Year : 2023

Molecular assessment of wild vervet monkeys' diet to highlight social factors

Schneider Judith

Schneider Judith, 2023, Molecular assessment of wild vervet monkeys' diet to highlight social factors

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_21AB646BB1586

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Département d'Ecologie et Evolution

Molecular assessment of wild vervet monkeys' diet to highlight social factors

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Judith SCHNEIDER

Master de l'Université de Lausanne in Behaviour, Ecology and Evolution

Jury
Prof. Liliane MICHALIK, Présidente
Dr. Luca FUMAGALLI, Directeur de thèse
Prof. Erica VAN DE WAAL, Co-directrice de thèse
Prof. Lucie ZINGER, Experte
Prof. Redouan BSHARY, Expert

Lausanne
2023

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Madame	Prof.	Liliane	**Michalik**
Directeur·trice de thèse	Monsieur	Dr	Luca	**Fumagalli**
Co-directeur·trice	Madame	Prof.	Erica	**Van de Waal**
Expert·e·s	Madame	Prof.	Lucie	**Zinger**
	Monsieur	Prof.	Redouan	**Bshary**

le Conseil de Faculté autorise l'impression de la thèse de

Judith Schneider

Maîtrise universitaire ès Sciences en comportement, évolution et conservation, Université de Lausanne

intitulée

Molecular assessment of wild vervet monkeys' diet to highlight social factors

Lausanne, le 24 novembre 2023

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Liliane Michalik

Table of Contents

TABLE OF CONTENTS **3**

ACKNOWLEDGEMENTS **5**

SUMMARY **6**

RÉSUMÉ **6**

GENERAL INTRODUCTION **8**

A BRIEF HISTORY OF EDNA 8

EDNA APPLIED, SCOPE AND LIMITS 10

THE BEHAVIOURAL ECOLOGY OF VERVET MONKEYS 13

THE INKAWU VERVET PROJECT 14

EVALUATING THE INFLUENCE OF SOCIAL FACTORS VIA DIETARY ASSESSMENTS 17

THE INTEREST TO STUDY INTERGROUP VARIATION (IGV) 20

THESIS OUTLINE 22

CONTRIBUTIONS TO THE CHAPTERS 23

CHAPTER 1: FOCAL VS. FECAL: SEASONAL VARIATION IN THE DIET OF WILD VERVET MONKEYS FROM OBSERVATIONAL AND DNA METABARCODING DATA **24**

CHAPTER 2: MOLECULAR ASSESSMENT OF DIETARY VARIATION IN NEIGHBOURING PRIMATE GROUPS **41**

CHAPTER 3: INTERGROUP VARIATION IN WILD ADULT FEMALE VERVET MONKEYS **54**

ABSTRACT 55

INTRODUCTION 56

METHODS 58

Field work and samples 58

Laboratory and bioinformatics 58

Data analyses 60

RESULTS 61

DISCUSSION 68

ACKNOWLEDGEMENTS 71

GENERAL DISCUSSION 73

REFERENCES 81

APPENDICES 103

SUPPLEMENTARY DATA CHAPTER 1 103

SUPPLEMENTARY DATA CHAPTER 2 126

SUPPLEMENTARY DATA CHAPTER 3 135

PROJECT+: BAITS FOR INSECT FISHING 149

ABSTRACT 150

INTRODUCTION 151

BAIT DESIGN 155

EXPERIMENTAL DESIGN 158

Mock community preparation 158

eDNA samples 159

DNA enrichment 160

Perspective data analyses 163

CONCLUSION 164

ACKNOWLEDGEMENTS 167

SUPPLEMENTARY DATA PROJECT+ 168

ADDITIONAL ARTICLE 1: COMPREHENSIVE COVERAGE OF HUMAN LAST MEAL COMPONENTS REVEALED BY A FORENSIC DNA METABARCODING APPROACH 170

ADDITIONAL ARTICLE 2: ASSESSING ENVIRONMENTAL DNA METABARCODING AND CAMERA TRAP SURVEYS AS COMPLEMENTARY TOOLS FOR BIOMONITORING OF REMOTE DESERT WATER BODIES 178

Acknowledgements

I begin, of course, by thanking Luca, for years of support and empathy and everything I've learned since I've started my master's project in the group. And for giving me the opportunity to do this project and his unequivocal confidence; I am very conscious that it is (unfortunately) not common to sign a PhD contract seven months pregnant. And thanks to Erica, I am grateful for her steady encouragements and little pushes towards conference presentations, which I (almost) enjoyed in the end, and her contagious passion for her research. I really hope we get to go to the field one day together. You make a good supervising team and I am glad I could be part of both of your research groups. I also like to thank Lucie Zinger and Redouan Bshary for being the experts on my committee and for their precious advice during the mid-thesis exam and Liliane Michalik for still finding the time to chair the committee.

A huge thank you goes to the IVP team in the field for the sampling and organisation of samples delivery and to Loïc in particular for his great master project. And to all other group members of both research groups. Eduard, I am very grateful that we did this PhD journey together. Molly, thanks for digging through thousands of (sometimes smelly) monkey scats with me and making me laugh so much. Julie, we will manage to have our weekly coffee break, one day for sure. And Céline, your good mood and optimism got me through this last year. Many more colleagues, supervisors, friends throughout the master and PhD have accompanied and helped me at certain times or over the whole time, especially Camille-Sophie and Sarah, Fred and Pierre, Céline, Nadège, Eléonore, Roberto, Jérôme, Anne-Lyse and Lena, Pooja, Charlotte and Rachel, and Olivier.

I am grateful for my long-term friendships that survived international moves or constant lack of time; about voice messages that so often made my days, about summer traditions, about long walks along the lake. And for my whole big, loud and loving family. My four amazing sisters who are always there for me (I know that you might be tempted to call me Doc Snyder for ever now…but please don't!). To Gertrud and especially to Achim who thought a PhD was totally unnecessary and would be very proud no matter what. And finally, I am happy every day to have you in my life, Alain, Noé and Louna. Thank you for putting up with me; I only made it because of you.

Summary

The long-term survival of many species is challenged by the need to adapt to changing environments on evolutionary short timeframes. Behavioural diversity may enhance the capacity of adaptation to instable living conditions or, on the contrary, cause a certain inflexibility to respond to changes. Behavioural studies on broad spatial and temporal scales represent great challenges in the wild for direct observational methods. eDNA-based studies may provide complementary tools and help including behavioural or even cultural diversity into conservation management. The term eDNA refers to the genetic material that organisms leave in their environment. eDNA can be recovered by collecting any type of environmental sample, providing a wealth of information at the scales of species, populations, communities, complete ecosystems and biodiversity. In addition, being a non-invasive method, organisms can be studied without the need to observe, disturb or harm them.

For this PhD project, faecal samples of wild vervet monkeys (*Chlorocebus pygerythrus*) living in neighbouring groups in South Africa were analysed using DNA metabarcoding to link behavioural diversity to social factors. Chapter 1 compared arthropod and plant sequences from faecal samples and observational data of foraging on plants and insects, on the population level. We observed strong seasonal effects and a correlation between both methods, validating here the use of relative read abundances (RRA) as semi-quantitative measurements. In Chapters 2 and 3, the plant-based diet was assessed to investigate the part of intergroup variation (IGV) that cannot be explained by environment or genetics, indicating effects of social factors. In summer, we found IGV for all individuals, but also a positive correlation between group size and intragroup variation, indicating the importance of social dynamics. Therefore, Chapter 3 focused on adult females, the philopatric sex, to inquire about IGV across seasons. In addition, we developed a set of baits for targeted sequence capture (TSC) of insects (Project+). In this project, we have shown the potential of eDNA-based analyses to inform on behavioural diversity. This opens the possibility for further research on our study species as well as others. In addition, the combination of eDNA samples and TSC is a promising approach to comble the lack of absolute quantification of DNA metabarcoding. Overall, these are interesting tools and perspectives for behavioural ecology, evolutionary biology, and cultural evolution studies, as well as conservation management.

Résumé

La survie à long terme de nombreuses espèces est compromise par la nécessité de s'adapter à des environnements changeants dans des délais très courts. La diversité comportementale peut augmenter la capacité d'adaptation à des conditions de vie instables ou au contraire induire une

certaine rigidité face aux changements. Les études comportementales sur de grandes échelles spatiales et temporelles représentent un grand défi en milieu naturel pour les méthodes d'observation directe. Les études basées sur l'ADN environnemental (ADNe) peuvent fournir des outils complémentaires et aider à inclure la diversité comportementale ou même culturelle dans la gestion de la conservation. Le terme ADNe fait référence au matériel génétique que les organismes laissent dans leur environnement. L'ADNe peut être récupéré en collectant n'importe quel type d'échantillon environnemental, ce qui fournit une mine d'informations à l'échelle des espèces, des populations, des communautés et même des écosystèmes complets, ainsi que sur la biodiversité. En outre, la méthode étant non invasive, les organismes peuvent être étudiés sans qu'il soit nécessaire de les observer ou perturber.

Dans le cadre de ce projet de doctorat, des échantillons fécaux de singes vervet sauvages (*Chlorocebus pygerythrus*) vivant dans des groupes voisins en Afrique du Sud ont été analysés à l'aide du métabarcoding de l'ADN afin d'établir un lien entre la diversité comportementale et les facteurs sociaux. Dans le chapitre 1, nous avons comparé les séquences d'arthropodes et de plantes trouvés dans les échantillons fécaux et les données d'observation concernant la recherche de nourriture de plantes et d'insectes au niveau de la population. Nous avons observé de forts effets saisonniers et une corrélation entre les deux méthodes. Dans les chapitres 2 et 3, le régime alimentaire à base de plantes a été évalué pour investiguer la part de variation intergroupe (IGV) qui ne peut pas être expliquée par l'environnement ou la génétique, indiquant des effets de facteurs sociaux. En été, nous avons trouvé de l'IGV pour tous les individus, mais aussi une corrélation positive entre la taille du groupe et l'IGV, ce qui indique l'importance de la dynamique sociale. Par conséquent, le chapitre 3 s'est concentré sur les femelles adultes, le sexe philopatrique, pour analyser l'IGV à travers les saisons. En outre, nous avons développé un ensemble de sondes pour la capture de séquences ciblées (TSC) d'insectes (Project+). Dans ce projet, nous avons montré le potentiel des analyses basées sur l'ADNe pour informer sur la diversité comportementale. Cela ouvre la voie à des recherches plus approfondies sur les espèces étudiées et sur d'autres. En outre, la combinaison d'échantillons d'ADNe et de TSC est une approche prometteuse pour pallier le manque de quantification absolue du métabarcoding de l'ADN. Dans l'ensemble, il s'agit d'outils et de perspectives intéressants pour l'écologie comportementale, la biologie évolutive et les études sur l'évolution culturelle, ainsi que pour la gestion de la conservation.

General introduction

A brief history of eDNA

The multidisciplinary field of *environmental DNA* (eDNA) has been rapidly evolving and eDNA applications have become key to biomonitoring, research and conservation efforts[1]. The term refers to the genetic material that organisms leave, actively or passively, in their respective environment and which can be recovered by collecting environmental samples (such as soil, water, air). The source and state of this DNA can be manifold, e.g., urine, cell debris, faeces, hair, skin or carcasses[2]. After the extraction of eDNA from environmental samples it can be analysed using different molecular techniques, and assigned to taxa using genetic reference databases or be analysed as molecular operational taxonomic units (MOTUs). Consequently, eDNA-based approaches are considered non-invasive, since organisms can be detected without the need to isolate or disturb them in their respective environments. Starting from a single environmental sample, eDNA approaches open a wealth of information regarding species and populations, communities, and complete ecosystems and biodiversity.

The concept of eDNA was first used in metagenomics to study the microbial communities of marine sediments[3]. It took another decade to apply eDNA techniques to macroorganisms[4] and to sequence DNA extracted from environmental samples[5]. While early eDNA studies were mostly based on PCR followed by Sanger sequencing or on quantitative-PCR (qPCR), the advent of next-generation sequencing (NGS) technologies enabled the simultaneous sequencing of complex DNA mixtures in environmental samples, and therefore more exhaustive assessments[1].

Hereafter, the development of this field of research was facilitated and then rapidly implemented for various purposes such as the detection of invasive, rare and elusive species[6], or as a biomonitoring tool, thus representing today an important area of research for the future management and conservation of biodiversity[1,7–9]. In particular, DNA metabarcoding (PCR amplification with *universal* primers and next generation sequencing (NGS) of DNA mixtures[10]) is widely applied for comprehensive taxonomic identification of environmental samples.

DNA collected from environmental samples is often degraded and fragmented due to e.g. enzymatic and microbial activity, and the respective techniques of analyses need to be adapted to account for these characteristics. For example, the primers used for DNA metabarcoding are developed for short but informative genetic regions (metabarcodes). While called 'universal', they do in most cases target more restrained taxonomic groups, however, allowing for broad assessments of DNA pools. The fact that they are designed for taxonomic groups rather than single species, and therefore for genetically diverse organisms, leads to compromises between specificity, coverage and resolution. Before library preparation, PCR amplicons are usually pooled together (this is not per se necessary but increases sample throughput efficiently). To enable the attribution of sequences to samples after sequencing, metabarcoding primers are tagged with variable combinations of oligonucleotides[1].

eDNA applied, scope and limits

The multitude of eDNA studies continuously published these last years, highlights the methods value and interest (reviewed in e.g.[1,8]). In particular, DNA metabarcoding approaches have been widely used, providing valuable information on ecosystems, food webs, species' distributions and behaviour. After a first proof of successful species detection using eDNA in water samples[6], many applications of invasive species detection followed (e.g.[11–13]), as well as for the monitoring of rare and elusive species[14] and/or in difficult accessible terrain[15,16]. Research aims go from monitoring of reproduction activities[17], diet assessments[18–21] to simultaneous assessments of predators and prey, or endoparasites[22,23]. Multispecies assessments looked at the composition of communities[24], temporal changes in communities[25] or seasonal variations[26,27]. Among the most fascinating applications are inferences of population genetics of whale sharks using sea water samples[28,29] and the extraction of DNA from ancient lake sediments to deduce anthropologic developments[30]. The subfield of ancient, environmental DNA (aDNA) is a very active research field (reviewed in[31]). Besides commonly used sample types as water, soil or faeces, there are also more creative applications extracting DNA from honey[32], air[33], or leeches[34,35]. eDNA sampling may thus cover huge spatio-temporal scopes, and provide data to study multiple taxonomic groups and their interactions relying on the same sample collection.

DNA metabarcoding often provides higher taxonomic resolution and coverage than traditional methods[8,36]. Traditional survey methods, mainly based on direct observation of target organisms and their identification (in traps, faecal samples, or camera trap images) are often time-consuming and labour-intensive since they depend heavily on taxonomic expertise, long-time observation and tedious image coding, among others. eDNA-based approaches on the

contrary rely on standardisable, straightforward and cost-effective sample collection methods. The sensitivity of the method allows to detect organisms at low frequencies or which are difficult to observe as cryptic species or life stages, elusive and nocturnal species, representing a valuable complementary tool to traditional methods. Another important advantage of eDNA methods is that it is non-invasive for organisms as these are not physically damaged or disturbed, and ecosystems[2]. In a certain sense this is valid also for behaviour, in the sense that these are less affected by observational methods or experimental settings[37].

However, there is possibly variability in eDNA detectability, preservation and transport due to biotic and abiotic ecological factors and many studies inquired about such effects. In particular, the persistence as a proxy for the temporal signal of eDNA has been assessed, as well as the effect on DNA degradation of factors as pH and temperature (reviewed in[38]). And since eDNA can be shed by dead organisms as well as by all life stages, based on genetic information only, it is not possible to distinguish between e.g. eggs, larvae and adults or flowers, fruits and bark or alive and dead organisms[2]. Furthermore, the method implies certain biases, in particular concerning the PCR amplification step[39–41] which leads to problematic biomass quantifications[42–44]. Different primer affinities are caused by the characteristics of 'universal' metabarcoding primers, between specificity and coverage, and the genetically diverse DNA mixtures competing for the primers. Reliable measurements of abundances and biomass are useful or even indispensable in ecological studies, therefore quantifying DNA present in eDNA pools is very relevant. In diet studies, digestion adds more possible biases for quantitative interpretations[45–47]. In fact, the only abundance measure is relative, i.e. based on the relative number of reads per sequence per sample. Reliable correlations between relative sequence reads and initial abundances (e.g. number of individuals or foraged items) are

complicated. A number of studies that used qPCR to assess the correlation between eDNA and biomass confirm this[48,49]. Nevertheless, a correlation between read counts and biomass cannot absolutely be neglected[42] and relative read abundances (RRA; the number of a specific sequence divided by the total number of reads within a sample) has been employed to use DNA metabarcoding semi-quantitatively to inquire about foraging niches[19,50,51]. Beside ecological factors, different laboratory protocols and bioinformatic choices can play a role for the outcome of eDNA studies, therefore it is crucial to keep conditions stable and to standardise and control every step from the field to data analyses[52]. PCR amplification is considered causing the most biases in DNA metabarcoding studies since different primer affinity can lead to preferred amplification of similar sequences and at the same time non-amplification of more divergent ones[40,53]. The composition of eDNA pools might also play a role via primer competition or inhibition[39]. To overcome these issues, the multiplexing of primers or the use of degenerate primers has been proposed[54–56]. Alternatively, approaches are also proposed that skip the amplification process, and use sets of baits instead of primers that hybridise to and capture DNA (see Project+). The combination of target sequence capture (TSC) and NGS has been suggested as convenient alternative to overcome PCR-induced biases and provide more reliable quantification measures[57,58]. The principle here is to create DNA or RNA baits that are complementary to target sequences. Single stranded DNA is then hybridised to these biotinylated oligonucleotides and physically bound (“captured”) and off-target DNA subsequently washed away before being processed for sequencing[59,60]. As such, TSC follows a different logic to achieve target enrichment than PCR amplification, since the increase in target sequences is achieved by reducing non-target sequences[61].

The behavioural ecology of vervet monkeys

Here, we used DNA metabarcoding sequence data to study the diet of wild vervet monkeys (*Chlorocebus pygerythrus*) for three out of the four present chapters to answer various research questions. Vervet monkeys are non-threatened primates widely distributed throughout Southern-Eastern Africa below the equator (Fig. 1). Their variable habitats include savannah biome, coastal zones as well as urban areas. This semiarboreal species lives in multimale multifemale groups. The social system of vervet monkeys is in general based on a complex hierarchy organised by matrilines, with females being philopatric. In the studied population, female dominance has been often documented in direct interactions between all group members and females held the alpha position in most groups[62]. Male vervet monkeys disperse often multiple times throughout their lives. The constant gene flow between groups due to male dispersal reduces genetic intergroup differences[63]. Many experiments have demonstrated the high social learning abilities of vervet monkeys[64,65] (see below). Vervet monkeys are omnivorous feeders; and selective behaviours, i.e. preferential feeding on certain resources, is supposable. Behavioural traits not correlated to environmental or genetic factors might be transmitted socially through social learning[66], according to the method of exclusion that postulates that by excluding ecological and genetic explanations, sole sociality remains[67]. The study of dietary variation might therefore allow to trace the existence of socially learnt transmission in a social group.

Figure 1. Distribution of vervet monkeys (Chlorocebus pygerythrus) in Africa; ©Maphobbyist, with distribution data from the IUCN Red List. The red monkey symbol indicates location of the field site of the iNkawu Vervet Project (IVP).

The iNkawu Vervet Project

Faecal samples have been collected over a period of four years at the iNkawu Vervet Project (IVP), Mawana game reserve (28°00.327S, 31°12.348E), South Africa (Fig. 1). At this research station up to seven groups of vervet monkeys have been habituated and studied for behavioural data collection since 2010 (Fig. 2). Here data was included for five of these groups: Ankhase (AK), Baie Dankie (BD), Kubu (KB), Lemon Tree (LT) and Noha (NH). All individuals are recognisable by specific body and face characteristics (e.g. facial features, scars and fur colours) and are named at birth and life history data is continuously collected. The monkeys are closely followed by the field team for various research activities (e.g. focal observations or behavioural experiments). Faecal samples are assigned to individuals by observation and can hence be linked to the individuals' life history data (always sex and current group, and most often also available age, natal group, mother). Age categories have been assigned to

predefined rules for females as for males (infant <1 year, juvenile 1-2 years for females, and 1-4 years for males (if not dispersed), adult 3 years for females and 4 years for males if they dispersed, otherwise 5 years[68]). Different research questions can hence be studied at different levels: from the population to the group and even the individual. In this thesis, we assessed both the plant and arthropod components of the diet for 1745 and 823 faecal samples, respectively. In addition, 12 months of focal (observational) data have been analysed. We also used environmental data such as plant cover, as these can be linked to the groups' territories, for which home ranges have been established (Fig. 3, see all methodological details in Chapters 1 and 2). In the home ranges there is a mix of semiarid riparian woodland and mosaic savannah biome, that is open grasslands with clusters of trees, thornveld and bushveld[69]. The assessment of plant cover data indicated no significant differences between the territories of four groups (AK, BD, KB, NH); these data were not available for LT. The homogenous environment and the constant gene flow between groups, which limits the influence of genetic factors, allow to study other causes of behavioural variations[68].

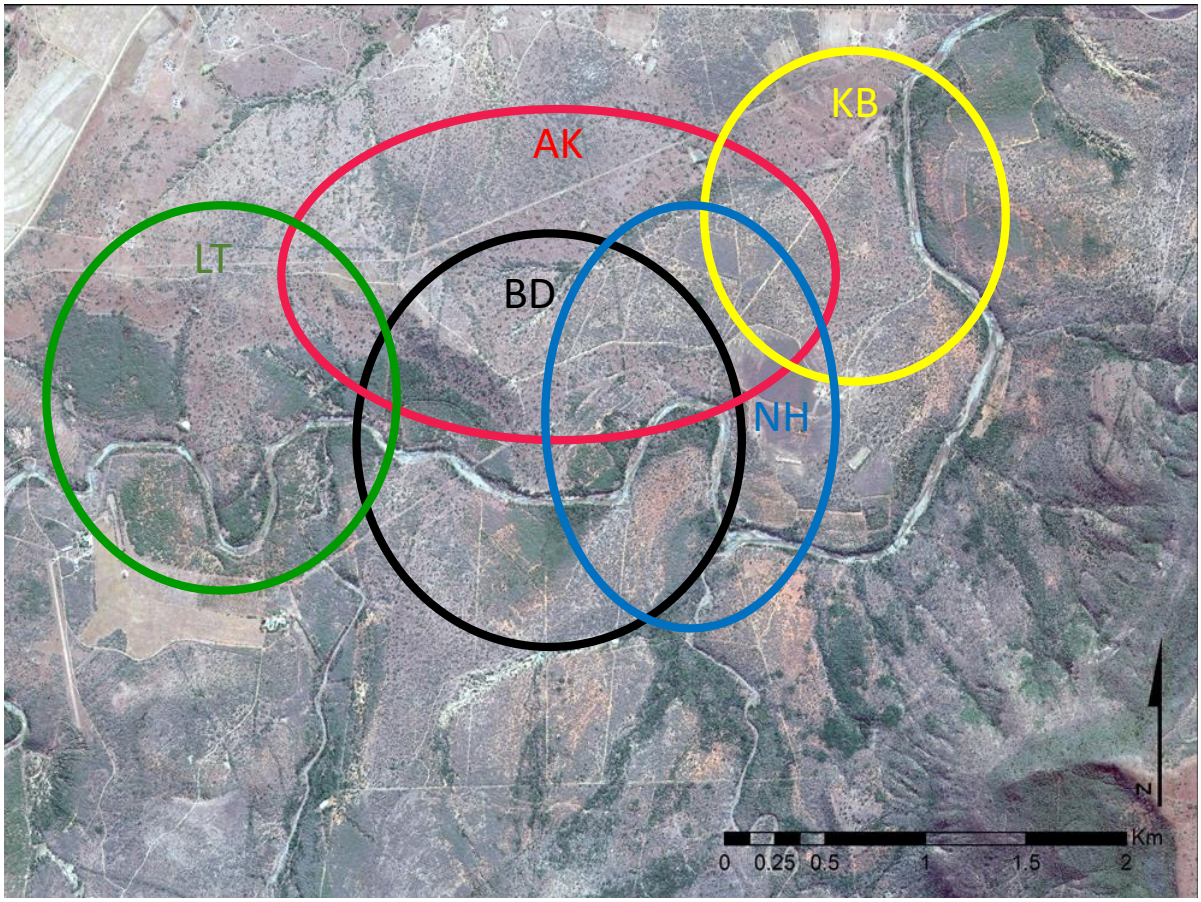


Figure 2: Aerial view of the average home ranges of the five studied groups at iNkawu Vervet Project (IVP). Ankhase (AK), Baie Dankie (BD), Kubu (KB), Lemon Tree (LT) and Noha (NH).

In addition to providing valuable insights on the feeding behaviour of vervet monkeys, our study system enabled us to assess the power of DNA metabarcoding as semi-quantitative measurement. In a second time we were interested in the transferability of the method to similar research questions for other species in different contexts (e.g. with less accessible field and/or less observation-prone organisms or behaviour). If the method allows to highlight the effect of social factors on behavioural dietary plasticity in our study system, it could be employed as a means of capturing cultural diversity in a wider context for consideration in conservation measures.

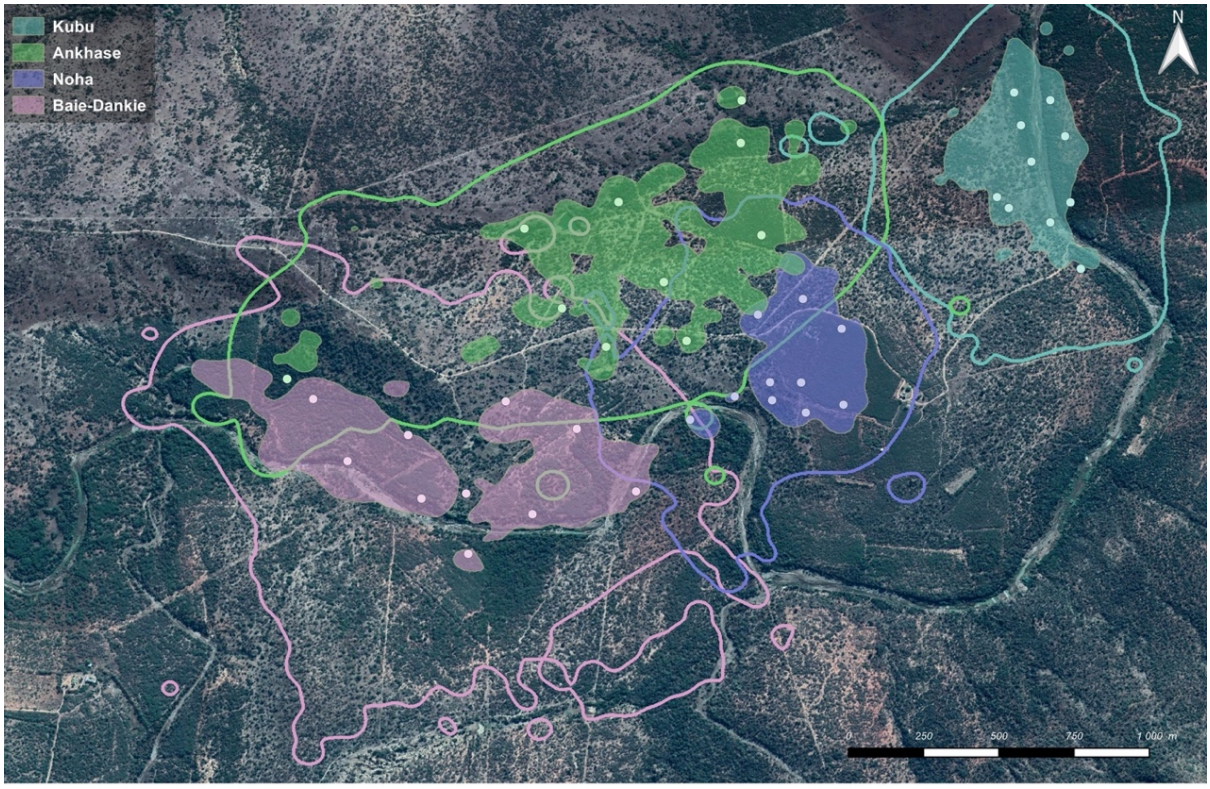


Figure 3: Home ranges (50 % and 95 % isopleths) of four groups based on GPS data of faecal and focal sampling locations. Points indicate the location of the 40 vegetation plots where plant coverage has been assessed (Figure taken from Supplementary Material of[70]). The group Lemon Tree (LT) was not included in the vegetation assessment.

Evaluating the influence of social factors via dietary assessments

In contrast to individual learning, social learning is defined as "learning that is influenced by observation of, or interaction with, another animal (typically a conspecific) or its products"[71]. Analogous to genetic transmission, social transmission can influence behaviour (albeit along many more axes than only from parents to offspring) and create traditions if the behaviour persists over time (also labelled "secondary inheritance system"[72]). Distinct traditions in turn may define cultural traits of e.g. populations or groups. Social learning is thus a prerequisite for animal culture, being the mechanism by which cultural traits are acquired from

conspecifics[73]. Behavioural plasticity can be an adaptive potential when facing changing environments, e.g. mediated through climate change or anthropogenic habitat destruction[74,75]. In this sense, and similar to hotspots of genetic diversity, cultural traits represent specific features of populations/groups relevant for conservation efforts and for the identification of the most vulnerable or conservation-prone population/group[76–78].

However, studying mechanisms and biases of social learning is a major challenge, all the more difficult in the wild. Many studies are conducted in captive environments (reviewed for zoos in[79]). While captive settings allow to control confounding factors and mostly the tested individuals, there might be issues concerning the representativity of captive studies for natural behaviour[80]. In an attempt to discern this issue, Harrison et al. (2023) tested a cognitive experiment using touchscreen technology on wild and captive vervet monkeys and found no difference in performance but lower participation rates in the wild correlated with sex and age[80]. Most social learning studies in the wild rely on either indirect observations, i.e. the observation of behavioural variation not driven by environment or genetics, or on experimental testing of transmission mechanisms. Observational evidence for social transmission of behavioural variants have been documented first in chimpanzees[81] and then in many species including in orangutans[82,83], cockatoos[84], whales and dolphins[73], as well as by acoustical observations of songs in humpback whales[85]. Studies using experimental testing to trace social learning mechanism and diffusion of social information in the wild have been done for example with great tits[86–88].

Foraging behaviour is particularly well suited for the study of social learning. Indeed, it is essential that offspring is quickly able to extract food resources from their environment, with their mothers most of the times serving as their primary model for acquiring the most adaptive

foraging behaviours[89,90]. Generally in primates, infants learn from their mothers what to eat, where to find food items and how to manipulate them, i.e. by vertical transmission of knowledge[64,83,91]. An individual's diet may be influenced by demographic parameters such as rank or group size in the case of strong competition, or when foraging behaviour is strongly mediated by social learning[92,93]. That means in a reverse thinking that social learning could be deduced by studying the diet and possible effects of social factors. In great tits, social transmission of a specific foraging behaviour (in that case, sliding the door of a foraging device in the same way as a trained model) was documented in the wild[88]. In vervets, field experiments showed as well that social learning can efficiently be assessed via the study of feeding behaviours. For example, a selective attention towards female demonstrators was shown in a puzzle box experiment, where the side to open the box displayed by dominant females was copied, i.e. socially learned, by conspecifics but not when the demonstrator was a dominant male[94]. Another experiment used different coloured corn, with different taste during the training phase (normal vs bitter), to show that infant vervet monkeys rely on their mothers experience regarding food choice and also that dispersing males conformed to their new group food preference[95]. Whereas a group effect on the infants, i.e. horizontal transmission, could not be excluded in the corn study, an experiment on food-manipulation focused specifically on naïve infants and highlights the vertical transmission of knowledge from mothers to infants[96].

While all of these studies were conducted in the wild, they relied on artificial foraging tasks or altered food items. On the one hand, these may also influence the studied behaviour and on the other hand, they do not reflect natural feeding behaviour. For instance, social tolerance, competition and variable degrees of habituation to humans are factors that might play a role

for participation in experimental testing and setups[80]. While natural foraging may also be influenced by e.g. rank and social tolerance, an approach based on eDNA can benefit from the temporal scale of eDNA samples and it may provide a good means to study natural behaviour in the wild. The non-invasiveness of eDNA sampling is another advantage in this context since in particular for endangered species the exposure to artificial disturbances of their behaviour should be carefully considered[80]. We understand invasive as an alteration or damaging of the organism, environment or behaviour, and not the simple presence of observers as has been argued elsewhere[37]. Harrison et al. (2023) pointed out that different habituation degrees between the studied groups may impact participation rates in their touchscreen study[80]. The observation in our study concerns defecation events only and not the behaviour under study, i.e. the feeding events which take place mostly during unobserved periods. However, a possible bias due to different habituation degrees of groups/individuals cannot be totally excluded in the sense that certain individuals may be easier to observe and sample. It remains challenging to study social learning in the wild; nevertheless, the potential of an eDNA approach to assess the influence of social factors on dietary variation as a proximate can be evaluated in our study system. If successful and depending on the accessibility of faecal samples, the concept could be employed to assess behavioural plasticity in foraging behaviour of species subject to learn socially but difficult to observe or test experimentally[68].

The interest to study intergroup variation (IGV)

In Chapters 2 and 3, we assess diet at the group-level to determine whether there is intergroup variation (IGV) that cannot be explained by environmental or genetic factors, and which might therefore indicate the effects of social factors. IGV is understood here as the variation in

behaviour that can be observed between social groups, i.e. intraspecies plasticity at the group level[92]. The group variable is an influential social factor in group living species. And the group level is an important intermediate level between individuals and populations, that should be considered more extensively to avoid species-wide generalisations[92,97]. Kaufhold and van Leeuwen (2019) define *cultural* intergroup variation as variation in behaviours across groups due to social learning, which can be qualitative, i.e. presence or absence of behaviours, and quantitative, expressed through the frequency of behaviours. To better understand the dynamics behind IGV, it is therefore also necessary to consider the driving forces at play at the intragroup level. Ecological, genetic of social explanations may be at cause for the emergence of group-specific traditions. Behavioural plasticity in terms of foraging between groups that share similar environments and genetics can serve here as a proxy indicating possible social transmission. Therefore small-scale studies are very relevant here and comparisons between groups in the wild can serve to highlight the influence of social factors on behaviour and cognition[80]. In vervet monkeys there seems to be a group conformity effect, which underlines the importance of group integration as social factor. In the above mentioned corn study, it was also observed that newly immigrant males abandoned the food preference of their natal group to conform to the one of their new group[95]. Such greater behavioural flexibility of the dispersing sex than of the philopatric one regarding role models was confirmed in an artificial fruit experiment[98]. In turn, the social structure in philopatric species favours distinct group identities[97]. In vervets, the philopatric sex, females, have been shown experimentally to be the preferred role models[94,96]. Therefore, we focus on the samples of adult females to discern a group identity and possible intergroup variation in foraging behaviour.

There are few eDNA-based studies so far that assessed differences beyond the population level or linked individual samples to social factors. As an exception, Voelker et al. (2020) used it to assess intrapopulation feeding diversity of harbor seals (*Phoca vitulina*) based on the predators' sex[99]. We suggest that assessing the effect of social factors on behaviour via eDNA data can provide a valuable tool for conservation management, behavioural ecology and even cultural studies.

Thesis outline

It this thesis, we first assessed the validity and reliability of eDNA data by comparing it to observational data. Thereafter, we used DNA metabarcoding data to inquire about variation in feeding behaviour between social groups. Despite the above-mentioned uncertainties of the eDNA approach, we consider that by standardising field and laboratory conditions for all samples allows for semi-quantitative comparisons based on RRA[70]. Nonetheless, we also developed bait set to capture the taxonomic group of insects to assess the impact and possibly overcome the impact of certain biasing factors of DNA metabarcoding.

In Chapter 1, we compared arthropod and plant sequence data of 823 faecal samples and focal data on foraging on plants and insects both sampled over 12 months, in order to analyse seasonal variation in the diet of vervet monkeys. On the population-level, we show a strong seasonal effect on vervet monkeys' diet based on both datasets, and a strong correlation between plant RRA and observational data, validating the use of RRA as a semi-quantitative measure of consumption in this system.

The results of the seasonal assessment (Chapter 1) showed that selective feeding behaviours are more likely to occur in summer when resources are more abundant than in scarcer seasons[70]. Chapter 2 therefore focused on assessing IGV between four neighbouring groups of vervets during the summer season. For this mostly conceptual study, we assessed a) whether IGV in the diet was greater than intragroup variation; b) whether IGV was greater when considering all individuals in the group or only adult females, the philopatric sex; and c) whether there was a relation between the availability of single food items per home range and their consumption.

Following the results of Chapter 2, we increased the sample size and focused on adult females to inquire about IGV on a broader scale and all seasons in Chapter 3. Here we included 1226 samples of 87 individuals, sampled from January 2018 to September 2022.

In Project+, we describe the development of a set of baits designed to capture the *Folmer* region of COI of the whole taxonomic groups of insects and outline the experimental design to compare targeted sequence capture (TSC), DNA metabarcoding and shotgun sequencing using artificial mock communities and eDNA samples.

Contributions to the chapters

I contributed to the generation of the datasets and data analyses for Chapters 1-2-3. I was responsible for the DNA extraction, amplification, library preparation and sequencing for Chapter 3. I contributed to the design, insect sampling and preparatory laboratory analyses, as well as the bait design of Project+. I wrote the first draft of all chapters.

Chapter 1: Focal vs. fecal: Seasonal variation in the diet of wild vervet monkeys from observational and DNA metabarcoding data

Loïc Brun[1,§], Judith Schneider[1,§], Eduard Mas Carrió[1], Pooja Dongre[2,3], Pierre Taberlet[4,5], Erica van de Waal[2,3,‡] and Luca Fumagalli[1,6,‡*]

[§] joint first authors

[‡] joint last authors

[1]Laboratory for Conservation Biology, Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland.

[2]Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland.

[3]Inkawu Vervet Project, Mawana Game Reserve, Swart Mfolozi, KwaZulu Natal, South Africa.

[4]Université Grenoble Alpes, CNRS, Laboratoire d'Ecologie Alpine, 38000 Grenoble, France.

[5]UiT – The Arctic University of Norway, Tromsø Museum, Tromsø, Norway.

[6]Swiss Human Institute of Forensic Taphonomy, University Centre of Legal Medicine Lausanne-Geneva, Lausanne University Hospital and University of Lausanne, Ch. de la Vulliette 4, 1000 Lausanne 25, Switzerland.

*corresponding author: Luca Fumagalli, luca.fumagalli@unil.ch.

Chapter status: published, https://doi.org/10.1002/ece3.9358

Author contribution
EW and LF designed the study and supervised all the analyses. LB and PD conducted the field work. EM-C, JS, LB and PT performed the laboratory work and bioinformatics. LB and JS conducted the data analyses. LB and JS wrote the first draft of the manuscript, with input from all other authors.

Received: 14 January 2022 | Revised: 12 August 2022 | Accepted: 5 September 2022

DOI: 10.1002/ece3.9358

Ecology and Evolution Open Access WILEY

RESEARCH ARTICLE

Focal vs. fecal: Seasonal variation in the diet of wild vervet monkeys from observational and DNA metabarcoding data

Loïc Brun[1] | Judith Schneider[1] | Eduard Mas Carrió[1] | Pooja Dongre[2,3] | Pierre Taberlet[4,5] | Erica van de Waal[2,3] | Luca Fumagalli[1,6]

[1]Laboratory for Conservation Biology, Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Switzerland

[2]Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Switzerland

[3]Inkawu Vervet Project, Mawana Game Reserve, Swart Mfolozi, KwaZulu Natal, South Africa

[4]Laboratoire d'Ecologie Alpine, Université Grenoble Alpes, CNRS, Grenoble, France

[5]UiT – The Arctic University of Norway, Tromsø Museum, Tromsø, Norway

[6]Swiss Human Institute of Forensic Taphonomy, University Centre of Legal Medicine Lausanne-Geneva, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

Correspondence
Luca Fumagalli, Laboratory for Conservation Biology, Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland.
Email: luca.fumagalli@unil.ch

Funding information
Branco-Weiss Fellowship-Society in Science; H2020 European Research Council, Grant/Award Number: 949379; Swiss National Science Foundation (SNSF), Grant/Award Number: 310030_192512 and PP00P3_170624; University of Lausanne

Abstract

Assessing the diet of wild animals reveals valuable information about their ecology and trophic relationships that may help elucidate dynamic interactions in ecosystems and forecast responses to environmental changes. Advances in molecular biology provide valuable research tools in this field. However, comparative empirical research is still required to highlight strengths and potential biases of different approaches. Therefore, this study compares environmental DNA and observational methods for the same study population and sampling duration. We employed DNA metabarcoding assays targeting plant and arthropod diet items in 823 fecal samples collected over 12 months in a wild population of an omnivorous primate, the vervet monkey (*Chlorocebus pygerythrus*). DNA metabarcoding data were subsequently compared to direct observations. We observed the same seasonal patterns of plant consumption with both methods; however, DNA metabarcoding showed considerably greater taxonomic coverage and resolution compared to observations, mostly due to the construction of a local plant DNA database. We found a strong effect of season on variation in plant consumption largely shaped by the dry and wet seasons. The seasonal effect on arthropod consumption was weaker, but feeding on arthropods was more frequent in spring and summer, showing overall that vervets adapt their diet according to available resources. The DNA metabarcoding assay outperformed also direct observations of arthropod consumption in both taxonomic coverage and resolution. Combining traditional techniques and DNA metabarcoding data can therefore not only provide enhanced assessments of complex diets and trophic interactions to the benefit of wildlife conservationists and managers but also opens new perspectives for behavioral ecologists studying whether diet variation in social species is induced by environmental differences or might reflect selective foraging behaviors.

KEYWORDS
diet estimation, DNA metabarcoding, environmental DNA, method comparison, primates, seasonal variation

Loïc Brun and Judith Schneider joint first authors; Erica van de Waal and Luca Fumagalli joint senior authors.

TAXONOMY CLASSIFICATION
Behavioural ecology, Biodiversity ecology, Conservation ecology, Conservation genetics, Ecological genetics, Genetics, Population ecology, Trophic interactions

1 | INTRODUCTION

Assessing a wild organism's diet is key to understanding its ecology and to highlight dynamics of communities and ecosystems through species' trophic interactions (Duffy et al., 2007). Traditionally employed methods, e.g. direct observations, microhistology of feces or gut contents, fatty acid, and stable isotope analysis, encounter certain limits when analyzing the diet of generalist and omnivorous species or attempting to disentangle the structure of complex food webs (Nielsen et al., 2018; Pompanon et al., 2012). The advent of DNA metabarcoding (Taberlet et al., 2012) and the simultaneous assessment of heterogeneous species mixes provide a valuable technique to open new perspectives in ecological network analysis (Clare, 2014). DNA metabarcoding studies using feces cover a range of different aims, such as diet characterization (Burgar et al., 2014; De Barba et al., 2014; Shehzad et al., 2012), parallel prey and predator identification (Galan et al., 2018; Gillet et al., 2015), or biodiversity assessment (Nørgaard et al., 2021; Shao et al., 2021). Some studies include different variables such as endoparasites and sex ratios along with the diet (Swift et al., 2018), or the predator's population structure (Bohmann et al., 2018). For many research questions in ecology, robust estimations of biomass or abundances are necessary for meaningful results going beyond simple detection or non-detection (Pimm et al., 2014). Therefore, a number of studies show the method's potential for assessing complex correlations relying on its semi-quantitative explanatory power when studying, for example, niche partitioning (Arrizabalaga-Escudero et al., 2018; Kartzinel et al., 2015; Pansu et al., 2019; Vesterinen et al., 2018) or intrapopulation variation (Voelker et al., 2020).

In many cases, reliable abundance data can be obtained by observation; however, there is an ongoing debate about the quantification potential of eDNA-based methods (Deagle et al., 2019; Zinger et al., 2019). For example, PCR primer-induced biases, i.e. the preferential amplification of certain taxa and the under- or non-representation of others, are considered a main source of biases in DNA metabarcoding (Jusino et al., 2019; Piñol et al., 2015; Piñol et al., 2019). Data treatment also influences the outcome (Calderón-Sanou et al., 2019); occurrence data supposedly inflate rare taxa but are less sensitive to PCR-introduced biases whereas the use of relative read abundances (RRA) may better account for variations in biomass (Deagle et al., 2019). RRA correspond to the number of reads of a sequence in a sample divided by the total number of reads of the same sample. Relative data do not only account for the presence of taxa in a sample but are expected to correlate to some extent with the amount of DNA present in the sample, therefore representing a semi-quantitative approach. In this study, we used RRA data, maintaining identical experimental conditions for all samples to minimize biases and to allow for comparisons.

FIGURE 1 Juvenile vervet monkey (*Chlorocebus pygerythrus*) feeding on fruits of *Ziziphus mucronata*. © Michael Henshall.

The DNA metabarcoding approach has been used only recently for diet studies in primatology, as the research field has traditionally relied on various observational methods for behavioral studies (but see Lyke et al., 2018; Mallott et al., 2017, 2018; Mallott et al., 2015; Osman et al., 2020; Quéméré et al., 2013; Rowe et al., 2021). Inter-method comparisons are useful to test different methods' reliabilities and congruencies to assess consistency of results. However, the aim is not only to compare performances but also to determine under which circumstance the complementary use of these methods is advisable to allow their optimal application in future studies. Since in many cases observational feeding data are available, but with weak taxonomic resolution and/or with a limitation due to feeding habits that are difficult to observe, complementing these data by a DNA metabarcoding approach may be beneficial.

To this aim, we compared dietary variation inferred from DNA metabarcoding to direct observations, in an opportunistic and generalist primate, the vervet monkey (*Chlorocebus pygerythrus*, Figure 1). Vervet monkeys are omnivorous and previous observational studies found that they feed mainly on trees, invertebrates, and occasionally small vertebrates (Barrett, 2009; Tournier et al., 2014). We analyzed 823 fecal samples of 130 individuals from four neighboring wild groups collected over 1 year, with two DNA metabarcoding assays targeting plant and arthropod components of the diet. The study of omnivorous species represents certain challenges (Tercel et al., 2021) that will be addressed in the discussion. The aim of the present study was threefold: (a) compare taxonomic coverage and resolution between observational and DNA metabarcoding data, (b) establish the most complete dietary profile in a wild vervet monkeys' population, and (c) assess resource use by vervet monkeys across seasons.

2 | MATERIALS AND METHODS

2.1 | Study site and subjects

The study was conducted between 09/2017 and 02/2019 as part of the Inkawu Vervet Project (IVP) in the Mawana game reserve (28°00.327S, 031°12.348E), KwaZulu Natal, South Africa. IVP was founded in 2010 and research has been conducted ever since on wild vervet monkeys mainly in the field of behavioral ecology, demonstrating the high social learning capacity of this species (Whiten & van de Waal, 2018). Our study includes four neighboring groups that are routinely followed by researchers. All individuals were identified using specific bodily and facial features (e.g. scars, colors, shape). The vegetation of the study site is classed as Savannah biome, characterized by areas of grasslands with dispersed singular or clusters of trees forming a mosaic with the typical savannah thornveld, bushveld, and thicket patches (Mucina & Rutherford, 2006). Each dataset, observational and DNA metabarcoding data, covered a period of 12 months, but they overlapped for 6 months only due to temporary constraints on focal sampling activities. Meteorological data assessed for the whole sampling period do not show major variation between the two sampled years for rainfall and temperature (Appendix S1: Figure S1). Therefore, we expected season to have a greater impact in terms of vegetation variation than the year of sampling and we consequently compared the data per month/season regardless of the year. Seasons were defined as follows, with the middle of a month as the seasonal delimitation (van Wyk & van Wyk, 2013): August–November (spring), November–March (summer), March–May (autumn), and May–August (winter).

2.2 | Observational data

The observational data used for this study were obtained by instantaneous focal animal sampling methods on 101 adult group members between 09/2017 and 08/2018. In focal samplings, the focal individual is followed for a defined period and occurrences of (inter)actions are recorded, but parameters can vary according to specific study designs (Altmann, 1974). Here, each focal sample lasted 20 min and the focal animal's behavior was recorded instantaneously every 2 min resulting in 10 data points per focal sample (6176 focal screenings in total). Observers chose focal animals opportunistically, with the aim to collect one full focal sample per individual across three different time windows (morning, midday, afternoon), every 10 days. Total length of the data collection periods per day varied throughout the year according to sunrise and sunset times, while being equally distributed between the three daily time windows covering all daylight hours. Prior to data collection, all IVP observers had to pass an inter-observer reliability test with a minimum Cohen's kappa value of 0.8 for each data category with an experienced researcher. Data were collected on tablets (Vodacom Smart Table 2, equipped with Pendragon Forms version 8). From the complete dataset, we extracted all feeding observations and created separate datasets for plant and arthropod items. The focal dataset for plants contained 19,406 observations, of which 12,315 identified plant genera or species (63.46%). The arthropod dataset contained 1359 observations (of which 15.82% indicated broad insect categories, i.e. termites or grasshoppers). Plant and arthropod observations that only occurred once were omitted from the final dataset.

2.3 | Local plant database

In the field, 54 plant species were morphologically identified and collected (van Wyk & van Wyk, 2013). These include all species confirmed by previous observation of feeding behavior in the area and other frequently occurring plants that could potentially be consumed. Sampled material from each species was stored in silica gel until DNA extraction using the DNeasy Plant Mini Kit (Qiagen) with a final elution in 100 µl AE buffer. To construct a local database, the whole chloroplast *trn*L (UAA) intron, which comprises the P6 loop targeted in the DNA metabarcoding assay described below, was amplified with primers c/d (Taberlet et al., 2007). The PCR reactions were performed in 25 µl. The mixture contained 1× PCR Gold Buffer (Thermo Fisher Scientific), 2 mM $MgCl_2$, 0.2 mM of dNTPs, 0.04 µg of bovine serum albumin (Roche Diagnostics), 0.5 µM of forward and reverse primers, 1 U of AmpliTaq Gold (Thermo Fischer Scientific), and 2 µl of template DNA. PCR cycling conditions were 10 min denaturation at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at 50°C, and 1 min at 72°C, with a final elongation step of 5 min at 72°C. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) before Sanger sequencing in both directions at Microsynth AG. The obtained P6 loop sequences were used for our reference database. The final database consisted of 48 sequences matching 54 species (i.e. 43 unique sequences, four sequences shared between two species and one sequence shared between three species, Appendix S1).

2.4 | Fecal sample collection

A total of 823 fecal samples of 130 known individuals were collected during a 12-month period (03/2018 to 02/2019, Figure 2). Whenever a specific individual was observed defecating, the inner part of the scat was immediately collected unless it had already been sampled the same day or if an experiment involving food rewards had been conducted with the group in the 48 preceding hours. Approximately 0.5 cm^3 was collected with gloves and a disposable plastic spoon from inside the scat into 20 ml HDPE scintillation vials (Carl Roth GmbH) and covered with 10 ml absolute ethanol. After 24–36 h, the ethanol was replaced by silica gel beads and samples stored until DNA extraction.

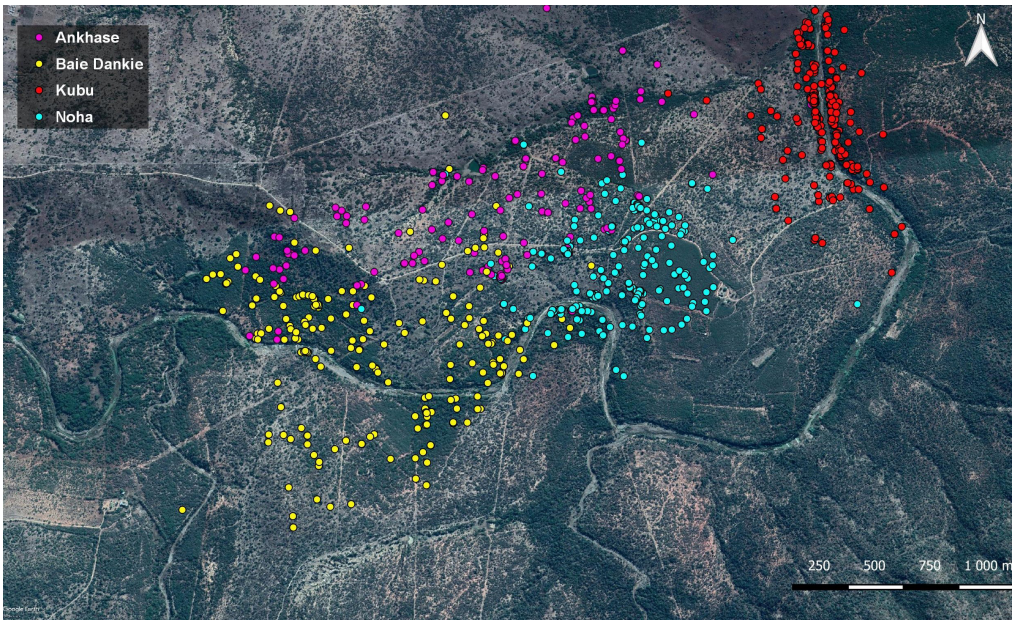


FIGURE 2 The map indicates the sampling locations of the 823 fecal samples of 130 individuals in the Inkawu Vervet Project, South Africa. The different groups are represented by different colored dots: Ankhase: purple (n = 146), Baie Dankie: yellow (n = 212), Kubu: red (n = 224), Noha: blue (n = 241).

2.5 | DNA metabarcoding

2.5.1 | DNA extraction

DNA extraction of scat samples was performed using a phosphate buffer-based approach (Taberlet et al., 2018) following a modified protocol of the NucleoSpin Soil Kit (Macherey-Nagel). Scats were directly transferred from the scintillation vials into 2 ml Eppendorf tubes with 1.3 ml of saturated phosphate buffer. For a better absorption of the DNA, the samples were homogenized by vortexing before spinning on a tube rotator for 15 min. The suite of the protocol was as recommended using the QIAvac technology (Qiagen), with a final elution in 100 µl of SE buffer. Extractions were performed in a pre-PCR laboratory exclusively dedicated to low DNA-content analyses (Laboratory for Conservation Biology, University of Lausanne). A subset of the extractions was tested for inhibitors with real-time quantitative PCR (qPCR) applying different dilutions in triplicates. qPCR reagents and conditions were the same as in DNA metabarcoding PCR (see below), but for 45 cycles and with the addition of SybrGreen (Thermo Fischer Scientific). Following these analyses, all samples were diluted 5-fold.

2.5.2 | DNA metabarcoding assay

DNA extracts were amplified in triplicates with two sets of primers. The first one targets the plant components of the diet amplifying the P6 loop of the *trn*L intron (UAA) of chloroplast DNA (10–220 bp, Sper01 (Taberlet et al., 2018) corresponding to g/h (Taberlet et al., 2007)). The second primer pair amplifies a fragment of 16S mitochondrial rDNA within the phylum Arthropoda (76–168 bp, Arth02 (Taberlet et al., 2018)). For the latter, one blocking oligonucleotide (5′-AGGGATAACAGCGCAATYCTAT TCTAGAGTC-C3-3′) was added, designed specifically for this study to limit the amplification of both human and vervet monkey DNA (for specifications see Appendix S1: Figure S2 and Taberlet et al., 2018). PCR reactions were performed in a final volume of 20 µl in 384-well plates. The mixture contained 1 U AmpliTaq Gold 360 mix (Thermo Fischer Scientific), 0.04 µg of bovine serum albumin (Roche Diagnostics), 2 µM of human-blocking primer (coupled with Arth02 primers only), 0.2 µM of tagged forward and reverse primers (i.e. primers with eight variable nucleotides added to their 5′-end, allowing sample identification), and 2 µl of template DNA. PCR cycling conditions were 10 min at 95°C, followed by 40 cycles of 30 s at 95°C, 30 s at 49 or 52°C for Arth02 and Sper01, respectively, and 1 min at 72°C, with a final elongation step of 7 min at 72°C. For each assay, extraction negative, PCR negative (H_2O), and positive controls as well as blanks were included. The positive controls of DNA mixtures of known concentrations were added in order to control for amplification success and were composed of species not expected in the study site (Appendix S1: Table S2), sequences were added to the respective databases. The inclusion of blanks, i.e. completely empty wells, allows to detect

artifactual sequences after tag switches during the sequencing process (Schnell et al., 2015). Amplification success was verified for a subset of samples, using the QIAxel technology (Qiagen). All PCRs were performed at the Laboratoire d'Ecologie Alpine (LECA).

PCR reactions were pooled per replicate before library preparation, i.e. resulting in six separate libraries (i.e. three per metabarcode) each containing 823 samples plus controls. Amplicon pools were purified using the MinElute PCR Purification Kit (Qiagen) and quantified using a Qubit 2.0 Fluorometer (Life Technology Corporation). Library preparation was performed using the TruSeq DNA PCR-Free Library Prep Kit (Illumina) starting at the repair ends and library size selection step with an adjusted beads ratio of 1.8 to remove small fragments. After adapter ligation, libraries were validated on a fragment analyzer (Advanced Analytical Technologies). Final libraries were quantified by qPCR, normalized and pooled before 150 paired-end sequencing on the Illumina Miniseq Sequencing System with a High-Output Kit, yielding up to 25 million reads (Illumina).

2.5.3 | Bioinformatic data treatment

Bioinformatic processing of raw sequences was conducted separately for each library using the *OBITools* package (Boyer et al., 2016). Forward and reverse reads were assembled with a minimum quality score of 40 and assigned to samples based on unique tag and primer combinations, allowing two mismatches with primer, and identical sequences were clustered. All sequences with <10 reads per library were discarded as well as those not corresponding to primer specific barcode lengths, i.e., 10–220 bp for Sper01 and 76–168 bp for Arth02 (Taberlet et al., 2018). This was followed by two different clustering methods. First, pairwise dissimilarities between reads were computed and lesser abundant sequences were clustered into the most abundant ones. Second, we reduced remaining clusters based on a sequence similarity of 97% using the *sumaclust* algorithm (Mercier et al., 2013). For taxonomic assignment of sequences, three different reference databases were used. The *local* database for Sper01 was based on the local plant collection (see 2.3). Furthermore, to construct *global* databases, both primer sets were used to simulate in silico PCRs on GenBank using the *ecoPCR* software (Ficetola et al., 2010) to select all sequences corresponding to our primers (restrained to three mismatches, the targeted barcode lengths and to Metazoa and Viridiplantae, respectively). Sper01 sequences were first assigned to the local database and non-assigned sequences were subsequently run against the global Sper01 database, both with 97% thresholds. In addition, in order to test the effect of the local database, we did the taxonomic assignment of Sper01 sequences with only the global database and assessed the ratio of assigned sequences. Arthropod sequences were directly run against the global Arth02 database with a 97% similarity threshold.

Additional filtering of sequences and subsequent data analyses were performed in R (version 4.0.2; R Core Team, 2022). Sequences that were more abundant in extraction and PCR controls than in samples were considered as contaminants and removed. To account for tag switching, we considered the leaking of a sequence to be directly linked to its abundance. To test this, we performed Wilcoxon signed-rank tests to assess the relationship between samples and blanks and a ratio was defined independently for each library to remove likely leaked sequences, as implemented in the R package *metabar* (Zinger et al., 2021). Replicates per sample were compared and the mean number of reads was retained if a sequence was present in at least two out of three replicates, in line with Ficetola et al. (2015) and a minimum of five reads. All plant species-level assignments were manually verified and re-assigned to genus level if the known geographic species range did not match but the genus was known to occur in South Africa, else to family. For Arth02, we retained only the family level to avoid any taxonomic ambiguities (Meiklejohn et al., 2019) and all sequences assigned to vervets and humans were discarded.

2.6 | Data analyses

Analyses on the sequence data were conducted using RRA if not stated otherwise. In order to treat the observational data similarly, the sum of observations of each consumed item per day was divided by the total number of focal screenings conducted that day. Sample numbers varied between months/seasons and methods, hence for subsequent analyses mean values were taken per temporal unit. Since data were not normally distributed (according to Shapiro–Wilk's tests), we employed non-parametric tests. The impact of seasons on dietary variation was determined by principal coordinates analyses (PCoA) using the ade4 package (Dray & Dufour, 2007). To account for pseudo-replication, the same weight was given here to all individuals, i.e. replicate samples sum up to 1 per specific individual, while observational data were aggregated per focal individual/season and transformed to relative abundances. We identified plant indicators for seasons using *Indicator value analyses* (Indval; Dufrêne & Legendre, 1997). Shannon–Wiener diversity indices were calculated per season (genera/species for plants, family level for arthropods) and Hutcheson *t* tests performed to test for significant differences between seasons (Hutcheson, 1970). We performed Mantel's tests (Mantel, 1967) implemented in the *vegan* package with 9999 permutations to compare the correlation between datasets with data aggregated per month and transformed to Bray–Curtis dissimilarity matrices. Spearman rank correlations were calculated for all plant species present in both datasets and with a minimum count of 350 in the focal dataset (with the exception of *V. nilotica*/*C. decapetala* and *E. crispa*/*E. undulata*/*D. dichrophylla* since sequence data matched two different species in the focal dataset).

3 | RESULTS

The final dataset for Sper01 contained 5,275,361 reads assigned to 22 orders, 43 families, 61 genera, and 35 species. Of these 4,599,838 reads were assigned to 31 items with the local database, including 25

identifications at species level. Most of the plant genera and species consumed by this species are not only trees and shrubs but also cactuses, herbs, and grasses (Appendix S1: Table S3). Taxonomic assignment with solely the global database resulted in 330,612 reads assigned to 15 different species; however, only 10 species were reliable (Appendix S1: Figure S3). The taxonomic resolution was hence greatly increased with the local database allowing for more detailed analyses.

During focal follows, vervet monkeys were observed feeding on 27 different plant species and two plant genera. Mean observations per month of the eight most frequent plant species in the focal dataset showed similar temporal patterns as the DNA metabarcoding data (Figure 4a) and a Mantel's test of Bray–Curtis dissimilarity matrices of data aggregated per month indicated a high correlation between methods ($r = .62$, $p = 1e\text{-}04$). There was no positive correlation between methods for numbers of different diet items detected/observed per month (Appendix S1: Figure S4). However, positive Spearman rank correlations were observed when comparing single plant species, among which the most consumed ones (Appendix S1: Figure S5). In addition to the plant genera and species that were identified by both methods, DNA metabarcoding revealed 41 supplementary dietary items at this taxonomic level of which 21 at species level (Figures 3a and 5, Appendix S1: Table S3). The Shannon diversity did not differ significantly between both methods for plant genera and species observations/detections (Hutcheson t-tests not significant) despite the variable total numbers, i.e. richness (Figure 4c). Seasonal shifts were most pronounced between the wet and the dry season for *B. zeyheri* and *Z. mucronata* indicating that one substitutes the other as principal food resource (Figure 4a). Season explained a lot of the variation in both datasets as illustrated by PCoAs (Figure 6a,b) and confirmed by *ANOSIM* with $R = .51$ and $R = .57$, both $p = 1e\text{-}04$, for eDNA and observational data, respectively. Figure S6 shows observations and RRA over 12 months for seven plant species that were season indicators based on observational data. All except one, *C. jamacaru*, were indicator species in the metabarcoding dataset as well. The latter revealed several additional season indicator species (Appendix S1: Table S3).

Over 12 months of observational focal sampling, there were in total 1359 foraging events for arthropods (1142 undetermined insects, 191 termites, 24 grasshoppers; Figure 3b). We investigated in particular the temporal dimension of the "termites" category since vervets feed on termites extensively during swarming periods, which can be easily observed. Figure 4b shows percentages of the occurrences of these categories together with the combined RRA data for the families Hodotermitidae and Termitidae ("termites") as well as all taxa of the order Orthoptera ("grasshoppers"), and all other sequences combined ("others"). While a consistent trend was observed between methods, observations and DNA sequence data are not significantly correlated (Appendix S1: Figure S7).

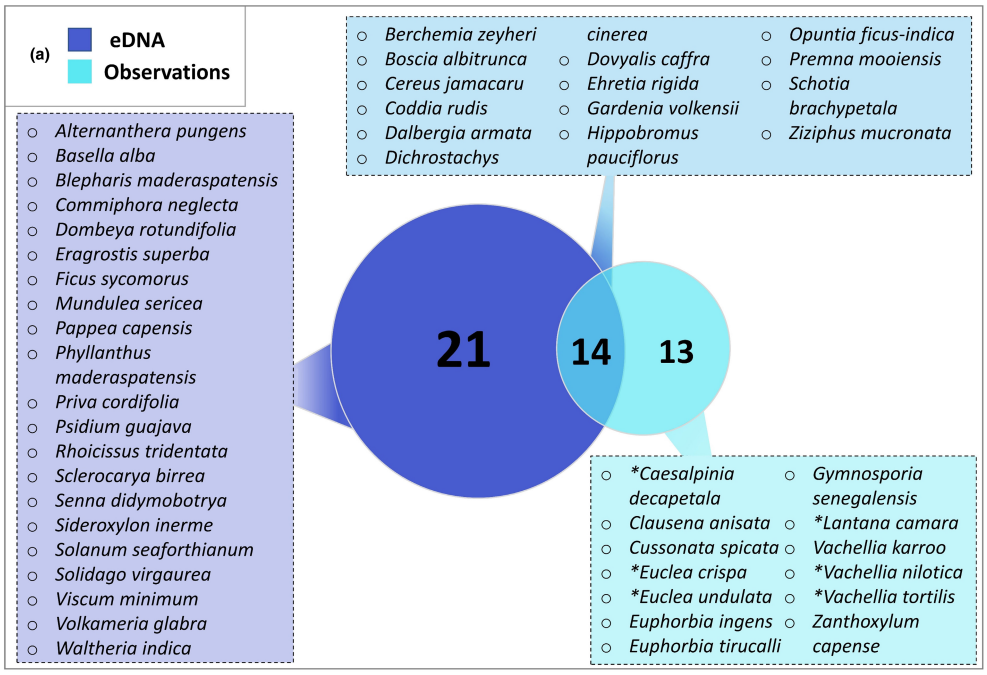
Without relying on a reference database for taxonomic assignments, the Arth02 assay resulted in 1,698,439 sequences in total whereof, however, 961,542 belonged to vervets, leaving 736,897 reads clustered to 404 presumed arthropod operational taxonomic units (OTUs) (Appendix S1: Figure S3). By relying on the global database, the number of DNA sequences after final data filtering was 360,040 assigned to 11 orders and 35 families (Appendix S1: Table S4), i.e. 48.86% of reads were taxonomically assigned (not considering those of *C. pygerythrus*). The most abundant arthropod orders in terms of read counts and frequencies were Coleoptera, Blattodea, and Lepidoptera. We detected arthropod sequences in 96% of the samples in spring, 89.15% in summer, 58.59% in autumn, and 82.72% in winter, whereas the highest number of different orders and families was detected in summer, also showing the highest Shannon diversity (Figure 4c). While we observed monthly variation for certain taxa (Figure 7), there was overall a significant yet weak seasonal effect (Appendix S1: Figure S8).

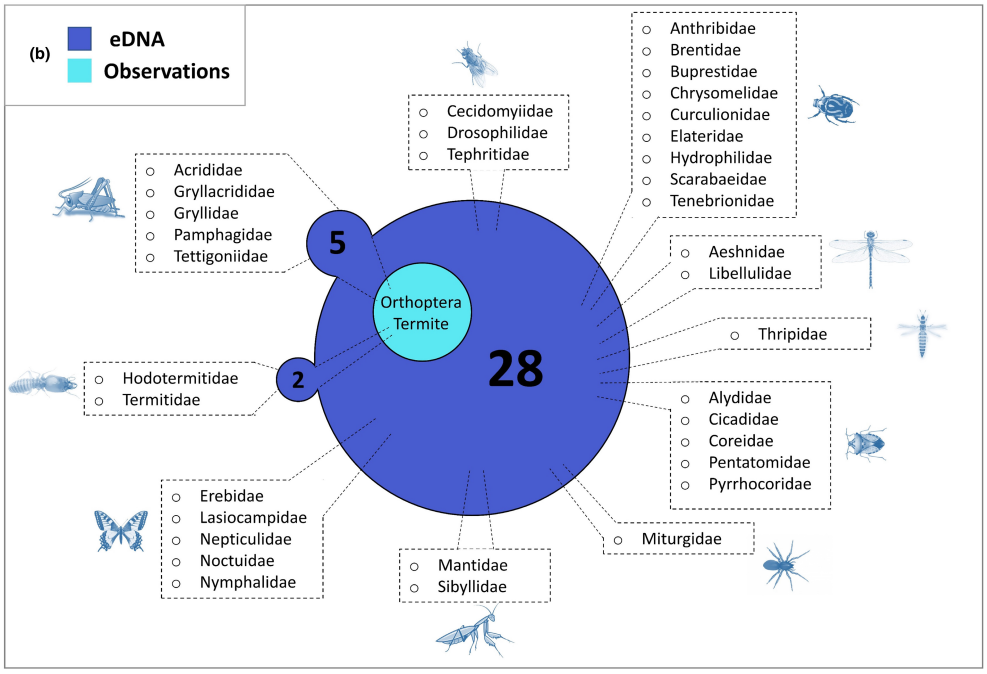
4 | DISCUSSION

The present study of vervet monkeys' diet over a 12-month period highlights strong seasonal variation in consumed plants and less pronounced variation in arthropod consumption across seasons. The comparison of DNA metabarcoding data of plant diet components with field observational data shows similar patterns, in particular regarding relative abundances and seasonal variation. However, whilst observations captured well the main plant diet components, DNA sequencing data showed improved taxonomic coverage and resolution. With respect to arthropod consumption, DNA metabarcoding outperformed observations, allowing for a considerable expansion of the range of dietary items identified and demonstrating the clear advantages of this method to describe cryptic feeding behavior. Both methods have certain advantages and shortcomings as further discussed below, and genetic data are increasingly merged for network analyses with data from different sources to be used in a complementary way. For example, observational data provide in many cases more information regarding state and life stage of consumed items. While this may lead to more complete datasets, it implies also specific challenges as discussed by Cuff et al. (2022).

For plants items, our DNA metabarcoding assay detected many additional species and genera that had not been observed

FIGURE 3 Venn diagrams. (a) Between consumed plant items at the taxonomic level of species detected by observation and eDNA. Plant species beginning with an asterisk (*) correspond to species for which the sequences amplified by the Sper01 metabarcode were identical between species as shown in Table S1. (b) Between arthropods detected by observation and eDNA. For eDNA data, the family level is included, whereas observations were limited to the order level for orthopterans and the infraorder level for termites. The two bubbles on the left side of the diagram indicate the families detected by eDNA that compose these two taxonomic groups. The category "undetermined insects" is not included for observations (see text). Rectangles separate the different orders illustrated by icons.


(a)
eDNA
Observations
Alternanthera pungens
Basella alba
Blepharis maderaspatensis
Commiphora neglecta
Dombeya rotundifolia
Eragrostis superba
Ficus sycomorus
Mundulea sericea
Pappea capensis
Phyllanthus maderaspatensis
Priva cordifolia
Psidium guajava
Rhoicissus tridentata
Sclerocarya birrea
Senna didymobotrya
Sideroxylon inerme
Solanum seaforthianum
Solidago virgaurea
Viscum minimum
Volkameria glabra
Waltheria indica
21
14
13
Berchemia zeyheri
Boscia albitrunca
Cereus jamacaru
Coddia rudis
Dalbergia armata
Dichrostachys cinerea
Dovyalis caffra
Ehretia rigida
Gardenia volkensii
Hippobromus pauciflorus
Opuntia ficus-indica
Premna mooiensis
Schotia brachypetala
Ziziphus mucronata
*Caesalpinia decapetala
Clausena anisata
Cussonata spicata
*Euclea crispa
*Euclea undulata
Euphorbia ingens
Euphorbia tirucalli
Gymnosporia senegalensis
*Lantana camara
Vachellia karroo
*Vachellia nilotica
*Vachellia tortilis
Zanthoxylum capense


(b)
eDNA
Observations
Orthoptera
Termite
5
2
28
Acrididae
Gryllacrididae
Gryllidae
Pamphagidae
Tettigoniidae
Hodotermitidae
Termitidae
Cecidomyiidae
Drosophilidae
Tephritidae
Anthribidae
Brentidae
Buprestidae
Chrysomelidae
Curculionidae
Elateridae
Hydrophilidae
Scarabaeidae
Tenebrionidae
Aeshnidae
Libellulidae
Thripidae
Alydidae
Cicadidae
Coreidae
Pentatomidae
Pyrrhocoridae
Miturgidae
Mantidae
Sibyllidae
Erebidae
Lasiocampidae
Nepticulidae
Noctuidae
Nymphalidae

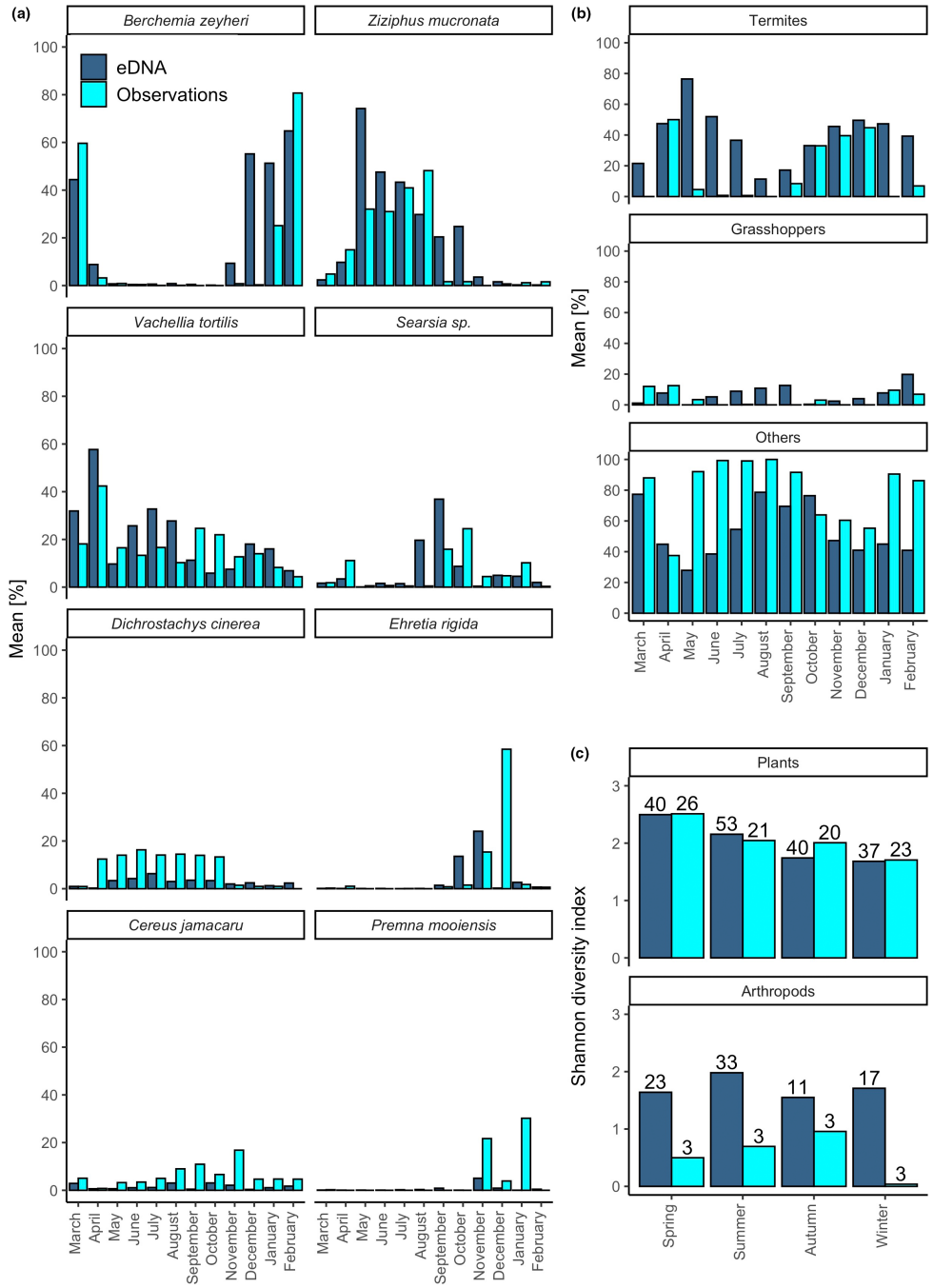
or identified to this level, as well as most of the species observed during focal follows. The taxonomic resolution was excellent for the plant assay due to the use of the local database (see also Quéméré et al., 2013). The increased detection by metabarcoding is likely due to observational difficulties in recording certain food items that are hard to identify or to observe, e.g. taking place in inaccessible or dense terrain (Matthews et al., 2020; Su & Lee, 2001). In our study, DNA metabarcoding further revealed consumption of otherwise well-documented species in periods when they were missed during observations, likely due to the consumption of less visible parts, e.g. tree sap, or dried seeds or fruits collected from the ground.

All new information made available by DNA metabarcoding could imply important trophic relations that have been overlooked so far. This is particularly relevant for arthropod items, a food type rich in proteins and lipids (Rothman et al., 2014), for which feeding habits are poorly studied in primatology. Previous observational studies indicate feeding of vervets on arthropods with varying degrees of precision (Barrett, 2005; Struhsaker, 1967; Tournier et al., 2014) but detailed records have so far been missing. Here, with DNA metabarcoding, 35 different families representing 11 orders were identified and demonstrate increased diversity of arthropod consumption in vervets' diets compared to the three broad taxonomic categories grouping termites, orthopterans, and others as identified with observations (Figures 3b and 7 and Appendix S1: Table S4). For arthropods, dietary diversity and richness are hence markedly higher when relying on DNA metabarcoding (Figure 4c). Accordingly, we found no correlation between observational and genetic data (Appendix S1: Figure S7), indicating the aptitude of the latter to unmask new trophic interactions and to shed light on cryptic feeding behavior. A good example illustrated by our dataset is that of the twice-yearly termite swarming, a major ecological event in South-Africa (Lesnik, 2014), which was adequately captured by both our methods (Figure 4b). Although showing a similar trend, the observations and DNA sequence data are not significantly correlated. During swarming, the large number of flying termites emerging from the nest makes them highly visible to observers. However, during the rest of the year, when monkeys forage directly on the ground or in dead wood and in lesser quantities, most of these foraging events are cryptic or difficult to identify and thus missed by observers but documented by genetics. In general, observation of feeding on arthropods is particularly challenging (Pickett et al., 2012) and this is the likely cause of the minimal detail available from our observational data and previous observational studies on vervets. A comparison between observations and DNA metabarcoding yielded similar results for white-faced capuchins (*Cebus capucinus*), with eight arthropod orders observed against 29 orders detected (Mallott et al., 2017). Furthermore, recent genetic studies on other primate species have similarly contributed to a better representation of arthropod diet components, either by using a cloning approach (Pickett et al., 2012), DNA metabarcoding (Lyke et al., 2018; Mallott et al., 2017, 2015; Rowe et al., 2021), or metagenomic sequencing (Srivathsan et al., 2015). This study demonstrates the advantages of using DNA metabarcoding alongside observations, adding to previous findings for the part of plants and arthropods of the diet of wild vervets.

In line with previous work showing that movements of vervets were mostly driven by plant resource availability, and therefore seasonality (Barrett, 2009), we found significant variation in plant consumption, largely shaped by the dry and wet seasons (Figure 6). For the plant genera and plant species that have been recorded with both methods, we found comparable abundances, similar seasonal patterns, and season indicator species (Appendix S1: Figure S6, Table S3). Our inter-method comparison illustrates for certain plant species very clear temporal correlations (Figure 4a, Appendix S1: Figure S5). Regarding plants, both methods indicated similar Shannon indices per season but the genetic approach resulted in higher dietary richness (Figure 4c). While some plants are consumed continuously (different parts may be eaten over the year), the consumption of others was associated with particular seasons (e.g. strong association of *Z. mucronata* with winter). Previous studies on vervets found that they spend more time foraging in the dry season because of resource scarcity (Arseneau-Robar et al., 2017; Canteloup et al., 2019). They can hence be expected to be more opportunistic feeders in the dry season than when food is abundant in the wet season and the opportunity to engage in selective foraging behaviors arises. During wet, food-abundant summer, we detected a higher diversity of consumed items in the scat samples. This shows that vervets adapt their diet according to available resources.

Concerning arthropod consumption, although the statistical effect of season on arthropod consumption was weak, the highest percentage of samples containing arthropod sequences was found in spring and summer, as well as the highest (family) richness and Shannon diversity (Figure 4c). Given the very different numbers

FIGURE 4 (a) Monthly comparison of DNA metabarcoding and observational data for the most frequent species in the focal dataset (>350 observations), with the exception of those that had identical metabarcodes and matched several species in the focal dataset. The observed plant *V. tortilis* corresponds to *V. tortilis/sieberiana* in the DNA metabarcoding dataset. Metabarcoding data are represented by the mean RRA and observational data by the mean count, both in percent. (b) Monthly comparison of DNA metabarcoding and observational data for "termites" (RRA of Hodotermitidae and Termitidae combined), "grasshoppers" (RRA of all detected families belonging to the order Orthoptera), and "others" (RRA of all remaining items). Metabarcoding data are represented by the mean RRA and observational data by the mean count, both in percent. (c) Shannon diversity index per season for observations and eDNA. There was no significant difference in diversity between methods (Hutcheson *t* test). Numbers on the bars indicate numbers of different observed/detected items per season. For plants, the included items are all observed/detected species and genera. For arthropods, the Shannon diversity was measured at family level for the metabarcoding data and for observational data based on the three categories (b).


(a)
Berchemia zeyheri
Ziziphus mucronata
eDNA
Observations
Vachellia tortilis
Searsia sp.
Dichrostachys cinerea
Ehretia rigida
Cereus jamacaru
Premna mooiensis
Mean [%]
March
April
May
June
July
August
September
October
November
December
January
February
(b)
Termites
Grasshoppers
Others
Mean [%]
(c)
Plants
40 26
53 21
40 20
37 23
Arthropods
23 3
33 3
11 3
17 3
Shannon diversity index
Spring
Summer
Autumn
Winter

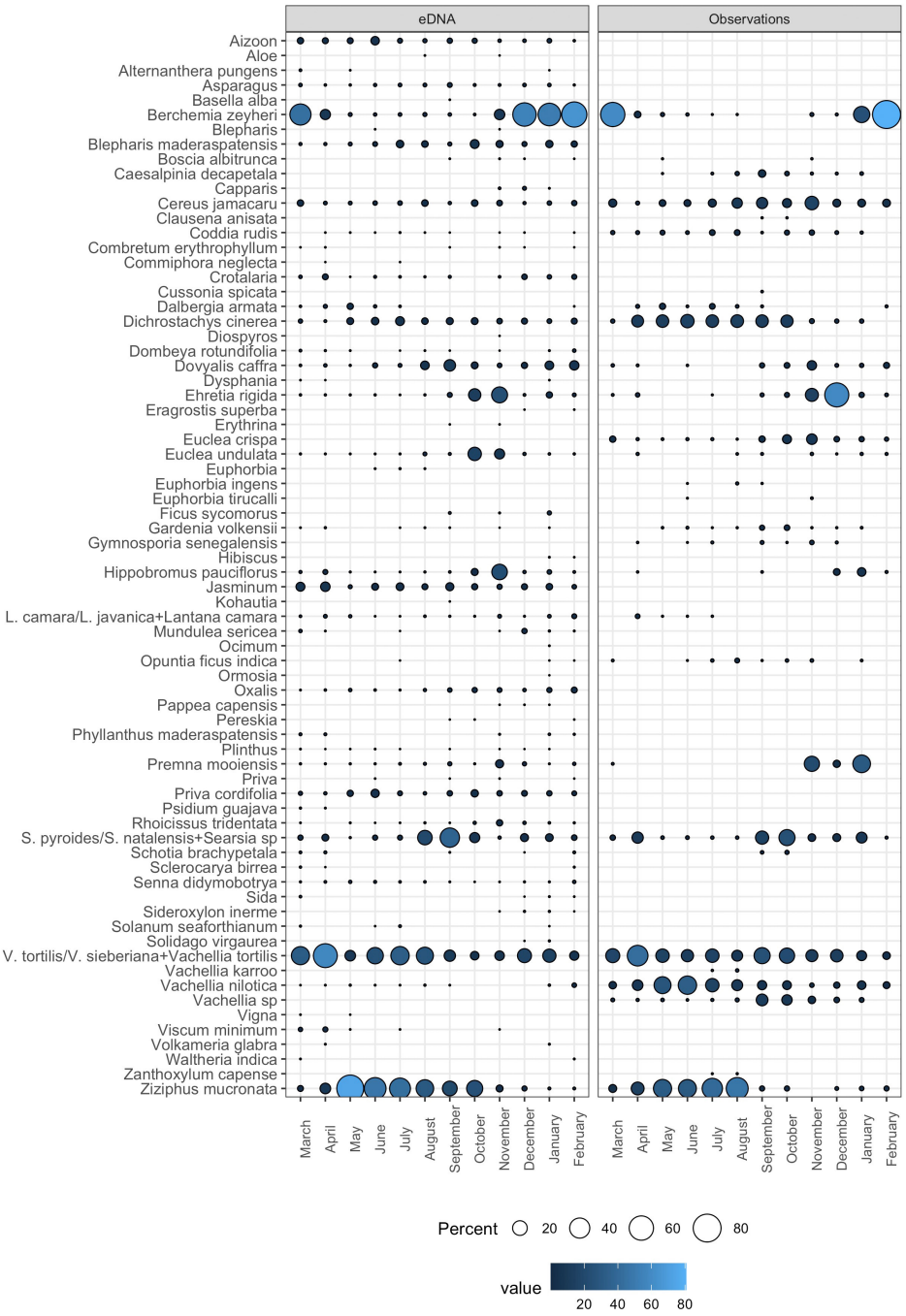

eDNA
Observations
Aizoon
Aloe
Alternanthera pungens
Asparagus
Basella alba
Berchemia zeyheri
Blepharis
Blepharis maderaspatensis
Boscia albitrunca
Caesalpinia decapetala
Capparis
Cereus jamacaru
Clausena anisata
Coddia rudis
Combretum erythrophyllum
Commiphora neglecta
Crotalaria
Cussonia spicata
Dalbergia armata
Dichrostachys cinerea
Diospyros
Dombeya rotundifolia
Dovyalis caffra
Dysphania
Ehretia rigida
Eragrostis superba
Erythrina
Euclea crispa
Euclea undulata
Euphorbia
Euphorbia ingens
Euphorbia tirucalli
Ficus sycomorus
Gardenia volkensii
Gymnosporia senegalensis
Hibiscus
Hippobromus pauciflorus
Jasminum
Kohautia
L. camara/L. javanica+Lantana camara
Mundulea sericea
Ocimum
Opuntia ficus indica
Ormosia
Oxalis
Pappea capensis
Pereskia
Phyllanthus maderaspatensis
Plinthus
Premna mooiensis
Priva
Priva cordifolia
Psidium guajava
Rhoicissus tridentata
S. pyroides/S. natalensis+Searsia sp
Schotia brachypetala
Sclerocarya birrea
Senna didymobotrya
Sida
Sideroxylon inerme
Solanum seaforthianum
Solidago virgaurea
V. tortilis/V. sieberiana+Vachellia tortilis
Vachellia karroo
Vachellia nilotica
Vachellia sp
Vigna
Viscum minimum
Volkameria glabra
Waltheria indica
Zanthoxylum capense
Ziziphus mucronata
March
April
May
June
July
August
September
October
November
December
January
February
Percent
20
40
60
80
value
20
40
60
80

FIGURE 5 Mean RRA of plants genera and species in fecal samples per month (left) and mean of observations in focal follows per month (right). Note that the obtained sequence for Euphorbia is different from *E. ingens* and *E. tirucalli*. Also, *E. crispa* and *E. undulata* were identified to species level in the field but have identical sequences, the same is true for *V. nilotica* and *C. decapetala*; therefore, both entries for observations were kept but only one for eDNA. Several names in one line indicate identical sequences as well (on the left), but only one observed genus/species (on the right).

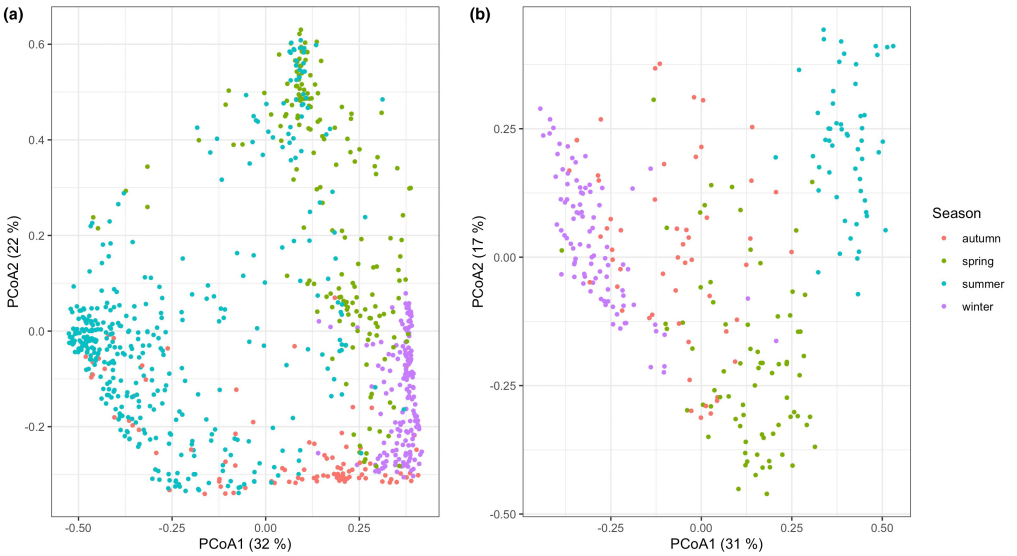


FIGURE 6 Principal coordinates analyses (PCoA) based on (a) relative read abundances (RRA) of consumed plants detected in fecal samples (n = 823) and (b) observational plant data of focal follow transformed to relative abundances per individual/season (n = 279). In brackets the relative Eigenvalues in percent.

of arthropod items detected per method, the comparable diversity might surprise but can be explained by the dominance of few abundant families/categories; this may be different in other study contexts. Overall, our results show that season is an important variable for diet choice; therefore, sampling designs should take it into account when this is relevant for the research question. Here, selective behaviors are most likely in the wet season when differences are the most accentuated and resources are not limiting, hence future sampling could focus on that season to capture most efficiently any behavioral differences that are not driven by resource availability, as discussed below.

DNA metabarcoding approaches do nonetheless entail their own limitations, some are marker specific and some are methodological. Primer-induced biases may have led to under- or non-representation of certain arthropod taxa in this study. The study of omnivorous species is often neglected and thus highly necessary but requires in most cases the combination of different primer sets, which increases study cost and introduces new challenges (Tercel et al., 2021). Plants and arthropods were considered the most important targets based on observational data; however, our marker choice excluded the detection of other dietary items (i.e. feeding on birds, eggs, and mushrooms was occasionally observed). Some plant species shared identical sequences in the metabarcode we amplified, making it impossible to differentiate genetically between them (Taberlet et al., 2007). For plants observed only in small numbers and not detected (false negatives), this may be due to stochastic reasons and the fact that observations and scat samplings were not conducted at the same time. For the observed but not detected *V. karroo* and *Z. capense* there is no sequence available in our databases. While this can be overcome by including further sequences, it points to the issue of incomplete databases in metabarcoding studies (Furlan et al., 2020; Taberlet et al., 2012). A local database would certainly increase the taxonomic coverage and resolution as well for the Arth02 assay and would have allowed the attribution of some abundantly represented OTUs, in particular since our research is pursued in a geographic region underrepresented in genetic databases (Kvist, 2013; Marques et al., 2021). In addition, unlike observational data, genetic data cannot detail which part and state of the plant or which life stage of an arthropod has been consumed (Pompanon et al., 2012; Rees et al., 2014). Parts of the sequences may be due to secondary ingestion, accidental consumption, or of parasitic origin and not represent (intentionally) consumed items

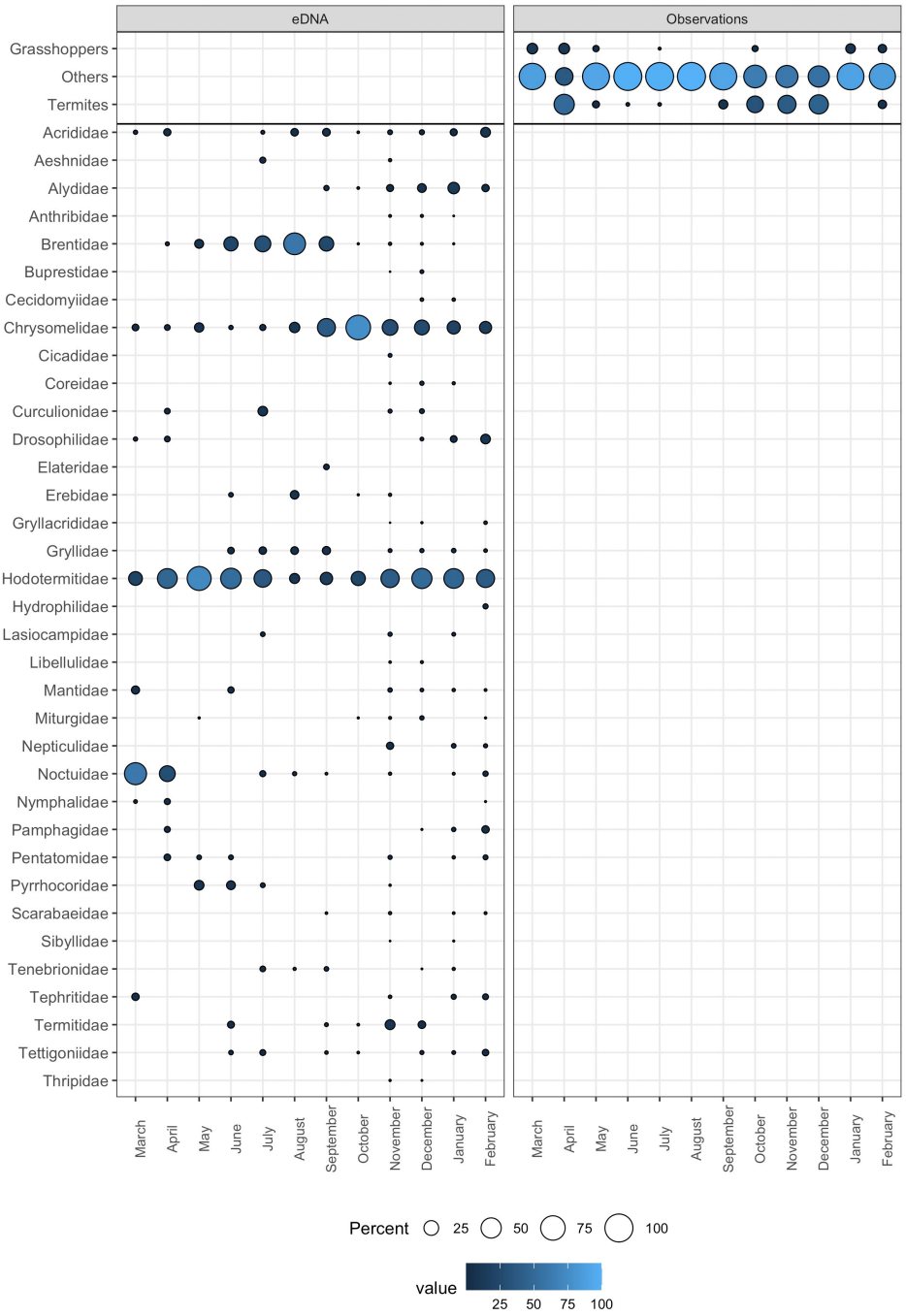

eDNA
Observations
Grasshoppers
Others
Termites
Acrididae
Aeshnidae
Alydidae
Anthribidae
Brentidae
Buprestidae
Cecidomyiidae
Chrysomelidae
Cicadidae
Coreidae
Curculionidae
Drosophilidae
Elateridae
Erebidae
Gryllacrididae
Gryllidae
Hodotermitidae
Hydrophilidae
Lasiocampidae
Libellulidae
Mantidae
Miturgidae
Nepticulidae
Noctuidae
Nymphalidae
Pamphagidae
Pentatomidae
Pyrrhocoridae
Scarabaeidae
Sibyllidae
Tenebrionidae
Tephritidae
Termitidae
Tettigoniidae
Thripidae
March
April
May
June
July
August
September
October
November
December
January
February
Percent
25
50
75
100
value
25
50
75
100

FIGURE 7 Mean RRA of arthropod families in fecal samples per month (left) and mean of observations in focal follows per month (right). The category "others" includes all insect observations that were neither identified as grasshoppers nor as termites. The families in the order Orthoptera ("grasshoppers") are Acrididae, Gryllacrididae, Gryllidae, Pamphagidae, and Tettigoniidae. The families in the order Blattodea (equivalent to "termites") are Hodotermitidae and Termitidae.

(Tercel et al., 2021); therefore, interdisciplinary studies with parasitology may be fruitful. Arthropods may have ingested plant DNA that we thus falsely detected as part of vervet diet, and at the other end of the spectrum unintentional feeding of arthropods is possible, e.g. of small Thripidae. The feeding on termites and grasshoppers is confirmed by observations, and also active foraging (i.e. vervets searching for insects), showing once more the benefit of complementary use of methods.

Choices made during the processing of DNA metabarcoding data may influence the outcome of these studies (Calderón-Sanou et al., 2019). In this study, we applied a stringent filtering of the data to avoid spurious DNA, using percentual and absolute thresholds. It has been argued that arbitrary minimum copy thresholds might omit true sequences (Littleford-Colquhoun et al., 2022) and that percentual thresholds were more suitable in case of uneven sequencing depths (Drake et al., 2022). To avoid the generation of supplementary biases, it is recommended to normalize PCR amplicons before pooling. Here we accepted the risk of missing some true detections by omitting items with very small read counts, which may also affect samples with uneven sequencing depths differently. Another point is the transformation of read counts; while most studies traditionally rely on occurrence data, others argue that RRA data might better capture ecological signals (Deagle et al., 2019; Kartzinel et al., 2015; Voelker et al., 2020). Here, we chose RRA and although it may entail biases, the comparison to observational data validates this choice. For example, two of the most consumed plants throughout the year, *B. zeyheri* and *Z. mucronata*, represent very variable proportions of the diet depending on the season. Categorical data would not show any variation here; however, we observed strong seasonal patterns with both RRA and observational data (Figure S6). A recent diet study targeting the same genetic region found positive correlations between the RRA of plant families in fecal samples and the observed duration spent feeding on those (Mallott et al., 2018).

The taxonomic coverage and resolution as well as the methodological standardization (including no inter-observer variability) point to the benefits of environmental DNA (eDNA)-based surveys. Depending on the species studied, DNA metabarcoding represents cost- and labor-effective alternatives or complements to traditional methods (Mena et al., 2021) and sequencing costs are likely to further decrease in the near future. The sensitivity, taxonomic resolution, and non-invasiveness of the method are major advantages in conservation research (Thomsen & Willerslev, 2015). There is great potential to learn more about, for example, nocturnal, arboreal, and other elusive species and/or the adaptive potential of fragmented populations (Quéméré et al., 2013). Many primates are threatened and of high conservation concern (IUCN, 2020; Schwitzer et al., 2017). There is thus a need for robust data to inform empirically based conservation strategies (Pimm et al., 2014), where diet studies are undoubtedly of primary interest. Although it remains challenging to properly assess to what extent the final data represent the biomass of food items initially ingested, controls incorporated throughout the study and appropriate knowledge of the ecology enable valuable insights going beyond traditional approaches. DNA metabarcoding has thus great potential to bring new insights on foraging behaviors and ultimately, on the underlying mechanisms shaping such behaviors.

Our study demonstrates benefits of an interdisciplinary approach. Moreover, this study being the first validating the use of eDNA to assess diet in our system, future analyses may investigate whether variation in individual or group diet is induced by environmental differences or if it might reflect selective foraging behaviors. Therefore, the application of a DNA metabarcoding approach can be useful not only for conservation studies aimed at disentangling complex diets or reveal trophic interactions but also opens new perspectives for behavioral ecologists and cultural evolutionists studying social species in the wild.

AUTHOR CONTRIBUTIONS

Loic Brun: Formal analysis (equal); investigation (equal); writing – original draft (equal). **Judith Schneider:** Formal analysis (equal); investigation (equal); writing – original draft (equal). **Eduard Mas-Carrió:** Investigation (equal); writing – review and editing (equal). **Pooja Dongre:** Investigation (equal); writing – review and editing (equal). **Pierre Taberlet:** Investigation (equal); supervision (supporting); writing – review and editing (equal). **Erica van de Waal:** Conceptualization (equal); data curation (equal); funding acquisition (equal); project administration (equal); resources (equal); supervision (equal); validation (equal); writing – review and editing (equal). **Luca Fumagalli:** Conceptualization (equal); data curation (equal); funding acquisition (equal); project administration (equal); resources (equal); supervision (equal); validation (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

We would like to thank the van der Walt family for permission to conduct the study on their land. We are very grateful to the whole IVP team who assisted in the field work and especially the on-site manager Arend van Blerk who organized the permits and shipment of the samples. The observational research was approved by the relevant local authority "Ezemvelo KZN Wildlife" in South Africa. The field costs during data and sample collection were funded by grants to EW from the Branco-Weiss Fellowship-Society in Science, the 'ProFemmes' grant of the Faculty of Biology and Medicine of the University of Lausanne and the Swiss National Science Foundation (grant number PP00P3_170624). At the time of writing, EW was supported by the European Research Council under the European

Union's Horizon 2020 research and innovation programme for the ERC 'KNOWLEDGE MOVES' starting grant (grant agreement No. 949379). This project was supported by a Swiss National Science Foundation grant to LF (nr. 310030_192512). EM-C was supported by a fellowship in Life Sciences (Faculty of Biology and Medicine, University of Lausanne).

CONFLICT OF INTEREST

The authors note that PT is co-inventor of a patent related to the Sper01 primers and the use of the P6 loop of the chloroplast trn*L* (UAA) intron for plant identification using degraded template DNA. This patent only restricts commercial applications and has no impact on the use of this locus by academic researchers.

DATA AVAILABILITY STATEMENT

The DNA metabarcoding data generated for this study are available on DRYAD (10.5061/dryad.6q573n621). Sanger sequences for the local database have been deposited in GenBank under accession numbers OL898555-OL898608.

ORCID

Loïc Brun https://orcid.org/0000-0003-3009-9091
Judith Schneider https://orcid.org/0000-0003-2663-6130
Eduard Mas Carrió https://orcid.org/0000-0002-5875-0424
Pierre Taberlet https://orcid.org/0000-0002-3554-5954
van de Waal https://orcid.org/0000-0001-7778-418X
Luca Fumagalli https://orcid.org/0000-0002-6648-2570

REFERENCES

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, *49*, 227–267. https://doi.org/10.1163/156853974X00534

Arrizabalaga-Escudero, A., Clare, E. L., Salsamendi, E., Alberdi, A., Garin, I., Aihartza, J., & Goiti, U. (2018). Assessing niche partitioning of co-occurring sibling bat species by DNA metabarcoding. *Molecular Ecology*, *27*, 1273–1283. https://doi.org/10.1111/mec.14508

Arseneau-Robar, T. J. M., Taucher, A. L., Schnider, A. B., van Schaik, C. P., & Willems, E. P. (2017). Intra- and interindividual differences in the costs and benefits of intergroup aggression in female vervet monkeys. *Animal Behaviour*, *123*, 129–137. https://doi.org/10.1016/j.anbehav.2016.10.034

Barrett, A. (2005). *Foraging ecology of the vervet monkey (Chlorocebus aethiops) in mixed lowveld bushveld and sour lowveld bushveld of the blydeberg conservancy, Northern Province, South Africa*. University of South Africa.

Barrett, A. (2009). *Spatial and temporal patterns in resource dispersion and the structure of range use and co-existence in a social omnivore Chlorocebus aethiops*. Ph.D Dissertation. University of South Africa.

Bohmann, K., Gopalakrishnan, S., Nielsen, M., Nielsen, L. D. S. B., Jones, G., Streicker, D. G., & Gilbert, M. T. P. (2018). Using DNA metabarcoding for simultaneous inference of common vampire bat diet and population structure. *Molecular Ecology Resources*, *18*, 1050–1063. https://doi.org/10.1111/1755-0998.12891

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*, 176–182. https://doi.org/10.1111/1755-0998.12428

Burgar, J. M., Murray, D. C., Craig, M. D., Haile, J., Houston, J., Stokes, V., & Bunce, M. (2014). Who's for dinner? High-throughput sequencing reveals bat dietary differentiation in a biodiversity hotspot where prey taxonomy is largely undescribed. *Molecular Ecology*, *23*, 3605–3617. https://doi.org/10.1111/mec.12531

Calderón-Sanou, I., Münkemüller, T., Boyer, F., Zinger, L., & Thuiller, W. (2019). From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices? *Journal of Biogeography*, *47*, 193–206. https://doi.org/10.1111/jbi.13681

Canteloup, C., Borgeaud, C., Wubs, M., & van de Waal, E. (2019). The effect of social and ecological factors on the time budget of wild vervet monkeys. *Ethology*, *125*, 902–913. https://doi.org/10.1111/eth.12946

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

Cuff, J. P., Windsor, F. M., Tercel, M. P. T. G., Kitson, J. J. N., & Evans, D. M. (2022). Overcoming the pitfalls of merging dietary metabarcoding into ecological networks. *Methods in Ecology and Evolution*, *13*, 545–559. https://doi.org/10.1111/2041-210X.13796

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*, 306–323. https://doi.org/10.1111/1755-0998.12188

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*, 391–406. https://doi.org/10.1111/mec.14734

Drake, L. E., Cuff, J. P., Young, R. E., Marchbank, A., Chadwick, E. A., & Symondson, W. O. C. (2022). An assessment of minimum sequence copy thresholds for identifying and reducing the prevalence of artefacts in dietary metabarcoding data. *Methods in Ecology and Evolution*, *13*, 694–710. https://doi.org/10.1111/2041-210X.13780

Dray, S., & Dufour, A. B. (2007). The ade4 Package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*, 1–20. https://doi.org/10.18637/jss.v022.i04

Duffy, J. E., Cardinale, B. J., France, K. E., McIntyre, P. B., Thébault, E., & Loreau, M. (2007). The functional role of biodiversity in ecosystems: Incorporating trophic complexity. *Ecology Letters*, *10*, 522–538. https://doi.org/10.1111/j.1461-0248.2007.01037.x

Dufrêne, M., & Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs*, *67*, 345–366. https://doi.org/10.2307/2963459

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., & Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. *BMC Genomics*, *11*, 434. https://doi.org/10.1186/1471-2164-11-434

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

Furlan, E. M., Davis, J., & Duncan, R. P. (2020). Identifying error and accurately interpreting environmental DNA metabarcoding results: A case study to detect vertebrates at arid zone waterholes. *Molecular Ecology Resources*, *20*, 1259–1276. https://doi.org/10.1111/1755-0998.13170

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

Gillet, F., Tiouchichine, M.-L., Galan, M., Blanc, F., Némoz, M., Aulagnier, S., & Michaux, J. R. (2015). A new method to identify

the endangered *Pyrenean desman* (*Galemys pyrenaicus*) and to study its diet, using next generation sequencing from faeces. *Mammalian Biology*, *80*, 505–509. https://doi.org/10.1016/j.mambio.2015.08.002

Hutcheson, K. (1970). A test for comparing diversities based on the shannon formula. *Journal of Theoretical Biology*, *29*, 151–154. https://doi.org/10.1016/0022-5193(70)90124-4

IUCN. (2020). The IUCN red list of threatened species, version 2020-1.

Jusino, M. A., Banik, M. T., Palmer, J. M., Wray, A. K., Xiao, L., Pelton, E., Barber, J. R., Kawahara, A. Y., Gratton, C., Peery, M. Z., & Lindner, D. L. (2019). An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Molecular Ecology Resources*, *19*, 176–190. https://doi.org/10.1111/1755-0998.12951

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

Kvist, S. (2013). Barcoding in the dark?: A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. *Molecular Phylogenetics and Evolution*, *69*, 39–45. https://doi.org/10.1016/j.ympev.2013.05.012

Lesnik, J. J. (2014). Termites in the hominin diet: A meta-analysis of termite genera, species and castes as a dietary supplement for South African robust australopithecines. *Journal of Human Evolution*, *71*, 94–104. https://doi.org/10.1016/j.jhevol.2013.07.015

Littleford-Colquhoun, B. L., Freeman, P. T., Sackett, V. I., Tulloss, C. V., McGarvey, L. M., Geremia, C., & Kartzinel, T. R. (2022). The precautionary principle and dietary DNA metabarcoding: Commonly used abundance thresholds change ecological interpretation. *Molecular Ecology*, *31*, 1615–1626. https://doi.org/10.1111/mec.16352

Lyke, M. M., Di Fiore, A., Fierer, N., Madden, A. A., & Lambert, J. E. (2018). Metagenomic analyses reveal previously unrecognized variation in the diets of sympatric Old World monkey species. *PLoS One*, *14*, e0218245. https://doi.org/10.1371/journal.pone.0218245

Mallott, E. K., Garber, P. A., & Malhi, R. S. (2017). Integrating feeding behavior, ecological data, and DNA barcoding to identify developmental differences in invertebrate foraging strategies in wild white-faced capuchins (*Cebus capucinus*). *American Journal of Physical Anthropology*, *162*, 241–254. https://doi.org/10.1002/ajpa.23113

Mallott, E. K., Garber, P. A., & Malhi, R. S. (2018). trnL outperforms rbcL as a DNA metabarcoding marker when compared with the observed plant component of the diet of wild white-faced capuchins (*Cebus capucinus*, Primates). *PLoS One*, *13*, e0199556. https://doi.org/10.1371/journal.pone.0199556

Mallott, E. K., Malhi, R. S., & Garber, P. A. (2015). Brief communication: High-throughput sequencing of fecal DNA to identify insects consumed by wild weddell's saddleback tamarins (*Saguinus weddelli*, Cebidae, Primates) in Bolivia. *American Journal of Physical Anthropology*, *156*, 474–481. https://doi.org/10.1002/ajpa.22654

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220.

Marques, V., Milhau, T., Albouy, C., Dejean, T., Manel, S., Mouillot, D., & Juhel, J.-B. (2021). GAPeDNA: Assessing and mapping global species gaps in genetic databases for eDNA metabarcoding. *Diversity and Distributions*, *27*, 1880–1892. https://doi.org/10.1111/ddi.13142

Matthews, J. K., Ridley, A., Kaplin, B. A., & Grueter, C. C. (2020). A comparison of fecal sampling and direct feeding observations for quantifying the diet of a frugivorous primate. *Current Zoology*, *66*, 333–343. https://doi.org/10.1093/CZ/ZOZ058

Meiklejohn, K. A., Damaso, N., & Robertson, J. M. (2019). Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. *PLoS One*, *14*, e0217084. https://doi.org/10.1371/journal.pone.0217084

Mena, J. L., Yagui, H., Tejeda, V., Bonifaz, E., Bellemain, E., Valentini, A., Tobler, M. W., Sánchez-Vendizú, P., & Lyet, A. (2021). Environmental DNA metabarcoding as a useful tool for evaluating terrestrial mammal diversity in tropical forests. *Ecological Applications*, *31*, e02335. https://doi.org/10.1002/eap.2335

Mercier, C., Boyer, F., Bonin, A., & Coissac, E. (2013). SUMATRA and SUMACLUST: Fast and exact comparison and clustering of sequences. *Programs and Abstracts of the SeqBio 2013 Workshop, GdRBIM and GdrIM*, Montpellier, France.

Mucina, L., & Rutherford, M. C. (2006). *The vegetation of South Africa, Lesotho and Swaziland (Strelitzia)*. South African National Biodiversity Institute.

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

Nørgaard, L., Olesen, C. R., Trøjelsgaard, K., Pertoldi, C., Nielsen, J. L., Taberlet, P., Ruiz-González, A., De Barba, M., & Iacolina, L. (2021). eDNA metabarcoding for biodiversity assessment, generalist predators as sampling assistants. *Scientific Reports*, *11*, 6820. https://doi.org/10.1038/s41598-021-85488-9

Osman, N. A., Abdul-Latiff, M. A. B., Mohd-Ridwan, A. R., Yaakop, S., Nor, S. M., & Md-Zain, B. M. (2020). Diet composition of the wild stump-tailed macaque (*Macaca arctoides*) in Perlis State Park, Peninsular Malaysia, using a chloroplast tRNL DNA metabarcoding approach: A preliminary study. *Animals*, *10*, 1–14. https://doi.org/10.3390/ani10122215

Pansu, J., Guyton, J. A., Potter, A. B., Atkins, J. L., Daskin, J. H., Wursten, B., Kartzinel, T. R., & Pringle, R. M. (2019). Trophic ecology of large herbivores in a reassembling African ecosystem. *Journal of Ecology*, *107*, 1355–1376. https://doi.org/10.1111/1365-2745.13113

Pickett, S. B., Bergey, C. M., & Di Fiore, A. (2012). A metagenomic study of primate insect diet diversity. *American Journal of Primatology*, *74*, 622–631. https://doi.org/10.1002/ajp.22014

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M., & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*, 1246752. https://doi.org/10.1126/science.1246752

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

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

Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: Diet assessment using next generation sequencing. *Molecular Ecology*, *21*, 1931–1950. https://doi.org/10.1111/j.1365-294X.2011.05403.x

Quéméré, E., Hibert, F., Miquel, C., Lhuillier, E., Rasolondraibe, E., Champeau, J., Rabarivola, C., Nusbaumer, L., Chatelain, C., Gautier, L., Ranirison, P., Crouau-Roy, B., Taberlet, P., & Chikhi, L. (2013). A DNA metabarcoding study of a primate dietary diversity and plasticity across its entire fragmented range. *PLoS One*, *8*, e58971. https://doi.org/10.1371/journal.pone.0058971

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). The detection of aquatic animal species using environmental DNA-A review of eDNA as a survey tool in

ecology. *Journal of Applied Ecology*, *51*, 1450–1459. https://doi.org/10.1111/1365-2664.12306

Rothman, J. M., Raubenheimer, D., Bryer, M. A. H., Takahashi, M., & Gilbert, C. C. (2014). Nutritional contributions of insects to primate diets: Implications for primate evolution. *Journal of Human Evolution*, *71*, 59–69. https://doi.org/10.1016/j.jhevol.2014.02.016

Rowe, A. K., Donohue, M. E., Clare, E. L., Drinkwater, R., Koenig, A., Ridgway, Z. M., Martin, L. D., Nomenjanahary, E. S., Zakamanana, F., Randriamanandaza, L. J., Rakotonirina, T. E., & Wright, P. C. (2021). Exploratory analysis reveals arthropod consumption in 10 lemur species using DNA metabarcoding. *American Journal of Primatology*, *83*, e23256. https://doi.org/10.1002/ajp.23256

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

Schwitzer, C., Mittermeier, R. A., Rylands, A. B., Chiozza, F., Williamson, E. A., Macfie, E. J., Wallis, J., & Cotton, A. (2017). *Primates in Peril: The World's 25 Most Endangered Primates 2016–2018*. IUCN SSC Primate Specialist Group (PSG), International Primatological Society (IPS), Conservation International (CI), and Bristol Zoological Society. https://doi.org/10.1896/052.022.0101

Shao, X., Lu, Q., Liu, M., Xiong, M., Bu, H., Wang, D., Liu, S., Zhao, J., Li, S., & Yao, M. (2021). Generalist carnivores can be effective biodiversity samplers of terrestrial vertebrates. *Frontiers in Ecology and the Environment*, *19*, 557–563. https://doi.org/10.1002/fee.2407

Shehzad, W., Riaz, T., Nawaz, M. A., Miquel, C., Poillot, C., Shah, S. A., Pompanon, F., Coissac, E., & Taberlet, P. (2012). Carnivore diet analysis based on next-generation sequencing: Application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, *21*, 1951–1965. https://doi.org/10.1111/j.1365-294X.2011.05424.x

Srivathsan, A., Sha, J. C. M., Vogler, A. P., & Meier, R. (2015). Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (*Pygathrix nemaeus*). *Molecular Ecology Resources*, *15*, 250–261. https://doi.org/10.1111/1755-0998.12302

Struhsaker, T. T. (1967). Ecology of vervet monkeys (*Cercopithecus Aethiops*) in the Masai-Amboseli game reserve, Kenya. *Ecology*, *48*, 891–904. https://doi.org/10.2307/1934531

Su, H.-H., & Lee, L.-L. (2001). Food habits of Formosan rock macaques (*Macaca cyclopis*) in Jentse, Northeastern Taiwan, assessed by fecal analysis and behavioral observation. *International Journal of Primatology*, *22*, 359–377. https://doi.org/10.1023/A:1010799410911

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

Taberlet, P., Bonin, A., Zinger, L., & Coissac, É. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press. https://doi.org/10.1093/oso/9780198767220.001.0001

Taberlet, P., Coissac, E., Gielly, L., Miquel, C., Brochmann, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., & Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, *35*, e14. https://doi.org/10.1093/nar/gkl938

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*, 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tercel, M. P. T. G., Symondson, W. O. C., & Cuff, J. P. (2021). The problem of omnivory: A synthesis on omnivory and DNA metabarcoding. *Molecular Ecology*, *30*, 2199–2206. https://doi.org/10.1111/mec.15903

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Tournier, E., Tournier, V., van de Waal, E., Barrett, A., Brown, L., & Bshary, R. (2014). Differences in diet between six neighbouring groups of vervet monkeys. *Ethology*, *120*, 471–482. https://doi.org/10.1111/eth.12218

van Wyk, B., & van Wyk, P. (2013). *Field guide to trees of Southern Africa* (2nd ed.). Penguin Random House South Africa.

Vesterinen, E. J., Puisto, A. I. E., Blomberg, A. S., & Lilley, T. M. (2018). Table for five, please: Dietary partitioning in boreal bats. *Ecology and Evolution*, *8*, 10914–10937. https://doi.org/10.1002/ece3.4559

Voelker, M. R., Schwarz, D., Thomas, A., Nelson, B. W., & Acevedo-Gutiérrez, A. (2020). Large-scale molecular barcoding of prey DNA reveals predictors of intrapopulation feeding diversity in a marine predator. *Ecology and Evolution*, *10*, 9867–9885. https://doi.org/10.1002/ece3.6638

Whiten, A., & van de Waal, E. (2018). The pervasive role of social learning in primate lifetime development. *Behavioral Ecology and Sociobiology*, *72*, 80. https://doi.org/10.1007/s00265-018-2489-3

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., Chariton, A. A., Creer, S., Coissac, E., Deagle, B. E., De Barba, M., Dickie, I. A., Dumbrell, A. J., Ficetola, G. F., Fierer, N., Fumagalli, L., Gilbert, M. T. P., Jarman, S., Jumpponen, A., & Taberlet, P. (2019). DNA metabarcoding–Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, *28*, 1857–1862. https://doi.org/10.1111/mec.15060

Zinger, L., Lionnet, C., Benoiston, A. S., Donald, J., Mercier, C., & Boyer, F. (2021). metabaR: An R package for the evaluation and improvement of DNA metabarcoding data quality. *Methods in Ecology and Evolution*, *12*, 586–592. https://doi.org/10.1111/2041-210X.13552

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Brun, L., Schneider, J., Carrió, E. M., Dongre, P., Taberlet, P., Waal, v. d., & Fumagalli, L. (2022). Focal vs. fecal: Seasonal variation in the diet of wild vervet monkeys from observational and DNA metabarcoding data. *Ecology and Evolution*, *12*, e9358. https://doi.org/10.1002/ece3.9358

Chapter 2: Molecular assessment of dietary variation in neighbouring primate groups

Judith Schneider[1,§], Loïc Brun[1,2§], Pierre Taberlet[3,4], Luca Fumagalli[1,5,‡] and Erica van de Waal[2,6,7,‡]

[§] joint first authors

[‡] joint last authors

[1]Laboratory for Conservation Biology, Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland.

[2]Inkawu Vervet Project, Mawana Game Reserve, Swart Mfolozi, KwaZulu Natal, South Africa.

[3]Université Grenoble Alpes, CNRS, Laboratoire d'Ecologie Alpine, F-38000 Grenoble, France.

[4]UiT – The Arctic University of Norway, Tromsø Museum, Tromsø, Norway.

[5]Swiss Human Institute of Forensic Taphonomy, University Centre of Legal Medicine Lausanne-Geneva, Lausanne University Hospital and University of Lausanne, Ch. de la Vulliette 4, 1000 Lausanne 25, Switzerland.

[6]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland.

[7]Centre for Functional Biodiversity, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

corresponding author: Judith Schneider, judith.schneider@unil.ch.

Chapter status: published, https://doi.org/ 10.1111/2041-210X.14078

Author contribution
EW and LF designed the study and supervised all the analyses. LB conducted the field work. JS, LB and PT performed the laboratory work and bioinformatics. LB and JS conducted the data analyses. LB and JS wrote the first draft of the manuscript, with input from all other authors.

Received: 25 April 2022 | Accepted: 2 February 2023

DOI: 10.1111/2041-210X.14078

Methods in Ecology and Evolution

RESEARCH ARTICLE

Contemporary Methods for Studying Animal Sociality in the Wild

Molecular assessment of dietary variation in neighbouring primate groups

Judith Schneider[1] | **Loïc Brun**[1,2] | **Pierre Taberlet**[3,4] | **Luca Fumagalli**[1,5] | **Erica van de Waal**[2,6,7]

[1]Laboratory for Conservation Biology, Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland; [2]Inkawu Vervet Project, Mawana Game Reserve, Swart Mfolozi, South Africa; [3]Laboratoire d'Ecologie Alpine, Université Grenoble Alpes, CNRS, Grenoble, France; [4]UiT – The Arctic University of Norway, Tromsø Museum, Tromsø, Norway; [5]Swiss Human Institute of Forensic Taphonomy, University Centre of Legal Medicine Lausanne-Geneva, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland; [6]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland and [7]Centre for Functional Biodiversity, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Correspondence
Judith Schneider
Email: judith.schneider@unil.ch

Funding information
Branco Weiss Fellowship - Society in Science; H2020 European Research Council, Grant/Award Number: 949379; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 310030_192512 and PP00P3_170624; Université de Lausanne

Handling Editor: Thibaud Gruber

Abstract

1. Facing rapid environmental changes and anthropogenic habitat destruction, animal behavioural plasticity becomes an adaptive potential that needs to be considered in conservation strategies along with, for example, genetic diversity. Here, we evaluate to what extent non-invasive environmental DNA (eDNA) methods may contribute to the assessment of intraspecies behavioural plasticity in terms of foraging behaviour.
2. We analysed DNA metabarcoding data for plant components in the diet of four neighbouring groups of wild vervet monkeys *Chlorocebus pygerythrus* to identify intergroup variation (IGV). The faecal samples considered for the analyses were limited to the summer season to minimise the impact of seasonality. Each sample was attributed by observation to individuals with known life history data. A plant survey was conducted in each group home range during the study period to account for environmental variation.
3. We observed mixed results when testing whether IGV in plant consumption was greater than intragroup variation, indicating that the influence of social dynamics must be considered. Intragroup variation was positively correlated with group size. We observed IGV in diet composition among all groups as well as in some pairwise comparisons. We found significant dietary differences between two group pairs when considering only adult females. Lastly, we observed IGV in foraging of specific plants that were not explained by their distribution, suggesting behavioural differences in selectivity between groups.

Judith Schneider and Loïc Brun joint first authors.

Luca Fumagalli and Erica van de Waal joint last authors.

4. Our study system and organism, being a highly social and non-threatened primate species, with constant gene flow and overlapping territories between groups, provides an ideal model to evaluate the usage of eDNA-based methods to better understand the impact of social factors on IGV. Our results highlight the need to consider social and demographic factors, the impact of which remains complicated to disentangle from environmental factors. However, we emphasise the great potential for studying social groups using eDNA and that such studies are needed to better understand intraspecific behavioural plasticity in wild populations.

KEYWORDS
DNA metabarcoding, environmental DNA, foraging behaviour, intergroup variation, primate population, vervet monkeys

1 | INTRODUCTION

The analysis of DNA extracted from environmental samples (i.e. environmental DNA; eDNA) has seen a rapid implementation in various research fields (Bohmann et al., 2014; Ruppert et al., 2019; Taberlet et al., 2018; Thomsen & Willerslev, 2015). In particular, the development of DNA metabarcoding (PCR amplification of short but informative metabarcodes with universal primers and next generation sequencing [NGS] of DNA mixtures, Taberlet et al., 2012) enables comprehensive taxonomic identification of complex environmental samples.

DNA metabarcoding often provides higher taxonomic resolution and coverage than traditional methods (Nørgaard et al., 2021; Ruppert et al., 2019). For terrestrial species, studies commonly rely on faecal sampling for diet characterisation (De Barba et al., 2014; Shehzad et al., 2012), parallel prey and predator identification (Galan et al., 2018; Gillet et al., 2015) and biodiversity assessment (Nørgaard et al., 2021; Shao et al., 2021). In the field of primatology, the most commonly used methods for dietary analyses are direct observation of feeding events and microhistology of faecal samples as discussed in (Matthews et al., 2020). Both are time- and labour-intensive, rely heavily on taxonomic expertise and results are often constrained in taxonomic resolution and coverage (Nielsen et al., 2018), as the identification of consumed items or the observation of feeding events themselves are often challenging (Matthews et al., 2020; Pickett et al., 2012). Depending on the studied organism and field conditions, observations on broad temporal and spatial scales are complicated. Recently, the use of DNA metabarcoding in the field of primatology has enabled new insights, in particular regarding the consumption of arthropods (Lyke et al., 2018; Mallott et al., 2015, 2017; Rowe et al., 2021) but also plants (Brun, Schneider, Mas Carrió, Dongre, van de Waal, et al., 2022; Mallott et al., 2018; Quéméré et al., 2013). The sampling procedure of eDNA promises new opportunities to investigate behavioural plasticity through the study of foraging behaviour of species that are challenging to observe but for which faecal samples can be obtained.

DNA metabarcoding is constrained, however, in that it cannot, for example, identify different life stages or states of detected organisms, and it is more complicated to estimate abundances than with observations due to several potential biases that need to be considered (Piñol et al., 2019). Nevertheless, eDNA approaches can add valuable information for ecological network analysis (Clare, 2014) and several studies, for example in the context of niche partitioning, have used relative read abundance (RRA; the number of a specific sequence divided by the total read of reads within a sample) to use DNA metabarcoding semi-quantitatively (Deagle et al., 2019; Kartzinel et al., 2015; Pansu et al., 2019; Vesterinen et al., 2018). However, there are few studies that assess differences beyond the population level (see the work of Voelker et al., 2020). DNA metabarcoding can hence be useful to assess feeding patterns between different organisms, groups or populations.

Considering the intergroup-level, the unit of analysis avoids species-wide generalisations based on behavioural studies of single populations (Kaufhold & van Leeuwen, 2019; van de Waal, 2018). To date, most studies assessing intergroup variation (IGV) in primates have set the focus on tool-use (Luncz et al., 2012; Tan et al., 2015), social behaviour (Borgeaud et al., 2016; DeTroy et al., 2021; van Leeuwen et al., 2018, 2021) or both (Santorelli et al., 2011; Whiten et al., 1999). There are few studies on IGV in foraging behaviour (Quéméré et al., 2013; Samuni et al., 2020; Tournier et al., 2014). The majority of studies on IGV assess differences qualitatively rather than quantitatively (see, e.g. the studies of Luncz et al., 2012; Samuni et al., 2020; van Leeuwen et al., 2021). This study investigates to what extent DNA metabarcoding data allows assessment of IGV quantitatively in foraging behaviour, that is, intraspecies behavioural variation at the group level. The aim is to assess whether the method works effectively in our study system, and to evaluate its potential as a means of capturing cultural diversity in a wider context for consideration in conservation measures.

Research at the Inkawu Vervet Project (IVP), Mawana game reserve (28°00.327S, 031°12.348E), South Africa, focuses on the behavioural ecology of wild vervet monkeys *Chlorocebus pygerythrus*.

This species lives in social groups with female philopatry and male dispersal. The constant gene flow between groups reduces genetic differences within the population (Cheney & Seyfarth, 1983). The high social learning capacity of vervet monkeys has been demonstrated in manifold experiments (Mertz et al., 2019; Whiten & van de Waal, 2018). Indeed, in philopatric species, the social structure promotes the development of distinct group cultures (van de Waal, 2018) and females, the philopatric sex in vervets, have been shown to be preferred role models under experimental conditions (van de Waal et al., 2010, 2014). Therefore, we predict that the diet of adult females best represents that of a group, and hence IGV (measured by RRA) will be accentuated when focussing on adult females only.

We have shown previously a strong seasonal effect on vervet monkeys' diet analysing 823 faecal samples collected over a year, and a strong correlation between plant RRA and observational data, validating the use of RRA as a semi-quantitative measure of consumption in this system. The results indicate that selective feeding behaviours are more likely to occur in summer when resources are more abundant than in scarcer seasons (Brun, Schneider, Mas Carrió, Dongre, van de Waal, et al., 2022). The current study therefore focuses on assessing IGV between four neighbouring groups of vervets during the summer season. We used the RRA of dietary items detected in faecal samples to investigate intrapopulation behavioural plasticity, possibly learnt and transmitted through social learning. According to the 'exclusion principle' variation in a behaviour that is not induced by genetic or environmental differences is likely to result from social learning (van de Waal et al., 2015). Therefore, we assessed (a) whether IGV in the diet was greater than intragroup variation; (b) whether IGV was greater when considering all individuals in the group or only adult females, the philopatric sex; and (c) we investigated the relation between the availability of single food items per home range and their consumption. We use these data to illustrate the potential of the method, in particular the use of an eDNA approach as a promising tool to go beyond classical observational analyses of diet composition. Finally, we discuss challenges arising from the method broadening the perspectives on how to assess intraspecies foraging behavioural variation in wild animals using eDNA approaches.

2 | MATERIALS AND METHODS

2.1 | Studied vervet monkey groups and sampling

Our data were collected from four IVP long-term studied groups with overlapping territories (Figure S1): Ankhase (AK), Baie Dankie (BD), Kubu (KB) and Noha (NH). All individuals were identified using specific bodily and facial features (e.g. scars, colours, shape) and each sample was assigned by experienced field assistants to a specific vervet monkey and consequently linked to available life history data. We defined adulthood for males and females separately, as their life cycles follow different patterns: 3 years for females, and 4 years for males if they dispersed, otherwise 5 years. To avoid redundancy, the social factors sex and age were combined into one category with three levels (female adults, male adults, juveniles) during analyses. Infants were not sampled because they are born at the start of the summer and only feed by nursing for the first 3 months. Infants of the previous year were already 1 year old, and thus in the juvenile class. Approximately 0.5 cm^3 from inside a scat were collected with gloves and a disposable plastic spoon into 20 mL HDPE scintillation vials (Carl Roth GmbH, Karlsruhe, Germany) and covered with 10 mL absolute ethanol, immediately after an individual was observed defecating. After 24–36 h, the ethanol was replaced by silica gel beads and samples stored until DNA extraction. For this study we use the sequence data of 372 faecal samples collected in summer, from mid-November 2018 to mid-March 2019 (Table S1).

Fifty and 95% core areas of each group's home range were calculated based on GPS positions (639 scat sampling and 4669 locations of observations, Table S2) using the *Brownian bridge movement model* (Horne et al., 2007). A full year's data was included, as the use of more data provided better estimates of the model parameters for the different home ranges and their respective core areas. Furthermore, these remain stable throughout the year in this species (Cheney, 1981; Struhsaker, 1967). Subsequently, to account for variable plant distributions across groups' home ranges, ten square vegetation plots (each 1600 m^2) were randomly allocated per 50% core area of each group using QGIS software, and the vegetation cover of 52 presumed forage plants was recorded to estimate local abundance (Figure S1). The final dataset comprised coverage data of 27 plants that the monkeys consumed, and which were also detected in faecal samples. Species accumulation curves (SACs) made with the VEGAN package (Oksanen et al., 2014) showed the adequacy of this survey for representing the distribution of plants in the study area (Figure S2). We controlled for homogeneity of group dispersions with the *betadisper* function (*vegan*) before investigating potential variation in plant coverage between groups' territories with a permutational multivariate analysis of variance (*PERMANOVA* with Bonferroni correction, pseudo $F_{40} = 1.44$, $R^2 = 0.11$, $p = 0.12$) on Bray–Curtis dissimilarity matrices, and pairwise tests also revealed no significant difference (Figure S3).

2.2 | DNA metabarcoding

DNA extraction of scat samples was performed using a phosphate buffer-based approach (Taberlet et al., 2018) following a modified protocol of the NucleoSpin Soil Kit (Macherey-Nagel, Düren, Germany), as described in (Brun, Schneider, Mas Carrió, Dongre, van de Waal, et al., 2022). Extractions were performed in a pre-PCR laboratory exclusively dedicated to low DNA-content analyses (Laboratory for Conservation Biology, University of Lausanne, Switzerland). To assess the plant part of the diet, DNA extracts were amplified in triplicates with a primer pair (Sper01) targeting the P6 loop of the *trn*L intron (UAA) of chloroplast DNA (Taberlet et al., 2018). Library preparation was performed using the TruSeq

DNA PCR-Free Library Prep Kit (Illumina) and libraries were 150 paired-end sequenced on the Illumina Miniseq Sequencing System (Illumina). Bioinformatic processing of raw sequences was conducted with the OBITOOLS package (Boyer et al., 2016) and in R (version 4.0.2). All details of experimental conditions and sequence alignment, filtering, clustering, data cleaning based on controls and taxonomic assignments are described in (Brun, Schneider, Mas Carrió, Dongre, van de Waal, et al., 2022).

2.3 | Data analyses

In order to test the assumption that intergroup variation was greater than intragroup variation, we used the weighted means of Bray–Curtis dissimilarities ranging from 0 (complete overlap) to 1 (complete nonoverlap). For all analyses, data was aggregated as the mean of RRA per plant item and per monkey to account for pseudoreplication. Dietary patterns of the four groups were visualised with non-metrical dissimilarity scaling (NMDS). The assumption of homogeneity of group dispersions was tested with the *betadisper* function in VEGAN package (Oksanen et al., 2014). If these were homogenous, we performed *PERMANOVA* with 9999 permutations and Bonferroni correction on Bray–Curtis dissimilarity matrices to test the effect of the group and sex/age variables on diet composition. If dispersions were heterogenous, we used beta regression models taking into account these dispersions to assess differences in proportions of dietary plant species per group with the package BETAREG (Cribari-Neto & Zeileis, 2010) and the *joint-tests* function in the EMMEANS package (Lenth, 2019) to assess the main effects of the models.

For certain plant species ($n = 9$), we investigated the extent to which consumption was related to environmental factors, that is, plant coverage in the groups' home ranges. In a first step, a feature selection analysis based on a random forest algorithm with 2000 permutations was conducted in the R BORUTA package (Kursa & Rudnicki, 2010) to determine, which species were significantly impacting on IGV in diet. Of the 61 plant species consumed, the random forest algorithm identified 16 species that differed significantly between groups and for nine of these distribution data was available (Figure S4, Table S3). Subsequently, for these nine species RRA data was corrected for the coverage in a group's territory to account for environmental differences (RRA*(1-(percentage of coverage in the group's territory/100))). The Jacob's D index was calculated for these species ranging from −1 (avoidance) to +1 (preference) to visualise differences in selectivity between the groups.

3 | RESULTS

Regarding our hypothesis that intragroup variation was lower than intergroup variation, we found inconsistent results when all individuals were included, since it holds for the smaller groups (AK, KB) but not for the larger ones (BD, NH), as shown in Table 1. A Pearson's product–moment correlation confirmed this positive relationship between group size and increased intragroup variation (0.97, $p = 0.03$). The intragroup dispersions of the four groups were heterogenous ($F = 27.15$, $p = 0.001$). However, we observed group clustering in the NMDS (Figure 1a) and the boxplots of centroids (Figure S5) also indicated location effects; the heterogenous dispersions might have been caused by the unbalanced sample size. Using beta regression, the variable group was significant for all groups ($F_{ratio} = 8.49$, $p < 0.0001$). Testing the groups pairwise, we observed significant effects of the factor group for AK/BD ($F_{ratio} = 4.43$, $p = 0.0353$), AK/KB ($F_{ratio} = 4.62$, $p = 0.0316$), BD/KB ($F_{ratio} = 22.95$, $p < 0.0001$) and KB/NH ($F_{ratio} = 13.17$, $p = 0.0003$). PERMANOVA showed no significant effect of the variable sex/age on dietary variations between groups ($F = 1$, $R^2 = 0.02$, $p = 0.4021$).

TABLE 1 Weighted means of Bray–Curtis dissimilarities of intragroup and intergroup dietary variation using plant RRA data. Green colour indicates that our results were in line with the hypothesis that intergroup variation was greater than intragroup variation (compared to intragroup RRA per row) and red colour that they were not in line with our hypothesis meaning that intragroup variation was greater than intergroup variation. Stars indicate significance level for the group variable in pairwise regression models (*** < 0.001, ** < 0.01, * < 0.05).

Group	AK	BD	KB	NH	Intragroup
AK		0.59*	0.36*	0.48	0.37
BD	0.59*		0.57***	0.58	0.60
KB	0.36*	0.57***		0.46***	0.28
NH	0.48	0.58	0.46***		0.50

The results above point out that group demographics and social dynamics are important factors influencing the foraging behaviour of the studied groups, highlighting the need of further analyses for certain classes of individuals or on the intragroup level. Our hypothesis that intergroup variation was higher than intragroup variation was more accurate when including only adults of the philopatric sex (Table 2), illustrated also by the NMDS (Figure 1b). While for adult females of all groups, dispersions were also heterogenous ($F = 5.46$, $p = 0.005$), the results differed for pairwise tests as these were homogenous for KB/AK and BD/NH. PERMANOVA showed that group explained part of the variance in diet composition for KB/AK ($F = 3.69$, $R^2 = 0.22$, $p = 0.0024$) and BD/NH ($F = 3.76$, $R^2 = 0.13$, $p = 0.0056$). With the beta regression model, the variable group was neither significant for all groups ($F_{ratio} = 1.5$, $p = 0.21$) nor for the other group combinations. We found no significant effect of the variable sex/age at the intragroup level. The small sample size for adult males biases comparisons but there is nevertheless some structuring in the NMDS plots per group (Figure S6).

Figure 2 shows the group-specific selection patterns between resource availability and consumption (measured by Jacob's D index), for those plant species indicated by random forest analysis as being variably consumed between groups and for which distribution data were available. We observed that *Berchemia zeyheri*, a tree whose fruits are a main resource in summer, is highly consumed by all groups, but the least by BD. The resource distribution does not in

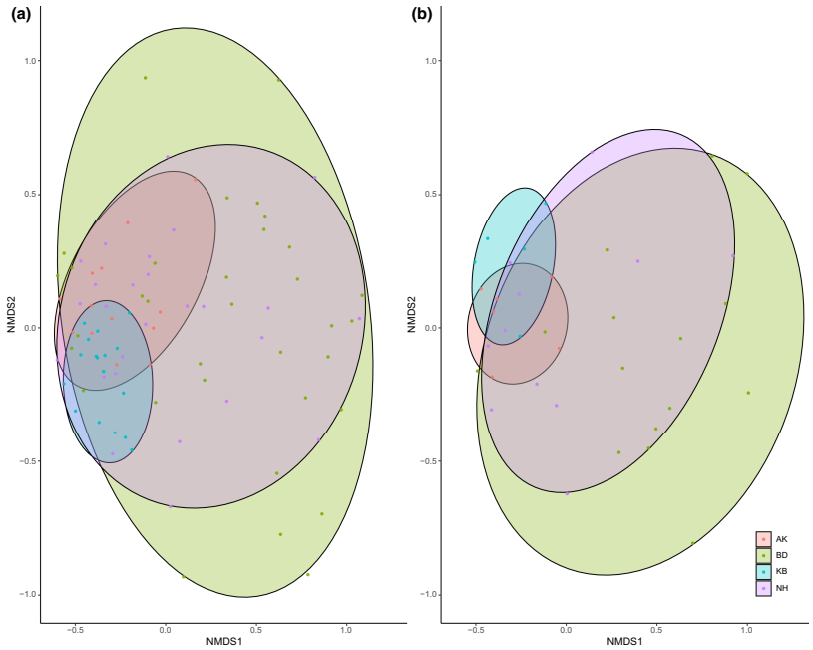


FIGURE 1 Nonmetric multidimensional scaling (NMDS), based on relative read abundances (RRA) of consumed plants detected in faecal samples aggregated per monkey per group in summer for (a) all individuals ($n = 103$) and (b) adult females ($n = 42$). The colours represent the four groups (Ankhase, AK; Baie Dankie, BD; Noha, NH and Kubu, KB).

all cases explain consumption patterns, as it is the case for example for *Cereus jamacaru*, *Hippobromus pauciflorus* and *Premna mooeinsis*. *Ziziphus mucronata* is a main food source in winter, explaining here the pattern of low consumption in summer.

4 | DISCUSSION

Behavioural diversity can be a product of both ecological factors and of cultural traits and it is very difficult to distinguish the effect of even small-scale variation (Brakes et al., 2021; Samuni et al., 2020). We have shown previously that there are strong seasonal patterns in the diet of our population, thus being strongly correlated with resource availability (Brun, Schneider, Mas Carrió, Dongre, van de Waal, et al., 2022). Therefore, to assess IGV, we focused here on one season to reduce environmental variation, selecting summer when food resources are more abundant as it increases individuals' opportunities to select food according to their preferences. The results did not fully match our assumption (IGV was not consistently higher than intragroup variation). Intragroup variation was correlated with group size and may reflect higher inter-individual competition for resources; higher in larger groups (BD and NH) and lower in smaller ones (AK and KB). Both smaller groups in our study had only one adult male at that time and it was therefore not possible to disentangle the relative impact of group size and sex on the observed reduction of intragroup variation. We found an effect of group by assessing the differences in proportions of consumed plants. We suppose that the heterogenous dispersions of group variances were a consequence of the unbalanced design and that both the NMDS (Figure 1a) and boxplots of centroids (Figure S5) indicated true differences in centroids. Future individual-level analyses with more balanced sample numbers will be beneficial to investigate the relationship between group

TABLE 2 Weighted means of Bray–Curtis dissimilarities of intragroup and intergroup dietary variation using plant RRA data for adult females only. Green colour indicates that our results were in line with the hypothesis that intergroup variation was greater than intragroup variation and red colour that they were not in line with our hypothesis meaning that intragroup variation was greater than intergroup variation. Triangles indicate significance level for differences between groups in pairwise *PERMANOVA* with 9999 permutations for the group variable, where model assumption where fulfilled (▲▲ < 0.01).

Group	AK	BD	KB	NH	Intragroup
AK		0.56	0.37▲▲	0.46	0.34
BD	0.56		0.57	0.56▲▲	0.55
KB	0.37▲▲	0.57		0.47	0.33
NH	0.46	0.56▲▲	0.47		0.49

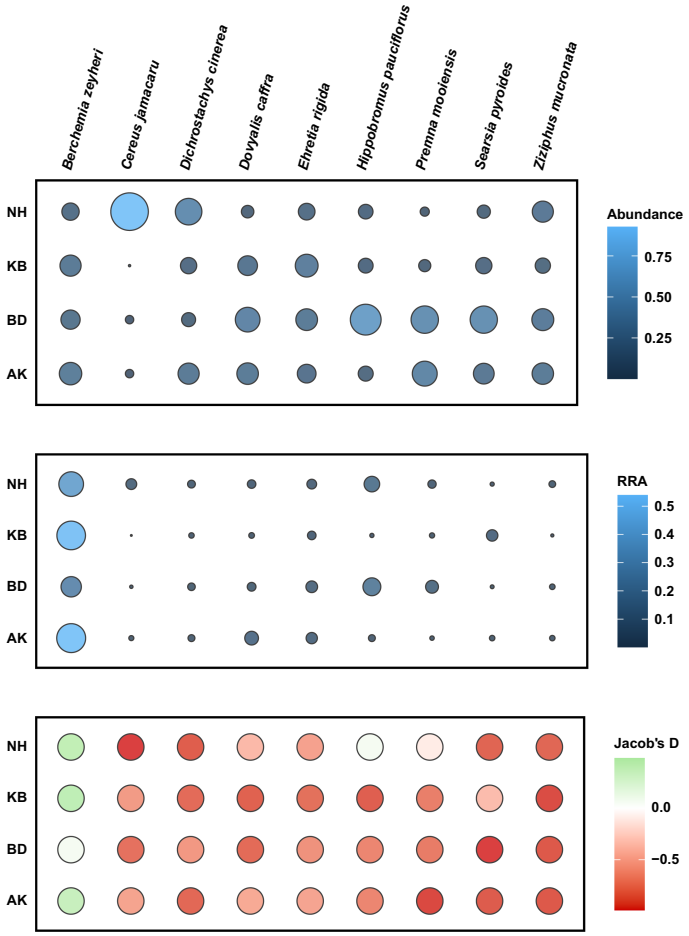


FIGURE 2 Resource availability, consumption, and selectivity of plant species indicated by random forest analysis and present in both plant cover assessments and sequence data. On the y-axis of each plot are the four groups: Noha (NH), Kubu (KB), Baie Dankie (BD), Ankhase (AK). (a) Proportional coverage per group home range; circle size and colour reflect relative abundance. (b) Proportional contribution of each plant to the diet of each group; circle size and colour reflect the mean of relative read abundances. (c) Jacob's D selectivity index for each plant taxon, ranging from −1 (low consumption compared to abundance, red) to 1 (strongest selection or high consumption compared to abundance, green).

size, composition, and diet. Demographic parameters may have an impact on an individual's diet, particularly when social interactions are strongly involved, as in the case of strong competition or when food acquisition is strongly mediated by social learning (Kaufhold & van Leeuwen, 2019; van Boekholt et al., 2021). Obtaining this kind of information in the wild is challenging, yet it is helpful to better understand the driving forces of IGV (DeTroy et al., 2021). While our results show that DNA metabarcoding is a useful approach when studying IGV, and that it may also bring insights into the study of social learning in non-experimental settings, it also emphasises the importance and difficulty of having balanced sampling when demographic parameters must be accounted for.

In this context, the focus on the philopatric sex also produced interesting results. Previous studies on social learning in our population demonstrated in the context of foraging experiments a bias to copy dominant females (van de Waal et al., 2010) and vertical transmission happening primarily in mother-infant dyads where naive infants first rely on their mothers' experience regarding foraging choices (van de Waal et al., 2013, 2014). In the context of a female philopatric system, these premises lay the necessary foundation for

the evolution of cultural variants of foraging behaviours that would essentially be maintained and transmitted by adult females. The heterogenous dispersions prevented unequivocal conclusions about all groups at once, but we observed the factor group to significantly affect the diet of two group pairs. It should be noted that these were on the one hand the smaller groups (AK/KB) and on the other hand the larger ones (BD/NH). A more balanced design may thus allow us to overcome the statistical issue of heterogenous dispersions and to also trace IGV between the remaining groups. However, the observed homogeneity could either be an effect of sample size reduction or of the exclusion of individuals foraging more diversely. When we looked at the results for all individuals, we saw the greatest variation between the smallest group with only one adult male (KB) and the largest group with the most adult males (BD), while we did not see any effect when only looking at adult females.

There is a range of interesting research questions regarding the effect of social factors that could be studied using eDNA data. Future analyses at the intragroup level could assess whether individual foraging is consistent over time and if matrilines (mother–offspring) show foraging behaviour distinct from other matrilines or if different energetic needs due to the reproductive state of females influence the diet. Albeit, sex was not a significant predictor for the present data, it could be worthwhile to study possibly greater behavioural flexibility of the dispersing group members, adult males, as shown previously in behavioural experiments and discussed above (Bono et al., 2018; van de Waal et al., 2013). Furthermore, it has been found that the sex of an individual can already lead to dietary variations in juveniles as described for orangutans (Ehmann et al., 2021). Here, this effect was not significant but the structure within the NMDS still suggests differences between male and female juveniles (Figure S7).

An observational study reported variable feeding on tree species between neighbouring groups of vervet monkeys with multiple mismatches between the dietary importance of a species and its local abundance (Tournier et al., 2014). After taking into account the differences in plant species distribution in the groups' home ranges, we observed variation in consumption of certain plant species that is not entirely explained by a difference in abundance in the territories (Figure 2). For instance, our data showed that monkeys in NH are more selective towards *P. mooiensis* and *H. pauciflorus* relative to its abundance in their home range than the other groups, and the same pattern can be observed for KB with *C. jamacaru*. A possible explanation for variation in selectivity could be that different plant species provide different nutrients, and when species are less abundant in a group's home range, the monkeys must compensate by consuming other species providing similar nutrient intakes. Alternatively, monkeys might have developed distinct food preferences not constrained by physiological needs but rather by social learning. Whenever possible, sampling of environmental factors should also respect seasonal variations, for example the phenology of plants. Here, the plant census was done solely in terms of coverage in the same season as the analysed samples, assuming equal intraspecies-phenology for the presented data. One option is to sample across seasons and to control for this factor in data analyses, but as we have seen, limited sample numbers and unbalanced sampling may be an issue and thus within season designs are a good alternative despite the risk of overlooking important patterns (Matthews et al., 2020).

Reliable and robust estimations of biomass or abundances are crucial to many research questions in ecology and conservation biology (Pimm et al., 2014). Often these revolve not around simple detection or non-detection but need abundance measures to lead to meaningful results. While abundances can be measured by observational methods, the quantification potential of eDNA-approaches is an unresolved debate (Calderón-Sanou et al., 2019; Cuff et al., 2022; Zinger et al., 2019). For example, PCR primer-induced biases, that is, the preferential amplification of certain taxa and the under- or non-representation of others, are considered a main source of biases in PCR-based target enrichment approaches such as DNA metabarcoding (Jusino et al., 2019; Piñol et al., 2015, 2019), and multiplexing of primers or the use of degenerated primers has been proposed as alternative (Dowle et al., 2016; Krehenwinkel et al., 2017). Diet studies are faced with the additional challenge of possible digestion-related biases (Clare, 2014). Macroscopic studies in primates provided evidence for different digestibility of different items (e.g. Matthews et al., 2020). A feeding trial with little penguins *Eudyptula minor* in captivity indicated that the initially fed proportions were not directly correlated to sequence counts (Deagle et al., 2010). Nonetheless, DNA metabarcoding offers the potential to semi-quantitatively study IGV provided that the same experimental conditions are maintained for all groups. Vervet monkeys are omnivorous, however, plants represent the main food source which makes them the first choice for assessment but also implies certain challenges (Cuff et al., 2022). We acknowledge that while we relied on a promising study system, the limited sample numbers and probably the targeted diet components did not allow us to draw the conclusions we had hoped for. An alternative is to study food items that are less frequently consumed or that are more difficult to prey on to inquire about IGV using occurrence data; for example, vertebrates in vervet monkeys or crabs in chimpanzees *Pan troglodytes verus* (Koops et al., 2019). In a study on bonobos *Pan paniscus*, Samuni et al. (2020) observed variation in hunting behaviour of mammalian prey of two groups sharing the same habitat.

The main advantages of an eDNA-based approach are that no experimental setup is needed, it is non-invasive, studied animals are hence not disturbed or influenced in their natural behaviour, and it can provide a thorough picture including both wide temporal and spatial scales (reviewed in Bohmann et al., 2014; Taberlet et al., 2012, 2018; Thomsen & Willerslev, 2015). DNA metabarcoding can also serve to study population structure (Bohmann et al., 2018) and in principle, using SNPs or microsatellite genotyping of faeces DNA would allow assessment of relatedness and genetic structure. A wealth of information can be extracted from samples particularly when conservation status and ethical issues prevent invasive tissue sampling. eDNA samples can *ideally* provide data on genetic *and* behavioural diversity.

To conclude, eDNA-based approaches offer new research opportunities to assess the influence of social factors on dietary variation, in particular for species that are not prone to observation, such as rare and endangered, nocturnal, elusive or dangerous ones. Obtaining information at the individual level might not always be feasible but investigating whether there are dietary variations between males and females, whether diets differ between groups that share a similar environment or, in contrast, that live in very different ones, would greatly contribute to our current knowledge of dietary IGV which has been little studied so far. When different foraging behaviours are detected, possible social transmission, and sometimes even cultural traits, can be studied later. Understanding the driving force and the circumstances that regulate IGV in different populations or species could provide significant insights for various fields of research, including behavioural ecology and cultural evolution but also for applied conservation. In light of climate change and increased anthropogenic habitat destruction, behavioural plasticity might be an important means of responding to rapid disturbances (Brakes et al., 2019, 2021; Gruber et al., 2019). Similarly to the identification of hotspots of genetic diversity to prioritise conservation efforts, cultural traits should be taken into consideration to define populations with the greatest potential for survival (Keith & Bull, 2017; Kühl et al., 2019; Sih, 2013). Cultural transmission of different behaviours through social learning may establish distinct traditions that define a culture, differentiating populations or subpopulations, leading implicitly to varying adaptive potentials. However, the identification of socially transmitted variants and the subsequent potential for cultural differentiation remains challenging to observe in the wild and in this context eDNA techniques might prove valuable.

AUTHOR CONTRIBUTIONS

Erica van de Waal and Luca Fumagalli conceived the ideas and designed methodology; Loïc Brun collected the data; Loïc Brun, Judith Schneider and Pierre Taberlet performed lab work; Judith Schneider and Loïc Brun analysed the data; Judith Schneider led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

The authors thank the van der Walt family for permission to conduct the study on their land. The authors are very grateful to Pooja Dongre for coordinating the start of sample collection and to the whole IVP team who assisted in the field work. The authors thank Eduard Mas Carrió for his precious help during data curation. The authors also thank the Editor and three anonymous reviewers for their critical and valuable comments on earlier versions of the manuscript. The observational research was approved by the relevant local authority "Ezemvelo KZN Wildlife" in South Africa. The field costs during data and sample collection were funded by grants to EW from the Branco Weiss Fellowship-Society in Science and the Swiss National Science Foundation (grant number PP00P3_170624), and the grant 'ProFemmes' of the Faculty of Biology and Medicine, University of Lausanne. At the time of writing, Erica van de Waal and Loïc Brun were supported by the European Research Council under the European Union's Horizon 2020 research and innovation programme for the ERC 'KNOWLEDGE MOVES' starting grant (grant agreement No. 949379). This project was supported by a Swiss National Science Foundation grant to Luca Fumagalli (nr. 310030_192512).

CONFLICT OF INTEREST STATEMENT

The authors note that Pierre Taberlet is co-inventor of a patent related to the Sper01 primers and the use of the P6 loop of the chloroplast *trnL* (UAA) intron for plant identification using degraded template DNA. This patent only restricts commercial applications and have no impact on the use of this loci by academic researchers.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14078.

DATA AVAILABILITY STATEMENT

Data available via the Dryad Digital Repository https://doi.org/10.5061/dryad.6q573n621 (Brun, Schneider, Mas Carrió, Dongre, Taberlet, et al., 2022). Sanger sequences for the local database have been deposited in GenBank under accession numbers OL898555-OL898608.

ORCID

Judith Schneider https://orcid.org/0000-0003-2663-6130
Loïc Brun https://orcid.org/0000-0003-3009-9091
Pierre Taberlet https://orcid.org/0000-0002-3554-5954
Luca Fumagalli https://orcid.org/0000-0002-6648-2570
Erica van de Waal https://orcid.org/0000-0001-7778-418X

REFERENCES

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, *29*(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Bohmann, K., Gopalakrishnan, S., Nielsen, M., Nielsen, L. D. S. B., Jones, G., Streicker, D. G., & Gilbert, M. T. P. (2018). Using DNA metabarcoding for simultaneous inference of common vampire bat diet and population structure. *Molecular Ecology Resources*, *18*, 1050–1063. https://doi.org/10.1111/1755-0998.12891

Bono, A. E. J., Whiten, A., van Schaik, C., Krützen, M., Eichenberger, F., Schnider, A., & van de Waal, E. (2018). Payoff- and sex-biased social learning interact in a wild primate population. *Current Biology*, *28*, 2800–2805. https://doi.org/10.1016/j.cub.2018.06.015

Borgeaud, C., Sosa, S., Bshary, R., Sueur, C., & van de Waal, E. (2016). Intergroup variation of social relationships in wild vervet monkeys: A dynamic network approach. *Frontiers in Psychology*, *7*, 915. https://doi.org/10.3389/fpsyg.2016.00915

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*, 176–182. https://doi.org/10.1111/1755-0998.12428

Brakes, P., Carroll, E. L., Dall, S. R. X., Keith, S. A., McGregor, P. K., Mesnick, S. L., Noad, M. J., Rendell, L., Robbins, M. M., Rutz, C., Thornton, A., Whiten, A., Whiting, M. J., Aplin, L. M., Bearhop, S., Ciucci, P., Fishlock, V., Ford, J. K. B., Notarbartolo di Sciara, G., & Garland, E. C. (2021). A deepening understanding of animal culture suggests lessons for conservation. *Proceedings of the Royal Society B: Biological Sciences*, *288*, 20202718. https://doi.org/10.1098/rspb.2020.2718

Brakes, P., Dall, S. R. X., Aplin, L. M., Bearhop, S., Carroll, E. L., Ciucci, P., Fishlock, V., Ford, J. K. B., Garland, E. C., Keith, S. A., McGregor, P. K., Mesnick, S. L., Noad, M. J., Notarbartolo di Sciara, G., Robbins, M. M., Simmonds, M. P., Spina, F., Thornton, A., Wade, P. R., & Rutz, C. (2019). Animal cultures matter for conservation. *Science*, *363*(6431), 1032–1034. https://doi.org/10.1126/science.aaw3557

Brun, L., Schneider, J., Mas Carrió, E., Dongre, P., Taberlet, P., van de Waal, E., & Fumagalli, L. (2022). Focal vs. faecal: Seasonal variation in the diet of wild vervet monkeys from observational and DNA metabarcoding data. *Dryad Digital Repository*, *12*, 9358. https://doi.org/10.5061/dryad.6q573n621

Brun, L., Schneider, J., Mas Carrió, E., Dongre, P., van de Waal, E., & Fumagalli, L. (2022). Focal vs. faecal: Seasonal variation in the diet of wild vervet monkeys from observational and DNA metabarcoding data. *Ecology and Evolution*, *12*, e9358. https://doi.org/10.1002/ece3.9358

Calderón-Sanou, I., Münkemüller, T., Boyer, F., Zinger, L., & Thuiller, W. (2019). From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices? *Journal of Biogeography*, *47*, 193–206. https://doi.org/10.1111/jbi.13681

Cheney, D. L. (1981). Intergroup encounters among free-ranging vervet monkeys. *Folia Primatologica*, *35*, 124–146. https://doi.org/10.1159/000155970

Cheney, D. L., & Seyfarth, R. M. (1983). Nonrandom dispersal in free-ranging vervet monkeys: Social and genetic consequences. *The American Naturalist*, *122*(3), 392–412.

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

Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, *34*(2), 1–24. https://doi.org/10.18637/jss.v034.i02

Cuff, J. P., Windsor, F. M., Tercel, M. P. T. G., Kitson, J. J. N., & Evans, D. M. (2022). Overcoming the pitfalls of merging dietary metabarcoding into ecological networks. *Methods in Ecology and Evolution*, *13*, 545–559. https://doi.org/10.1111/2041-210X.13796

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*, 306–323. https://doi.org/10.1111/1755-0998.12188

Deagle, B. E., Chiaradia, A., McInnes, J., & Jarman, S. N. (2010). Pyrosequencing faecal DNA to determine diet of little penguins: Is what goes in what comes out? *Conservation Genetics*, *11*(5), 2039–2048. https://doi.org/10.1007/s10592-010-0096-6

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*, 391–406. https://doi.org/10.1111/mec.14734

DeTroy, S. E., Ross, C. T., Cronin, K. A., van Leeuwen, E. J. C., & Haun, D. B. M. (2021). Cofeeding tolerance in chimpanzees depends on group composition: A longitudinal study across four communities. *IScience*, *24*(3), 102175. https://doi.org/10.1016/j.isci.2021.102175

Dowle, E. J., Pochon, X., Banks, J. C., Shearer, K., & Wood, S. A. (2016). Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: A case study using freshwater macroinvertebrates. *Molecular Ecology Resources*,*16*(5), 1240–1254. https://doi.org/10.1111/1755-0998.12488

Ehmann, B., van Schaik, C. P., Ashbury, A. M., Mörchen, J., Musdarlia, H., Atmoko, S. U., van Noordwijk, M. A., & Schuppli, C. (2021). Immature wild orangutans acquire relevant ecological knowledge through sex-specific attentional biases during social learning. *PLoS Biology*, *19*(5), 1–21. https://doi.org/10.1371/journal.pbio.3001173

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

Gillet, F., Tiouchichine, M.-L., Galan, M., Blanc, F., Némoz, M., Aulagnier, S., & Michaux, J. R. (2015). A new method to identify the endangered Pyrenean desman (*Galemys pyrenaicus*) and to study its diet, using next generation sequencing from faeces. *Mammalian Biology*, *80*, 505–509. https://doi.org/10.1016/j.mambio.2015.08.002

Gruber, T., Luncz, L., Mörchen, J., Schuppli, C., Kendal, R. L., & Hockings, K. (2019). Cultural change in animals: A flexible behavioural adaptation to human disturbance. *Palgrave Communications*, *5*(64), 1–9. https://doi.org/10.1057/s41599-019-0271-4

Horne, J. S., Garton, E. O., Krone, S. M., & Lewis, J. S. (2007). Analyzing animal movements using Brownian bridges. *Ecology*, *88*(9), 2354–2363. https://doi.org/10.1890/06-0957.1

Jusino, M. A., Banik, M. T., Palmer, J. M., Wray, A. K., Xiao, L., Pelton, E., Barber, J. R., Kawahara, A. Y., Gratton, C., Peery, M. Z., & Lindner, D. L. (2019). An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Molecular Ecology Resources*, *19*, 176–190. https://doi.org/10.1111/1755-0998.12951

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

Kaufhold, S. P., & van Leeuwen, E. J. C. (2019). Why intergroup variation matters for understanding behaviour. *Biology Letters*,*15*(11), 20190695. https://doi.org/10.1098/rsbl.2019.0695

Keith, S. A., & Bull, J. W. (2017). Animal culture impacts species' capacity to realise climate-driven range shifts. *Ecography*,*40*(2), 296–304. https://doi.org/10.1111/ecog.02481

Koops, K., Wrangham, R. W., Cumberlidge, N., Fitzgerald, M. A., van Leeuwen, K. L., Rothman, J. M., & Matsuzawa, T. (2019). Crab-fishing by chimpanzees in the Nimba Mountains, Guinea. *Journal of Human Evolution*, *133*, 230–241. https://doi.org/10.1016/j.jhevol.2019.05.002

Krehenwinkel, H., Kennedy, S., Pekár, S., & Gillespie, R. G. (2017). A cost-efficient and simple protocol to enrich prey DNA from extractions of predatory arthropods for large-scale gut content analysis by Illumina sequencing. *Methods in Ecology and Evolution*, *8*(1), 126–134. https://doi.org/10.1111/2041-210X.12647

Kühl, H. S., Boesch, C., Kulik, L., Haas, F., Arandjelovic, M., Dieguez, P., Bocksberger, G., McElreath, M. B., Agbor, A., Angedakin, S., Ayimisin, E. A., Bailey, E., Barubiyo, D., Bessone, M., Brazzola, G., Chancellor, R., Cohen, H., Coupland, C., Danquah, E., & Kalan, A. K. (2019). Human impact erodes chimpanzee behavioral diversity. *Science*, *363*(6434), 1453–1455. https://doi.org/10.1126/science.aau4532

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, *36*(11), 1–13. https://doi.org/10.18637/jss.v036.i11

Lenth, R. (2019). Emmeans: Estimated marginal means, aka least-squares mean. R package version 1.4.1.

Luncz, L. V., Mundry, R., & Boesch, C. (2012). Evidence for cultural differences between neighboring chimpanzee communities. *Current Biology*, *22*, 922–926. https://doi.org/10.1016/j.cub.2012.03.031

Lyke, M. M., Di Fiore, A., Fierer, N., Madden, A. A., & Lambert, J. E. (2018). Metagenomic analyses reveal previously unrecognized variation in the diets of sympatric Old World monkey species. *PLoS ONE*, *14*, e0218245. https://doi.org/10.1371/journal.pone.0218245

Mallott, E. K., Garber, P. A., & Malhi, R. S. (2017). Integrating feeding behavior, ecological data, and DNA barcoding to identify developmental differences in invertebrate foraging strategies in wild white-faced capuchins (*Cebus capucinus*). *American Journal of Physical Anthropology*, *162*, 241–254. https://doi.org/10.1002/ajpa.23113

Mallott, E. K., Garber, P. A., & Malhi, R. S. (2018). trnL outperforms rbcL as a DNA metabarcoding marker when compared with the observed plant component of the diet of wild white-faced capuchins (*Cebus capucinus*, Primates). *PLoS ONE*, *13*, e0199556. https://doi.org/10.1371/journal.pone.0199556

Mallott, E. K., Malhi, R. S., & Garber, P. A. (2015). Brief communication: High-throughput sequencing of fecal DNA to identify insects consumed by wild Weddell's saddleback tamarins (*Saguinus weddelli*, Cebidae, Primates) in Bolivia. *American Journal of Physical Anthropology*, *156*, 474–481. https://doi.org/10.1002/ajpa.22654

Matthews, J. K., Ridley, A., Kaplin, B. A., & Grueter, C. C. (2020). A comparison of fecal sampling and direct feeding observations for quantifying the diet of a frugivorous primate. *Current Zoology*, *66*, 333–343. https://doi.org/10.1093/CZ/ZOZ058

Mertz, J., Surreault, A., van de Waal, E., & Botting, J. (2019). Primates are living links to our past: The contribution of comparative studies with wild vervet monkeys to the field of social cognition. *Cortex*, *118*, 65–81. https://doi.org/10.1016/j.cortex.2019.03.007

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

Nørgaard, L., Olesen, C. R., Trøjelsgaard, K., Pertoldi, C., Nielsen, J. L., Taberlet, P., Ruiz-Gonzáles, A., De Barba, M., & Iacolina, L. (2021). eDNA metabarcoding for biodiversity assessment, generalist predators as sampling assistants. *Scientific Reports*, *11*, 6820. https://doi.org/10.1038/s41598-021-85488-9

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2014). Vegan: Community ecology package. R package version 2.5–7.

Pansu, J., Guyton, J. A., Potter, A. B., Atkins, J. L., Daskin, J. H., Wursten, B., Kartzinel, T. R., & Pringle, R. M. (2019). Trophic ecology of large herbivores in a reassembling African ecosystem. *Journal of Ecology*, *107*, 1355–1376. https://doi.org/10.1111/1365-2745.13113

Pickett, S. B., Bergey, C. M., & Di Fiore, A. (2012). A metagenomic study of primate insect diet diversity. *American Journal of Primatology*, *74*, 622–631. https://doi.org/10.1002/ajp.22014

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M., & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*, 1246752. https://doi.org/10.1126/science.1246752

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

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

Quéméré, E., Hibert, F., Miquel, C., Lhuillier, E., Rasolondraibe, E., Champeau, J., Rabarivola, C., Nusbaumer, L., Chatelain, C., Gautier, L., Ranirison, P., Crouau-Roy, B., Taberlet, P., & Chikhi, L. (2013). A DNA metabarcoding study of a primate dietary diversity and plasticity across its entire fragmented range. *PLoS ONE*, *8*, e58971. https://doi.org/10.1371/journal.pone.0058971

Rowe, A. K., Donohue, M. E., Clare, E. L., Drinkwater, R., Koenig, A., Ridgway, Z. M., Martin, L. D., Nomenjanahary, E. S., Zakamanana, F., Randriamanandaza, L. J., Rakotonirina, T. E., & Wright, P. C. (2021). Exploratory analysis reveals arthropod consumption in 10 lemur species using DNA metabarcoding. *American Journal of Primatology*, *83*, e23256. https://doi.org/10.1002/ajp.23256

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*, *17*, e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Samuni, L., Wegdell, F., & Surbeck, M. (2020). Behavioural diversity of bonobo prey preference as a potential cultural trait. *eLife*,*9*, 1–10. https://doi.org/10.7554/ELIFE.59191

Santorelli, C. J., Schaffner, C. M., Campbell, C. J., Notman, H., Pavelka, M. S., Weghorst, J. A., & Aureli, F. (2011). Traditions in spider monkeys are biased towards the social domain. *PLoS ONE*, *6*(2), e16863. https://doi.org/10.1371/journal.pone.0016863

Shao, X., Lu, Q., Liu, M., Xiong, M., Bu, H., Wang, D., Liu, S., Zhao, J., Li, S., & Yao, M. (2021). Generalist carnivores can be effective biodiversity samplers of terrestrial vertebrates. *Frontiers in Ecology and the Environment*, *19*, 557–563. https://doi.org/10.1002/fee.2407

Shehzad, W., Riaz, T., Nawaz, M. A., Miquel, C., Poillot, C., Shah, S. A., Pompanon, F., Coissac, E., & Taberlet, P. (2012). Carnivore diet analysis based on next-generation sequencing: Application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, *21*, 1951–1965. https://doi.org/10.1111/j.1365-294X.2011.05424.x

Sih, A. (2013). Understanding variation in behavioural responses to human-induced rapid environmental change: A conceptual overview. *Animal Behaviour*, *85*(5), 1077–1088. https://doi.org/10.1016/j.anbehav.2013.02.017

Struhsaker, T. T. (1967). Ecology of vervet monkeys (*Cercopithecus aethiops*) in the Masai-Amboseli game reserve, Kenya. *Ecology*, *48*, 891–904. https://doi.org/10.2307/1934531

Taberlet, P., Bonin, A., Zinger, L., & Coissac, É. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press. https://doi.org/10.1093/oso/9780198767220.001.0001

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*, 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tan, A., Tan, S. H., Vyas, D., Malaivijitnond, S., & Gumert, M. D. (2015). There is more than one way to crack an oyster: Identifying variation in burmese long-tailed macaque (*Macaca fascicularis aurea*) stone-tool use. *PLoS ONE*, *10*(5), 1–25. https://doi.org/10.1371/journal.pone.0124733

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA—An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Tournier, E., Tournier, V., van de Waal, E., Barrett, A., Brown, L., & Bshary, R. (2014). Differences in diet between six neighbouring groups of vervet monkeys. *Ethology*, *120*, 471–482. https://doi.org/10.1111/eth.12218

van Boekholt, B., van de Waal, E., & Sterck, E. H. M. (2021). Organized to learn: The influence of social structure on social learning opportunities in a group. *iScience*, *24*(2), 102117. https://doi.org/10.1016/j.isci.2021.102117

van de Waal, E. (2018). On the neglected behavioural variation among neighbouring primate groups. *Ethology*, *124*, 845–854. https://doi.org/10.1111/eth.12815
van de Waal, E., Borgeaud, C., & Whiten, A. (2013). Potent social learning and conformity shape a wild primate's foraging decisions. *Science*, *340*(6131), 483–485. https://doi.org/10.1126/science.1232769
van de Waal, E., Bshary, R., & Whiten, A. (2014). Wild vervet monkey infants acquire the food-processing variants of their mothers. *Animal Behaviour*, *90*, 41–45. https://doi.org/10.1016/j.anbehav.2014.01.015
van de Waal, E., Claidière, N., & Whiten, A. (2015). Wild vervet monkeys copy alternative methods for opening an artificial fruit. *Animal Cognition*, *18*, 617–627. https://doi.org/10.1007/s10071-014-0830-4
van de Waal, E., Renevey, N., Favre, C. M., & Bshary, R. (2010). Selective attention to philopatric models causes directed social learning in wild vervet monkeys. *Proceedings of the Royal Society B: Biological Sciences*,*277*, 2105–2111. https://doi.org/10.1098/rspb.2009.2260
van Leeuwen, E. J. C., Cronin, K. A., & Haun, D. B. M. (2018). Population-specific social dynamics in chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), 11393–11400. https://doi.org/10.1073/pnas.1722614115
van Leeuwen, E. J. C., Van Donink, S., Eens, M., & Stevens, J. M. G. (2021). Group-level variation in co-feeding tolerance between two sanctuary-housed communities of chimpanzees (*Pan troglodytes*). *Ethology*, *127*(7), 517–526. https://doi.org/10.1111/eth.13154
Vesterinen, E. J., Puisto, A. I. E., Blomberg, A. S., & Lilley, T. M. (2018). Table for five, please: Dietary partitioning in boreal bats. *Ecology and Evolution*, *8*, 10914–10937. https://doi.org/10.1002/ece3.4559
Voelker, M. R., Schwarz, D., Thomas, A., Nelson, B. W., & Acevedo-Gutiérrez, A. (2020). Large-scale molecular barcoding of prey DNA reveals predictors of intrapopulation feeding diversity in a marine predator. *Ecology and Evolution*,*10*, 9867–9885. https://doi.org/10.1002/ece3.6638
Whiten, A., Goodall, J., Mcgrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. F. G., Wrangham, R. W., & Boesch, C. (1999). Cultures in chimpanzees. *Nature*, *399*, 15–18. https://doi.org/10.1038/21415
Whiten, A., & van de Waal, E. (2018). The pervasive role of social learning in primate lifetime development. *Behavioral Ecology and Sociobiology*, *72*, 80. https://doi.org/10.1007/s00265-018-2489-3
Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., Chariton, A. A., Creer, S., Coissac, E., Deagle, B. E., De Barba, M., Dickie, I. A., Dumbrell, A. J., Ficetola, G. F., Fierer, N., Fumagalli, L., Gilbert, T. P., Jarman, S., Jumpponen, A., & Taberlet, P. (2019). DNA metabarcoding–Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, *28*, 1857–1862. https://doi.org/10.1111/mec.15060

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Table S1. Sample counts per group in summer: number of samples/number of individuals sampled (number of individuals present in group at the start of summer; 15th November 2018). Discrepancies between samples/individuals and individuals/group possible when age categories changed between 15th of November and the time of sampling (e.g. AK adult females and BD one individual sampled as juvenile F and adult F). Pearson's Chi-square tests confirmed the goodness of fit per age/sex category of sampled individuals per group for AK, KB and NH. For BD the null hypothesis of good fit was only confirmed when excluding the infant category. Age categories were defined as follows: infant <1 year, juvenile 1-2 years for females and 1-4 years for males (if not dispersed), adult 3 years for females and 4 years for males if they dispersed, otherwise 5 years.

Table S2. Number of GPS locations sampled per group from the faecal sample spots and through observations, used to calculate the 50 % and 95 % isopleths of the groups' home ranges (*total* rows). The *faecal samples* and *observations* rows indicate the number of these found within their respective isopleths.

Table S3. Comparison of the predicted group (from the random forest model) and the observed group from which the faecal samples were originating for the test dataset (25 % of the observations). Observed samples are in columns and predicted samples in rows. Values on the diagonals represent the true positive of the models. Sensitivity (True positives/(True positives + False negatives)) and Specificity (True negatives/(True negatives + False positives)) are indicated for each group.

Figure S1. Home ranges (50 % and 95 % isopleths) of the four groups based on GPS data of sampling locations of faecal samples and observations. Points indicate the location of the 40 vegetation plots.

Figure S2. Species accumulation curves (SACs) of plant abundance in terms of coverage for the 40 vegetation plots. On the x-axis are the ten vegetation plots per groups and on the y-axis the cumulative number of plant species, grey shading indicates 95 % confidence intervals.

Figure S3. Nonmetric multidimensional scaling (NMDS) with *envfit* function of plant abundance in terms of coverage for the 40 vegetation plots taken in the respective 50 % home ranges of the four groups. Vectors of plants are shown for those with $p<0.005$. *PERMANOVA* indicates no significant difference in plant coverage between groups' territories (pseudo$F40 = 1.44$, $R^2 = 0.11$, $P = 0.12$).

Figure S4. The plot shows the plant species that were selected by the random forest algorithms implemented in the Boruta R package as relevant features to explain dietary variation between groups. The higher the importance, the higher the group specificity of the corresponding species. Blue shows the minimum, average and maximum importance scores obtained by chance after 2000 random row permutations. Species in red were below the maximum threshold and considered not specific to any of the groups. *E. undulata* in yellow was very close to the maximum threshold and also not kept for further analyses. For species in green the group specificity was higher than that obtained by chance. Species above the threshold were corrected according to their respective abundance in the different groups' home ranges when available.

Figure S5. Boxplots displaying the dispersion from the centroids for each group, for (a) all individuals (average distance to median: AK = 0.25, BD = 0.43, KB = 0.19, NH = 0.35; $p<0.001$) and (b) only adult females (average distance to median: AK = 0.23, BD = 0.38, KB = 0.21, NH = 0.33; $p<0.005$).

Figure S6. Nonmetric multidimensional scaling (NMDS), based on relative read abundances (RRA) of consumed plants detected in faecal samples aggregated per monkey in summer for all groups (Ankhase, AK; Baie Dankie, BD; Noha, NH and Kubu, KB) with variable sex/age with three factor levels. The colours represent female adults (red), male adults (green) and juveniles (blue).

Figure S7. Nonmetric multidimensional scaling (NMDS), based on relative read abundances (RRA) of consumed plants detected in faecal samples aggregated per monkey in summer for all groups (Ankhase, AK; Baie Dankie, BD; Noha, NH and Kubu, KB) with variable sex/age with four factor levels. The colours represent female adults (red), male adults (green), female juveniles (blue) and male juveniles (violet).

How to cite this article: Schneider, J., Brun, L., Taberlet, P., Fumagalli, L., & van de Waal, E. (2023). Molecular assessment of dietary variation in neighbouring primate groups. *Methods in Ecology and Evolution*, *00*, 1–12. https://doi.org/10.1111/2041-210X.14078

Chapter 3: Intergroup variation in wild adult female vervet monkeys

Judith Schneider[1], Molly Baur[1], Loïc Brun[2,3], Erica van de Waal[2,3,4,‡], Luca Fumagalli[1,5,‡]

[‡] joint last authors

[1]Laboratory for Conservation Biology, Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland.

[2]Inkawu Vervet Project, Mawana Game Reserve, Swart Mfolozi, KwaZulu Natal, South Africa.

[3]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland.

[4]Centre for Functional Biodiversity, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

[5]Swiss Human Institute of Forensic Taphonomy, University Centre of Legal Medicine Lausanne-Geneva, Lausanne University Hospital and University of Lausanne, 1000 Lausanne, Switzerland.

Chapter status: unpublished

Author contribution
EW and LF designed the study and supervised all the analyses. JS, LB and MB performed the laboratory work. JS conducted bioinformatics, data analyses and wrote the first draft of the manuscript, with input from all other authors.

Abstract

1. In order to inform conservation programmes, intergroup comparisons are useful to inquire about variable adaptation potentials in terms of behaviour.
2. To identify intergroup variation (IGV), we analysed DNA metabarcoding data for plant components in the diet of adult female vervet monkeys (*Chlorocebus pygerythrus*). Each of the 1226 faecal samples was attributed by observation to 87 individuals with known life history data belonging to five neighbouring groups in the wild.
3. The results did not confirm our hypothesis of distinct intergroup variation between members of the philopatric sex, here limited to adult females. The effect of season was always greater than the effect of group. However, these analyses open up the scope for further research questions to be investigated.

Introduction

Climate change and progressive habitat disturbance and fragmentation threaten the long-term survival of many species. Nonhuman primates are particularly vulnerable to anthropogenic activities; about 60 % of primate species risk extinction and 75 % face population declines[100]. The value of primates and their interest for biodiversity conservation actions are uncounted: ecological (often key role in ecosystems), social, cultural and economic (important role for human traditional knowledge, religion, local economies and tourism), and scientific (closest living relatives to humans, model species to understand evolutionary processes)[100]. In addition, primates represent the group where cultural traits have been expressed the most (admittedly also because of the increased interest in the closest living relatives of humans). But not the only one, cultural behaviour was observed in notably avian and other mammalian species[101]. This is at the same time an adaptive potential for long-lived species when facing rapidly changing conditions and represents also valuable cultural traits like a species cultural heritage[77]. In order to decide on conservation priorities, populations are often defined in terms of their genetic diversity. Analogous to these genetic assessments, behavioural diversity needs to be included in conservation programmes[76,78]. Group-level assessments can avoid species-wide generalisations[92,97] and provide data on a finer scale. In this sense, to discern groups of specific conservation concern, it is worthwhile to assess intergroup variation in terms of behaviour.

Group-level assessments allow at the same time to discern the drivers of behavioural diversity as it helps in certain cases to disentangle ecological, genetic and social factors. As a result, intergroup comparisons, in the wild in particular, are useful to assess the influence of social factors on behaviour[80]. For example, in a study system with similar environment and constant

gene flow between groups due to male dispersal[63], variation in behaviour could be ascribed to social factors according to the exclusion principle[67,81]. Environmental DNA (eDNA) methods sample an organism's environment in order to analyse the genetic fragments left in this environment[1]. It seems an interesting complementary method to study behavioural diversity in the wild, which on top of that, is ideally based on the same samples as those to assess genetic diversity[68]. We therefore assessed the plant-based diet of vervet monkeys (*Chlorocebus pygerythrus*) relying on DNA metabarcoding[10] of faecal samples at the group-level to assess if there is intergroup variation (IGV) in foraging behaviour that cannot be explained by environmental or genetic factors and might thus indicate effects of social factors.

Vervet monkeys live in social groups and the group affiliation is a strong social factor, and they likely acquire behaviour through social learning with a possible conformity effect on behaviour[95]. The social structure in philopatric species in general favours distinct group identities[97]. Female vervets, the philopatric sex, have been shown experimentally to be the preferred role models[94,96] and male vervets, the dispersing sex, showed greater behavioural flexibility[98]. In Chapter 2 of this manuscript, we showed, with a limited sample number and by focusing on just one season, that adult females represented well a group's diet. Therefore, here we used a larger sample size and focused on the samples of adult females to discern possible group-specific patterns and intergroup variation in terms of foraging behaviour. For this we analysed 1226 samples collected over four years and all seasons assigned to 87 adult females of five groups.

Methods

Field work and samples

These faecal samples have been collected at the iNkawu Vervet Project (IVP), Mawana game reserve, South Africa. Samples have been attributed to individuals by observation; recognisable for the trained field team by specific body and face characteristics (e.g. facial features, scars and fur colours). All individuals are named at birth and the continuously collected life history data (e.g. sex, age category, group affiliation) can be linked to each sample. Here sequences amplified with primers targeting the plant part of the diet of 1226 faecal samples from 87 adult female vervet monkeys of five groups were analysed, including the 126 samples of 50 adult females assessed in Schneider et al. (2023; i.e. Chapter 2 of the present manuscript): Ankhase (AK, 273 samples/12 individuals), Baie Dankie (BD, 383 samples/29 individuals), Kubu (KB, 111 samples/11 individuals), Lemon Tree (LT, 183 samples/14 individuals) and Noha (NH, 276 samples/21 individuals; see Table S1 and Fig. S1, Supplementary data). Adult females include all females aged three years and older. Faecal samples were collected between January 2018 and September 2022. The field sampling protocol was the same as described in Brun et al. (2022; i.e. Chapter 1 of the present manuscript).

Laboratory and bioinformatics

The DNA extraction, PCR amplification with primers Sper01[1], and sequencing of the faecal samples was conducted as in Brun et al. (2022; i.e. Chapter 1 of the present manuscript), with 150 paired-end sequencing on the Illumina Miniseq Sequencing System (Illumina, San Diego, CA, USA). All raw sequences were analysed following the same bioinformatic pipeline as

described in[70], except for the *obiclean* command (-r option 0.25 instead of 1), and a threshold of 500 reads to omit empty PCR reactions. For the taxonomic assignment, all sequences were dereplicated and then manually blasted against GenBank with a minimum sequence similarity threshold of 97 %. The sequences of the previous local database[70] had been included in GenBank by the time. Priority was given to those if several hits had identical identity percentages and geographical distribution was ambivalent. Bacterial sequences were omitted as well as only order level plant assignments. In cases with unclear plant assignment, only the family level was retained. This procedure ensured a better controlled taxonomic assignment and the most possible resolution and information for the dataset. The taxonomic information was merged back with the original dataset and assignations to the same taxonomic units aggregated thereafter. The number of OTUs was normally distributed over samples (Fig. S2, Supplementary data). Following final data cleaning and aggregating by taxonomic assignment, 110 OTUs remained, of which 27 assigned to family level, 41 to genus and 33 to species and 7 to species complexes (Figs. S3 and S4, Supplementary data). The two mainly consumed species, *Berchemia zeyheri* and *Ziziphus mucronata* belong both to the plant family Rhamnaceae, therefore we chose to include genus and species level assignments. And family level assignments, representing ~10% of total reads, were finally not considered for statistical analyses in order not to increase variation between samples by bioinformatic choices. Relative read abundance (RRA; the number of a specific sequence divided by the total number of reads within a sample) were used for analyses if not stated otherwise.

Data analyses

All analyses were done with R software (Version 4.0.2). We defined season as follows, with the middle of a month as the seasonal delimitation[102]: August – November (spring), November – March (summer), March – May (autumn), May – August (winter). The wet season comprises summer and autumn and the dry season winter and spring. As such we considered seasonality in terms of seasons but not possible fluctuations between years since we expected no significant annual variation in vegetation (see[70]).

The variation in composition of consumed plants between groups was assessed and represented graphically by non-metrical dissimilarity scaling (NMDS) using RRA data and Bray-Curtis dissimilarity matrixes with one data point per sample. In addition, we aggregated data per adult female as the mean of RRA, to decrease effects of unbiased sampling, and repeated the NMDS. Data was subset by season and investigated for group effects on dietary variation. In addition, the data was aggregated per adult female after sub-setting the data by season to account for seasonal effects and increase the resolution. Pairwise comparisons between groups were deemed appropriate to assess variation between pairs of groups specifically. The homogeneity of group dispersions was tested with the *betadisper* function in the VEGAN package[103]. If these were homogenous, we performed a permutational multivariate analysis of variance (PERMANOVA with 9999 permutations and Bonferroni correction) with Bray–Curtis dissimilarity matrices of RRA data as dependent variable. To test the effect of the group on diet composition the variable group was included in the model as well as sample nested in individual monkeys to account for pseudo-replication and, if applicable, season. Heterogenous dispersions, were taken into account by using beta regression models with the BETAREG[104] package. Thus differences in proportions of dietary plant species per group were assessed and

the *joint-tests* function in the EMMEANS package[105] indicated the main effects of the models. To investigate variation between adult females at the individual level and to follow up on the analyses in Chapter 2, intragroup variation was assessed for summer.

In addition, individual specialisation (V) was calculated, frequency-based, per individual and per sample with the RInSp package[106]. The proportional similarity index PS_i is defined here as dietary overlap between an individual/a sample and its group and estimates the diet similarity in terms of consumed plants between a sample and the average diet of its group:

$$PS_i = 1 - 0.5 \sum_j | p_{ij} - q_j |$$

with p_{ij} being the frequency of category j in the individual i's diet, and q_j the frequency of category j in the group[106]. The index V

$$V = 1 - PS_i$$

goes from 0 to 1, with 0 defining more similar and 1 more dissimilar diets. First, V was calculated per individual in relation to the respective group. A single value was retained here, the mean per individual, and the inter-individual variation plotted in relation to the group's mean (coloured by group). Second, V was calculated for each sample and the inter-sample variation plotted per group (coloured by individual), again in relation to the group's mean.

Results

Figure 1 illustrates the consumed plant items composition per sample and individual and its respective group affiliation across all seasons. In Fig. 1A all samples are considered independently not taking into account pseudo-replication. Intragroup dispersions were

heterogenous and IGV was not observed between the five studied groups using beta regression models (F_{ratio} = 1.09, p-value = 0.3614). When data was aggregated per individual (Fig. 1B), a significant effect of the variable group was observed by PERMANOVA (R2 = 0.12, p-value < 0.0001).

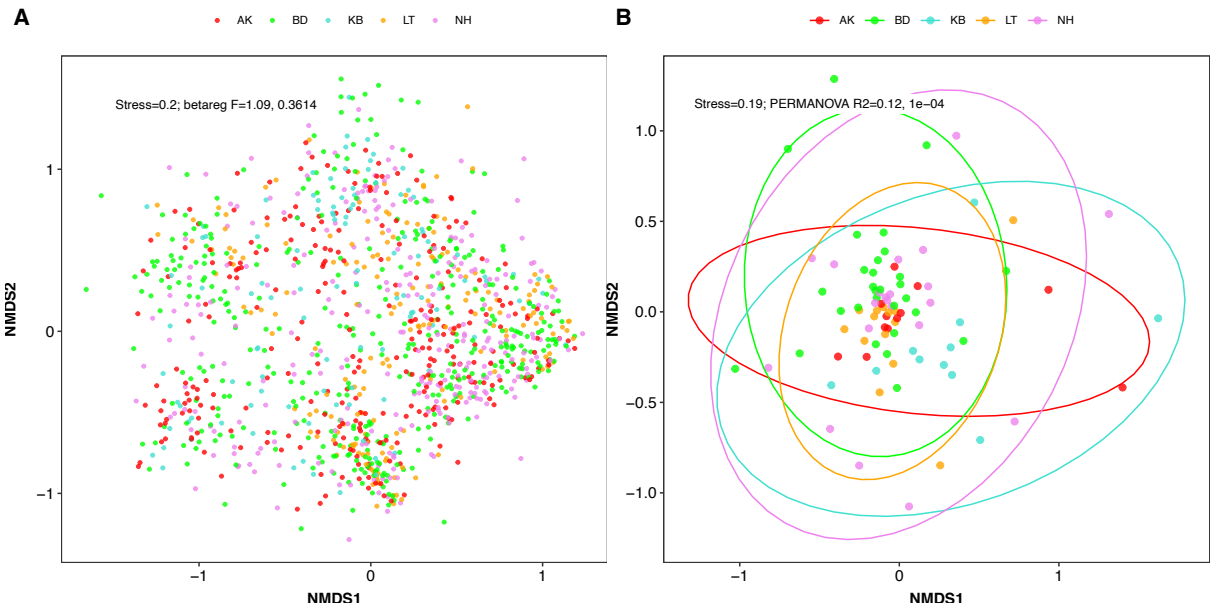


Figure 1. Non-metrical dissimilarity scaling (NMDS) of all groups over all seasons using Bray-Curtis distances for A) non-aggregated relative read abundance (RRA) data, i.e. showing 1226 data points and B) RRA data aggregated per monkey, i.e. showing 87 data points.

In order to minimise seasonal effects, the data was subsequently subset per season. For non-aggregated data, the betadisper function indicated heterogenous dispersions between groups for all seasons. For none of the four seasons, we observed a significant effect of the variable group when the five groups were considered together (Fig. 2A-D). Testing the groups pairwise all seasons combined, we observed significant effects of the factor group for all group comparisons (Table S2, Supplementary data). However, the seasonal effect was always larger and no distinct clustering was observed in the graphical representation by NMDS (Fig. S5, Supplementary data).

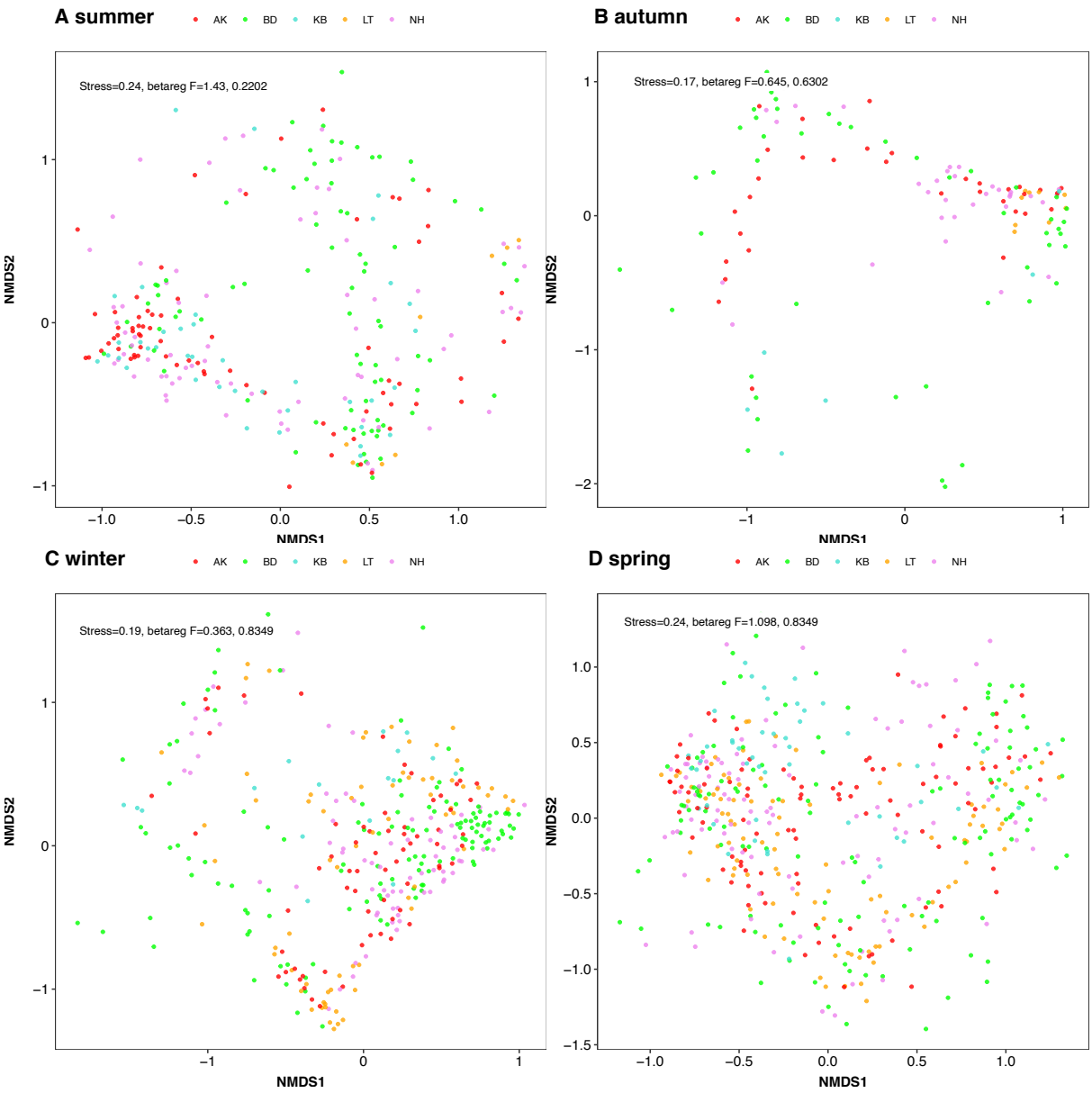


Figure 2. Non-metrical dissimilarity scaling (NMDS) using Bray-Curtis distances of relative read abundance (RRA) data, subset by season. For A. summer, B. autumn, C. winter and D. spring.

For data aggregated per individual, group dispersions were homogenous for summer and winter and the group variable was significant, measured by PERMANOVA (R2 = 0.20, p-value < 0.0001; Figs. S6A, S6C, Supplementary data). No significant effect of the group variable was found in autumn and spring using beta regression models (autumn F_{ratio} = 1.82, p-value = 0.1226; spring F_{ratio} = 2.1, p-value = 0.0734; Figs. S6B, S6D, Supplementary data).

As a follow up analysis of Chapter 2, we focused on summer for intragroup analyses and pairwise tests. Figure 3A illustrates for Ankhase (AK) the clustering per season on the intragroup level for all samples independently, however, the effect is not significant. Figure 3B shows inter-individual variation for the season where we expect the most distinct patterns[68]. For AK, the sample distribution between individuals was relatively even. PERMANOVA was not significant (R2 = 0.10666, P = 0.3857) for sample with the individual monkey as nested factor taking into account pseudo-replication, suggesting relatively homogenous intragroup feeding behaviour for adult females. However, without nesting samples in individuals, there was significant variation between samples, indicating individual effects.

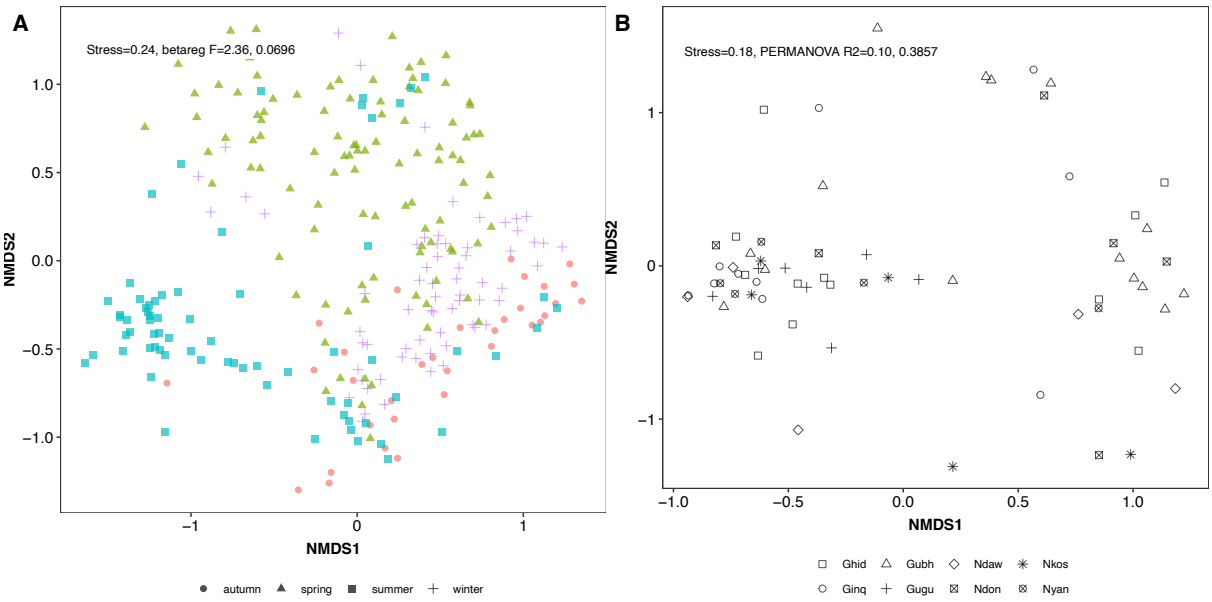


Figure 3: Non-metrical dissimilarity scaling (NMDS) using Bray-Curtis distances of RRA data. A. Ankhase (AK) across seasons, different shapes and colours represent different seasons (273 samples of 12 individuals). B. Ankhase in summer, included are all individuals with ≥ 5 samples, different shapes represent different individuals (65 samples of 8 individuals).

The same pattern has been observed for BD, KB, NH and LT, although for the latter group there were few samples in summer (Fig. S7, Supplementary data). The observed homogenous

feeding behaviour at the intragroup level between adult females allowed to investigate if intergroup variation was greater than intragroup variation. Figure 4 illustrates the results of pairwise testing focused on summer with violet arrows indicating observed IGV and orange arrows group pairs with non-significant effect of group (Table S3, Supplementary data). However, these results have to be taken with caution due to strong sampling biases in particular for Lemon Tree.

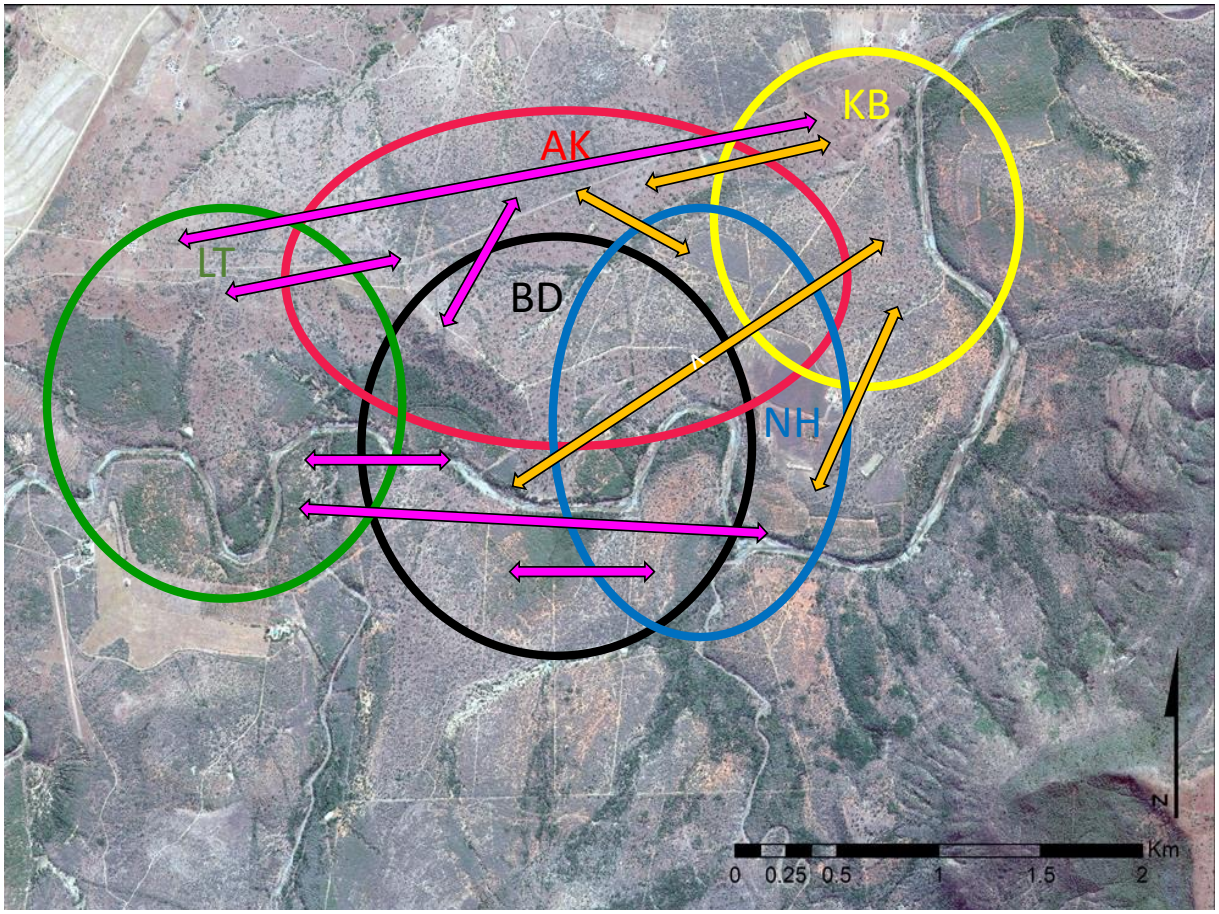


*Figure 4: Aerial view of the average home ranges of the five studied groups at iNkawu Vervet Project (IVP). Ankhase (AK), Baie Dankie (BD), Kubu (KB), Lemon Tree (LT) and Noha (NH). The arrows show if pairwise comparisons were significant for the group variable in summer. Violet arrows indicate significant effect of the group variable, orange arrows indicate not significant effect of the group variable (**Table S3, Supplementary data**).*

Figure 5 shows the inter-sample variation (V) in relation to the respective group, each sample was considered independently despite possible effects of pseudo-replicates. Fig. 5A shows the

variation score of all samples coloured by social group. IGV in terms of differences between groups' means of V were not observed. LT disposes the lowest mean and BD the highest, which might indicate higher and lower group cohesion, respectively, in terms of foraging behaviour. BD and NH seem to show more diverse foraging tendencies (more samples with relative high V) but different sample numbers affect these results. To inquire about individual niches compared to the group mean, Fig. 5B-F show these variation scores per group with samples coloured by individuals. There seems to be more variation between individuals, i.e. that certain individuals forage differently from the rest of the group (V closer to 1), although sampling biases skew the representation. In addition, seasonal variation is not considered here, although overall specialisation scores was generally higher in the wet season (Figs. S8 and S9, Supplementary data).

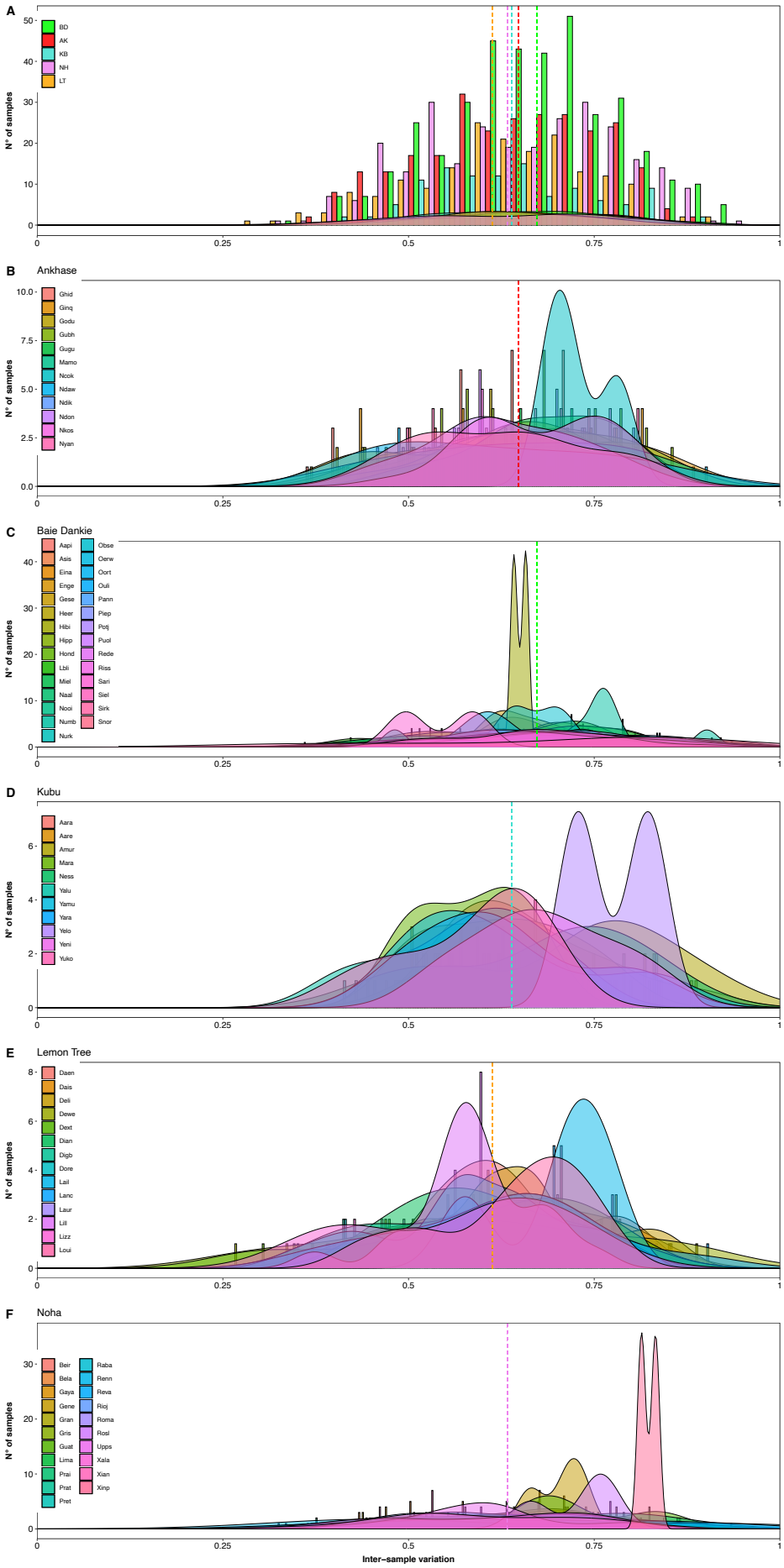

A
BD
AK
KB
NH
LT
N° of samples
B
Ankhase
Ghid
Ginq
Godu
Gubh
Gugu
Mamo
Ncok
Ndaw
Ndik
Ndon
Nkos
Nyan
C
Baie Dankie
Aapi
Asis
Eina
Enge
Gese
Heer
Hibi
Hipp
Hond
Lbli
Miel
Naal
Nooi
Numb
Nurk
Obse
Oerw
Oort
Ouli
Pann
Piep
Potj
Puol
Rede
Riss
Sari
Siel
Sirk
Snor
D
Kubu
Aara
Aare
Amur
Mara
Ness
Yalu
Yamu
Yara
Yelo
Yeni
Yuko
E
Lemon Tree
Daen
Dais
Deli
Dewe
Dext
Dian
Digb
Dore
Lail
Lanc
Laur
Lill
Lizz
Loui
F
Noha
Beir
Bela
Gaya
Gene
Gran
Gris
Guat
Lima
Prai
Prat
Pret
Raba
Renn
Reva
Rioj
Roma
Rosl
Upps
Xala
Xian
Xinp
Inter−sample variation

Figure 5: Inter-sample variation (V) in relation to social group. The vertical lines represent the groups' means. A) Inter-sample variation coloured by groups Ankhase (AK), Baie Dankie (BD), Kubu (KB), Lemon Tree (LT) and Noha (NH). B-F) Inter-sample variation in relation to the respective group coloured by individual. B) Ankhase, C) Baie Dankie, D) Kubu, E) Lemon Tree and F) Noha.

Discussion

A previous assessment of variation in foraging behaviour in summer between four of the neighbouring groups living in IVP (AK, BD, KB and NH) indicated that the social factor group may have an effect on feeding habits and that adult females well represent a group's diet[68]. Following these preliminary analyses on limited sample numbers, the aim of the present study was to increase sample numbers and to assess IGV across seasons for adult females only, including with LT an additional group. However, while we observed in some cases significant results for the variable group, in particular in pairwise comparisons, there was less variation between the groups than expected. Overall, based on DNA metabarcoding of faecal samples targeting the plant part of the diet, we observed relatively homogenous foraging behaviour between the groups and individuals. Our results confirmed the strong effect of season, that was already observed in[70] and that was larger than any group effect. Subsetting the data per season allowed a more detailed assessment of the latter effect, however, it was only statistically significant for data aggregated per individual or for certain pairwise group comparisons. It should be noted here that the statistical tests were not always the same depending on the distribution of dispersions between samples and the centroid. While of course heterogenous distribution of dispersions are in principle not favouring distinctions between groups, it may also be that one test is more sensitive to variation than the other one, due to logit transformation of the data in the regression models (i.e. PERMANOVA > regression models).

The relatively small effect of the variable group between female adults in this population does not allow to reject the assumption of intergroup variation. However, it could indicate that our hypothesis of more distinct IGV for the philopatric sex was disconfirmed and that on the contrary the diet of adult females is homogenous across groups. This could be explained by similar nutrient needs of this social category, in particular due to pregnancy and nursing. By limiting the analyses to adult females, we controlled for variation due to sex and partly age. Another explanation could be that the plant part of the diet of all vervets sharing the same environment is more or less the same, responding in the first place to seasonal fluctuation and plant availability. Vervet monkeys are known to be opportunistic feeders.

Both hypotheses could be assessed prospectively in our project. To inquire about different feeding behaviour between different social categories, the focus on a single group and inclusion of all individuals could be fruitful. There is a complementary dataset available that comprises plant sequences of individuals of all age and sex classes of the largest group, BD (N = 950, including adult females). Based on these data, it is possible to analyse at the intragroup level the effect of more social factors, e.g. sex, age and social networks and investigate possibly about sex effects already in juveniles. To assess the second hypothesis, that plant items might not be the most relevant part of the diet to trace IGV, feeding on arthropod or maybe also vertebrates could be assessed. The sequencing of the above-mentioned BD samples targeting arthropods is in preparation as part of a master project. These parts of the diet may reveal more pronounced differences, either between social categories depending on sex and age or between individuals or matrilines, depending on rank. For the prospective analyses, we also need to include more variables and do these also per plant species. The group is a strong social structure in vervet monkeys. But it is a dynamic structure that needs

to be conceptualised as such; we need to include group size and group composition as factors (e.g. number of individuals per age and sex category and consequent ratios). Regarding single plant species, we have seen that seasons were often dominated by foraging of single plant species, e.g. *Berchemia zeyeri* in summer and *Ziziphus mucronata* in autumn and winter. Single plant assessments for the most consumed plants might show more distinct feeding patterns between groups or again also between social categories.

Due to the opportunistic way of sampling, there is a sampling bias in the study. Given the 1226 samples of 87 individuals, ideally each female adult should have been sampled 14.09 times equally distributed over all seasons. By looking at Table S1 and Fig. S1 this was not the case here. For example, LT is less habituated than other groups and the overall seasonal sampling bias (more samples in the dry season) may be caused by the fact that more field experiment are undertaken in winter and spring and thus more people are in the field and/or because the scarcer vegetation allows to spot the faecal samples more easily. Biases in the sampling entail inevitably biases in the results, as may be the case for pairwise comparisons in summer. A key question is how to deal with these skewed sample numbers and how to account for replicates per individual. Aggregated RRA data per individual or per season and individual resulted in more distinct differences between the groups for the presented data than per sample assessments (Figs. 1, 2 and S6, Supplementary data). A possible explanation is that aggregation of the data allowed to level down sampling biases, resulting in more equal data points per individual, another one is that dispersions were homogenous due to aggregation and PERMANOVA could be used (see above). Also in the intragroup analyses in summer, we observed different results whether samples were nested in individuals or not. While that might indicate effects of the individual, we have also seen that intraindividual variation is high,

i.e. samples of the same individual show different patterns. The question remains open if the group affiliation is more important than individual preferences or feeding habits. In any case, social structure plays a key role in this species; therefore we could also question the independence of the data points in terms of matrilines (mothers are indicated in Table S1, Supplementary data). Other levels of analyses may be of interest below the group-level, such as matrilines and dyads or individuals. We could test for a social learning effect on dietary choice from mothers to offspring, or among samples of other members of the same matriline provided that there is enough variation for plant foraging, otherwise this could be undertaken using arthropod sequence data for BD. Combining a social network approach with dietary sequence data could highlight whether physical proximity equals similar diets. Despite the mixed results of the present chapter, we think that eDNA-based analyses can complement other study methods in a fruitful way and that generally the study of behavioural plasticity at the group level can add much to the conservation of cultural diversity of species.

Acknowledgements

We would like to thank the van der Walt family for permission to conduct the study on their land. We are very grateful to the whole IVP field team for the tremendous sampling effort and in particular to Loïc Brun. We thank Eduard Mas Carrió for his precious help. The observational research was approved by the relevant local authority “Ezemvelo KZN Wildlife” in South Africa. The field costs during data and sample collection were funded by grants to EW from the Branco Weiss Fellowship-Society in Science and the Swiss National Science Foundation (grant number PP00P3_170624), and the grant `ProFemmes’ of the Faculty of Biology and Medicine, University of Lausanne. At the time of writing, EW was supported by the European Research

Council under the European Union's Horizon 2020 research and innovation programme for the ERC 'KNOWLEDGE MOVES' starting grant (grant agreement No. 949379). This project was supported by a Swiss National Science Foundation grant to LF (nr. 310030_192512).

General discussion

Our planet faces rapid changes, entailing population declines and the challenge to adapt on evolutionary short timeframes. In order to direct and prioritise conservation measures, the most vulnerable groups, habitats or zones need to be defined. Conservation management already uses measurements of genetic diversity ("genetic hotspots") to target the most threatened populations and areas. The idea to define such targets based on behavioural or even cultural diversity is relatively recent[74–76,107], as is the idea of animal culture (and also still debated). Definitions of animal culture include the persistence of traditions and knowledge over time, that is, transmitted by social learning, see for example[108,109]. While social learning of non-group living or solitary animals is possible[110], culture seems as such reserved to social, group-living species. To study culture in the wild is a challenge that seems beyond the scope of eDNA-based methods alone; a focus on the influence of social factors, however, may highlight the causes of behavioural diversity. Sociality is in this sense a prerequisite for culture (and less debatable) and gained recently increasing interest for research.

Although it is also generally challenging to study sociality in the wild, methodological progress and the emergence of alternative methods contributed valuable insights into the behavioural ecology of species and sometimes cultural evolution[111]. In particular, technical progress has allowed the development of a wider variety of methods to study behaviour in the wild. There are now valid alternatives at choice to direct observational monitoring, which is extremely time-consuming, complicated in case of rare and elusive animals or even dangerous or unethical in some cases. Assessments of whole ecosystems as well as behavioural studies on broad spatial and temporal scales represent great challenges for traditional observational methods. Among these techniques are e.g. camera-traps that represent certainly advantages

for the mere detection of organisms in difficult accessible field[16,112] and belong to the lesser invasive methods adapted for monitoring of unhabituated organisms. For the study of sociality, they can provide valuable information and avoid habituation effects to human observers but are also somehow limited geographically and temporally as well as in terms of flexibility[37]. The use of drones is another example that may help to study sociality efficiently albeit likely inflicting stress on the studied animals[37].

Like the above-mentioned approaches, eDNA-based studies may add valuable information and opportunities to study sociality in the wild. The term eDNA refers to the genetic material that organisms leave in their respective environment and which can be recovered by collecting environmental samples, the DNA subsequently extracted and analysed[2]. The sampling of environmental samples (e.g. soil, water or faeces) is standardisable[113] and time-effective compared to traditional methods as observation. Technical progresses and ever-decreasing costs, particularly in next-generation sequencing technologies, have made eDNA a valuable study tool in many research fields, among others ecology, invasive species management and biomonitoring[1]. Starting from a single environmental sample, eDNA approaches open a wealth of information regarding species and populations, communities, and complete ecosystems and biodiversity. In addition, eDNA-based approaches are considered non-invasive, since organisms can be detected without the need to observe or disturb them in their respective environments, which opens exciting opportunities for behavioural ecologists. In the best case scenario, the same samples can be used to assess genetic and behavioural diversity[68]. eDNA analyses allow not only to study elusive species, among others, but also elusive behaviours. Indeed, it enables sometimes to complement observational data as in the case of insect consumption that is a feeding behaviour usually difficult to observe with high taxonomic

resolution. Our comparison between DNA metabarcoding and traditional focal data sampling has illustrated inter-method congruency and in addition higher taxonomic resolution and coverage for DNA metabarcoding of plant and arthropod diet items than behavioural observations[70].

The iNkawu Vervet Project (IVP) in Mawana Game Reserve, South Africa, represents a very special research context to study sociality. Genetics, ecology and sociality are considered the main explanations for behaviour[81]. In the present study population, there is constant gene flow between the groups due to male dispersal[63], limiting the influence of genetic explanations. Although we must admit that our vegetation assessment was temporally constrained, the ecology seems similar for all groups living in close proximity. According to the method of exclusion[67,81], behavioural variation might therefore be explained by social factors. A major advantage of the study system is the assignment of faecal samples to individuals as well as the available long-term (over ten years) life history data. This allows to directly link social data to dietary information based on sequence data. The different assessments based on the collected faecal samples provide for interesting analyses. In Chapter 1, we described the diet of vervet monkeys studied at IVP. For this we compared focal and DNA metabarcoding data for the plant and insect part of the diet. A strong seasonal effect was documented; with the two dominantly consumed plant species *Berchemia zeyheri* in summer and *Ziziphus mucronata* in winter. Shannon diversity was the highest in summer, therefore the analysis of intergroup variation (IGV) in Chapter 2 focused on the samples collected during this season because selective behaviour were more likely[68]. We observed significant IGV between the four studied groups, which was more accentuated for certain pairwise comparisons of adult females than for all individuals combined. We postulated, based on the social organisation of

the species, that the philopatric sex represents more homogenously a group's diet. Therefore, we focused on adult females and an enlarged the sample size and temporal scope in Chapter 3. However, the analyses in this chapter did not confirm the hypothesis of a more distinct feeding behaviour of the member of the philopatric sex, here limited to adult females, and instead suggested relatively homogenous foraging behaviour of adult females across groups. We have seen with our data that it is complicated to measure the influence of social factors on dietary variation for opportunistic and omnivorous feeders. We may have the ideal study system, but not imperatively the ideal diet to study. Our endeavour to link social factors to dietary variation may be more fruitful for other species foraging more diversely and/or where simple presence/absence data is sufficient for clear findings. For example, the detection of nuts and crabs in the diet of chimpanzees (*Pan troglodytes verus*) as an indicator of different group traditions may be such a case[114]. For the vervets in our study groups, an assessment of dietary items that are rarely consumed, e.g. vertebrates, could also yield interesting results.

In addition, many factors influence the results of DNA metabarcoding and we have discussed the limits of DNA metabarcoding at length above. Both bioinformatic choices and data treatment affect the results[43,115]. RRA data is the most adapted to reflect abundances and mirror true biomass[42]. However, RRA is the most sensitive to biases due to different primer affinities of organisms present in eDNA pools[44,116]. Among potential biases, we highlighted the PCR-induced ones also since these are considered the most important. In order to overcome those, we have designed the bait set for the TSC approach (see Project+). TSC follows a different logic to achieve target enrichment than PCR amplification, since the increase in target sequences is achieved by reducing non-target sequences[61]. The main challenge for the design of bait sets targeting broader taxonomic groups is to find the best compromise

between specificity and flexibility. We are confident that this approach has the potential to yield more realistic biomass estimates and to provide a reliable tool for quantitative assessments using eDNA.

While DNA metabarcoding implies the above-mentioned limitations, it is nevertheless a valuable tool for assessments of ecosystems, diets or population monitoring. Behavioural ecologists can rely on it for many different research questions, among which the study of sociality. There are further research questions beyond the group level and specifically targeting the influence of social factors on behaviour. This always leads implicitly to the question where behavioural plasticity ends and variation begins. Unlike the transmission of genetic information, social transmission of information may follow other than vertical pathway, i.e. from parents to offspring. It can be reverse-vertically from offspring to parents, horizontally among one generation (related or unrelated) or obliquely across generations and unrelated individuals[72]. Although social learning can reduce the costs of asocial learning, it is not *per se* a warranty to increase fitness[110]. To be optimal, and hence adaptive, it is hypothesised that individuals are guided by `social learning strategies' (SLSs) that influence when and which social information to copy and from whom[117]. SLSs, also named `transmission biases', can be content or context dependent; they can supposedly change over time, be combined and employed simultaneously (see for review[110]). Context dependent SLSs are more relevant in our study system; e.g. to study if the sex of an individual can influence the choice of the role model (in terms of sex of the role model) since copied behaviour may vary depending on sex and therefore differ in usefulness depending on sex[118]. Such sex effect on dietary choices has been observed already for juvenile orangutans[82]. In vervets, a female-model transmission bias has been documented in general[94,96], however, also a higher

behavioural flexibility of the dispersing sex, i.e. males[95,98]. Additional sequence data is available, in particular for the largest group, BD, with the most intragroup variation observed for the samples tested restricted to one summer[68]. Plant diet data is already available for additional 474 samples of males, juveniles and infants (total number of analysed samples for BD 950). Based on these samples, a variety of research questions could be assessed, as a model-based bias or sex-effect focusing on juveniles. Analogous to the sex effect in juveniles, studying a general sex effect is worthwhile. Adult males can disperse several times in their lifetime; in addition to (for now hypothetic) sex-specific knowledge acquired through model-biased learning when juvenile, these males can potentially learn different behaviour in each group and may hence dispose a greater behavioural repertoire than females that remain in their natal group. Faecal samples of adult males would consequently dispose a more diverse diet than those assigned to females and juveniles. In addition, studying behavioural plasticity of dispersing males can highlight the interplay of foraging behaviour learnt socially in the natal group and potential conformist food choice to their new group as highlighted experimentally[95]. Another hypothesis is to consider immigrant males as source of information and innovation. Adaptive behavioural responses to fast changing environments, information, food sources, new predators, etc. can be critical for the survival of long-lived species[119]. Innovation is therefore evolutionary important and in particular for social groups. Dongre et al. (2021) studied experimentally innovation and the spread of novel information in IVP by exposing five groups to a novel food (i.e. peanuts). In two groups the innovators were immigrant males, in the third an infant, and in the two remaining groups, the behaviour was imported by immigrant males. It was also observed that the new behaviour was copied faster from immigrant males than from the infant[119]. Adult males seem to play a critical role for innovation and spread of information over group boundaries through dispersal. While the

study of innovation in social groups using social network approaches is promising to learn more about social learning strategies[120], it represents a great challenge in the wild. eDNA-based techniques have the advantage to `show' also non-observed behaviour and can increase spatio-temporally the study coverage, but sampling strategies need to be adapted (e.g. focus on small numbers of individuals but increase samples numbers per individual). Once more, eDNA can be a valuable complement to other methods and thus interdisciplinary approaches may lead to great study outcomes.

Social learning is not the only factor to consider; foraging strategies can be influenced by a number of other factors, some of which extrinsic (food abundance, diversity and distribution, seasonality or predation risk but also conspecifics) and others intrinsic to the forager (body mass and social status in the group, i.e. rank, or the developmental stage). While we tried to control for some of these like plant abundances and seasonality, others can be confounding and should be assessed specifically. For example, there are life history traits that likely influence foraging strategies. The reproductive status of females in particular does infer specific needs in nutrients since pregnancy and lactating can increase energy demands of e.g. nitrogen and calcium. Vervets are seasonal breeders[121] and females at IVP give birth roughly from October to December which coincides with the beginning of summer when food is diverse and abundant. Since insects contain high levels of proteins[122], it could be assessed if pregnant or lactating females consume significantly more insects relative to males or female juveniles. With DNA metabarcoding here again a measure of absolute biomass is not possible to verify if these females increase their food intake in general. However, physiological measurements could be a valuable complement here (e.g. nitrogen stable isotopes) to assess their interplay with dietary choices[101]. With arthropod feeding behaviour, other hypotheses

can be tested that are relevant to assess the influence of social factors on diet. Catching and handling arthropods requires skills that may depend on age and experience or be socially learnt, it could be used to assess an age effect. Another hypothesis is that the consumption of highly nutrient food as arthropods and vertebrates depends on the rank of an individual, i.e. the social status in the group. Mannion et al. (2022) favour combinations of ecological, physiological and behavioural assessments to inquire about cultural foraging practices, in particular when animal culture is not only accepted but gains attention for conservation actions[101]. eDNA data may not be sufficient alone to answer complex questions in the realm of animal culture but can serve as valuable complement to multidisciplinary approaches. In this thesis, the interdisciplinary approach between the fields of behavioural ecology and molecular biology yielded interesting insights into the behavioural ecology of vervet monkeys, with potentially more in the near future.

References

1. Taberlet, P., Bonin, A., Zinger, L. & Coissac, É. *Environmental DNA: For Biodiversity Research and Monitoring*. (Oxford University Press, 2018). doi:10.1093/oso/9780198767220.001.0001

2. Thomsen, P. F. & Willerslev, E. Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* **183**, 4–18 (2015).

3. Ogram, A., Sayler, G. S. & Barkay, T. The extraction and purification of microbial DNA from sediments. *J. Microbiol. Methods* **7**, 57–66 (1987).

4. Paget, E., Lebrun, M., Freyssinet, G. & Simonet, P. The fate of recombinant plant DNA in soil. *Eur. J. Soil Biol.* **34**, 81–88 (1998).

5. Willerslev, E., Hansen, A. J., Christensen, B., Steffensen, J. P. & Arctander, P. Diversity of Holocene life forms in fossil glacier ice. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8017–8021 (1999).

6. Ficetola, G. F., Miaud, C., Pompanon, F. & Taberlet, P. Species detection using environmental DNA from water samples. *Biol. Lett.* **4**, 423–425 (2008).

7. Bohmann, K. *et al.* Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* **29**, 358–367 (2014).

8. Ruppert, K. M., Kline, R. J. & Rahman, M. S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* **17**, e00547 (2019).

9. Pascher, K., Švara, V. & Jungmeier, M. Environmental DNA-Based Methods in Biodiversity Monitoring of Protected Areas: Application Range, Limitations, and Needs. *Diversity* **14**, (2022).

10. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050 (2012).

11. Schneider, J. *et al.* Detection of Invasive Mosquito Vectors Using Environmental DNA (eDNA) from Water Samples. *PLoS One* **11**, e0162493 (2016).

12. Valentin, R. E. *et al.* Moving eDNA surveys onto land: Strategies for active eDNA aggregation to detect invasive forest insects. *Mol. Ecol. Resour.* **20**, 746– 755 (2020).

13. Maslo, B. *et al.* Chirosurveillance: The use of native bats to detect invasive agricultural pests. *PLoS One* **12**, 1–10 (2017).

14. Shehzad, W. *et al.* Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (Prionailurus bengalensis) in Pakistan. *Mol. Ecol.* **21**, 1951–1965 (2012).

15. Palacios Mejia, M. *et al.* The utility of environmental DNA from sediment and water samples for recovery of observed plant and animal species from four Mojave Desert springs. *Environ. DNA* **3**, 214–230 (2021).

16. Mas-Carrió, E. *et al.* Assessing environmental DNA metabarcoding and camera trap surveys as complementary tools for biomonitoring of remote desert water bodies. *Environ. DNA* **4**, 580– 595 (2021).

17. Bylemans, J. *et al.* An environmental DNA-based method for monitoring spawning activity: a case study, using the endangered Macquarie perch (Macquaria australasica). *Methods Ecol. Evol.* **8**, 646–655 (2017).

18. De Barba, M. *et al.* DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Mol. Ecol. Resour.* **14**, 306–323 (2014).

19. Pansu, J. *et al.* Trophic ecology of large herbivores in a reassembling African ecosystem. *J. Ecol.* **107**, 1355–1376 (2019).

20. Potter, A. B. *et al.* Mechanisms of dietary resource partitioning in large-herbivore assemblages: A plant-trait-based approach. *J. Ecol.* **110**, 817–832 (2022).

21. Soininen, E. M. *et al.* Shedding new light on the diet of Norwegian lemmings: DNA metabarcoding of stomach content. *Polar Biol.* **36**, 1069–1076 (2013).

22. Swift, J. F. *et al.* Multifaceted DNA metabarcoding: Validation of a noninvasive, next-generation approach to studying bat populations. *Evol. Appl.* **11**, 1120–1138 (2018).

23. Bohmann, K. *et al.* Using DNA metabarcoding for simultaneous inference of common vampire bat diet and population structure. *Mol. Ecol. Resour.* **18**, 1050–1063 (2018).

24. Zinger, L. *et al.* Body size determines soil community assembly in a tropical forest. *Mol. Ecol.* **28**, 528–543 (2019).

25. Bista, I. *et al.* Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nat. Commun.* **8**, (2017).

26. Schiøtt, S. *et al.* Environmental DNA metabarcoding reveals seasonal and spatial

variation in the vertebrate fauna of Ilulissat Icefjord, Greenland. *Mar. Ecol. Prog. Ser.* **706**, 91–108 (2023).

27. Berry, T. E. *et al.* Marine environmental DNA biomonitoring reveals seasonal patterns in biodiversity and identifies ecosystem responses to anomalous climatic events. *PloS Genet.* **15**, e1007943 (2019).

28. Sigsgaard, E. E. *et al.* Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nat. Ecol. Evol.* **1**, 0004 (2016).

29. Dugal, L. *et al.* Individual haplotyping of whale sharks from seawater environmental DNA. *Mol. Ecol. Resour.* **22**, 56–65 (2022).

30. Pansu, J. *et al.* Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. *Mol. Ecol.* **24**, 1485–1498 (2015).

31. Pedersen, M. W. *et al.* Ancient and modern environmental DNA. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, (2015).

32. Harper, L. R. *et al.* BeeDNA: Microfluidic environmental DNA metabarcoding as a tool for connecting plant and pollinator communities. *Environ. DNA* **5**, 191–211 (2023).

33. Roger, F. *et al.* Airborne environmental DNA metabarcoding for the monitoring of terrestrial insects—A proof of concept from the field. *Environ. DNA* **4**, 790–807 (2022).

34. Nguyen, T. V. *et al.* Using terrestrial leeches to assess the genetic diversity of an elusive species: The Annamite striped rabbit Nesolagus timminsi. *Environ. DNA* **3**, 780–791 (2021).

35. Wilting, A. *et al.* Creating genetic reference datasets: Indirect sampling of target species using terrestrial leeches as sample "collectors". *Environ. DNA* **4**, 311–325 (2022).

36. Nørgaard, L. *et al.* eDNA metabarcoding for biodiversity assessment, generalist predators as sampling assistants. *Sci. Rep.* **11**, 6820 (2021).

37. Gruber, T. An ethical assessment of the use of old and new methods to study sociality in wild animals. *Methods Ecol. Evol.* **14**, 1842–1851 (2023).

38. Barnes, M. A. & Turner, C. R. The ecology of environmental DNA and implications for conservation genetics. *Conserv. Genet.* **17**, 1–17 (2016).

39. Piñol, J., Senar, M. A. & Symondson, W. O. C. The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Mol. Ecol.* **28**, 407–419 (2019).

40. Piñol, J., Mir, G., Gomez-Polo, P. & Agustí, N. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Mol. Ecol. Resour.* **15**, 819–830 (2015).

41. Bista, I. *et al.* Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Mol. Ecol. Resour.* **18**, 1020–1034 (2018).

42. Deagle, B. E. *et al.* Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Mol. Ecol.* **28**, 391–406 (2019).

43. Calderón-Sanou, I., Münkemüller, T., Boyer, F., Zinger, L. & Thuiller, W. From environmental DNA sequences to ecological conclusions: How strong is the influence of

methodological choices? *J. Biogeogr.* **47**, 193–206 (2019).

44. Fonseca, V. G. Pitfalls in relative abundance estimation using edna metabarcoding. *Mol. Ecol. Resour.* **18**, 923–926 (2018).

45. Clare, E. L. Molecular detection of trophic interactions: emerging trends, distinct advantages, significant considerations and conservation applications. *Evol. Appl.* **7**, 1144–1157 (2014).

46. Matthews, J. K., Ridley, A., Kaplin, B. A. & Grueter, C. C. A comparison of fecal sampling and direct feeding observations for quantifying the diet of a frugivorous primate. *Curr. Zool.* **66**, 333–343 (2020).

47. Deagle, B. E., Chiaradia, A., McInnes, J. & Jarman, S. N. Pyrosequencing faecal DNA to determine diet of little penguins: Is what goes in what comes out? *Conserv. Genet.* **11**, 2039–2048 (2010).

48. Sassoubre, L. M., Yamahara, K. M., Gardner, L. D., Block, B. A. & Boehm, A. B. Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. *Environ. Sci. Technol.* **50**, 10456–10464 (2016).

49. Rojahn, J. *et al.* The value of quantitative environmental DNA analyses for the management of invasive and endangered native fish. *Freshw. Biol.* **66**, 1619–1629 (2021).

50. Kartzinel, T. R. *et al.* DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8019–8024 (2015).

51. Vesterinen, E. J., Puisto, A. I. E., Blomberg, A. S. & Lilley, T. M. Table for five, please:

Dietary partitioning in boreal bats. *Ecol. Evol.* **8**, 10914–10937 (2018).

52. Zinger, L. *et al.* DNA metabarcoding–Need for robust experimental designs to draw sound ecological conclusions. *Mol. Ecol.* **28**, 1857–1862 (2019).

53. Jusino, M. A. *et al.* An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Mol. Ecol. Resour.* **19**, 176–190 (2019).

54. Krehenwinkel, H., Kennedy, S., Pekár, S. & Gillespie, R. G. A cost-efficient and simple protocol to enrich prey DNA from extractions of predatory arthropods for large-scale gut content analysis by Illumina sequencing. *Methods Ecol. Evol.* **8**, 126–134 (2017).

55. Alberdi, A., Aizpurua, O., Gilbert, M. T. P. & Bohmann, K. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods Ecol. Evol.* **9**, 134–147 (2018).

56. Galan, M. *et al.* Metabarcoding for the parallel identification of several hundred predators and their prey: Application to bat species diet analysis. *Mol. Ecol. Resour.* **18**, 474–489 (2018).

57. Creer, S. *et al.* The ecologist's field guide to sequence-based identification of biodiversity. *Methods Ecol. Evol.* **7**, 1008–1018 (2016).

58. Jones, M. R. & Good, J. M. Targeted capture in evolutionary and ecological genomics. *Mol. Ecol.* **25**, 185–202 (2016).

59. Gasc, C., Peyretaillade, E. & Peyret, P. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic Acids*

Res. **44**, 4504–4518 (2016).

60. Mamanova, L. *et al.* Target-enrichment strategies for next-generation sequencing. *Nat. Methods* **7**, 111–118 (2010).

61. Paijmans, J. L. A., Fickel, J., Courtiol, A., Hofreiter, M. & Förster, D. W. Impact of enrichment conditions on cross-species capture of fresh and degraded DNA. *Mol. Ecol. Resour.* **16**, 42–55 (2016).

62. Hemelrijk, C. K., Wubs, M., Gort, G., Botting, J. & van de Waal, E. Dynamics of Intersexual Dominance and Adult Sex- Ratio in Wild Vervet Monkeys. *Front. Psychol.* **11**, 1–11 (2020).

63. Cheney, D. L. & Seyfarth, R. M. Nonrandom Dispersal in Free-Ranging Vervet Monkeys: Social and Genetic Consequences. *Am. Nat.* **122**, 392–412 (1983).

64. Whiten, A. & van de Waal, E. The pervasive role of social learning in primate lifetime development. *Behav. Ecol. Sociobiol.* **72**, 80 (2018).

65. Mertz, J., Surreault, A., van de Waal, E. & Botting, J. Primates are living links to our past: The contribution of comparative studies with wild vervet monkeys to the field of social cognition. *Cortex* **118**, 65–81 (2019).

66. van de Waal, E., Claidière, N. & Whiten, A. Wild vervet monkeys copy alternative methods for opening an artificial fruit. *Anim. Cogn.* **18**, 617–627 (2015).

67. Krützen, M., van Schaik, C. & Whiten, A. The animal cultures debate: response to Laland and Janik. *Trends Ecol. Evol.* **22**, 6 (2007).

68. Schneider, J., Brun, L., Taberlet, P., Fumagalli, L. & van de Waal, E. Molecular assessment of dietary variation in neighbouring primate groups. *Methods Ecol. Evol.* **14**, 1925–1936 (2023).

69. *The vegetation of South Africa, Lesotho and Swaziland*. (South African National Biodiversity Institute, Pretoria, 2006).

70. Brun, L. *et al.* Focal vs. faecal: seasonal variation in the diet of wild vervet monkeys from observational and DNA metabarcoding data. *Ecol. Evol.* **12**, e9358 (2022).

71. Heyes, C. M. Social learning in animals: categories and mechanisms. *Biol. Rev.* **69**, 207–231 (1994).

72. Whiten, A. The second inheritance system of chimpanzees and humans. *Nature* **437**, 52–55 (2005).

73. Rendell, L. & Whitehead, H. Culture in whales and dolphins. *Behav. Brain Sci.* **24**, 309–382 (2001).

74. Brakes, P. *et al.* A deepening understanding of animal culture suggests lessons for conservation. *Proc. R. Soc. B Biol. Sci.* **288**, 1–10 (2021).

75. Gruber, T. *et al.* Cultural change in animals: a flexible behavioural adaptation to human disturbance. *Palgrave Commun.* **5**, (2019).

76. Keith, S. A. & Bull, J. W. Animal culture impacts species' capacity to realise climate-driven range shifts. *Ecography (Cop.).* **40**, 296–304 (2017).

77. Kühl, H. S. *et al.* Human impact erodes chimpanzee behavioral diversity. *Science (80-.).*

363, 1453–1455 (2019).

78. Sih, A. Understanding variation in behavioural responses to human-induced rapid environmental change: A conceptual overview. *Anim. Behav.* **85**, 1077–1088 (2013).

79. Hopper, L. M. Cognitive research in zoos. *Curr. Opin. Behav. Sci.* **16**, 100–110 (2017).

80. Harrison, R. A., Mohr, T. & van de Waal, E. Lab cognition going wild: Implementing a new portable touchscreen system in vervet monkeys. *J. Anim. Ecol.* **92**, 1545–1559 (2023).

81. Whiten, A. *et al.* Cultures in chimpanzees. *Nature* **399**, 15–18 (1999).

82. Ehmann, B. *et al.* Immature wild orangutans acquire relevant ecological knowledge through sex-specific attentional biases during social learning. *PLoS Biol.* **19**, 1–21 (2021).

83. Jaeggi, A. V. *et al.* Social Learning of Diet and Foraging Skills by Wild Immature Bornean Orangutans: Implications for Culture. *Am. J. Primatol.* **72**, 62–71 (2010).

84. Klump, B. C. *et al.* Innovation and geographic spread of a complex foraging culture in an urban parrot. *Science (80-.).* **373**, 456–460 (2021).

85. Owen, C. *et al.* Migratory convergence facilitates cultural transmission of humpback whale song. *R. Soc. Open Sci.* **6**, (2019).

86. Cauchoix, M., Hermer, E., Chaine, A. S. & Morand-Ferron, J. Cognition in the field: Comparison of reversal learning performance in captive and wild passerines. *Sci. Rep.* **7**, 12945 (2017).

87. Aplin, L. M., Sheldon, B. C. & McElreath, R. Conformity does not perpetuate suboptimal

traditions in a wild population of songbirds. *Proc. Natl. Acad. Sci. U. S. A.* **114**, (2017).

88. Aplin, L. M. *et al.* Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature* **518**, 538–541 (2015).

89. Slagsvold, T. & Wiebe, K. L. Social learning in birds and its role in shaping a foraging niche. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 969–977 (2011).

90. Thornton, A. Social learning about novel foods in young meerkats. *Anim. Behav.* **76**, 1411–1421 (2008).

91. Rapaport, L. G. & Brown, G. R. Social influences on foraging behavior in young nonhuman primates: Learning what, where, and how to eat. *Evol. Anthropol.* **17**, 189–201 (2008).

92. Kaufhold, S. P. & van Leeuwen, E. J. C. Why intergroup variation matters for understanding behaviour. *Biol. Lett.* **15**, (2019).

93. van Boekholt, B., van de Waal, E. & Sterck, E. H. M. Organized to learn: the influence of social structure on social learning opportunities in a group. *iScience* **24**, 102117 (2021).

94. van de Waal, E., Renevey, N., Favre, C. M. & Bshary, R. Selective attention to philopatric models causes directed social learning in wild vervet monkeys. *Proc. R. Soc. B Biol. Sci.* **277**, 2105–2111 (2010).

95. van de Waal, E., Borgeaud, C. & Whiten, A. Potent social learning and conformity shape a wild primate's foraging decisions. *Science (80-.).* **340**, 483–485 (2013).

96. van de Waal, E., Bshary, R. & Whiten, A. Wild vervet monkey infants acquire the food-

processing variants of their mothers. *Anim. Behav.* **90**, 41–45 (2014).

97. van de Waal, E. On the neglected behavioural variation among neighbouring primate groups. *Ethology* **124**, 845–854 (2018).

98. Bono, A. E. J. *et al.* Payoff- and Sex-Biased Social Learning Interact in a Wild Primate Population. *Curr. Biol.* **28**, 2800–2805 (2018).

99. Voelker, M. R., Schwarz, D., Thomas, A., Nelson, B. W. & Acevedo-Gutiérrez, A. Large-scale molecular barcoding of prey DNA reveals predictors of intrapopulation feeding diversity in a marine predator. *Ecol. Evol.* **10**, 9867–9885 (2020).

100. Estrada, A. *et al.* Impending extinction crisis of the world's primates: Why primates matter. *Sci. Adv.* **3**, (2017).

101. Mannion, K. R., Ballare, E. F., Marks, M. & Gruber, T. A multicomponent approach to studying cultural propensities during foraging in the wild. *J. Anim. Ecol.* **92**, 1478–1488 (2023).

102. van Wyk, B. & van Wyk, P. *Field Guide to Trees of Southern Africa*. (Penguin Random House South Africa, 2013).

103. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.5-7. (2014).

104. Cribari-Neto, F. & Zeileis, A. Beta Regression in R. *J. Stat. Softw.* **34**, 129–150 (2010).

105. Lenth, R. emmeans: Estimated Marginal Means, aka Least-Squares Mean. in (R package version 1.4.1, 2019).

106. Zaccarelli, N., Bolnick, D. I. & Mancinelli, G. RInSp: An r package for the analysis of

individual specialization in resource use. *Methods Ecol. Evol.* **4**, 1018–1023 (2013).

107. Brakes, P. *et al.* Animal cultures matter for conservation. *Science (80-.).* **363**, 1032–1034 (2019).

108. van Schaik, C. P. *et al.* Orangutan Cultures and the Evolution of Material Culture. *Science (80-.).* **299**, 102–106 (2003).

109. Fragaszy, D. & Perry, S. *The Biology of Traditions: Models and Evidence*. (Cambridge University Press., 2003). doi:10.1017/CBO9780511584022.002

110. Kendal, R. L. *et al.* Social Learning Strategies: Bridge-Building between Fields. *Trends Cogn. Sci.* **22**, 651–665 (2018).

111. Gruber, T. & van de Waal, E. Striking pay dirt: Contemporary methods for studying animal sociality in the wild. *Methods Ecol. Evol.* **14**, 1838–1841 (2023).

112. Caravaggi, A. *et al.* A review of factors to consider when using camera traps to study animal behavior to inform wildlife ecology and conservation. *Conserv. Sci. Pract.* **2**, 1–9 (2020).

113. Dickie, I. A. *et al.* Towards robust and repeatable sampling methods in eDNA-based studies. *Mol. Ecol. Resour.* **18**, 940–952 (2018).

114. Koops, K. *et al.* Crab-fishing by chimpanzees in the Nimba Mountains, Guinea. *J. Hum. Evol.* **133**, 230–241 (2019).

115. Furlan, E. M., Davis, J. & Duncan, R. P. Identifying error and accurately interpreting environmental DNA metabarcoding results: A case study to detect vertebrates at arid

zone waterholes. *Mol. Ecol. Resour.* **20**, 1259–1276 (2020).

116. Wilcox, T. M. *et al.* Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity. *PLoS One* **8**, e59520 (2013).

117. Laland, K. N. Social learning strategies. *Anim. Learn. Behav.* **32**, 4–14 (2004).

118. Kendal, R. L., Coolen, I., van Bergen, Y. & Laland, K. N. Trade-Offs in the Adaptive Use of Social and Asocial Learning. in *Advances in the Study of Behavior* 333–379 (Press, Academic, 2005). doi:10.1016/S0065-3454(05)35008-X

119. Dongre, P., Canteloup, C., Lanté, G., Cantat, M. & van de Waal, E. Role of immigrants and muzzle contacts in the uptake of a novel food by wild vervet monkeys. *bioRxiv* 1–27 (2021). doi:10.1101/2021.12.16.472640

120. Beck, K. B., Sheldon, B. C. & Firth, J. A. Social learning mechanisms shape transmission pathways through replicate local social networks of wild birds. *Elife* **12**, 1–23 (2023).

121. Schwegel, M. A., Filazzola, A., Chapman, C. A. & Schoof, V. A. M. Breeding Seasonality in Female Vervet Monkeys (Chlorocebus pygerythrus) Living in an Anthropogenic Landscape. *Int. J. Primatol.* **44**, 45–70 (2023).

122. Rothman, J. M., Raubenheimer, D., Bryer, M. A. H., Takahashi, M. & Gilbert, C. C. Nutritional contributions of insects to primate diets: Implications for primate evolution. *J. Hum. Evol.* **71**, 59–69 (2014).

123. Dufrêne, M. & Legendre, P. Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecol. Monogr.* **67**, 345–366 (1997).

124. Vestheim, H. & Jarman, S. N. Blocking primers to enhance PCR amplification of rare sequences in mixed samples - A case study on prey DNA in Antarctic krill stomachs. *Front. Zool.* **5**, 1–11 (2008).

125. Hallmann, C. A. *et al.* More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS One* **12**, (2017).

126. Wagner, D. L., Grames, E. M., Forister, M. L., Berenbaum, M. R. & Stopak, D. Insect decline in the Anthropocene: Death by a thousand cuts. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–10 (2021).

127. Pickett, S. B., Bergey, C. M. & Di Fiore, A. A Metagenomic Study of Primate Insect Diet Diversity. *Am. J. Primatol.* **74**, 622–631 (2012).

128. Rowe, A. K. *et al.* Exploratory analysis reveals arthropod consumption in 10 lemur species using DNA metabarcoding. *Am. J. Primatol.* **83**, (2021).

129. Lyke, M. M., Di Fiore, A., Fierer, N., Madden, A. A. & Lambert, J. E. Metagenomic analyses reveal previously unrecognized variation in the diets of sympatric Old World monkey species. *PLoS One* **14**, e0218245 (2018).

130. Camila, L., Dejean, T., Valentini, A., Pauline, J. & Pierre, J. R. A novel trap design for non-lethal monitoring of dung beetles using eDNA metabarcoding. *J. Insect Conserv.* **25**, 629–642 (2021).

131. Dowle, E. J., Pochon, X., C Banks, J., Shearer, K. & Wood, S. A. Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. *Mol. Ecol. Resour.* **16**, 1240–1254 (2016).

132. Gregorič, M. *et al.* Spider webs as eDNA samplers: Biodiversity assessment across the tree of life. *Mol. Ecol. Resour.* **22**, 2534–2545 (2022).

133. Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A. & Johnson, E. A. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* **17**, 240–248 (2007).

134. Bi, K. *et al.* Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics* **13**, (2012).

135. Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C. & Brumfield, R. T. Sequence Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. *Syst. Biol.* **65**, 910–924 (2016).

136. Andermann, T. *et al.* A Guide to Carrying Out a Phylogenomic Target Sequence Capture Project. *Front. Genet.* **10**, 1–20 (2020).

137. Hedtke, S. M., Morgan, M. J., Cannatella, D. C. & Hillis, D. M. Targeted Enrichment : Maximizing Orthologous Gene Comparisons across Deep Evolutionary Time. *PLoS One* **8**, e67908 (2013).

138. Mason, V. C., Li, G., Helgen, K. M. & Murphy, W. J. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Res.* **21**, 1695–1704 (2011).

139. Gnirke, A. *et al.* Solution Hybrid Selection with Ultra-long Oligonucleotides for Massively Parallel Targeted Sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).

140. Okou, D. T. *et al.* Microarray-based genomic selection for high-throughput resequencing. *Nat. Methods* **4**, 907–909 (2007).

141. Mohandesan, E. *et al.* Combined hybridization capture and shotgun sequencing for ancient DNA analysis of extinct wild and domestic dromedary camel. *Mol. Ecol. Resour.* **17**, 300–313 (2017).

142. Gauthier, J. *et al.* Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland. *Mol. Ecol. Resour.* **20**, 1191–1205 (2020).

143. Liu, S. *et al.* Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Mol. Ecol. Resour.* **16**, 470–479 (2016).

144. Mariac, C. *et al.* Metabarcoding by capture using a single COI probe (MCSP) to identify and quantify fish species in ichthyoplankton swarms. *PLoS One* **13**, 1–15 (2018).

145. Maggia, M. E. *et al.* DNA metabarcoding of amazonian ichthyoplankton swarms. *PLoS One* **12**, 1–14 (2017).

146. Wilcox, T. M. *et al.* Capture enrichment of aquatic environmental DNA: A first proof of concept. *Mol. Ecol. Resour.* **18**, 1392–1401 (2018).

147. Gauthier, M. *et al.* Enhancing DNA metabarcoding performance and applicability with bait capture enrichment and DNA from conservative ethanol. *Mol. Ecol. Resour.* **20**, 79–96 (2020).

148. Seeber, P. A. *et al.* Terrestrial mammal surveillance using hybridization capture of environmental DNA from African waterholes. *Mol. Ecol. Resour.* **19**, 1486–1496 (2019).

149. Giebner, H. *et al.* Comparing diversity levels in environmental samples: DNA sequence capture and metabarcoding approaches using 18S and COI genes. *Mol. Ecol. Resour.* **20**, 1333–1345 (2020).

150. Beaudry, M. S. *et al.* Improved Microbial Community Characterization of 16S rRNA via Metagenome Hybridization Capture Enrichment. *Front. Microbiol.* **12**, 1–14 (2021).

151. Schulte, L. *et al.* Hybridization capture of larch (Larix Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. *Mol. Ecol. Resour.* **21**, 801–815 (2021).

152. Miles Zhang, Y., Williams, J. L. & Lucky, A. Understanding UCEs: A Comprehensive Primer on Using Ultraconserved Elements for Arthropod Phylogenomics. *Insect Syst. Divers.* **3**, 1–12 (2019).

153. Saenz Manchola, O. F. *et al.* Mining Ultraconserved Elements From Transcriptome and Genome Data to Explore the Phylogenomics of the Free-living Lice Suborder Psocomorpha (Insecta: Psocodea). *Insect Syst. Divers.* **6**, 1–11 (2022).

154. Liu, D., Niu, M., Lu, Y., Wei, J. & Zhang, H. Taxon-specific ultraconserved element probe design for phylogenetic analyses of scale insects (Hemiptera: Sternorrhyncha: Coccoidea). *Front. Ecol. Evol.* **10**, 1–14 (2022).

155. Gustafson, G. T., Glynn, R. D., Short, A. E. Z., Tarasov, S. & Gunter, N. L. To design, or not to design? Comparison of beetle ultraconserved element probe set utility based on phylogenetic distance, breadth, and method of probe design. *Insect Syst. Divers.* **7**, 1–14 (2023).

156. Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* **270**, 313–321 (2003).

157. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).

158. Glenn, T. C., Faircloth, B. C. & Rouge, B. Capturing Darwin's Dream. **16**, 1051–1058 (2017).

159. Hutter, C. R. *et al.* FrogCap: A modular sequence capture probe-set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales. *Mol. Ecol. Resour.* **22**, 1100–1119 (2022).

160. Hoffberg, S. L. *et al.* RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Mol. Ecol. Resour.* **16**, 1264–1278 (2016).

161. Suchan, T. *et al.* Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS One* **11**, 1–22 (2016).

162. Schmid, S. *et al.* HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. *Methods Ecol. Evol.* **8**, 1374–1388 (2017).

163. Peñalba, J. V. *et al.* Sequence capture using PCR-generated probes: A cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Mol. Ecol. Resour.* **14**, 1000–1010 (2014).

164. Mayer, C. *et al.* BaitFisher: A Software Package for Multispecies Target DNA Enrichment Probe Design. *Mol. Biol. Evol.* **33**, 1875–1886 (2016).

165. Jiménez-Mena, B. *et al.* Fishing for DNA? Designing baits for population genetics in target enrichment experiments: Guidelines, considerations and the new tool supeRbaits. *Mol. Ecol. Resour.* **22**, 2105–2119 (2022).

166. Edgar, R. C. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).

167. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).

168. Tournier, E. *et al.* Differences in Diet Between Six Neighbouring Groups of Vervet Monkeys. *Ethology* **120**, 471–482 (2014).

169. Zeale, M. R. K., Butlin, R. K., Barker, G. L. A., Lees, D. C. & Jones, G. Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Mol. Ecol. Resour.* **11**, 236–244 (2011).

170. Gillet, F. *et al.* A new method to identify the endangered Pyrenean desman (Galemys pyrenaicus) and to study its diet, using next generation sequencing from faeces. *Mamm. Biol.* **80**, 505–509 (2015).

171. Corse, E. *et al.* A from-benchtop-to-desktop workflow for validating HTS data and for taxonomic identification in diet metabarcoding studies. *Mol. Ecol. Resour.* **17**, e146–e159 (2017).

172. Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F. & Taberlet, P. DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biol. Lett.* **10**, 2–5

(2014).

173. Clarke, L. J., Soubrier, J., Weyrich, L. S. & Cooper, A. Environmental metabarcodes for insects: In silico PCR reveals potential for taxonomic bias. *Mol. Ecol. Resour.* **14**, 1160–1170 (2014).

174. Meiklejohn, K. A., Damaso, N. & Robertson, J. M. Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. *PLoS One* **14**, e0217084 (2019).

175. Elbrecht, V. *et al.* Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *PeerJ* **2016**, 1–12 (2016).

176. Tournayre, O. *et al.* In silico and empirical evaluation of twelve metabarcoding primer sets for insectivorous diet analyses. *Ecol. Evol.* **10**, 6310–6332 (2020).

177. Aizpurua, O. *et al.* Agriculture shapes the trophic niche of a bat preying on multiple pest arthropods across Europe: Evidence from DNA metabarcoding. *Mol. Ecol.* **27**, 815–825 (2018).

178. Shively, R., Barboza, P., Doak, P. & Jung, T. S. Increased diet breadth of little brown bats (*Myotis lucifugus*) at their northern range limit: a multimethod approach. *Can. J. Zool.* **96**, 31–38 (2018).

179. Clare, E. L. *et al.* The diet of Myotis lucifugus across Canada: Assessing foraging quality and diet variability. *Mol. Ecol.* **23**, 3618–3632 (2014).

180. Arrizabalaga-Escudero, A. *et al.* Assessing niche partitioning of co-occurring sibling bat species by DNA metabarcoding. *Mol. Ecol.* **27**, 1273–1283 (2018).

181. Burgar, J. M. *et al.* Who's for dinner? High-throughput sequencing reveals bat dietary differentiation in a biodiversity hotspot where prey taxonomy is largely undescribed. *Mol. Ecol.* **23**, 3605–3617 (2014).

182. Razgour, O. *et al.* High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. *Ecol. Evol.* **1**, 556–570 (2011).

183. Vesterinen, E. J., Lilley, T., Laine, V. N. & Wahlberg, N. Next generation sequencing of fecal DNA reveals the dietary diversity of the widespread insectivorous predator Daubenton's bat (Myotis daubentonii) in southwestern Finland. *PLoS One* **8**, 1–14 (2013).

Appendices

Supplementary data Chapter 1

Appendix S1

Focal vs. faecal: seasonal variation in the diet of wild vervet monkeys from observational and DNA metabarcoding data

Loïc Brun, Judith Schneider, Eduard Mas Carrió, Pooja Dongre, Pierre Taberlet, Erica van de Waal and Luca Fumagalli

Ecology and Evolution

Appendix S1: Table S1: Species of local database, in bold species with identical sequences.

Family	Species
Anacardiaceae	***Searsia pyroides, Searsia natalensis***
Anacardiaceae	*Sclerocarya birrea*
Araliaceae	*Cussonia spicata*
Arecaceae	*Phoenix reclinata*
Bignoniaceae	*Jacaranda mimosifolia*
Boraginaceae	*Ehretia rigida*
Burseraceae	*Commiphora neglecta*
Cactaceae	*Cereus jamacaru*
Cactaceae	*Opuntia ficus-indica*
Caesalpiniaceae	*Schotia brachypetala*
Capparaceae	*Boscia albitrunca*
Celastraceae	*Gymnosporia senegalensis*
Celtidaceae	*Celtis africana*
Combretaceae	*Combretum erythrophyllum*
Combretaceae	*Combretum apiculatum*
Ebenaceae	***Euclea crispa, Euclea undulata, Diospyros dichrophy***
Euphorbiaceae	*Euphorbia ingens*
Euphorbiaceae	*Euphorbia tirucalli*
Fabaceae	*Dalbergia armata*
Fabaceae	*Dalbergia obovata*
Fabaceae	*Senna didymobotrya*
Fabaceae	*Sesbania punicea*

Appendix S1: Table S1: Species of local database, in bold species with identical sequences.

Family	Species
Anacardiaceae	***Searsia pyroides, Searsia natalensis***
Anacardiaceae	*Sclerocarya birrea*
Araliaceae	*Cussonia spicata*
Arecaceae	*Phoenix reclinata*
Bignoniaceae	*Jacaranda mimosifolia*
Boraginaceae	*Ehretia rigida*
Burseraceae	*Commiphora neglecta*
Cactaceae	*Cereus jamacaru*
Cactaceae	*Opuntia ficus-indica*
Caesalpiniaceae	*Schotia brachypetala*
Capparaceae	*Boscia albitrunca*
Celastraceae	*Gymnosporia senegalensis*
Celtidaceae	*Celtis africana*
Combretaceae	*Combretum erythrophyllum*
Combretaceae	*Combretum apiculatum*
Ebenaceae	***Euclea crispa, Euclea undulata, Diospyros dichrophylla***
Euphorbiaceae	*Euphorbia ingens*
Euphorbiaceae	*Euphorbia tirucalli*
Fabaceae	*Dalbergia armata*
Fabaceae	*Dalbergia obovata*
Fabaceae	*Senna didymobotrya*
Fabaceae	*Sesbania punicea*
Fabaceae	*Mundulea sericea*

Flacourtiaceae	*Dovyalis caffra*
Lamiaceae	*Premna mooiensis*
Lamiaceae	*Volkameria glabra*
Meliaceae	*Melia azedarach*
Mimosaceae	***Vachellia tortilis, Vachellia sieberiana***
Mimosaceae	*Dichrostachys cinerea*
Mimosaceae, Caesalpiniaceae	***Vachellia nilotica, Caesalpinia decapetala***
Moraceae	*Ficus sycomorus*
Myrtaceae	*Psidium guajava*
Myrtaceae	*Eucalyptus camaldulensis*
Olacaceae	*Ximenia caffra*
Oleaceae	*Olea europaea*
Pentapetaceae	*Dombeya rotundifolia*
Rhamnaceae	*Berchemia zeyheri*
Rhamnaceae	*Ziziphus mucronata*
Rubiaceae	*Gardenia volkensii*
Rubiaceae	*Coddia rudis*
Rutaceae	*Citrus limon*
Sapindaceae	*Hippobromus pauciflorus*
Sapindaceae	*Pappea capensis*
Sapotaceae	*Sideroxylon inerme*
Solanaceae	*Solanum seaforthianum*
Solanaceae	*Solanum aculeastrum*
Verbenaceae	***Lantana camara, Lippia javanica***
Vitaceae	*Rhoicissus tridentata*

Appendix S1: Table S2 Species included in positive controls for Sper01 and Arth02 assays, in the order of 2-fold dilutions.

Metabarcode	Species
Sper01	*Taxus baccata, Salvia pratensis, Populus tremula, Rumex acetosa, Carpinus betulus, Fraxinus excelsior, Picea abies, Lonicera xylosteum, Abies alba, Acer campestre, Briza media, Rosa canina, Capsella bursa-pastoris, Geranium robertianum, Rhododendron ferrugineum, Lotus corniculatus*
Arth02	*Acheta domesticus, Timema douglasi, Harmonia axyridis, Galleria mellonella, Pyrrhocoris apterus, Blaptica dubia, Isoperla rivulorum, Silo pallipes*

Appendix S1: Table S3 Plant genus and species in observational focal follows and detected in faecal samples. Plant indicators for seasons were identified using *Indicator value analyses* (Indval; Dufrêne & Legendre, 1997). For database assignments: L = assigned with local database, G = assigned with global database, NA = no sequence available for the used metabarcode.

Observational Data			**DNA metabarcoding data**					**Plant category**
Genus/species observed	**Frequency/ 12315**	**Season indicator > 0.2, *****	**Genus/species detected**	**Frequency/ 823**	**Total read counts**	**Season indicator RRA > 0.2, *****	**Data-base**	
Berchemia zeyheri	704	summer+autumn, 0.397	*Berchemia zeyheri*	811	1285128	summer+autumn, 0.460	L	tree
Boscia albitrunca	2		*Boscia albitrunca*	15	417		L	tree
Caesalpinia decapetala	84	spring, 0.255	*V. nilotica/C. decapetala*	82	4621		L	tree/shrub
Cereus jamacaru	752	spring, 0.324	*Cereus jamacaru*	365	53092		L	cactus
Clausena anisata	2		*Clausena anisata*	0	0		G	shrub
Coddia rudis	145	autumn, 0.236	*Coddia rudis*	31	324		L	shrub
Cussonia spicata	2		*Cussonia spicata*	0	0		L	tree
Dalbergia armata	92		*Dalbergia armata*	80	8113	autumn+winter, 0.235	L	liane

Dichrostachys cinerea	1346		*Dichrostachys cinerea*	685	95218	winter, 0.295	L	small tree/shrub
Dovyalis caffra	132	spring, 0.340	*Dovyalis caffra*	559	163527		L	small tree/shrub
Ehretia rigida	785	spring, 0.233	*Ehretia rigida*	523	236935	spring, 0.585	L	small tree/shrub
Euclea crispa	249	spring, 0.332	*E. crispa/E. undulata/ D. dichrophylla*	321	98890	spring, 0.446	L	tree
Euclea undulata	16		*E. crispa/E. undulata/ D. dichrophylla*	321	98890	spring, 0.446	L	small tree/shrub
Euphorbia ingens	6		*Euphorbia ingens*	0	0		L	cactus
Euphorbia tirucalli	3		*Euphorbia tirucalli*	0	0		L	shrub
Gardenia volkensii	51		*Gardenia volkensii*	23	483		L	small tree/shrub
Gymnosporia senegalensis	29		*Gymnosporia senegalensis*	0	0		L	small tree/shrub
Hippobromus pauciflorus	68	summer, 0.244	*Hippobromus pauciflorus*	444	225809	spring, 0.345	L	tree
Lantana camara	15	autumn, 0.216	*L. camara/Lippia javanica*	202	14426	autumn, 0.238	L	shrub

Opuntia ficus-indica	37		*Opuntia ficus-indica*	3	34		L	cactus
Premna mooiensis	391	spring+summer, 0.276	*Premna mooiensis*	309	48541	spring+summer, 0.223	L	tree
Schotia brachypetala	15		*Schotia brachypetala*	18	2490		L	tree
Searsia sp	688	autumn+spring, 0.278	*S. pyroides/S. natalensis*	638	256910	spring, 0.339	L	tree/shrub
Vachellia karroo	7		*NA*	0	0		NA	tree
Vachellia nilotica	1763	autumn+winter, 0.400	*V. nilotica/C. decapetala*	82	4621		L	tree/shrub
Vachellia sp	323	spring, 0.464	*NA*	0	0		NA	tree
Vachellia tortilis	2109	autumn, 0.423	*V. tortilis/V. sieberiana*	817	772456	autumn, 0.768	L	tree
Zanthoxylum capense	2		*NA*	0	0		NA	tree
Ziziphus mucronata	2497	autumn+winter, 0.490	*Ziziphus mucronata*	765	511273	winter, 0.681	L	tree
NA	0		*Aizoon*	467	50557	autumn+winter, 0.203	G	herb/shrub

NA	0	*Aloe*	3	37	G	various
NA	0	*Alternanthera pungens*	4	208	G	herb/shrub
NA	0	*Asparagus*	309	18695	G	herb/shrub
NA	0	*Basella alba*	4	26	G	climbing plant
NA	0	*Blepharis*	3	41	G	herb/shrub
NA	0	*Blepharis maderaspatensis*	447	87987	G	herb
NA	0	*Capparis*	45	7161	G	herb/shrub
NA	0	*Combretum*	9	190	G	shrub
NA	0	*Commiphora neglecta*	2	17	G	small tree/shrub
NA	0	*Crotalaria*	205	38455	G	herb/shrub
NA	0	*Diospyros*	4	30	G	tree
NA	0	*Dombeya rotundifolia*	31	2112	L	tree
NA	0	*Dysphania*	5	65	G	herb

NA	0	*Eragrostis superba*	2	51		G	grass
NA	0	*Erythrina*	2	104		G	tree
NA	0	*Euphorbia*	9	343		L	various
NA	0	*Ficus sycomorus*	21	6481		L	tree
NA	0	*Hibiscus*	4	197		G	shrub
NA	0	*Jasminum*	699	114664	autumn, 0.452	G	shrub
NA	0	*Kohautia*	2	16		G	herb/shrub
NA	0	*Mundulea sericea*	43	12193		L	shrub
NA	0	*Ocimum*	2	128		G	herb/shrub
NA	0	*Ormosia*	2	11		G	small tree/shrub
NA	0	*Oxalis*	325	36621	spring, 0.218	G	herb/shrub
NA	0	*Pappea capensis*	6	144		L	tree
NA	0	*Pereskia*	4	24		G	cactus

NA	0	*Phyllanthus maderaspatensis*	19	2943		G	herb
NA	0	*Plinthus*	17	141		G	herb/shrub
NA	0	*Priva*	7	109		G	herb/shrub
NA	0	*Priva cordifolia*	555	69468	winter+spring, 0.245	G	herb/shrub
NA	0	*Psidium guajava*	5	272		L	shrub
NA	0	*Rhoicissus tridentata*	212	31942		L	shrub
NA	0	*Sclerocarya birrea*	7	586		L	tree
NA	0	*Senna didymobotrya*	225	4949	autumn+winter, 0.324	L	herb/shrub
NA	0	*Sida*	6	1040		G	herb/shrub
NA	0	*Sideroxylon inerme*	11	1098		L	tree
NA	0	*Solanum seaforthianum*	5	696		L	shrub
NA	0	*Solidago virgaurea*	3	524		G	herb/shrub

NA	0	*Viscum minimum*	37	6829	autumn, 0.242	G	hemi-parasite
NA	0	*Volkameria glabra*	7	431		L	tree
NA	0	*Vigna*	2	47		G	herb/shrub
NA	0	*Waltheria indica*	6	217		G	herb/shrub

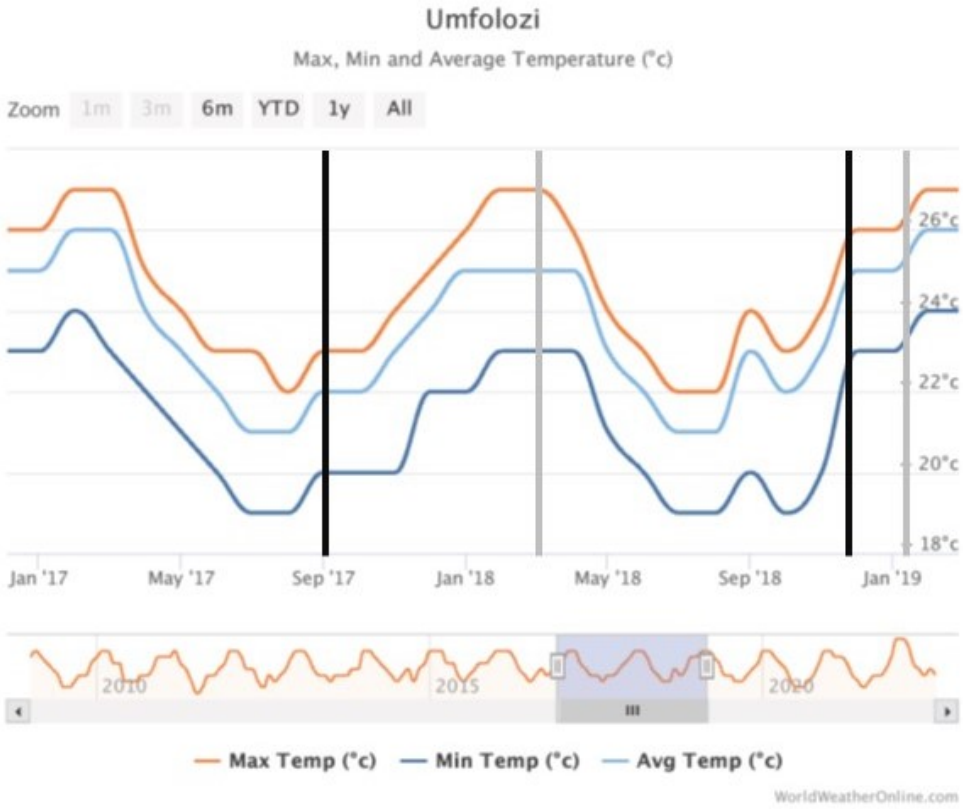
Appendix S1: Table S4 Taxonomy of detected arthropod items in 823 faecal samples, total read counts over all samples.

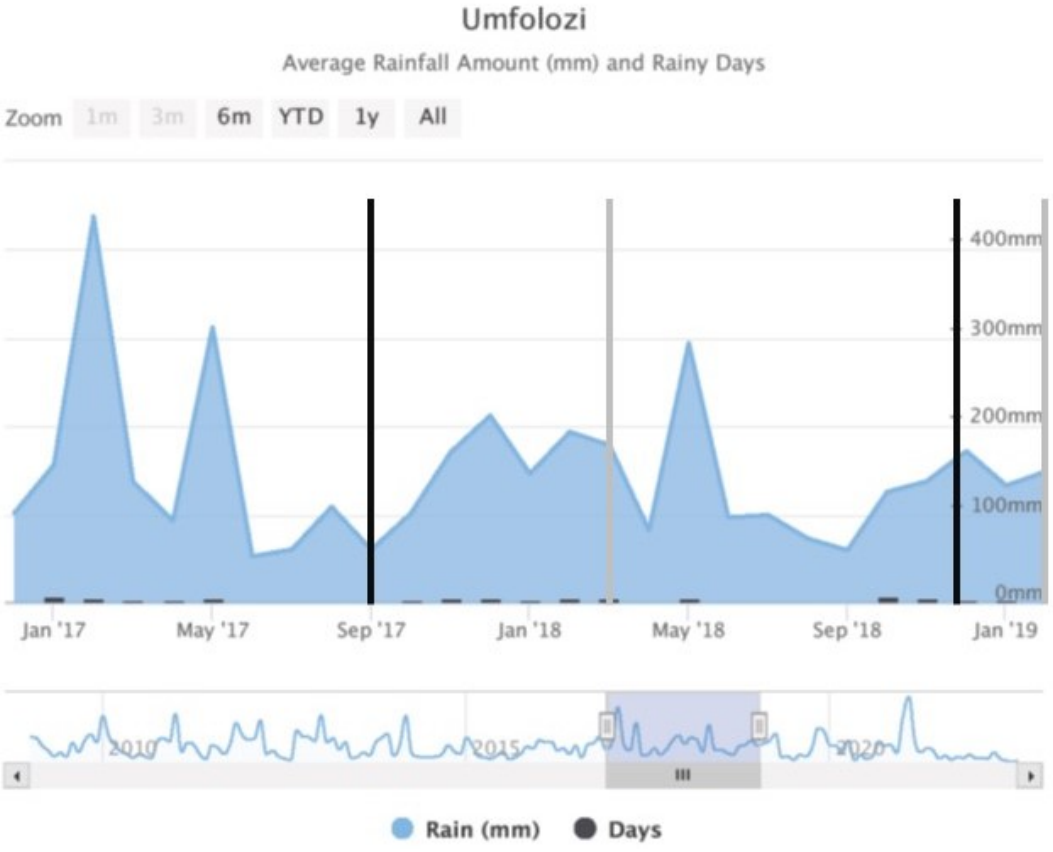
Order	Family	Count	Frequency
Araneae	Miturgidae	3734	7
Blattodea	Hodotermitidae	98835	438
	Termitidae	21619	46
Coleoptera	NA	59596	265
	Anthribidae	297	3
	Brentidae	12321	77
	Buprestidae	295	4
	Chrysomelidae	38345	300
	Curculionidae	8126	18
	Elateridae	321	2
	Hydrophilidae	212	3
	Scarabaeidae	572	6
	Tenebrionidae	1076	8
Diptera	NA	3069	7
	Cecidomyiidae	408	6

	Drosophilidae	2702	20
	Tephritidae	1645	9
Hemiptera	Alydidae	9594	131
	Cicadidae	1286	2
	Coreidae	250	7
	Pentatomidae	1012	26
	Pyrrhocoridae	197	7
Lepidoptera	NA	71362	372
	Erebidae	464	5
	Lasiocampidae	458	6
	Nepticulidae	1446	24
	Noctuidae	4724	52
	Nymphalidae	248	5
Mantodea	NA	178	2
	Mantidae	927	19
	Sibyllidae	69	2
Neuroptera	NA	180	5

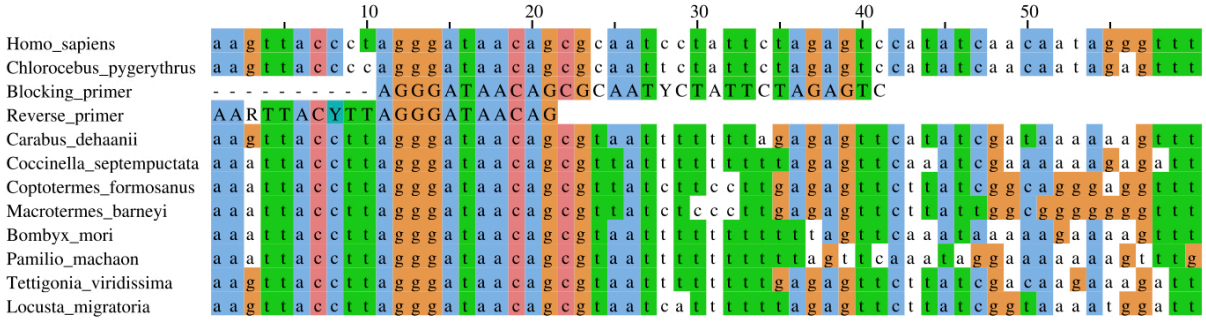
Odonata	NA	283	2
	Aeshnidae	247	2
	Libellulidae	149	2
Orthoptera	NA	611	6
	Acrididae	7196	81
	Gryllacrididae	170	4
	Gryllidae	3636	29
	Pamphagidae	783	11
	Tettigoniidae	1368	19
Thysanoptera	Thripidae	29	2

eDNA sampling period
Observation sampling period

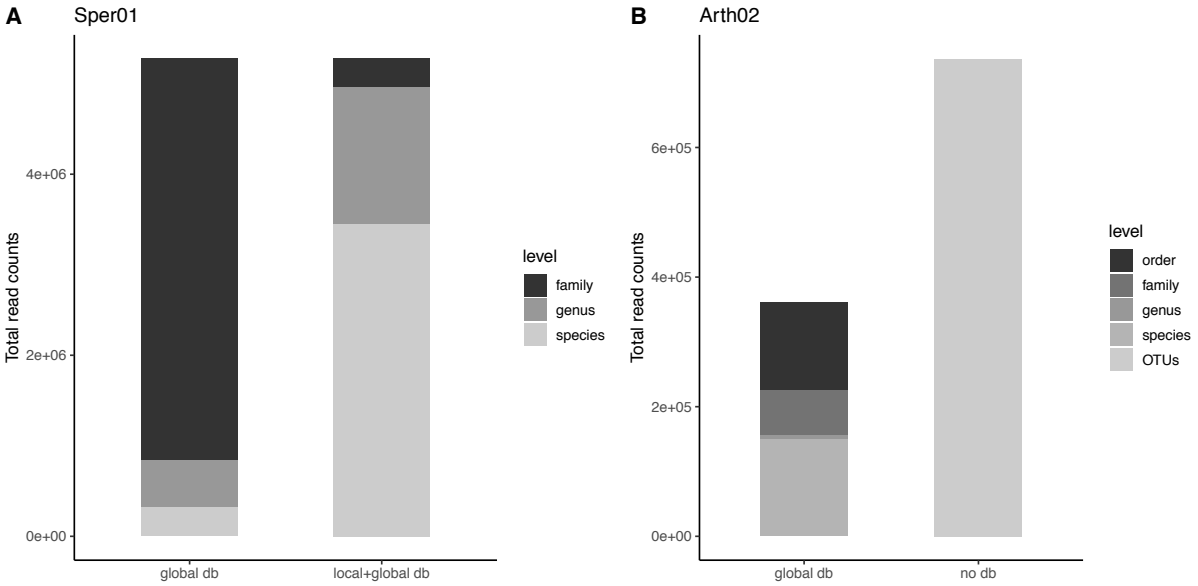




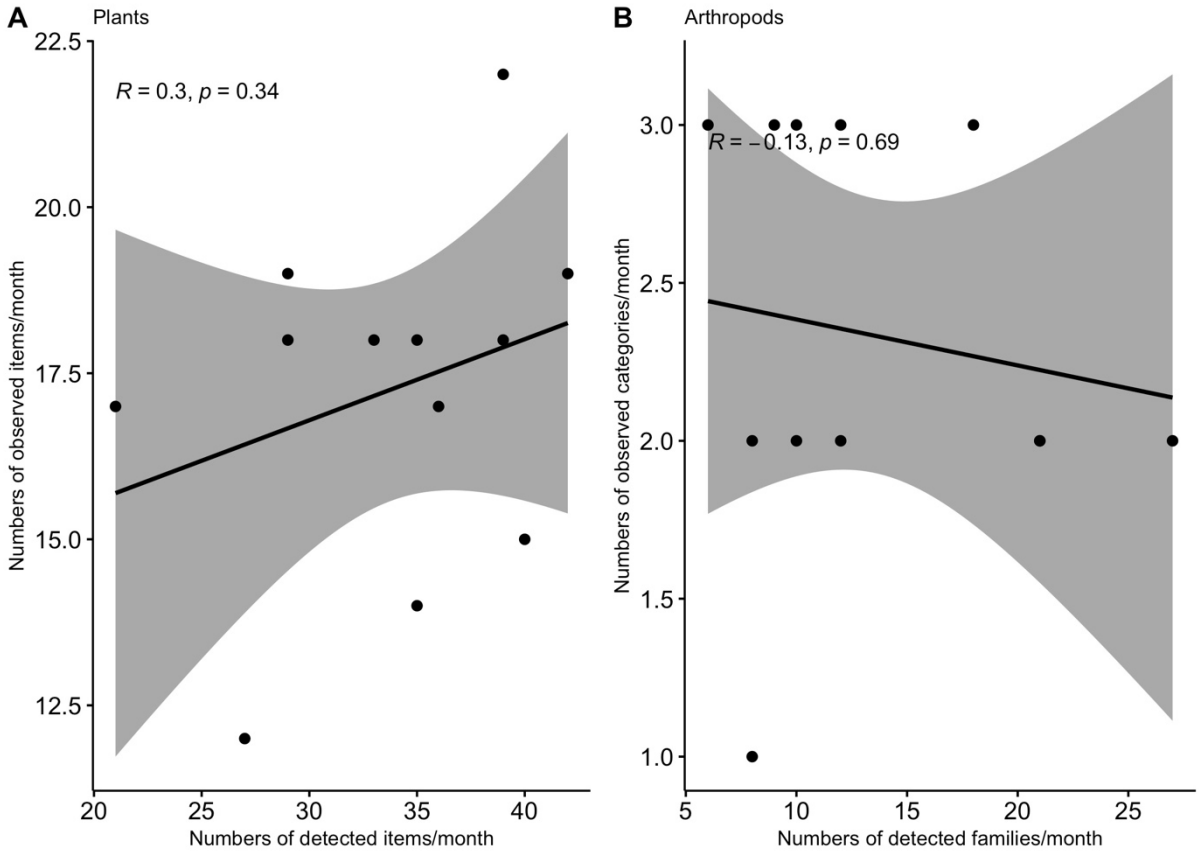
Appendix S1: Figure S1: Meteorological conditions have been assessed for the entire sampling period in terms of temperature (at the top; https://www.worldweatheronline.com/umfolozi-weather-averages/kwazulu-natal/za.aspx) and rainfall (at the bottom; https://www.weathersa.co.za/home/historicalrain). The period when focal screenings have been conducted is indicated in red, the one for eDNA sampling in green.



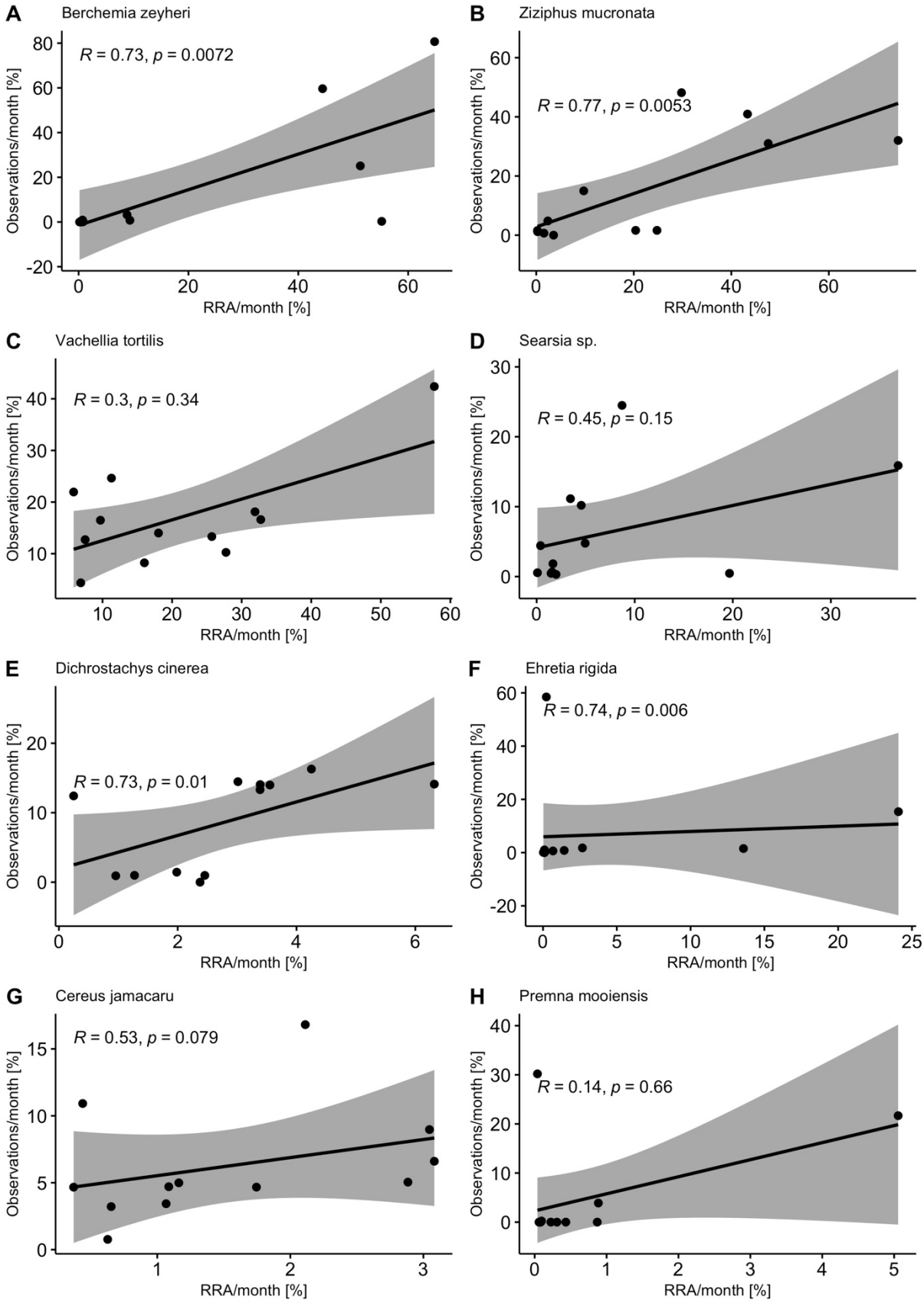
Appendix S1: Figure S2. Design of a blocking primer for vervet and human DNA for the Arth02 primer pair, allowing the amplification of target arthropod species. In this case, the 5'end of the blocking primer overlaps eleven nucleotides of the 3'-end of the reverse Arth02 amplification primer (Reverse_primer). The ideal position for the blocking primer is at the end of the PCR fragment where the variation between the sequence of the species to be blocked and the sequences of the target species is highest[1]. A C3 carbon spacer must be added on the 3'-end of the blocking primer to prevent its 3'-extension[124]. In addition of human, vervet, blocking primer and Arth02 reverse primer, representative arthropod sequences are also shown in the alignment.



Appendix S1: Figure S3. Stacked bar plots resuming the proportion of read counts assigned to different taxonomic levels by using particular database options for **A.** the Sper01 assay and **B.** the Arth02 assay.

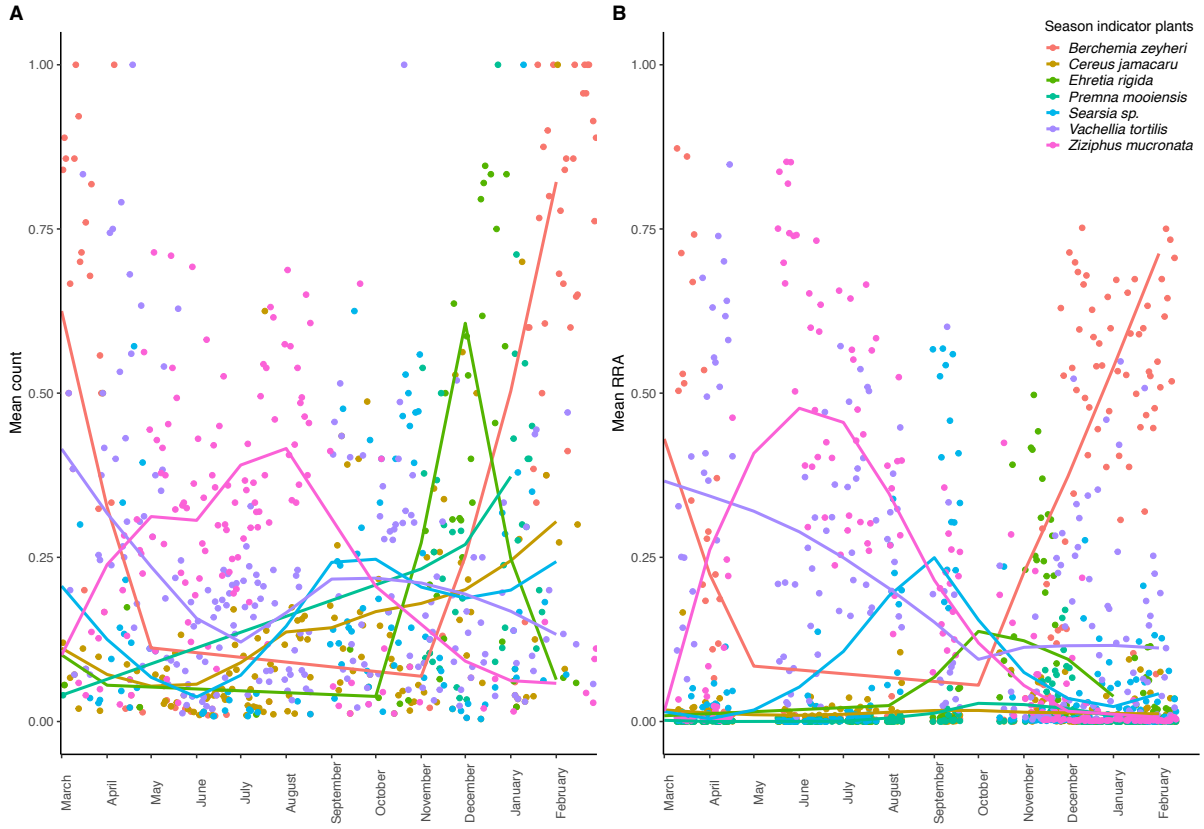


Appendix S1: Figure S4: Spearman rank correlations and coefficients based on total numbers per month of different dietary items as observed during focal follows and detected in faecal samples for **A.** plant data and **B.** arthropod data.

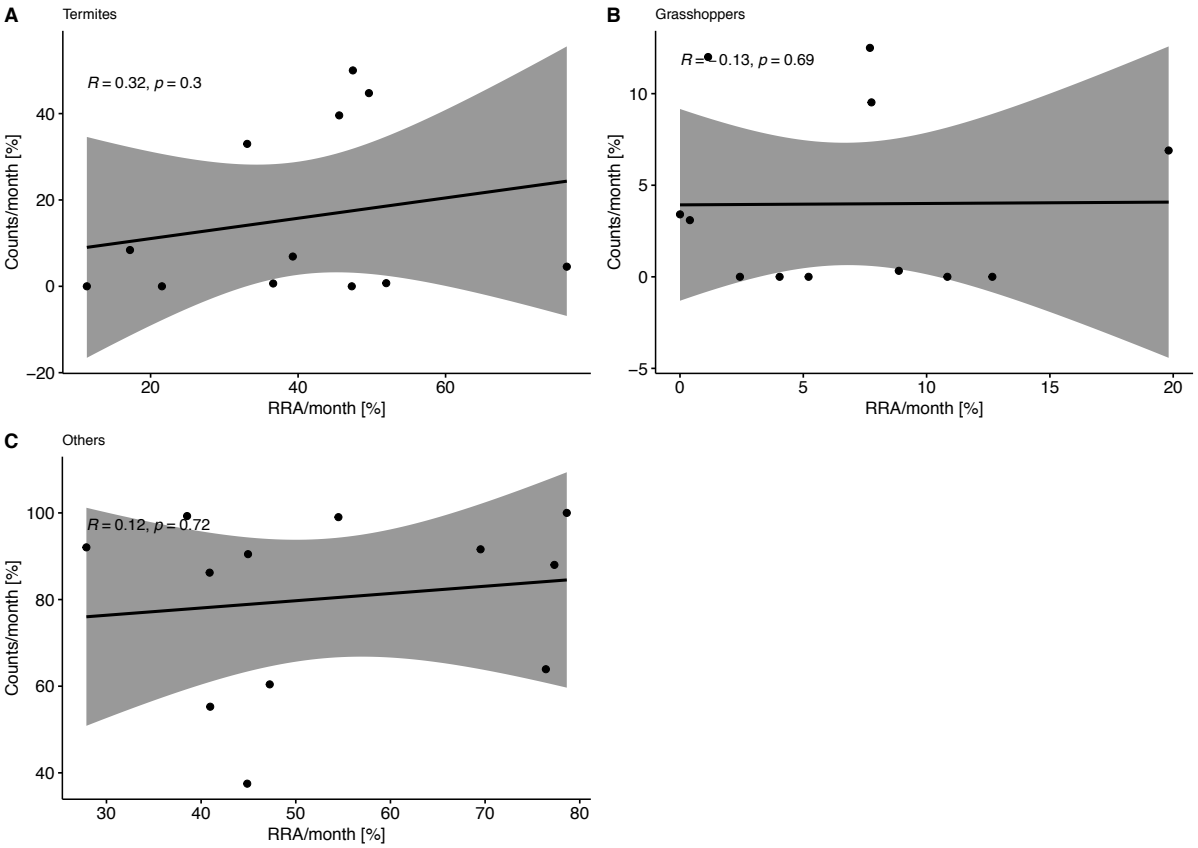


Appendix S1: Figure S5: Spearman rank correlations and coefficients based on mean count and RRA per month for all plant species present in both datasets and with a minimum of 350

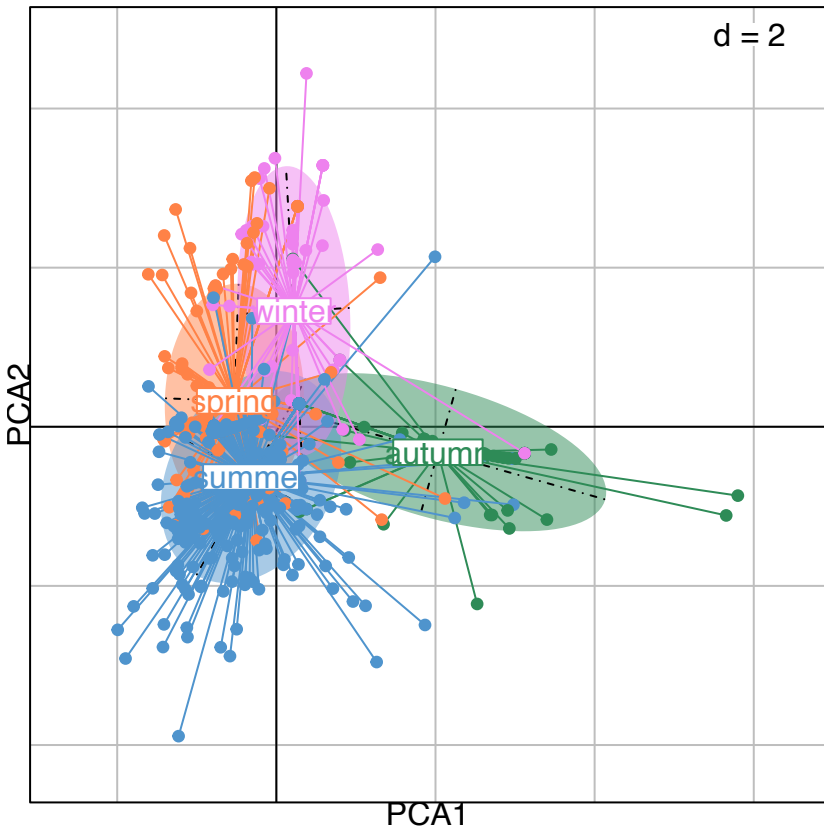
observations, with the exception of those that had identical metabarcodes and matched several species in the focal dataset (i.e. *E. crispa*/*E. undulata*/*D. dichrophylla* and *V. nilotica*/*C. decapetala*). The observed plant *V. tortilis* corresponds to *V. tortilis/sieberiana* in the eDNA dataset.



Appendix S1: Figure S6. A. Mean counts per date of seven consumed plant species observed during focal screenings that are indicators for seasons (shown are those plants with Indval value > 0.2, which were observed > 350 times and which do not share sequences with other observed species, Appendix S1: Table S3). **B.** Mean RRA per date of the same seven consumed plant species in faecal samples, only included RRA > 0.001. All species, except *C. jamacaru*, are also season indicator species (> 0.2) in the RRA dataset (Appendix S1: Table S3). The observed plant *V. tortilis* corresponds to *V. tortilis/sieberiana* in the eDNA dataset. Note that this representation serves to compare methods and that there are a number of additional indicator species and genus in the metabarcoding dataset not included here (Appendix S1: Table S3).



Appendix S1: Figure S7: Spearman rank correlations and coefficients based on mean count and RRA per month for all arthropod categories as shown in Fig. 2B.



Appendix S1: Figure S8: Principal component analysis (PCA), based on relative read abundances (RRA) of consumed arthropod families detected in faecal samples (R2 = 3.6 %). The four seasons are represented by different colours and the texts represent the centroids.

eDNA methods for intergroup dietary variation

Supporting Information

Table S1: Sample counts per group in summer: number of samples/number of individuals sampled (number of individuals present in group at the start of summer; 15th November 2018). Discrepancies between samples/individuals and individuals/group possible when age categories changed between 15th of November and the time of sampling (e.g. AK adult females and BD one individual sampled as juvenile F and adult F). Pearson's Chi-square tests confirmed the goodness of fit per age/sex category of sampled individuals per group for AK, KB and NH. For BD the null hypothesis of good fit was only confirmed when excluding the infant category. Age categories were defined as follows: infant <1 year, juvenile 1-2 years for females and 1-4 years for males (if not dispersed), adult 3 years for females and 4 years for males if they dispersed, otherwise 5 years.

Group	Total	Adult F	Adult M	Juvenile F	Juvenile M	Infants
AK	64/19 (26)	34/9 (8)	5/1 (1)	9/4 (5)	16/5 (5)	0/0 (7)
BD	78/38 (66)	42/16 (22)	16/11 (12)	4/4 (5)	16/8 (10)	0/0 (17)
KB	165/18 (22)	24/6 (6)	14/1 (1)	65/6 (6)	62/5 (5)	0/0 (4)
NH	65/28 (43)	26/11 (14)	4/3 (4)	9/3 (3)	26/11 (12)	0/0 (10)
Total	372/104 (157)	126/42 (50)	39/14 (18)	87/17 (19)	120/31 (32)	0/0 (38)

Table S2: Number of GPS locations sampled per group from the faecal sample spots and through observations, used to calculate the 50 % and 95 % isopleths of the groups' home ranges (*total* rows).

The *faecal samples* and *observations* rows indicate the number of these found within their respective isopleths.

GPS	AK50	AK95	BD50	BD95	KB50	KB95	NH50	NH95
Faecal samples	84	107	99	188	91	120	133	213
Faecal samples total	108	108	190	190	120	120	221	221
Observations	461	604	1280	1946	401	657	920	1436
Observations total	605	605	1957	1957	662	662	1445	1445

Table S3: Comparison of the predicted group (from the random forest model) and the observed group from which the faecal samples were originating for the test dataset (25 % of the observations). Observed samples are in columns and predicted samples in rows. Values on the diagonals represent the true positive of the models. Sensitivity (True positives/(True positives + False negatives)) and Specificity (True negatives/(True negatives + False positives)) are indicated for each group.

Confusion matrix	AK	BD	NH	KB	Sensitivity	Specificity
AK	7	1	0	0	0.54	0.96
BD	3	11	3	0	0.92	0.79
KB	3	0	7	0	0.70	0.90
NH	0	0	0	6	1.00	1.00

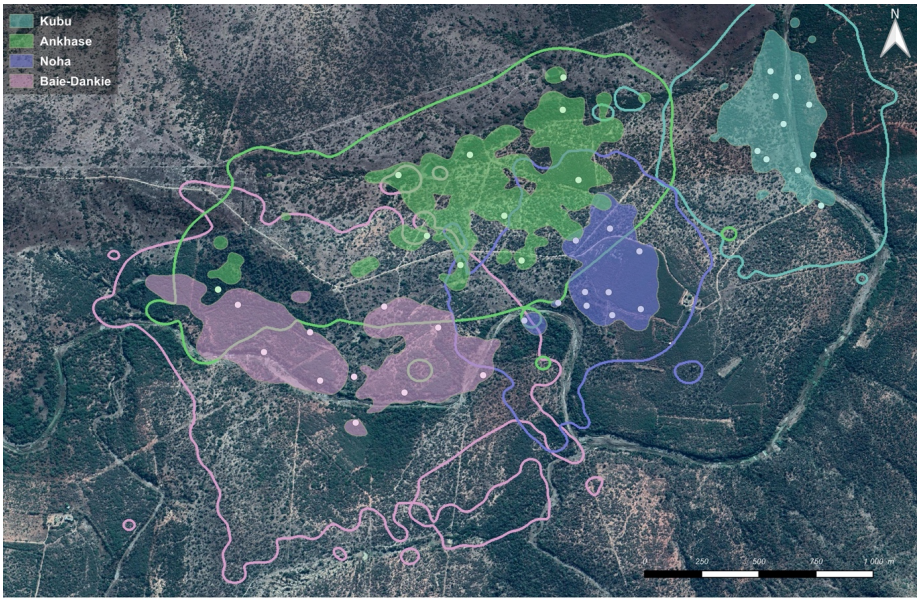


Figure S1: Home ranges (50 % and 95 % isopleths) of the four groups based on GPS data of sampling locations of faecal samples and observations. Points indicate the location of the 40 vegetation plots.

23

3

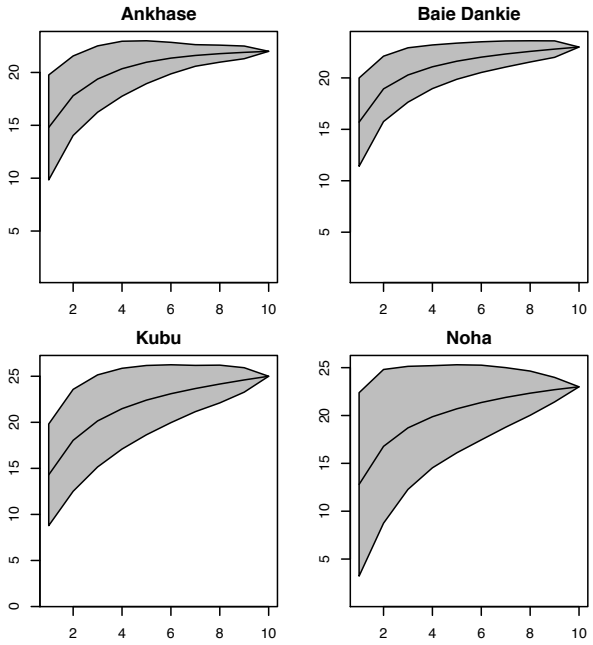


Figure S2: Species accumulation curves (SACs) of plant abundance in terms of coverage for the 40 vegetation plots. On the x-axis are the ten vegetation plots per groups and on the y-axis the cumulative number of plant species, grey shading indicates 95 % confidence intervals.

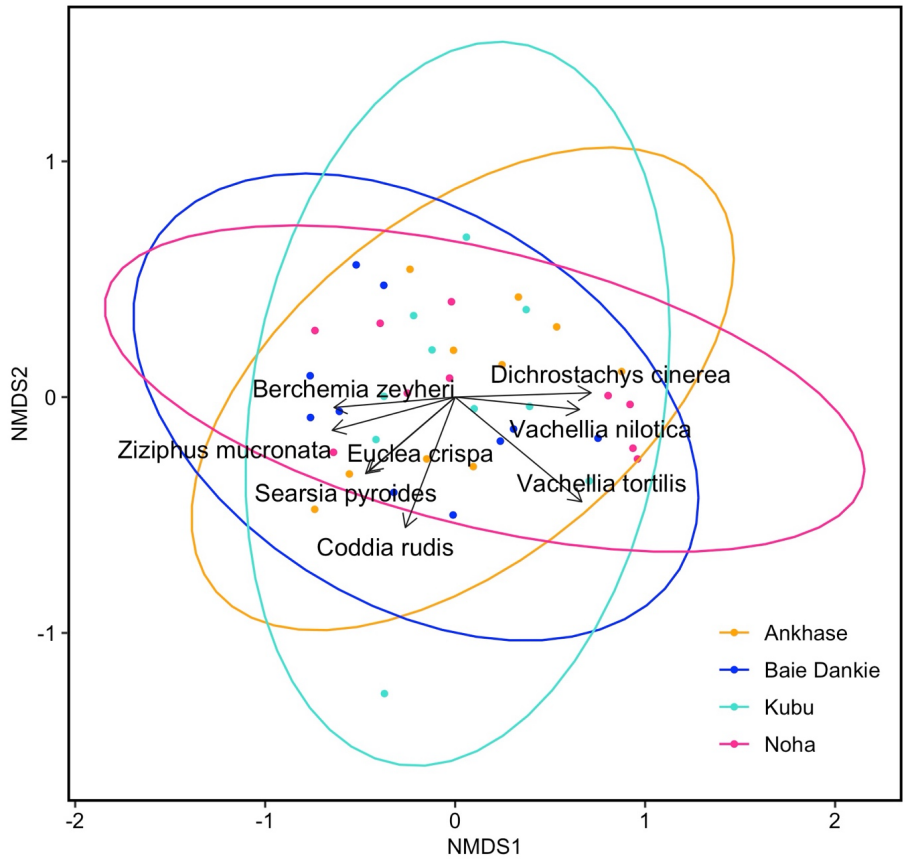


Figure S3: Nonmetric multidimensional scaling (NMDS) with *envfit* function of plant abundance in terms of coverage for the 40 vegetation plots taken in the respective 50 % home ranges of the four groups. Vectors of plants are shown for those with p<0.005. *PERMANOVA* indicates no significant difference in plant coverage between groups' territories (pseudo $F_{40} = 1.44$, $R^2 = 0.11$, $P = 0.12$).

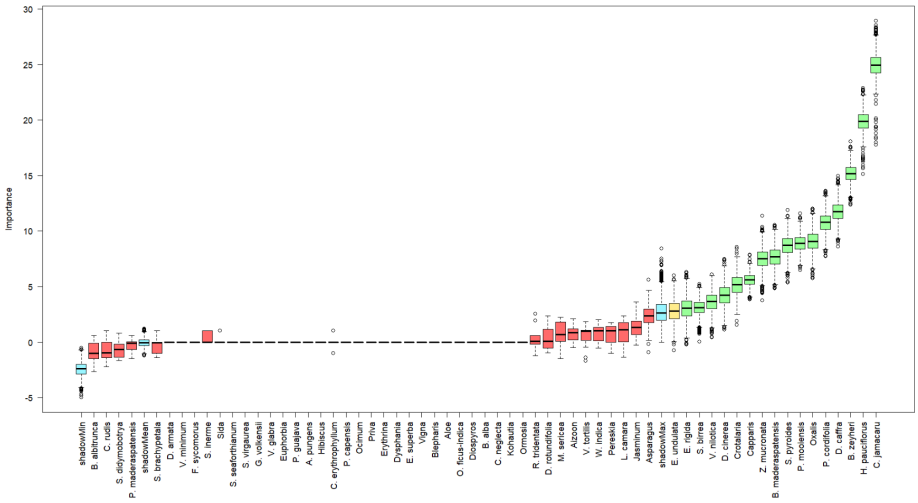


Figure S4: The plot shows the plant species that were selected by the random forest algorithms implemented in the *Boruta* R package as relevant features to explain dietary variation between groups. The higher the importance, the higher the group specificity of the corresponding species. Blue shows the minimum, average and maximum importance scores obtained by chance after 2000 random row permutations. Species in red were below the maximum threshold and considered not specific to any of the groups. *E. undulata* in yellow was very close to the maximum threshold and also not kept for further analyses. For species in green the group specificity was higher than that obtained by chance. Species above the threshold were corrected according to their respective abundance in the different groups' home ranges when available.

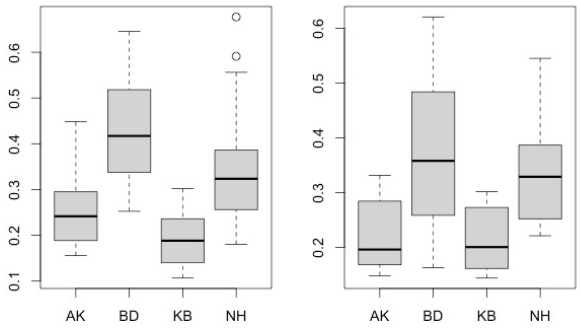


Figure S5: Boxplots displaying the dispersion from the centroids for each group, for a) all individuals (average distance to median: AK = 0.25, BD = 0.43, KB = 0.19, NH = 0.35; $p<0.001$) and b) only adult females (average distance to median: AK = 0.23, BD = 0.38, KB = 0.21, NH = 0.33; $p<0.005$).

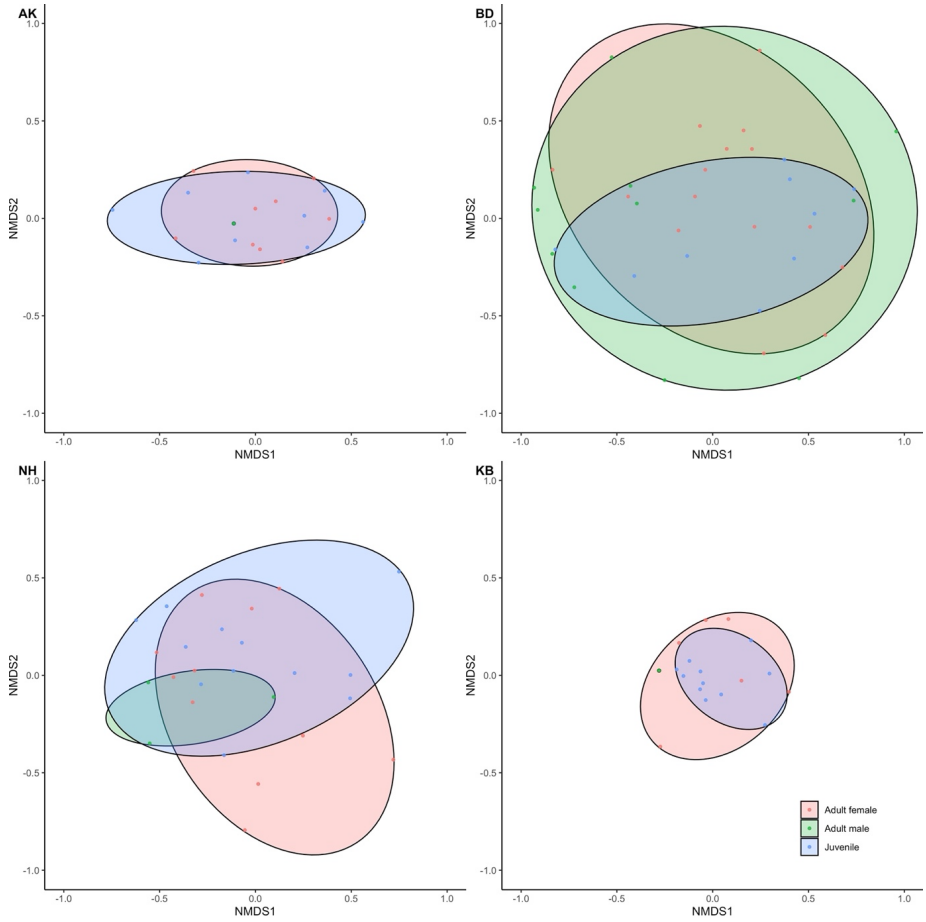


Figure S6: Nonmetric multidimensional scaling (NMDS), based on relative read abundances (RRA) of consumed plants detected in faecal samples aggregated per monkey in summer for all groups (Ankhase, AK; Baie Dankie, BD; Noha, NH and Kubu, KB) with variable sex/age with three factor levels. The colours represent female adults (red), male adults (green) and juveniles (blue).

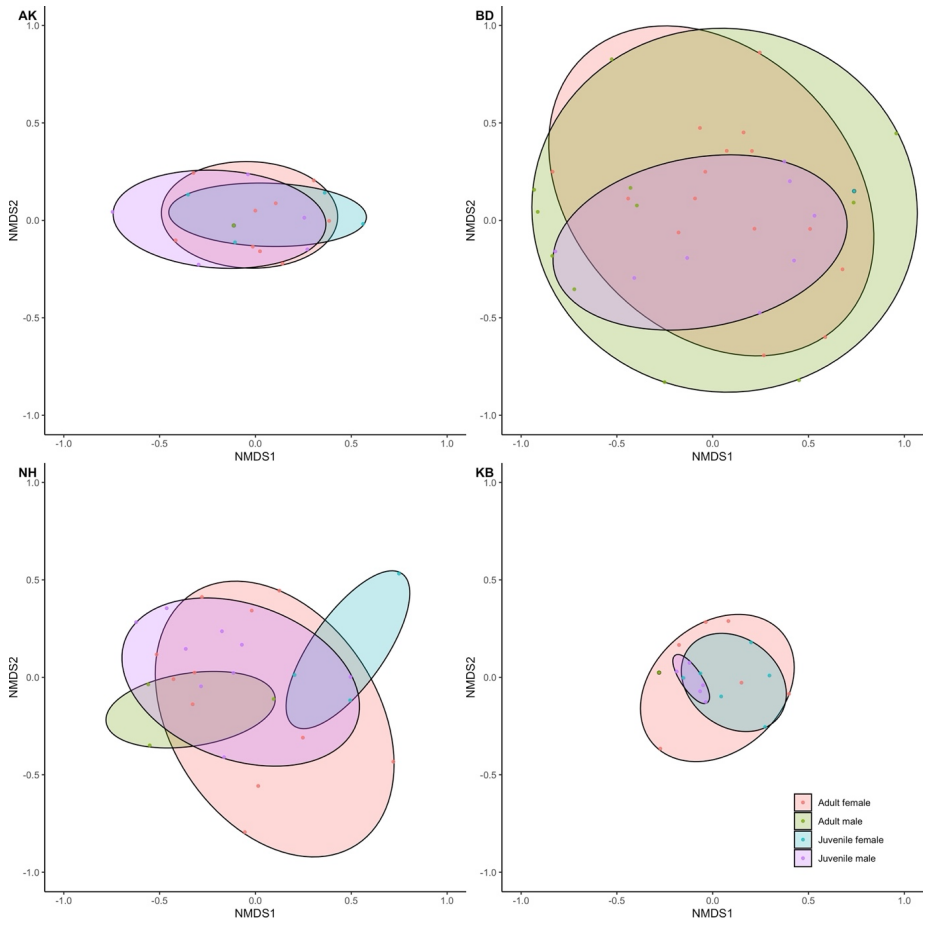


Figure S7: Nonmetric multidimensional scaling (NMDS), based on relative read abundances (RRA) of consumed plants detected in faecal samples aggregated per monkey in summer for all groups (Ankhase, AK; Baie Dankie, BD; Noha, NH and Kubu, KB) with variable sex/age with four factor levels. The colours represent female adults (red), male adults (green), female juveniles (blue) and male juveniles (violet).

Supplementary data Chapter 3

Table S1: Sample numbers per adult female, group and season. IDs of females are represented by the first four letters of their names. Different matrilines start with different letters.

Group	Monkey ID	Mother	Total	summer	autumn	winter	spring
Ankhase	*Ghid*	*Ghan*	44	12	7	9	16
	Ginq	*Gaga*	35	10	4	8	13
	Godu	*Gele*	22	2	2	8	10
	Gubh	*Gugu*	41	15	4	8	14
	Gugu	*Gaga*	18	7	2	3	6
	Mamo	*NA*	9	3	1	4	1
	Ncok	*Nkos*	3	0	0	0	3
	Ndaw	*Ndon*	17	5	1	5	6
	Ndik	*Ndon*	1	1	0	0	0
	Ndon	*Nkos*	32	6	5	6	15
	Nkos	*NA*	22	5	3	4	10
	Nyan	*Nkos*	29	5	4	12	8
	12	*TOTAL*	273	73	32	66	102
Baie Dankie	*Aapi*	*Asis*	20	2	5	6	7
	Asis	*NA*	10	2	1	7	0
	Eina	*Enge*	27	8	5	6	8
	Enge	*NA*	9	1	0	5	3
	Gese	*NA*	6	0	0	3	3
	Heer	*Haki*	17	1	4	7	5
	Hibi	*Hipp*	2	2	0	0	0
	Hipp	*Haki*	4	1	1	0	4
	Hond	*Heer*	15	3	2	6	4
	Lbli	*NA*	5	2	1	1	1
	Miel	*Mooi*	31	7	5	12	7
	Naal	*Numb*	1	0	1	0	0
	Nooi	*Numb*	21	8	1	7	5
	Numb	*NA*	7	2	0	2	3
	Nurk	*Numb*	24	5	4	7	8
	Obse	*Ouli*	4	2	0	1	1
	Oerw	*Ouli*	3	1	0	0	2
	Oort	*Ouli*	14	3	3	3	5
	Ouli	*NA*	18	2	3	6	7
	Pann	*Snor*	8	4	1	0	3
	Piep	*Prin*	22	4	3	6	9
	Potj	*Prin*	25	6	2	11	6

	Puol	*Prin*	17	3	2	10	2
	Rede	*Riss*	18	4	2	5	7
	Riss	*NA*	24	3	2	8	11
	Sari	*Snor*	2	0	0	0	2
	Siel	*Snor*	20	3	2	3	12
	Sirk	*Siel*	6	2	1	1	2
	Snor	*NA*	1	0	0	1	0
	29	*TOTAL*	383	99	58	123	133
Kubu	*Aara*	*Aare*	13	1	1	2	9
	Aare	*Amaz*	17	6	0	5	6
	Amur	*Amaz*	8	4	2	0	2
	Mara	*NA*	7	2	0	1	4
	Ness	*NA*	10	8	0	0	2
	Yalu	*Yeni*	8	6	1	0	1
	Yamu	*Yalu*	12	2	0	3	7
	Yara	*Yalu*	18	1	1	3	13
	Yelo	*Yeni*	2	1	1	0	0
	Yeni	*NA*	6	6	0	0	0
	Yuko	*Yeni*	10	2	0	3	5
	11	*TOTAL*	111	39	6	17	48
Lemon Tree	*Daen*	*Dian*	12	0	0	5	7
	Dais	*Dian*	17	1	2	6	8
	Deli	*Dian*	13	0	1	5	7
	Dewe	*Deli*	16	1	0	7	8
	Dext	*Deli*	8	1	1	2	4
	Dian	*NA*	11	1	1	4	5
	Digb	*Dian*	9	0	0	3	6
	Dore	*Dian*	18	1	1	7	9
	Lail	*Laur*	4	0	0	0	4
	Lanc	*Laur*	20	0	0	11	9
	Laur	*Lizz*	1	0	0	0	1
	Lill	*Lizz*	22	1	1	8	12
	Lizz	*NA*	12	0	0	7	5
	Loui	*Lizz*	20	2	2	5	11
	14	*TOTAL*	183	8	9	69	96
Noha	*Beir*	*Bela*	1	1	0	0	0
	Bela	*Bogo*	33	8	3	8	14
	Gaya	*Gene*	34	8	5	5	16
	Gene	*NA*	3	2	0	1	0

	Gran	*Gene*	28	7	2	8	11
	Gris	*Gran*	1	1	0	0	0
	Guat	*Gene*	3	1	0	0	2
	Lima	*Lhas*	9	2	1	3	3
	Prai	*Pret*	20	7	1	5	7
	Prat	*Prai*	1	1	0	0	0
	Pret	*Pari*	15	3	2	2	8
	Raba	*Roma*	16	4	3	5	4
	Renn	*Roma*	8	1	0	3	4
	Reva	*Roma*	9	4	0	3	2
	Rioj	*Roma*	12	1	2	6	3
	Roma	*NA*	1	0	0	1	0
	Rosl	*Roma*	3	3	0	0	0
	Upps	*NA*	21	3	3	5	10
	Xala	*Xaix*	24	5	4	7	8
	Xian	*Xaix*	32	7	7	10	8
	Xinp	*Xian*	2	2	0	0	0
	21	*TOTAL*	276	88	41	72	97

Table S2: Results of permutational multivariate analysis of variance (PERMANOVA with 9999 permutations and Bonferroni correction) for pairwise comparison between all five groups (Bray-Curtis dissimilarity matrix ~ Group + Seasons + Sample/Monkey_ID).

Source	**df**	**SumOfSqs**	**R2**	**PseudoF**	**p**
AK, BD					
Group	1	1.192	0.00552	4.3607	2e-04 ***
Season	3	20.841	0.09651	25.4144	1e-04 ***
Sample:Monkey_ID	40	14.461	0.06697	1.3226	3e-04 ***
Residual	610	166.741	0.77216		
Total	655	215.942	1.00000		
AK, KB					
Group	1	0.583	0.00473	2.4104	0.0198 *
Season	3	14.518	0.11790	20.0222	1e-04 ***
Sample:Monkey_ID	22	7.752	0.06296	1.4579	6e-04 ***
Residual	356	86.043	0.69878		
Total	383	123.133	1.00000		
AK, LT					
Group	1	1.692	0.01188	6.8653	1e-04 ***
Season	3	16.738	0.11748	22.6368	1e-04 ***
Sample:Monkey_ID	25	8.646	0.06069	1.4031	4e-04 ***
Residual	425	104.750	0.73524		
Total	455	142.469	1.00000		
AK, NH					

Group	1	1.052	0.00610	4.3402	2e-04 ***
Season	2	25.582	0.14842	35.1801	1e-04 ***
Sample:Monkey_ID	32	10.118	0.05870	1.3045	6e-04 ***
Residual	511	123.860	0.71860		
Total	548	172.361	1.00000		
BD, KB					
Group	1	1.047	0.00629	3.8176	6e-04 ***
Season	3	12.336	0.07415	14.9965	1e-04 ***
Sample:Monkey_ID	39	14.772	0.08879	1.3814	1e-04 ***
Residual	449	123.113	0.74001		
Total	493	166.366	1.00000		
BD, LT					
Group	1	2.992	0.01620	10.8139	1e-04 ***
Season	3	13.055	0.07069	15.7285	1e-04 ***
Sample:Monkey_ID	42	16.165	0.08753	1.3911	1e-04 ***
Residual	518	143.321	0.77605		
Total	565	184.680	1.00000		
BD, NH					
Group	1	1.846	0.00863	6.7994	1e-04 ***
Season	3	20.316	0.09492	24.9383	1e-04 ***
Sample:Monkey_ID	49	16.771	0.07836	1.2604	8e-04 ***
Residual	604	164.014	0.76631		
Total	658	214.032	1.00000		
KB, LT					
Group	1	2.608	0.02829	11.2201	1e-04 ***
Season	3	8.000	0.08679	11.4740	1e-04 ***
Sample:Monkey_ID	24	10.678	0.11584	1.9144	1e-04 ***
Residual	264	61.355	0.66562		
Total	293	92.177	1.00000		
NH, KB					
Group	1	0.440	0.00359	1.8445	0.0744
Season	3	13.723	0.11171	19.1543	1e-04 ***
Sample:Monkey_ID	31	9.577	0.07796	1.2936	0.0020 **
Residual	350	83.585	0.68039		
Total	386	122.849	1.00000		
NH, LT					
Group	1	2.652	0.01881	10.8038	1e-04 ***
Season	3	15.370	0.10897	20.8693	1e-04 ***
Sample:Monkey_ID	34	12.428	0.08811	1.4889	1e-04 ***
Residual	419	102.865	0.72931		
Total	458	141.046	1.00000		

Significance codes: *** 0.001; ** 0.01; * 0.05.

Table S3: Results of permutational multivariate analysis of variance (PERMANOVA with 9999 permutations and Bonferroni correction) for pairwise comparison between all five groups in summer (Bray-Curtis dissimilarity matrix ~ Group + Sample/Monkey_ID).

Source	df	SumOfSqs	R2	PseudoF	p
AK, BD					
Group	1	1.555	0.03896	5.4090	0.0005 ***
Sample:Monkey_ID	20	5.479	0.13726	0.9529	0.6321
Residual	103	29.612	0.74188		
Total	125	39.916	1.00000		
AK, KB					
Group	1	0.2384	0.00865	0.9681	0.4186
Sample:Monkey_ID	15	4.3001	0.15598	1.1641	0.1856
Residual	83	20.4393	0.74140		
Total	100	27.5683	1.00000		
AK, LT					
Group	1	0.5407	0.02424	2.0050	0.0020 **
Sample:Monkey_ID	14	4.1338	0.18529	1.0949	0.2688
Residual	56	15.1018	0.67691		
Total	72	22.3099	1.00000		
AK, NH					
Group	1	0.503	0.01351	1.7664	0.0971
Sample:Monkey_ID	18	5.230	0.14035	1.0194	0.4317
Residual	103	29.359	0.78779		
Total	123	37.267	1.00000		
BD, KB					
Group	1	0.4249	0.01432	1.5841	0.1507
Sample:Monkey_ID	20	5.4484	0.18368	1.0157	0.4392
Residual	74	19.8481	0.66912		
Total	96	29.6629	1.00000		
BD, LT					
Group	1	0.5028	0.02309	1.6287	0.0345 *
Sample:Monkey_ID	19	5.1929	0.23845	0.8853	0.8320
Residual	47	14.5106	0.66629		
Total	68	21.7782	1.00000		
BD, NH					
Group	1	1.210	0.03149	3.9537	0.0013 **
Sample:Monkey_ID	23	6.525	0.16980	0.9270	0.7248
Residual	94	28.768	0.74866		
Total	119	38.426	1.00000		
KB, LT					
Group	1	0.5148	0.04242	2.6043	0.0040 **
Sample:Monkey_ID	14	3.9973	0.32935	1.4443	0.0288 *
Residual	27	5.3374	0.43977		
Total	43	12.1368	1.00000		
NH, KB					
Group	1	0.1782	0.00650	0.6731	0.6842
Sample:Monkey_ID	18	4.8557	0.17697	1.0188	0.4317
Residual	74	19.5947	0.71414		

Total	94	27.4382	1.00000		
NH, LT					
Group	1	0.4915	0.02302	1.6202	0.0496 *
Sample:Monkey_ID	17	4.8372	0.22661	0.9380	0.6768
Residual	47	14.2572	0.66792		
Total	66	21.3459	1.00000		

Significance codes: *** 0.001; ** 0.01; * 0.05.

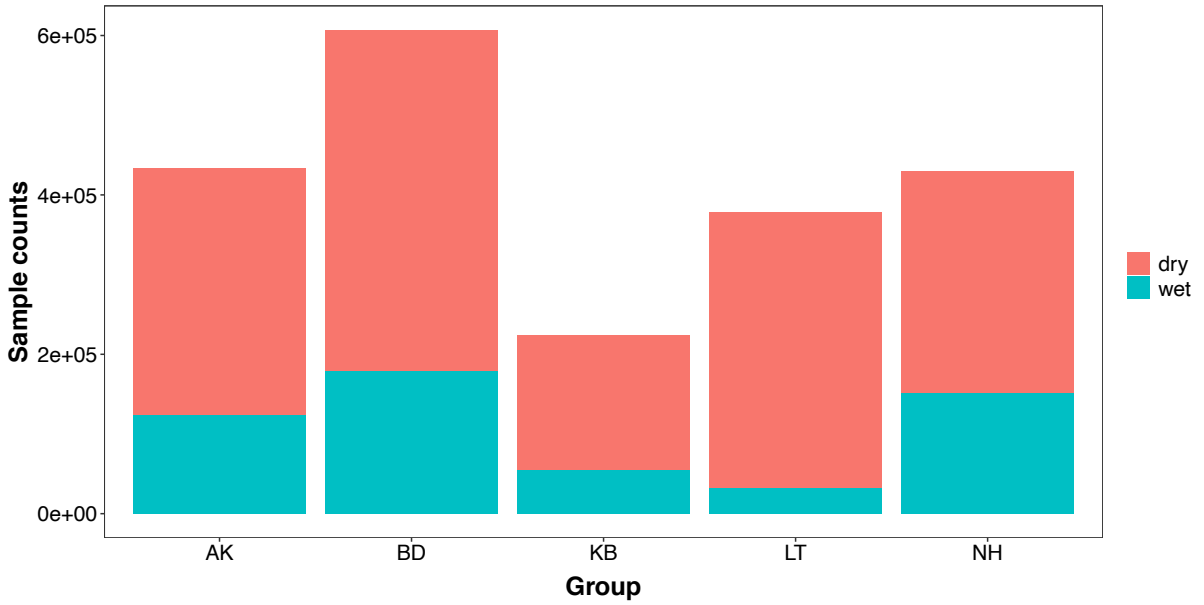


Figure S1: Cumulative sample numbers per group, wet (summer and autumn) and dry season (winter and spring).

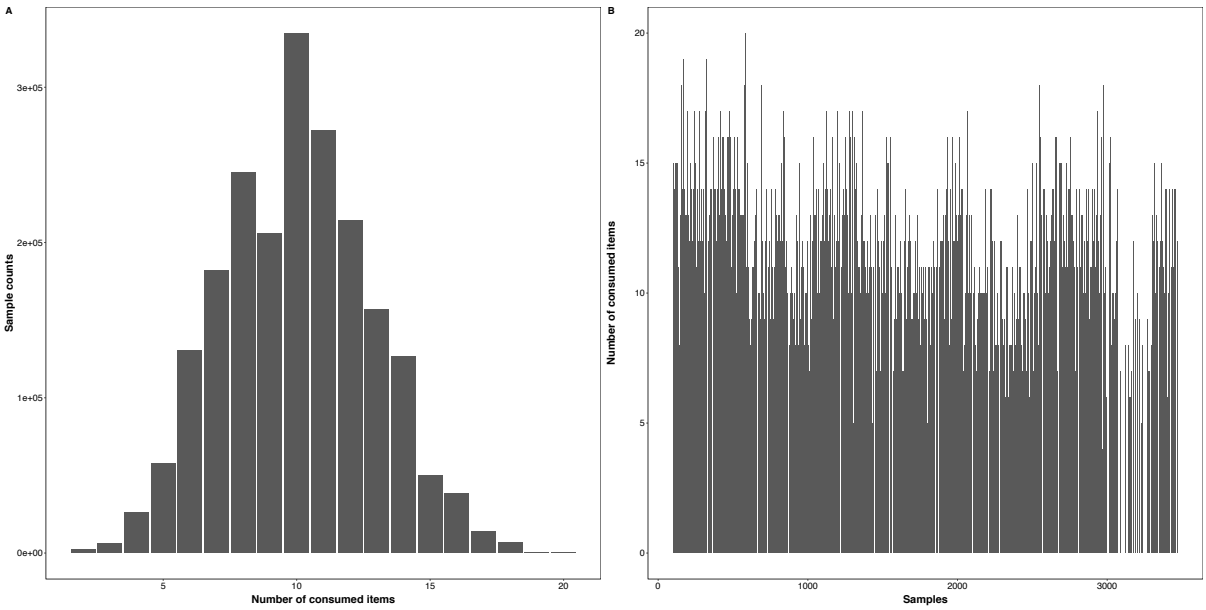


Figure S2: A) Distribution of number of consumed items (genus and species level). B) Number of consumed items per sample (genus and species level).

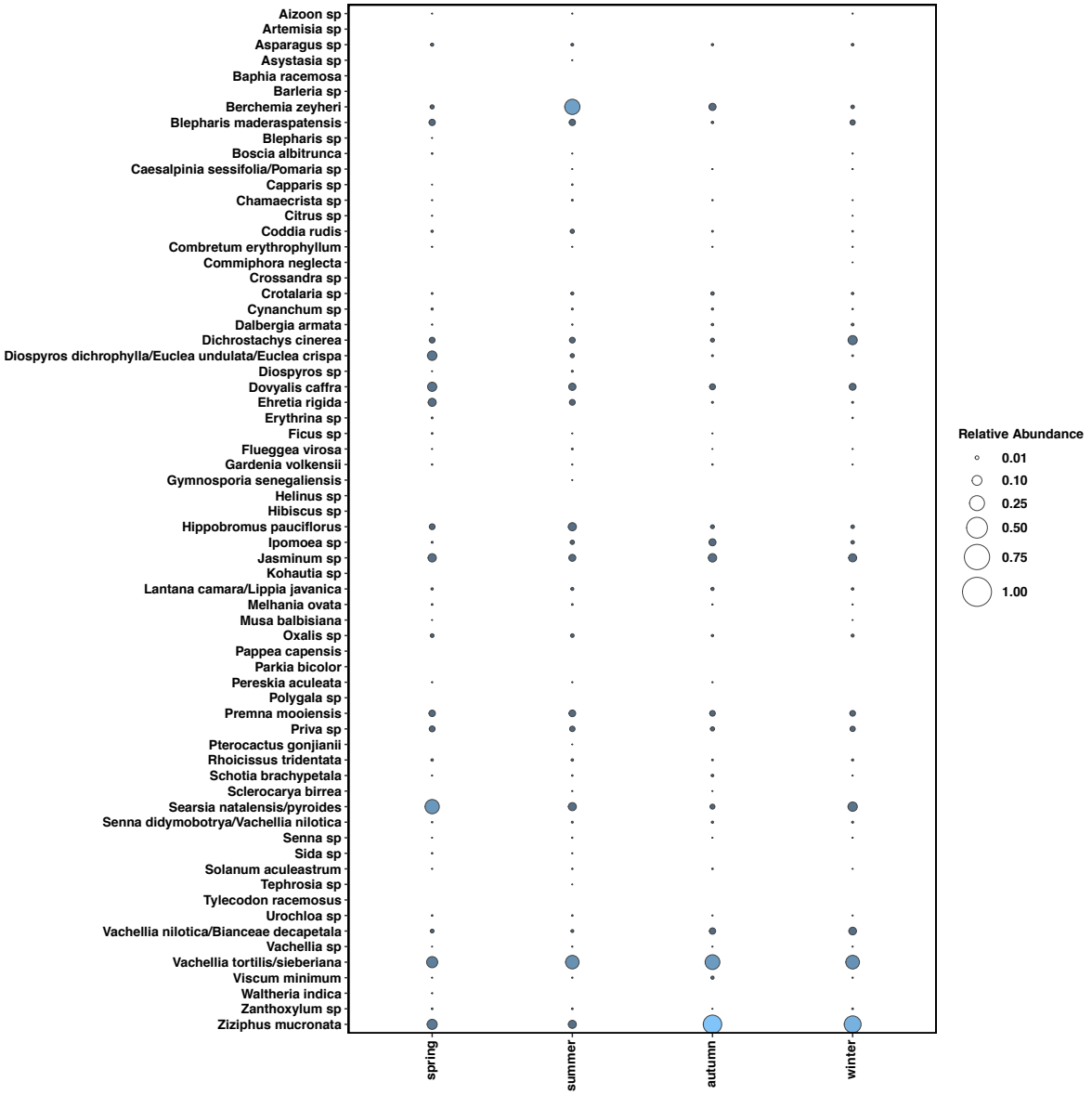


Figure S3: Relative read abundance (RRA) data of consumed plant genera and species cumulated by season. Plant items consumed only once were omitted from this graph.

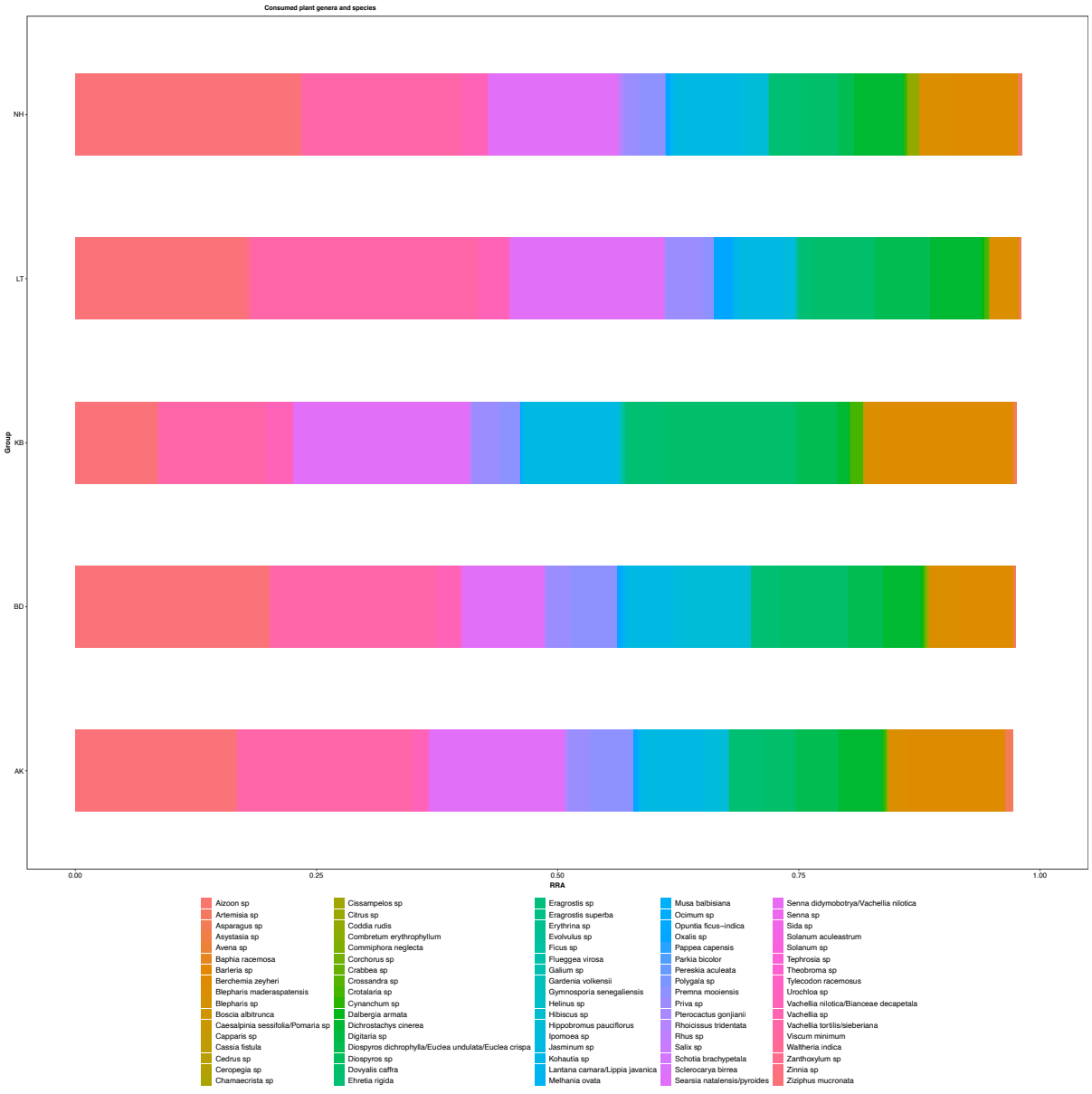


Figure S4: Relative read abundance (RRA) data of consumed plant genera and species cumulated by season. Plant items consumed only were included in this graph. Family-level assignations were not shown but included in the calculation of RRA, i.e. representing the missing data to attain 1. The 27 recorded plant families were: Anacardiaceae, Acanthaceae, Adoxaceae, Amaryllidaceae, Arecaceae, Apocynaceae, Asteraceae, Cactaceae, Chenopodiaceae, Commelinaceae, Convolvulaceae, Crassulaceae, Cucurbitaceae, Euphorbiaceae, Fabaceae, Hyacinthaceae, Malvaceae, Orchidaceae, Pedaliaceae, Poaceae, Rhamnaceae, Rosaceae, Rubiaceae, Salicaceae, Solanaceae, Verbenaceae, Vitaceae.

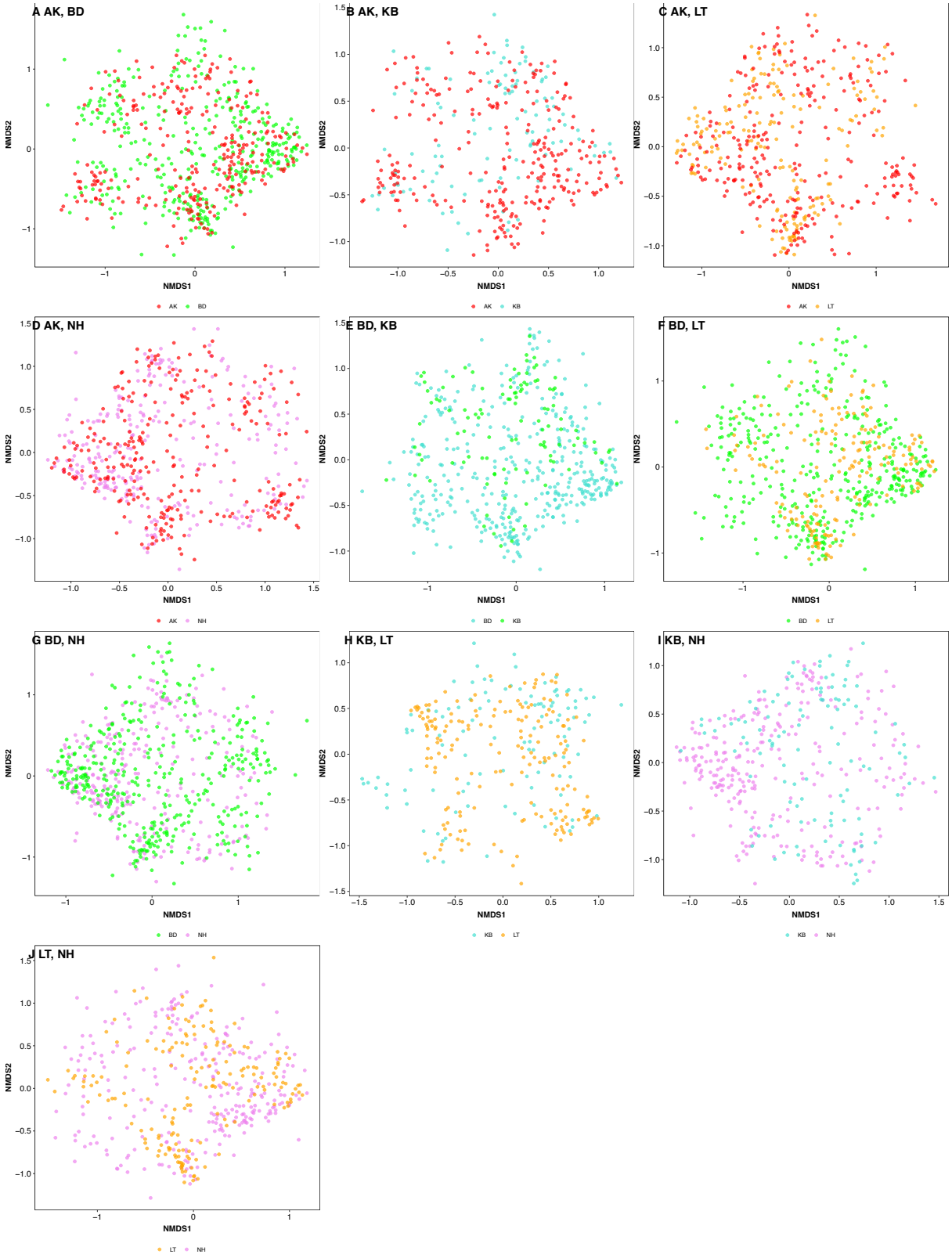


Figure S5: Non-metrical dissimilarity scaling (NMDS) of pairwise comparisons between all groups over all seasons using Bray-Curtis distances with non-aggregated relative read abundance (RRA) data. A) AK, BD; B) AK, KB; C) AK, LT; D) AK, NH; E) BD, KB; F) BD, LT; G) BD, NH; H) KB, LT; I) KB, NH and J) LT, NH.

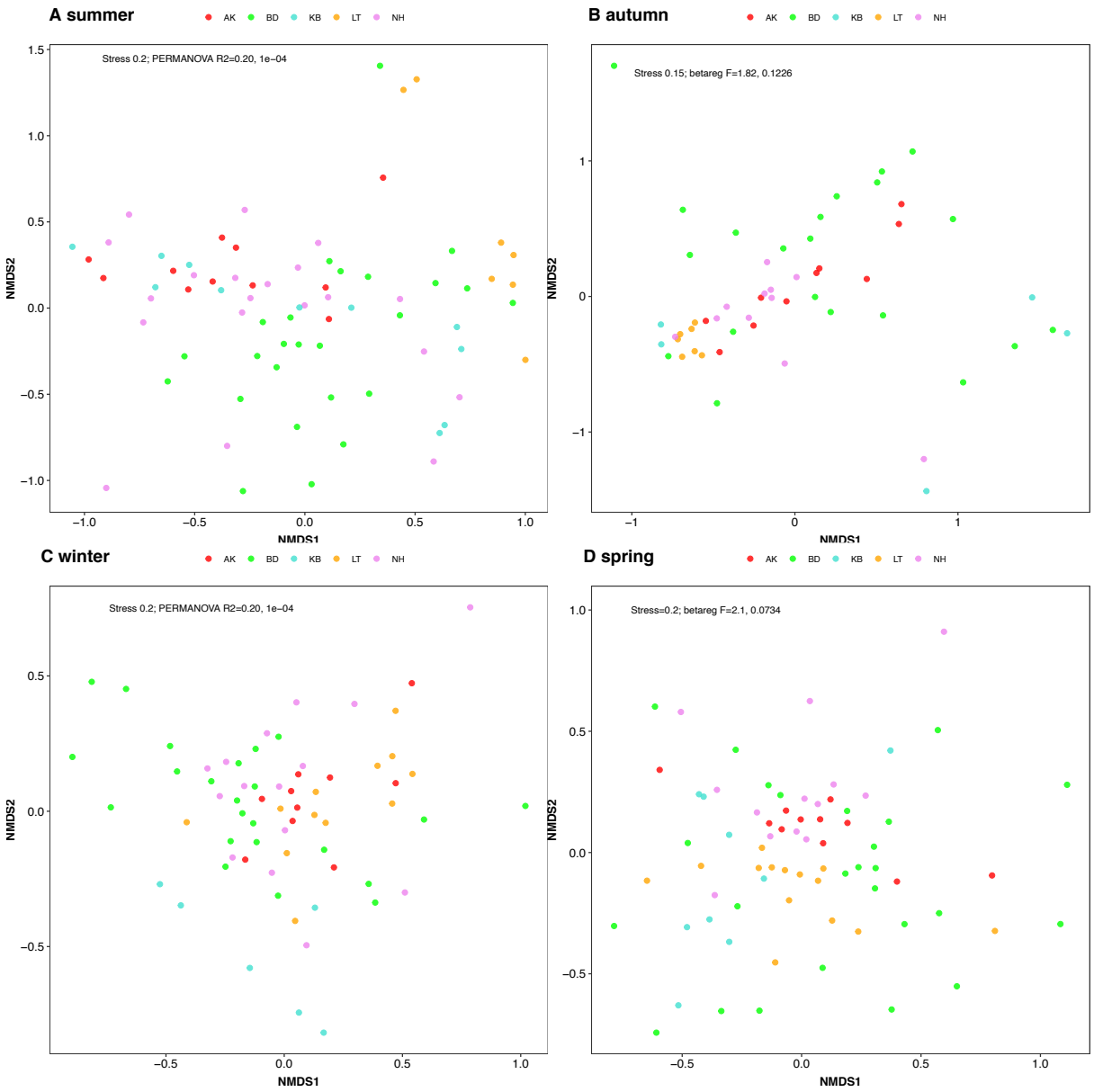


Figure S6. Non-metrical dissimilarity scaling (NMDS) using Bray-Curtis distances of relative read abundance (RRA) data aggregated by individual and season. For A) summer, B) autumn, C) winter and D) spring.

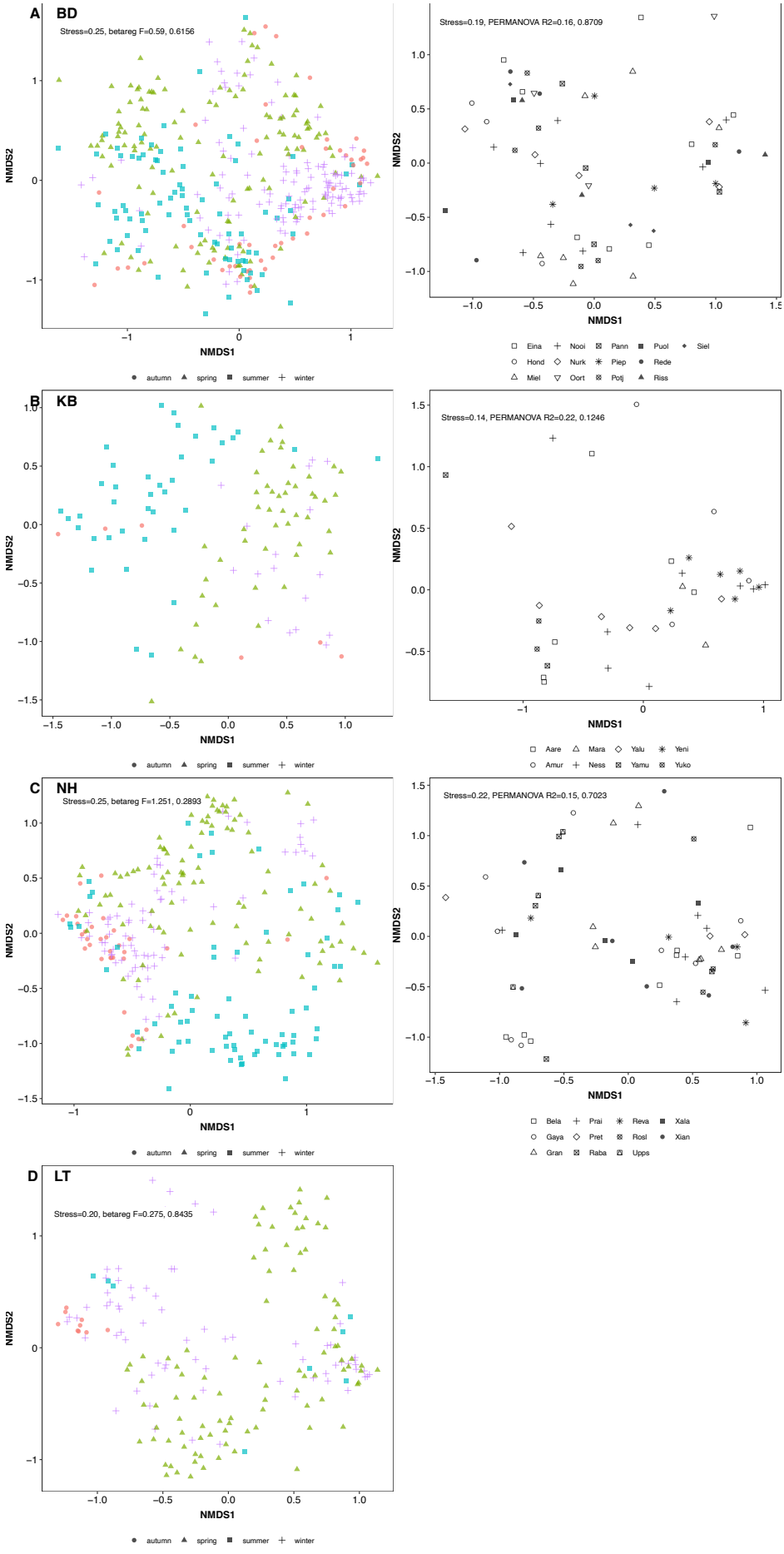

A
BD
Stress=0.25, betareg F=0.59, 0.6156
Stress=0.19, PERMANOVA R2=0.16, 0.8709
NMDS1
NMDS2
autumn spring summer winter
Eina Nooi Pann Puol Siel
Hond Nurk Piep Rede
Miel Oort Potj Riss
B
KB
Stress=0.14, PERMANOVA R2=0.22, 0.1246
Aare Mara Yalu Yeni
Amur Ness Yamu Yuko
C
NH
Stress=0.25, betareg F=1.251, 0.2893
Stress=0.22, PERMANOVA R2=0.15, 0.7023
Bela Prai Reva Xala
Gaya Pret Rosl Xian
Gran Raba Upps
D
LT
Stress=0.20, betareg F=0.275, 0.8435

Figure S7: Non-metrical dissimilarity scaling (NMDS) per group over all seasons using Bray-Curtis distances with non-aggregated relative read abundance (RRA) data (on the left) and in summer at the intragroup level with shape indication individual monkeys (on the right, except for LT due to low sample numbers). A) BD: 383 of 29 individuals; in summer, included are all individuals with ≥ 3 samples, different shapes represent different individuals (61 samples of 13 individuals), B) KB: 111 of 11 individuals; in summer, included are all individuals with ≥ 2 samples, different shapes represent different individuals (36 samples of 8 individuals), C) NH: 276 of 21 individuals; in summer, included are all individuals with ≥ 3 samples, different shapes represent different individuals (59 samples of 11 individuals) and D) LT: 183 of 14 individuals; in summer to low sample number (8 samples of 7 individuals).

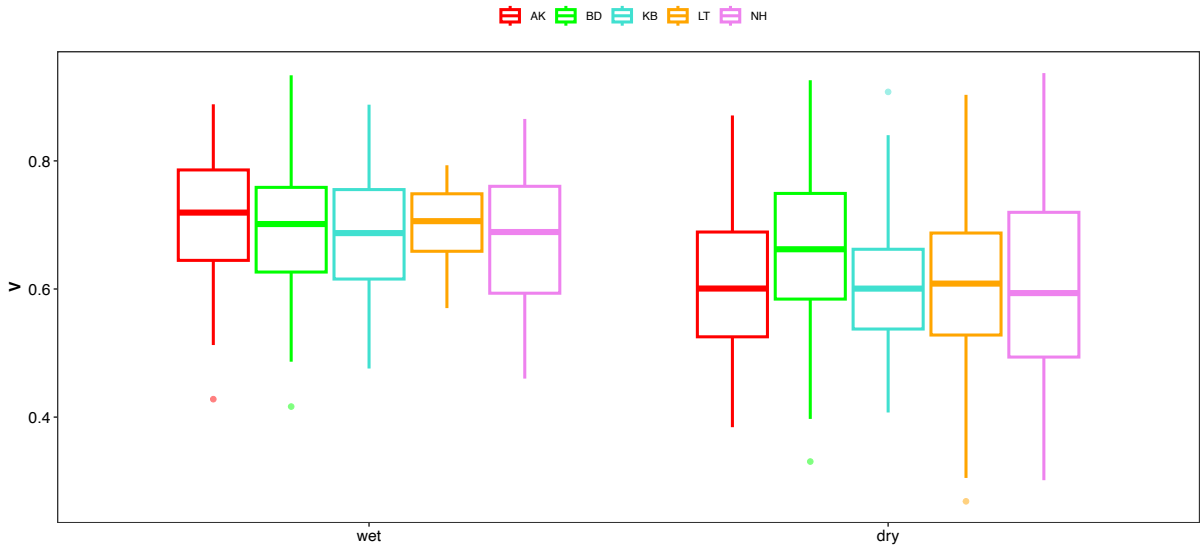


Figure S8: Inter-sample variation (V) compared to the group and wet/dry seasons. The wet season comprises summer and autumn and the dry season winter and spring. Colours indicate the social group.

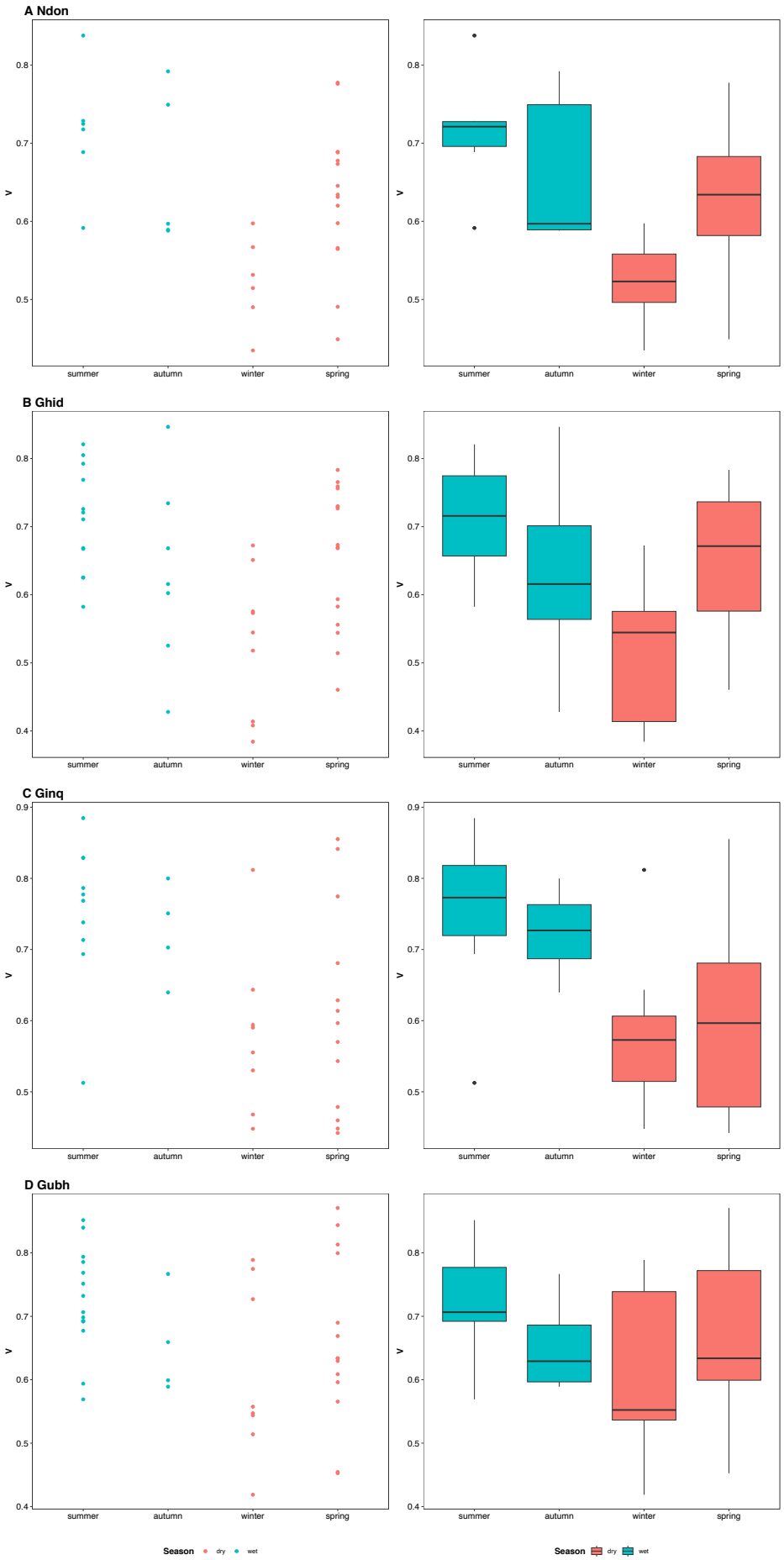

A Ndon
B Ghid
C Ginq
D Gubh
v
summer
autumn
winter
spring
Season
dry
wet

Figure S9: Inter-sample variation compared to the group was generally higher in wet season (summer and autumn) than in dry season (winter and spring). For illustration we chose individuals with high samples numbers and relatively evenly distributed among seasons; all individuals belong to the group Ankhase (AK). The plots on the left and right side represent the same data.

Project+: Baits for insect fishing

Judith Schneider[1], Erica van de Waal[2,3,4], Frédéric Boyer[5], Pierre Taberlet[5,6], Luca Fumagalli[1,7]

[1]Laboratory for Conservation Biology, Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland.

[2]Inkawu Vervet Project, Mawana Game Reserve, Swart Mfolozi, KwaZulu Natal, South Africa.

[3]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland.

[4]Centre for Functional Biodiversity, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

[5]Université Grenoble Alpes, CNRS, Laboratoire d'Ecologie Alpine, 38000 Grenoble, France.

[6]UiT – The Arctic University of Norway, Tromsø Museum, Tromsø, Norway.

[7]Swiss Human Institute of Forensic Taphonomy, University Centre of Legal Medicine Lausanne-Geneva, Lausanne University Hospital and University of Lausanne, 1000 Lausanne, Switzerland.

Chapter status: unpublished

Author contribution

JS and LF designed the study together with PT and FB. FB and JS developed the strategy for bait design. JS conducted the insect sampling and laboratory analyses, coordinated the sampling of eDNA samples and wrote the first draft of the manuscript, with input from all other authors.

Abstract

1. Ecological studies rely greatly on the measurements of abundances and biomass. Quantifying these based on DNA present in environmental (eDNA) samples is an important step towards more effective biomonitoring, but it is also complicated. For example, PCR-induced biases represent a major challenge for DNA metabarcoding studies.

2. The combination of targeted sequence capture and NGS (TCS) provides a promising alternative to other reduced-representation sequencing approaches that may result in more realistic biomass representations. Therefore we describe here the design process of a bait set for TCS targeting insect cytochrome c oxidase subunit I (COI) DNA.

3. The main challenge for the design of bait sets targeting broad taxonomic groups is to find the best compromise between specificity and flexibility.

Introduction

The use of environmental DNA (eDNA) has become a key component of various ecological studies, and it represents an important research domain for future biodiversity and conservation management[2,7,9]. While the current biodiversity crisis affects most if not all taxonomic groups, insects exhibit particularly high declines[125,126]. Thus monitoring of insect diversity and biomass is needed. Biodiversity data for insects still often relies on traditional sampling methods, i.e. the morphological identification of trapped specimens. These traditional methods are time- and labour-intense, and depend on taxonomic expertise, which is a declining skill and may e.g. not always identify cryptic life stages. The observation of feeding on arthropods is challenging and taxonomic resolution of observational studies mostly limited[127]. Macroscopic identification of insect remains in faeces or stomach contents is complicated, rare or soft-bodied taxa are consequently missed out[45]. Molecular methods like DNA barcoding have also been applied to achieve the identification of organisms. One important advantage of eDNA methods is that it is non-invasive and does not imply any harm for ecosystems[2]. While early eDNA studies were mostly based on PCR followed by Sanger sequencing or on qPCR assays, the advent of next-generation sequencing (NGS) technologies enabled the simultaneous sequencing of complex DNA mixtures in environmental samples, and therefore more exhaustive assessments[1]. DNA metabarcoding has recently provided valuable insights and thus highlighted the importance to consider arthropod diet components, e.g. in primatology[128,129]. Different types of environmental samples have been used to explore insect diversity illustrating the huge potential spectrum of applications (e.g. water[11], faeces[53], soil[130], bulk tissue samples that is mixtures of insects specimens[131], spider webs[132] or air[33] and plant surface substrates[12]).

However, while DNA metabarcoding[10] is widely used and enables valuable insights, the method implies certain biases, in particular concerning DNA amplification[39–41] and hence quantification issues[42,43]. Measurements of abundances and biomass are highly relevant in ecological studies, therefore quantifying DNA present in eDNA pools is needed yet at the same time extremely difficult. A number of studies resort to relative read abundances (RRA) as semi-quantitative proxy, e.g. for network approaches and food web analyses[19,50]. In order to monitor and report insect decline, reliable measurements of abundance, biomass and biodiversity are needed, and hence also quantitative assessments.

While whole-genome sequencing is possible nowadays, it remains demanding in terms of DNA quality, cost, time and data storage, and to obtain the requested information complete genomes are not always needed[58]. In order to find good compromises in terms of costs, sample numbers and loci, there are many different strategies of reduced-representation sequencing or genome-subsampling, aiming to obtain subsets of genomic data. Among those, the most often applied are restriction-site associated digestion methods (RADseq), which targets the sites next to restriction enzyme sites[133], transcriptome sequencing (RNAseq), which targets the expressed exome of tissues[134], and targeted sequence capture (from now on TSC, also known as hybridisation capture, sequence capture, target capture). The usefulness of RADseq and RNAseq for the analyses of DNA mixtures is limited, therefore we do not discuss these further (but see for a comparison between RADseq and sequence capture[135]).

The combination of targeted sequence capture and NGS has been suggested as convenient alternative to overcome PCR-induced biases and provide more reliable quantification measures[57,58]. The principle here is to create DNA or RNA baits that are complementary to

target sequences. Single stranded DNA is then hybridised to these biotinylated oligonucleotides and physically bound ("captured"). That means the baits are marked with biotin that can bind to biotin receptors located on magnetic beads[136]. Off-target DNA is subsequently washed away after the hybridisation of target DNA to the baits which is then processed for sequencing[59,60]. As such, TSC follows a different logic to achieve target enrichment than PCR amplification, since the increase in target sequences is achieved by reducing non-target sequences[61]. It has been shown that TSC enables sequencing of distantly related species by using only one bait species for probe design[137,138]. This is due to the relative acceptance of mismatches between probes and target sequences thus rendering the approach "flexible"[61]. For a comprehensive review of the TSC approach, see[58].

The first TSC assays were developed for biomedical research purposes[139,140] and are applied widely nowadays in this field. The applications of TSC, targeted sample types and taxonomic groups are diverse, e.g. used for the analysis of ancient DNA (aDNA)[141], museomics[138,142] and non-model organisms[137]. The main features of sequence capture that make it a suitable method for analyses of degraded and low-quality DNA samples are: 1) little input DNA needed, e.g. for species of conservation concern or museum specimens; 2) the flexibility of capture, i.e. the relative acceptance of mismatches between probes and target sequences; 3) that short fragments are targeted with short baits and 4) that coverage of targeted loci can be increased specifically compared to WGS[136].

The majority of studies is based on tissue or blood samples, in the biomedical field, and aims at single species or tagged samples in mixed libraries in phylogenomic studies. A few studies analysed multiple species assemblies based on bulk tissue samples of invertebrates[131,143] and fish larvae[144,145] and the methodology is tested and applied increasingly to eDNA samples. For

example, Wilcox et al. (2018) analysed water samples using one PCR-generated probe per targeted taxon, the results showing a positive correlation between target sequences and initial DNA after correcting for mitochondrial copy number variations among taxa[146]. Gauthier et al. (2020) compared TSC targeting COI to DNA metabarcoding on freshwater zoobenthos mock communities. They considered dissimilarity to bait (<15 %) as crucial for successful enrichment[147]. Seeber et al. (2019) designed RNA baits based on 38 selected mammal mitogenomes and compared this assay to classical PCR using water and sediment samples. Interestingly, while this study highlights the potential to amplify divergent and rare species with TSC, a quantitative mock community assay included in the study indicates possible biases regarding quantitative assessments depending on the divergence between baits and targets[148] (similarly to PCR-induced biases). Giebner et al. (2020) compare TSC to DNA metabarcoding on Malaise trap samples and mock communities for cytochrome c oxidase subunit I (COI) and 18S[149]. Beaudry et al. (2021) designed a 16S rRNA bait set for microbial communities, tested on faecal samples and mock communities[150]. Successful characterisation of *Larix* population dynamics (~6700 years ago) using lake sediment cores, show the utility of TSC for genomic analyses of *ancient* eDNA[151].

A number of studies have developed bait sets for insects, mostly per insect order, using ultra-conserved elements (UCE), see references in[152] and more recent for Psocodea[28], Hemiptera[154], and Coleoptera[155]. These markers have the advantage of being conserved across distant taxa, thus increasing the capture success, however, offering limited taxonomic resolution when applied simultaneously to closely related taxa. Here the focus is not the phylogeny or phylogeography of restrained taxonomic groups but DNA mixes with the prior aim of universal detection i.e. identification of insects, therefore restricting the genetic target

regions to standard markers traditionally used for species identification and thus present in public sequence databases. The *Folmer* region is a fragment of COI that has been used widely in animal DNA barcoding[156,157]. This fragment combines the advantages of conserved priming sites and highly variable regions, as it is protein-coding, and allows hence for species-level identifications in many cases. We describe here the design strategy for a set of baits targeting the *Folmer* region across the taxonomic group of Insecta and outline an experimental design to assess its specificity, sensitivity and quantification potential.

Bait Design

The capture with baits of TCS protocols can be done array-based or in-solution. However, in-solution sequence capture has become predominant[58,158]. There is a multitude of convictions concerning the choice of target loci. It goes from genomic, e.g. single nucleotide polymorphisms (SNPs) or UCEs, to genetic approaches targeting few genetic regions; there is no consensus over coding vs. non-coding regions either, and it depends on for example the research question and sample numbers. For further discussion about marker types and their usefulness in phylogenetic studies, as well as an interesting concept of a composable bait set, for frogs in this case, see[159]. As for loci, there are many different approaches for bait design, the choice depends ultimately on each project (research question, budget, available genomic resources, etc.). For example, if reference genomes are available, these can be used for the design of baits for the same species or closely related species[137,138,148]. That also allows for the development of TSC assays for ancient DNA[59,141]. If such genomic resources are unavailable, the first step of a design can be RADseq to obtain sequences as preprint for RNA baits (RADcap[160]) or that are directly transformed to biotinylated RNA-baits (HyRAD[142,161,162]).

In a similar logic, other studies have used PCR products as probes instead of RADseq output[144–146], an approach also suited for non-model species[163]. The mentioned designs are suited for single species or a group of closely related species. Attempts to design baits for broader taxonomic groups can either combine different sets of baits designed per taxonomic unit[148] or often draw on multisequence alignments, for a comprehensive review of bait design strategies and all other study steps, see[136]. The "wide" approach of Giebner et al. (2020) is tempting: to target two genes for a broad range of phyla in one bait set with a design based on a limited number of publicly available sequences. They use 4,970 sequences of 10 phyla for mitochondrial COI and 300 sequences of 20 phyla for nuclear 18S, which are clustered and then aligned to serve as input for BaitFisher[164]. However, the results point out that the capture success in terms of coverage is limited indicating that a too wide design of baits fails to capture all present organisms. And yet, the experimental design of the study has a flaw which impedes assessing the full potential of the baits, i.e. while the baits are designed for a range of phyla, the positive control includes only arthropod species[149].

The aim of the present study is to design a set of baits that, preferentially unbiased and completely, captures mitochondrial COI DNA of the whole taxonomic group of Insecta. We dismissed the approach to design baits or cluster per taxonomic subgroup, for example per insect order, since we wanted to rely on sequence diversity for the targeted genes and not on the taxonomic system. We downloaded all available insect sequences for the target gene in GenBank (~2.4M for COI). We annotated these sequences to keep only *Folmer* fragments of ~650 bp, resulting in 1,258,592 dereplicated sequences. There are a number of softwares that can help for the design, e.g. BaitFisher[164], that uses multisequence alignments as input to create bait sets depending on user-defined parameters, or R packages as SupeRbaits[165].

Although the latter seems a promising tool in principle, it is not practical for our purpose because it considers every input sequence independently and would thus design redundant baits. In addition, these tools were not useful for our data since they rely on multisequence alignments and we had too many sequences for alignment. Instead the bait design was done following the methodology of Beaudry et al. (2021). For this final design, input sequences are clustered with USEARCH[166] and one centroidal sequence per cluster retained. For the centroidal sequences, baits are designed, which are then once more clustered with USEARCH, again retaining one representative bait sequence per cluster[150]. The main parameters to choose for the design concerned bait length, tiling and clustering threshold. While the standard recommendation for bait sets by a company such as Daicel Arbor Biosciences (Ann Arbor, MI) is 80 bp baits, 2x tiling and 95 % clustering, we had to find the best compromise between number of baits and efficiency of the bait set. The tiling density indicates how much the baits overlap. The higher the tiling, the higher the coverage of targeted regions. That means higher tiling density is better for degraded DNA. The clustering threshold describes the sequence identity cut-off used for a group of related sequences for which to keep one representative sequence. The challenge is to ensure, on the one hand, that all genetic divergence is covered equally by baits to avoid capture biases or non-capture. No sequence should be more divergent than the chosen threshold. And on the other hand, to prevent the capturing of off-target DNA if the chosen threshold is too low. After assessment of diversity scores and sequence similarities in the compiled sequence data, chosen parameters were 80 bp baits with 90 % overlap and 80 % sequence identity. The application of these parameters resulted in a bait set of ~56,626 baits, which is ~92 baits per target nucleotide (pers. communication Brian Brunelle). To limit non-target capture of organisms likely present in our samples, the baits were checked against the RefSeq bacterial database as well as primate and

bat genomes. The filtered sequences were then send to BioCat GmbH (Heidelberg, Germany) for final design and synthesis of a *myBaits* Hybridization Capture Kit (40-60K).

Experimental design

Mock community preparation

For Phase I of the experiment (see Table 1), we will analyse artificial mock communities of tissue-derived insect DNA. Insect specimens were identified morphologically and total genomic DNA was extracted using the Qiagen DNeasy Blood and Tissue DNA Purification Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. To confirm morphological identification, all DNA extracts were sequenced on both strands using standard Sanger sequencing technology at Microsynth AG (Balgach, Switzerland) and checked manually for errors using MEGA version 10.1.0[167]. PCR reactions were done with primers LCO1490/HCO2198[157] following the conditions of[53]. Amplification success and fragment size was verified on a 1.5 % agarose gel stained with ethidium bromide before purification of the PCR products using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). In addition, the extractions were quantified using real-time qPCR targeting mitochondrial DNA (ARTH02 primers[1], conditions as for DNA metabarcoding, but for 45 cycles plus 10,000 fold diluted SybrGreen, Thermo Fischer Scientific, USA). Four different tissue-derived insect mock communities of known mitochondrial DNA concentrations will be prepared: 10 species equimolar, 10 species variable concentration, 40 species equimolar, 40 species variable concentration. In addition, non-insect taxa will be added to MCs to test for the amplification/capture of off-target DNA (e.g. gastropods, annelids and frogs).

eDNA samples

For Phase II of the experiment (see Table 1), eDNA samples will be included. eDNA extracts of scat samples from different organisms are available, disposing different dietary characteristics:

Vervet monkeys (*Chlorocebus pygerythrus*), an omnivorous species but mostly feeding on plants and occasionally on arthropods[168]. Samples were collected in at the iNkawu Vervet Project (IVP), Mawana game reserve, South Africa. Based on the results of Brun et al. (2022, i.e. Chapter 1 of the present manuscript), samples could be chosen showing high diversity of arthropod species with DNA metabarcoding.

Various bat species (see Table S1), presumed insectivorous and thus limiting off-target capture and likely representing the most diverse insect detection. Samples have been collected in the region of Grenoble, France, from August to September 2018.

A single macaque (*Macaca fascicularis*) living in a separate enclosure at the *Bioparc Genève*, Switzerland. In a controlled feeding trial, the macaque was given desert locusts (*Schistocerca gregaria*) and mealworm beetle larvae (*Tenebrio molitor*) in different weighted proportions (Table S2) for ten successive afternoons, respectively, with two-days-breaks in-between conditions. The insects were given in addition to the usual diet including various items as fruits, vegetables, meat and eggs. However, the enclosure comprises an outdoor area, feeding on other insects can hence not be excluded. The experiment was performed from June to August 2021. For these samples the proportions of original biomass are known.

From faecal samples of vervets and the macaque, 0.5 cm^3 of scat were collected with gloves and a plastic spoon from inside the scat into 20 mL HDPE scintillation vials (Carl Roth GmbH,

Karlsruhe, Germany), and whole bat samples were collected into 2 mL Eppendorf tubes, then covered with silica gel beads and stored at ambient temperature until DNA extraction. DNA extractions were performed as in Brun et al. (2022; i.e. Chapter 1 of the present manuscript).

DNA enrichment

For the DNA metabarcoding assay samples will be amplified with a primer pair targeting mitochondrial COI insect DNA (ZBJ-ArtF1c/ZBJ-ArtR2c[169], hereafter ZBJ), according to the recommended thermal profile. The PCR reactions will be performed in triplicates in a final volume of 20 µL. The mixture contains 1 U AmpliTaq Gold 360 mix (Thermo Fischer Scientific, USA), 0.04 µg of bovine serum albumin (Roche Diagnostics, Basel, Switzerland), 2 µM of human-blocking primer (i.e. 5'-AGGGATAACAGCGCAAYTCTATTCTAGAGTC-C3-3'[70]) and 0.2 µM of tagged forward and reverse primers and 2 µL of template DNA. For ZBJ the PCR consists of a two-cycle step-down PCR: 3 mins at 94 °C, 16 cycles of 30 s at 94 °C, 30 s at 61 °C (decreasing 0.5 °C/cycle) and 30 s at 72 °C, followed by 24 cycles of 30 s at 94 °C, 30 s at 53 °C and 30 s at 72 °C with a final elongation of 10 mins at 72 °C. Extraction, non-template and positive controls as well as blanks will be included. Amplicon pools will be purified using the MinElute PCR Purification Kit (Qiagen, Hilden, Germany) and quantified using a Qubit 2.0 Fluorometer (Life Technology Corporation, USA). Library preparation will be performed using e.g. a TruSeq DNA Library Prep Kit (Illumina, San Diego, CA, USA). Final libraries will be quantified by qPCR, normalized and pooled before 150 paired-end sequencing on the Illumina Miniseq Sequencing System (Illumina, San Diego, CA, USA).

The TSC will be performed with the manufacturer's protocol, *myBaits Standard* or *myBaits High Sensitivity* for Hybridization Capture for Targeted NGS version 5.01 (Daicel Arbor Biosciences, USA) after library preparation with an Illumina protocol using dual indexing (e.g. TruSeq Library Prep Kit; Illumina). While for the mock communities the *Standard* protocol can be used (if at least 100 ng/library), the *High Sensitivity* protocol with two capture rounds is recommended for eDNA samples. Each capture reaction can be done with a maximum of 7 µL library volume, which should be suspended in either nuclease-free buffer or water and concentrated if needed. In a first step, libraries are mixed with various components (adapter blockers, etc.) and then denatured. After denaturation libraries are combined with other hybridisation reagents, among which the biotinylated baits. Hybridisation temperature depends on presumed target divergence to baits (60/62/65°C or 55/61/63°C for *Standard* or *High Sensitivity*; the higher the temperature, the less flexible the assay). The baits hybridise 16h-24h to complementary sequences and are then bound to streptavidin-coated beads during a warm wash, i.e. at the same temperature as the hybridisation step. For the *Standard* protocol, libraries are then ready for amplification, quantification and sequence preparation. For the *High Sensitivity* protocol, libraries are amplified before a second round of hybridisation capture, wash and amplification and only then double-captured libraries are quantified and prepared for sequencing as usual. Enriched libraries will be sequenced on an Illumina Miniseq platform (2x 150 bp paired-ends). One of each MCs will be prepared for shotgun sequencing (non-enriched libraries). For an outline of the experimental phases, see Table 1.

Table 1: Overview over the experimental design.

	PHASE I: 8 reactions **Mock communities (MCs)**	PHASE II: 8 reactions **eDNA samples**
SAMPLES	1 MC equimolar (10 species) 1 MC equimolar (40 species) 1 MC variable concentrations (10 species) 1 MC variable concentrations (40 species)	Available to include: N vervet scats N bat scats N macaque scats
SAMPLE PREPARATION	DNeasy Blood and Tissue extractions of insect tissue Sanger COI Quantification of mtDNA by qPCR Preparation of various MCs	NucleoSpin Soil Kit DNA extraction
DNA ENRICHMENT	TSC *Standard* with COI baits • 4 MCs in duplicates (N=8) DNA metabarcoding with COI DNA, ZBJ primers • 4 MCs in triplicates (N=24) Shotgun sequencing of non-enriched libraries, one of each MCs (N=4).	TSC *High Sensitivity* with COI baits • 4 samples in duplicates (N=8) DNA metabarcoding with COI DNA, ZBJ primers • 4 samples in triplicates (N=24) Shotgun sequencing of non-enriched libraries (N=4).
SEQUENCING	Illumina MiniSeq 2x 150 bp	Illumina MiniSeq 2x 150 bp

Perspective data analyses

The presented experimental design allows to assess and compare the assays' results in terms of specificity, sensitivity, quantification and consider effects of MC composition.

SPECIFICITY: Assay specificity can be assessed in terms of detection probabilities of the target taxa for Phase I, since the composition is known. We expect TSC to be more flexible and less specific, i.e. the detection of all taxa is expected but also more non-target DNA capture. The DNA metabarcoding assay risks not to detect all taxa due to amplification biases but should not amplify non-target taxa. Therefore DNA metabarcoding may outcompete TSC in terms of specificity.

SENSITIVITY: In particular relevant to assess the sensitivity are MCs with variable concentrations in Phase I. For TSC there is a trade-off between sensitivity vs specificity[147]. We expect it to be more sensitive but less specific than the DNA metabarcoding assay. Taxa present in low concentrations should be detected with TSC while target competition risks to limit amplifications in PCR.

QUANTIFICATION: The quantification potential of DNA metabarcoding and TSC can be compared, i.e. if linear corelations of relative DNA concentration can be established between the sample and sequence results[39]. PCR-induced and TSC-induced biases can be compared. Here, the baseline of the shotgun sequencing for eDNA samples and known concentrations of MCs can serve as reference. We hypothesise that TSC will show higher fidelity to the true relative abundance of each sequence in the MCs and higher magnitudes compared to the DNA metabarcoding approach. As primers do not equally align to different sequences, we anticipate a bias of over-representation towards taxa whose sequences are more similar to

the primer sequence and analogously an under-representation with increasing number of mismatches. As for direct shotgun sequencing we expect an unbiased representation of sequences but, depending on the sequencing depths, rare sequences may not be represented.

COMPOSITION: In addition, the effect of different species assemblies of MCs in terms of numbers and concentrations will be assessed. As argued by Piñol et al. (2019), the composition of a mixture presumably influences the quantification potential in DNA metabarcoding; i.e. mixtures with i) more species and ii) variable concentrations perform better[39]. Our design permits to test for both hypotheses. For TSC we expect no difference in terms of detection.

Conclusion

The mitochondrial gene region COI has been established as the standard marker for animal barcoding due to its high interspecific taxonomic resolution[156]. Although its length (658 bp) makes it less suitable for DNA metabarcoding studies where DNA is mostly degraded and hence fragmented[1], this can be overcome by using COI mini barcodes[170,171]. Despite the popularity of COI markers for eDNA studies, it is nonetheless difficult to design truly universal PCR primers for the highly variable protein coding gene COI. Divergence in priming sites may imply mismatches between primers and targets and hence PCR-induced biases, resulting in non-amplification of degraded or too divergent DNA and thus skewed taxa representation[172,173]. Furthermore, COI is not as specific for short barcodes as for the complete *Folmer* region. Therefore including a second gene region, more specific for shorter

barcodes, in the experimental set up could be desirable. Additionally, with the gradual development from morphological identifications of specimen over DNA barcoding of isolated specimen to DNA metabarcoding of complex DNA mixtures, the importance of reference databases steadily increased. Without any verification by morphological assessment, the correct assignment of sequences becomes a crucial step for reliable results. While public sequence databases (namely GenBank and BOLD) are most comprehensive for the *Folmer* region for animals, a multi-locus approach is nonetheless recommended to increase taxonomic coverage and avoid ambiguous assignments[53,174]. 16S mitochondrial rDNA (16S) has been proposed as powerful alternative to COI for insect detection[173]. Comparisons between COI and 16S primers resulted in less biased detection scores for aquatic insects using 16S, which translates into more realistic biomass representations[175]. The disadvantage of this marker, however, is that reference databases are not as representative as for COI. The risk of incorrect assignment of sequences is higher the shorter the barcode if reference databases are incomplete[176]. Therefore, a combination of COI and 16S genes has been recommended for DNA metabarcoding studies, e.g.[23,177].

We attempted to include baits targeting 16S in our design, however, abandoned this idea for the moment due to a lack of genetic reference material. There were ~51K sequences available for 16S in Genbank. We reduced these to 29,176 sequences, all of them ~650 bp long, based on targeted fragment and length. There were clear biases in abundances of sequences between taxonomic families, depending on the marker (e.g. more than 10,000 COI sequences for the Psychodidae family and not one corresponding sequence for 16S). In order not to bias our TSC assay due to database issues, we therefore abandoned the idea of a combined bait set for both genes. However, a subsequent design of 16S baits could serve to expand the

experiment and to test for the effects of targeted genetic region and databases. For this, the DNA metabarcoding assay could be complemented by using a primer targeting 16S, e.g. INSE01[1] and conduct complementary data analyses as described below.

METABARCODES AND GENETIC REGIONS: The performance of the different primers and probes for COI and 16S could be assessed in terms of specificity and taxonomic resolution. The ZBJ primers have originally been designed for the analysis of bat guano and extensively used for bat diet studies (e.g.[23,178–183]). The diet of many bat species is dominated by Lepidopteran and Dipteran species, both orders amplify well with this primer set. There has been critics presuming the bias of these primers towards these orders[173] and effectively not all orders seem to amplify equally[176,177]. The eDNA samples allow for comprehensive methodological tests since included bat species are strictly insectivorous whereas vervet monkeys mostly feed on plants and occasionally on arthropods[168]. Therefore, the latter samples are supposed to contain lower concentrated target insect DNA and more diverse components. The comparison between sample types for this primer as well as of different markers for the same sample type may yield interesting results, i.e. comparing COI and 16S for insects in bat guano (for a previous comparison showing complementarity, see[177]). We expect more reliable biomass estimation with Inse01[1] than ZBJ, whereas the taxonomic resolution remains unclear.

DATABASES: The influence of reference databases can be evaluated for the MCs by using local reference databases containing all sequences vs public databases. We postulate the taxonomical assignment to be better for COI than for 16S when relying on public reference databases only[174]. In a recent study, Tournayre et al. (2020) compared the performance of 12 DNA metabarcodes, whereof ten COI and two 16S, on arthropod mock communities and bat guano samples. While COI outperformed 16S for bat guano, the authors related this to less

complete public databases, *in silico* analyses showing a good performance for 16S[176]. It would be interesting to see if this effect remained the same when using complete local databases. However, we assume that taxonomic resolution is higher for COI given its interspecific variability.

The main challenge for the design of bait sets targeting broader taxonomic groups is, in general, to find the best compromise between specificity and flexibility. If target DNA is divergent from available baits, i.e. if the clustering threshold is too high, there is a risk not to capture target species equally well resulting in biases or false-negative detections. And if the threshold is too low, i.e. the assay is too flexible, there is a risk to capture non-target DNA. The presented bait set was designed to detect – ideally unbiased – all insect species with correct biomass correlations. The experimental design was conceived to evaluate it the best possible. Overall, TSC assays present the potential to overcome some biases typical to DNA metabarcoding, that is mainly PCR-induced biases, and hence provide the opportunity to result in reliable biomass representations. When eDNA samples analysed by TSC provide such biomass information, eDNA can reveal even more insights into ecosystems. Our specific bait set could advise on conservation efforts for a taxonomic group of high conservation concern playing a key role for many organisms and ecosystems.

Acknowledgements

We are grateful to Dr Brian Brunelle from Daicel Arbor Biosciences (Ann Arbor, MI) for helpful discussion during the design process and the final bait design. We also thank Myrtille Bérenger, Grenoble, for sampling of bat faecal samples, the IVP field team for the sampling

of vervet monkey faecal samples and the Bioparc staff for conducting the feeding trial, including the sampling of macaque faecal samples.

Supplementary data Project+

Table S1: Available bat scats of various or unknown bat species, sampled in France.

ID	Commune	Lieu	Date	Species	Method
1	Polignac, FR	Grotte de la Denise	25.08.18	*Myotis daubentonii*	capture
2	Polignac, FR	Grotte de la Denise	25.08.18	*Rhinolophus ferrumequnum*	capture
3	Polignac, FR	Grotte de la Denise	25.08.18	*Myotis daubentonii*	capture
4	Polignac, FR	Grotte de la Denise	25.08.18	*Plecotus auritus*	capture
5	Polignac, FR	Grotte de la Denise	25.08.18	*Myotis daubentonii*	capture
6	Polignac, FR	Grotte de la Denise	25.08.18	*Myotis myotis*	capture
7	Polignac, FR	Grotte de la Denise	25.08.18	*Myotis daubentonii*	capture
8	Polignac, FR	Grotte de la Denise	25.08.18	*Myotis daubentonii*	capture
9	Polignac, FR	Grotte de la Denise	25.08.18	*Myotis daubentonii*	capture
10	Polignac, FR	Grotte de la Denise	25.08.18	*Myotis daubentonii*	capture
11	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
12	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
13	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
14	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
15	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
16	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
17	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
18	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
19	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
20	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
21	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
22	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
23	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
24	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
25	Choranche, FR	Grotte de Gournier	16.09.18	*Pipistrellus pipistrellus*	capture
26	Choranche, FR	Grotte de Gournier	16.09.18	*Nyctalus noctula*	capture
27	Choranche, FR	Grotte de Gournier	16.09.18	*Myotis myotis*	capture
28	Choranche, FR	Grotte de Gournier	16.09.18	*Plecotus austriacus*	capture
29	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
30	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
31	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
32	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
33	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
34	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
35	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture

36	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
37	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
38	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
39	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
40	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
41	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
42	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
43	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
44	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
45	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
46	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	capture
52	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection
53	Châtelus, FR	Grotte de Bournillon	17.09.18	NA	collection

Table S2: Macaque (Macaca fascicularis) faecal samples collected during a controlled feeding trial conducted in the Bioparc, Geneva, Switzerland. The macaque was given desert locusts (Schistocerca gregaria) and mealworm beetle larvae (Tenebrio molitor) in different weighted proportions for ten successive afternoons, respectively, with two-days-breaks in-between conditions.

ID	Date	Desert locusts [%]	Mealworm beetle larvae [%]
MAC01	29.06.21	50	50
MAC02	30.06.21	50	50
MAC03	01.07.21	50	50
MAC04	02.07.21	50	50
MAC05	03.07.21	50	50
MAC06	04.07.21	50	50
MAC07	05.07.21	50	50
MAC08	06.07.21	50	50
MAC09	07.07.21	50	50
MAC10	08.07.21	50	50
MAC11	11.07.21	20	80
MAC12	12.07.21	20	80
MAC13	13.07.21	20	80
MAC14	14.07.21	20	80
MAC15	15.07.21	20	80
MAC16	16.07.21	20	80
MAC17	17.07.21	20	80
MAC18	18.07.21	20	80
MAC19	20.07.21	20	80
MAC21	23.07.21	90	10
MAC22	24.07.21	90	10
MAC23	25.07.21	90	10
MAC24	26.07.21	90	10
MAC25	27.07.21	90	10
MAC26	28.07.21	90	10
MAC27	29.07.21	90	10
MAC28	30.07.21	90	10
MAC29	31.07.21	90	10
MAC30	01.08.21	90	10

Additional article 1: Comprehensive coverage of human last meal components revealed by a forensic DNA metabarcoding approach

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

Comprehensive coverage of human last meal components revealed by a forensic DNA metabarcoding approach

Judith Schneider[1,5], Eduard Mas-Carrió[1,5], Catherine Jan[2], Christian Miquel[3], Pierre Taberlet[3,4], Katarzyna Michaud[2] & Luca Fumagalli[1,2]✉

Stomach content analyses are a valuable tool in human forensic science to interpret perimortem events. While the identification of food components of plant and animal origin has traditionally been conducted by macro- and microscopical approaches in case of incomplete digestion, molecular methods provide the potential to increase sensitivity and taxonomic resolution. In particular, DNA metabarcoding (PCR-amplification and next generation sequencing of complex DNA mixtures) has seen a rapid growth in the field of wildlife ecology to assess species' diets from faecal and gastric samples. Despite clear advantages, molecular approaches have not yet been established in routine human forensics to investigate the last meal components of deceased persons. In this pilot study we applied for the first time a DNA metabarcoding approach to assess both plant and vertebrate components of 48 human stomach content samples taken during medicolegal autopsies. We obtained a final dataset with 34 vertebrate and 124 vegetal unique sequences, that were clustered to 9 and 33 operational taxonomic units (OTUs), respectively. Our results suggest that this approach can provide crucial information about circumstances preceding death, and open promising perspectives for biomedical dietary surveys based on digested food items found in the gastrointestinal tract.

Postmortem stomach content analyses are an essential tool in forensic science. In addition to e.g. pathological or toxicological investigations, the identification of organic material of plant and animal origin may give valuable information not only about the last meal components but also the last hours surrounding death and its time-frame, as well as establishing a link between a victim and a suspect or a location[1–3]. Macroscopic and microscopic inspection is the standard method to morphologically identify food items found in the stomach of deceased persons when autopsies are performed. However, this approach is of limited efficiency especially if food components have been rendered non-identifiable due to chewing and the digestive processes occurring in the highly acidic environment of the stomach. In addition, the structure of food items can be too similar between different taxa to allow unambiguous taxonomic identification.

Over the last decade, molecular methods have increasingly been employed to study the diet components of several non-human organisms, due to the advances in DNA amplification and sequencing technologies. In particular, DNA metabarcoding (i.e. the simultaneous PCR-amplification with universal primers and next generation sequencing (NGS) of complex DNA mixtures[4]) has been used in the field of wildlife ecology to assess a species diet and to infer prey-predator relationships or ecological networks based on faecal samples (e.g.[5–10]). Short DNA metabarcodes, usually less than 150 base pairs (bp)[11], ideally combine high taxonomic coverage and resolution, and have the great advantage to be applicable to degraded DNA, which is the very characteristic of digested food samples. So far, the only study available on humans tested in a clinical context the methodological feasibility of using DNA metabarcoding of faecal samples to compare the inferred plant components with self-reported lists of eaten items, highlighting the potential of this approach[12]. Alternatively, the analysis of stomach content samples, although invasive, provides the advantage that aliments (and consequently DNA) are

[1]Laboratory for Conservation Biology, Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland. [2]University Center of Legal Medicine Lausanne and Geneva, Lausanne University Hospital (CHUV) and University of Lausanne, Ch. de la Vulliette 4, 1000 Lausanne 25, Switzerland. [3]Laboratoire d'Ecologie Alpine, CNRS, Université Grenoble Alpes, 38000 Grenoble, France. [4]UiT, The Arctic University of Norway, Tromsø Museum, Tromsø, Norway. [5]These authors contributed equally: Judith Schneider and Eduard Mas-Carrió. ✉email: luca.fumagalli@unil.ch

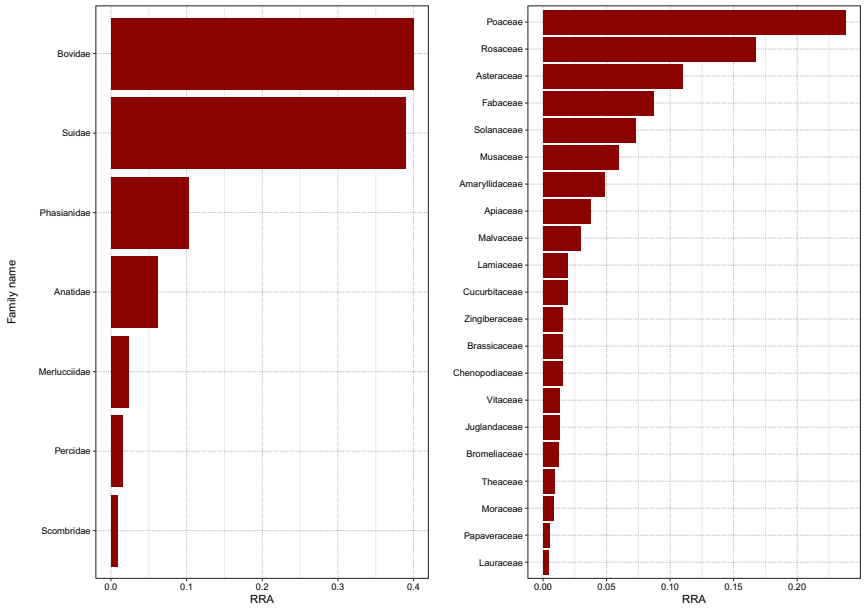


Figure 1. Barplot representing the sum of relative read abundances (RRA) for vertebrate (left panel) and plant (right panel) items across all samples at family level. Human reads were removed before calculating RRA scores.

less digested and degraded than after their passage through the intestinal tract[13]. DNA metabarcoding studies of stomach content samples have already been done to assess the diet of wildlife taxa such diverse as e.g. krill[14], Norwegian lemmings[15], Antarctic toothfish[16], spiders[17], Pygmy devil rays[18] and bugs[19].

Despite their clear advantages, molecular approaches have surprisingly not yet been established in routine human forensics to comprehensively investigate the last meal components of deceased persons (but see[20] in a very different context). The few studies published to date focused on the identification of a single taxon or food items (i.e. tomato and pepper seeds in faeces, mushrooms in clinical forensic samples, dandelion juice in the stomach of a presumed murder victim) using amplified fragment length polymorphism (AFLP) analysis[21,22] or PCR followed by Sanger sequencing[23,24]. While these studies demonstrate the interest to genetically identify digested food for forensic purposes, the scope of species-specific assays remains limited and their respective development time-consuming. In addition, the use of solid or structural intact particles as a source of DNA is not always possible when dealing with (partially) digested stomach contents.

To overcome these limitations, in this study we applied for the first time a DNA metabarcoding approach to test its potential to assess both plant and vertebrate components of human stomach content samples, taken during medicolegal autopsies. We identified several plant and animal taxa, consistent with previous food consumption descriptions in the studied region. Our results suggest that this method could reveal crucial information in providing corroborative evidence about the last hours preceding death. Besides being useful for purely forensic objectives, our study opens promising perspectives in the wider context of human dietary surveys based on digested food items found in the gastrointestinal tract or in faecal samples.

Results

After all quality filtering steps and merging of the data of all 48 samples, the final dataset for the Vert01 assay contained 34 different vertebrate sequences, clustered into 9 operational taxonomic units (OTUs), excluding human DNA. The Sper01 assay contained 124 different plant sequences, clustered into 33 OTUs. The relative read abundance (RRA) of animal and plant items is summarised per assay for all samples combined (Fig. 1). The heatmap shows RRA of all OTUs found per individual sample (Fig. 2). Total RRA of all OTUs per sample and replicate can be found in Supplementary Table S2.

We obtained plant sequences for all samples but one (CG16) that did not retain any OTUs after all filtering steps. We successfully amplified non-human vertebrate DNA in 34 samples, the remaining 14 resulted only in human DNA sequences.

We statistically tested for an effect of the digestion degree on the amount of different sequences retained per sample (both plants and animals) using a Pearson correlation test, but found no significant correlation

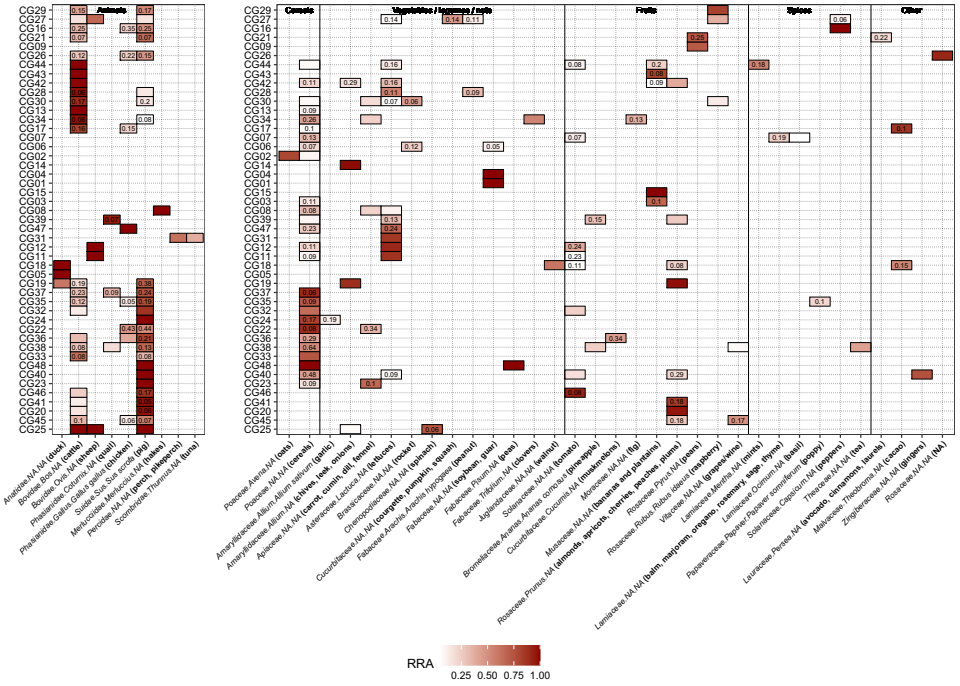


Figure 2. Heatmap representing relative read abundances (RRA) of detected items per sample. Values inside each box show the standard deviation of the mean between replicates. RRA scores have been calculated separately for vertebrates and plants. Within each sample, Sper01 OTUs not constituting at least 10% of RRA and Vert01 OTUs below 5% are not shown in the heatmap. We indicate Family.Genus.species assignments according to the *ecotag* command, along with the common name of all edible species or group of species (written in bold in brackets) which resulted in a 100% match with the NCBI database after manually blasting every sequence. NAs are shown in order to better visualise taxonomic resolution of the ecotag assignments.

($r(42) = 0.20$, $p = 0.18$). We also tested for the effect of days since death (time between death and autopsy), but found no significant correlation ($r(42) = 0.10$, $p = 0.49$) neither for plants nor for animals.

Discussion

In this study we successfully applied a DNA metabarcoding approach to identify consumed food items of plant and animal origin in human stomach content samples, even when digestion was advanced and macroscopic inspection no longer possible. A wide panel of common and less common edible food items were found, including meat, fish, legumes, cereals, nuts, fruits and spices. So far, gastric content analyses in a forensic context are typically based on microscopic and macroscopic identification of food items (reviewed e.g. in[1]). However, this approach is characterised by low taxonomic resolution, low sensitivity, and proves ineffective when meal leftovers are rendered unidentifiable due to chewing and digestive processes. In the field of molecular ecology, studies on animals have shown that morphological identification of prey items in the stomach underestimates prey diversity, which is particularly true when digestion is advanced (e.g.[25]). The only study to date applying DNA metabarcoding to infer human diet was based on faecal samples and did not assess any animal components of diet, although including a controlled feeding trial of an animal-based diet[12]. The comparison of the obtained plant DNA sequences to self-reporting indicated that, while some items were not reported but detected by DNA metabarcoding, all but one self-reported items were detected (the only exception being coffee), thus highlighting the sensitivity of the method. The present study, based on a random sampling of 48 human stomach contents collected during routine autopsies, includes a higher number of vegetal items and shows for the first time the successful detection of dietary items of animal origin. We found no correlation between the diversity of species detected and the time since death or digestion degree, which advocates for the utility of this methodology. The Vert01 primer set, highly specific to vertebrates, enables to distinguish between commonly eaten animal taxa and is clearly advantageous over morphological identification. In line with regional eating habits and previously published diet surveys[26], we found within the 48 samples mainly pig, cattle/dairy and OTUs assigned to the plant families Poaceae, Rosaceae and Asteraceae (likely cereals, fruits, lettuces; Fig. 1). We did not detect coffee

(*Coffea spp.*) in any of the stomach content samples, in line with[12], which might be due to a degrading effect of roasting procedures on DNA, the absence of this popular beverage in all of the stomach samples being unlikely. Similarly, although common in Swiss eating habits, we also did not detect potato, which is usually eaten boiled or baked. Note that additional edible plant species, not listed in Fig. 2 since not constituting at least 10% of RRA but with 100% match with the database, were also detected (e.g. buckwheat, citrus fruits, flax, mangoes, sesame; Supplementary Table S2). Because we could obviously not compare our results to self-reported diets, we applied very stringent filtering parameters to avoid the occurrence of false positives (see Bioinformatic data treatment). It is beyond the approach of this study to distinguish between the animal source and a final processed food item (e.g. dairy or egg products) based on the obtained DNA sequences. However, this could be achieved by complementing the primer set with a bacterial marker (to e.g. identify the presence of a particular cheese[27]) or using proteomics (see below).

Overall, the Vert01 metabarcode is able to discriminate well among commonly eaten genera. However, owing to its limited taxonomic resolution (72.4% at the species level, based on in silico testing[11]), species-level distinction is not always possible (e.g. between perch and pikeperch) or between potentially-eaten wild species and their conspecific domestic counterparts (e.g. wild boar and pig). In Fig. 2, we present the taxonomical assignation done using *ObiTools* together with a common name, selected after manually inspecting each sequence using BLAST and only considering 100% matches with edible species. In some cases, the common name refers to a group of species because the barcode was not specific enough to distinguish between genera or species. This is more relevant concerning plants, as the Sper01 metabarcode length ranges from 10 to 220 bp, implying that some items with shorter metabarcode and/or closely related phylogenetically could not be distinguished to genus or species level due to limited resolutive power. This is related to the nature of this universal plant marker, which has been designed to target a region of the *trn*L intron of chloroplast DNA which lacks taxonomic resolution within several plant families (only 21.5% resolution at the species level[9,11]) but has wide taxonomic coverage. This trade-off meant for our study that we could genetically not distinguish between some close species which are clearly different morphologically (e.g. stone fruits, cucurbits). To overcome this issue and increase the taxonomic resolution of the results, it is possible to envisage multiplexing within the same PCR of additional primers specifically targeting groups of species that cannot be identified at the species level by the P6 loop of the trn*L* intron. Such a strategy has already been implemented to distinguish between *Carpinus betulus* and *Corylus avellana* in bison diet[28]. Furthermore, it must be outlined that by using these primer sets only, diet assessment is not comprehensive as it does not target all possibly present food products. Even so-called universal primers may result in preferential amplification of some taxa over others and non-amplification of target taxa[29,30]. For this pilot study, we chose to use two universal PCR primer pairs with wide taxonomic coverage but limited specific resolution, in order to detect a broad range of items. To gain resolution for specific vertebrate or plant taxonomic groups (e.g. fish, birds, cereals) or target taxa not covered by these primers and which could be of forensic interest (e.g. marine crustaceans and molluscs, algae, fungi), it is possible to complement Vert01 and Sper01 with additional, taxonomically-restricted PCR metabarcoding primers described in the literature (e.g.[31]; examples reviewed in[11]). Taxonomic assignation of an unknown DNA sequence strongly depends on the exhaustiveness and quality of a reference database, either public as e.g. GenBank or custom-made/local (reviewed in[32]). In case of a priori knowledge of the overall consumed diet in samples, local databases may be restrained to the expected DNA sequences, which subsequently improves taxonomic assignment. For this study we in silico compiled databases containing all possible sequences amplified by our markers, but restricted these to vertebrates and spermatophytes (i.e. seed plants), respectively.

The duration of stomach emptying has been estimated by the percentage of a meal present in a stomach[3], but this process is influenced by several variables including the type and volume of consumed food, lifestyle and health, and can therefore last from few hours to days[2]. While one could argue that plant items usually remain longer in the stomach, our findings do not allow to draw robust conclusions about correlations of certain food items and digestion times. In order to establish hypotheses useful for time-frame estimations, additional experiments are necessary. In a controversial case of death, MS-based proteomics provided additional information through the analysis of food-derived proteins and peptides in the gastric content sampled at autopsy, indicating a last breakfast of milk and bread. While this method is certainly promising, it might reveal difficult if digestion is in an advanced stage, and has a less comprehensive scope than a DNA metabarcoding assay[33]. Furthermore, the effect of food processing techniques on DNA quality must be taken into account since cooking denatures e.g. proteins which in turn renders DNA amplification preferential to immunological approaches[1]. Different cooking treatments (variable duration of boiling, frying, baking) of tomato seeds showed that DNA extraction yielded in good quality DNA only for fresh seeds[34], while digestion did not destroy DNA[21]. Hence, there might be an implicit bias of DNA metabarcoding to preferentially detect non-processed food (i.e. raw versus cooked). Another issue of environmental DNA-based methods is that it is not possible to distinguish between different states of food products based on DNA sequences. As mentioned before, we could not discriminate between e.g. grapes/wine, fruits/juices, beef meat/dairy products or chicken meat/eggs, since the DNA sequence of a derived product is identical to the DNA sequence of its source. While it is less common to encounter such biases for plants, mainly in cereal-derived products, it has to be taken into account when extrapolating diet patterns from DNA metabarcoding results.

Stomach content sampling is invasive, but advantageous or even required with certain animal species and in particular circumstances, including definitely the human forensic context. An advantage of stomach content over faecal samples is that food is in an early stage of digestion before passing through the pyloric sphincter into the intestines, thus the effects of inhibition by bacteria or enzymes and degradation of DNA are less significant[11,18]. While some food particles such as seeds sometimes remain identifiable, even morphologically, after passing through the digestive system[21], others do not and the same applies to DNA which is degraded by the digestive processes taking place in the intestinal tract. In a controlled feeding experiment on insects, the detectability of

food DNA in different types of dietary samples showed that regurgitates and entire animals (including stomach content) outperformed faeces regarding detectability of prey DNA[13]. While food journals in dietary surveys may contain errors or deliberate omissions[12], they are a comprehensive and easily accessible method of human diet assessment. However, in case of deceased persons that option is no longer available.

Stomach content analyses provided crucial information for criminal investigations about cases of sudden and unexplained death on numerous occasions in recent years, enabling investigators to interpret perimortem events in detail (case examples reviewed in[2]). The results of this pilot study show that human stomach content analyses by DNA metabarcoding can be used as a complementary tool to traditional forensic macro- and microscopic approaches, with clear advantages such as an almost unlimited flexibility in terms of nature and range of taxa targeted, as well as high sensitivity and taxonomic resolution. Consequently, information that might otherwise remain undetected can be revealed, highlighting timings and circumstances surrounding the last hours of a person and his/her food intake. In a broader perspective, taking into account the potential improvements and refinements described above, and the growing amount of research literature available for wildlife species (i.e. environmental DNA-based studies), our results open up promising and novel prospects in the broader framework of human biomedical investigations of dietary patterns, based on partially or fully digested food found in the gastrointestinal tract or in faecal samples.

Methods

Sample collection. In this proof of concept study, we selected 48 anonymised frozen stomach content samples collected during medicolegal autopsies performed at the Lausanne University Center of Legal Medicine (Switzerland) in 2015. The inferred time span between death and autopsy was noted. For each sample, a degree of digestion as defined in[35], ranging from 1 (no signs of digestion, mainly solid components) to 4 (complete digestion, only liquid content), was given after macroscopic inspection (Supplementary Table S1). Following gentle mixing to ensure representativity, subsamples of 45–50 mL were transferred to BMT-50-S tubes for grinding with stainless steel beads (IKA, Staufen, Germany) and stored at -20 °C until DNA extraction.

DNA extraction. Two independent extractions per sample were performed using the DNeasy *mericon* Food Kit (Qiagen, Hilden, Germany) recommended for this sample type[35,36]. A subset of the extractions was tested for inhibitors with quantitative real-time PCR (qPCR) applying different dilutions in triplicates. qPCR reagents and conditions were the same as in DNA metabarcoding PCR reactions (see below), with the addition of 10,000 fold diluted SybrGreen (Thermo Fisher Scientific, USA). Following these analyses, all samples were diluted fivefold before PCR amplification. All extractions were performed in a laboratory restricted to forensic or low DNA-content analyses.

DNA metabarcoding assay. In order to assess a broad range of potential food components in human diet, samples were amplified using two different primer pairs, targeting taxa of both animal and vegetal origin. The first primer pair targets a 56–132 bp gene fragment of the 12S mitochondrial DNA gene in vertebrates (Vert01[11]; corresponding to 12SV5F/R[37]), allowing the amplification of animal-derived components of human diet. A human-blocking primer[9] that binds human DNA sequences to limit their amplification was added. As shown in a previous study[8], the chosen concentration of the human-blocking primer (see below) corresponds to the best compromise between the efficiency of the amplification of the vertebrate species and the blocking effect over the unwanted target. The second primer pair amplifies the P6 loop of the *trn*L intron (UAA) of chloroplast DNA (10–220 bp) and targets plant components of the diet (Sper01[11]; corresponding to g/h[38]). To allow attribution of DNA sequences to samples, primers were tagged with eight variable nucleotides added to their 5′-end with at least five differences between tags. The PCR reactions were performed in a final volume of 20 µL, using 96-well plates. The mixture contained 1 U AmpliTaq Gold 360 mix (Thermo Fisher Scientific, USA), 0.04 µg of bovine serum albumin (Roche Diagnostics, Basel, Switzerland), 2 µM of human-blocking primer (coupled with Vert01 primers only), 0.2 µM of tagged forward and reverse primers and 2 µL of fivefold diluted template DNA. PCR cycling conditions were denaturation for 10 min at 95 °C, followed by 40 cycles of 30 s at 95 °C, 30 s at 49 °C (Vert01) or 52 °C (Sper01), and 1 min at 72 °C, with a final elongation step of 7 min at 72 °C. For each assay, we included: (i) extraction negative controls; (ii) PCR negative and positive controls; (iii) blanks. Blanks correspond to empty wells on the PCR plate (i.e. no primer, no template) enabling to estimate the percentage of tag switches[39]. Both DNA extraction duplicates were amplified in triplicate (i.e. six PCR amplifications per sample were performed in total). Amplification success and fragment sizes were confirmed on a 2% agarose gel. Amplicons were pooled per plate, purified using a MinElute PCR Purification Kit (Qiagen, Hilden, Germany) and quantified using a Qubit 2.0 Fluorometer (Life Technologies Corporation, USA).

Two sequencing runs were performed, the first one to test the method on 12 samples (CG01-CG04, CG06-CG12, CG14), the second run including the remaining 36 samples. For the first run, library preparation and sequencing were performed at Fasteris facilities (Geneva, Switzerland). Libraries were prepared using the Metafast protocol (https://www.fasteris.com). A paired-end sequencing was carried out in an Illumina HiSeq 2500 (2 × 125 bp; Illumina, San Diego, CA, USA) using the HiSeq SBS Kit v4 (Illumina, San Diego, CA, USA) following the manufacturer's instructions. For the second run, library preparation was performed using the TruSeq DNA PCR-Free Library Prep Kit (Illumina, San Diego, CA, USA) with an adjusted beads ratio of 1.8 to remove small fragments. After adapter ligation, libraries were validated on a fragment analyser (Advanced Analytical Technologies, USA). Final libraries were quantified, normalised and pooled before 150 paired-end sequencing on an Illumina MiniSeq sequencing system with a Mid Output Kit (Illumina, San Diego, CA, USA).

Bioinformatic data treatment. The bioinformatic processing of the raw sequences output was performed using the *ObiTools* package[40]. The following steps were done separately for each library (i.e. per PCR plate containing each 12 samples and controls). Initially, forward and reverse reads were assembled with a minimum quality score of 40. The joined sequences were assigned to samples based on unique tag and primer combinations allowing two mismatches on primers and no mismatches on tags. Assigned sequences were then dereplicated, retaining only unique sequences. All sequences with less than 100 reads per library were discarded as well as those not fitting the above stated metabarcode lengths. This was followed by two different clustering methods. First, pairwise dissimilarities between reads were computed with the *obiclean* command and lesser abundant sequences with single nucleotide dissimilarity were clustered into the most abundant ones[40]. Second, we used the *sumaclust* algorithm[41] to further refine the resulting clusters based on a sequence similarity of 97%. Using the program ecoPCR[42] on the EMBL 2019 release, we built two databases by running in silico PCRs based on primer sequences and expected metabarcode lengths for Vert01 (16,292 sequences, Supplementary Data S1) and Sper01 (18,636 sequences, Supplementary Data S2). These databases were restricted to vertebrate and spermatophyte taxa, respectively. Sequences were assigned to taxa present in the database using the *ecotag* command, with a similarity threshold of 97%. Operational taxonomic units (OTUs) with similarity lower than 97% were eliminated from the dataset.

Further data cleaning and filtering was done in R (version 3.6.2). Sequences that were more abundant in extraction and PCR controls than in samples were considered as contamination and removed. To account for tag switching, we considered the leaking of a sequence to be directly linked to its abundance. To test this, we performed Wilcoxon signed-rank tests between samples and blanks and consequently removed from all samples a given ratio of presumed tag-leaked sequences. Dysfunctional PCR replicates were also discarded, i.e. with too small overall reads count based on library-dependant thresholds[11]. Final count of reads was transformed to RRA in order to have a normalised and comparable dataset between samples and sequencing runs. In the next step, PCR replicates were merged by sequence and extraction. Sequences that were present in only one out of three PCR replicates were removed, in line with[42]. This approach allowed us to discard single OTUs instead of whole PCR replicates. Finally, we combined the extraction duplicates of a sample, and calculated the mean count per OTU for each sample as well as the standard deviation.

For this resulting dataset (124 plant and 34 animal sequences), we re-assessed the taxonomic assignment done by *ObiTools (ecotag)* in order to assign a common name to each OTU, acknowledging in particular the limitations of the Sper01 metabarcode for domesticated varieties that share identical sequences for the *trn*L-P6 locus. We blasted[43] each sequence (Sper01 and Vert01) on the NCBI database and compared the results with the *ecotag* assignments.

Ethical statement. Study protocol was approved by the *Cantonal Commission on Ethics in Human Research* (Lausanne, Switzerland). Since the analyses do not concern the human genome but only aim at amplifying and analysing animal and plant DNA for research purposes, and since samples were completely anonymised, the study protocol does not require informed consent.

Data availability

The datasets analysed during the current study are available from the corresponding author on reasonable request.

Received: 26 April 2020; Accepted: 12 April 2021
Published online: 23 April 2021

References

1. Legge, C. M., Payne-James, J. J., Puntis, J. W. & Short, S. L. Postmortem gastric content analysis. In *Current Practice in Forensic Medicine* Vol. 2 (ed J. A. M. Gall and J. J. Payne-James) 47–66 (John Wiley & Sons, 2016).
2. Miller, J. Forensic Botany and Stomach Content Analysis: Established Practice and Innovation. In *Taphonomy of Human Remains: Forensic Analysis of the Dead and the Depositional Environment* (eds E. M. J. Schotsmans, N. Márquez-Grant, & S. L. Forbes) 187–200 (Wiley-Blackwell, 2017).
3. Tröger, H. D., Baur, C. & Spann, K. W. *Mageninhalt und Todeszeitbestimmung.* (Schmidt-Römhild, 1987).
4. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x (2012).
5. Deagle, B. E., Eveson, J. P. & Jarman, S. N. Quantification of damage in DNA recovered from highly degraded samples–a case study on DNA in faeces. *Front. Zool.* **3**, 11. https://doi.org/10.1186/1742-9994-3-11 (2006).
6. Valentini, A. *et al.* New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Mol. Ecol. Resour.* **9**, 51–60. https://doi.org/10.1111/j.1755-0998.2008.02352.x (2009).
7. Pompanon, F. *et al.* Who is eating what: diet assessment using next generation sequencing. *Mol. Ecol.* **21**, 1931–1950. https://doi.org/10.1111/j.1365-294X.2011.05403.x (2012).
8. Shehzad, W. *et al.* Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (Prionailurus bengalensis) in Pakistan. *Mol. Ecol.* **21**, 1951–1965. https://doi.org/10.1111/j.1365-294X.2011.05424.x (2012).
9. De Barba, M. *et al.* DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Mol. Ecol. Resour.* **14**, 306–323. https://doi.org/10.1111/1755-0998.12188 (2014).
10. Pansu, J. *et al.* Trophic ecology of large herbivores in a reassembling African ecosystem. *J. Ecol.* **107**, 1355–1376. https://doi.org/10.1111/1365-2745.13113 (2019).
11. Taberlet, P., Bonin, A., Zinger, L. & Coissac, E. *Environmental DNA: For Biodiversity Research and Monitoring.* (Oxford University Press, 2018).
12. Reese, A. T. *et al.* Using DNA metabarcoding to evaluate the plant component of human diets: a proof of concept. *mSystems* **4**, e00458–00419, doi:https://doi.org/10.1128/mSystems.00458-19 (2019).
13. Kamenova, S. *et al.* Comparing three types of dietary samples for prey DNA decay in an insect generalist predator. *Mol. Ecol. Resour.* **18**, 966–973. https://doi.org/10.1111/1755-0998.12775 (2018).

14. Vestheim, H. & Jarman, S. N. Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Front. Zool.* **5**, 12. https://doi.org/10.1186/1742-9994-5-12 (2008).
15. Soininen, E. M. *et al.* Shedding new light on the diet of Norwegian lemmings: DNA metabarcoding of stomach content. *Polar Biol.* **36**, 1069–1076. https://doi.org/10.1007/s00300-013-1328-2 (2013).
16. Yoon, T. H. *et al.* Metabarcoding analysis of the stomach contents of the Antarctic Toothfish (Dissostichus mawsoni) collected in the Antarctic Ocean. *PeerJ* **5**, e3977. https://doi.org/10.7717/peerj.3977 (2017).
17. Kennedy, S., Lim, J. Y., Clavel, J., Krehenwinkel, H. & Gillespie, R. G. Spider webs, stable isotopes and molecular gut content analysis: Multiple lines of evidence support trophic niche differentiation in a community of Hawaiian spiders. *Funct. Ecol.* **33**, 1722–1733. https://doi.org/10.1111/1365-2435.13361 (2019).
18. Comparing tissue and ethanol preservative-derived DNA. Barbato, M., Kovacs, T., Coleman, M. A., Broadhurst, M. K. & de Bruyn, M. Metabarcoding for stomach-content analyses of Pygmy devil ray (Mobula kuhlii cf. eregoodootenkee). *Ecol. Evol.* **9**, 2678–2687. https://doi.org/10.1002/ece3.4934 (2019).
19. Masonick, P., Hernandez, M. & Weirauch, C. No guts, no glory: Gut content metabarcoding unveils the diet of a flower-associated coastal sage scrub predator. *Ecosphere* **10**, e02712. https://doi.org/10.1002/ecs2.2712 (2019).
20. Maixner, F. *et al.* The Iceman's last meal consisted of fat, wild meat, and cereals. *Curr. Biol.* **28**, 2348-2355.e2349. https://doi.org/10.1016/j.cub.2018.05.067 (2018).
21. Lee, C. L. *et al.* DNA analysis of digested tomato seeds in stomach contents. *Am. J. Foren. Med. Path.* **27**, 121–125. https://doi.org/10.1097/01.paf.0000202722.21167.47 (2006).
22. Lee, C.-L., Coyle, H. M. & Lee, H. C. Genetic analysis of individual seeds by amplified fragment length polymorphism. *Croat. Med. J.* **48**, 563–565 (2007).
23. Kowalczyk, M. *et al.* Practical aspects of genetic identification of hallucinogenic and other poisonous mushrooms for clinical and forensic purposes. *Croat. Med. J.* **56**, 32–40. https://doi.org/10.3325/cmj.2015.56.32 (2015).
24. Lee, E. J. *et al.* The identification of ingested dandelion juice in gastric contents of a deceased person by direct sequencing and GC-MS methods. *J. Forensic Sci.* **54**, 721–727. https://doi.org/10.1111/j.1556-4029.2009.01019.x (2009).
25. Harms-Tuohy, C. A., Schizas, N. V. & Appeldoorn, R. S. Use of DNA metabarcoding for stomach content analysis in the invasive lionfish Pterois volitans in Puerto Rico. *Mar. Ecol. Prog. Ser.* **558**, 181–191. https://doi.org/10.3354/meps11738 (2016).
26. Chatelan, A. *et al.* Major differences in diet across three linguistic regions of Switzerland: Results from the first national nutrition survey menuCH. *Nutrients* **9**, 1163. https://doi.org/10.3390/nu9111163 (2017).
27. Moser, A., Berthoud, H., Eugster, E., Meile, L. & Irmler, S. Detection and enumeration of Lactobacillus helveticus in dairy products. *Int. Dairy J.* **68**, 52–59. https://doi.org/10.1016/j.idairyj.2016.12.007 (2017).
28. Kowalczyk, R. *et al.* Foraging plasticity allows a large herbivore to persist in a sheltering forest habitat: DNA metabarcoding diet analysis of the European bison. *Forest Ecol. Manag.* **449**, 117474. https://doi.org/10.1016/j.foreco.2019.117474 (2019).
29. Piñol, J., Senar, M. A. & Symondson, W. O. C. The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Mol. Ecol.* **28**, 407–419. https://doi.org/10.1111/mec.14776 (2019).
30. Jusino, M. A. *et al.* An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Mol. Ecol. Resour.* **19**, 176–190. https://doi.org/10.1111/1755-0998.12951 (2019).
31. Miya, M. *et al.* MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *R. Soc. Open Sci.* **2**, 150088. https://doi.org/10.1098/rsos.150088 (2015).
32. Zinger, L. *et al.* DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Mol. Ecol.* **28**, 1857–1862. https://doi.org/10.1111/mec.15060 (2019).
33. Pieri, M., Lombardi, A., Basilicata, P., Mamone, G. & Picariello, G. Proteomics in forensic sciences: Identification of the nature of the last meal at autopsy. *J. Proteome Res.* **17**, 2412–2420. https://doi.org/10.1021/acs.jproteome.8b00159 (2018).
34. Lee, C.-L., Huang, Y.-H., Hsu, I. C. & Lee, H. C. Evaluation of plant seed DNA and botanical evidence for potential forensic applications. *Forensic Sci. Res.*, 1–9, doi:https://doi.org/10.1080/20961790.2019.1594599 (2019).
35. Schwerer, M. J., Vogl, S. & Graw, M. Extraction of DNA from gastric content in different stages of digestion. *Forensic Sci. Int. Genet. Suppl. Ser.* **5**, e625–e626. https://doi.org/10.1016/j.fsigss.2015.10.004 (2015).
36. Zarzoso-Lacoste, D., Corse, E. & Vidal, E. Improving PCR detection of prey in molecular diet studies: importance of group-specific primer set selection and extraction protocol performances. *Mol. Ecol. Resour.* **13**, 117–127. https://doi.org/10.1111/1755-0998.12029 (2013).
37. Riaz, T. *et al.* ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Res.* **39**, e145. https://doi.org/10.1093/nar/gkr732 (2011).
38. Taberlet, P. *et al.* Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **35**, e14. https://doi.org/10.1093/nar/gkl938 (2007).
39. Schnell, I. B., Bohmann, K. & Gilbert, M. T. Tag jumps illuminated–reducing sequence-to-sample misidentifications in metabarcoding studies. *Mol. Ecol. Resour.* **15**, 1289–1303. https://doi.org/10.1111/1755-0998.12402 (2015).
40. Boyer, F. *et al.* obitools: a unix-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* **16**, 176–182. https://doi.org/10.1111/1755-0998.12428 (2016).
41. Mercier, C., Boyer, F., Bonin, A. & Coissac, E. SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. In Programs and Abstracts of the SeqBio. *workshop (Abstract)* 2013 (France, 2013).
42. Ficetola, G. F. *et al.* An In silico approach for the evaluation of DNA barcodes. *BMC Genomics* **11**, 434. https://doi.org/10.1186/1471-2164-11-434 (2010).
43. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421. https://doi.org/10.1186/1471-2105-10-421 (2009).

Acknowledgements

We would like to thank A. Bouvet, J. Guenat, N. Remollino and C. Stoffel for help in the laboratory. This study was supported by the University of Lausanne. E.M.-C. was supported by a fellowship in Life Sciences (Faculty of Biology and Medicine, University of Lausanne).

Author contributions

L.F. conceived the project and designed the experiments. L.F. and P.T. supervised molecular and statistical work. C.J. and C.M. performed laboratory work. J.S. and E.M.-C. performed bioinformatics and analysed the data. K.M. coordinated sample collection. J.S., E.M.-C. and L.F. wrote the manuscript with the input of all authors.

Competing interests

The authors note that P.T. is co-inventor of a patent related to the Sper01 primers and the use of the P6 loop of the chloroplast trn*L* (UAA) intron for plant identification using degraded template DNA. P.T. and C.M. are co-inventors of a patent related to the Vert01 primers for vertebrate identification using degraded template DNA.

These patents only restrict commercial applications and have no impact on the use of these loci by academic researchers. All other authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-88418-x.

Correspondence and requests for materials should be addressed to L.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 6 October 2021 | Revised: 11 December 2021 | Accepted: 13 December 2021

DOI: 10.1002/edn3.274

ORIGINAL ARTICLE

Environmental DNA — Dedicated to the study and use of environmental DNA for basic and applied sciences — WILEY

Assessing environmental DNA metabarcoding and camera trap surveys as complementary tools for biomonitoring of remote desert water bodies

Eduard Mas-Carrió[1] | Judith Schneider[1] | Battogtokh Nasanbat[2,3,4] | Samiya Ravchig[4] | Mmabaledi Buxton[5] | Casper Nyamukondiwa[5] | Céline Stoffel[1] | Claudio Augugliaro[6] | Francisco Ceacero[2] | Pierre Taberlet[7,8] | Olivier Glaizot[6,9] | Philippe Christe[6] | Luca Fumagalli[1,10]

[1]Laboratory for Conservation Biology, Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Switzerland

[2]Department of Animal Science and Food Processing, Faculty of Tropical AgriSciences, Czech University of Life Sciences Prague, Prague, Czech Republic

[3]Laboratory of Mammalian Ecology, Institute of Biology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia

[4]Department of Biology, School of Arts and Sciences, National University of Mongolia, Ulaanbaatar, Mongolia

[5]Department of Biological Sciences and Biotechnology, Botswana International University of Sciences and Technology (BIUST), Palapye, Botswana

[6]Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Switzerland

[7]Laboratoire d'Ecologie Alpine, CNRS, Université Grenoble Alpes, Grenoble, France

[8]UiT – The Arctic University of Norway, Tromsø Museum, Tromsø, Norway

[9]Musée Cantonal de Zoologie, Lausanne, Switzerland

[10]Swiss Human Institute of Forensic Taphonomy, University Centre of Legal Medicine Lausanne-Geneva, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

Correspondence
Luca Fumagalli, Laboratory for Conservation Biology, Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne 1015, Switzerland.
Email: luca.fumagalli@unil.ch

Funding information
Université de Lausanne

Abstract

Biodiversity assessments are indispensable tools for planning and monitoring conservation strategies. Camera traps (CT) are widely used to monitor wildlife and have proven their usefulness. Environmental DNA (eDNA)-based approaches are increasingly implemented for biomonitoring, combining sensitivity, high taxonomic coverage and resolution, non-invasiveness and easiness of sampling, but remain challenging for terrestrial fauna. However, in remote desert areas where scattered water bodies attract terrestrial species, which release their DNA into the water, this method presents a unique opportunity for their detection. In order to identify the most efficient method for a given study system, comparative studies are needed. Here, we compare CT and DNA metabarcoding of water samples collected from two desert ecosystems, the Trans-Altai Gobi in Mongolia and the Kalahari in Botswana. We recorded with CT the visiting patterns of wildlife and studied the correlation with the biodiversity captured with the eDNA approach. The aim of the present study was threefold: (a) to

Mas-Carrió and Schneider joint first authors.

Christe and Fumagalli joint last authors.

investigate how well waterborne eDNA captures signals of terrestrial fauna in remote desert environments, which have been so far neglected in terms of biomonitoring efforts; (b) to compare two distinct approaches for biomonitoring in such environments; and (c) to draw recommendations for future eDNA-based biomonitoring. We found significant correlations between the two methodologies and describe a detectability score based on variables extracted from CT data and the visiting patterns of wildlife. This supports the use of eDNA-based biomonitoring in these ecosystems and encourages further research to integrate the methodology in the planning and monitoring of conservation strategies.

KEYWORDS
biomonitoring, camera traps, deserts, DNA metabarcoding, eDNA, water bodies

1 | INTRODUCTION

Large-scale biodiversity loss has been documented in all types of ecosystems around the globe due to anthropogenic and climate change effects (Butchart et al., 2010; Díaz et al., 2019; Rosenzweig et al., 2008; WWF, 2020). Reliable biodiversity surveys are therefore needed to assess species conservation status over time and to plan and monitor management measures (Pimm et al., 2014), including the identification of biodiversity hotspots (Brooks et al., 2006; Myers et al., 2000). Desert ecosystems have been neglected in terms of scientific and monitoring efforts, resulting in knowledge gaps particularly for remote areas with difficult access, although they harbor diverse biological assemblages (Brito et al., 2014; Durant et al., 2012, 2014) and cover almost one fifth of the earth's land (Safriel et al., 2005). As climate change may impact environmental conditions in desert ecosystems disproportionately faster (Loarie et al., 2009), these unique systems should be placed at the center of attention.

Biomonitoring aims to provide detailed data on species' distribution, abundance, and diversity. Conventional, observer-based methods, such as visual censuses and systematic trapping, are time- and labor-intensive and mostly focus on a limited number of taxa (Thomsen & Willerslev, 2015). Technology-based methods gain importance in conservation research (Stephenson, 2020), whereof we compare two in this study. Non-invasive camera traps (CT) with infrared sensors are widely employed for conservation research and monitoring, in particular for larger terrestrial mammals (Caravaggi et al., 2017; Salvatori et al., 2021). They are becoming less effort-intensive thanks to AI-based tools to sort CT datasets. Nevertheless, there are also limiting factors for unbiased detectability of species, such as movement range (Burton et al., 2015; Caravaggi et al., 2020) or body mass. Small mammals are usually underrepresented in CT because their size is insufficient to trigger the camera sensor (Leempoel et al., 2020; but see Littlewood et al., 2021). They provide information on species' abundances, density, and richness while allowing for multispecies monitoring (Rovero & Zimmermann, 2016). Environmental DNA (eDNA) approaches offer valuable biodiversity assessment tools given the simultaneous analyses of complex DNA mixtures that enable to detect species' presences, estimate diversities, and relative abundances. Advances in DNA sequencing technologies facilitated an increase in eDNA studies over the last decade (Bohmann et al., 2014; Ruppert et al., 2019; Taberlet et al., 2018; Thomsen & Willerslev, 2015). In particular, DNA metabarcoding (i.e., PCR amplification of short but informative barcodes with universal primers and next-generation sequencing (NGS) of DNA mixtures (Taberlet et al., 2012)) allows the simultaneous assessment of whole communities. Most of these studies focus on aquatic organisms from freshwater ecosystems (Belle et al., 2019; Rees et al., 2014). Water samples are well suited to collecting eDNA due to high distribution capabilities of eDNA in water bodies (Rodgers & Mock, 2015; Valentini et al., 2016). Sampling being standardizable and relatively fast, the method requires only single visits to study sites (or repeated visits for temporal monitoring). Waterborne eDNA reflects temporally accurate biodiversity information due to the limited persistence of free eDNA in water for days or maximally weeks (Barnes & Turner, 2016). DNA degradation is the main cause impeding detection by eDNA-based techniques. Experimental studies have shown that the persistence of free aqueous eDNA (not bound to particles, i.e., sedimentary eDNA) depends on a number of factors, with, for example, higher temperatures, more solar radiation, and neutral or acidic *pH* leading to shorter detection periods (Pilliod et al., 2014; Strickler et al., 2015). However, its persistence depends on dynamic interactions of various biotic (e.g., rate of DNA shedding, microbial activity) and abiotic factors, hindering the drawing of general patterns. While most studies have been carried out in temperate areas, but see (Coutant et al., 2021; Ishige et al., 2017; Mena et al., 2021; Sales et al., 2020; Seeber et al., 2019), sampling eDNA from remote desert water bodies is particularly challenging because of DNA degradation, which is expected to be accelerated by extreme seasonal and daily temperature variations and high UV-B exposure found in this type of environment, technical difficulties caused by the filtration of typically turbid water samples (Egeter et al., 2018) and restrained accessibility. There are few studies to date using water

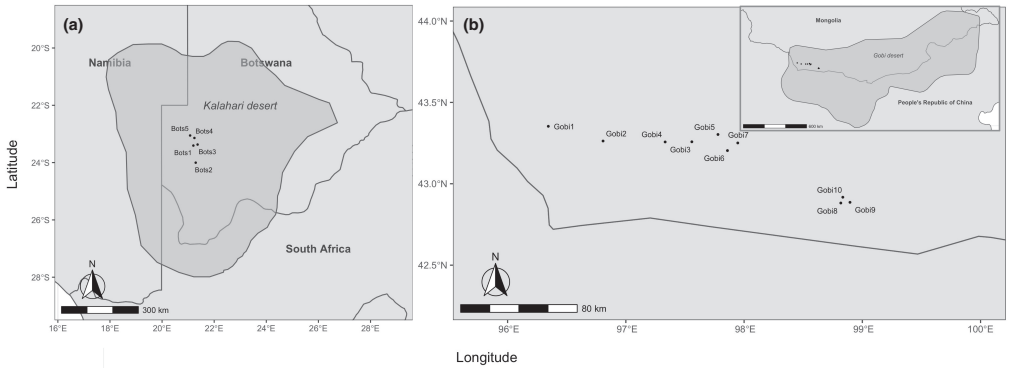


FIGURE 1 Sampling locations in (a) the Kalahari Desert and (b) the Trans-Altai Gobi Desert. Dark gray areas in each map indicate the extension of the Kalahari Desert and the Trans-Altai Gobi Desert, respectively

samples to assess biodiversity in an arid or semi-arid environment, but see (Egeter et al., 2018; Seeber et al., 2019). Despite these challenges, eDNA methods may reveal valuable tools for general biodiversity assessments and the monitoring of iconic and threatened species in precious ecosystems with reduced accessibility.

Environmental DNA from terrestrial animals has been mostly assessed by analyzing scats (De Barba et al., 2014; Kartzinel et al., 2015; Swift et al., 2018), soil (Leempoel et al., 2020; Yoccoz et al., 2012; Zinger et al., 2019), stomach content samples (Kennedy et al., 2019; Masonick et al., 2019; Soininen et al., 2013), leeches blood meals (Abrams et al., 2019; Nguyen et al., 2021; Tilker et al., 2020; Weiskopf et al., 2018; Wilting et al., 2021), or carrion flies (Calvignac-Spencer et al., 2013; Gogarten et al., 2020; Rodgers et al., 2017; Schubert et al., 2015). Bulk tissue samples (mixtures of, e.g., insects or other macroinvertebrate specimens) are also increasingly used not only to assess invertebrate diversity but also as an indirect way to sample vertebrate DNA (Lynggaard et al., 2019). However, animals also leave DNA traces in water while drinking or bathing, which means this water can be sampled and analyzed to detect non-aquatic organisms. This has first been proven using PCR and Sanger sequencing in an experimental setting with coyote DNA (Rodgers & Mock, 2015). Further studies successfully analyzed eDNA of terrestrial animals shed in water bodies sampled across different natural environments, from salt-licks in a Bornean tropical forest (Ishige et al., 2017), water bodies (Seeber et al., 2019; Ushio et al., 2017, 2018) and ponds (Harper et al., 2019), stagnant and running water combined (Mena et al., 2021), to rivers and streams (Coutant et al., 2021; Sales et al., 2020; Sales et al., 2020). This approach is particularly relevant for desert ecosystems with extreme conditions, where waterholes are small and scattered, leading to a spatial concentration of terrestrial animals that must gather and use the few available water sources (Davis et al., 2017; Razgour et al., 2018; Vale et al., 2015). Albeit the close association of water resources and desert species, there are also numerous adaptations to reduce their dependence and some species, such as gazelles, do not always comply with this expectation, as documented in the Trans-Altai Gobi (Nasanbat et al., 2021).

While CT and eDNA are two key tools available for species monitoring, there is limited information available to help researchers choose the most appropriate method for their needs, to compare performance, and decide whether and when methods can be used together (Stephenson, 2020). In order to enable inter-method comparability and their complementary use, comparative studies are therefore needed. Here, we compare CT and eDNA, with a focus on vertebrate terrestrial taxa in two desert ecosystems. While eDNA approaches are still relatively recent, CT have been used far longer but are undergoing increased attraction for conservation monitoring (Rovero & Zimmermann, 2016). The complementary use of these two methods can be appropriate for many situations and in particular for environments that are not favorable for observer-based monitoring. Analyzing images of CT allows us to quantify relative densities of species per sampling site. Based on these data, we can identify variables that best describe visiting patterns and assess whether they are mirrored by eDNA sequence data. We expect, for example, to find DNA of those taxa that visit regularly, in great numbers and shortly before sampling. The aim of the present study was threefold: (a) to investigate how well waterborne eDNA captures signals of terrestrial fauna in remote desert environments, (b) to compare two approaches for biomonitoring in such environments, and (c) to draw recommendations for future eDNA-based biomonitoring.

2 | MATERIALS AND METHODS

2.1 | Sampling sites

Camera traps and water sampling were conducted at 10 different sites in the Gobi Desert in Mongolia and four different sites in the Central Kalahari Game Reserve, within the Kalahari Desert in

Botswana (Figure 1 and Table S1). The water bodies were natural in Mongolia and artificial in Botswana.

The Great Gobi A Strictly Protected Area (SPA) in Trans-Altai Gobi was created in 1975, covers 44,000 km^2 and hosts emblematic species such as the snow leopard (*Panthera uncia*), the brown bear (*Ursus arctos*), the Asian wild ass (*Equus hemionus*) and the Bactrian camel (*Camelus ferus*). For a detailed description of the environmental conditions of this ecosystem, see Nasanbat et al. (2021).

The Central Kalahari Game Reserve was created in 1961 and covers 52,800 km^2, where ecotourism is a fundamental source of income (Stone et al., 2017). It hosts a great diversity of emblematic African large mammals including the African elephant (*Loxodonta Africana*), the cheetah (*Acinonyx jubatus*), the lion (*Panthera leo*), the brown hyena (*Parahyaena brunnea*) and the South African oryx (*Oryx gazella*).

2.2 | Water sampling

At each sampling location (Figure 1), one water sample was taken following the methodology of Pont et al. (2018), with modifications. The water was filtered through a VigiDNA 0.45 μM cross-flow filtration capsule (SPYGEN), with disposable 200 ml sterile syringes for each filtration capsule. For the Mongolian samples, 10 L of water was filtered at each location. For the Kalahari samples, filtered volumes varied from 1 to 10 L (average 6.3 L) depending on water body size and water turbidity. To avoid eDNA degradation, water in the capsule was replaced by 80 ml of CL1 conservation buffer (SPYGEN) and stored at room temperature. Sampling in Mongolia took place in August 2018 and in Botswana in May 2019 (for all details on sampling locations and samples see Table S1).

2.3 | DNA extraction

Extractions were performed in a pre-PCR laboratory dedicated to low DNA-content analyses, using a protocol modified from Pont et al. (2018). The filtration capsules were shaken for one hour at 420 rpm and agitated manually for 2 min to ensure a maximum DNA yield from the filter. From each capsule, 45 ml was poured into three separate 50 ml Falcon tubes (15 ml each) and 33 ml of 96% ethanol and 1.5 ml of 3 M sodium acetate (*pH* 5.2) were added before overnight incubation at −20°C. After precipitation, tubes were centrifuged at 7000 *g* for 30 min at 6°C. Supernatants were discarded, and tubes were incubated at 56°C for 10 min to evaporate residual ethanol. 720 μl of ATL buffer from the DNeasy Blood & Tissue Extraction Kit (Qiagen) and 40 μl of proteinase K were added, and the mixture was transferred to 2 ml Eppendorf tubes for at least 2 h of incubation at 56°C. The DNA extraction was pursued at step 6 of the NucleoSpin Soil Kit protocol (Macherey-Nagel). The three sub-samples were pooled in the extraction column. Elution was done with 2 x 100 μl of SE buffer. Negative controls were included at all steps.

The extractions were tested for inhibitors with real-time quantitative PCR (qPCR) applying different dilutions in triplicates. qPCR reagents and conditions were the same as in PCR amplification (see below), with the addition of SybrGreen (Thermo Fisher Scientific). Based on the results, all samples were subsequently diluted 10-fold before PCR amplification.

2.4 | DNA metabarcoding

DNA extracts were amplified with two primer sets. The first primer pair targets a fragment of the mitochondrial 12S rRNA gene in vertebrates (Vert01 (Taberlet et al., 2018), corresponding to 12SV5F/R in (Riaz et al., 2011)), the second targets a fragment of the mitochondrial 16S rRNA gene of mammals (Mamm02 (Giguet-Covex et al., 2014; Taberlet et al., 2018)). Human-blocking primers were added to the PCR mixes to prevent amplification of human DNA contaminants (for details of all primers, see Table S2). The total PCR volume was 20 μl, including 2 μl of template DNA and 1 U AmpliTaq Gold 360 mix (Thermo Fisher Scientific), 0.16 mg/ml of bovine serum albumin (BSA, Roche Diagnostics), 2 μM of human-blocking primer, and 0.5 μM of each tagged forward and reverse primer (i.e., primers with eight variable nucleotides added to their 5' end, allowing further sample identification, see (Taberlet et al., 2018)). Each sample was amplified in 12 replicates per primer in three separate PCR plates. Thermocycling conditions were as follows: denaturation at 95°C for 10 min, followed by 40 cycles of 30 s at 95°C, 30 s at 49°C and 57°C for Vert01 and Mamm02, respectively, 1 min at 72°C, with a final elongation step of 7 min at 72°C. Each 96-well PCR plate contained 12 blanks, eight negative extraction controls, eight negative PCR controls, and eight positive controls (DNA assemblies of species not present in the studied regions). Blanks correspond to empty wells and allow to estimate the proportion of tag switches (i.e., false combination of tags, generating chimeric sequences) occurring during the sequencing process (Schnell et al., 2015). Successful amplification was confirmed on a 1.5% agarose gel, and PCR products were subsequently pooled per PCR plate. Pooled amplicons were purified using a MinElute PCR Purification Kit (Qiagen). Purified PCR products were quantified using a Qubit 2.0 Fluorometer (Life Technology Corporation).

Library preparation was performed using a TruSeq DNA PCR-Free Library Prep Kit (Illumina) with an adjusted beads ratio of 1.8 to remove small fragments. After adapter ligation, libraries were validated on a fragment analyzer (Advanced Analytical Technologies). Since larger and smaller fragments besides the target size remained after this step, additional post library bead purifications were performed. To remove large and small fragments from Mamm02 libraries, a ratio of 0.7 was used followed by a ratio of 1.1. For Vert01 libraries, a ratio of 1 was used to remove small fragments. Final libraries were quantified by qPCR, normalized to 1 nM and

pooled before 150 paired-end sequencing on an Illumina Miniseq Sequencing System with a Mid-Output Kit (Illumina).

2.5 | Bioinformatic data analyses

The bioinformatic processing of the raw sequence output was conducted using the *OBITools* package (Boyer et al., 2016). The subsequent steps were followed separately for each library. Forward and reverse reads were assembled with a minimum quality score of 40 and assigned to samples based on unique tag and primer combinations, allowing two mismatches with primer, and identical sequences were clustered. All sequences with less than five reads per library were discarded as well as those not corresponding to primer specific barcode lengths, that is, 56–132 bp for Vert01 and 53–84 bp for Mamm02 (Taberlet et al., 2018). This was followed by two different clustering methods. First, pairwise dissimilarities between reads were computed and lesser abundant sequences with single nucleotide dissimilarity were clustered into the most abundant ones. Second, using the *sumaclust* algorithm, we reduced remaining clusters based on a sequence similarity of 97% (Mercier et al., 2013). Sequences were then assigned to a taxon using a reference database in two steps. First, in silico PCRs were performed with the *ecoPCR* software (Ficetola et al., 2010) on the whole EMBL repository to build reference databases for both metabarcodes (4455 Mamm02 sequences; 16,292 Vert01 sequences, Appendix S1). Taxonomic assignments with these databases were performed using a 95% sequence similarity threshold. Unassigned sequences after this step were discarded from downstream analyses. Second, each taxonomic assignment was manually inspected, and each sequence queried using the BLAST algorithm of GenBank to account for potential mis-assignation, because we used a relatively low similarity threshold. This is more likely to occur for the Vert01 metabarcode, because the amplicons can be very similar between close species. For cases with multiple candidate species, the geographic range was taken into account to select the correct species.

Further data cleaning and statistical analyses were conducted in R (version 4.0.2). Sequences that were more abundant in extraction controls as well as in negative and positive PCR controls than in samples were considered contaminants and removed as well as known common contaminants that were not expected in the study areas (Furlan et al., 2020). To account for tag switching, we considered the leaking of a sequence to be directly linked to its abundance. We performed Wilcoxon signed-rank tests to assess the relationship between samples and blanks. Removal of tag-leaked sequences was done independently per library. Dysfunctional PCR replicates with too small read counts were also discarded. Absolute sequence read counts were transformed to relative read abundance (RRA). Sequences not present in at least two PCR replicates were discarded from downstream analyses. Finally, RRA values were grouped across replicates to obtain a mean value per sample (for a reference data cleaning workflow, see Axtner et al. (2019)). Environmental DNA was considered both as presence/absence and as RRA data for comparison with CT data.

One location in Kalahari (Bots1) had to be excluded from the analyses due to insufficient amplification, possibly due to a problem during sampling, storage, and/or the DNA extraction step.

2.6 | Camera trapping and image coding

Camera traps were set up on sampling sites (Figure 1) between 40 and 70 days before water sampling, in such a way as to cover a maximum of the water bodies and shorelines. In the Trans-Altai Gobi Desert, we used Reconyx HyperFire HC600 (Reconyx), with trigger time of 0.2 s, recovery speed of 0.9 s, sensitivity set to "medium," and detection range/field of view of 30 m/42°. We also used Scout guard 565F model camera, with a trigger time of 1.2 s, a recovery speed of 1 s, and a detection range/field of view of 10 m/52°. In the Kalahari Desert, we used Reconyx Professional HP2X HyperFire 2, with a trigger time of 0.2 s, a recovery speed of 1 s, and a detection range/field of view of 24 m/40°.

We retained images spanning up to 40 days before water sampling to be able to compare between locations. Images were manually examined and the number of individuals per taxon, the time and date of visit recorded in hourly intervals (to minimize the risk of counting several times the same individuals). We coded all individuals in the pictures, regardless of their interaction with the water body, assuming that their presence implied a need for water. Animals were identified to the lowest possible taxonomic rank (species or genus). For each taxon, we recorded *body mass* extracted from PanTHERIA (Jones et al., 2009; Pigot et al., 2020), the *total number of visits*, the number of *days of last visit before sampling*, and the *mean frequency of visits* as potential explanatory variables.

Based on these variables, we built an equation to evaluate the quantity of DNA of a given species in a given location, under the assumption that taxa frequenting a water body more often would be more likely to be detected and yield more reads. To this end, we used the maximum number of individuals recorded in a single picture for each taxon i at each station j, within each hour interval t (N_{it}). First, we calculated a CT based DNA detectability score with:

$$D_{ij} = m_i \sum_t \frac{N_{ijt}}{t} \quad (1)$$

where m_i is the body mass of species i and N_{ijt} is the maximum number of individuals recorded in a single picture of species i at station j and at time t [days] before water sampling.

This equation gives a value of the cumulative DNA detectability (D_{ij}) for each taxon in each location at a particular time, which we then summed up over the 40 days before water sampling to obtain a cumulative DNA detectability score through time. This approach assumes that the eDNA added by a taxon in the water remains constant until its next visit and reaches its maximum concentration at

the end of the recording period. We assume that the quantity of released DNA is proportional to species body mass m_i.

The second index (Dr_{ij}) is based on Equation (1) but considers the relative cumulative quantity of DNA in the water body at the end of the recorded period.

$$\mathrm{Dr}_{ij} = \frac{\sum D_{ij}}{\sum D_j} \quad (2)$$

The above-mentioned indexes do not account for the presence and constant turnover of eDNA of other taxa over the days preceding a visit, that is, the pool of eDNA in the water body. We recalculated the scores from Equation (2) to account only for the last 5 days before sampling ($Dr5_{ij}$), to reduce the pooling effect. Our three detectability formulas were tested using the Mamm02 dataset only. Overall, we excluded the Vert01 data because the amplification of mammals and birds using this primer yields mainly mammal and only a limited number of bird sequences (see Figure S2 for an overview on bird detection).

2.7 | Statistical analyses

We investigated the individual effect of all CT-derived variables (*days of last visit before sampling*, *total number of visits* separately, *body mass*, and *mean frequency of visit*) on the eDNA data, both qualitatively (presence/absence) and quantitatively (logit RRA).

The relative read abundances (RRA) were transformed to avoid zero values using Equation (3) with a sample size (S) of 12 samples per site (Smithson & Verkuilen, 2006):

$$\mathrm{RRA}' = \frac{(\mathrm{RRA} * (S-1) + 1/2)}{S} \quad (3)$$

A logit transformation was subsequently used to achieve normality:

$$\mathrm{logit}\,(\mathrm{RRA}') = \log \frac{\mathrm{RRA}'}{1-\mathrm{RRA}'} \quad (4)$$

The non-linear correlations between some of the CT-derived variables suggested a more complex role of each variable to explain detectability (Figures S1 and S2). We aimed at disentangling from our hypothesis which variables were best explaining the observed eDNA detection. First, we used Kendall rank correlations to test separately the effect of each variable. Second, we used generalized linear mixed models (GLMM), with the *lme4* package (Bates et al., 2015), to investigate to what extent the variables (scaled and centered) influenced the likelihood of eDNA being detected. A logistic regression was performed on the presence/absence eDNA data and a normal regression for the RRA data (logit RRA) on the CT-derived variables. eDNA data (present/absent) were fitted to the explanatory variables extracted from CT data with a binomial distribution (Model 1, Supporting Information). Third, we used linear mixed-effect models (LMM), with the *lmer* package (Kuznetsova et al., 2017), to investigate the influence of the variables on the RRA data (Model 2, Supporting Information). For both model selections, we used Akaike information criterion (AIC) to select the model with the best fit, that is, the lowest AIC value, to reduce overfitting or underfitting the model (Burnham & Anderson, 2002). We chose the qualitative approach (Table 1, Model 1) to further investigate eDNA detection probability based on CT-derived variables because of its better explanatory power and ecological significance compared to the quantitative approach (Table 1, Model 2). We recalculated the predicted values of Model 1 for the three variables separately (*days of last visit before sampling*, *total number of visits*, and *body mass*) and combined through the cumulative detectability (D_{ij}, Equation 1).

TABLE 1 Results of the logistic regression for 0/1 eDNA data (categorical approach) and linear regression for the RRA data (quantitative approach)

	Presence/absence	Logit(RRA')
D_{ij} (Equation 1)	$p < 0.001$	$p < 0.001$
	AIC = 186.62	$R^2 = 0.136$
Dr_{ij} (Equation 2)	$p = 0.372$	$p = 0.879$
	AIC = 238.85	n. s.
$Dr5_{ij}$	$p = 0.526$	$p = 0.157$
	AIC = 60.79	n. s.
Days of last visit before sampling	$p < 0.001$	$p = 0.065$
		n. s.
Total number of visits	$p < 0.001$	$p < 0.001$
		$R^2 = 0.29$
Mean frequency of visits	$p < 0.001$	$p < 0.001$
		$R^2 = 0.133$
Body mass	$p < 0.01$	$p < 0.001$
		$R^2 = 0.12$

Note: We used logit transformation on the RRA data and removed 0 and 1 values from the dataset to test for the linear regression, as shown in Equations 3 and 4. Significant *p*-values are shown in bold, n. s., stands for *not significant*. R^2 values show the Adjusted R^2.

3 | RESULTS

3.1 | DNA metabarcoding

After all quality filtering steps, we retained 1,254,585 reads of 93 different OTUs for the Vert01 assay that were assigned to 37 taxa. 747,628 reads of 51 different OTUs were assigned to 36 taxa for the Mamm02 assay (all species detected by eDNA can be found in Table S4, Supporting Information). We detected 18 taxa in the Trans-Altai Gobi and 21 in the Kalahari. Vert01 and Mamm02 primers are overlapping for some taxa, that is, these taxa can be amplified by both primer sets. Bird species detected with both eDNA and CT can be

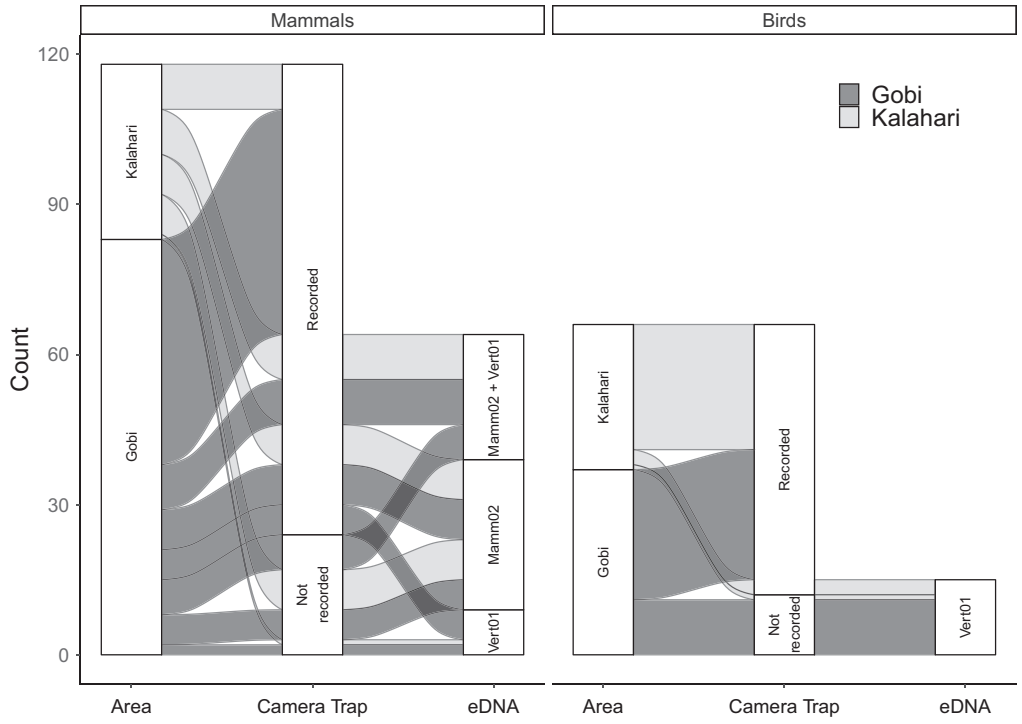


FIGURE 2 Comparison between areas of study, detection methods, and primers used for birds and mammals separately. The y-axis represents the number of different species in each particular location

found in Figure S2. While most mammal taxa were detected by only one primer pair, nine were shared between assays in Kalahari and 16 in Gobi (Figure 2). Despite not being a prior goal of this project, we compared primer specificity and found that eight mammal taxa in Gobi and one in Kalahari were detected exclusively with the Vert01 primer set. The numerous presences of birds in the Vert01 dataset contributed to the variable detection score between primers, as we did not detect bird sequences in the Mamm02 results.

3.2 | Camera traps

We identified 38 taxa in Kalahari and 22 in Gobi with CT (Figure 3, Table S3, Table S4, Supporting Information). One camera from Kalahari could not be recovered, and this location was therefore excluded from all analyses (Bots2). Using the variables retrieved from the images, we assessed the correlations between them to better understand the visiting patterns of the recorded species. We observed a negative exponential correlation between *total number of visits* and *days of last visit before sampling* ($R^2 = 0.35$, p-value < 0.001, Figure S1A) and between *total number of visits* and *mean frequency of visits* ($R^2 = 0.31$, p-value < 0.001, Figure S1B).

3.3 | Comparison between eDNA and camera trap data

In total, 84 taxa were identified combining data from CT and eDNA, 59 in Kalahari and 31 in Gobi. Some species were present in both areas. Detailed overview on the performance of detection methods can be found in Figures 2 and 3, as well as primer differences within the successfully amplified species. Note that each occurrence in Figure 2 indicates a single species for each particular location.

3.4 | Detectability score and eDNA

Camera traps results were used as a reference to compare the detectability score of the eDNA approach in these environments. Using the score from Equation (1), we separated the cumulative curves by positive and negative eDNA results (Figure 4). D_{ij} increases drastically if there were visits to the water body the day before sampling or the same day. We found significant correlations with eDNA data for the raw CT variables and with the detectability score D_{ij} (Equation 1), both for the quantitative (RRA) or the

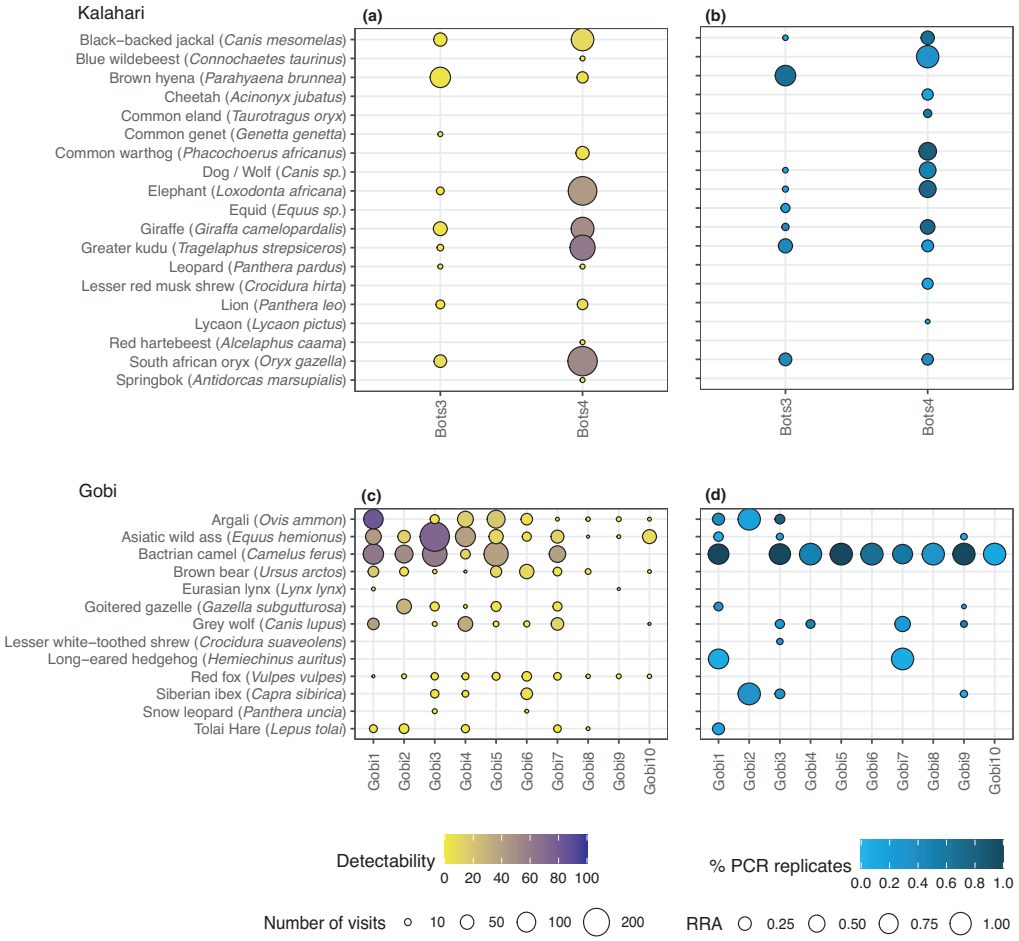


FIGURE 3 Bubble plot representing the detected mammalian species by each method. (a) Species recorded with camera traps in the Kalahari Desert. (b) Species detected with eDNA in the Kalahari Desert. (c) Species recorded with camera traps in the Trans-Altai Gobi Desert. (d) Species detected with eDNA in the Trans-Altai Gobi Desert

presence/absence measure. Relative detectability approaches (Dr_{ij} and $Dr5_{ij}$) poorly explained the eDNA results compared to the absolute D_{ij} approach (Table 1).

3.5 | Modeling eDNA detection

The explanatory variables showed significant correlations with eDNA results when tested individually, except *days of last visit before sampling* for RRA (Table 1). We first used eDNA presence/absence data as our model response variable (Model 1). Variance in Model 1 was significantly explained by *days of last visit before sampling*, *total number of visits* but not *body mass (log transformed)*, which also had the lowest AIC score. In Model 2, the quantitative response variable (logit RRA) was not significant for any possible combination (Model 2, Supporting Information).

The lower fit of the RRA data led us to further explore the presence/absence approach (Model 1) regarding detectability by eDNA. Figure 5 illustrates the relationship between CT data and eDNA detection for the cumulative detectability score (D_{ij}) (Figure 5a) and for each of the three variables independently (Figure 5b-d), which were used to build the detectability score equation (Equation 1).

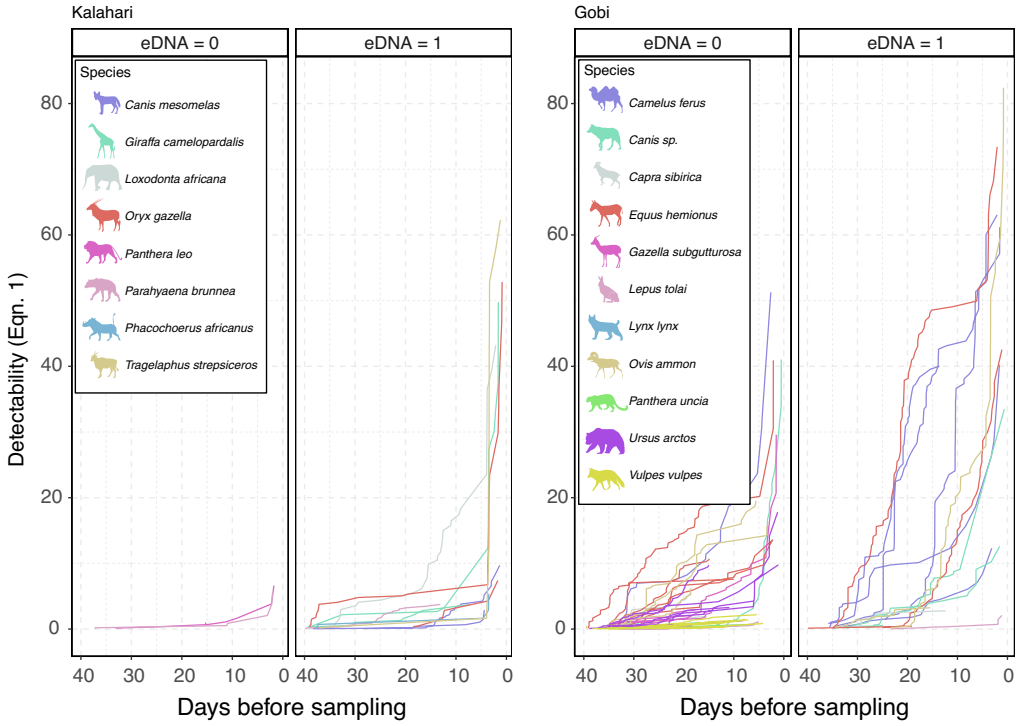


FIGURE 4 Detectability curves for each of the sampled areas calculated using Equation (1). Lines in the "0" box indicate that no eDNA could be recovered. Lines in the "1" box yield eDNA. This figure does not show species which were recorded only once, but these were included in Figure 3. Each species is represented by a silhouette

4 | DISCUSSION

Studying fauna in terrestrial environments using eDNA is generally more challenging than in aquatic habitats because the presence and concentration of eDNA is less homogeneous across the area of sampling (Leempoel et al., 2020; Lyet et al., 2021). To study terrestrial and semiaquatic environments using eDNA, sampling water can therefore be more advantageous over other sample types (Harper et al., 2019; Rodgers & Mock, 2015). Biomonitoring in arid or semi-arid ecosystems represents an additional challenge given the harsh environmental conditions and often vast spatial areas. However, the rare and spatially dispersed water bodies attract organisms and operate as DNA pools that record and temporarily preserve the information of visiting animals, thus representing unique opportunities for sampling. In this study, we successfully show the detection of terrestrial fauna using desert water. We analyzed the correlation between CT image data and eDNA sequence data and showed the interplay between the visiting patterns of species and the probability to detect them using eDNA. We detected terrestrial organisms in all our water samples with both primer sets, which illustrates the capacity of our assays to detect terrestrial fauna using waterborne eDNA from desert ecosystems. While CT is widely used for biomonitoring, only few studies compare CT and eDNA data for terrestrial animals (Leempoel et al., 2020; Lyet et al., 2021; Mena et al., 2021; Sales et al., 2020; Sales et al., 2020) and fewer explored the potential of waterborne eDNA for arid and semi-arid ecosystems. Seeber et al. (2019) used samples of water bodies in two African ecosystems to study the presence of mammal species, comparing a hybridization capture approach to conventional PCR. Furthermore, Egeter et al. (2018) sampled water in three Saharan water bodies for a mainly methodological DNA metabarcoding study focusing on the water filtration process. We observed clogging of the filters during filtration in the Kalahari due to the turbidity of the water resulting in variable volumes (Table S1), although we used the filter pore size as recommended by Egeter et al. (2018). Alternatively, Abrams et al. (2019) and Weiskopf et al. (2018) investigated the suitability of leech-derived eDNA as a survey tool for vertebrate species by comparing it to CT detections. Taking the analyses one step further, Tilker et al. (2020) combined the data obtained with these approaches to identify species responses to environmental factors.

DNA metabarcoding relies on "universal" primers that are designed for restrained taxa or groups. Variable numbers of

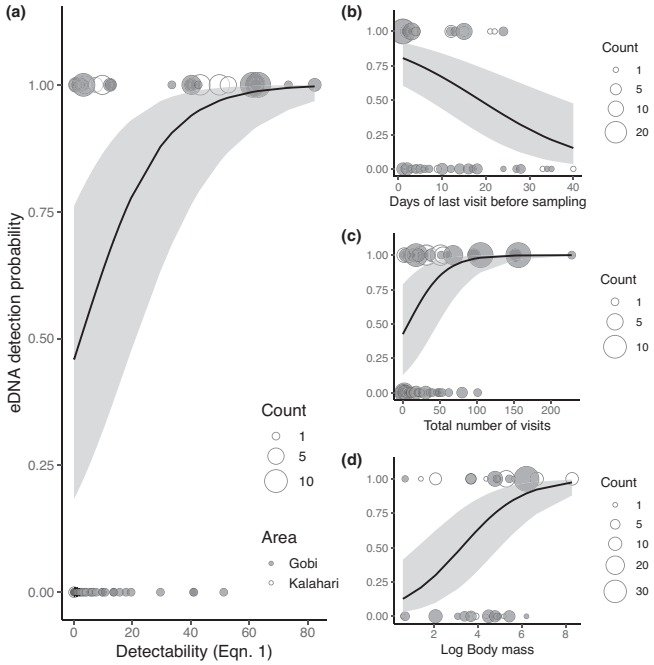


FIGURE 5 (a) Modeled prediction of eDNA detection for cumulative detectability (Equation (1)). (b–d) Modeled prediction for each variable involved in Equation (1) separately. All with *p*-value below 0.001. Size of the dots indicates count of occurrences, and gray scale indicates the area. The black line indicates the model's predicted values with its confidence interval in gray

mismatches between primers and templates are the presumed main cause for the preferential amplification of certain taxa and the under- or non-representation of others (Piñol et al., 2015; Piñol et al., 2019). The multiplexing of primers and barcodes is an attempt to minimize these effects (Alberdi et al., 2018; Galan et al., 2018; Jusino et al., 2019; Krehenwinkel et al., 2017). However, in our study, while we chose the Vert01 primers to target birds and reptiles in addition to mammals, mammal sequences were also the most amplified with Vert01 primer, duplicating eDNA detections. The redundancy of the information raises the question of the utility of primer multiplexing in this context (Figure 2).

The presence of trace DNA of humans and domesticated animals in eDNA studies represents a common issue, the possible sources of such contaminations being numerous (Furlan et al., 2020). We removed the obvious cases (pig, cow); however, we kept two taxa in the Kalahari dataset which are ambiguous and unconfirmed by CT. The detection of *Equus* sp. DNA in one Kalahari sample could be explained by the presence of horse, zebra, or donkey or the transport of such DNA, for example, by a predator. In addition, in the Kalahari samples, we amplified DNA assigned to *Canis* sp. (dog/wolf). The sequences differ from the ones found in the Gobi samples, therefore excluding a cross-contamination, but its source remains unclear.

Previous CT vs. eDNA studies found that smaller animals are less likely to trigger CT and risk being overlooked in this type of biomonitoring studies (Leempoel et al., 2020; Lyet et al., 2021; Sales et al., 2020). Smaller species probably release less DNA into the water than larger species and are hence less likely to be detected due to the presence of other species' DNA. Mena et al. (2021) conducted a comparative study of different traditional survey methods (pitfalls, grids, mist nets, and CT) and aquatic eDNA for the detection of terrestrial mammals in tropical forests. The overall results point out the benefits of eDNA surveys, in terms of detection scores, labor-effort and costs, but depend very much on the species and sampling area. It must be noted that in the present study, we analyzed 40 days of CT data and compared it to the results of only one water sampling event for each locality (Figures 2 and 3). This has to be kept in mind when evaluating the performance of the eDNA assays, since a 40 days period is well beyond the persistence of free eDNA in water according to literature (Barnes & Turner, 2016). In the light of these shortcomings, we consider it encouraging to have detected relying on eDNA more than half of the mammals recorded with CT (Figure 2). We also acknowledge the limitations of taking a single sample per water body, as it may be underrepresenting the wildlife diversity. Additional samples from the same water body would likely increase detected species

numbers, and we advocate that future studies aim to determine the optimal number of samples per location.

In this study, DNA was already present in the water prior to the placement of CT and the a priori DNA composition is unknown. We used the RRA approach as a proxy for species abundances, but it could be biased by the variable *body mass* of a species or by its behavior (e.g., drinking, bathing, and defecating), which affects the release of DNA (Harper et al., 2019). We did not find different detection scores between methods for nocturnal/diurnal species, but this is due to the capacity of CT to detect nocturnal species, contrasting to line transects (Coutant et al., 2021). For Gobi, the coverage of species is biased by the overwhelming presence of Bactrian camels (*Camelus ferus*). These animals are big and recurrent in the area, as confirmed by CT. To overcome this limitation, we suggest adding blocking primers for dominant species, if known beforehand. The usage of blocking primers is recommended for this type of study (De Barba et al., 2014; Vestheim & Jarman, 2008). Egeter et al. (2018) used Vert01 primers without human-blocking primers and obtained 68% of total reads assigned to hominids. In our study, we had fewer human sequences (9.36% of sequences for Mamm02 and 37.5% for Vert01), indicating a good performance of the blocking primer. A higher concentration of blocking primer than the one we used would increase the risk of co-blocking targeted taxa (Shehzad et al., 2012; Taberlet et al., 2018). In fact, the high abundance of camel and human DNA could explain the low detection rates of other less recurrent species, mainly carnivore species, despite the frequent CT recordings. For example, the locally rare and emblematic brown bear (*Ursus arctos*), whose residual population in the Gobi Desert numbers a few dozen individuals, was detected only once with the Vert01 primer pair (and therefore is not present in Figure 3d). This low detection of carnivore eDNA was also observed for Botswana and could indicate that further factors, such as drinking technique or contact with water (bathing), are involved besides the species body mass, see for example (Lyet et al., 2021). We therefore encourage future metabarcoding studies to investigate the factors associated with successful eDNA detection of carnivores.

The negative correlation between the variables *total number of visits* and *days of last visit before sampling* (Figure S1A) was expected because species visiting a water body many times are also likely to have visited it recently, and it could only be biased by migratory or nomad species that visit a water body in great numbers but low frequency. We excluded this potential confounding effect because taxa with high *total number of visits* were also the ones with the lowest *mean frequency between visits* (Figure S1B). Hence, we were able to use these variables as predictors to calculate and model eDNA detection probabilities.

We tested several equations to combine the explanatory variables retrieved from CT data into a comprehensive index to account for the expected detectability of species in each location, both using a categorical (0/1) and quantitative (RRA) approach (Table 1). When visualizing the increasing detectability score D_{ij} (Equation 1) by eDNA detection (Figure 4), there was a sudden increase for some species at the end of the monitored period. Such increase is due to the visits occurring not long before water sampling, which have a major impact on the overall score through time. Dr_{ij} and $Dr5_{ij}$ could not be visualized across time because only single values per species and location can be obtained. We found better correlation between RRA and $Dr5_{ij}$ compared to Dr_{ij} (Table 1), probably due to lower DNA degradation, but none were significant, which highlights the complexity of defining detectability scores. However, these scores are a simplification of reality, as they do not account among other factors for DNA decay (Barnes & Turner, 2016). Using RRA as a proxy for species relative abundance must, however, be taken with caution because of the biases that DNA extraction, amplification, and sequencing imply (for an overview of the biasing factors, see (Fonseca, 2018). Furthermore, the categorical approach homogenizes the coverage of each species and, in practice, increases the weight of low RRA species in our test (Deagle et al., 2019). This is of particular interest when using RRA data, as it provides a more realistic proxy in terms of abundances. Overall, the better correlation of RRA for $Dr5_{ij}$ suggests the RRA approach to be a better proxy for species detection when water samples are taken frequently. The categorical approach is recommended when the sampling is done only once (as in the present study) or sparsely. In addition, these scores only make sense when comparing CT and eDNA data. Still, the characteristics of the sampled area need to be accounted for when drawing detectability scores. For instance, Lyet et al. (2021) sampled river water to detect mammal species and they defined their detectability score based on the camera trap detection rate and the pluviometry of the day. Nonetheless, our results are promising, and optimizing the accuracy of these scores will improve cross-validation of both methodologies, both for comparative studies and when using eDNA as a complementary tool to CT. However, the complexity of interacting ecological factors complicates building a simple equation to reliably infer eDNA detection probability. In this line, a purpose-built experiment should be carried out to tackle this matter, with a limited number of species in a controlled environment.

Another goal of this study was to draw guidelines for future studies aiming to use eDNA as a biomonitoring tool in desert environments. We built models based both on eDNA 0/1 data (Model 1) and RRA data (Model 2). All variables used in Model 1 except *body mass* were significant, suggesting that the categorical transformation of our eDNA data is more advisable when the goal is species detection rather than its relative abundance. We used this model as a reference to calculate detection probabilities for each variable and area independently, in order to disentangle the effect of each and visualize them (Figure 5). Interestingly, the positive eDNA detections based on *Final cumulative detectability* are clearly divided into two groups (Figure 5a). This suggests that our detectability score fails to properly reflect the true detectability of some species, which could be explained by the poor correlation observed for *days of last visit before sampling* (sparse distribution of non-detection occurrences, Figure 5b) and *body mass* (Figure 5d). In this line, these last two variables should be studied more in depth to properly understand their impact on eDNA detectability. The good fit of *total number of visits* (Figure 5c) is, as mentioned before, influenced by *days of last visit before sampling* (Figure S1B).

Modeling with RRA data (logit transformed) was more ambitious because we also had to cope with the issue of PCR-introduced biases, which were minimized when transforming our eDNA data to a categorical approach. The best model fit was obtained with *total number of visits* and *days of last visit before sampling* (Model 2), but none of these variables were significant. The RRA per taxa is assumed to be correlated to the released DNA, that is, the initial biomass of taxa in a sample (Deagle et al., 2019). It was difficult to correlate RRA data to visiting patterns to the water body. In fact, the degradation of DNA in the water combined with the continuous turnover of new DNA creates a complex multivariate dynamic system of DNA concentration and quality in the water body which is captured only once at the moment of sampling. In our study, the 40 days range of CT monitoring exceeds the duration free eDNA remains detectable in water. This could explain the lack of significance for the explanatory variables of Model 2. We tested this same model only with data from the last 5 days before sampling ($Dr5_{ij}$), and we obtained a better fit but remaining non-significant. Nevertheless, the proposed DNA detectability scores calculated from CT data successfully represented the detection of species through eDNA, surpassing 75% of positive detection for D_{ij} scores above 25 (Figure 5a). Furthermore, species that visited the water bodies more than 25 days before sampling were never detected (Figure 5b), which indicates the maximum day-span between sampling events. However, these numbers apply only to our particular study system. DNA degradation and its detectability through DNA metabarcoding are very sensitive to environmental conditions of the sampled area, and future studies should target the effects of additional biophysical (such as *pH*, temperature, UV radiation, water body size, and depth) or biological variables (such as bacterial activity). Increasing the resolution and ecotype range of this kind of study will contribute to defining the probability of species detection through eDNA and contribute to improving sampling strategy for future research.

The direct comparison of detection success is strongly biased by the different survey lengths of this study and we found, therefore uncontested, CT to detect more taxa than the eDNA approach. Both methods have undoubtedly pros and cons, and it is crucial to weigh those and adapt the sampling strategy to the respective study system. Our results on detectability suggest that CT is preferential over eDNA for monitoring species when eDNA sampling cannot be made at regular intervals or to cover long monitoring periods. It ultimately also depends on the targeted taxa, being preferential when you study medium- to large-sized organisms which can be easily differentiated morphologically (Mena et al., 2021). Species abundances and densities can also be deduced from image data more easily. However, to rely on cameras implies the risk of losing data (over 40 days, 1/14 cameras were lost) and demands repeated visits which may be complicated in remote areas (at least installation and recovery of cameras), where eDNA sampling offers valuable study opportunities (McInnes et al., 2021). Also, CT does not offer the same options as eDNA in terms of possible population genetic studies (Bohmann et al., 2018; Nguyen et al., 2021; Sigsgaard et al., 2016, 2020; Tilker et al., 2020; Wilting et al., 2021) and the extension to other phyla (as, e.g., invertebrates, plants, and bacteria). eDNA sampling is appealing due to its sensitivity, standardization, and non-invasiveness, as well as the independence from taxonomic experts for taxa identification. eDNA techniques may enable the detection of elusive species and the taxonomic differentiation of morphologically similar as well as of cryptic species (Thomsen & Willerslev, 2015). In terms of quantifying the abundance of species, the eDNA approach has to cope with the dynamics of DNA in the water body released by the visiting individuals. They create a continuous turnover of available DNA that can easily bias the inference from amplified DNA to species relative abundances. Furthermore, genetic data at this scale are per se unable to provide certain population dynamic parameters (sex, age, and absolute numbers of individuals). As we have seen in our data, regular sampling in short intervals is necessary to provide a complete picture in terms of species richness. Alternatively, eDNA surveys can be used, for example, on a broad geographic scale to get a first glimpse of the biodiversity of the area that can be locally refined with CT (Sales et al., 2020). In fact, the combination of the two methods is increasingly used in biomonitoring studies (Sales et al., 2020; Sales et al., 2020), which is advantageous due to their complementary strengths. This could be particularly beneficial when the study system includes both large and small taxa of interest. The water samples are going to be dominated by the DNA of large taxa, which are easily recorded by CT, but they complicate the amplification of small taxa. Dividing methodological efforts, for instance by including blocking primers of the large dominant species, will facilitate the detection of smaller species, which are more likely to be missed by CT.

Our results show that water bodies concentrate information about large remote regions that are difficult to access and monitor using conventional, observer-based methods. They function as eDNA reservoirs containing information about valuable ecosystems. In light of the manifold risks facing wildlife—particularly species in remote and threatened ecosystems—further cross-method tests are needed, to validate their parallel application and support their integration into conservation monitoring strategies.

ACKNOWLEDGMENTS

We thank the Ministry of Environment, Natural Resources Conservation and Tourism (Botswana) for the issuing of a research permit (ENT 8/36/4 XXX II [10]), and the Bavarian Association for Bird Protection (LBV, Germany) for fieldwork and logistic support in Gobi. We sincerely acknowledge the contributions made by Kealeboga Matshameko, Norman Keoagile, and Mogomotsi Mokgadi in water collection and field work proceedings. We thank Julie Guenat and Jérôme Wassef for their contribution in the laboratory and image coding, respectively. We thank Jérémy Gremion and Ara Monadjem for bird species identifications. We also thank Jonas Bylemans for the constructive discussions during data analyses and Peter James Stephenson for comments on an earlier version of the manuscript.

FUNDING INFORMATION

EM-C was supported by a fellowship in Life Sciences (Faculty of Biology and Medicine, University of Lausanne). We thank the Botswana International University of Science and Technology (BIUST) for providing funding and infrastructure support to MB and CN. This project was supported by a Swiss National Science Foundation grant to PC (nr. 31003A_179378) and to LF (nr. 310030_192512).

CONFLICT OF INTERESTS

PT is co-inventor of a patent related to the Vert01 primers for vertebrate identification using degraded template DNA. This patent only restricts commercial applications and has no impact on the use of this locus by academic researchers. All other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in the paper.

AUTHOR CONTRIBUTIONS

LF, PC, and OG designed the study and supervised all analyses. BN, SR, OG, PC, MB, and CN conducted fieldwork. CA provided logistic support. CS carried out laboratory work. BN, EM-C, SR, FC, and PC carried out species identifications and image coding based on camera trap data. EM-C conducted bioinformatics and data analyses and prepared the figures. EM-C and JS wrote the paper, with input from all other authors.

DATA AVAILABILITY STATEMENT

The DNA metabarcoding data generated for this study is available on DRYAD (https://doi.org/10.5061/dryad.wpzgmsbp2).

ORCID

Luca Fumagalli https://orcid.org/0000-0002-6648-2570

REFERENCES

Abrams, J. F., Hörig, L. A., Brozovic, R., Axtner, J., Crampton-Platt, A., Mohamed, A., Wong, S. T., Sollmann, R., Yu, D. W., & Wilting, A. (2019). Shifting up a gear with iDNA: From mammal detection events to standardised surveys. *Journal of Applied Ecology*, *56*(7), 1637–1648. https://doi.org/10.1111/1365-2664.13411

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

Axtner, J., Crampton-Platt, A., Hörig, L. A., Mohamed, A., Xu, C. C. Y., Yu, D. W., & Wilting, A. (2019). An efficient and robust laboratory workflow and tetrapod database for larger scale environmental DNA studies. *GigaScience*, *8*(4), giz029. https://doi.org/10.1093/gigascience/giz029

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, *17*(1), 1–17. https://doi.org/10.1007/s10592-015-0775-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Belle, C. C., Stoeckle, B. C., & Geist, J. (2019). Taxonomic and geographical representation of freshwater environmental DNA research in aquatic conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *29*(11), 1996–2009. https://doi.org/10.1002/aqc.3208

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, *29*(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Bohmann, K., Gopalakrishnan, S., Nielsen, M., Nielsen, L., Jones, G., Streicker, D. G., & Gilbert, M. T. P. (2018). Using DNA metabarcoding for simultaneous inference of common vampire bat diet and population structure. *Molecular Ecology Resources*, *18*(5), 1050–1063. https://doi.org/10.1111/1755-0998.12891

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. https://doi.org/10.1111/1755-0998.12428

Brito, J. C., Godinho, R., Martínez-Freiría, F., Pleguezuelos, J. M., Rebelo, H., Santos, X., Vale, C. G., Velo-Antón, G., Boratyński, Z., Carvalho, S. B., Ferreira, S., Gonçalves, D. V., Silva, T. L., Tarroso, P., Campos, J. C., Leite, J. V., Nogueira, J., Álvares, F., Sillero, N., … Carranza, S. (2014). Unravelling biodiversity, evolution and threats to conservation in the Sahara-Sahel. *Biological Reviews*, *89*(1), 215–231. https://doi.org/10.1111/brv.12049

Brooks, T. M., Mittermeier, R. A., da Fonseca, G. A. B., Gerlach, J., Hoffmann, M., Lamoreux, J. F., Mittermeier, C. G., Pilgrim, J. D., & Rodrigues, A. S. L. (2006). Global biodiversity conservation priorities. *Science*, *313*(5783), 58–61. https://doi.org/10.1126/science.1127609

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference. A practical information-theoretic approach*. Springer-Verlag.

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., & Boutin, S. (2015). Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Butchart, S. H. M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J. P. W., Almond, R. E. A., Baillie, J. E. M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K. E., Carr, G. M., Chanson, J., Chenery, A. M., Csirke, J., Davidson, N. C., Dentener, F., Foster, M., Galli, A., … Watson, R. (2010). Global biodiversity: Indicators of recent declines. *Science*, *328*(5982), 1164–1168. https://doi.org/10.1126/science.1187512

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*(4), 915–924. https://doi.org/10.1111/mec.12183

Caravaggi, A., Banks, P. B., Burton, A. C., Finlay, C. M. V., Haswell, P. M., Hayward, M. W., Rowcliffe, M. J., & Wood, M. D. (2017). A review of camera trapping for conservation behaviour research. *Remote Sensing in Ecology and Conservation*, *3*(3), 109–122. https://doi.org/10.1002/rse2.48

Caravaggi, A., Burton, A. C., Clark, D. A., Fisher, J. T., Grass, A., Green, S., Hobaiter, C., Hofmeester, T. R., Kalan, A. K., Rabaiotti, D., & Rivet, D. (2020). A review of factors to consider when using camera traps to study animal behavior to inform wildlife ecology and conservation. *Conservation Science and Practice*, *2*(e239). https://doi.org/10.1111/csp2.239

Coutant, O., Richard-Hansen, C., de Thoisy, B., Decotte, J.-B., Valentini, A., Dejean, T., Vigouroux, R., Murienne, J., & Brosse, S. (2021). Amazonian mammal monitoring using aquatic environmental DNA. *Molecular Ecology Resources*, *21*(6), 1875–1888. https://doi.org/10.1111/1755-0998.13393

Davis, J. A., Kerezsy, A., & Nicol, S. (2017). Springs: Conserving perennial water is critical in arid landscapes. *Biological Conservation*, *211*, 30–35. https://doi.org/10.1016/j.biocon.2016.12.036

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. https://doi.org/10.1111/1755-0998.12188

Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarke, L. J., Vesterinen, E. J., Clare, E. L., Kartzinel, T. R. & Eveson, J. P. (2019). Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology*, *28*(2), 391–406. https://doi.org/10.1111/mec.14734

Díaz, S., Settele, J., Brondízio, E., Ngo, H., Guèze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K., Butchart, S., Chan, K., Garibaldi, L., Ichii, K., Liu, J., Subrmanian, S., Midgley, G., Miloslavich, P., Molnár, Z., Obura, D., …, Zayas, C. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. IPBES Secretariat.

Durant, S. M., Pettorelli, N., Bashir, S., Woodroffe, R., Wacher, T., De Ornellas, P., Ransom, C., Abáigar, T., Abdelgadir, M., El Alqamy, H., Beddiaf, M., Belbachir, F., Belbachir-Bazi, A., Berbash, A. A., Beudels-Jamar, R., Boitani, L., Breitenmoser, C., Cano, M., Chardonnet, P., … Baillie, J. E. M. (2012). Forgotten biodiversity in desert ecosystems. *Science*, *336*(6087), 1379–1380. https://doi.org/10.1126/science.336.6087.1379

Durant, S. M., Wacher, T., Bashir, S., Woodroffe, R., De Ornellas, P., Ransom, C., Newby, J., Abáigar, T., Abdelgadir, M., El Alqamy, H., Baillie, J., Beddiaf, M., Belbachir, F., Belbachir-Bazi, A., Berbash, A. A., Bemadjim, N. E., Beudels-Jamar, R., Boitani, L., Breitenmoser, C., … Pettorelli, N. (2014). Fiddling in biodiversity hotspots while deserts burn? Collapse of the Sahara's Megafauna. *Diversity and Distributions*, *20*(1), 114–122. https://doi.org/10.1111/ddi.12157

Egeter, B., Peixoto, S., Brito, J. C., Jarman, S., Puppo, P., & Velo-Anton, G. (2018). Challenges for assessing vertebrate diversity in turbid Saharan water-bodies using environmental DNA. *Genome*, *61*(11), 807–814. https://doi.org/10.1139/gen-2018-0071

Ficetola, G., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., & Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. *BMC Genomics*, *11*(1), 434. https://doi.org/10.1186/1471-2164-11-434

Fonseca, V. G. (2018). Pitfalls in relative abundance estimation using eDNA metabarcoding. *Molecular Ecology Resources*, *18*(5), 923–926. https://doi.org/10.1111/1755-0998.12902

Furlan, E. M., Davis, J., & Duncan, R. P. (2020). Identifying error and accurately interpreting environmental DNA metabarcoding results: A case study to detect vertebrates at arid zone waterholes. *Molecular Ecology Resources*, *20*(5), 1259–1276. https://doi.org/10.1111/1755-0998.13170

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

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., Domaizon, I., Coissac, E., David, F., Choler, P., Poulenard, J., & Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications*, *5*, 3211. https://doi.org/10.1038/ncomms4211

Gogarten, J. F., Hoffmann, C., Arandjelovic, M., Sachse, A., Merkel, K., Dieguez, P., Agbor, A., Angedakin, S., Brazzola, G., Jones, S., Langergraber, K. E., Lee, K., Marrocoli, S., Murai, M., Sommer, V., Kühl, H., Leendertz, F. H., & Calvignac-Spencer, S. (2020). Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity. *Environmental DNA*, *2*(1), 63–76. https://doi.org/10.1002/edn3.46

Harper, L. R., Lawson Handley, L., Carpenter, A. I., Ghazali, M., Di Muri, C., Macgregor, C. J., Logan, T. W., Law, A., Breithaupt, T., Read, D. S., McDevitt, A. D., & Hänfling, B. (2019). Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals. *Biological Conservation*, *238*, 108225. https://doi.org/10.1016/j.biocon.2019.108225

Ishige, T., Miya, M., Ushio, M., Sado, T., Ushioda, M., Maebashi, K., Yonechi, R., Lagan, P., & Matsubayashi, H. (2017). Tropical-forest mammals as detected by environmental DNA at natural saltlicks in Borneo. *Biological Conservation*, *210*, 281–285. https://doi.org/10.1016/j.biocon.2017.04.023

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., … Purvis, A. (2009). PanTHERIA: A species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, *90*(9), 2648. https://doi.org/10.1890/08-1494.1

Jusino, M. A., Banik, M. T., Palmer, J. M., Wray, A. K., Xiao, L., Pelton, E., Barber, J. R., Kawahara, A. Y., Gratton, C., Peery, M. Z., & Lindner, D. L. (2019). An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Molecular Ecology Resources*, *19*(1), 176–190. https://doi.org/10.1111/1755-0998.12951

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

Kennedy, S., Lim, J. Y., Clavel, J., Krehenwinkel, H., & Gillespie, R. G. (2019). Spider webs, stable isotopes and molecular gut content analysis: Multiple lines of evidence support trophic niche differentiation in a community of Hawaiian spiders. *Functional Ecology*, *33*(9), 1722–1733. https://doi.org/10.1111/1365-2435.13361

Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B., & Gillespie, R. G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, *7*(1), 17668. https://doi.org/10.1038/s41598-017-17333-x

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *2*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Leempoel, K., Hebert, T., & Hadly, E. A. (2020). A comparison of eDNA to camera trapping for assessment of terrestrial mammal diversity. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1918), 20192353–https://doi.org/10.1098/rspb.2019.2353

Littlewood, N. A., Hancock, M. H., Newey, S., Shackelford, G., & Toney, R. (2021). Use of a novel camera trapping approach to measure small mammal responses to peatland restoration. *European Journal of Wildlife Research*, *67*(12). https://doi.org/10.1007/s10344-020-01449-z

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B., & Ackerly, D. D. (2009). The velocity of climate change. *Nature*, *462*(7276), 1052–1055. https://doi.org/10.1038/nature08649

Lyet, A., Pellissier, L., Valentini, A., Dejean, T., Hehmeyer, A., & Naidoo, R. (2021). eDNA sampled from stream networks correlates with camera trap detection rates of terrestrial mammals. *Scientific Reports*, *11*(1):11362. https://doi.org/10.1038/s41598-021-90598-5

Lynggaard, C., Nielsen, M., Santos-Bay, L., Gastauer, M., Oliveira, G., & Bohmann, K. (2019). Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests. *Environmental DNA*, *1*(4), 329–341. https://doi.org/10.1002/edn3.34

Masonick, P., Hernandez, M., & Weirauch, C. (2019). No guts, no glory: Gut content metabarcoding unveils the diet of a flower-associated coastal sage scrub predator. *Ecosphere*, *10*(5), e02712. https://doi.org/10.1002/ecs2.2712

McInnes, J. C., Bird, J. P., Deagle, B. E., Polanowski, A. M., & Shaw, J. D. (2021). Using DNA metabarcoding to detect burrowing seabirds

in a remote landscape. *Conservation Science and Practice*, *3*(7), e4391–e4396. https://doi.org/10.1111/csp2.439

Mena, J. L., Yagui, H., Tejeda, V., Bonifaz, E., Bellemain, E., Valentini, A., Tobler, M. W., Sánchez-Vendizú, P., & Lyet, A. (2021). Environmental DNA metabarcoding as a useful tool for evaluating terrestrial mammal diversity in tropical forests. *Ecological Applications*, *31*(5), e02335. https://doi.org/10.1002/eap.2335

Mercier, C., Boyer, F., Bonin, A., & Coissac, E. (2013). *SUMACLUST: Fast and exact comparison and clustering of sequences*. Paper presented at the Programs and Abstracts of the SeqBio 2013 workshop (Abstract), GdRBIM and gdrIM, Montpellier, France.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Nasanbat, B., Ceacero, F., & Ravchig, S. (2021). A small neighborhood well-organized: seasonal and daily activity patterns of the community of large and mid-sized mammals around waterholes in the Gobi Desert, Mongolia. *Frontiers in Zoology*, *18*(25). https://doi.org/10.1186/s12983-021-00412-1

Nguyen, T. V., Tilker, A., Nguyen, A., Hörig, L., Axtner, J., Schmidt, A., Le, M., Nguyen, A. H. Q., Rawson, B. M., & Wilting, A. & Fickel, J. (2021). Using terrestrial leeches to assess the genetic diversity of an elusive species: The Annamite striped rabbit *Nesolagus timminsi*. *Environmental DNA*, *3*(4), 780–791. https://doi.org/10.1002/edn3.182

Pigot, A. L., Sheard, C., Miller, E. T., Bregman, T. P., Freeman, B. G., Roll, U., Seddon, N., Trisos, C. H., Weeks, B. C., & Tobias, J. A. (2020). Macroevolutionary convergence connects morphological form to ecological function in birds. *Nature Ecology & Evolution*, *4*(2), 230–239. https://doi.org/10.1038/s41559-019-1070-4

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., & Waits, L. P. (2014). Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular Ecology Resources*, *14*(1), 109–116. https://doi.org/10.1111/1755-0998.12159

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M., & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*(6187), 1246752. https://doi.org/10.1126/science.1246752

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

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

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., Roset, N., Schabuss, M., Zornig, H., & Dejean, T. (2018). Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports*, *8*(1), 10361. https://doi.org/10.1038/s41598-018-28424-8

Razgour, O., Persey, M., Shamir, U., & Korine, C. (2018). The role of climate, water and biotic interactions in shaping biodiversity patterns in arid environments across spatial scales. *Diversity and Distributions*, *24*(10), 1440–1452. https://doi.org/10.1111/ddi.12773

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, *51*(5), 1450–1459. https://doi.org/10.1111/1365-2664.12306

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. https://doi.org/10.1093/nar/gkr732

Rodgers, T. W., & Mock, K. E. (2015). Drinking water as a source of environmental DNA for the detection of terrestrial wildlife species. *Conservation Genetics Resources*, *7*(3), 693–696. https://doi.org/10.1007/s12686-015-0478-7

Rodgers, T. W., Xu, C. C. Y., Giacalone, J., Kapheim, K. M., Saltonstall, K., Vargas, M., Yu, D. W., Panu Somervuo, W., & McMillan, O. & Jansen, P. A. (2017). Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community. *Molecular Ecology Resources*, *17*(6), e133–e145. https://doi.org/10.1111/1755-0998.12701

Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Menzel, A., Root, T. L., Estrella, N., Seguin, B., Tryjanowski, P., Liu, C., Rawlins, S., & Imeson, A. (2008). Attributing physical and biological impacts to anthropogenic climate change. *Nature*, *453*(7193), 353–357. https://doi.org/10.1038/nature06937

Rovero, F., & Zimmermann, F. (Eds.) (2016). *Camera trapping for wildlife research*. Pelagic Publishing.

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*, *17*, e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Safriel, U., Adeel, Z., Niemeijer, D., Puigdefabregas, J., White, R., Lal, R., Winslow, M., Ziedler, J., Prince, S., Archer, E., King, C., Shapiro, B., Wessels, K., Nielsen, T. T., Portnov, B., Reshef, I., Thornell, J., Lachman, E., & McNab, D. (2005). Dryland systems. In R. Hassan, R. Scholes, & N. Ash (Eds.), *Ecosystems and human well-being: Current state and trends: Findings of the condition and Trends Working Group* (pp. 623–662). Island Press.

Sales, N. G., Kaizer, M. D. C., Coscia, I., Perkins, J. C., Highlands, A., Boubli, J. P., Magnusson, W. E., Da Silva, M. N. F., Benvenuto, C., & Mcdevitt, A. D. (2020). Assessing the potential of environmental DNA metabarcoding for monitoring Neotropical mammals: A case study in the Amazon and Atlantic Forest, Brazil. *Mammal Review*, *50*(3), 221–225. https://doi.org/10.1111/mam.12183

Sales, N. G., McKenzie, M. B., Drake, J., Harper, L. R., Browett, S. S., Coscia, I., Wangensteen, O. S., Baillie, C., Bryce, E., Dawson, D. A., Ochu, E., Hänfling, B., Lawson Handley, L., Mariani, S., Lambin, X., Sutherland, C., & McDevitt, A. D. (2020). Fishing for mammals: Landscape-level monitoring of terrestrial and semi-aquatic communities using eDNA from riverine systems. *Journal of Applied Ecology*, *57*(4), 707–716. https://doi.org/10.1111/1365-2664.13592

Salvatori, M., Tenan, S., Oberosler, V., Augugliaro, C., Christe, P., Groff, C., Krofel, M., Zimmermann, F., & Rovero, F. (2021). Co-occurrence of snow leopard, wolf and Siberian ibex under livestock encroachment into protected areas across the Mongolian Altai. *Biological Conservation*, *261*, 109294. https://doi.org/10.1016/j.biocon.2021.109294

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

Schubert, G., Stockhausen, M., Hoffmann, C., Merkel, K., Vigilant, L., Leendertz, F. H., & Calvignac-Spencer, S. (2015). Targeted detection of mammalian species using carrion fly-derived DNA. *Molecular Ecology Resources*, *15*(2), 285–294. https://doi.org/10.1111/1755-0998.12306

Seeber, P. A., McEwen, G. K., Löber, U., Förster, D. W., East, M. L., Melzheimer, J., & Greenwood, A. D. (2019). Terrestrial mammal surveillance using hybridization capture of environmental DNA from African waterholes. *Molecular Ecology Resources*, *19*(6), 1486–1496. https://doi.org/10.1111/1755-0998.13069

Shehzad, W., Riaz, T., Nawaz, M. A., Miquel, C., Poillot, C., Shah, S. A., Pompanon, F., Coissac, E., & Taberlet, P. (2012). Carnivore diet analysis based on next-generation sequencing: Application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, *21*(8), 1951–1965. https://doi.org/10.1111/j.1365-294X.2011.05424.x

Sigsgaard, E. E., Jensen, M. R., Winkelmann, I. E., Møller, P. R., Hansen, M. M., & Thomsen, P. F. (2020). Population-level inferences from environmental DNA-Current status and future perspectives. *Evolutionary Applications*, *13*(2), 245–262. https://doi.org/10.1111/eva.12882

Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., Pedersen, M. W., Jaidah, M. A., Orlando, L., Willerslev, E., Møller, P. R., & Thomsen, P. F. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, *1*(1), 4. https://doi.org/10.1038/s41559-016-0004

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54–71. https://doi.org/10.1037/1082-989X.11.1.54

Soininen, E. M., Zinger, L., Gielly, L., Bellemain, E., Bråthen, K. A., Brochmann, C., Epp, L. S., Gussarova, G., Hassel, K., Henden, J.-A., Killengreen, S. T., Rämä, T., Stenøien, H. K., Yoccoz, N. G., & Ims, R. A. (2013). Shedding new light on the diet of Norwegian lemmings: DNA metabarcoding of stomach content. *Polar Biology*, *36*(7), 1069–1076. https://doi.org/10.1007/s00300-013-1328-2

Stephenson, P. J. (2020). Technological advances in biodiversity monitoring: Applicability, opportunities and challenges. *Current Opinion in Environmental Sustainability*, *45*, 36–41. https://doi.org/10.1016/j.cosust.2020.08.005

Stone, L. S., Stone, T. M., & Mbaiwa, J. E. (2017). Tourism in Botswana in the last 50 Years: A review. *Botswana Notes and Records*, *49*, 57–72.

Strickler, K. M., Fremier, A. K., & Goldberg, C. S. (2015). Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation*, *183*, 85–92. https://doi.org/10.1016/j.biocon.2014.11.038

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

Taberlet, P., Bonin, A., Zinger, L., & Coissac, É. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press.

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

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Tilker, A., Abrams, J. F., Nguyen, A. N., Hörig, L., Axtner, J., Louvrier, J., Rawson, B. M., Nguyen, H. A. Q., Guegan, F., Nguyen, T. V., Le, M., Sollmann, R., & Wilting, A. (2020). Identifying conservation priorities in a defaunated tropical biodiversity hotspot. *Diversity and Distributions*, *26*(4), 426–440. https://doi.org/10.1111/ddi.13029

Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., & Miya, M. (2017). Environmental DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources*, *17*(6):e63–e75. https://doi.org/10.1111/1755-0998.12690

Ushio, M., Murata, K., Sado, T., Nishiumi, I., Takeshita, M., Iwasaki, W., & Miya, M. (2018). Demonstration of the potential of environmental DNA as a tool for the detection of avian species. *Scientific Reports*, *8*(1), 4493. https://doi.org/10.1038/s41598-018-22817-5

Vale, C. G., Pimm, S. L., & Brito, J. C. (2015). Overlooked mountain rock pools in deserts are critical local hotspots of biodiversity. *PLoS One*, *10*(2), e0118367. https://doi.org/10.1371/journal.pone.0118367

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.-M., … Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*(4), 929–942. https://doi.org/10.1111/mec.13428

Vestheim, H., & Jarman, S. N. (2008). Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, *5*(1), 12. https://doi.org/10.1186/1742-9994-5-12

Weiskopf, S. R., McCarthy, K. P., Tessler, M., Rahman, H. A., McCarthy, J. L., Hersch, R., Faisal, M. M., & Siddall, M. E. (2018). Using terrestrial haematophagous leeches to enhance tropical biodiversity monitoring programmes in Bangladesh. *Journal of Applied Ecology*, *55*(4), 2071–2081. https://doi.org/10.1111/1365-2664.13111

Wilting, A., Nguyen, T. V., Axtner, J., Nguyen, A. N., Schmidt, A., Le, M., Nguyen, A. H. Q., Rawson, B. M., Tilker, A., & Fickel, J. (2021). Creating genetic reference datasets: Indirect sampling of target species using terrestrial leeches as sample "collectors". *Environmental DNA*, 1–15. https://doi.org/10.1002/edn3.256

WWF (2020). In R. E. A. Almond, M. Grooten, & T. Petersen (Eds.), *Living Planet Report 2020 - Bending the curve of biodiversity loss*. WWF.

Yoccoz, N. G., Bråthen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., Von Stedingk, H., Brysting, A. K., Coissac, E., Pompanon, F., Sønstebø, J. H., Miquel, C., Valentini, A., De Bello, F., Chave, J., Thuiller, W., Wincker, P., Cruaud, C., Gavory, F., … Taberlet, P. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, *21*(15), 3647–3655. https://doi.org/10.1111/j.1365-294X.2012.05545.x

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., Gaucher, P., Gielly, L., Giguet-Covex, C., Iribar, A., Réjou-Méchain, M., Rayé, G., Rioux, D., Schilling, V., Tymen, B., Viers, J., Zouiten, C., Thuiller, W., Coissac, E., & Chave, J. (2019). Body size determines soil community assembly in a tropical forest. *Molecular Ecology*, *28*(3), 528–543. https://doi.org/10.1111/mec.14919

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Mas-Carrió, E., Schneider, J., Nasanbat, B., Ravchig, S., Buxton, M., Nyamukondiwa, C., Stoffel, C., Augugliaro, C., Ceacero, F., Taberlet, P., Glaizot, O., Christe, P., & Fumagalli, L. (2022). Assessing environmental DNA metabarcoding and camera trap surveys as complementary tools for biomonitoring of remote desert water bodies. *Environmental DNA*, 4, 580–595. https://doi.org/10.1002/edn3.274