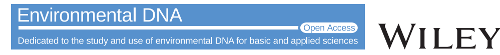
Check for updates

Received: 26 August 2022 | Revised: 15 May 2023 | Accepted: 17 May 2023

DOI: 10.1002/edn3.436

SPECIAL ISSUE ARTICLE



Mammal mitogenomics from invertebrate-derived DNA

Renita Danabalan[1,2] | Kevin Merkel[3,4] | Ida Bærholm Schnell[5] | Mimi Arandjelovic[6,7] | Christophe Boesch[6] | Gregory Brazzola[6] | Paula Dieguez[7] | Jef Dupain[8] | Magloire Kambale-Vyalengerera[6] | Hjalmar S. Kühl[6,7] | Constanze Hoffmann[3] | Juan Lapuente[6] | Van Ngoc Thinh[9,10] | Fee Zimmermann[3,11,12] | Fabian H. Leendertz[3,12] | M. Thomas P. Gilbert[5,13] | Christian Roos[9] | Camila Mazzoni[1,2] | Jan F. Gogarten[3,4,12,14] | Sébastien Calvignac-Spencer[3,4]

[1]Institute for Zoo and Wildlife Research, Berlin, Germany
[2]Berlin Centre for Genomics and Biodiversity Research, Berlin, Germany
[3]Epidemiology of Highly Pathogenic Organisms, Robert Koch Institute, Berlin, Germany
[4]Viral Evolution, Robert Koch Institute, Berlin, Germany
[5]Centre for Evolutionary Hologenomics, The Globe Institute, University of Copenhagen, Copenhagen, Denmark
[6]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[7]German Centre for Integrative Biodiversity Research (iDiv) Halle-Leipzig-Jena, Leipzig, Germany
[8]Antwerp Zoo Foundation, Royal Zoological Society Antwerp, Antwerp, Belgium
[9]German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany
[10]World Wide Fund for Nature (WWF) Vietnam, Hanoi, Vietnam
[11]Bundeswehr Institute of Microbiology, Munich, Germany
[12]Helmholtz Institute for One Health, Greifswald, Germany
[13]NTNU University Museum, Trondheim, Norway
[14]Applied Zoology and Nature Conservation, University of Greifswald, Greifswald, Germany

Correspondence
Jan F. Gogarten and Sébastien Calvignac-Spencer, Epidemiology of Highly Pathogenic Organisms, Robert Koch Institute, Berlin, Germany.
Email: jan.gogarten@gmail.com and calvignacs@rki.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: FZT-118 and GO 3443/1-1; Heinz L. Krekeler Foundation

Abstract

The metabarcoding of vertebrate DNA found in invertebrate-derived DNA (iDNA) has proven a powerful tool for monitoring biodiversity. To date, iDNA has primarily been used to detect the presence/absence of particular taxa using metabarcoding, though recent efforts demonstrated the potential utility of these data for estimating relative animal abundance. Here, we test whether iDNA can also be used to reconstruct complete mammalian mitogenomes and therefore bring the field closer to population-level analyses. Specifically, we used mitogenomic hybridization capture coupled with high-throughput sequencing to analyze individual ($N = 7$) or pooled ($N = 5$) fly-derived DNA extracts, and individual ($N = 7$) or pooled ($N = 1$) leech-derived DNA extracts, which were known a priori to contain primate DNA. All sources of iDNA showed their ability to generate large amounts of mammalian mitogenomic information and deeper

Jan F. Gogarten and Sébastien Calvignac-Spencer contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Environmental DNA* published by John Wiley & Sons Ltd.

sequencing of libraries is predicted to allow for even more complete recovery of primate mitogenomes from most samples (90%). Sixty percent of these iDNA extracts allowed for the recovery of (near) complete mammalian mitochondrial genomes (hereafter mitogenomes) that proved useable for phylogenomic analyses. These findings contribute to paving the way for iDNA-based population mitogenomic studies of terrestrial mammals.

KEYWORDS
fly, hybridization capture, iDNA, leech, mitochondrial genome, non-human primates

JEL CLASSIFICATION
phylogeny, primates

1 | INTRODUCTION

Persistent and growing threats to biological diversity necessitate the development of tools to document biodiversity in rapidly changing environments (Butchart et al., 2010; Mantyka-Pringle et al., 2015). Molecular advances have enabled the detection of DNA that organisms shed into their environment (i.e., environmental DNA, eDNA; Rondon et al., 2000; Taberlet et al., 2018), expanding the biodiversity monitoring toolkit. Notably, metabarcoding of vertebrate mitochondrial DNA (mtDNA) found in water (Andruszkiewicz et al., 2017; Port et al., 2016), sediments (Holman et al., 2019) and soils (Andersen et al., 2012) has proven effective for characterizing community composition in a diversity of ecosystems. A similar use of air has recently been suggested based on promising initial results in zoos (Clare et al., 2022; Lynggaard, Bertelsen, et al., 2022).

Vertebrate DNA is also actively sampled by invertebrates that come into contact with them or their byproducts. Like eDNA, invertebrate-derived DNA (iDNA) often allows for the detection of vertebrate mtDNA. Leeches (e.g., Drinkwater, Jucker, et al., 2021; Lynggaard, Oceguera-Figueroa, et al., 2022; Schnell et al., 2012, 2015, 2018), carrion flies (e.g., Calvignac-Spencer, Merkel, et al., 2013; Gogarten et al., 2020; Hoffmann et al., 2018; Rodgers et al., 2017), sandflies (Kocher et al., 2017; Massey et al., 2022), mosquitos (e.g., Kocher et al., 2017; Massey et al., 2022), and dung beetles (Drinkwater, Williamson, et al., 2021) have all been successfully used in metabarcoding studies targeting mtDNA.

The metabarcoding of short mitochondrial eDNA/iDNA fragments is useful to determine vertebrate species presence/absence, and may even be used to estimate species abundance when combined with careful experimental design and modeling (Carvalho et al., 2021; Fonseca, 2018; Luo et al., 2022). Yet, short mitochondrial eDNA/iDNA fragments are poorly suited to investigate within-species, population-level characteristics, including genetic diversity, population size, or the connectivity between populations (Adams et al., 2019; Sigsgaard et al., 2020). For such purposes, nuclear genomes (or subsamples thereof) would be a desirable target; yet, the retrieval of nuclear DNA from eDNA/iDNA extracts is still a challenging task (Sigsgaard et al., 2020). For example, one iDNA study predicted that microsatellite genotyping would only be possible from <5% of mtDNA-positive flies (Schubert et al., 2015).

While mtDNA, a single linked locus characterized by strict maternal inheritance, has inherent limitations for studying within-species, population-level characteristics, sequencing relatively large fragments of its fast-evolving regions or even the entire mitogenome generates valuable information for population genetics. Indeed, both strategies have been successfully employed in eDNA studies. For example, Sigsgaard et al. (2016) amplified and sequenced ca. 450 bp D-loop fragments of whale shark mtDNA using seawater eDNA and showed that estimates of haplotype frequency and abundance matched those derived from direct sampling of individuals partaking in aggregations. Similarly, eDNA from ancient sediments was used to reconstruct full mitogenomes of archaic hominins and other mammals, generating evidence of population turnovers at Denisova cave (Slon et al., 2017; Zavala et al., 2021). Other eDNA sources (e.g., waterholes) have also allowed for the reconstruction of vertebrate mitogenomes (Seeber et al., 2019).

Here, we set out to test whether iDNA is also amenable to mitogenomic analyses. Depending on invertebrate ecology and physiology, iDNA could be sampled in very small quantities (e.g., when collected from dung by carrion flies or dung beetles) and subject to active degradation through digestion processes (though this might be slow, as demonstrated in leeches; Schnell et al., 2012). This could mean that iDNA is too rare and/or degraded for the systematic generation of mitogenomic data for population genetics. To our knowledge attempts at generating mitogenome-wide information from iDNA have thus far been limited to a single study that targeted multiple mtDNA fragments of the Annamite striped rabbit (*Nesolagus timminsi*; Nguyen et al., 2021).

Here, we explore the potential of in-solution hybridization capture (the use of biotinylated nucleic acid baits to selectively enrich target library molecules in solution) as a tool to facilitate the sequencing of vertebrate mitogenomes from iDNA isolated from carrion flies and leeches. These invertebrates arguably represent the best explored iDNA sources, while also representing ecologically and physiologically divergent organisms (Calvignac-Spencer, Leendertz, et al., 2013). Specifically, we employ a bait set that spans

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

nonhuman primate (NHP) mitochondrial diversity (excluding humans) to enrich libraries constructed from DNA of individual flies and terrestrial leeches (hereafter leeches), as well as from pools of flies or leeches, all of which contained NHP mtDNA.

2 | METHODS AND MATERIALS

2.1 | Samples

From leeches, we selected individual DNA extracts derived from 7 individuals, as well as 1 pool of DNA containing 10 leeches in the pool (Schnell et al., 2018). From flies, we selected individual DNA extracts derived from 7 individuals as well as 5 pools of 6 individual DNA extracts (Table 1). We were unable to sort leeches or flies morphologically into taxonomic groups, and did not attempt to identify taxa using molecular approaches. Previous studies of the cytochrome oxidase *c* subunit I (COI) barcode diversity of flies captured for biomonitoring with these baits in Sub-Saharan Africa ecosystems have shown a diversity of species assigned to the families Calliphoridae, Sarcophagidae, and Muscidae (Gogarten et al., 2022; Hoffmann et al., 2017). Similarly, an analysis of COI barcodes of leeches collected for mammal metabarcoding also suggested a diversity of species from many genera and families are often collected (Schnell et al., 2018). All individual extracts and extract pools showed signs of containing NHP DNA in earlier experiments (Table 1; CR personal communication; Gogarten et al., 2020; Hoffmann et al., 2017; Schnell et al., 2018). Mammal mtDNA had also been quantified for individual fly DNA extracts using the quantitative PCR assay described in Schubert et al. (2015); Table 1, which comprised amplifying primers from Taylor (1996) and a human blocking primer from Boessenkool et al. (2012).

2.2 | Hybridization capture

The iDNA extracts were fragmented to 550bp using an S220 Focused-ultrasonicator (Covaris Inc) and Illumina sequencing libraries then prepared using the NebNext Ultra II Kit (New England Bioloabs) and dual indexed using the NEBNext Multiplex Oligos for Illumina (New England Biolabs). Libraries were quantified and equal masses were pooled to generate pools for hybridization capture (to avoid confusion with the extracted pools of flies and leeches discussed throughout the text as pools, we refer to these pools of libraries as "capture pools"). Specifically, we generated capture pools of leeches collected in Madagascar (one pool of two libraries), capture pools of leeches from Vietnam (two pools containing two and four libraries, respectively), and capture pools of flies from mainland Africa (three pools containing two, five, and five libraries, respectively; see Table 1 for details). For libraries with mammal quantification data available, we created capture pools with samples of similar concentrations together (Table 1). Specifically, where the information was available, we used the estimated mammalian mitogenome to background ratio to place libraries into capture pools within one $\log_{10}$ difference of one another.

In solution hybridization capture was performed on capture pools using RNA baits designed to span primate mitochondrial diversity (excluding humans; Daicel Arbor Biosciences). With this approach, DNA or RNA probes that are biotinylated are mixed to denatured library molecules, which then enables the selection of targeted fragment-probe heteroduplexes with the help of magnetic streptavidin beads (Enk et al., 2014). Nontargeted fragments in the library are removed from the solution through repeated washings, while probe/target library fragment heteroduplexes are retained on the magnetic beads. Afterwards, the targeted fragments can then be eluted from the beads. The bait design consisted of 19,552 80bp RNA baits; 3189 baits to span the selected gorilla, chimpanzee and bonobo mitogenomes, and 16,363 to target the remaining 81 NHP mitogenomes with an 75+% overlap between baits and clustering of baits with an 88+% identity (Table S1).

Capture reactions were set up according to the manufacturer's protocol. Briefly, for each capture pool, prior to capture a 20μL hybridization mix was prepared using 9μL of HYB #1, 0.5μL of HYB #2, 3.5μL of HYB #3, 0.5μL of HYB #4, 1μL of RNase Block and 5.5μL of the baits, which were diluted with water in a 1:4 ratio. A 5.5μL blocking mix was then prepared using 2.5μL of Block #1, 2.5μL of Block #2, 0.5μL of Block #3 and we added 7μL of each capture pool to a prepared blocking mix and incubated it at 95°C for 5min. The blocking mix targets the Illumina adapters and indices and prevents binding of library molecules to one another. The hybridization mix was then heated to the hybridization temperature of 65°C for 5min, then 18μL was added to the blocking-sample mix and thoroughly mixed. The hybridization capture took place at 65°C in a thermocycler for at least 24h. Capture products were cleaned using Dynabeads MyOne Streptavidin C1 Magnetic beads (Thermofisher). We prepared 800μL of wash buffer 2.2 using 6.4μL of HYB #4, 633.6μL of nuclease free water and 160μL of Wash Buffer 2, which was used for up to four washes of each capture pool. An aliquot of 30μL of beads was prepared with three 200μL washes of Binding buffer. Following the third wash, beads were resuspended in 70μL of Binding buffer. Resuspended beads were heated to the hybridization temperature (65°C) for 2min. All 25μL of the capture mix was transferred to the heated beads and incubated for 30min. Cleaning of the bead-sample mix comprised three steps which were repeated for a total of four times; first, the addition of 180μL of Wash Buffer, followed by incubation at 65°C for 10min and finally removal of the supernatant. Following the removal of the supernatant in the final wash, the pellet was resuspended in 30μL of TET. The cleaned product was then amplified using the KAPA Hot Start Library Amplification Kit to generate at least 200ng of library that was used for the second round of capture. Concentrations were assessed using the KAPA HiFi Library Quantification Kit; briefly, 1mL of the amplified libraries was taken and diluted 1:1000 and 1:8000 and quantified with a mix consisting of 12μL of KAPA Master Mix, 4μL of water and 4μL of diluted samples with the following thermocycling conditions: 95°C for 5min, then 35cycles of 95° for 30s and 60°C for 45s, ending with a

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Samples used for hybridization capture and summary of sequencing results.

Sample	Location	DNA conc. (ng/μL)	Mammal mt conc. (copies/μL)	Capture pool	Total *N* reads	Primate reads (%)	Human reads (%)	Reads mapped (%)	Mt genome coverage ≥10× and 95% agreement (%)	Best hit in BLAST search (% identity)	Closest NHP relative present in respective locality
Individual leeches											
53	Bach Ma NP, Vietnam	8	-	A	1,709,483	86.1	1.1	56.7	91.3	*Macaca mulatta* (94.3)	*Macaca mulatta*
177	Bach Ma NP, Vietnam	6.8	-	A	1,885,645	86.1	0.9	37.2	94.0	*Macaca arctoides* (95.9)	*Macaca arctoides*
181	Bach Ma NP, Vietnam	7.6	-	A	1,458,247	88.4	0.8	38.2	96.5	*Macaca arctoides* (97.0)	*Macaca arctoides*
315	Bach Ma NP, Vietnam	16.1	-	B	1,315,860	90.5	2.3	66.2	99.3	*Macaca arctoides* (97.9)	*Macaca arctoides*
393	Bach Ma NP, Vietnam	6.7	-	B	6,590,944	91.9	0.5	67.9	98.9	*Macaca mulatta* (94.3)	*Macaca mulatta*
394[a]	Bach Ma NP, Vietnam	10.9	-	A	2122	79.5	0.1	71.5	17.6	*Macaca arctoides* (95.6–99.2[c])	*Macaca arctoides*
2_2[a]	Ranomafana NP, Madagascar	4.2	-	C	0	-	-	-	-	-	
Leech pool											
17_4[a]	Andasibe Madagascar	3.6	-	C	886,816	91.3	62.2	7.7	11.4	*Eulemur fulvus* (92.3–99.4[c])	*Eulemus fulvus*
Individual flies											
1184[a]	Bili, DRC	5	6	D	599,132	82.9	0.9	43.0	40.1	*Lophocebus aterrimus* (94.8–100[c])	*Lophocebus albigena*
1251[a,b]	Bili, DRC	23	6	D	530,786	62.6	3.0	52.8	48.5	*Cercocebus chrysogaster* (94.5–97.8[c])	*Cercocebus agilis*
666[a]	Taï NP, Côte d'Ivoire	7	1	D	0	-	-	-	-	-	
1340	Taï NP, Côte d'Ivoire	190	78	D	1,898,379	79.5	3.9	68.0	95.2	*Cercocebus atys* (91.8–96.1[c])	*Cercocebus atys*
1458[b]	Dzanga-Sangha NP, CAF	6	2469	E	6,230,515	87.0	1.7	65.1	97.5	*Cercocebus chrysogaster* (95.7)	*Cercocebus agilis*
1522[b]	Dzanga-Sangha NP, CAF	3	171	E	3,607,011	87.1	1.6	60.3	94.5	*Cercocebus chrysogaster* (94.7)	*Cercocebus agilis*
1542[b]	Dzanga-Sangha NP, CAF	102	324	D	1,562,749	79.7	1.2	59.5	97.3	*Cercocebus chrysogaster* (95.7)	*Cercocebus agilis*
Fly pools											
79[a]	Grebo-Krahn NP, Liberia	5.5	-	F	433,392	29.1	8.5	17.4	8.0	*Cercocebus atys* (93.9–96.5[c])	*Cercocebus atys*

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

TABLE 1 (Continued)

Sample	Location	DNA conc. (ng/µL)	Mammal mt conc. (copies/µL)	Capture pool	Total N reads	Primate reads (%)	Human reads (%)	Reads mapped (%)	Mt genome coverage ≥10× and 95% agreement (%)	Best hit in BLAST search (% identity)	Closest NHP relative present in respective locality
81	Grebo-Krahn NP, Liberia	2.6	-	F	438,304	75.8	6.3	53.3	67.6	*Cercopithecus petaurista* (92.5–99.2[c])	*Cercopithecus petaurista*
86[a]	Grebo-Krahn NP, Liberia	4.1	-	F	4156	62.1	5.6	26.6	13.9	*Cercopithecus petaurista* (95.7[d])	*Cercopithecus cephus*
104	Taï NP, Côte d'Ivoire	4.4	-	F	1,515,441	76.7	0.3	49.7	86.8	*Cercopithecus petaurista* (94.3–98.4[c])	*Cercopithecus petaurista*
114	Sobeya, Guinea	4.6	-	F	2,033,336	64.6	6.9	50.2	88.3	*Papio papio* (94.8–98.4[c])	*Papio papio*

Abbreviations: CAF, Central African Republic; DRC, Democratic Republic of Congo; NP, National Park.

[a]Not included in Figure 1d as we only considered fly- or leech-derived mitogenomes with more than 50% coverage.

[b]No complete mitogenome of *Cercocebus agilis* is currently available thus BLAST results return *C. chrysogaster* as the closest match to the consensus sequence.

[c]Lowest and highest hit among multiple short sequences.

[d]A single short sequence returned a best hit on *C. petaurista*.

dissociation cycle. Samples with less than 200 ng of DNA were amplified further for a maximum of up to 10 cycles and re-quantified. We then used at least 200 ng of the cleaned and enriched capture product for the second capture reaction. Cleaned and quantified products from the second capture reaction were then pooled in equimolar ratios and sequenced on an Illumina NextSeq using V2 chemistry (2 × 150 cycles; Illumina).

2.3 | Bioinformatic and statistical analysis

For each sample, the 150 bp forward and reverse reads were trimmed using Trimmomatic v.038 (Bolger et al., 2014) to remove the adapters (ILLUMINACLIP:2:30:10:8), low-quality bases at both ends as well as within the reads (LEADING:30 TRAILING:30 SLIDINGWINDOW:4:30) and to keep reads of a minimum length (MINLEN:40). Trimmed reads were then merged using clip&merge v1.7.8 (Peltzer et al., 2016) and all merged, unmerged and unpaired reads were concatenated into a single file per sample.

As a preliminary exploration of the vertebrate reads generated in these experiments, we classified reads using Kraken (Wood & Salzberg, 2014). For this, a custom database of all complete mitochondrial vertebrate sequences was created from GenBank. Trimmed reads were mapped to a human reference sequence to remove human reads using bwa MEM (Li, 2013). Unmapped reads were then competitively mapped to 220 unique NHP mitogenomes (Table S1). Each sample was then remapped to the single best reference, which was selected on the basis of having the most mapped reads in the initial mapping. Bam files were then sorted and duplicate reads removed using the *SortSam* and *MarkDuplicates* functions in the Picard package of tools (The Broad Institute; http://broadinstitute.github.io/picard/), with coverage calculated using samtools' depth function (Li et al., 2009). Average insert size was determined with the *CollectInsertSizeMetrics* function in Picard, using only merged and unmerged reads. To explore the accumulation of unique molecules in libraries at different sequencing depths, we used preseq's *lc extrap* function, which bootstraps (BSs) the observed duplicate counts histogram (Daley & Smith, 2014). Libraries are amplified during their preparation and after capture, creating the potential for resequencing of the same amplified initial fragment multiple times, and errors introduced during the PCR can thus be sequenced multiple times. As such, when considering the coverage and certainty in calling bases of the target genome, there is considerable interest in acquiring as many unique on-target reads as possible. In essence, such efforts explore when sequencing efforts have exhausted the library's molecular complexity, beyond which resequencing will not generate novel information about the target of interest (Daley & Smith, 2014; Enk et al., 2014).

A consensus sequence for each sample was generated using Geneious Prime (Biomatters Limited) with a threshold of 95% identity among ≥10 unique read coverage for a base to be called. For a first preliminary species identification, we ran a BLAST search of the consensus sequences against the NCBI nonredundant database and

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

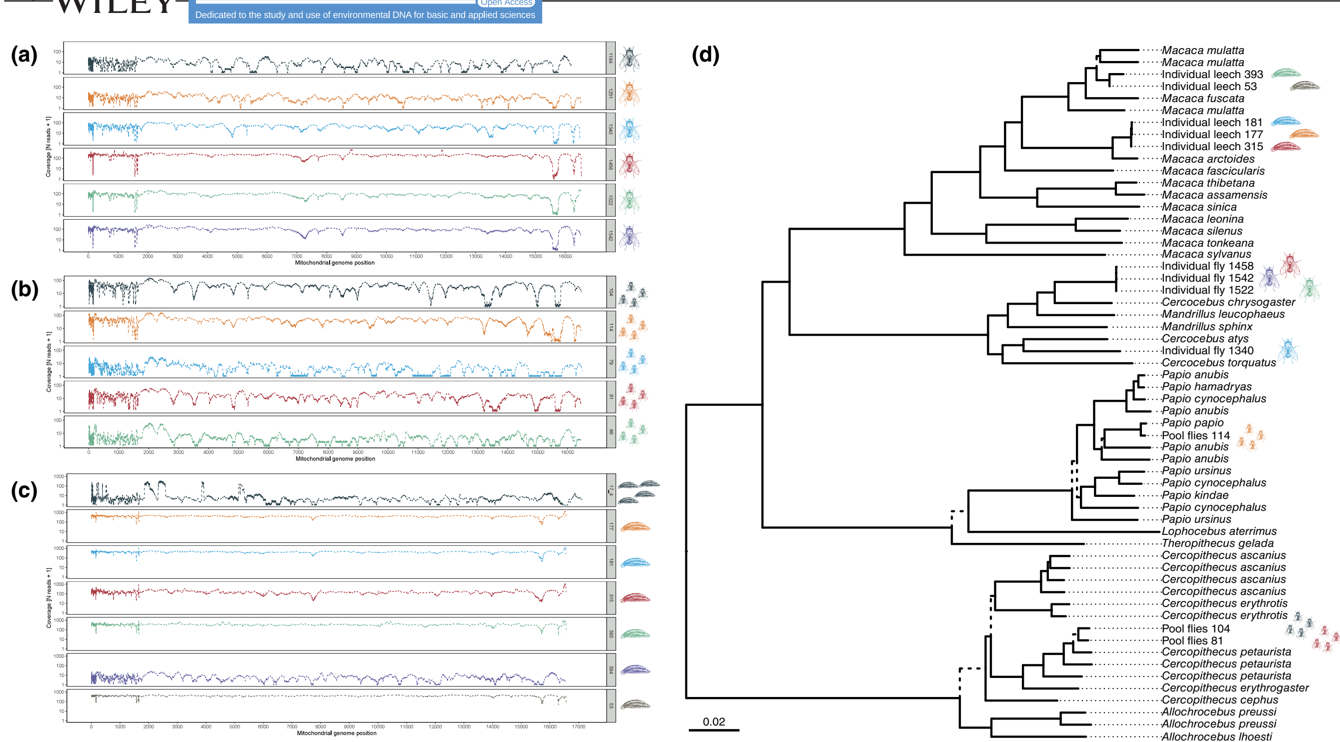


FIGURE 1 Mitogenome coverage generated from libraries from (a) individual flies, (b) pools of flies, or (c) individual or pools of leeches. The dashed lines indicate the moving 10 bp average across the genome. The fly or leech images next to the plots indicate a unique color/shape combination and whether individual extracts or pools were used. (d) A maximum likelihood phylogeny inferred from an alignment of 91 primate mitogenomes genomes from GenBank and the high-quality mitogenomes generated from flies, leeches, or pools of flies. The phylogeny presented is pruned to aid in interpretation, but the full tree including additional information about the nodes and branch lengths is available in the Data S1. A minimum 10× coverage was required to call a base for these mitogenomes with a threshold of 95% identity for a base to be called. Nodes that appeared in >95% of bootstrap replicates shown as solid lines. The scale shows nucleotide substitutions per site.

recorded the best hit species name and sequence identity (Altschul et al., 1990). For a better, phylogeny-informed species assignment, we downloaded mitochondrial genomes of Cercopithecinae from GenBank (for GenBank accession numbers see Data S1). The dataset was filtered to contain at least one representative of all available species, subspecies and lineages of Cercopithecinae in GenBank, but sequences with less than 80% sequence coverage were excluded (the selected sequence set was independent from the set of reference sequences used for capture bait design). All fly-derived mitogenomes with more than 50% coverage were included. The resulting 91 sequences from GenBank and 12 newly generated mitogenomes were aligned with MAFFT v7.307 (Katoh & Standley, 2013). We identified conserved blocks using Gblocks (Talavera & Castresana, 2007) as implemented in SeaView v5 (Gouy et al., 2021), reducing the alignment to 15,040 positions. A maximum-likelihood tree was reconstructed in IQ-TREE 1.5.3 (Nguyen et al., 2014) using the best-fit substitution model (TIM3+I+G4) as automatically calculated in IQ-TREE using the option-m TEST (similar to jModelTest) and the Bayesian information criterion. Node support was obtained via 10,000 ultrafast BS replicates (Hoang et al., 2018).

To explore the utility of iDNA to describe within-species variation, we calculated the pairwise divergence between the high-quality iDNA mitogenomes included in the phylogenetic analysis described above. We did this with Geneious Prime and considered only unambiguous differences between pairs of sequences. This analysis compared the three individual leech-derived *Macaca arctoides* mitogenomes from Bach Ma NP, Vietnam, the two leech-derived *Macaca mulatta* mitogenomes from Bach Ma NP, as well as the three fly-derived *Cercocebus agilis* mitogenomes from Dzanga-Sangha NP, CAF and the *C. agilis* mitogenome from Bili, DRC.

Previously, mammal mtDNA estimates were generated for 1154 individual flies captured in a diversity of Sub-Saharan African ecosystems (Gogarten et al., 2020; Hoffmann et al., 2017). Here, we compared these concentrations to the concentrations of individual flies from which we were successfully able to generate usable mitogenome coverage, to estimate the proportion of flies amenable for mitogenomic analyses.

Figures were created in R v4.1.2 (R Core Team, 2021), using the ggplot2 (Wickham, 2016) and ggtree2 packages (Yu et al., 2017).

3 | RESULTS

All but two of the samples generated reads that could be subsequently analyzed (Table 1). We generated a total of 32,700,196 paired end reads ($\bar{x}_{\text{per sample}}$ = 1,721,063 paired reads, range = 0–6,685,581). Average read lengths ranged from 40 bp to 151 bp ($\bar{x}_{\text{per sample}}$ = 95.5). The Kraken analysis revealed that the majority of the reads originated

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

from primates ($\bar{x}$ = 77.8% of total reads, range = 29.1%–91.9% of total reads; Table 1). The Kraken analysis also suggested that typically, only a small number of the total reads sequenced were of human origin ($\bar{x}$ = 6.0% of total reads, range = 0.1%–62.1% of reads), with only a single sample (the leech pool) having more than 10% of reads assigned as human (Table 1). All but one of the individual leeches generated significant amounts of NHP reads and the single pooled leech sample contained the least NHP and the most human reads (Table 1). Similarly, it appeared that individual flies generated relatively more NHP reads than pools of flies (Table 1).

Mapping of non-human reads from each sample to the best candidate primate mitogenome, revealed that it was possible to recover partial mitogenomes from 18 of the 20 samples tested (Table 1; Figure 1a–c). The proportion of reads mapping to the best primate mitogenome seemed to be higher for individual flies and leeches than for pools of flies or leeches (Table 1). Insert size only slightly varied by source, with fly pools having the largest insert sizes ($\bar{X}$ = 273 bp), followed by individual flies ($\bar{X}$ = 253 bp), and leeches generating the smallest insert sizes ($\bar{X}$ = 235 bp; Figure 2d). As DNA was fragmented prior to library construction, these insert sizes are underestimates of the condition of vertebrate DNA originally found in these flies and leeches and represent a lower bound for the fragment sizes of the DNA in the samples themselves.

Among individual fly samples, the proportion of reads mapping to the best primate mitogenome was at its lowest for samples with extremely low mtDNA copy numbers. While sample sizes are small, we were able to generate considerable genomic data from individual flies that had more than six mitochondrial copies/μL. Comparing this rough threshold to estimates of mitochondrial copy numbers in our collection of individual flies from various ecosystems in Sub Saharan Africa revealed that 24.1% of flies (278/1154) in these ecosystems are suitable for mitogenomic analysis (Figure 3).

Of the 18 samples with detectable mapping to the primate mitogenome, it was possible to recover nearly complete genomes from a majority of samples ($\bar{x}$ = 71.8% of primate mitogenome covered at ≥10×, range = 8.9%–100.0% of primate mitogenome covered at ≥10×; Table 1). The completeness of genomes recovered was higher for individual flies and leeches than from pools of flies and leeches and was again at its lowest for the three individual flies with very low mtDNA concentration (Table 1). Samples that generated lower completeness also tended to exhibit more uneven coverage (Figure 1a–c). The accumulation of unique molecules in libraries generated from all types of samples (except for those two samples that did not generate any reads) suggests that further sequencing of these libraries would continue to improve mitogenome completeness (Figure 2a–c). BLAST search using the 18 consensus sequences returned best hits consistent with initial metabarcoding results, local species composition, and mitogenome availability in public databases (Table 1). Finally, the 12 non-chimeric mtDNA consensus sequences covering more than 50% of the mitogenome clustered with sequences from species present in the ecosystem where the invertebrates were collected (or the closest species from which a mitogenome was available; Figure 1d).

iDNA detected mitogenomic variation within NHP species. Two of the three individual leech-derived *M. arctoides* mitogenomes were identical (Individual leech 177 and 181), while one (Individual leech 315) differed from these two genomes at three sites. The two leech-derived *M. mulatta* mitogenomes from Bach Ma NP, Vietnam (Individual leech 53 and 393) differed from one another at 12 sites. The three fly-derived *C. agilis* mitogenomes from Dzanga-Sangha NP, CAF were nearly identical, with two of the *C. agilis* mitogenomes different from the others at a single position each (Individual fly 1542 and 1458). The *C. agilis* mitogenome from Bili, DRC differed from the other mitogenomes assigned to the same species at 219 sites.

4 | DISCUSSION

Our results demonstrate that hybridization capture allows for the reconstruction of mammalian mitogenomes from leeches and flies, which are then useable for phylogenomic analyses. This joins a growing body of evidence suggesting that iDNA not only has the potential to serve as a biomonitoring tool to assess the presence/absence of species, but also to document a variety of population genomic characteristics (Keven et al., 2019; Nguyen et al., 2021; Schubert et al., 2015).

We found that many of our primate positive iDNA samples generated mitogenomic information useful for phylogenomic analyses. Perhaps unsurprisingly, mammalian mtDNA concentrations in flies were related to the performance of hybridization capture, with the fly extract with the lowest concentration being the only individual fly sample that failed to generate mitogenomic data. Given the seeming importance of mitogenome concentrations on our ability to reconstruct mitogenomes, the poorer performance of pooling extracts of flies and leeches that we observed might simply reflect the dilution of the positive extract(s) in these pools (Mata et al., 2019). Despite this limitation, iDNA analyses often work with pools of flies or leeches that are extracted together to reduce time and costs of extractions and for metabarcoding approaches, this appears to not reduce sensitivity greatly. Here, we set out to test whether an hybridization capture based approach also worked with these pools of insects. Our results demonstrated that it is possible to generate mitogenomic data from these pools. Furthermore, using target concentration information to optimize pooling schemes of libraries generated from pools prior to hybridization capture, so that samples with more similar concentrations of mammalian mtDNA are captured together, might improve success rates. Our preseq analysis also suggests that further sequencing of poorly performing samples would generally continue to improve mitogenome recovery (clearly the case for six of the eight poorly performing samples). Further sequencing is complicated by the fact that capture was performed on pooled libraries, so a resequencing effort would likely benefit from a new capture experiment on those samples needing further sequencing. Importantly, our results suggest that a sizeable fraction of individual flies positive for mammalian species in a diversity of sub-Saharan African ecosystems are amenable for such mitogenomic

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

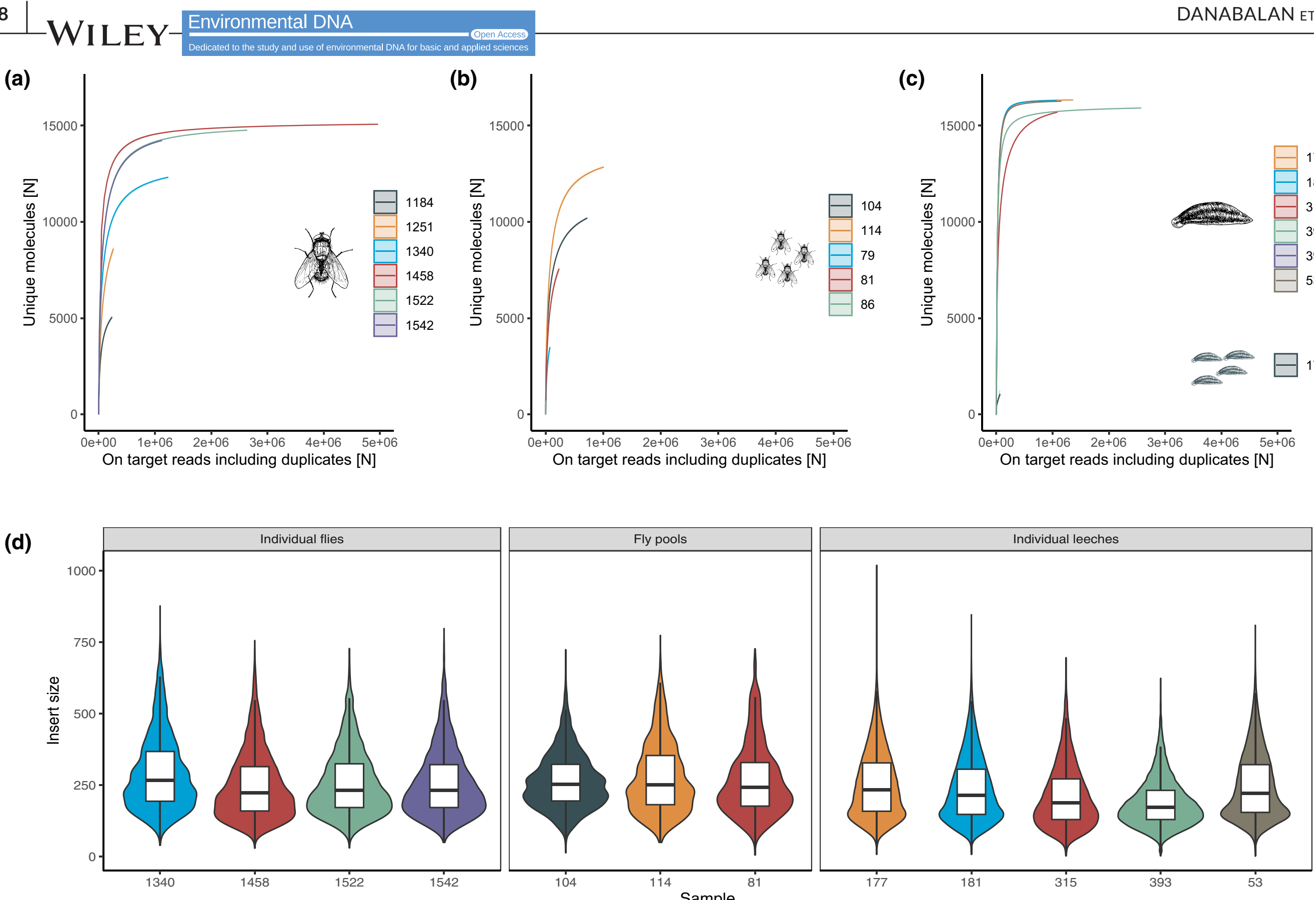


FIGURE 2 The accumulation of unique molecules in libraries generated from (a) individual flies, (b) pooled flies, and (c) individual and pooled leaches. The accumulation of unique molecules is shown in relation to the number of on target reads from the experiment, including duplicates. The shaded regions show 95% confidence intervals for each curve which is computed by bootstrapping the observe counts histogram using the program preseq. (d) Violin plots and overlaid box plots showing the insert size for each of the samples that are shown in Figure 1d, using the corresponding color code. As DNA was fragmented as part of the library preparation procedure, these insert size distributions represent a lower bound for the fragment size distribution of the samples prior to fragmentation.

analyses, showcasing the feasibility of large-scale mitogenomic studies based on iDNA.

We found variation between mitogenomes assigned to the same species and detected within an ecosystem, and also identified considerable divergence between distant populations from what has historically been considered a single species, *C. agilis*. These findings demonstrate the potential of iDNA to contribute data to explore within-species, population-level characteristics, in particular genetic diversity and has clear potential to be applied to estimating population sizes and the connectivity between populations (Adams et al., 2019; Sigsgaard et al., 2020).

The promising performance of iDNA for mitogenomic analyses observed here may be due in part to the large average fragment sizes observed, despite having included a fragmentation step in our library preparation protocol. The expectation that iDNA is highly fragmented represents a largely untested hypothesis that has driven many studies to target short fragments of mtDNA (<150 bp). For example, a study of 216 leech samples found that 144 of them were amenable to the amplification and sequencing of mtDNA 16S rRNA and cytochrome *b* (cytb) fragments of 461 to 486 bp lengths (Morishima et al., 2020). In carrion flies, a study comparing the detection of shorter (130 bp) and longer mammal mtDNA fragments (300 bp) did not reveal dramatic difference in the detection rates (25% for short fragment vs. 40% for the long fragment; Calvignac-Spencer, Merkel, et al., 2013). In contrast, a study with tsetse flies found that the blood meal could be identified in 35% of samples when targeting long fragments of the cytb (450 bp) and COI genes (330 bp or 660 bp); by using a shorter fragment (150 bp of the 16S rRNA), they were able to recover the blood meal in 68 of the 89 (76%) samples that could not be identified with the longer PCR systems, which suggests more fragmented iDNA in tsetse flies (Bitome-Essono et al., 2017). A more thorough examination of the fragmentation of iDNA in different invertebrates and environments will help guide the selection of PCR target lengths and library preparation methods to maximize detection rates and mitogenomic reconstructions. A major benefit of targeting mammal mtDNA with hybridization capture is that it performs well with even highly fragmented samples; if library building costs decrease, this may in some cases represent a cost-effective and sensitive screening tool.

A potential limiting factor for using iDNA for population-level inferences is that samples can contain the mtDNA not just of multiple species, but also from multiple individuals of a single species

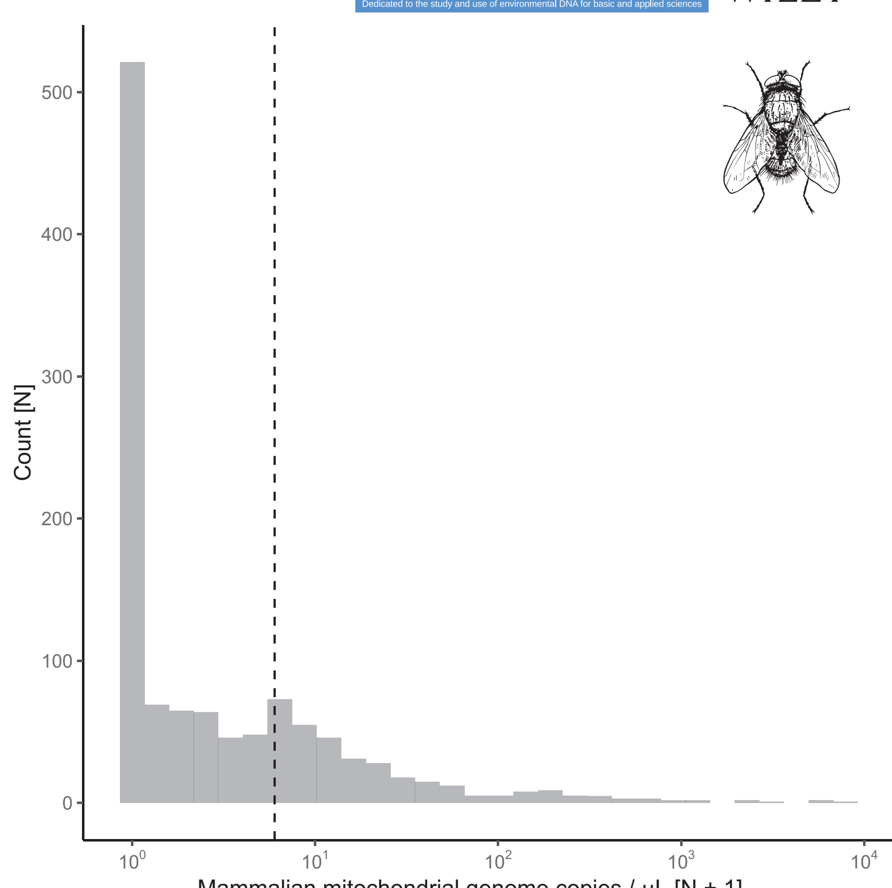


FIGURE 3 Histogram of the mammalian mitogenome concentrations from 1154 individual flies captured in a diversity of Sub-Saharan African ecosystems. The dashed line indicates our rough estimate for a threshold above which we were able to consistently generate useful mammalian mitochondrial genomic information using hybridization capture. To enable plotting on a log scale with zero values, we added one to each copy number estimate.

(Sigsgaard et al., 2020). The magnitude of the problem is uncertain and will likely depend on the invertebrate sampler. The impact of this problem may be mitigated by: (i) favoring the analysis of individual invertebrates, in that it typically isolates one or at most a few mitogenomes, (ii) focusing mitogenomic efforts on metabarcoded specimens that do not show preliminary evidence for multiple variants' being present, and, possibly, (iii) applying long-read sequencing to resolve mixed mitochondrial haplotypes exhibiting even low divergence (Sigsgaard et al., 2020), if raw read accuracy continues improving (Wang et al., 2021). Ultimately, the risk of inadvertent assembly of chimeric mitogenomes can likely be kept at quite low levels, given the nonrecombinant nature of mtDNA and the availability of many chimerism and recombination detection tools.

The cost of metabarcoding invertebrates for vertebrate biodiversity monitoring have declined to the point where they could effectively be routinely used to complement traditional survey techniques (e.g., a recent estimate of the price per fly pool was €31.93: Gogarten et al., 2020). Pools containing the DNA of animals of interest could then be targeted using the methods presented here to generate mitogenomes for an additional cost of ~100–150€/pool for library preparation and hybridization capture. The cost of sequencing the capture products will then depend on the depth of coverage needed and the sequencing platform used, but the high proportion of reads on target typically observed here, suggest the costs to generate mitogenomes will be on the order of ~100–200€/pool (e.g., a MiSeq V2 chemistry 2 × 150 costs 1035€ and generates 7.5 million paired-end reads, with library diversity typically plateaued after ~1 million reads on targeted). We conclude that, although more detailed exploration will help optimize these tools, mitogenomic studies are probably already within the reach of iDNA users.

AUTHOR CONTRIBUTIONS

Conception of the study: Renita Danabalan, M. Thomas P. Gilbert, Christian Roos, Camila Mazzoni, Jan F. Gogarten, Sébastien Calvignac-Spencer. *Acquisition, analysis and interpretation of the data*: Renita Danabalan, Kevin Merkel, Ida Bærholm Schnell, Mimi Arandjelovic, Christophe Boesch, Gregory Brazzola, Paula Dieguez, Jef Dupain, Magloire Kambale-Vyalengerera, Hjalmar S. Kühl, Constanze Hoffmann, Juan Lapuente, Van Ngoc Thinh, Fee Zimmermann. *Writing of the initial draft of the manuscript*: Renita Danabalan, Christian Roos, Jan F. Gogarten, Sébastien Calvignac-Spencer.

ACKNOWLEDGMENTS

We also would like to thank the following authorities who granted permission to conduct research: Ministere des Eaux et Forets, Cote d'Ivoire, Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, Côte d'Ivoire Institut Congolais pour la Conservation de la Nature, DR-Congo, Ministere de la Recherche Scientifique, DR-Congo, Ministere de l'Agriculture de l'Elevage et des Eaux et Forets, Guinea, the Forestry Development Authority, Liberia and the government of the Central African Republic for their long term support, particularly the Ministère des Eaux, Fôrets, Chasse et Peche and the

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Ministère de l'Enseignement Superieur, de la Recherche Scientifique et de l'Innovation Technologique. We are grateful Karsten Dierks, Dervla Dowd, John Hart, Thurston Cleveland Hicks, Vincent Lapeyre, Sergio Marocculi, Amelia Meier, Mizuki Murai, Emmanuelle Normand, Virginie Vergnes and Roman M. Wittig for their help with data collection, field logistics and/or support as part of the PanAf. We thank the WWF forest guards and rangers of Bach Ma National Park, Vietnam for collecting leeches. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

JFG was supported by funding from the Deutsche Forschungsgemeinschaft for the study entitled, the ecology and evolution of primate phageomes (DFG; GO 3443/1-1; Project number 453352748). The Pan African Programme: The Cultured Chimpanzee (PanAf) is supported by the Max Planck Society, Max Planck Society Innovation Fund, Heinz L. Krekeler Foundation and inter alia by the Deutsche Forschungsgemeinschaft (DFG; FZT-118).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All sequence data have been uploaded to Zenodo with the following DOI: https://doi.org/10.5281/zenodo.7018975.

ORCID

Jan F. Gogarten https://orcid.org/0000-0003-1889-4113
Sébastien Calvignac-Spencer https://orcid.org/0000-0003-4834-0509

REFERENCES

Adams, C. I., Knapp, M., Gemmell, N. J., Jeunen, G.-J., Bunce, M., Lamare, M. D., & Taylor, H. R. (2019). Beyond biodiversity: Can environmental DNA (eDNA) cut it as a population genetics tool? *Genes*, *10*(3), 192.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403–410.

Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjaer, K. H., Orlando, L., Gilbert, M. T., & Willerslev, E. (2012). Metabarcoding of 'dirt'DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, *21*(8), 1966–1979.

Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., & Boehm, A. B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLoS One*, *12*(4), e0176343.

Bitome-Essono, P.-Y., Ollomo, B., Arnathau, C., Durand, P., Mokoudoum, N. D., Yacka-Mouele, L., Okouga, A. P., Boundenga, L., Mve-Ondo, B., Obame-Nkoghe, J., Mbehang-Nguema, P., Njiokou, F., Makanga, B., Wattier, R., Ayala, D., Ayala, F. J., Renaud, F., Rougeron, V., Bretagnolle, F., … Paupy, C. (2017). Tracking zoonotic pathogens using blood-sucking flies as 'flying syringes'. *eLife*, *6*, e22069. https://doi.org/10.7554/eLife.22069

Boessenkool, S., Epp, L. S., Haile, J., Bellemain, E., Edwards, M., Coissac, E., Willerslev, E., & Brochmann, C. (2012). Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology*, *21*(8), 1806–1815.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Butchart, S. H. M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J. P. W., Almond, R. E. A., Baillie, J. E., Bomhard, B., Brown, C., Bruno, J., Carpenter, K. E., Carr, G. M., Chanson, J., Chenery, A. M., Csirke, J., Davidson, N. C., Dentener, F., Foster, M., Galli, A., … Watson, R. (2010). Global biodiversity: Indicators of recent declines. *Science*, *328*(5982), 1164–1168. https://doi.org/10.1126/science.1187512

Calvignac-Spencer, S., 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*, *35*(11), 1004–1013.

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P. M., Metzger, S., Schubert, G., & Leendertz, F. H. (2013). Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology*, *22*(4), 915–924.

Carvalho, C. S., De Oliveira, M. E., Rodriguez-Castro, K. G., Saranholi, B. H., & Galetti, P. M., Jr. (2021). Efficiency of eDNA and iDNA in assessing vertebrate diversity and its abundance. *Molecular Ecology Resources*, *22*(4), 1262–1273.

Clare, E. L., Economou, C. K., Bennett, F. J., Dyer, C. E., Adams, K., McRobie, B., Drinkwater, R., & Littlefair, J. E. (2022). Measuring biodiversity from DNA in the air. *Current Biology*, *32*, 693–700.

Daley, T., & Smith, A. D. (2014). Modeling genome coverage in single-cell sequencing. *Bioinformatics*, *30*(22), 3159–3165.

Drinkwater, R., Jucker, T., Potter, J. H. T., Swinfield, T., Coomes, D. A., Slade, E. M., Gilbert, M. T. P., Lewis, O. T., Bernard, H., Struebig, M. J., Clare, E. L., & Rossiter, S. J. (2021). Leech blood-meal invertebrate-derived DNA reveals differences in Bornean mammal diversity across habitats. *Molecular Ecology*, *30*(13), 3299–3312. https://doi.org/10.1111/mec.15724

Drinkwater, R., Williamson, J., Clare, E. L., Chung, A. Y., Rossiter, S. J., & Slade, E. (2021). Dung beetles as samplers of mammals in Malaysian Borneo—A test of high throughput metabarcoding of iDNA. *PeerJ*, *9*, e11897.

Enk, J. M., Devault, A. M., Kuch, M., Murgha, Y. E., Rouillard, J.-M., & Poinar, H. N. (2014). Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution*, *31*(5), 1292–1294. https://doi.org/10.1093/molbev/msu074

Fonseca, V. G. (2018). Pitfalls in relative abundance estimation using eDNA metabarcoding. *Molecular Ecology Resources*, *18*(5), 923–926. https://doi.org/10.1111/1755-0998.12902

Gogarten, J. F., Hoffmann, C., Arandjelovic, M., Sachse, A., Merkel, K., Dieguez, P., Agbor, A., Angedakin, S., Brazzola, G., Jones, S., Langergraber, K. E., Lee, K. C., Marrocoli, S., Murai, M., Sommer, V., Kühl, H. S., Leendertz, F. H., & Calvignac-Spencer, S. (2020). Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity. *Environmental DNA*, *2*(1), 63–76. https://doi.org/10.1002/edn3.46

Gogarten, J. F., Jahan, M., Calvignac-Spencer, S., Chapman, C. A., Goldberg, T. L., Leendertz, F. H., & Rothman, J. M. (2022). The cost of living in larger primate groups includes higher fly densities. *EcoHealth*, *19*(2), 290–298. https://doi.org/10.1007/s10393-022-01597-5

Gouy, M., Tannier, E., Comte, N., & Parsons, D. P. (2021). Seaview version 5: A multiplatform software for multiple sequence alignment, molecular phylogenetic analyses, and tree reconciliation. *Methods in Molecular Biology (Clifton, N.J.)*, *2231*, 241–260.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522.

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Hoffmann, C., Zimmermann, F., Biek, R., Kuehl, H., Nowak, K., Mundry, R., Agbor, A., Angedakin, S., Arandjelovic, M., Blankenburg, A., Brazolla, G., Corogenes, K., Couacy-Hymann, E., Deschner, T., Dieguez, P., Dierks, K., Düx, A., Dupke, S., Eshuis, H., … Leendertz, F. H. (2017). Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest. *Nature*, *548*(7665), 82–86.

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2019). Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. *Scientific Reports*, *9*(1), 11559. https://doi.org/10.1038/s41598-019-47899-7

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Keven, J. B., Walker, E. D., & Venta, P. J. (2019). A microsatellite multiplex assay for profiling pig DNA in mosquito bloodmeals. *Journal of Medical Entomology*, *56*(4), 907–914.

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

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. arXiv preprint arXiv:1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Luo, M., Ji, Y., Warton, D., & Yu, D. W. (2022). Extracting abundance information from DNA-based data. *Molecular Ecology Resources*, *23*, 174–189. https://doi.org/10.1111/1755-0998.13703

Lynggaard, C., Bertelsen, M. F., Jensen, C. V., Johnson, M. S., Frøslev, T. G., Olsen, M. T., & Bohmann, K. (2022). Airborne environmental DNA for terrestrial vertebrate community monitoring. *Current Biology*, *32*, 701–707.e5.

Lynggaard, C., Oceguera-Figueroa, A., Kvist, S., Gilbert, M. T. P., & Bohmann, K. (2022). The potential of aquatic bloodfeeding and nonbloodfeeding leeches as a tool for iDNA characterisation. *Molecular Ecology Resources*, *22*(2), 539–553. https://doi.org/10.1111/1755-0998.13486

Mantyka-Pringle, C. S., Visconti, P., Di Marco, M., Martin, T. G., Rondinini, C., & Rhodes, J. R. (2015). Climate change modifies risk of global biodiversity loss due to land-cover change. *Biological Conservation*, *187*, 103–111.

Massey, A. L., Bronzoni, R. V. M., da Silva, D. J. F., Allen, J. M., de Lázari, P. R., Dos Santos-Filho, M., Canale, G. R., Bernardo, C. S. S., Peres, C. A., & Levi, T. (2022). Invertebrates for vertebrate biodiversity monitoring: Comparisons using three insect taxa as iDNA samplers. *Molecular Ecology Resources*, *22*(3), 962–977. https://doi.org/10.1111/1755-0998.13525

Mata, V. A., Rebelo, H., Amorim, F., McCracken, G. F., Jarman, S., & Beja, P. (2019). How much is enough? Effects of technical and biological replication on metabarcoding dietary analysis. *Molecular Ecology*, *28*(2), 165–175. https://doi.org/10.1111/mec.14779

Morishima, K., Nakano, T., & Aizawa, M. (2020). Sika deer presence affects the host–Parasite interface of a Japanese land leech. *Ecology and Evolution*, *10*(12), 6030–6038. https://doi.org/10.1002/ece3.6344

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nguyen, T. V., Tilker, A., Nguyen, A., Hörig, L., Axtner, J., Schmidt, A., Le, M., Nguyen, A. H. Q., Rawson, B. M., Wilting, A., & Fickel, J. (2021). Using terrestrial leeches to assess the genetic diversity of an elusive species: The Annamite striped rabbit *Nesolagus timminsi*. *Environmental DNA*, *3*(4), 780–791.

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: Efficient ancient genome reconstruction. *Genome Biology*, *17*(1), 1–14.

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., Yamahara, K. M., & Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, *25*(2), 527–541.

R Core Team. (2021). *R: A language and environment for statistical computing. R foundation for statistical computing*. Available from: https://www.R-project.org/

Rodgers, T. W., Xu, C. C., Giacalone, J., Kapheim, K. M., Saltonstall, K., Vargas, M., Somervuo, P., McMillan, W. O., & Jansen, P. A. (2017). Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community. *Molecular Ecology Resources*, *17*(6), e133–e145.

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., MacNeil, I. A., Minor, C., Tiong, C. L., Gilman, M., Osburne, M. S., Clardy, J., Handelsman, J., & Goodman, R. M. (2000). Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology*, *66*(6), 2541–2547. https://doi.org/10.1128/aem.66.6.2541-2547.2000

Schnell, I. B., Bohmann, K., Schultze, S. E., Richter, S. R., Murray, D. C., Sinding, M.-H. S., Bass, D., Cadle, J. E., Campbell, M. J., Dolch, R., Edwards, D. P., Gray, T. N. E., Hansen, T., Hoa, A. N. Q., Noer, C. L., Heise-Pavlov, S., Pedersen, A. F. S., Ramamonjisoa, J. C., Siddall, M. E., … Gilbert, M. T. P. (2018). Debugging diversity – A pan-continental 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), 1–14.

Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R. D., Willerslev, E., Bertelsen, M. F., & 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.2012.02.058

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.

Seeber, P. A., McEwen, G. K., Löber, U., Förster, D. W., East, M. L., Melzheimer, J., & Greenwood, A. D. (2019). Terrestrial mammal surveillance using hybridization capture of environmental DNA from African waterholes. *Molecular Ecology Resources*, *19*(6), 1486–1496.

Sigsgaard, E. E., Jensen, M. R., Winkelmann, I. E., Møller, P. R., Hansen, M. M., & Thomsen, P. F. (2020). Population-level inferences from environmental DNA—Current status and future perspectives. *Evolutionary Applications*, *13*(2), 245–262. https://doi.org/10.1111/eva.12882

Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., Pedersen, M. W., Jaidah, M. A., Orlando, L., Willerslev, E., Møller, P. R., & Thomsen, P. F. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, *1*(1), 1–5.

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B., Roberts, R. G., Shunkov, M. V., de Lumley, H., Perrenoud, C., Gušić, I., … Meyer, M. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. *Science*, *356*(6338), 605–608. https://doi.org/10.1126/science.aam9695

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press.

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*(4), 564–577.

Taylor, P. G. (1996). Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Molecular Biology and Evolution*, *13*(1), 283–285.

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, *39*(11), 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3), 1–12.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36.

Zavala, E. I., Jacobs, Z., Vernot, B., Shunkov, M. V., Kozlikin, M. B., Derevianko, A. P., Essel, E., de Fillipo, C., Nagel, S., Richter, J., Romagné, F., Schmidt, A., Li, B., O'Gorman, K., Slon, V., Kelso, J., Pääbo, S., Roberts, R. G., & Meyer, M. (2021). Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova cave. *Nature*, *595*(7867), 399–403. https://doi.org/10.1038/s41586-021-03675-0

SUPPORTING INFORMATION

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

How to cite this article: Danabalan, R., Merkel, K., Bærholm Schnell, I., Arandjelovic, M., Boesch, C., Brazzola, G., Dieguez, P., Dupain, J., Kambale-Vyalengerera, M., Kühl, H. S., Hoffmann, C., Lapuente, J., Ngoc Thinh, V., Zimmermann, F., Leendertz, F. H., Gilbert, M. T. P., Roos, C., Mazzoni, C., Gogarten, J. F., & Calvignac-Spencer, S. (2023). Mammal mitogenomics from invertebrate-derived DNA. *Environmental DNA*, *00*, 1–12. https://doi.org/10.1002/edn3.436

26374943, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/edn3.436 by Ntnu Norwegian University Of Science & Technology, Wiley Online Library on [11/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License