

## ORIGINAL ARTICLE


Environmental DNA

Open Access

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

WILEY

# Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity

Jan F. Gogarten<sup>1,2,3,4</sup>  | Constanze Hoffmann<sup>1</sup> | Mimi Arandjelovic<sup>5</sup> |  
 Andreas Sachse<sup>1</sup> | Kevin Merkel<sup>1</sup> | Paula Dieguez<sup>5</sup> | Anthony Agbor<sup>5</sup> |  
 Samuel Angedakin<sup>5,6</sup> | Gregory Brazzola<sup>5</sup> | Sorrel Jones<sup>5</sup> | Kevin E. Langergraber<sup>6,7</sup> |  
 Kevin Lee<sup>5</sup> | Sergio Marrocoli<sup>5</sup> | Mizuki Murai<sup>5</sup> | Volker Sommer<sup>8,9</sup> | Hjalmar Kühl<sup>5,10</sup> |  
 Fabian H. Leendertz<sup>1</sup> | Sébastien Calvignac-Spencer<sup>1,4</sup>

<sup>1</sup>Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Berlin, Germany

<sup>2</sup>Institute of Microbiology and Epizootics, Freie Universität Berlin, Berlin, Germany

<sup>3</sup>Center for Infection and Immunity, Columbia University Mailman School of Public Health

<sup>4</sup>Viral Evolution, Robert Koch Institute Berlin, Berlin, Germany

<sup>5</sup>Max Planck Institute for Evolutionary Anthropology (MPI EVAN), Leipzig, Germany

<sup>6</sup>School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA

<sup>7</sup>Institute of Human Origins, Arizona State University, Tempe, AZ, USA

<sup>8</sup>Department of Anthropology, University College London, London, UK

<sup>9</sup>Gashaka Primate Project, Taraba State, Serti, Nigeria

<sup>10</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

## Correspondence

Sébastien Calvignac-Spencer, Viral Evolution, Robert Koch Institute, Seestraße 10, Berlin 13353, Germany.  
 Email: CalvignacS@rki.de

## Funding information

Max Planck Society Innovation Fund; German Academic Exchange Service

## Abstract

**Background:** Metabarcoding of vertebrate DNA found in invertebrates (iDNA) represents a potentially powerful tool for monitoring biodiversity. Preliminary evidence suggests fly iDNA biodiversity assessments compare favorably with established approaches such as camera trapping or line transects.

**Aims and Methods:** To assess whether fly-derived iDNA is consistently useful for biodiversity monitoring across a diversity of ecosystems, we compared metabarcoding of the mitochondrial 16S gene of fly pool-derived iDNA (range = 49–105 flies/site,  $N = 784$  flies) with camera traps (range = 198–1,654 videos of mammals identified to the species level/site) at eight sites, representing different habitat types in five countries across tropical Africa.

**Results:** We detected a similar number of mammal species using fly-derived iDNA (range = 8–15 species/site) and camera traps (range = 8–27 species/site). However, the two approaches detected mostly different species (range = 6%–43% of species detected/site were detected with both methods), with fly-derived iDNA detecting on average smaller-bodied species than camera traps. Despite addressing different phylogenetic components of local mammalian communities, both methods resulted in similar beta-diversity estimates across sites and habitats.

**Conclusion:** These results support a growing body of evidence that fly-derived iDNA is a cost- and time-efficient tool that complements camera trapping in assessing mammalian biodiversity. Fly-derived iDNA may facilitate biomonitoring in terrestrial ecosystems at broad spatial and temporal scales, in much the same way as water eDNA has improved biomonitoring across aquatic ecosystems.

Jan F. Gogarten and Constanze Hoffmann contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

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

London, Grant/Award Number: REA grant agreement n° 605728 ; Heinz L. Krekeler Foundation; Max-Planck-Gesellschaft; German Federal Ministry of Education; European Union's Seventh Framework Programme, Grant/Award Number: FP7/2007-2013

## KEYWORDS

Africa, biodiversity, environmental monitoring, invertebrates, mammals

## 1 | INTRODUCTION

Understanding the distribution of life on this planet is an important first step toward deciphering the processes that create, maintain, and now threaten biological diversity (Steffen, Crutzen, & McNeill, 2007). Existing resources for monitoring global biodiversity (e.g., the Global Biodiversity Information Facility) have large gaps where no data are available; this is particularly true for large parts of tropical Africa. Tools that allow for rapid and cost-effective biodiversity assessment can aid in identifying the drivers of biodiversity declines and be used to monitor the effects of different conservation strategies and fill the gaps in these global biomonitoring efforts (Nicholson et al., 2012).

DNA barcoding approaches, which use short DNA sequences to assign unknown samples to a particular species based on comparison with a reference database of homologous sequences (Moritz & Cicero, 2004), have been explicitly developed to accelerate biodiversity assessments. The advent of high-throughput sequencing (HTS) technologies has enabled the extension of this concept of barcoding to the automated identification of many species from a single sample (i.e., metabarcoding; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012).

Metabarcoding now allows researchers to use a variety of substrates as sources of DNA to determine many of the species present in an ecosystem. For example, environmental DNA (eDNA) in water has proven useful to monitor fish and invertebrates in aquatic ecosystems (Valentini et al., 2016). Although soil (Andersen et al., 2012) and water from watering holes (Rodgers & Mock, 2015) and rivers (Deiner, Fronhofer, Mächler, Walsler, & Altermatt, 2016) have been shown to contain amplifiable DNA from terrestrial vertebrates, eDNA has been less frequently used to characterize terrestrial ecosystems.

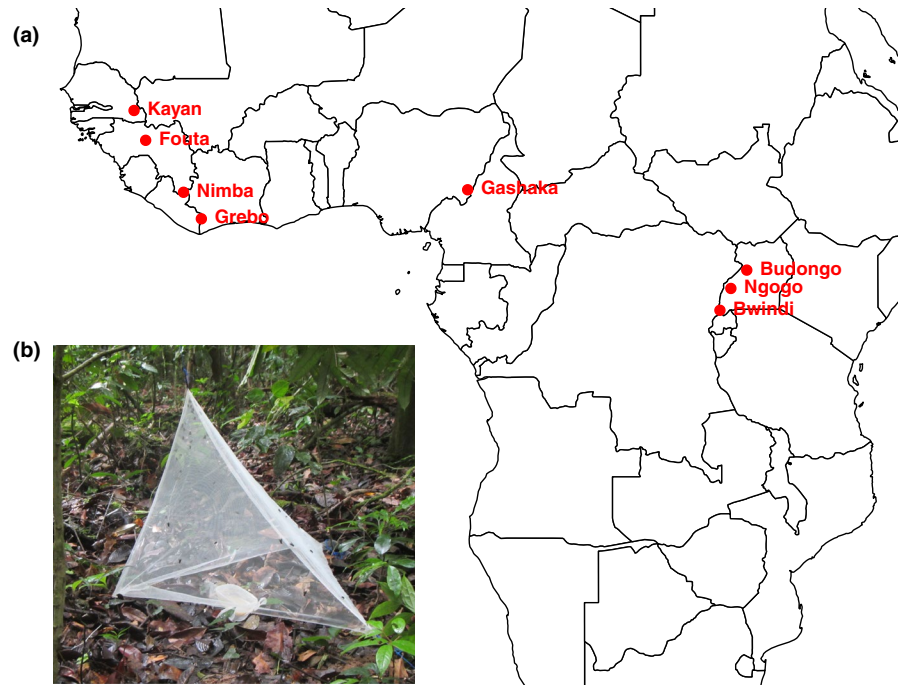
Several factors may explain the higher appeal and faster development of eDNA biomonitoring in water than in terrestrial ecosystems. Plant DNA has been shown to persist in soils for up to a few hundred years (Yoccoz et al., 2012), while vertebrate DNA was shown to remain detectable for several years in soil when conditions are favorable (Andersen et al., 2012); although this facilitates reconstruction of past biodiversity and allows researchers to escape temporal or seasonal variation affecting other survey methods, it prevents distinguishing modern and historic signals (Andersen et al., 2012; Taberlet, Bonin, Zinger, & Coissac, 2018). Once in soil, eDNA is generally bound to substrates such as clay particles or organic material that hinder its spatial diffusion; large volumes of soil should be analyzed to compensate for such heterogeneity, which entails

significant costs and logistic difficulties (Taberlet et al., 2018). In contrast, DNA tends to persist for shorter periods in water (Dejean et al., 2011; Thomsen et al., 2012). In addition, vertebrate DNA can diffuse through a broader body of water, which allows for cost-effective sampling of large volumes of water when coupled with filtration techniques (Valentini et al., 2016). These differences have likely contributed to the higher appeal and faster development of water eDNA approaches and highlight the need for the development of techniques amenable to terrestrial ecosystems.

Invertebrates that come into contact with vertebrates or their by-products as part of their life cycle represent a promising alternative source of vertebrate DNA for metabarcoding-based assessments of terrestrial biodiversity. This is particularly true for invertebrates that are easily trapped in large numbers in field settings by nonexperts. Invertebrate-derived DNA (iDNA) obtained from terrestrial leeches and carrion flies (families Calliphoridae and Sarcophagidae; hereafter referred to as flies) has already shown promise as a tool for rapid, cost-efficient detection of a broad range of terrestrial vertebrates (Bohmann, Schnell, & Gilbert, 2013; Calvignac-Spencer et al. 2013; Schnell et al., 2018; Schnell et al., 2012). Studies of medical entomology have also repeatedly demonstrated that other easily trapped invertebrates contain amplifiable vertebrate DNA (Kent, 2009), including but not limited to, mosquitos (e.g., Bataille et al., 2012), biting midges (Lassen, Nielsen, Skovgård, & Kristensen, 2011), and sandflies (Abbasi et al., 2009); indeed, iDNA from sandflies and mosquitoes was recently used to detect a diversity of Amazonian vertebrates in the context of assessing biodiversity (Kocher et al., 2017).

It is however unclear how well iDNA, and in particular fly iDNA, describes vertebrate biodiversity in an ecosystem and how it compares with other approaches such as camera traps (Schnell et al., 2015). The study that first proposed to use fly iDNA for biodiversity assessments showed that this method detected many of the known mammalian species present in two tropical ecosystems in Côte d'Ivoire and Madagascar, even though it relied on small numbers of flies (Calvignac-Spencer et al. 2013). More recently, a comparison of fly iDNA and camera traps at Selangor and Tembat Forest Reserve in Peninsular Malaysia found that fly iDNA detected a few more species than camera traps but that only a single species was detected by both approaches (Lee, Gan, Clements, & Wilson, 2016). Similarly, a fly iDNA metabarcoding study conducted at Barro Colorado Island, Panama, identified roughly the same number of species as camera traps and transects, but also only found a moderate species overlap between these methods (Rodgers et al., 2017). This preliminary evidence suggests that camera trapping and fly iDNA might

**FIGURE 1** (a) Map of field sites across sub-Saharan Africa showing locations sampled. The map was created using the R package *rworldmap* (South, 2011). (b) A fly trap deployed in the field



represent complementary methods for describing mammalian species assemblages.

Here, we implemented a metabarcoding approach to validate the use of flies as a tool for assessing mammalian biodiversity in a broader range of ecosystems, focusing on eight sites across sub-Saharan Africa. We directly compared the performance of fly iDNA metabarcoding with camera trapping at all sites to assess whether these methods generally tend to be complementary or usually detect the same species. To understand how these approaches might complement one another, we also examined whether flies and camera traps differed in their ability to detect small- and large-bodied mammal species, whether the species detected by these methods were phylogenetically distinct, and whether the mammalian communities they describe differed by habitat type.

## 2 | MATERIALS AND METHODS

### 2.1 | Study sites

Fly collection and camera trapping were undertaken within the framework of the Pan African Programme: The Cultured Chimpanzee (PanAf) ([panafrican.eva.mpg.de](http://panafrican.eva.mpg.de)). The program studies the diversity and diversifying mechanism in chimpanzees and involves 39 temporary research sites in 16 countries. Material for this study was collected at eight of these sites located in five countries and represented three different habitat types: two in forest-savannah ecosystems (Gashaka Gumti National Park, Nigeria; Sobeya, Guinea), one in a savannah ecosystem (Kayan, Senegal), three in East African rainforests (Budongo Forest, Uganda; Bwindi Impenetrable National Forest, Uganda; Ngogo East in Kibale National Park, Uganda), and two located in West African rainforests (Grebo National Forest, Liberia; East Nimba Nature Reserve, Liberia; Figure 1a, Table 1). At each

research site, we defined a data collection zone (range = 12–96 km<sup>2</sup>), based on direct and indirect observations of chimpanzee activity in the area. This collection zone was then divided into a grid with 1 × 1 km cells. Collection of flies and camera trapping were performed in parallel in 2013, with additional camera trap data collected in 2012 and 2014 at some sites (Table 1).

### 2.2 | Camera traps

Camera traps were installed in each data collection zone with the aim of covering the entire grid evenly. Installation locations within the grids were not chosen randomly, but rather with regard to wildlife activity (e.g., travel paths, natural bridges, and feeding sites) with an effort to maximize the number of animal species detected. For protection against wildlife damage and humidity, the cameras (Bushnell) were kept inside plastic boxes sealed with cling film that also contained silica. Maintenance (i.e., exchange of memory cards, recharging of batteries, and exchange of silica) was performed every month or every second month as required by the climate. Details of camera trapping effort are presented in Table 1. Species assignment from video clips was performed by experts at each field site, and we only included videos where a species-level detection was possible.

### 2.3 | Fly trapping

Flies were caught at one-kilometer intervals along the grid system. Fly trapping was performed as described by Calvignac-Spencer et al. (2013). In short, fly traps consisted of a pyramidal net and a plastic bowl containing commercially available bait based on animal proteins (Unkonventionelle Produkte Feldner, Waldsee, Germany; Figure 1b). Plastic bowls were covered with a net to prevent contact

**TABLE 1** Sampling of flies and camera trap data was performed in parallel at 8 different research sites

Country	Research Site	Habitat type	Years of sampling	N files	N traps	N 16S amplicons	N subsampled reads analyzed	N assignable reads <sup>a</sup> [c10/c25]	N traps with mammal DNA <sup>a</sup> [c10/c25]	N videos <sup>b</sup>
Guinea	Sobeya	Forest-savannah	2012–2014	105	15	28	672,000	78,310/75,874	10/8	1,170
Liberia	Grebo National Forest	Rainforest	2013–2014	105	15	28	672,000	270,188/250,656	13/13	1,538
	East Nimba Nature Reserve	Rainforest	2013–2014	49	7	14	336,000	95,639/89,985	7/7	585
Nigeria	Gashaka Gumti National Park	Forest-savannah	2012–2014	105	15	29	651,141	419,352/393,021	15/15	1,217
Senegal	Kayan	Savannah	2012–2014	105	15	26	598,528	128,078/121,729	11/11	1,654
Uganda	Budongo Forest Reserve	Rainforest	2012–2014	105	15	28	660,449	325,026/299,764	14/14	198
	Bwindi Impenetrable National Park	Rainforest	2012–2014	105	15	20	480,000	44,778/42,827	9/7	1,469
	Ngogo East (Kibale National Park)	Rainforest	2012–2014	105	15	23	552,000	126,403/118,926	11/11	475
	Total			784	112	196	4,622,118	1,487,774/1,392,782		8,306

Note: c10/25: minimum of 10/25 sequences representing one unique sequence for the reads to be considered.

<sup>a</sup>After removal of reads assigned to humans or to species that were not on the list of mammalian species present in the country according to the IUCN Red List of Threatened Species.

<sup>b</sup>After removal of multiple videos of the same species taken on the same day and removing all videos of animals that were not on the list of mammalian species present in the country according to the IUCN Red List of Threatened Species, as well as all videos of humans.

between flies and the bait. Traps were left open for 20 min with a maximum of 20 flies collected per trap. Flies were killed by placing flies into a container with a cotton swab soaked in 95% ethanol or ether. Where possible, flies were collected twice at 25 locations at each site, once in the dry season and again in the wet season. Flies were collected in 50-ml Falcon tubes (Carl Roth) and stored on silica at ambient temperature (in the field) and subsequently at 4°C (at the Robert Koch Institute). The baits used here were previously shown to attract flies belonging to a minimum of three families: Calliphoridae, Sarcophagidae, and Muscidae (Calvignac-Spencer et al. 2013; Hoffmann et al., 2018; Hoffmann et al., 2017; Schubert et al., 2015). We confirmed that this was also the case across these field sites by using soup CO1 metabarcoding (Folmer, Black, Hoeh, Lutz, & Vrijenhoek, 1994; Yu et al., 2012) of our fly pools; of the 12,768 sequences assignable to the family level using OBITools (Boyer et al., 2016; Fujisawa & Barraclough, 2013), 8,101 sequences were assigned to Calliphoridae, 4,653 to Sarcophagidae, and 14 to Muscidae.

Due to monetary and time constraints, we selected 15 fly traps per site for further molecular analyses, and traps were selected to maximize the space sampled across the study area (for East Nimba Nature Reserve, only seven suitable traps were available due to a shortage of material). Previous work by our group suggested that at least 1 in 7 flies tested contained mammal DNA (C. Hoffmann, S. Calvignac-Spencer, & F. H. Leendertz, unpublished data), so from each of these traps, seven flies were randomly selected and used for subsequent analysis (105 flies per site, with the exception of East Nimba Nature Reserve for which only 49 were collected). In total, 784 flies were selected for further molecular analyses.

## 2.4 | Molecular and bioinformatic methods

Extraction of DNA was performed on each individual fly using the GeneMATRIX Stool DNA Purification Kit (Roboklon) by a service provider (GenExpress), as described in detail by Calvignac-Spencer et al. (2013). Flies were extracted individually to enable their use in studies linking the detection of pathogens in a particular fly with a particular species of vertebrate, but we have previously shown the feasibility of extracting from pools (Hoffmann et al., 2018). DNA of flies collected from a trap was pooled in equal volumes, resulting in 15 DNA pools per site (i.e., each pool represented one trap) and was then used for metabarcoding. Metabarcoding was based on a 130-bp fragment of the mitochondrial 16S gene, which was amplified from each DNA pool using primers described by Taylor (1996). Amplicons from each pool were prepared and dual-indexed for deep sequencing in different runs on an Illumina MiSeq sequencer (Illumina) using the MiSeq Reagent Kit v2 (2 × 150 bp) or v3 (2 × 300 bp; Illumina). PCR conditions and details regarding the preparation of amplicons for deep sequencing are described in detail by Hoffmann, Stockhausen, Merkel, Calvignac-Spencer, and Leendertz (2016) and in the Supplementary Material. Briefly, the first PCR was performed using the primers described by Taylor (1996) to limit amplification biases and improve sensitivity, and then, Illumina-specific adapters

were added to these amplicons with a PCR performed with fusion primers. Subsequently, Nextera indexes and Illumina sequencing adapters were appended with a PCR. Per pool, two sets of 16S amplicons were generated, processed, and independently indexed, resulting in total of 196 amplicons that were deep sequenced. We sequenced one negative control, which comprised very few sequences ( $N = 244$ ), of which 15 sequences were assigned to *Colobus guereza*, likely reflecting low-level laboratory contamination.

The pipeline used for the analysis of our metabarcoding data is described in detail by Hoffmann et al. (2016). In short, paired-end raw reads were joined (*illumina-paired-end*) in the software package OBITools (v1.1.18), setting the minimum alignment score to 40, and primer and adapter sequences were subsequently removed (Cutadapt v1.2.1) and reads were quality-trimmed, setting the quality score to 30 over a sliding window of four bases (Trimmomatic v0.35; Bolger, Lohse, & Usadel, 2014; Boyer et al., 2016; Martin, 2011). The dataset was de-replicated using the OBITools software package, and only sequences represented by at least 10 or 25 reads (c10 and c25 thresholds) were used for further taxonomic assignment (*obiuniq*, *obigrep*). Reference databases were built by performing an in silico PCR on all vertebrate sequences available in GenBank using the program ecoPCR v0.2, allowing three mismatches between primers and reference sequences. Taxonomic assignment was implemented with the OBITools *ecotag* command, which uses the Needleman–Wunsch algorithm to map query sequences against an ecoPCR database and provides taxonomic assignments at species, genus, and family levels, with a minimum identity level of 0.95 (Bellemain et al., 2010; Ficetola et al., 2010; Needleman, 1970). The OBITools *ecotag* program first looks for the reference sequence that has the highest similarity to the query sequence and then identifies other reference sequences with similar identities. It then assigns the query sequence to the taxonomic rank that encompasses all of these similar reference sequences. All downstream analyses were based on sequences assigned at the species level. Analyses were also performed at the genus level, but trends did not differ from the species level; genus-level results are thus not presented in the manuscript.

## 2.5 | Quality control

Two different thresholds were used to reduce the likelihood of contaminant sequences leading to false species detection. We sought to explore the importance of these thresholds in determining the species present at a particular site, and we therefore present results from both threshold combinations. Specifically, we applied two thresholds to the metabarcoding data: a minimum of 10 sequences representing one unique sequence (c10) and a minimum of 25 sequences (c25). To further exclude likely contaminant sequences and potential misidentification of camera trap videos, we also filtered both the iDNA and camera trap datasets by excluding those mammalian species not present in the country according to the IUCN Red List of Threatened Species (lists of native mammalian species present in each country downloaded on 11 May 2018, Tables S7–S11). This approach de facto also excluded humans and domestic animal species, which do not

appear on the Red List. Applying the c10 or c25 thresholds (followed by filtering of species not on the Red List in at least one site studied here) to our negative control would identify it as positive for *C. guereza* or negative, respectively.

## 2.6 | Body size and biomass estimates

To examine whether species detected with flies or camera traps differed in either their body size or biomass, we used estimates of adult body mass and population density from the PanTHERIA database (Jones et al., 2009). To estimate biomass for a species, we multiplied the average adult body mass by the average population density. Adult body mass estimates were available for 87 of the 92 species detected here, while we were only able to estimate biomass for 51 of the 92 species detected based on available data. These biomass estimates were systematically missing for smaller-bodied species (Figure S1), precluding a further analysis of this variable here.

## 2.7 | Statistical analysis

To examine whether camera traps and flies tended to detect the same species, we tested for a relationship between the fly detection rate (proportion of fly traps detecting a species) and the camera detection rate (proportion of videos at a site detecting a species) using generalized linear mixed models (Baayen, 2008) with a binomial error structure and logit link function. In these models, we included the camera trap detection rate (i.e., the proportion of camera trap videos at a site detecting a given species) as a fixed effect and species and site as random effects, allowing for random slopes of the site random effect. As an overall test of the effect of the fixed effect, we compared the full model with a null model lacking the fixed effect, but comprising the same random effects structure as the full model (Forstmeier & Schielzeth, 2011) using a likelihood ratio test (Dobson, 2002). We checked models for overdispersion and found no evidence for overdispersion in these models. We assessed model stability by comparing the estimates obtained from a model based on all data with those obtained from models with the levels of the random effects excluded one at a time and found no major issues with model stability, indicating that no influential sites or species existed. Models were implemented using the function *glmer* of the R package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). The samples for these models consisted of 92 species detected across the eight field sites with 112 fly traps and 8,306 camera trap videos.

We compared the average adult body mass of species detected with camera traps and flies using generalized linear mixed models (Baayen, 2008), including site as a random effect. We log-transformed average adult body mass to improve the normality of the dependent variable. Species could not be included as a random effect in this model due to model convergence issues. To allow for a likelihood ratio test, we fitted the models using maximum likelihood (rather than restricted maximum likelihood; Bolker et al., 2008). We checked whether the assumptions of normally distributed and homogeneous residuals were fulfilled by visually inspecting a qqplot

and the residuals plotted against fitted values, both of which indicated no obvious deviations from these assumptions. We tested for significance as described above. Confidence intervals were derived using the function `bootMer` of the package *lme4*, using 1,000 parametric bootstraps and bootstrapping over the random effects. The samples for these models consisted of 87 species detected across the eight field sites with 112 fly traps and 8,306 camera trap videos for which body mass estimates were available. Statistical modeling was conducted in R version 3.4.0 (R Core Team, 2017), and data visualization was performed using the `ggplot2` package (Wickham, 2016).

To analyze how mammal species detected with camera traps or fly-derived iDNA were distributed on the phylogeny, we used a Bayesian inference method for examining the evolution of a phenotype (in this case, the detectability of a species only with camera traps, only with fly-derived iDNA, or with both methods) on a phylogenetic tree. For this, we used the *TreeBreaker* program (Ansari & Didelot, 2016), which is able to break down a tree into components for which the phenotype distributions differ, on a phylogeny of the species detected in this studies, which we downloaded from the *TimeTree* project (Kumar, Stecher, Suleski, & Hedges, 2017). We modified this phylogeny by adding (a) *Mungos gambianus*, whose phylogenetic placement and branch length were drawn from Nyakatura and Bininda-Emonds (2012), (b) *Aonyx congicus*, which we placed as the sister taxon of *Aonyx capensis*; no phylogenetic data were available to assess the branch length for their divergence, but as there is some debate as to whether these taxa should be subspecies or species, we arbitrarily estimated that they diverged 1 million years ago. Plots of these phylogenies and the posterior probability of a trait change on branches were generated using version 5.3 of the *ape* R package (Paradis & Schliep, 2018). Following Kass and Raftery (1995), we considered values of  $2 * \ln$  (Bayes factor) that were greater than 10 to indicate decisive support for the alternative hypothesis, in this case that there were differences in phenotype distributions on the phylogeny.

To compare the species communities detected with the fly-derived iDNA and camera traps in different habitats, we used nonmetric multidimensional scaling (nMDS) on the presence or absence of species, comparing both the Bray–Curtis dissimilarity metric and the phylogenetically informed UniFrac distance metric. We used *adonis*, a permutational multivariate analysis of variance (Anderson, 2001), to test for differences in mammal community composition based on site, and detection method. Small sample sizes precluded a formal test of significance of differences in beta diversity between communities by habitat type using *adonis*. In addition, we compared the similarity of the species communities detected at these eight sites with these two detection methods using hierarchical clustering with the unweighted pair group method with arithmetic mean method (UPGMA) on both the Bray–Curtis and UniFrac distance metrics of the presence or absence of species. These analyses were performed using version 2.5-5 of the *vegan* R package (Oksanen et al. 2018).

### 3 | RESULTS

In total, camera trapping across all sites resulted in 9,224 nonhuman mammal videos that were assignable to the species level, after having removed multiple videos of the same species taken on the same day. Of the according detection events, 918 were excluded as the identified species was not thought to be a native mammal (mostly domestic animals) according to the IUCN Red List of Threatened Species (Table 1). After filtering, we detected between 8 and 27 mammalian species at each site ( $\bar{x}_{\text{number of species}} = 20.1$ ; Tables S1, S2).

From 112 fly pools analyzed, 101 (90%) contained amplifiable mammalian DNA. For 95 fly pools, it was possible to generate an amplicon from both replicates; for 6 fly pools, this was possible only from one of the replicates. Raw reads from fly amplicons ( $N = 196$ ) were generated using different MiSeq runs and resulted in uneven numbers of raw reads per amplicon ranging from 1,378 to 743,372. This broad range was due to problems quantifying DNA concentrations of libraries for pooling on our initial MiSeq run (August 2014); this problem was fixed on the subsequent run (April 2015). To account for the uneven number of raw reads per amplicon, reads were resampled to a maximum of 24,000 raw reads per amplicon, resulting in a total of 4,622,118 raw reads ( $\bar{x}_{\text{number of raw reads per amplicon}} = 23,582$ ). For the c10 threshold, the pipeline generated 3,160,393 usable reads, and for the c25 threshold, 2,930,018 reads. For the c10 threshold, 75.4% of these reads were assigned at species level, 93.3% at the genus level, 94.2% at the family level, and 5.7% of the reads were not assignable. For the c25 threshold, 76.2% of the reads could be assigned at the species level, 93.5% at the genus level, 94.3% at the family level, and 5.6% of the reads were not assignable. For the c10 threshold, 27.8% of reads that were assignable to the species level were assigned to humans and were excluded from this dataset (for the c25 threshold, 27.6% of reads). In addition, 7.1% and 7.5% of the reads assignable to the species level, for the c10 and c25 thresholds, respectively, were excluded from the dataset because they were assigned to species not listed as native mammalian species on the IUCN Red List of Threatened Species for that country (including humans and domestic animals). Following filtration using the c10 threshold, we detected between 8 and 15 mammalian species per site ( $\bar{x}_{\text{number of species}} = 10.3$ ; Table S1), while using the c25 threshold we detected between 5 and 15 mammalian species per site ( $\bar{x}_{\text{number of species}} = 8.6$ ; Table S2). Generally, camera traps therefore detected more species than 16S metabarcoding of fly-derived iDNA (Figure 2a; Table S1, S2).

The camera trap and iDNA approaches detected mostly different species at any given site (Figure 2a; Tables S1, S2). This pattern was apparent regardless of the thresholds used to filter the data. For those species detected by both methods, we found no evidence for a relationship between the proportion of camera trap days a species was observed and the proportion of fly traps from which this species was detected, when using either the c10 (Figure 3a; likelihood ratio test:  $\chi^2 = 3.12$ ,  $df = 1$ ,  $p = .0774$ ) or c25 threshold (Figure S2B; likelihood ratio test:  $\chi^2 = 2.43$ ,  $df = 1$ ,  $p = .119$ ).

The three smallest animals detected with camera traps had an estimated adult body mass of 66 g (*Galagoides demidovii*), 215 g (*Galago senegalensis*), and 243 g (*Funisciurus pyrropus*), while the largest three were 271 kg (*Tragelaphus eurycerus*), 593 kg (*Syncerus caffer*), and 3,825 kg (*Loxodonta africana*). In contrast, the three smallest animals detected with flies had an estimated adult body mass of 27 g (*Mops condylurus*), 39 g (*Praomys jacksoni*), and 66 g (*Galagoides demidovii*), while the largest three had an estimated adult body mass of 198 kg (*Hylochoerus meinertzhageni*), 264 kg (*Hippotragus equinus*), and 3,825 kg (*Loxodonta africana*), respectively. The average adult body mass of species detected with camera traps was significantly higher than that of mammalian species detected using flies using the c10 (Figure 3b; Table S3,  $\bar{x}_{\text{camera trap}} = 93.2$  kg,  $\bar{x}_{\text{flies}} = 69.5$  kg; likelihood ratio test:  $\chi^2 = 23.63$ ,  $df = 1$ ,  $p < .001$ ) and c25 thresholds (Figure S3; Table S4,  $\bar{x}_{\text{camera trap}} = 93.2$  kg,  $\bar{x}_{\text{flies}} = 80.4$  kg; likelihood ratio test:  $\chi^2 = 18.84$ ,  $df = 1$ ,  $p < .001$ ).

We detected statistical support for a change in phenotype distribution (i.e., detection with camera traps, detection with fly-derived iDNA, or detection with both methods) across the phylogeny of mammals detected in this study. The Bayes factor of the TreeBreaker analysis of the model with one or more change points compared with a model including no change points was 249;  $2 * \ln(\text{Bayes factor}) = 11.0$ , which indicates decisive support for the alternative hypothesis that there were differences in phenotype distributions on the phylogeny. The mean number of change points estimated by the TreeBreaker analysis was 4.6, while the 95% credibility interval for the number of change points was 1–11. We detected the strongest statistical support for a change in phenotype distribution in the descending lineages of two branches on the phylogeny of mammals detected in this study (Figure 2b; Figures S5 and S6). Bats were only detected with fly-derived iDNA and camera traps seemed better able to detect artiodactyls and carnivores, though fly-derived iDNA did detect some of these species as well.

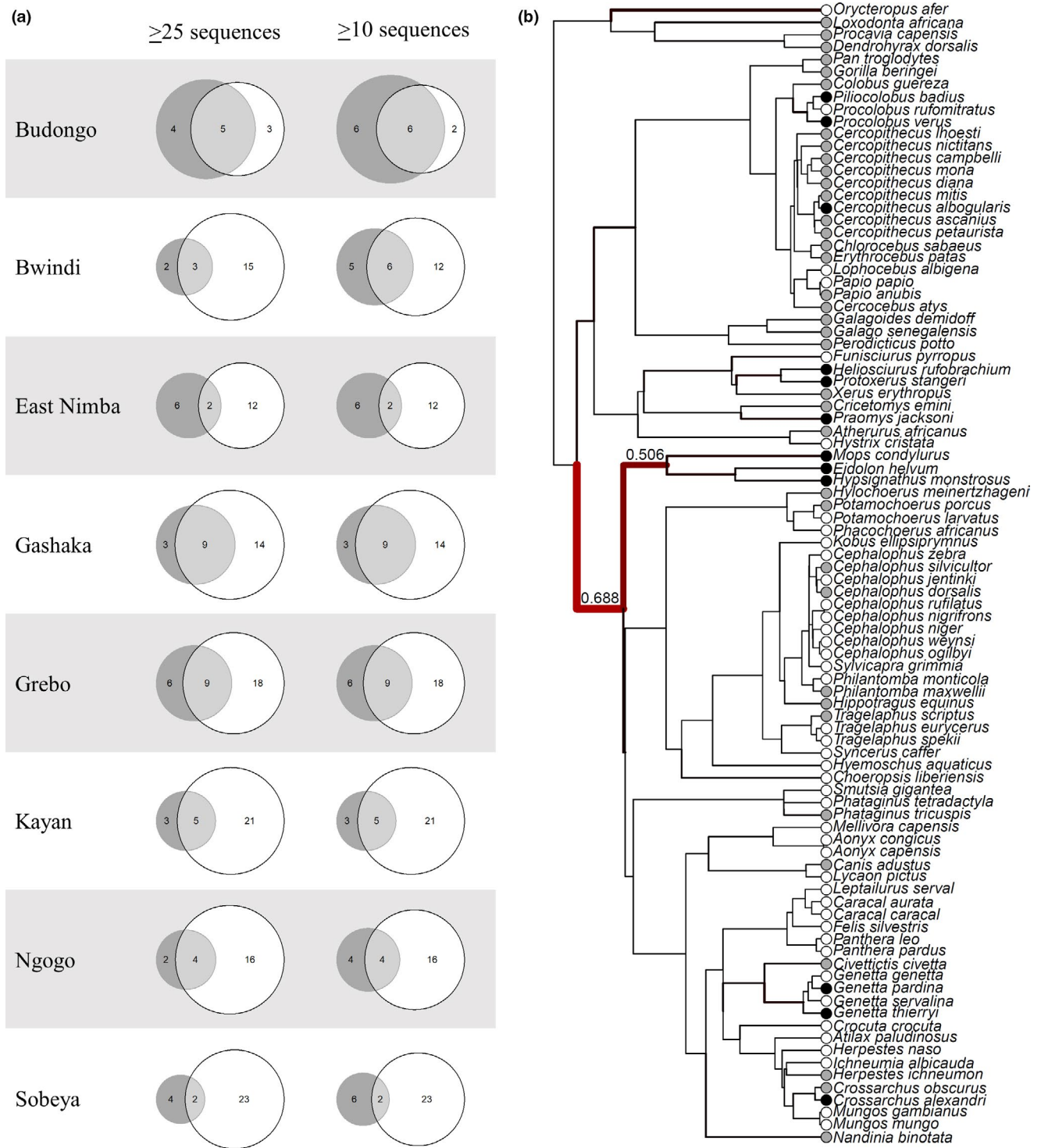
The UPGMA analysis based on Bray–Curtis distances revealed that fly (c10)- and camera trap-derived community compositions resulted in identical overall clustering of sites, regardless of the detection method (Figure 4a). In line with this, a permutational multivariate analysis of variance of Bray–Curtis distances failed to show any statistically significant difference between the communities described by the two methods (adonis  $F = 1.669$ ,  $R^2 = 0.116$ ,  $p = .093$ ), as illustrated by largely overlapping 95% confidence ellipses in the nMDS plot (Figure 4b). The UPGMA analysis based on Bray–Curtis distances only showed moderate support for clustering according to habitat (Figure 4a). Species composition appeared to differ by habitat type in the nMDS plot, though the low number of sites representing each habitat type precluded a formal test using adonis (Figure 4c). When examining camera- and fly-derived species compositions, there was a significant difference in community composition by site, suggesting that iDNA and camera traps detected similar communities (adonis  $F = 2.114$ ,  $R^2 = 0.649$ ,  $p = .003$ ), highlighted by the proximity of sites in the nMDS plots (Figure 4b, c). A similar pattern was observed using the c25 threshold (Figure S4A–C).

In contrast, the UPGMA based on the UniFrac distance metric revealed that fly (c10)- and camera trap-derived community compositions differed in the overall clustering of sites, with community similarity rather appearing to be structured by detection method (Figure 4d). In line with this finding, a permutational multivariate analysis of variance of the UniFrac distance metric found a statistically significant difference between communities detected by these two methods (adonis  $F = 2.449$ ,  $R^2 = 0.148$ ,  $p = .006$ ), as illustrated by largely nonoverlapping 95% confidence ellipses in the nMDS plot (Figure 4e). Unlike what was observed using Bray–Curtis distances, communities' differences described by the UniFrac distance metric did not appear to be structured by habitat type, though again the low number of sites representing each habitat type precluded a formal test using adonis (Figure 4f). When examining camera- and fly-derived species compositions based on the UniFrac distance, there was no significant difference by site, collectively suggesting that iDNA and camera traps captured different parts of the mammalian phylogeny (adonis  $F = 1.261$ ,  $R^2 = 0.525$ ,  $p = .102$ ). A similar pattern for UniFrac distance analyses was observed using the c25 threshold dataset (Figure 4SD–F).

## 4 | DISCUSSION

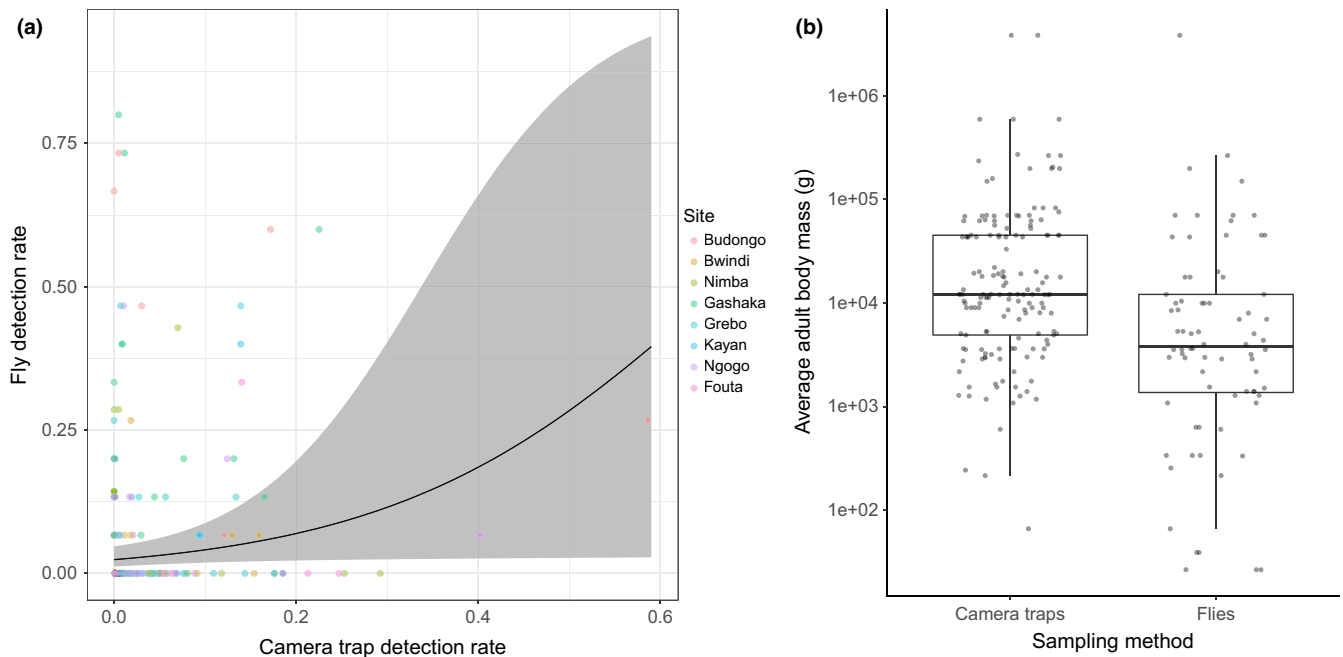
Across eight locations in sub-Saharan Africa, we detected a large number of mammalian species using metabarcoding of iDNA derived from a small number of flies. Fly-derived mammalian DNA detected species in habitat types where the feasibility of this approach has not previously been demonstrated (i.e., in savannah and highland tropical rainforests). Thus, the broad geographic range of our study adds to a growing body of evidence that wherever flies are present in the tropics, they represent a useful tool for assessing mammalian biodiversity (Calvignac-Spencer et al. 2013; Lee et al., 2016; Rodgers et al., 2017; Schubert et al., 2015). A recent study also showed the feasibility of using fly-derived mammalian DNA for biomonitoring in a temperate urban area (Hoffmann et al., 2018), suggesting this approach may prove useful across most terrestrial ecosystems where flies are present.

In agreement with two studies comparing camera traps with fly-derived mammalian DNA (Lee et al., 2016; Rodgers et al., 2017), our results confirm on a larger scale that these approaches are complementary, each detecting various unique taxa. Fly-derived mammalian DNA detected smaller-bodied species than camera traps, though fly-derived mammalian DNA also detected many large-bodied species. As described previously, camera traps seemed better able to detect larger species (Tobler et al. 2008), though they can be modified to increase the detection of smaller-bodied species (Melidonis & Peter, 2015) and we detected some small-bodied species in this study as well. A species-level delineation of smaller-bodied species from a video might be more difficult than for larger-bodied species. Indeed, arboreal and volant mammals are generally precluded from having as large body masses as terrestrial species (Bakker & Kelt, 2000; Smith & Lyons, 2011) and



**FIGURE 2** (a) Euler diagrams showing the number of mammalian species detected at each field site (rows) using iDNA (gray) and camera traps (white) with different thresholds for determining the presence at the field site (fly-derived DNA using 25 and 10 sequences assigned to a species as the minimum need to declare it present). The area of the circles and their overlap is proportional to the number of species detected using a method. This figure was created using the *Venvenable* R package (Swinton, 2016). (b) Phylogeny of mammals detected in the current study, with those species detected only using fly-derived iDNA (c10) highlighted with a black circle, those detected only with camera traps with a white circle, and those species detected with both methods highlighted with a gray circle. The only difference when using the c25 threshold was that *Praomys jacksoni* was no longer detected with flies, so this species was not present in the c25 phylogeny. The results of the Bayesian inference analysis of the evolution of the trait (i.e., detected with both methods, just camera traps, or just fly-derived iDNA) on the phylogeny are indicated through the thickness and redness of the branches, which are both proportional to the posterior probability of trait change on the given branch. To facilitate interpretation, for those branches with a posterior probability of a trait change on the branch >0.5, the posterior probability is shown





**FIGURE 3** (a) Scatter plot showing the mammal detection rate by camera traps (proportion of camera videos) and flies (proportion of fly traps using the  $c_{10}$  thresholds) for each species detected at a particular site (indicated by the color of the point) with at least one method. The solid black line represents the fit of the GLMM, while the gray shaded area represents the 95% confidence limits of the model. (b) Box plot showing the median average adult body mass of species detected by each sampling method ( $c_{10}$ ). Results are shown on a log scale and the lower and upper hinges correspond to the 25th and 75th percentiles, respectively, while the middle horizontal line represents the median. The upper and lower whiskers extend to the largest and lowest values no more than 1.5 times the interquartile range away from the hinge. Raw data are plotted in gray points

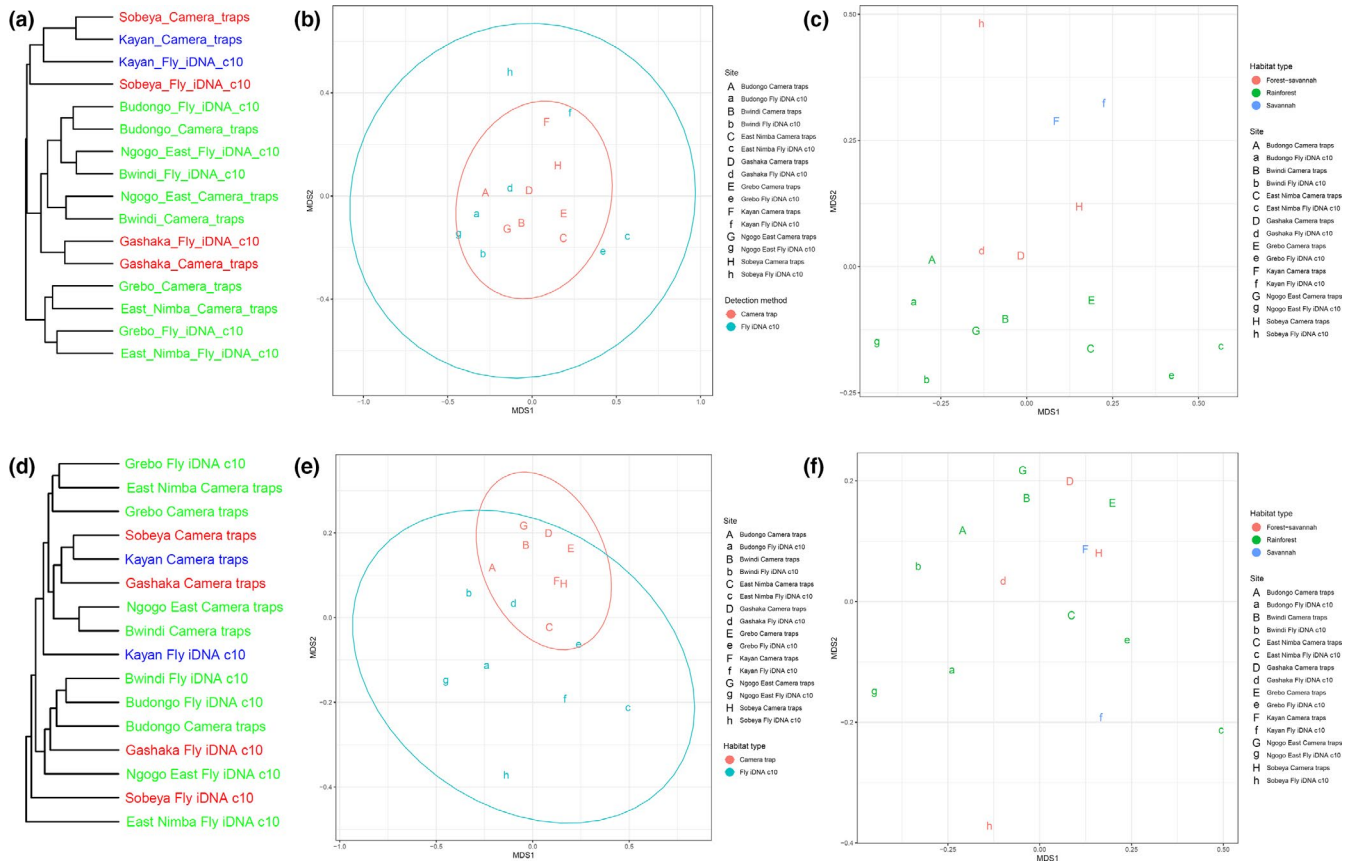
these mammals were detected less frequently with camera traps (e.g., in this study bats were only detected with fly-derived iDNA), which might be one factor involved in the observed differences in body mass of species detected with these two approaches. Our finding of a differential ability of the two methods to detect species across the mammal phylogeny fits well with these observations, since the clades where a change in detectability was observed are also characterized by distinct average body size and degree of terrestriality, that is, artiodactyls and carnivores are largely terrestrial, while many primates are largely arboreal. A similar ability to detect small-bodied species has been suggested for leech-derived iDNA (Weiskopf et al., 2018), though this pattern may not hold in all ecosystems (Abrams et al., 2019), perhaps due in part to differences in leech ecology and host preferences (Abrams et al., 2019; Drinkwater et al., 2018). Further research is also needed to disentangle whether and which other aspects of a mammalian species' biology beyond body mass may bias fly and other invertebrate iDNA analyses toward the detection of particular species.

While camera trapping and fly-derived mammalian DNA metabarcoding likely address different components of mammalian alpha diversity, our analyses suggest that they produce similar estimates of beta diversity. We attributed the differences observed between the Bray–Curtis and UniFrac-based beta-diversity estimates to the phylogenetic structure of detectability with both methods. Preliminary analyses based on Bray–Curtis distances suggested that beta diversity was structured by habitat types, though this result

should be interpreted with caution because of the small number of sites per habitat type available here. Confirming the ability of flies to capture aspects of mammalian alpha and beta diversity in different habitat types with larger sample sizes represents an exciting area of future research.

Given the complementary nature of iDNA and camera trapping, biodiversity monitoring efforts will however likely be most efficient when these approaches are combined (Abrams et al., 2019). Camera traps are on occasion damaged by animals and require regular upkeep to keep batteries charged and maintain an unobstructed view for the camera. The regular visits required for such maintenance represent an ideal opportunity to also trap flies and other invertebrates to maximize the number of species detected. A large number of flies can usually be collected in a short time window, particularly when using multiple traps at a location, and training field assistants to use fly traps is straight forward. Fly traps and baits are inexpensive (Table S5) and can be reused. Following collection, flies can be stored at room temperature on silica gel, meaning no electricity is needed to cool samples in the field. While both approaches currently have limitations in the field, this study shows that fly collection and camera trapping can easily be performed in parallel in field settings across a broad diversity of habitat types.

In most cases, the joint implementation of these approaches will consist of adding an iDNA layer to ongoing camera trap studies. Exporting flies from these countries to Germany was a straightforward process requiring minimal permits and bureaucracy, as



**FIGURE 4** UPGMA hierarchical clustering (a) and nMDS of Bray–Curtis dissimilarity metrics of the mammal species communities detected at each of the field sites with camera traps and fly-derived iDNA (c10), colored by the detection method (b) and habitat type (c). UPGMA hierarchical clustering (d) and nMDS of the UniFrac distance metrics of the mammal species communities detected at each of the field sites, with camera traps and fly-derived iDNA (c10), colored by the detection method (e) and habitat type (f). Solid colored lines indicate the 95% confidence ellipses for the grouping variable

the flies are not listed on the Convention on International Trade in Endangered Species of Wild Fauna and Flora. As discussed above, the main hurdle to such approaches will likely not be in the field; rather, we expect that the significant additional costs incurred by molecular analyses or scaling up camera trap efforts (considering both manpower and materials) will represent more significant obstacles. Following the collection of camera trap data, hundreds of hours of videos need to be looked at by experts to determine the species present. Citizen science projects can be used to analyze camera trap videos by nonexperts (e.g., chimpanzee.org), but require many viewers per video and therefore take a lot of time. Deep learning approaches show promise for automating this process, which would improve the scalability of camera trap surveys (Norouzzadeh et al., 2018). DNA analyses are already sufficiently cheap that their inclusion in many conservation biology projects would not require a major budget increase. In our study, the laboratory work to process the 784 flies was conducted in ~34.7 working days (extraction of 45 flies/day = 17.4 days for extraction; PCR for 112 DNA pools with 2 amplicons/pool at a rate of 50/day = 4.5 days; amplicon preparation for high-throughput sequencing = 12.8 days, data analysis = 1 day). Pooled extraction of flies would have reduced the number of working days needed to 16 (Table S5), and

automation of some steps might have further sped up the process. Material costs were also not prohibitive (Table S5: price per fly pool €31.93; price per site: €478.9), though they could also be reduced by further pooling and automation. In contrast, the costs for the equipment to conduct camera trapping at each site were significantly higher (€7,734.05 per site; Table S6). The time to look at and identify species from videos from camera traps was also intensive; those working at the field sites generally spent 1 to 3 months on this task at each site, though at some sites even this amount of effort was not sufficient and required the use of a citizen science project to pre-analyze the data (chimpanzee.org). Since further decreases in material costs can also be expected (e.g., sequencing reagent costs), collecting iDNA samples in parallel to other efforts may be a prudent way to maximize the long-term efficiency of biomonitoring efforts, regardless of the immediate availability of funds for molecular analyses.

In this study, we only analyzed a small number of flies at each location; as the cost of molecular analyses declines, it will be interesting to examine whether larger sampling efforts increase the overlap between camera traps and iDNA species detection. Larger sample sizes combined with appropriate sampling strategies (in particular repeated sampling) could also enable the use

of occupancy models and therefore the production of statistically robust biodiversity assessments for terrestrial mammals (Abrams et al., 2019; Schnell et al., 2015). While the collection of large numbers of some invertebrates can be quite time-consuming and difficult in some seasons (e.g., leeches; Abrams et al., 2019), the mobility and abundance of flies means that large numbers can be rapidly collected across seasons and habitat types. The high mobility of flies does however bring with it more uncertainty regarding the location where the fly's contact with mammal DNA actually occurred; such uncertainty could potentially be accounted for in occupancy models, so long as information about fly dispersal and mammalian DNA persistence in or on their bodies is obtained. Gaining more insight into fly ecology and iDNA persistence therefore represents a critical next step in developing the use of occupancy models with these types of data. Different invertebrates may ultimately prove useful for assessing mammalian biodiversity at different spatial scales, and occupancy models may also benefit from combining iDNA from a diversity of invertebrates. Further, there are a myriad of alternative approaches for assessing biodiversity (e.g., transects monitoring for signs or the animals themselves, soil or water eDNA, trapping of animals, and hunter self-monitoring) that it will be interesting to compare with camera traps and fly-derived iDNA to determine the most reliable and cost-efficient means of generating biodiversity assessments (Marrocoli et al., 2019; Newman, Buesching, & Macdonald, 2003; Taberlet et al., 2018).

Fly-derived iDNA metabarcoding faces the same limitations as other metabarcoding approaches. In particular, metabarcoding is only as good as the reference databases used to assign taxonomy to sequences (Bohmann et al., 2014; Bush et al., 2017; Pedersen et al., 2015; Schnell et al., 2015; Taberlet et al., 2018). The difficulty in assigning a significant fraction of the sequences generated in this study at the species-level probably resulted in part from such database gaps—many mammalian species have simply not yet been bar-coded. The use of zoo and museum specimens provides an exciting way to generate reference sequences at a minimal cost (Salleh et al. 2017) and might contribute to rapidly improving our reference databases, which will in turn improve estimates derived from metabarcoding experiments. Thus, fly-derived iDNA datasets will represent permanent, verifiable sources of information from which regular reanalysis will provide more and more accurate retrospective biodiversity assessments.

Flies not only contain the DNA of mammals found in these ecosystems; they also have been shown to carry the DNA of other vertebrates (Calvignac-Spencer et al. 2013; Rodgers et al., 2017). Fly-derived iDNA collections can potentially be screened using primers developed for the detection of other groups of organisms of interest (Taberlet et al., 2018). Flies also contain the DNA of many microorganisms, including wildlife and human pathogens, suggesting fly-derived iDNA has the potential to provide insights into the microbial diversity of ecosystems (Bitome-Esson et al., 2017; Hoffmann et al., 2016; Hoffmann et al., 2017; Knauf et al., 2016). Synanthropic flies associated with humans and their

livestock, including fly families beyond the Calliphoridae and Sarcophagidae, are thought to be vectors for a broad range of pathogens, including protozoan parasites (Graczyk et al., 1999), bacteria (Emerson, Bailey, Mahdi, Walraven, & Lindsay, 2000), and helminths (Monzon et al., 1991). These fly associations also occur in nonhuman primate groups, and flies in these associations carry bacterial pathogens causing major mortality in these primates (Gogarten et al., 2019; Hoffmann et al., 2017). This suggests fly-derived iDNA collections contain information that could be useful for monitoring pathogens in ecosystems and potentially for detecting outbreaks as well. Collectively, our results suggest that fly-derived iDNA can facilitate surveys of biodiversity in terrestrial ecosystems at broad spatial, temporal, and phylogenetic scales, in much the same way as water eDNA has improved biomonitoring across aquatic ecosystems (Taberlet et al., 2018).

## ACKNOWLEDGMENTS

JFG was supported by a grant from the German Academic Exchange Service, with funds from the German Federal Ministry of Education and Research (BMBF) and the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 605728 (P.R.I.M.E.—Postdoctoral Researchers International Mobility Experience). For funding support, we thank the Max Planck Society, the Max Planck Society Innovation Fund, and the Heinz L. Krekeler Foundation. We also thank the following collaborating ministries and organizations: Ministère de l'Agriculture de l'Élevage et des Eaux et Forêts, Guinée; Forestry Development Authority, Liberia; National Park Service, Nigeria; Direction des Eaux, Forêts et Chasses, Senegal; Makerere University Biological Field Station, Uganda; National Council for Science and Technology, Uganda; and Wildlife Authority and the National Forestry Authority, Uganda. We also thank the following nongovernmental organizations for their support: Wild Chimpanzee Foundation, Guinée; Wild Chimpanzee Foundation, Liberia; Budongo Conservation Field Station, Uganda; and Ngogo Chimpanzee Project, Uganda. We are extremely grateful for Christophe Boesch's vision and dedication in establishing the Pan African Programme: The Cultured Chimpanzee. We also thank Katherine Corogenes, Dervla Dowd, Henk Eshuis, Vincent Lapeyre, Yasmin Moebius, Geoffrey Muhanguzi, Erin G. Wessling, and Klaus Zuberbuehler for their help with data collection and field logistics.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

All sequence data have been uploaded to the European Reads Archive with the Sample ERS 3187718–3187914.

## ORCID

Jan F. Gogarten  <https://orcid.org/0000-0003-1889-4113>

## REFERENCES

- Abbasi, I., Cunio, R., & Warburg, A. (2009). Identification of blood meals imbibed by phlebotomine sand flies using cytochrome b PCR and reverse line blotting. *Vector-Borne and Zoonotic Diseases*, 9(1), 79–86.
- Abrams, J. F., Hörig, 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. *J Appl Ecol*, 56, 1637–1648. <https://doi.org/10.1111/1365-2664.13411>
- Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-madsen, H., Kjaer, K. H., ... Willerslev, E. (2012). Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, 21(8), 1966–1979. <https://doi.org/10.1111/j.1365-294X.2011.05261.x>
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1), 32–46.
- Ansari, M. A., & Didelot, X. (2016). Bayesian inference of the evolution of a phenotype distribution on a phylogenetic tree. *Genetics*, 204(1), 89–98. <https://doi.org/10.1534/genetics.116.190496>
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.
- Bakker, V. J., & Kelt, D. A. (2000). Scale-dependent patterns in body size distributions of neotropical mammals. *Ecology*, 81(12), 3530–3547. [https://doi.org/10.1890/0012-9658\(2000\)081\[3530:sdpibs\]2.0.co;2](https://doi.org/10.1890/0012-9658(2000)081[3530:sdpibs]2.0.co;2)
- Bataille, A., Fournié, G., Cruz, M., Cedeño, V., Parker, P. G., Cunningham, A. A., & Goodman, S. J. (2012). Host selection and parasite infection in *Aedes taeniorhynchus*, endemic disease vector in the Galápagos Islands. *Infection, Genetics and Evolution*, 12(8), 1831–1841. <https://doi.org/10.1016/j.meegid.2012.07.019>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., & Kausserud, H. (2010). ITS as an environmental DNA barcode for fungi: An in silico approach reveals potential PCR biases. *BMC Microbiology*, 10(1), 189. <https://doi.org/10.1186/1471-2180-10-189>
- Bitome-Essono, P.-Y., Ollomo, B., Arnathau, C., Durand, P., Mokoudoum, N. D., Yacka-Mouele, L., ... Paupy, C. (2017). Tracking zoonotic pathogens using blood-sucking flies as 'flying syringes'. *eLife*, 6, e22069. <https://doi.org/10.7554/eLife.22069>
- Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., ... 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., 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>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S.-S. (2008). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135. <https://doi.org/10.1016/j.tree.2008.10.008>
- 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.
- Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., ... Yu, D. W. (2017). Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology & Evolution*, 1(7), 0176. <https://doi.org/10.1038/s41559-017-0176>
- Calvignac-Spencer, S., Leendertz, F. H., Gilbert, M. T. P., & Schubert, G. (2013). An invertebrate stomach's view on vertebrate ecology. *Bioessays*, 35(11), 1004–1013. <https://doi.org/10.1002/bies.201300060>
- Calvignac-Spencer, S., Merkel, K., Kutzner, N., Köhl, H., Boesch, C., Kappeler, P. M., ... 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.
- Deiner, K., Fronhofer, E. A., Mächler, E., Walsler, J.-C., & Altermatt, F. (2016). Environmental DNA reveals that rivers are conveyor belts of biodiversity information. *Nature Communications*, 7, 12544. <https://doi.org/10.1038/ncomms12544>
- Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P., & Miaud, C. (2011). Persistence of environmental DNA in freshwater ecosystems. *PloS One*, 6(8), e23398. <https://doi.org/10.1371/journal.pone.0023398>
- Dobson, A. (2002). *An introduction to generalized linear models*. Boca Raton, FL: Chapman & Hall/CRC.
- Drinkwater, R., Schnell, I. B., Bohmann, K., Bernard, H., Veron, G., Clare, E., ... S. J. (2018). 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.
- Emerson, P. M., Bailey, R. L., Mahdi, O. S., Walraven, G. E., & Lindsay, S. W. (2000). Transmission ecology of the fly *Musca sorbens*, a putative vector of trachoma. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94(1), 28–32. [https://doi.org/10.1016/S0035-9203\(00\)90427-9](https://doi.org/10.1016/S0035-9203(00)90427-9)
- Ficetola, G., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessièrè, J., ... 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>
- Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5), 294–299.
- Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, 65(1), 47–55. <https://doi.org/10.1007/s00265-010-1038-5>
- Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: A revised method and evaluation on simulated data sets. *Systematic Biology*, 62(5), 707–724. <https://doi.org/10.1093/sysbio/syt033>
- Gogarten, J. F., Dux, A., Mubemba, B., Pléh, K., Hoffmann, C., Mielke, A., ... Leendertz, F. H. (2019). Tropical rainforest flies carrying pathogens form stable associations with social non-human primates. *Molecular Ecology*. <https://doi.org/10.1111/mec.15145>
- Graczyk, T. K., Fayer, R., Cranfield, M. R., Mhangami-Ruwende, B., Knight, R., Trout, J. M., & Bixler, H. (1999). Filth flies are transport hosts of *Cryptosporidium parvum*. *Emerging Infectious Diseases*, 5(5), 726.
- 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>
- Hoffmann, C., Stockhausen, M., Merkel, K., Calvignac-Spencer, S., & Leendertz, F. H. (2016). Assessing the feasibility of fly based surveillance of wildlife infectious diseases. *Scientific Reports*, 6, 37952. <https://doi.org/10.1038/srep37952>
- Hoffmann, C., Zimmermann, F., Biek, R., Kuehl, H., Nowak, K., Mundry, R., & ... Leendertz, F. H., (2017). Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest. *Nature*, 548(7665), 82–86.
- Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., ... Purvis, A. (2009). PanTHERIA: A species-level database

- of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 90(9), 2648–2648. <https://doi.org/10.1890/08-1494.1>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kent, R. J. (2009). Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. *Molecular Ecology Resources*, 9(1), 4–18. <https://doi.org/10.1111/j.1755-0998.2008.02469.x>
- Knauf, S., Raphael, J., Mitjà, O., Lejora, I. A. V., Chuma, I. S., Batamuzi, E. K., ... Lukehart, S. A. (2016). Isolation of *Treponema* DNA from necrophagous flies in a natural ecosystem. *EBioMedicine*, 11, 85–90. <https://doi.org/10.1016/j.ebiom.2016.07.033>
- 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>
- Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, 34(7), 1812–1819. <https://doi.org/10.1093/molbev/msx116>
- Lassen, S. B., Nielsen, S. A., Skovgård, H., & Kristensen, M. (2011). Molecular identification of bloodmeals from biting midges (Diptera: Ceratopogonidae: Culicoides Latreille) in Denmark. *Parasitology Research*, 108(4), 823–829. <https://doi.org/10.1007/s00436-010-2123-4>
- 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>
- Marrocoli, S., Nielsen, M. R., Morgan, D., van Loon, T., Kulik, L., & Kühl, H. (2019). Using wildlife indicators to facilitate wildlife monitoring in hunter-self monitoring schemes. *Ecological Indicators*, 105, 254–263. <https://doi.org/10.1016/j.ecolind.2019.05.050>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet.journal*, 17(1), 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Melidonis, C. A., & Peter, C. I. (2015). Diurnal pollination, primarily by a single species of rodent, documented in *Protea foliosa* using modified camera traps. *South African Journal of Botany*, 97, 9–15. <https://doi.org/10.1016/j.sajb.2014.12.009>
- Monzon, R., Sanchez, A., Tadiaman, B., Najos, O., Valencia, E., De Rueda, R., & Ventura, J. (1991). A comparison of the role of *Musca domestica* (Linnaeus) and *Chrysomya megacephala* (Fabricius) as mechanical vectors of helminthic parasites in a typical slum area of Metropolitan Manila. *Southeast Asian J Trop Med Public Health*, 22(2), 222–228.
- Moritz, C., & Cicero, C. (2004). DNA barcoding: Promise and pitfalls. *PLoS Biology*, 2(10), e354. <https://doi.org/10.1371/journal.pbio.0020354>
- Needleman, S. B. (1970). Needleman-Wunsch algorithm for sequence similarity searches. *Journal of Molecular Biology*, 48, 443–453.
- Newman, C., Buesching, C. D., & Macdonald, D. W. (2003). Validating mammal monitoring methods and assessing the performance of volunteers in wildlife conservation—“Sed quis custodiet ipsos custodiet?”. *Biological Conservation*, 113(2), 189–197. [https://doi.org/10.1016/S0006-3207\(02\)00374-9](https://doi.org/10.1016/S0006-3207(02)00374-9)
- Nicholson, E., Collen, B., Barausse, A., Blanchard, J. L., Costelloe, B. T., Sullivan, K. M. E., ... Milner-Gulland, E. J. (2012). Making robust policy decisions using global biodiversity indicators. *PLoS One*, 7(7), e41128. <https://doi.org/10.1371/journal.pone.0041128>
- Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115(25), E5716–E5725. <https://doi.org/10.1073/pnas.1719367115>
- Nyakatura, K., & Bininda-Emonds, O. R. (2012). Updating the evolutionary history of Carnivora (Mammalia): A new species-level supertree complete with divergence time estimates. *BMC Biology*, 10(1), 12. <https://doi.org/10.1186/1741-7007-10-12>
- Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGinn, D., ... P. (2018). *vegan: Community Ecology Package*. R package version 2.5-5. 2018.
- Paradis, E., & Schliep, K. (2018). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528.
- Pedersen, M. W., Overballe-Petersen, S., Ermini, L., Sarkissian, C. D., Haile, J., Hellstrom, M., ... Willerslev, E. (2015). Ancient and modern environmental DNA. *Phil. Trans. R. Soc. B*, 370(1660), 20130383. <https://doi.org/10.1098/rstb.2013.0383>
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- 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., ... 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>
- Salleh, F. M., Ramos-Madriral, J., Peñaloza, F., Liu, S., Sinding, M.-H.-S., Patel, R. P., ... Shamsir, M. S. (2017). An expanded mammal mitogenome dataset from Southeast Asia. *GigaScience*, 6(8), 1–8.
- 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., Douglas, W. Y., 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(1), 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. <https://doi.org/10.1016/j.cub.2012.10.014>
- 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>
- Smith, F. A., & Lyons, S. K. (2011). How big should a mammal be? A macroecological look at mammalian body size over space and time. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1576), 2364–2378.
- South, A. (2011). rworldmap: A new R package for mapping global data. *R Journal*, 3(1), 35–43. <https://doi.org/10.32614/RJ-2011-006>
- Steffen, W., Crutzen, P. J., & McNeill, J. R. (2007). The Anthropocene: Are humans now overwhelming the great forces of nature. *AMBIO: A Journal of the Human Environment*, 36(8), 614–621. [https://doi.org/10.1579/0044-7447\(2007\)36\[614:TAHNO\]2.0.CO;2](https://doi.org/10.1579/0044-7447(2007)36[614:TAHNO]2.0.CO;2)
- Swinton, J. (2016). *Venvenable: Venn and Euler area-proportional diagrams*. R package version 3.1. 0.9000. <https://github.com/js229/Venvenable>
- Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford, UK: Oxford University Press.
- 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>
- Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS One*, 7(8), e41732. <https://doi.org/10.1371/journal.pone.0041732>
- Tobler, M., Carrillo-Percegué, S., Pitman, R. L., Mares, R., & Powell, G. (2008). An evaluation of camera traps for inventorying large- and medium-sized terrestrial rainforest mammals. *Animal Conservation*, 11(3), 169–178. <https://doi.org/10.1111/j.1469-1795.2008.00169.x>
- Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., ... 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>
- Weiskopf, S. R., McCarthy, K. P., Tessler, M., Rahman, H. A., McCarthy, J. L., Hersch, R., ... Siddall, M. E. (2018). Using terrestrial haemaphysagous 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>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Berlin, Germany: Springer.
- Yoccoz, N. G., Bråthen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., ... 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>
- 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>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Gogarten JF, Hoffmann C, Arandjelovic M, et al. Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity. *Environmental DNA*. 2020;2:63–76. <https://doi.org/10.1002/edn3.46>