PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/161842

Please be advised that this information was generated on 2018-02-24 and may be subject to change.
Invited Review

Added value of metabarcoding combined with microscopy for evolutionary studies of mammals

ANNE-JIFKE HAARSMA, HENK SIEPEL & BARBARA GRAVENDEEL

Submitted: 29 March 2016
Accepted: 11 August 2016
doi:10.1111/zsc.12214


Metabarcoding – taxon identification from complex mixtures using a standard DNA region – is increasingly used in evolutionary studies. With this method, it is not only possible to delimit species or collect indirect observational data but also to determine diets. Caveats such as false negatives and skewed abundances can be overcome when metabarcoding is combined with traditional methods such as microscopy. Such a combined approach can help deducing why some species went extinct or became endangered whereas others evolved into new lineages. This review will focus on the added value of metabarcoding when combined with traditional methods for evolutionary studies of mammals.

Introduction

To study the evolution of the world’s biodiversity, species identification is an important first step. Traditionally, this was done using morphology, but since the field of molecular biology took off in the seventies, DNA-based identification is increasingly used. This is done by matching a standardized region of the genome (a ‘barcode’) from an unknown sample against all available barcodes in a reference library (Valentini et al. 2009). For this approach, the term ‘DNA barcoding’ was coined. The method was first developed based on Sanger sequencing, in which fragments of genome are separated with gel electrophoresis. Since 2005, metabarcoding was developed, which can be considered as a high-throughput way of DNA barcoding: multiple DNA barcodes are generated simultaneously from a mixture of species. Both methods are still used, although in some scientific fields next-generation sequencing (NGS) is quickly outcompeting the classic Sanger chain-termination method (Taylor & Harris 2012). Below, we will summarize the six main steps involved in barcoding.

DNA barcoding and metabarcoding – the six main steps

First of all, the most suitable DNA barcoding regions should be chosen to answer a specific research question (Fig. 1). The most commonly used DNA barcode region for animals is a segment of approximately 600 bp long of the mitochondrial gene cytochrome oxidase I (CO1) (Hebert et al. 2003). This locus provides large sequence variation between species yet relatively small amount of variation within species (Kress et al. 2015). Other commonly used barcode regions used for species identification...
of animals are ribosomal DNA regions such as 16S, 18S and 12S and mitochondrial regions such as cytochrome B (Galan et al. 2012; Ivanova et al. 2012; Clare 2014; Nowak et al. 2014). All these markers have their advantages and disadvantages and are therefore used for different purposes (e.g. Pompanon et al. 2012; Deagle et al. 2014). Longer barcode regions (i.e. at least 600 bp long) are often needed for accurate species delimitation, especially to differentiate close relatives. Identification of the producer of organism’s remains such as faeces, hairs and saliva can be used as a proxy measure to verify absence/presence of a species in an ecosystem. The DNA in these remains is usually of low
quality and quantity, and therefore, shorter barcodes of around 100 bp long are used in these cases. Similarly, DNA remains in dung are often degraded as well, so short barcodes are needed to identify prey consumed.

Secondly, a reference database needs to be built of all DNA barcodes likely to occur in a study. Ideally, these barcodes need to be generated from vouchered specimens deposited in a publicly accessible place, such as for instance a natural history museum or another research institute. Building up such reference databases is currently being done all over the world. Partner organizations collaborate in international projects such as ‘International barcode of Life project’ (iBOL) and ‘Consortium for the Barcode of Life’ (CBOL), aiming to construct a DNA barcode reference that will be the foundation for DNA-based identification of the world’s biome. Well-known barcode repositories are NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and the Barcode of Life Database (BOLD) (http://www.boldsystems.org).

Thirdly, the cells containing the DNA of interest must be broken open to expose its DNA. This step, DNA extractions and purifications, should be performed from the substrate under investigation. There are several procedures available for this (Dhaliwal 2015). Specific techniques must be chosen to isolate DNA from substrates with partly degraded DNA, for example fossil samples, and samples containing inhibitors, such as blood, faeces and soil. Extractions in which DNA yield or quality is expected to be low should be carried out in an Ancient DNA facility, together with established protocols to avoid contamination with modern DNA (Cooper & Poinar 2000; Willerslev & Cooper 2005). Experiments should always be performed in duplicate (Ficetola et al. 2015) and with positive controls included. Additional steps might involve the use of special kits, such as the Qiagen stool kit for dung or GFX purification kit for blood.

Fourthly, amplicons have to be generated from DNA extracted, either from a single specimen or from complex mixtures with primers based on DNA barcodes selected under step 1. To keep track of their origin, labelled nucleotides (molecular IDs or MID labels) need to be added in case of metabarcoding. These labels are needed later on in the analyses to trace reads from a bulk data set back to their origin.

Fifthly, the appropriate techniques should be chosen for DNA sequencing. The classic Sanger chain-termination method relies on the selective incorporation of chain-elongating inhibitors of DNA polymerase during DNA replication. These four bases are separated by size using electrophoresis and later identified by laser detection. The Sanger method is limited and can produce a single read at the same time and is therefore suitable to generate DNA barcodes from substrates that contain only a single species. Modern NGS technologies can handle thousands to millions reads in parallel and are therefore suitable for mass identification of a mix of different species present in a substrate, summarized as metabarcoding.

Lastly, bioinformatic analyses need to be carried to match DNA barcodes obtained with Barcode Index Numbers (BINs) in reference libraries. Each BIN, or BIN cluster, can be identified to species level when it shows high (>97%) concordance with DNA barcodes linked to a species present in a reference library, or when taxonomic identification to the species level is still lacking, an operational taxonomic unit (OTU), which refers to a group of species (i.e. genus, family or higher taxonomic rank). The results of the bioinformatics pipeline must be pruned, for example by filtering out unreliable singletons, superfluous duplicates, low-quality reads and/or chimeric reads. This is generally done by carrying out serial Blast searches in combination with automatic filtering and trimming scripts (Li 2015). Standardized thresholds are needed to discriminate between different species or a correct and a wrong identification.

Understanding evolution of the world’s biodiversity is hampered by a lack of data, and DNA barcoding and metabarcoding are promising tools to fill this knowledge gap. In this review, we provide an overview of the opportunities and added value of DNA barcoding and metabarcoding in combination with traditional methods, microscopy in particular, for evolutionary studies. We first outline how DNA barcoding and metabarcoding helped finding answers on important evolutionary questions. For this purpose, we performed a comprehensive literature review, covering 101 publications. We finish by describing challenges that still need to be overcome and providing an outlook on further improvements. We will focus this review on mammals because they play an important role in ecosystems by regulation of insects and plants as predator and/or pollinator. They can be used as indicators of the health of an ecosystem. Many of the larger mammals went extinct during the last ice age or currently teeter on the brink of extinction. Despite all this, lineages continue to evolve into new species as well. Understanding species delimitation, presence/absence and diets is vital to obtain more insight in the drivers of mammalian evolution and answer the question why some lineages went extinct or became endangered whereas other evolved into new lineages.

**Main challenges in evolutionary studies of mammals**

The first mammals evolved at the end of the Triassic period over 200 million years ago. They were often as small as modern mice and occupied available niches left over by the dinosaurs, at that time the dominant land animals. After
the extinction of the former, mammals further diversified and filled vacated ecological niches by evolving highly specialized morphological adaptations. Among others, limbs, teeth, hairs, horns and mammary glands evolved for digging, running, swimming, bulldozing, flying, eating and feeding. The extinct hornless rhino (Paraceratherium), with its weight of 15 to 20 tons, was one of the largest terrestrial mammals that ever existed. Marine mammals of the order Cetacea currently include the largest animals on the planet. Bats (order Chiroptera) and shrews (order Eulipotyphla) encompass the smallest mammals, with a weight of less than 2 grams.

Mammals have colonized all continents and live in all climates. Often, biological events such as climate change, change in habitat requirements and food sources cause shifts in geographical ranges of mammal species. An example is the past shift in geographical range of the woolly mammoth (Mammothus primigenius) that became extinct near the end of the last ice age (Burns et al. 1996). The triggers for its extinction remain fervently debated. Multiple events, such as rapid climate oscillations with regional temperature changes of up to 16 °C, a change in the habitat and consequentially food sources, in combination with human impact (De Vivo & Carmignotto 2004; Cooper et al. 2015) are considered the principal causes for extinction. The giant deer (Megaceros giganteus) is another example of a species that went extinct. In this case, its huge antlers excluded the males from woodland, which contained their main food source (Stuart et al. 2004).

Ongoing increase of human activity since the last ice age currently causes the sixth extinction event known to life on earth (Kolbert 2014). It causes range contraction of several large Asian mammals in China (Li et al. 2015), including the Asian elephant (Elephas maximus), which evolved from mammoths (Krause et al. 2006; Orlando et al. 2007). Despite all this, new mammalian species keep evolving as well. An example of a recent speciation trigger is artificial selection against tusked male Asian elephants. This selection is driven by ivory collecting poachers. The trigger seems to facilitate rapid evolution of tuskless males (Chelliah & Sukumar 2013) that might eventually evolve into a new species.

DNA barcoding and metabarcoding have proven useful in many contexts (Francis et al. 2010; Joly et al. 2014; Galimberti et al. 2015; Kress et al. 2015). Although the predominant application still focuses on describing biodiversity, DNA barcoding and metabarcoding are becoming increasingly important tools for evolutionary studies as well. We wrote this review to show how DNA barcode data can be used to obtain more knowledge about mammalian evolution, specifically to understand why some species went extinct or became endangered, whereas others continued to evolve into new species. We will focus this review on three aspects: (i) species delimitation, (ii) indirect observations and (iii) diet analyses. These three aspects can help demarcate boundaries among species, reconstruct latest date of occurrence of extinct species and identify processes that promote current lineage diversification.

Species delimitation

Traditionally, species were described mainly on morphological characters, sometimes in combination with behaviour (Fig. 2). Auditory signals used to delimit territory, such as made by bats, whales, lemurs and deer, are useful tools for species delimitation as well, especially to keep morphologically similar species apart (Jones 1997). Geographical differences in these traits can make correct species identification problematic. In such situations, DNA barcoding can complement traditional methods, especially for the distinction of species that are erroneously classified under one species name. Below, we will provide examples of how DNA-based species delimitation improved knowledge on the evolution of extinct, endangered and newly evolving species.

Morphological identification of extinct mammalian species is difficult, as often only fragments of an animal are found. With DNA barcoding, these can be identified to species level with much higher accuracy. Moreover, with additional phylogenetic analyses, the evolution of mammals can be traced and the timeline can be uncovered in which extinct species lived. This was for instance possible for the last cave bears (Ursus spelaeus) in Europe (Loreille et al. 2001; Bocherens et al. 2014). By identifying fossil bone material using a combination of morphology and DNA barcoding and linking these to radiocarbon dates, climatic cooling in combination with decreased vegetational productivity could be identified as the main triggers for the disappearance of this species. In these studies, the benefits of DNA barcoding consisted of distinguishing bones of cave bear from related bear species still existing today. It is sometimes also possible to link timelines to other changes in the habitat such as forest fragmentation (Wooding & Ward 1997; Bhagwat et al. 2014). Orlando et al. (2013) succeeded in revising the recent evolutionary history of wild horse (Equus) using ancient DNA. Their results suggest that climatic changes and related grassland contraction were major demographic drivers for extinctions of local horse populations.

For mammalian species that are not yet extinct but nearly so, DNA barcoding can help unravelling the main reason for this. This knowledge can be useful to save the species from extinction. DNA analysis of the almost extinct Arabian oryx (Oryx leucoryx) showed that the species was about to get extinct due to severe inbreeding. By tracing its
closest relatives and setting up a breeding programme, the genetic diversity of this species could be increased again and extinction was prevented (Elmeer et al. 2012). In this case, DNA barcodes obtained could be used to find the closest still existing wild relative to use in a breeding programme.

Despite the relatively small size of the order and high popularity, new mammal species are still being discovered, not only in remote regions of the world, but also in well-studied areas. An example of the first case is the olinguito (Bassaricyon neblina), a member of the racoon family that was discovered in 2014 in the Andes of western Colombia and Ecuador (Helgen et al. 2013). Animals had been erroneously assigned to a related species in the past but DNA barcoding recently revealed the olinguito to be an independently evolved lineage of at least 3.5 million years old. Another example are snouters (Hyorhinomys) found on Sulawesi, a remote mountainous island in Indonesia. DNA barcoding suggest that these snouters represent a new genus in a group of endemic rats. A third example is the discovery of the soprano pipistrelle bat (Pipistrellus pygmaeus). It had been noticed for some years that the common pipistrelle (Pipistrellus pipistrellus) used two distinct echolocation frequencies. Only after DNA barcoding had been carried out, it was confirmed that the common pipistrelle actually consists of two distinct species (Barratt et al. 1997), both of which differ not only in their DNA but also in morphology and behaviour. Another example is the Natterer’s bat (Myotis nattereri) assemblage. Until a few years ago, Escalera’s bat Myotis escalerai was included (as a synonym) in the Natterer’s bat. Recently, Ibáñez et al. (2006) discovered that this cryptic assemblage consists of at least four separate bat species, with distinct differences in DNA sequences, but with only slight morphological differences. Escalera’s bat and the Moroccan Natterer’s bat species (Myotis sp. B) are estimated to have diverged about 2 million years ago (Garcia-Mudarra et al. 2009). The phenomenon of cryptic species occurs quite often among bats. Over the past decade, DNA barcoding facilitated identification of cryptic taxa. The number of bat species in the western Palaearctic therefore increased with almost 50% due to the discovery of previously unrecognized bat species (e.g. Mayer et al. 2007; Garcia-Mudarra et al. 2009).

The more accurate delimitation of species as described above provided more insight into the mechanisms that trigger evolution of new species. An example is the finding that in autumn, many temperature zone bat species gather at underground sites in a behaviour known as swarming. Swarming acts as mating behaviour and plays a role in the assessment of the suitability of an underground site as a hibernaculum. Bogdanowicz et al. (2012) demonstrated that this swarming behaviour leads not only to breeding among bats of the same species, but occasionally also to breeding among different species, resulting in new hybrid lineages with fertile offspring. Hybridization seems quite common among bats. Berthier et al. (2006) and Larsen et al. (2010) discovered that sibling species of bats can produce viable

Fig. 2 Delimitation of subspecies of bats by DNA barcoding of their ectoparases.
offspring in areas of sympatry. This information could only be obtained with DNA barcoding.

**Indirect observational data**

Direct observations of mammals are often difficult to collect, especially for small, elusive or only night-active species (Fig. 3). Traditional indirect methods to assess the presence of such mammals include the identification of their remains such as faeces. Visual identification of faeces is often based on size, shape, texture and composition. As these characteristics are highly variable, correct identification of faeces has proven difficult. Confusion between groups of similar donor species, such as carnivores, bats and deer, is highly possible. Where traditional methods fail, DNA barcoding of mammal remains can be useful to confirm the occurrence of species (Sheppard & Harwood 2005; Waits & Paetkau 2005). Below, we will provide examples of how indirect observational data based on DNA barcoding improved knowledge on the evolution of extinct, endangered and newly evolving species.

With fossilized remains, it is possible to determine past communities of extinct mammals (De Vivo & Carmignotto 2004). However, difference in preservation abilities of the organism and its habitat, and several sampling biases, can determine the likelihood that an organism is preserved as a fossil (MacPhew et al. 2002; Turvey & Cooper 2009). Community ecology based on such incomplete records is perilous. New methods, such as metabarcoding of large

![Image of a church with bats and faeces](image)

**Fig. 3** Collecting indirect observational data of different bat species by DNA barcoding of morphologically similar dung pellets from roosts using either first-generation (Sanger) or second-generation (NGS) sequencing.
environmental samples, can limit these biases, as long as researchers are aware of mixing due to erosion or cases where sediment accumulation was not always linear (e.g. Van Bellen et al. 2011). Previous reconstructions of historical biodiversity based on fossil data predominantly focussed on higher taxonomic levels (order, family, genus). The large number of mammal species and their remains preserved in the fossil record offers opportunities to explore new evolutionary questions when DNA barcoding data are included in the analyses. The inclusion of DNA barcode data in evolutionary studies is therefore starting to become more common (e.g. Chan et al. 2005; De Bruyn et al. 2011).

An increasing number of studies show that DNA barcodes obtained from faeces are accurate tools for species-level identification of endangered mammalian carnivores (Hansen & Jacobsen 1999; Kurose et al. 2005; Chaves et al. 2012). Other successful methods include the use of antlers (Hoffmann et al. 2013), roadkill (Klippel et al. 2015) and other environmental samples such as faeces from predators known to ingest the target species (Galan et al. 2012). Microscopical analyses of remains from owl pellets were used to detect allegedly extinct rodents, such as the Australian desert mouse (Pseudomys desertor) (Bennett et al. 2006) and the birch mouse (Sicista subtilis trizona) (Cserkész et al. 2015). These studies can be applied in a high-throughput manner using DNA barcoding and metabarcoding (Guimaraes et al. 2016). As breeding owls are not always present in an area, including other environmental samples such as pellets of diurnal birds of prey or samples collected from hair traps (Harris & Nicol 2010) might be useful for obtaining year-round observations. Several studies (e.g. Ruibal et al. 2010; Henry et al. 2011) showed that sufficient DNA can be generated from hair samples of endangered mammal species. Here the key contribution of DNA barcoding is the completeness of taxonomic sampling as DNA data can provide information that is otherwise very time-consuming or impossible to obtain.

Introductions of non-native species are increasingly recognized as a threat to ecosystems (Pechar & Mooney 2009) because they have a negative effect on native species by outcompeting them, introducing illnesses or hybridizing with them. With DNA barcoding, the range extension of non-native species such as the brushtail possum (Trichosurus vulpecula) in New Zealand (Ramón-Laca & Gleeson 2014), rodent pests in India (Lakshminarayanan et al. 2015), American mink (Mustela vison) in Europe (Chaves et al. 2012), greater white toothed shrew (Crocidura russula) in Ireland/UK and grey and pallas squirrel (Sicista carolinensis and Callosciurus erythraeus) in Europe can be monitored and, if needed, controlled, to prevent local mammalian species from going extinct. Clearly, such application has the potential of increasing observational data, thereby allowing a better understanding of changes in species distributions.

Ectoparasites, such as mites, ticks and lice, can be adapted to only one or a few host species of mammals. Combined microscopical and DNA barcoding surveys of ectoparasites can be a helpful tool for studying evolution of mammals. Their quick reproduction means that they can be expected to maintain much higher levels of genetic variation than their host. Some ectoparasites, such as whale lice (Crustacea, Amphipoda: Cyamidae) and Spinturnix bat wing mites (Acari, Mesostigmata), migrate only between hosts that have physical contact and hence belong to the same population of interbreeding individuals (Kaliszewska et al. 2005; Bruyndonckx et al. 2010). Others, such as batbugs (Cimex pipistrelli group, Heteroptera: Cimicidae) and bat flies (Diptera, Streblidae: Nycteribiidae), have a free-living stage where they inhabit places where they are likely to encounter a new host, such as a roost (Balvin et al. 2013; Van Schaik et al. 2015). As a consequence, the life histories of ectoparasite and host are coupled and the genetic structure of the parasite reflects patterns of behavioural interaction of the host (Nieberding & Olivieri 2007; Criscione 2008). DNA barcodes obtained from ectoparasites can therefore be helpful in understanding how new host species are originating due to the evolution of reproductive isolation among populations (Fig. 2). Until DNA barcoding was developed, our knowledge of evolutionary processes within populations was restricted to data obtained with traditional methods. With DNA barcoding data, we can obtain a better understanding of processes underlying species diversification.

Diet analysis
The diet of mammals can consist of arthropods, blood, fruits, fish, frogs, fungi, molluscs, other mammals or plants (Fig. 4). Traditional methods to analyse diet are morphological analysis of remains in faeces, pellets and stomachs of dead animals. Other more invasive methods, such as direct analysis of stomach contents, cannot always be applied, especially when studying endangered species. Identification of remains with traditional methods is often only possible to a high (i.e. order, family, genus) taxonomic level. Digestion difference between hard- and soft-bodied prey, or the culling of hard indigestible remains, can lead to a biased picture of the actual diet (e.g. Rabinowitz & Tuttle 1982; Orr & Harvey 2001; Symondson 2002; Bowen and Iverson 2013). With traditional methods, it is, for example, very difficult to study the diet of vampire or nectarivorous bats. On the other hand, morphological analyses of diet components can identify life stage (e.g.
Vaughan 1997) and gender of prey (Acharya 1995) that are impossible to deduce with DNA data.

Although DNA barcoding has its bias towards certain prey species (Deagle et al. 2013; Clare 2014), it is able to bypass some of the drawbacks encountered with traditional morphological methods. Classical DNA sequencing techniques, such as Sanger, require additional steps of cloning to reveal all species present in a highly diverse substrate. NGS bypasses these cloning steps and can reveal a level of diversity that vastly exceeds morphological and Sanger-based surveys (Fig. 3). As each DNA marker has its own taxonomic-specific affinities, using multiple markers is the best strategy to unravel the broadest possible taxonomic diversity of dietary components when applying metabarcoding.

DNA analysis of carnivore diets used to be especially challenging because predator DNA can be simultaneously amplified with prey DNA. To avoid this problem, blocking primers can be applied (Shehzad et al. 2012). These primers prevent the amplification of the predator DNA but allow the amplification of DNA of other organisms. But as these blocking primers can cause mismatches that limit the accuracy of the results (Piñol et al. 2014a), other researchers (Piñol et al. 2014b) advice to use NGS as this technique produces so many reads that sufficient coverage is obtained for DNA barcodes from both predator and prey. Similar to carnivores, DNA barcoding of the diet of herbivores is also challenging, as barcoding of plants cannot be accomplished using a single DNA region only. Often, a combination of two barcode regions is needed (Veldman et al. 2014; Srivathsan et al. 2015). Even with all precautions as described above taken into account, some components of mammalian diets will still fail to be retrieved using a DNA approach. Only a combined approach in which DNA and morphological surveys are combined as proposed by a.o. Van Geel et al. (2008) and Krüger et al. (2014) will provide a full picture. Below, we will provide examples of how DNA-based dietary analyses improved knowledge on the evolution of extinct, endangered and newly evolving species.

Metabarcoding of coprolites (fossil faeces) of the Balearic mountain goat (Myotragus balearicus) and fossilized communal latrines (middens) of the pack rat (Neotoma cinerea) and ground sloth (Nothrotheriops shastensis) showed that these species went extinct because of climate-related habitat changes that eliminated their staple food species (Alcover et al. 1999; Hofreiter et al. 2000; Kuch et al. 2002; Welker et al. 2014). Some pack rat middens have accumulated deposits over 30,000 years. These deposits can provide a detailed and chronological view of habitat changes (Chase et al. 2012). Vegetation history can also be derived from DNA analyses of coprolite samples in combination with radiocarbon dating (Willerslev et al. 2014). Here, large-scale DNA barcoding can reveal patterns about dietary changes of herbivores over time not detectable with traditional methods. This is of great benefit for the reconstruction of main triggers of extinction.

Metabarcoding is also increasingly used as a tool to study dietary separation of sympatric endangered mammals such as African herbivores (Kartzin et al. 2015), carnivores from Venezuela (Farrell et al. 2000), Japan (Kurose et al. 2005), Poland (Posluszný et al. 2007) and lemmings on the Canadian arctic island Bylot (Soininen et al. 2015). These studies show that although the diet of sympatric species can overlap, there are also several differences in food...
species, in both time and space. As a result, competition is avoided. Assemblages of herbivorous mammals seem to be tightly linked to each other and local plant diversity (Kart- zinel et al. 2015). A decline in population numbers of one of the species of the assemblage can lead to a reduction in food availability for the others, in the long run causing extinction. Carnivorous mammals can face a similar fate, especially with diminishing availability of their dominant prey due to human impact (Shehzad et al. 2012) or rapid decline caused by a deadly disease of the prey (Sobrino et al. 2009).

DNA-based diet analyses can also provide more insight in the evolution of new subspecies of mammals. An example of such an application is research on killer whale (Orcinus orca) populations. These form stable matrifocal social groups. Hoelzel et al. (1998) found proof that two social groups, each with their own resource specialization, known as the residents and transients, show genetic differentiation. Moura et al. (2015) used a phylogenetic approach and found evidence that life history and behavioural changes associated with resource use led to lineage differentiation between both social groups, indicating incipient speciation. Here, phylogenetic analyses with DNA barcode data provided the opportunity to unravel the history of these social groups. This provided more insight in ongoing evolutionary processes already indicated by diet and behaviour.

**Future opportunities and challenges**

Despite the considerable progress that the development of new DNA techniques brought to evolutionary studies of mammals, several challenges still need to be tackled. Below, we will describe these in more detail.

For species delimitation, more DNA barcodes still need to be added to reference libraries. According to Bold (http://www.mammaliabol.org), about 2995 mammal species have currently been barcoded, while the IUCN global mammal assessment lists 5488 species. This means that some work still needs to be done to have all mammals in DNA barcode reference database eventually. Until this has happened, DNA barcoding will have limited usefulness as a tool for the discovery of new species, as this technique depends on comparing the barcodes of species sampled with barcodes tied to vouchered specimens (Rubinoff 2006). When no match is found, this does currently not mean that a new species is discovered as this conclusion can only be drawn once all possible matches can be made. Sharing databases and further centralizing still scattered information could be improved as well, as not everything is available in a publically accessible reference database yet. Finally, updating erroneous or incomplete taxonomic identifications in reference databases is needed (Groenenberg et al. 2011).

For indirect observational data, DNA extractions from challenging substrates such as air and water could be further improved. This could be done by optimizing DNA filtering and concentration techniques using pumps and filters and developing new kits and buffers and DNA techniques skipping amplification prior to sequencing (Prosser et al. 2015).

For diet analyses, the reliability of metabarcoding should be further validated by combining this approach with traditional morphological assays. Different results of molecular and morphological studies indicate that technical or biological biases are present causing absence of PCR success in the presence of a species (e.g. Cowart et al. 2015). More collaboration between molecular biologists and field-workers is recommended, as the latter still have the expertise needed to perform complex morphological surveys. This is also needed to solve current complications in relating different copy numbers of DNA barcode regions to estimates of abundance of individual dietary components (Deagle et al. 2013; Piñol et al. 2014a). A final challenge for DNA-based diet analyses lies in data processing. Bioinformatic pipelines are increasingly needed to identify reads due to the vast amount of data that cannot be handled by hand anymore. These pipelines currently require much scripting knowledge (Balzer et al. 2011) and should become more user-friendly. In view of ongoing developments, we envisage a rapid increase in overlap between the fields of bioinformatics, computational biology and evolutionary biology. This will encourage more scientists to apply DNA barcoding and metabarcoding in evolutionary studies.

**Acknowledgements**

We thank Lutz Bachmann, Per Ericson and Per Sundberg for their invitation to contribute to this special issue of *Zoologica Scripta* based on the contribution of BG to the symposium *Systematics and Biodiversity Research in the Era of Genomics* that was held in Oslo in November 2015. Bas Blankevoort and Erik-Jan Bosch are thanked for their help with the illustrations.

**Author contributions**

AH wrote a first draft of the manuscript; BG and HS helped her finish the final draft. All authors provided editorial assistance.

**References**


from coprolites. *Biological Journal of the Linnean Society, 66*, 57–74.


A.-J. Haarsma et al. • Added value of metabarcoding for evolutionary studies


