



## Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests

Lynggaard, Christina; Nielsen, Martin; Santosbay, Luisa; Gastauer, Markus; Oliveira, Guilherme; Bohmann, Kristine

*Published in:*  
Environmental Engineering

*DOI:*  
[10.1002/edn3.34](https://doi.org/10.1002/edn3.34)

*Publication date:*  
2019

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY](#)

*Citation for published version (APA):*  
Lynggaard, C., Nielsen, M., Santosbay, L., Gastauer, M., Oliveira, G., & Bohmann, K. (2019). Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests. *Environmental Engineering*, 1(4), 329-341. <https://doi.org/10.1002/edn3.34>

# Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests

Christina Lynggaard<sup>1</sup>  | Martin Nielsen<sup>1</sup>  | Luisa Santos-Bay<sup>1,2</sup>  |  
Markus Gastauer<sup>3</sup>  | Guilherme Oliveira<sup>3</sup>  | Kristine Bohmann<sup>1,4</sup> 

<sup>1</sup>Section for Evolutionary Genomics, The GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen K, Denmark

<sup>2</sup>Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

<sup>3</sup>Instituto Tecnológico Vale, Belém, Pará, Brazil

<sup>4</sup>School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, UK

## Correspondence

Martin Nielsen, Section for Evolutionary Genomics, The GLOBE Institute, Faculty of Health and Medical Sciences, Øster Farimagsgade 5a, entrance 7, 2. floor, University of Copenhagen, Copenhagen K., Denmark.  
Email: mnielsen@science.ku.dk

## Funding information

Dr. Bøje Benzons Støttefond; CABANA project, Grant/Award Number: RCUK (BB/P027849/1); Consejo Nacional de Ciencia y Tecnología, Grant/Award Number: CVU 582667; Brazilian Ministry of Environment, Grant/Award Number: SISBIO 67580-1 and 7079-00029B; Tanzania National Parks (TANAPA); Tanzania Commission for Science and Technology, Grant/Award Number: 2014-138-ER-2012-147; Tanzania Wildlife Research Institute; Vale (Canga Plant & Soil and Biodiversity Genomics); CNPq, Grant/Award Number: 444227/2018-0, 402756/2018-5 and 307479/2016-1; CAPES, Grant/Award Number: 88887.130628/2016-00; Danish Council for Independent Research, Grant/Award Number: SISGEN R04409B; Independent Research Fund Denmark, Grant/Award Number: 5051-00140

## Abstract

**Background:** Thousands of bulk arthropod samples are collected globally every year for monitoring programs, conservation efforts, and ecosystem assessments. The taxonomic contents of these samples can be assessed either morphologically or molecularly using DNA metabarcoding coupled with high-throughput sequencing, the latter of which has gained popularity in recent years. In a related field, only vertebrate-ingesting invertebrates, such as carrion flies and blood-feeding leeches, are targeted for collection, and metabarcoding is carried out on the vertebrate DNA in their gut contents to provide information on vertebrate diversity (invertebrate-derived DNA, iDNA).

**Aims:** Here, we show that the two approaches can be combined, that is, that vertebrate DNA can be detected through metabarcoding of bulk arthropod samples. **Materials and Methods:** Two metabarcoding primer sets were used to PCR amplify mammal and vertebrate DNA in DNA extracted from bulk arthropod samples collected with pitfall and Malaise traps in tropical forests in Brazil and Tanzania.

**Results:** In total, 32 vertebrate taxa were detected representing mammals, amphibians, and birds. Detected taxa were within, or close to, their known geographical distributions.

**Conclusion:** This study demonstrates that with a relatively small additional investment, information on vertebrate diversity can be obtained from bulk arthropod samples. This is of particular interest in projects where bulk arthropod samples are collected and extracted with the aim to use metabarcoding to assess arthropod taxa. In such studies, the additional information on vertebrates can further inform ecological assessments and monitoring programs and function as a supplement to traditional survey methods of vertebrates.

## KEYWORDS

biodiversity, environmental DNA, invertebrate-derived DNA, vertebrate monitoring

Lynggaard and Nielsen are joint first authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Environmental DNA* published by John Wiley & Sons Ltd

## 1 | INTRODUCTION

Fauna monitoring can be used to assess ecosystem health; detect invasive, rare, and indicator species; define areas for conservation priority settings; and inform biodiversity and ecosystem management decisions (Hajibabaei et al., 2011; Hilty & Merenlender, 2000; Ji et al., 2013; Liu, Guo, Zhong, & Shi, 2018). For this, arthropods are a suitable taxonomic group as they occur in habitats all over the world and are important in ecosystem functioning, for example, by decomposition of organic matter, pollination, and serving as a food source for many aquatic and terrestrial animals (Rosenberg, Danks, & Lehmkuhl, 1986; Siddig, Ellison, Ochs, Villar-Leeman, & Lau, 2016). On the other hand, vertebrate diversity information may improve the achievement of management goals and public acceptance of management decisions. This is because the presence of popular charismatic vertebrate species such as flagship, keystone, or umbrella birds and mammals (Heywood & Watson 1995; Simberloff, 1998) is perceived more positively than the occurrence of invertebrate taxa (Hunter et al., 2016).

The surveys of vertebrates and invertebrates present different challenges. Taxonomic identification of vertebrate species is generally not challenging, but low abundances and shy behavior of the often-crepuscular animals can make direct surveys time- and labor intense, especially in remote areas and in areas with dense vegetation. Therefore, indirect methods are often applied to monitor vertebrate fauna such as the collection of road kills (Teixeira, Coelho, Esperandio, & Kindel, 2013) and identification of signs such as tracks, nests, and scats (Hoffmann et al., 2010). In contrast, while arthropods occur in high abundance and are easily sampled (Rosenberg et al., 1986), morphology-based taxonomic identification of arthropod community samples requires not only taxonomic expertise for multiple taxonomic groups but also a significant time investment to identify all the taxonomic constituents (Basset et al., 2012). Given the difficulty of their identification, molecular analyses are increasingly applied to identify the taxonomic contents of bulk arthropod samples (Ashfaq et al., 2018; Elbrecht et al., 2017; Gibson et al., 2014; Kocher, Gantler et al., 2017; Liu et al., 2018; Morinière et al., 2016; Oliverio, Gan, Wickings, & Fierer, 2018; Shokralla et al., 2015; Yu et al., 2012). For this, DNA metabarcoding approaches are the most frequently applied. Metabarcoding principally relies on PCR amplification of DNA extracts using primers that are universal for a selected taxonomic group targeted by a taxonomically informative “barcode marker.” Unique identifiers are added to sample amplicons before they are sequenced in parallel on a high-throughput sequencing platform. Following sequencing, the identifiers are used to trace the marker sequences back to the samples they originated from (Binladen et al., 2007; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012). After additional computational processing, the taxa present in the samples can be identified by comparing the “barcode” sequences obtained to a DNA reference database (Ratnasingham & Hebert, 2007). The targeted nature of metabarcoding means that it is a cost-effective and efficient method for identifying the taxonomic contents of hundreds to thousands of samples in

parallel (Galan et al., 2018; Taberlet et al., 2012). Metabarcoding of bulk insect samples has, for example, been used to characterize the diversity of insects in montane landscapes in tropical southern China (Zhang et al., 2016), explore the insect diversity in a Saharo-Arabian region with otherwise sparse fauna information (Ashfaq et al., 2018), and monitor temporal changes in arthropod communities in different forest types (Brandon-Mong et al., 2018).

A large number of invertebrate species feed on vertebrates and thereby sample their DNA. Recently, this so-called iDNA, short for invertebrate-derived DNA, has been used to monitor vertebrates. In these iDNA studies, metabarcoding is typically used to target taxonomically informative vertebrate DNA markers in DNA extracted from individual or pooled samples of invertebrates known to feed on flesh, blood, feces, and/or dead or decaying organic matter (reviewed in Calvignac-Spencer et al., 2013; Schnell, Sollmann, et al., 2015). Since the field of iDNA originated, targeted collection followed by iDNA analyses of gut contents has been carried out on different invertebrate taxa such as leeches (Drinkwater et al., 2019; Pérez-Flores, Rueda-Calderon, Kvist, Siddall, & Ocegüera-Figueroa, 2016; Schnell et al., 2018; Weiskopf et al., 2018), sand flies (Kocher, De Thoisy, et al., 2017), blow and flesh flies (Calvignac-Spencer et al., 2013; Hoffmann et al., 2018; Lee, Gan, Clements, & Wilson, 2016; Lee, Sing, & Wilson, 2015; Rodgers et al., 2017; Schubert et al., 2015), mosquitoes (Kocher, De Thoisy, et al., 2017), ticks (Garipey et al., 2012), marine copepods (Meekan et al., 2017), and shrimps (Siegenthaler et al., 2019). This has offered a new and promising tool to complement traditional vertebrate monitoring methods, something of great value in the ongoing biodiversity monitoring efforts (Bohmann et al., 2013).

This raises the question: Can these two study fields be merged? That is, is it possible to use iDNA methods to detect vertebrate DNA in bulk arthropod samples without targeting a specific vertebrate-feeding invertebrate, and thereby optimize the biodiversity information gained from bulk arthropod samples? Based on the accomplishments of the field of iDNA, the answer should be yes. However, given the aforementioned targeted nature of metabarcoding, studies assessing arthropod taxa in bulk arthropod samples using metabarcoding have so far only identified arthropod taxa and therefore the two fields, metabarcoding of bulk arthropod samples and of iDNA, have until now been quite separated.

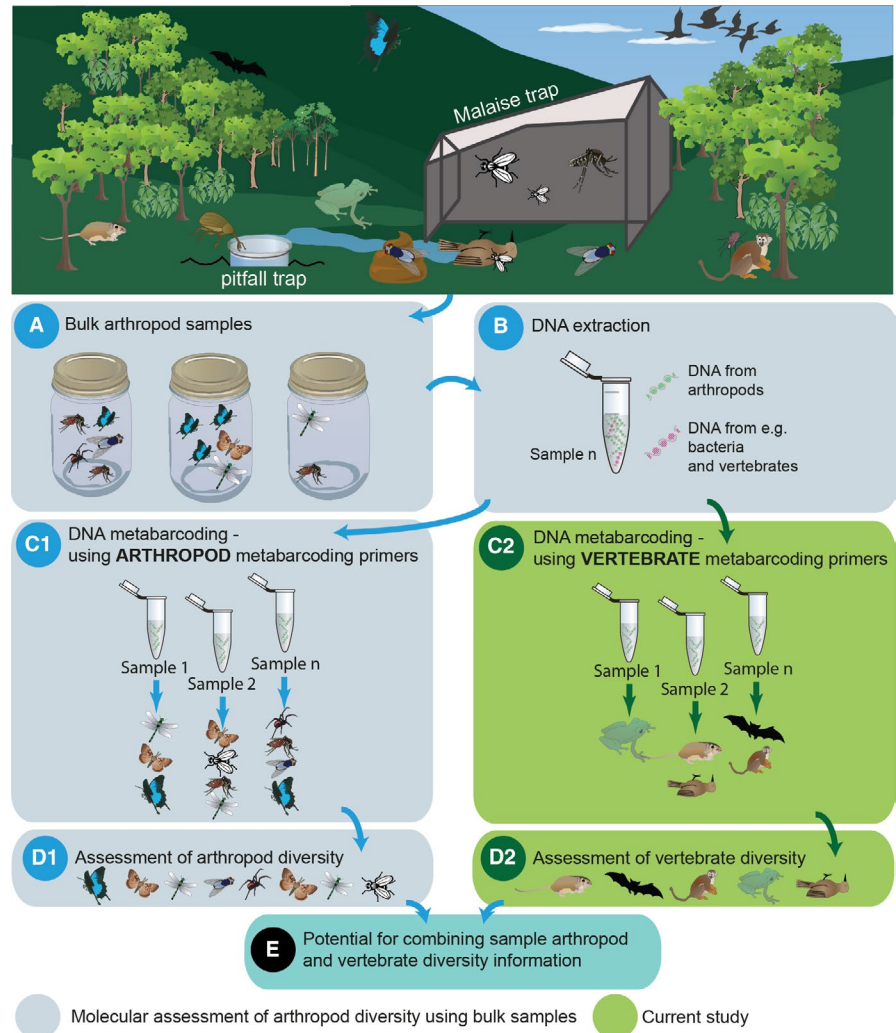
In this study, we evaluate whether it is possible to obtain information on vertebrate taxa through metabarcoding of bulk arthropod samples (Figure 1). To investigate this, we used vertebrate and mammal metabarcoding primers on DNA extracted from bulk arthropod samples collected with Malaise and pitfall traps in the Carajás National Forest in Brazil and the Udzungwa Mountains in Tanzania.

## 2 | MATERIALS AND METHODS

### 2.1 | Study sites and sample collection

Bulk arthropod samples were collected as part of ongoing diversity studies in Brazil and Tanzania. Bulk arthropod samples were collected in Malaise and pitfall traps. In Brazil, bulk arthropod

**FIGURE 1** The use of bulk arthropod samples to assess arthropod and vertebrate diversity starting from setting traps in the field (top), collection of bulk arthropod samples (A), extracting DNA (B), carrying out DNA metabarcoding (C1, C2) to assess diversity (D1, D2) and combining the obtained arthropod and vertebrate diversity data (E). Grey colour represents the general workflow when assessing arthropod diversity using DNA metabarcoding and green colour represents the current study. Many images within the illustrative figure courtesy of the Integration and Application Network, University of Maryland Center for Environmental Science ([ian.umces.edu/symbols/](http://ian.umces.edu/symbols/))



samples were collected in September 2017 (dry season) and April 2018 (wet season) in an iron mine area (06°03'31"S 50°10'37"W) in the Carajás National Forest, Pará State (Figure 2). Collection sites included pristine moist Amazonian equatorial forest, savanna-like Canga ecosystems (Mitre et al., 2019), and sites at different stages of environmental rehabilitation following mining. Malaise traps were left for 5 days, and pitfall traps were left for 24 hr before collection. Arthropods were collected in 70% ethanol in pitfall traps and in propylene glycol in Malaise traps. A total of 50 Malaise and 50 pitfall samples were collected. Samples were stored at room temperature for a maximum of a week before DNA extraction.

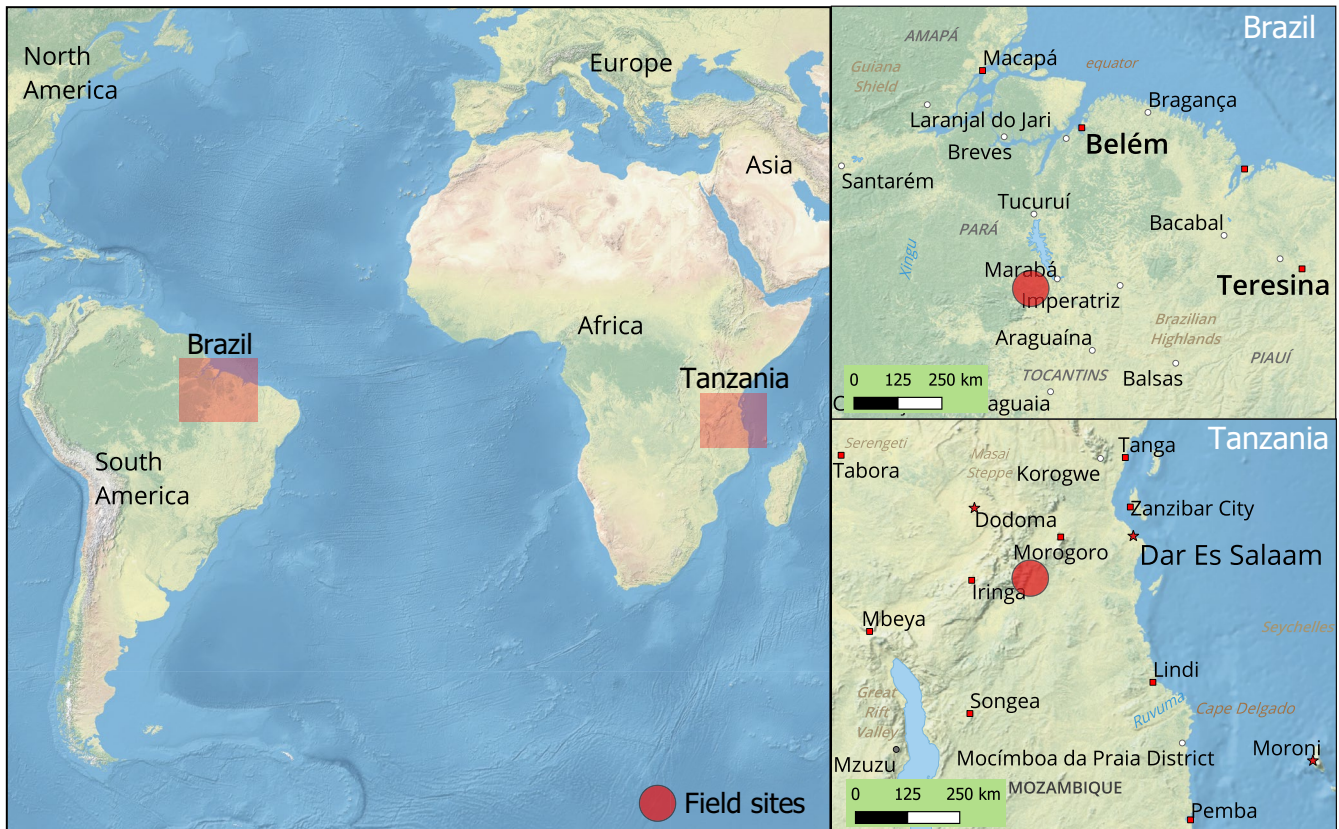
In Tanzania, bulk arthropod samples were collected in a mountainous rainforest about 1,000 m above sea level in the Udzungwa Mountains (07°41'07"S 36°55'49"E) (Figure 2). Samples were collected in September and October 2014 (the end of the dry season). Traps were placed in three locations, about 500 m apart, emptied every day for 7 days, and then emptied every other day for three collection events and finally every week for three collection events. Propylene glycol was used as collection fluid and samples transferred to 70% ethanol upon collection. A total of 78 Malaise and 78 pitfall samples were collected. After collection, samples were kept

at ambient temperature for a maximum of 2 weeks after which they were stored at  $-20^{\circ}\text{C}$  until DNA extraction.

## 2.2 | DNA extraction

For samples stored in ethanol, ethanol was carefully poured off the samples and arthropods were transferred to falcon tubes. Falcon tubes were placed in an oven at  $55^{\circ}\text{C}$  without lids to evaporate the remaining ethanol. For samples stored in propylene glycol, propylene glycol was carefully poured off with no further evaporation before DNA extraction. Samples with a volume larger than 30 ml were split into two before DNA extraction. This resulted in a total of 103 extracts from Brazil and 162 from Tanzania giving a total of 265 extracts. A negative extraction control was included for every 11–25 samples. Samples were extracted following a nondestructive protocol modified from Gilbert et al. (2007). In this protocol, a digest buffer is added to unsorted bulk arthropod samples that were not homogenized previously and therefore preserving the exoskeleton (Nielsen, Gilbert, Pape, & Bohmann, 2019). Following extraction, 200  $\mu\text{l}$  digest from samples and negative extraction controls were purified using the QiaQuick PCR Purification Kit (Qiagen) following





**FIGURE 2** Study sites in Brazil and Tanzania where bulk arthropod samples were collected. Maps created with QGIS (version 3.6.2)

the manufacturer's protocol with minor modifications. Specifically, after addition of elution buffer, samples were incubated for 15 min at 37°C after which they were eluted in 50 µl EB buffer and stored in Eppendorf LoBind tubes at -18°C.

DNA extractions of Tanzanian samples were carried out in a pre-PCR laboratory to minimize contamination risk. The Brazilian samples were extracted in a general use laboratory.

### 2.3 | DNA metabarcoding

Two metabarcoding primer sets were used to PCR amplify mammal and vertebrate DNA in the DNA extracted from the bulk arthropod samples. For mammals, a ca. 95 bp 16S rRNA mitochondrial marker was PCR-amplified with the primers 16Smam1 forward 5'-CGGTTGGGGTGACCTCGGA-3' and 16Smam2 reverse 5'-GCTGTTATCCCTAGGGTAACT-3' (Taylor, 1996). For vertebrates, a ca. 97 bp fragment of the 12S gene was PCR-amplified with the primer set 12SV05 forward 5'-TTAGATACCCCACTATGC-3' and 12SV05 reverse 5'-TAGAACAGGCTCCTCTAG-3' (Riaz et al., 2011). The two metabarcoding primer sets will be referred to as 16S mammal and 12S vertebrate primers, respectively. Nucleotide tags were added to the 5' ends of both forward and reverse primers to allow parallel sequencing (Binladen et al., 2007). Specifically, tags consisted of a total of 7–8 nucleotides of which 6 nucleotides were the tags and 1–2 were nucleotides added to increase complexity on the flow cell during sequencing (De Barba et al., 2014).

For each of the two primer sets, tagged PCRs were carried out with three PCR replicates for each of the 265 extracts and negative extraction controls. Furthermore, four to five positive controls were included, namely *Canis lupus* (wolf), *Ursus maritimus* (polar bear), *Zalophus californianus* (California sea lion), and *Ursus arctos* (brown bear), and additionally for the vertebrate 12S primer set, *Giraffa camelopardalis* (giraffe). Negative controls were included in PCR amplifications with both primer sets. PCR amplifications were performed with nonmatching nucleotide tags (e.g., forward primer tag 1- reverse primer tag 2, forward primer tag 1- reverse primer tag 3, and forward primer tag 1- reverse primer tag 4) to allow for more amplicons to be pooled together and reduce laboratory costs (Schnell, Bohmann, & Gilbert, 2015). Moreover, every PCR replicate for each sample was made with a different tag combination.

For the 16S mammal primers, the 25 µl reactions consisted of 1 µl DNA template, 1 U AmpliTaq Gold, 1× Gold PCR Buffer, and 2.5 mM MgCl<sub>2</sub> (all from Applied Biosystems); 0.6 µM each of 5' nucleotide tagged forward and reverse primer; 0.2 mM dNTP mix (Invitrogen); 0.5 mg/ml bovine serum albumin (BSA); and 5 µM human blocker (5'-3' GCGACCTCGGAGCAGAACCC-spacerC3) (Vestheim & Jarman, 2008). The thermal cycling profile was 95°C for 10 min, followed by 40 cycles of 94°C for 12 s, 59°C for 30 s, and 70°C for 25 s, with a final extension time of 72°C for 7 min. To evaluate the effect of using a human blocker, the extracts from Tanzania were also PCR-amplified with the same conditions but omitting human blocker (see Appendix S1).

For the 12S vertebrate primers, reactions were equal to that of the mammal primer except using 0.75 U AmpliTaq Gold, 20  $\mu$ l reaction volumes, and the human blocker (5'-3' TACCCCACTATGCTTAGCCCTAAACCTCAACAGTTAAATC-spacerC3) (Calvignac-Spencer et al., 2013). The thermal cycling profile was 95°C for 10 min, followed by 40 cycles of 94°C for 30 s, 59°C for 45 s, and 72°C for 60 s, with a final extension time of 72°C for 7 min.

Amplified PCR products were visualized on 2% agarose gels with GelRed against a 50 bp ladder. All negative controls appeared negative. Products with visible bands on the agarose gels, carrying different nucleotide tag combinations, were pooled as follows: For the 16S mammal primer, for Brazilian samples, any successfully amplified PCR product was pooled, while, for the Tanzanian samples, only PCR products from samples where all three PCR replicates had successfully amplified were pooled. For the 12S vertebrate primer, PCR replicates that successfully amplified were pooled if at least two out of the three PCR replicates for a sample had successfully amplified.

PCR products from positive controls and negative extraction controls were included in each amplicon pool. Amplicon pools were purified with SPRI beads (Rohland & Reich, 2012) with a 1.6 $\times$  bead-to-amplicon pool ratio and eluted in 35  $\mu$ l EB buffer. Purified amplicon pools were built into sequence libraries with an in-house protocol in which Illumina sequencing adapters and dual indices were ligated onto amplicons. The protocol omits the two steps that have been shown to cause tag jumps, that is, T4 DNA polymerase blunt ending and postligation PCR (Schnell, Bohmann, & Gilbert, 2015; van Orsouw et al., 2007). Libraries were purified with a 0.8 $\times$  bead-to-amplicon pool ratio and eluted in 30  $\mu$ l EB buffer and qPCR-quantified using the NEBNext Library Quant Kit for Illumina (New England Biolabs Inc.). Amplicon libraries were pooled and sequenced at the National High-throughput DNA Sequencing Centre, University of Copenhagen. Sequence libraries were sequenced 250 bp PE on an Illumina MiSeq sequencing platform using v2 chemistry, aiming for 25,000 paired reads per PCR replicate.

## 2.4 | Data processing and analyses

Sequence data were processed for each primer set separately. Using AdapterRemoval v2.2.2, sequence reads were trimmed to remove adaptors and low-quality bases and paired reads were merged (Schubert, Lindgreen, & Orlando, 2016). Sequences were sorted according to primers and tags using a modified version of DAME (Bohmann et al., 2018; Zepeda-Mendoza, Bohmann, Carmona Baez, & Gilbert, 2016, <https://github.com/shyamsg/DAME>). Thresholds for filtering sequences across the PCR replicates from each sample were guided by the sequenced negative and positive controls, and sequences were retained if having at least 15 and 37 sequence copies for the 16S mammal and 12S vertebrate primer sets, respectively (Alberdi et al., 2018). Further, sequences present in any of a sample's PCR replicates were kept. The filtered sequences were clustered using SUMACLUSt with a similarity score of 97% (Mercier, Boyer, Bonin, & Coissac, 2013). Postclustering curation of the operational

taxonomic unit (OTU) tables was carried out using the LULU algorithm with default settings (Frøslev et al., 2017). The LULU algorithm is independent of DNA reference databases and is composed of a core mechanism that retains rare and factual OTUs while discarding artifactual OTUs. It does so by identifying and merging the artifactual OTUs with factual abundant OTUs that are similar in sequence and that consistently co-occur.

The OTU sequences were compared against the NCBI Genbank database ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)) using BLASTn, and the output was imported into MEGAN Community Edition version 6.12.7 (Huson et al., 2016) using a weighted LCA algorithm with 90 as percent to cover, top percent of 2, and a min score of 150. The taxonomic assignment was complemented with a probabilistic taxonomic assignment method, PROTAX (Axtner et al., 2019; Somervuo, Koskela, Pennanen, Henrik Nilsson, & Ovaskainen, 2016), using the weighted model, which uses a list of the expected vertebrate species for each locality (script available from URL <https://github.com/dougwyu/screenforbio-mbc-ailaoshan>). Information from the PROTAX method was included if the family and genus probabilities were above .8. Any OTUs assigned to Hominidae were discarded from the datasets. For the 16S mammal dataset, our positive control contained an OTU assigned to gray wolf, *C. lupus*, as expected, but so did three samples from Brazil. For the 12S vertebrate dataset, an OTU assigned to Phasianidae was unexpectedly found in a positive control and in one sample from Tanzania. The wolf and Phasianidae OTUs were discarded before further analysis.

Taxonomy of all OTUs was further manually checked to validate assignments. A strict species assignment approach was applied so that species-level assignment was only performed when an OTU sequence had an identity of 100% to a NCBI reference sequence. However, one OTU sequence with 98% identity to *Nandinia binotata* (African palm civet) was assigned to species level as the taxonomic family that it belongs to consists of only this one species. All detected taxa were evaluated according to their known geographical distribution (<https://www.iucnredlist.org>). One OTU assigned to *Tapirus* sp. had 100% identity to two tapir species but was assigned to *Tapirus terrestris* based on the two species' known geographical distributions. A Krona chart (Ondov, Bergman, & Phillippy, 2011) was created for a visual representation of the taxonomic distribution of detected vertebrates. To test for differences in detection rates between trap types, primers, season (Brazilian dataset only), and countries, we built individual general linearized models for the entire dataset, as well as for the Tanzanian and the Brazilian dataset, fitted using binomial logit links. We adjusted *p*-values for multiple comparisons as suggested by Bonferroni (Amstrong, 2014).

## 3 | RESULTS

Of the analyzed 265 bulk arthropod sample DNA extracts, 76 (28.7%) were sequenced. Curation of the 16S mammal and 12S vertebrate OTU tables with the postclustering algorithm LULU (Frøslev et al.,





**TABLE 1** Vertebrate taxa detected in bulk arthropod samples from Brazil and Tanzania using DNA metabarcoding. N/A indicates that taxonomic identification to species or genus level was not possible. Information regarding trap type (Malaise or pitfall) and number of vertebrate detections, metabarcoding primer set (v: 12S vertebrate, Riaz et al., 2011; or m: 16S mammal, Taylor, 1996), whether the taxa were observed during sample collection, if its known geographical distribution falls within the collection site and finally the IUCN status (<https://www.iucnredlist.org/>), least concern (LC) or vulnerable (VU) of the taxa are shown

Taxonomy of the detected vertebrates										Sample type and no. of detections			Within geo. distribution	IUCN status							
Country	Class	Order	Family	Genus	Species	Common name	Malaise	Pitfall	Primer	Observed during sample collecting											
Brazil	Amphibia	Anura	Hylidae	N/A	sp. 1	Frog	1		v	Yes	Yes	LC									
						Frog	3		v	Yes	Yes										
						Frog	1		v	Yes	Yes										
	Mammalia	Artiodactyla	Carnivora	Cervidae	N/A	N/A	Deer	2		m	Yes	Yes	LC								
							Canid	1		v	Yes	Yes									
							Molossus	1		m	Velvety free-tailed bat	1		m	Yes	LC					
							Artibeus	2		m	Neotropical fruit bat	2		m	Yes	Yes					
							Tapirus	1		m	South American tapir	1		m	Yes	VU					
							Alouatta	2		m	Brown howler monkey	2	1	m	Yes	LC					
							Saguinus	1		m	Tamarins	1		m	Yes	Yes					
Rodentia	Cricetidae	Echimyidae	N/A	N/A	Oecomys rat	1		v	Yes	Yes											
					Proechimys rat	1		m	Yes	Yes											
					Rattus	1		m	Black rat	1		m	Yes	LC							
Tanzania	Amphibia	Anura	Arthroleptidae	Arthroleptis	A. xenodactyloides	Chirinda screeching frog	1	19	v	Yes	Yes	LC									
						N/A	1		v	Yes	Yes										
	Aves	Galliformes	Phasianidae	N/A	Guttera	G. pucherani	Crested guineafowl	1	1	v	Yes	Yes	LC								
							A. albifrons	1		v	Yes	Yes									
							B. indicator	1		v	Yes	Yes									
		Mammalia	Passeriformes	Pycnonotidae	Turdidae	N/A	N/A	Honeyguide Greenbul	1	1	v	Yes	Yes	LC							
								Thrush	1		v	Yes	Yes								
								African Barbets	1		v	Yes	Yes								
								Red forest duiker	2		m	Yes	Yes								
			Piciformes	Artiodactyla	Bovidae	Cephalophus	Neotragus	N. moschatus	Suni	1	1	m	Yes	Yes	LC						
									Pig	1		m	Yes	Yes							
									African palm civet	3		m	Yes	Yes							
			Primates	Carnivora	Nandinidae	Cercopithecoidea	Nandinia	N. binotata	White-eyed mangabey	1	1	m	Yes	Yes	LC						
									Cercopithecus	4		m	Yes	Yes							
									Colobus	1		m	Yes	Yes							
									Procolobus	2		m	Yes	Yes							
									Paragalago	3		m/v	Yes	Yes							
									Rodentia	Cricetidae	Muridae	N/A	N/A	P. zanzibaricus	Zanzibar bushbaby	2	2	v	Yes	Yes	LC
															N/A	1		v	Yes	Yes	
Grammomys	1														m	Yes	Yes				





**FIGURE 4** Visual observations during bulk arthropod sample collection confirmed the presence of some of the detected vertebrate taxa. Left: tapir footprint in one of the Brazilian study sites. Right: Paragalago zanzibaricus (Bushbaby) in the Udzungwa Mountains in the Tanzanian study site

per extract with either of the primer sets (mean of 1.28, *SD* 0.53). Of the 32 vertebrate taxa detected, 14 were identified to species level, 10 to genus level, and the remaining 8 to family level (Figure 3). Eight OTUs could not be identified to a lower taxonomic level than order and were discarded. Six samples only contained at least one of these OTUs and were therefore also discarded.

### 3.2 | Detected vertebrate taxa

The 32 detected vertebrate taxa encompassed 21 mammalian taxa spanning 15 families in 6 orders (Artiodactyla, Carnivora, Chiroptera, Perissodactyla, Primates, and Rodentia), 6 bird taxa spanning 6 families in 3 orders (Galliformes, Passeriformes, and Piciformes), and 5 amphibian taxa spanning 3 families in 1 order (Anura) (Figure 3, Table 1). Fourteen vertebrate taxa could be assigned to species level. Of these, 12 were known to occur in the study sites (Table 1). Although the two remaining species, *Alouatta guariba* (brown howler monkey) in Brazil and *Baeopogon indicator* (honeyguide greenbul) in Tanzania, are not known to occur within the study sites, their distribution falls close to these ([www.iucnredlist.org](http://www.iucnredlist.org)). Furthermore, five of the detected species were confirmed through visual observations during sample collection (Figure 4, Table 1). The detected vertebrates are ecologically diverse, ranging from large-bodied animals such as *T. terrestris* (South American tapir) to small-sized animals such as *Amblyospiza albifrons* (thick-billed weaver) and *Arthroleptis xenodactyloides* (Chirinda screeching frog). Furthermore, vertebrates encompassing different trophic levels with a range of different food preferences were identified, including ruminant herbivores such as Cervidae gen. sp. (deer), nonruminant herbivores such as *A. guariba* (brown howler monkey), omnivores such as *N. binotata* (African palm civet), frugivores such as *Artibeus* sp. (neotropical fruit bats), and insectivores such as the amphibian *Dendropsophus* sp. Their lifestyles also vary widely, from ground-dwelling Cervidae gen. sp. (deer) and *T. terrestris* (tapir); arboreal primates (e.g., *Colobus angolensis*) and rodents (e.g., *Oecomys* sp.); volant birds (e.g., *B. indicator*) and bats (e.g., *Molossus molossus*); nocturnal bats (e.g., *Artibeus* sp.), civets (*N. binotata*), and diurnal birds (e.g., *B. indicator*); and primates (e.g., *A. guariba*) (Figure 4, Table 1). The most detected vertebrate in Tanzania was *A. xenodactyloides* (Chirinda screeching frog), which was detected in 20 extracts. In Brazil, *A. guariba*

(brown howler monkey) was the most detected taxa, with detection in three extracts (Table 1).

### 3.3 | Vertebrate detections per trap type, the time before collection and season

When combining the two primer sets, no differences were found in detection rates of vertebrates in samples collected in Tanzania compared to Brazil. Further, differences between trap types were found, as the majority (13 out of 14) of samples in which vertebrates were detected in Brazil were collected with Malaise traps ( $t$ -value = 2.490, adjusted  $p$ -value = .037), whereas in Tanzania more vertebrates were detected in pitfall trap samples ( $t$ -value = 2.767, adjusted  $p$ -value = .0107). Adjusted  $p$ -values indicate no significant differences in detection rates between primer sets in both datasets. Regarding the amount of days that traps were left before collection in Tanzania, it was possible to detect vertebrate DNA in extracts originating from traps that were left for a longer time (2–7 days). Although not statistically significant due to small sample sizes, the success rate of vertebrate detection in these samples did not seem to decrease (Table S1). In spite of samples collected during the wet season in Brazil having a higher detection rate (23.5%), compared to the samples collected during dry season (0%), detection rate did not differ statistically between sampling season, most likely caused by the low number of samples.

### 3.4 | Primer performance

The 16S mammal primers amplified DNA in 44 (16.6%) of the 265 extracts, while the 12S vertebrate primers amplified DNA in 32 (12.1%) extracts. Of these sequenced extracts, vertebrate DNA, which could be taxonomically identified to family level or lower, was detected in 25 (56.9%) of the extracts amplified using the 16S mammal primer and in 29 (90.6%) of those amplified with the 12S vertebrate primer set. In five extracts, vertebrate taxa were identified with both primer sets. Of the 32 detected vertebrate taxa, 17 were detected with the 16S mammal primer set and 15 with the 12S vertebrate primer set. Thus, almost twice as many taxa were detected when using both primer sets as opposed to when only using one of them (Table 1). In one sample, the same vertebrate taxon was detected using the two different primer sets.

## 4 | DISCUSSION

While studies using invertebrate-derived DNA, iDNA, have so far focused on the targeted collection of invertebrates ingesting vertebrates and their genetic material, in this study we show that vertebrate DNA can also be detected in bulk arthropod samples, that is, without targeting a specific invertebrate taxon during collections and therefore with no prior knowledge of the collected arthropod taxa.

### 4.1 | Vertebrate detection

Through metabarcoding using 16S mammal and 12S vertebrate primers, vertebrate taxa were detected in 19.2% of all the analyzed bulk arthropod sample DNA extracts. This detection rate is below those reported by iDNA studies targeting specific invertebrate taxa where some studies reported detection rates of 21%–100% (reviewed in Calvignac-Spencer et al., 2013). Our relatively low vertebrate detection rate might simply be because the untargeted nature of the collection meant that not all samples contained invertebrates that had ingested vertebrate DNA. An additional explanation could lie in the complex mixture of DNA found in bulk arthropod samples. Specifically, as many invertebrates are pooled together, including both species that do and do not feed on vertebrate-derived samples, overall the invertebrate DNA will dominate any traces of vertebrate DNA. Some iDNA studies have found that the number of detected vertebrates increases when invertebrates are extracted individually in contrast to pooling the invertebrates (Rodgers et al., 2017; Schnell et al., 2012). As such, the pooled nature of the bulk arthropod samples in the present study could explain the low proportion of samples with vertebrate detections. However, sequencing pooled blowflies resulted in the detection of four additional vertebrate taxa as in comparison with individual sequencing (Calvignac-Spencer et al., 2013).

It should be noted that we only sequenced some DNA extracts, more specifically those with successful PCR amplification as assessed by gel electrophoresis, and that the vertebrate detection rate was relatively high in the sequenced sample extracts (90.6% and 56.9% for 12S vertebrate and 16S mammal primer set, respectively). Thirty-two vertebrate taxa were detected in the study sites in Brazil and Tanzania. Although study areas and design differ, this is comparable to the number of vertebrates detected in iDNA studies targeting specific invertebrate taxa such as carrion flies in Côte d'Ivoire and Madagascar (Calvignac-Spencer et al., 2013), carrion flies in Panama (Rodgers et al., 2017), leeches in Borneo (Schnell et al., 2018), blowflies in Malaysia (Lee et al., 2016), leeches in Vietnam (Schnell et al., 2012), and even higher than when targeting ticks in Canada (Garipey et al., 2012).

While the traditional survey methods are generally limited to vertebrate species from a single forest stratum (generally near ground level), we show that bulk arthropod samples can cover different forest strata, including the detection of canopy-occupying birds and primates (Figure 3 and Table 1). The wide range of vertebrates detected in the present study indicates that bulk arthropod

samples offer great potential for supplementing traditional methods for vertebrate surveying such as camera trapping, spoor tracking, and other visual surveys, as already shown for iDNA studies using leeches (Abrams et al., 2019).

Not all identified vertebrate taxa might originate from vertebrates that were ingested by invertebrates. For instance, 19 of the 20 frog detections in Tanzania originated from pitfall traps, which corresponds with the occasional observation of frogs in pitfall traps. These were, however, discarded from samples upon collection. Frog detections in the bulk samples might therefore be caused by invertebrates ingesting frog DNA but may also be caused by DNA traces of frogs in the collecting fluid.

While 12 of the 14 vertebrate taxa identified to species level had a geographical distribution within the study sites, two species did not (Table 1). Nevertheless, the known geographical distributions of these two species, *A. guariba* (brown howler monkey) in Brazil and *B. indicator* (honeyguide greenbul) in Tanzania, are close to the study sites, and therefore, it is not implausible that they might be found in the study areas.

Another explanation is that detection of vertebrate taxa in bulk arthropod samples provides evidence of the presence in a larger area. Because bulk arthropod samples likely consist of arthropod taxa occupying different habitats, having different feeding strategies and dispersal potentials, care should be taken when making inferences about the geographical location and temporal proximity of the detected vertebrates (Lee et al., 2016; Schnell, Sollmann, et al., 2015). A third explanation might be a relatively incomplete DNA reference database. The primate *Alouatta belzebul* (red-handed howler) belonging to the same genus as the detected *A. guariba* (brown howler monkey) is known to occur within the Brazilian sample site, but a DNA reference for the 16S marker used in this study has yet to be included in public databases such as Genbank. It is also possible that *A. belzebul* and *A. guariba* are identical over the 16S DNA barcode marker used to identify it in this study and that we have detected *A. belzebul* in our samples. This highlights the need for further development of DNA reference databases.

### 4.2 | Vertebrate detections per trap type, the time before collection, and season

In the present study, it was possible to detect vertebrate taxa in bulk arthropod samples from both Malaise and pitfall traps (Table 1). We therefore believe this can also be achieved from samples collected with other kinds of traps, such as light traps, pan traps, hand netting, and other mass collecting traps. More vertebrates were detected in samples collected by Malaise traps using the 16S mammal primer set, whereas the 12S vertebrate primers detected the most vertebrates in pitfall samples (Table 1). Although this might be location-specific, or as in our case influenced by frogs falling into pitfall traps in Tanzania, increasing the types of traps will naturally expand the invertebrate taxa collected and therefore likely also increase the diversity of the vertebrates detected. When collecting bulk arthropod

samples, the degradation rate of vertebrate DNA inside the arthropods has to be considered. It has been found that amplifiable vertebrate DNA in *Chrysomya megacephala* (blowflies) decreased markedly after only a few days postfeeding (Lee et al., 2015), but on the contrary, goat DNA has been found to persist in blood-feeding leeches for several months (Schnell et al., 2012). Additionally, in the present study, vertebrate DNA was detected in traps that had been left for several days (up to 7 days) before collection (Table S1). Nevertheless, it has been shown that arthropod DNA undergoes slight degradation when in a Malaise trap (Krehenwinkel et al., 2018) and most likely also the iDNA. To our knowledge, no studies have investigated the DNA degradation of prey DNA inside collected arthropods over time, but a trade-off likely exists between the number of days a trap can be left before collecting and the degradation of ingested vertebrate DNA, which depends on the invertebrate taxa. Although not statistically significant in this study, another factor affecting vertebrate detections in bulk arthropod samples is related to seasonality, as no vertebrates were detected in samples collected during the dry season in Brazil. This could indicate that arthropods ingesting vertebrate DNA are more abundant during the wet season in Brazil, and therefore, samples collected during the dry season are poor candidates for detecting vertebrates.

### 4.3 | Primer performance

When combining results from the 16S mammal (Taylor, 1996) and the 12S vertebrate (Riaz et al., 2011), primer sets almost twice as many vertebrate taxa were detected as compared to when either primer set was used alone (Table 1). This confirms the findings of Rodgers et al. (2017) that used the same primer sets on DNA extracted from pools of up to 16 carrion flies and found that more species were detected using both primers than using either marker alone. In agreement with our results, they also more often detected primates using the 16S mammal primer as opposed to the 12S vertebrate primer. Interestingly, we only obtained one overlapping vertebrate detection for the two primer sets and only in one sample. One explanation for this could be the incomplete reference database as both primer sets detected vertebrate OTUs which could not be taxonomically identified beyond order level and were therefore excluded for further analyses. Additionally, in some samples the 12S vertebrate primer only detected amphibian taxa, whereas the 16S primer detected mammal taxa in the same samples. Although not formally tested in this study, it seems like the 12S vertebrate primer set has an affinity toward amphibian DNA, which could also cause the discrepancy between the two primer sets. This shows the complementarity of the two primer sets and highlights the need to use both primer sets and potentially additional primer sets, to increase vertebrate detections in future studies. For example, no taxa of the class Reptilia were detected in the present study, which might be caused by the inability of the primers to amplify reptilian DNA. Therefore, we highly encourage research to optimize primer choice to enhance vertebrate detections.

### 4.4 | Technical considerations

When attempting to PCR amplify low amounts of template DNA, it is important to consider PCR stochasticity as it can lead to false negatives and thus the failure of amplification of certain taxa in some PCR replicates (Kebschull & Zador, 2015; Murray, Coghlan, & Bunce, 2015). Incorporation of additional PCR replicates can increase the probability of amplifying target DNA in low quantity (Alberdi et al., 2018). Although our study design did not allow a detailed assessment of how the number of PCR replicates influences vertebrate detection rates in bulk arthropod samples, some observations can be made. For the Brazilian samples, four vertebrate taxa would not have been detected if samples were only pooled when all three PCR replicates showed successful amplification when visualized on gel electrophoresis (data not shown). Similarly, for the 12S vertebrate primer, three vertebrate taxa would not have been detected in samples from Tanzania (data not shown). As for the 16S primer in the Tanzanian samples, it is possible that vertebrate taxa were missed, as the pooling strategy was stricter. This argues for careful consideration when deciding, based on the results of the gel electrophoresis, which and how many PCR replicates from each sample to pool and sequence when the aim is to detect vertebrate taxa in bulk arthropod samples. For example, even though vertebrate DNA is only successfully amplified in a single PCR replicate, our results highlight that this single PCR replicate should still be included in the following pooling. However, this requires careful consideration when filtering the DNA sequences during data analysis where it is necessary to keep sequences appearing in any of the sequenced PCR replicates, which may introduce false positives (Alberdi et al., 2018). A postclustering curation could therefore be applied to identify and delete some of the false positives (Alberdi et al., 2018; Frøslev et al., 2017). To optimize the potential of detecting vertebrate DNA and limit the risk of false positives, another but more costly approach can be to include additional PCR replicates (e.g., five) which then would permit a stricter filtering (e.g., only keeping sequences occurring in min. 3/5 PCR replicates). Finally, after deciding the number of PCR replicates to use and how to analyze the data, it is important to minimize type II errors (false negatives). Therefore, undetected species should not be treated as absent as the vertebrate species can be present in the study site but not have been fed upon by the sampled arthropods (Alberdi et al., 2019).

### 4.5 | Perspectives

We demonstrate that bulk arthropod samples should no longer be considered only to provide information about the arthropod communities but also as a source of vertebrate fauna information. It can require many field days to collect bulk arthropod samples, and once the samples have been brought into the laboratory, the hardship continues as sample preparation and DNA extraction of the sometimes hundreds of samples can require many man-hours. Therefore, once having collected and extracted bulk arthropod samples, researchers should obtain as much biodiversity information as possible from the samples. Using bulk arthropod samples to detect

vertebrate diversity can be of particular interest in larger projects (e.g., the Global Malaise Trap Program <http://biodiversitygenomics.net/projects/gmp/>) where samples have already been collected and DNA extracted. Moreover, this method could serve as a supplement to vertebrate monitoring such as camera trapping, visual surveys, and iDNA studies. By using the proposed approach, researchers can increase the value of bulk arthropod samples for ecological assessment and monitoring programs.

## ACKNOWLEDGMENTS

We thank Martin Ellegaard, Wesley de Melo Rangel, Gilliana Almeida da Rosa, Silvio Junio Ramos, and Cecílio Frois Caldeira for field assistance in Brazil, M. Thomas P. Gilbert for logistical and financial support, and Jagoba Malumbres-Olarte, Thomas Pape, Nikolaj Scharff, Richard M. L. Laizzer, and Aloyce Mwakisoma for logistics and fieldwork assistance in Tanzania. Furthermore, we thank the Udzungwa Ecological Monitoring Centre (UEMC) and its staff for logistical support in Tanzania. Collection permits in Brazil were granted by the Brazilian Ministry of Environment (SISBIO 67580-1 and SISGEN R04409B). Tanzanian export and collection permits were granted by Tanzania National Parks (TANAPA), the Tanzania Commission for Science and Technology (COSTECH: No. 2014-138-ER-2012-147), and the Tanzania Wildlife Research Institute (TAWIRI). Collections in Carajás, Brazil, received support from Vale (Canga Plant & Soil and Biodiversity Genomics), and CNPq to GO (444227/2018-0, 402756/2018-5, 307479/2016-1), CAPES to GO (88887.130628/2016-00), and the CABANA project (RCUK (BB/P027849/1)). We also thank Professor Douglas W. Yu for his valuable assistance with PROTAX. Finally, we thank the three reviewers for their valuable comments and suggestions. CL was supported by the National Council of Science and Technology in Mexico (CONACyT) Grant CVU 582667 and by the Danish Council for Independent Research, DFF grant 7079-00029B. KB was supported by the Independent Research Fund Denmark and DFF grant 5051-00140. MN was supported by Dr. Bøje Benzons Støttefond.

## CONFLICT OF INTEREST

None declared.

## AUTHOR CONTRIBUTIONS

CL, KB, and MN designed the study. CL, MG, and MN carried out sample collection. CL, LS-B, and MN generated the data. CL and MN analyzed the data. CL, KB, and MN wrote the original manuscript, and all authors contributed to revisions and accepted the final version.

## DATA AVAILABILITY STATEMENT

Sequence data and PROTAX weighted models are available at <http://www.doi.org/10.17894/ucph.263de3a8-9d79-46c9-9b5e-8adca7e6a716>.

## ORCID

- Christina Lynggaard  <https://orcid.org/0000-0003-4451-6642>  
 Martin Nielsen  <https://orcid.org/0000-0002-3718-526X>  
 Luisa Santos-Bay  <https://orcid.org/0000-0002-0554-6550>  
 Markus Gastauer  <https://orcid.org/0000-0002-9599-0902>  
 Guilherme Oliveira  <https://orcid.org/0000-0003-0054-3438>  
 Kristine Bohmann  <https://orcid.org/0000-0001-7907-064X>

## REFERENCES

- Abrams, J. F., Hoerig, L., Brozovic, R., Axtner, J., Crampton-Platt, A., Mohamed, A., ... Wilting, A. (2019). Shifting up a gear with iDNA: From mammal detection events to standardized surveys. *Journal of Applied Ecology*, *56*, 1637–1648. <https://doi.org/10.1111/1365-2664.13411>
- Alberdi, A., Aizpurua, O., Bohmann, K., Gopalakrishnan, S., Lynggaard, C., Nielsen, M., & Gilbert, M. T. P. (2019). Promises and pitfalls of using high-throughput sequencing for diet analysis. *Molecular Ecology Resources*, *19*, 327–348. <https://doi.org/10.1111/1755-0998.12960>
- Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), 134–147. <https://doi.org/10.1111/2041-210X.12849>
- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502–508. <https://doi.org/10.1111/opo.12131>
- Ashfaq, M., Sabir, J. S. M., El-Ansary, H. O., Perez, K., Levesque-Beaudin, V., Khan, A. M., ... Hebert, P. D. N. (2018). Insect diversity in the Saharo-Arabian region: Revealing a little-studied fauna by DNA barcoding. *PLoS One*, *13*(7), e0199965. <https://doi.org/10.1371/journal.pone.0199965>
- 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), 1–17. <https://doi.org/10.1093/gigascience/giz029>
- Basset, Y., Cizek, L., Cuénoud, P., Didham, R. K., Guilhaumon, F., Missa, O., ... Leponce, M. (2012). Arthropod diversity in a tropical forest. *Science*, *338*(6113), 1481–1484.
- Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., & Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One*, *2*(2), e197. <https://doi.org/10.1371/journal.pone.0000197>
- 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*(5), 1050–1063. <https://doi.org/10.1111/1755-0998.12891>
- Bohmann, K., Schnell, I. B., & Gilbert, M. T. P. (2013). When bugs reveal biodiversity. *Molecular Ecology*, *22*(4), 909–911. <https://doi.org/10.1111/mec.12221>
- Brandon-Mong, G. J., Littlefair, J. E., Sing, K. W., Lee, Y. P., Gan, H. M., Clare, E. L., & Wilson, J. J. (2018). Temporal changes in arthropod activity in tropical anthropogenic forests. *Bulletin of Entomological Research*, *108*(6), 792–799. <https://doi.org/10.1017/S000748531800010X>
- Calvignac-Spencer, S., Leendertz, F. H., Gilbert, M. T. P., & Schubert, G. (2013). An invertebrate stomach's view on vertebrate ecology:



- Certain invertebrates could be used as “vertebrate samplers” and deliver DNA-based information on many aspects of vertebrate ecology. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 35(11), 1004–1013.
- Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühn, H., Boesch, C., Kappeler, P. M., ... 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.
- 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>
- Drinkwater, R., Schnell, I. B., Bohmann, K., Bernard, H., Veron, G., Clare, E., ... Rossiter, S. J. (2019). Using metabarcoding to compare the suitability of two blood-feeding leech species for sampling mammalian diversity in North Borneo. *Molecular Ecology Resources*, 19(1), 105–117. <https://doi.org/10.1111/1755-0998.12943>
- Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, 8(10), 1265–1275. <https://doi.org/10.1111/2041-210X.12789>
- Frøsløv, T. G., Kjølner, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, 8(1), 1188. <https://doi.org/10.1038/s41467-017-01312-x>
- Galan, M., Pons, J. B., Tournayre, O., Pierre, E., 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>
- Garipey, T. D., Lindsay, R., Ogden, N., & Gregory, T. R. (2012). Identifying the last supper: Utility of the DNA barcode library for bloodmeal identification in ticks. *Molecular Ecology Resources*, 12(4), 646–652. <https://doi.org/10.1111/j.1755-0998.2012.03140.x>
- Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., ... Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasytematics. *Proceedings of the National Academy of Sciences USA*, 111(22), 8007–8012. <https://doi.org/10.1073/pnas.1406468111>
- Gilbert, M. T. P., Moore, W., Melchior, L., & Worobey, M. (2007). DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS One*, 2(3), e272. <https://doi.org/10.1371/journal.pone.0000272>
- Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C., & Baird, D. J. (2011). Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One*, 6(4), e17497. <https://doi.org/10.1371/journal.pone.0017497>
- Heywood, V. H., & Watson, R. T. (1995). *Global biodiversity assessment* (Vol. 1140). Cambridge, UK: Cambridge University Press.
- Hilty, J., & Merenlender, A. (2000). Faunal indicator taxa selection for monitoring ecosystem health. *Biological Conservation*, 92(2), 185–197. [https://doi.org/10.1016/S0006-3207\(99\)00052-X](https://doi.org/10.1016/S0006-3207(99)00052-X)
- Hoffmann, A., Decher, J., Rovero, F., Schaer, J., Voigt, C., & Wibbelt, G. (2010). Field methods and techniques for monitoring mammals. In D. VandenSpiegel, J. C. Monje, C. Häuser, J. Degreef, & J. Eymann, editors *Manual on field recording techniques and protocols for all taxa biodiversity inventories and monitoring* (Vol. 8, pp. 482–529). Brussels, Belgium: Abc Taxa.
- Hoffmann, C., Merkel, K., Sachse, A., Rodríguez, P., Leendertz, F. H., & Calvignac-Spencer, S. (2018). Blow flies as urban wildlife sensors. *Molecular Ecology Resources*, 18(3), 502–510. <https://doi.org/10.1111/1755-0998.12754>
- Hunter, M., Westgate, M., Barton, P., Calhoun, A., Pierson, J., Tulloch, A., ... Lindenmayer, D. (2016). Two roles for ecological surrogacy: Indicator surrogates and management surrogates. *Ecological Indicators*, 63, 121–125. <https://doi.org/10.1016/j.ecolind.2015.11.049>
- Huson, D. H., Beier, S., Flade, I., Górška, A., El-Hadidi, M., Mitra, S., ... Tappu, R. (2016). MEGAN community edition – Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Computational Biology*, 12(6), e1004957.
- Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., ... Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, 16(10), 1245–1257. <https://doi.org/10.1111/ele.12162>
- Kebschull, J. M., & Zador, A. M. (2015). Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Research*, 43(21), e143. <https://doi.org/10.1093/nar/gkv717>
- Kocher, A., De Thoisy, B., Catzeflis, F., Valière, S., Bañuls, A. L., & Murienne, J. (2017). iDNA screening: Disease vectors as vertebrate samplers. *Molecular Ecology*, 26(22), 6478–6486. <https://doi.org/10.1111/mec.14362>
- Kocher, A., Gantier, J.-C., Gaborit, P., Zinger, L., Holota, H., Valiere, S., ... Murienne, J. (2017). Vector soup: High-throughput identification of neotropical phlebotomine sand flies using metabarcoding. *Molecular Ecology Resources*, 17(2), 172–182. <https://doi.org/10.1111/1755-0998.12556>
- Krehenwinkel, H., Fong, M., Kennedy, S., Huang, E. G., Noriyuki, S., Cayetano, L., & Gillespie, R. (2018). The effect of DNA degradation bias in passive sampling devices on metabarcoding studies of arthropod communities and their associated microbiota. *PLoS One*, 13(1), e0189188. <https://doi.org/10.1371/journal.pone.0189188>
- Lee, P. S., Gan, H. M., Clements, G. R., & Wilson, J. J. (2016). Field calibration of blowfly-derived DNA against traditional methods for assessing mammal diversity in tropical forests. *Genome*, 59(11), 1008–1022. <https://doi.org/10.1139/gen-2015-0193>
- Lee, P. S., Sing, K. W., & Wilson, J. J. (2015). Reading mammal diversity from flies: The persistence period of amplifiable mammal mtDNA in blowfly guts (*Chrysomya megacephala*) and a new DNA mini-barcode target. *PLoS One*, 10(4), e0123871. <https://doi.org/10.1371/journal.pone.0123871>
- Liu, L., Guo, Z., Zhong, C., & Shi, S. (2018). DNA barcoding reveals insect diversity in the mangrove ecosystems of Hainan Island, China. *Genome*, 61(11), 797–806. <https://doi.org/10.1139/gen-2018-0062>
- Meekan, M., Austin, C. M., Tan, M. H., Wei, N.-W., Miller, A., Pierce, S. J., ... Gan, H. M. (2017). iDNA at Sea: Recovery of whale shark (*Rhincodon typus*) mitochondrial DNA sequences from the whale shark copepod (*Pandarus rhincodonicus*) confirms global population structure. *Frontiers in Marine Science*, 4, 420. <https://doi.org/10.3389/fmars.2017.00420>
- Mercier, C., Boyer, F., Bonin, A., & Coissac, E. (2013). SUMATRA and SUMACLUSt: fast and exact comparison and clustering of sequences. In *Programs and Abstracts of the SeqBio 2013 workshop*. Abstract (pp. 27–29).
- Mitre, S. K., Mardegan, S. F., Caldeira, C. F., Ramos, S. J., Furtini Neto, A. E., Siqueira, J. O., & Gastauer, M. (2019). Nutrient and water dynamics of Amazonian canga vegetation differ among physiognomies and from those of other neotropical ecosystems. *Plant Ecology*, 11, 1341–1353.
- Morinière, J., Cancian de Araujo, B., Lam, A. W., Hausmann, A., Balke, M., Schmidt, S., ... Haszprunar, G. (2016). Species identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. *PLoS One*, 11(5), e0155497. <https://doi.org/10.1371/journal.pone.0155497>
- Murray, D. C., Coghlan, M. L., & Bunce, M. (2015). From benchtop to desktop: Important considerations when designing amplicon sequencing workflows. *PLoS One*, 10(4), e0124671. <https://doi.org/10.1371/journal.pone.0124671>
- Nielsen, M., Gilbert, M. T. P., Pape, T., & Bohmann, K. (2019). A simplified DNA extraction protocol for unsorted bulk arthropod samples that maintains exoskeletal integrity. *Environmental DNA*, 1(2), 144–154. <https://doi.org/10.1002/edn3.16>
- Oliverio, A. M., Gan, H., Wickings, K., & Fierer, N. (2018). A DNA metabarcoding approach to characterize soil arthropod communities. *Soil Biology & Biochemistry*, 125, 37–43. <https://doi.org/10.1016/j.soilbio.2018.06.026>

- Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, 12, 385. <https://doi.org/10.1186/1471-2105-12-385>
- Pérez-Flores, J., Rueda-Calderon, H., Kvist, S., Siddall, M. E., & Ocegueda-Figueroa, A. (2016). From the worm in a bottle of Mezal: iDNA confirmation of a leech parasitizing the Antillean manatee. *Journal of Parasitology*, 102(5), 553–555. <https://doi.org/10.1645/16-46>
- Ratnasingham, S., & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 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., Xu, C. C. Y., Giacalone, J., Kapheim, K. M., Saltonstall, K., Vargas, M., ... 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>
- Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, 22(5), 939–946. <https://doi.org/10.1101/gr.128124.111>
- Rosenberg, D. M., Danks, H. V., & Lehmkuhl, D. M. (1986). Importance of insects in environmental impact assessment. *Environmental Management*, 10(6), 773–783. <https://doi.org/10.1007/BF01867730>
- Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated—reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, 15(6), 1289–1303. <https://doi.org/10.1111/1755-0998.12402>
- Schnell, I. B., Bohmann, K., Schultze, S. E., Richter, S. R., Murray, D. C., Sinding, M.-H., ... Gilbert, M. T. P. (2018). Debugging diversity—a pancontinental exploration of the potential of terrestrial blood-feeding leeches as a vertebrate monitoring tool. *Molecular Ecology Resources*, 18(6), 1282–1298. <https://doi.org/10.1111/1755-0998.12912>
- Schnell, I. B., Sollmann, R., Calvignac-Spencer, S., Siddall, M. E., Yu, D. W., Wilting, A., & Gilbert, M. T. P. (2015). iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool – prospects, pitfalls and avenues to be developed. *Frontiers in Zoology*, 12, 24. <https://doi.org/10.1186/s12983-015-0115-z>
- Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R. D., Willerslev, E., ... Gilbert, M. T. P. (2012). Screening mammal biodiversity using DNA from leeches. *Current Biology*, 22(8), R262–R263.
- 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>
- Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes*, 9, 88. <https://doi.org/10.1186/s13104-016-1900-2>
- Shokralla, S., Porter, T. M., Gibson, J. F., Dobosz, R., Janzen, D. H., Hallwachs, W., ... Hajibabaei, M. (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports*, 5, 9687. <https://doi.org/10.1038/srep09687>
- Siddig, A. A. H., Ellison, A. M., Ochs, A., Villar-Leeman, C., & Lau, M. K. (2016). How do ecologists select and use indicator species to monitor ecological change? Insights from 14 years of publication in Ecological Indicators. *Ecological Indicators*, 60, 223–230. <https://doi.org/10.1016/j.ecolind.2015.06.036>
- Siegenthaler, A., Wangenstein, O. S., Soto, A. Z., Benvenuto, C., Corrigan, L., & Mariani, S. (2019). Metabarcoding of shrimp stomach content: Harnessing a natural sampler for fish biodiversity monitoring. *Molecular Ecology Resources*, 19(1), 206–220. <https://doi.org/10.1111/1755-0998.12956>
- Simberloff, D. (1998). Flagships, umbrellas, and keystones: Is single-species management passé in the landscape era? *Biological Conservation*, 83(3), 247–257. [https://doi.org/10.1016/S0006-3207\(97\)00081-5](https://doi.org/10.1016/S0006-3207(97)00081-5)
- Somervuo, P., Koskela, S., Pennanen, J., Henrik Nilsson, R., & Ovaskainen, O. (2016). Unbiased probabilistic taxonomic classification for DNA barcoding. *Bioinformatics*, 32(19), 2920–2927. <https://doi.org/10.1093/bioinformatics/btw346>
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21(8), 2045–2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>
- Taylor, P. G. (1996). Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Molecular Biology and Evolution*, 13(1), 283–285. <https://doi.org/10.1093/oxfordjournals.molbev.a025566>
- Teixeira, F. Z., Coelho, A. V. P., Esperandio, I. B., & Kindel, A. (2013). Vertebrate road mortality estimates: Effects of sampling methods and carcass removal. *Biological Conservation*, 157, 317–323. <https://doi.org/10.1016/j.biocon.2012.09.006>
- van Orsouw, N. J., Hogers, R. C. J., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., ... van Eijk, M. J. T. (2007). Complexity reduction of polymorphic sequences (CROPS): A novel approach for large-scale polymorphism discovery in complex genomes. *PLoS One*, 2, e1172. <https://doi.org/10.1371/journal.pone.0001172>
- 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., ... 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>
- Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, 3(4), 613–623. <https://doi.org/10.1111/j.2041-210X.2012.00198.x>
- Zepeda-Mendoza, M. L., Bohmann, K., Carmona Baez, A., & Gilbert, M. T. P. (2016). DAME: A toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. *BMC Research Notes*, 9, 255. <https://doi.org/10.1186/s13104-016-2064-9>
- Zhang, K., Lin, S., Ji, Y., Yang, C., Wang, X., Yang, C., ... Yu, D. W. (2016). Plant diversity accurately predicts insect diversity in two tropical landscapes. *Molecular Ecology*, 25(17), 4407–4419. <https://doi.org/10.1111/mec.13770>

## SUPPORTING INFORMATION

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

**How to cite this article:** Lynggaard C, Nielsen M, Santos-Bay L, Gastauer M, Oliveira G, Bohmann K. Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests. *Environmental DNA*. 2019;1:329–341. <https://doi.org/10.1002/edn3.34>