and laboratory protocols* section, it was mentioned that eDNA samples need to be handled with special care to avoid contamination. This is particularly important when a single invasive or endangered species is targeted, or a matrix of presence/absence is constructed. Sources of false positives are, not only contamination in the field or the laboratory, but also PCR or sequencing errors (Ficetola *et al.* 2014). A recurrent strategy to deal with this problem is to use *ad hoc* solutions, such as ignoring those species that only appear in one or two samples,

considering them contaminations or PCR/sequencing artifacts. Comparison with traditional, unambiguous methods can help spotting those false positives, but the drawback of this approach is that such methods are not always feasible to conduct in the sampling site, or that eDNA is being used to avoid the slow and expensive traditional methods. Again, increasing sequencing depth would not solve this problem, as it can amplify the error by generating more sequences assigned to absent species, or create chimeras during sequencing (Ficetola *et al.* 2016). Statistical methods are being improved to solve the false positive and false negative obstacles, considering literature and past sampling results in modeling (Lahoz-Monfort *et al.* 2016).

Alternative markers for amplicon-based sequencing

Summarizing, one could argue that amplicon-based metabarcoding using COI is not the perfect tool for biodiversity assessment of macrofauna and macroflora using eDNA, but it is the best method we have. Nevertheless, a trend exists to question this statement, based on the above-mentioned facts.

Some metabarcoding studies have been carried out already using alternative markers. It is common, when working with a very broad taxonomic scope (up to phyla) to use a very conserved but easily amplified marker, such as the small nuclear ribosomal subunit (18S). Several examples can be found, mostly in soil/sediment biodiversity assessment (Drummond *et al.* 2015, soil; Brannock & Halanych 2015 and Dell'Anno *et al*. 2015, meiofauna). However, in a meiofaunal study, Tang *et al.* (2012) warned that 18S metabarcoding underestimates real biodiversity compared to traditional morphospecies determination and COI metabarcoding. For vertebrates, especially fishes, the most used alternative marker is the mitochondrial small

ribosomal subunit (12S), e.g. Valentini *et al.* (2016), Port *et al.* (2015) or Furlan *et al.* (2015). These studies are mainly focused on freshwater or marine fishes and amphibians, with very satisfactory results. Thomas *et al.* (2014, 2015) also focus on fishes, but in this case testing the capability of the mitochondrial large ribosomal subunit (16S) in the estimation of the fish biomass consumed by seals and cormorants. Performance of 16S has been tested in insect metabarcoding studies as well, showing very interesting results. Clarke *et al.* (2014) showed that 16S minibarcodes of less than 200 bp identified just little percentage less than minibarcodes of COI of a set of 315 species (constituting 264 genera and 23 orders) of insects, in an *in silico* simulation of a PCR reaction, while the taxonomic coverage (no. of species successfully amplified) was 75-90 % with 16S versus 50 % with COI. With minibarcodes longer than 175 bp, taxonomic resolution of COI was almost 100 %, while it decreased to 85 % with 16S[2]. Remarkably, taxonomic coverage and taxonomic resolution was consistent through the 23 orders when using 16S, while the best COI metabarcoding showed a coverage of more than 50 % within Diptera and Lepidoptera, but between 0-47 % within other insect orders. Similar results were obtained by Elbrecht *et al.* (2016): 16S amplified more species and more equally through orders, thus enhancing biomass estimation. They state that if the goal is to identify the species present in the sample, COI is still the best choice due to a better representation in public reference databases, but when the aim is to assess the biodiversity in numbers –rather than in identities-, 16S would be a better choice. Additionally, 16S metabarcoding has the advantage of amplicons not being mistaken with nuclear pseudogenes or *Wolbachia* reads, as can be the case with COI (Clarke *et al.* 2014).

[2] Considering the total as the sequences successfully amplified in the previous step. Thus, the 100 % of COI and 85 % are transformed into 50 % (50 % species amplified x 100 % species identified) and 66-77 % (75-90 % species amplified x 85 % species identified) respectively, of the total number of species present in the study.

Consistently with this statement, Deagle *et al*. (2014) suggest that the best strategy to follow in the future would be to build local databases of several markers, and conduct metabarcoding studies with these different markers simultaneously rather than focusing on a single universal marker (COI).

Conclusion

Metabarcoding is a very powerful tool for biodiversity analysis and assessing faunal composition. It is one of the fastest growing fields in science and methods development is increasing in quality and resources. However, we have only witnessed the dawn of metabarcoding, and there is yet much more work, testing and exploration left until we can call it an established field.

REFERENCES

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

Andersson, A. F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P., & Engstrand, L. (2008). Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PloS ONE*, 3(7): e2836.

Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjaer, K. H., ... & Willerslev, E. (2012). Meta-barcoding of 'dirt'DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, 21(8): 1966–1979.

Barrett, R. D. H., & Hebert, P. D. N. (2005). Identifying spiders through DNA barcodes. *Canadian Journal of Zoology*, 83(3): 481–491.

Bergsten, J., Bilton, D. T., Fujisawa, T., Elliott, M., Monaghan, M. T., Balke, M., ... & Ribera, I. (2012). The effect of geographical scale of sampling on DNA barcoding. *Systematic biology*, sys037.

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., Griffiths, R. A., ... & Williams, P. (2015). Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biological Conservation*, 183: 19–28.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). OBITools: a UNIX-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, 16(1): 176–182.

Brandon-Mong, G. J., Gan, H. M., Sing, K. W., Lee, P. S., Lim, P. E., & Wilson, J. J. (2015). DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of Entomological Research*, 105(06): 717–727.

Brannock, P. M., & Halanych, K. M. (2015). Meiofaunal community analysis by high-throughput sequencing: Comparison of extraction, quality filtering, and clustering methods. *Marine Genomics,* 23(C): 67–75.

Brown, D. S., Burger, R., Cole, N., Vencatasamy, D., Clare, E. L., Montazam, A. & Symondson, W. O. C. (2014), Dietary competition between the alien Asian Musk Shrew (*Suncus murinus*) and a re-introduced population of Telfair's Skink (*Leiolopisma telfairii*). *Molecular Ecology*, 23: 3695–3705.

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P. M., Metzger, S., Schubert, G. & Leendertz, F. H. (2013), Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology*, 22: 915–924.

Chase, M. W., Cowan, R. S., Hollingsworth, P. M., van den Berg, C., Madriñán, S., Petersen, G., ... & Pedersen, N. (2007). A proposal for a standardised protocol to barcode all land plants. *Taxon*, 56(2): 295–299.

Clare, E. L., Symondson, W. O., & Fenton, M. B. (2014). An inordinate fondness for beetles? Variation in seasonal dietary preferences of night-roosting big brown bats (*Eptesicus fuscus*). *Molecular Ecology*, 23(15): 3633–3647.

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.

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): 20140562.

Dell'Anno, A., Carugati, L., Corinaldesi, C., Riccioni, G., & Danovaro, R. (2015). Unveiling the Biodiversity of Deep-Sea Nematodes through Metabarcoding: Are We Ready to Bypass the Classical Taxonomy? *PloS ONE*, 10(12): e0144928.

Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C., ... & Park, D. (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience,* 4(1): 1–20.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST, *Bioinformatics*, 26(19): 2460–2461.

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods*, 10(10): 996–998.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16): 2194–2200.

Elbrecht V., Taberlet P., Dejean T., Valentini A., Usseglio-Polatera P., Beisel J., Coissac E., Boyer F., Leese F. (2016) Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *PeerJ* 4:e1966.

Emrich, M. A., Clare, E. L., Symondson, W. O., Koenig, S. E., & Fenton, M. B. (2014). Resource partitioning by insectivorous bats in Jamaica. *Molecular Ecology*, 23(15): 3648–3656.

Epp, L. S., Boessenkool, S., Bellemain, E. P., Haile, J., Esposito, A., Riaz, T., ... & Stenøien, H. K. (2012). New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular Ecology*, 21(8): 1821–1833.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., ... & Rayé, G. (2014). Replication levels, false presences and the estimation of the

presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, 15(3): 543–556.

Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding?. *Molecular Ecology Resources*, 16(3): 604–607.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5): 294–299.

Foote, A. D., Thomsen, P. F., Sveegaard, S., Wahlberg, M., Kielgast, J., Kyhn, L. A., ... & Gilbert, M. T. P. (2012). Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PloS ONE*, 7(8): e41781.

Furlan, E. M., Gleeson, D., Hardy, C. M., & Duncan, R. P. (2015). A framework for estimating the sensitivity of eDNA surveys. *Molecular Ecology Resources*.

Gariepy, T. D., Lindsay, R., Ogden, N., & Gregory, T. R. (2012). Identifying the last supper: utility of the DNA barcode library for bloodmeal identification in ticks. *Molecular Ecology Resources*, 12(4): 646–652.

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, 13(5): 851–861.

Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., Hallwachs, W. & Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, 111(22): 8007–8012.

Gibson, J. F., Shokralla, S., Curry, C., Baird, D. J., Monk, W. A., King, I., & Hajibabaei, M. (2015). Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. *PLoS ONE,* 10(10): e0138432–15.

Hajibabaei, M., Smith, M., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., & Hebert, P. D. (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, 6(4): 959–964.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A., & Baird, D. J. (2011). Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE,* 6(4): e17497.

Hajibabaei, M., Spall, J. L., Shokralla, S., & van Konynenburg, S. (2012). Assessing biodiversity of a freshwater benthic macroinvertebrate community through non- destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology,* 12(1): 1–1.

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.

Hebert, P. D., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, 101(41): 14812–14817.

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of Birds through DNA Barcodes. *PLoS Biology*, 2(10): e312.

Krüger, F., Clare, E. L., Symondson, W. O., Keišs, O., & Pētersons, G. (2014). Diet of the insectivorous bat *Pipistrellus nathusii* during autumn migration and summer residence. *Molecular Ecology*, 23(15): 3672–3683.

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Tingley, R. (2016). Statistical approaches to account for false-positive errors in environmental DNA samples. *Molecular Ecology Resources*, 16(3): 673–685.

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

Liles, M. R., Manske, B. F., Bintrim, S. B., Handelsman, J., & Goodman, R. M. (2003). A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Applied and Environmental Microbiology,* 69(5): 2684–2691.

Liu, S., Wang, X., Xie, L., Tan, M., Li, Z., Su, X., ... & Niehuis, O. (2015). Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources*, 16: 470–479.

Lücker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., Vacherie, B., ... & Daims, H. (2010). A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proceedings of the National Academy of Sciences,* 107(30): 13479–13484.

Meier, R., Zhang, G., & Ali, F. (2008). The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Systematic Biology*, 57(5): 809–813.

Meusnier, I., Singer, G. A., Landry, J. F., Hickey, D. A., Hebert, P. D., & Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC genomics*, 9(1): 214.

Meyer, C. P. & Paulay, G. (2005). DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology*, 3(12): e422.

Newmaster, S. G., Fazekas, A. J., & Ragupathy, S. (2006). DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. *Botany*, 84(3): 335–341.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., ... & Fiore-Donno, A. M. (2012). CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biology*, 10(11): e1001419.

Pompanon, F., Coissac, E., & Taberlet, P. (2011). Metabarcoding, une nouvelle façon d'analyser la biodiversité. *Biofutur,* 319: 30–32.

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., ... & Kelly, R. P. (2015). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, 25: 527–541.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., ... & Mende, D. R. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature,* 464(7285): 59–65.

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic acids research,* 39(21): e145.

Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R. D., Willerslev, E., Bertelsen, M. F. & Gilbert. M. P. T. (2012) Screening mammal biodiversity using DNA from leeches. *Current biology,* 22(8): R262–R263.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., ... & Miller, A. N. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, 109(16): 6241–6246.

Shokralla, S., Spall, J. L., Gibson, J. F., & Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, 21(8): 1794–1805.

Shokralla, S., Gibson, J. F., Nikbakht, H., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2014). Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources*, 14(5): 892-901.

Shokralla, S., Porter, T. M., Gibson, J. F., Dobosz, R., Janzen, D. H., Hallwachs, W., Golding, G. B. & Hajibabaei, M. (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific reports*, 5: 9687.

Shokralla, S., Gibson, J., King, I., Baird, D., Janzen, D., Hallwachs, W., & Hajibabaei, M. (2016). Environmental DNA Barcode Sequence Capture: Targeted, PCR-free Sequence Capture for Biodiversity Analysis from Bulk Environmental Samples. *bioRxiv*, 087437.

Symondson, W. O. C. & Harwood, J. D. (2014), Special issue on molecular detection of trophic interactions: Unpicking the tangled bank. *Molecular Ecology*, 23: 3601–3604.

Taberlet, P., Coissac, E., Hajibabaei, M., & Riesenberg, L. H. (2012). Environmental DNA. *Molecular Ecology,* 21: 1789–1793.

Tang, C. Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T. G., & Fontaneto, D. (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences*, 109(40): 16208–16212.

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.

Thomas, A. C., Deagle, B. E., Eveson, J. P., Harsch, C. H., & Trites, A. W. (2015). Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources*, 16(3): 714–726.

Townzen, J. S., Brower, A. V. Z., & Judd, D. D. (2008). Identification of mosquito bloodmeals using mitochondrial cytochrome oxidase subunit I and cytochrome b gene sequences. *Medical and Veterinary Entomology*, 22(4): 386–393.

Tringe, S. G., Von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., ... & Bork, P. (2005). Comparative metagenomics of microbial communities. *Science,* 308(5721): 554–557.

Tréguier, A., Paillisson, J. M., Dejean, T., Valentini, A., Schlaepfer, M. A., & Roussel, J. M. (2014). Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. Journal of Applied Ecology, 51(4): 871–879.

Turner, C. R., Barnes, M. A., Xu, C. C., Jones, S. E., Jerde, C. L., & Lodge, D. M. (2014). Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, 5(7): 676–684.

Turner, C. R., Uy, K. L., & Everhart, R. C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, 183(C): 93–102.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., ... & Gaboriaud, C. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology,* 25: 929–942.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., ... & Fouts, D. E. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science,* 304(5667): 66–74.

Welch, K. D., Schofield, M. R., Chapman, E. G. & Harwood, J. D. (2014), Comparing rates of springtail predation by web-building spiders using Bayesian inference. *Molecular Ecology*, 23: 3814–3825.

Wiemers, M., & Fiedler, K. (2007). Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, 4(1): 8–16.

Xu, C. C. Y., Yen, I. J., Bowman, D., & Turner, C. R. (2015). Spider Web DNA: A New Spin on Noninvasive Genetics of Predator and Prey. *PLoS ONE,* 10(11): e0142503–10.

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, 3(4): 613–623.

Zhu, L., Zhang, S., Gu, X., & Wei, F. (2011). Significant genetic boundaries and spatial dynamics of giant pandas occupying fragmented habitat across southwest China. *Molecular Ecology*, 20(6): 1122–1132.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., ... & Huang, Q. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *Gigascience*, 2(1): 4.

WEBSITES

Illumina: www.illumina.com

OBITools: http://metabarcoding.org/obitools

CBOL database: www.boldsystems.org