MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 915–924 doi: 10.1111/mec.12183

FROM THE COVER

# Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity

SÉBASTIEN CALVIGNAC-SPENCER,* KEVIN MERKEL,* NADINE KUTZNER,* HJALMAR KÜHL,† CHRISTOPHE BOESCH,† PETER M. KAPPELER,‡ SONJA METZGER,*† GRIT SCHUBERT* and FABIAN H. LEENDERTZ*

*Emerging Zoonoses, Robert Koch Institute, Nordufer 20, 13353 Berlin, Germany, †Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 4103 Leipzig, Germany, ‡Behavioral Ecology & Sociobiology Unit, German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany

## Abstract

**Large-scale monitoring schemes are essential in assessing global mammalian biodiversity, and in this framework, leeches have recently been promoted as an indirect source of DNA from terrestrial mammal species. Carrion feeding flies are ubiquitous and can be expected to feed on many vertebrate carcasses. Hence, we tested whether fly-derived DNA analysis may also serve as a novel tool for mammalian diversity surveys. We screened DNA extracted from 201 carrion flies collected in tropical habitats of Côte d'Ivoire and Madagascar for mammal DNA using multiple PCR systems and retrieved DNA sequences from a diverse set of species (22 in Côte d'Ivoire, four in Madagascar) exploiting distinct forest strata and displaying a broad range of body sizes. Deep sequencing of amplicons generated from pools of flies performed equally well as individual sequencing approaches. We conclude that the analysis of fly-derived DNA can be implemented in a very rapid and cost-effective manner and will give a relatively unbiased picture of local mammal diversity. Carrion flies therefore represent an extraordinary and thus far unexploited resource of mammal DNA, which will probably prove useful for future inventories of wild mammal communities.**

*Keywords*: biodiversity, carrion fly, environmental DNA, mammals, noninvasive sampling

*Received 1 November 2012; revision received 12 November 2012; accepted 14 November 2012*

## Introduction

The global decline of biodiversity calls for vigorous conservation initiatives for which large-scale monitoring remains an essential tool (Butchart *et al.* 2010). Even for comparatively well-studied groups of animals such as mammals, future efforts to assess biodiversity may not only document temporal changes in recognized species' ranges, but are also predicted to unveil the existence of more than 300 new species in the next 20 years (Jones & Safi 2011).

Traditional monitoring schemes involve the repeated collection of direct observational data or animal traces such as faeces, footprints or nesting sites, and require the concerted mobilization of numerous experts (e.g. indigenous people with biodiversity knowledge, taxonomists and ecologists) over extended periods of time (Campbell *et al.* 2011). Over the last decades, the need for improved accuracy and higher cost and time efficiency triggered the development of a number of novel tools. For instance, remote sensing through camera trapping enabled standardized data acquisition on relatively large-bodied animals in multiple biodiversity hotspots, with minimum logistic effort (Ahumada *et al.* 2011). Noninvasive sampling of hair, urine and faeces, or even freshwater, soil and sediment (containing the so-called environmental DNA) is increasingly employed to PCR-amplify DNA fragments that allow for individual or species identification (Taberlet *et al.* 1999, 2012). Through such molecular approaches, the size and viability of numerous mammal populations were estimated (e.g. primates; Vigilant & Guschanski 2009), and ancient

Correspondence: Sébastien Calvignac-Spencer, Fax: +4930187542181; E-mail: calvignacs@rki.de, Fabian Leendertz, Fax: +4930187542181; E-mail: leendertzf@rki.de

mammal communities described (Boessenkool *et al.* 2011). Recently, Schnell *et al.* (2012) took a step forward by depicting local mammal diversity using terrestrial leech DNA. Although molecular analyses of bloodmeals from insect vectors of human pathogens (i.e. mosquitoes, tsetse flies) have been used to recover vertebrate DNA before (Kent 2009), this study (Schnell *et al.* 2012) was the first to depict invertebrate-derived DNA as a tool for biodiversity assessment.

Blow and flesh flies (families Calliphoridae and Sarcophagidae; here collectively referred to as 'carrion flies') feed on carrion, open wounds of living animals and/or faecal matter and generally oviposit on the first two substrates (Norris 1965). Carrion flies are distributed worldwide in all kinds of habitats and are easy to collect (Norris 1965): they might represent another way of sampling DNA from wild mammals. Mammal monitoring through carrion flies may even overcome some of the impediments associated with the use of haematophagous insects and leeches (i.e. habitat restriction of tsetse flies and leeches, host preferences of many mosquito and tsetse fly species; Junglen *et al.* 2009; Lyimo & Ferguson 2009), and become an effective complement to traditional and other genetic monitoring methods.

In this article, we aim at exploring the potential of carrion fly DNA analysis in assessing mammalian biodiversity. We therefore analysed blow and flesh flies collected at two distinct tropical ecosystems harbouring a broad diversity of mammal taxa: the Taï National Park, Côte d'Ivoire, a moist forest, and Kirindy Forest, Madagascar, a dry deciduous forest. We show that carrion flies frequently contain mammalian DNA from a broad variety of taxa.

## Materials and methods

### *Field sites and sampling methods' description*

Carrion flies were captured at two tropical field sites. The Taï National Park, Côte d'Ivoire (thereafter Taï), is the largest remnant block of the Upper Guinean rainforest and a typical example of tropical moist forest. It houses a considerable diversity of mammals. Among these, nine nonhuman primate species can be found, many of which have been studied within the framework of a long-term research programme (Boesch & Boesch-Achermann 2001; McGraw *et al.* 2007). Kirindy Forest, western Madagascar (thereafter Kirindy), is a tropical dry deciduous forest, hosting a number of mammal species. Eight lemur species occur there, which are the focus of a long-term research programme as well (Kappeler & Fichtel 2012). Flies were either caught under mosquito nets protecting dissection areas or at random forest locations using mosquito nets thrown over a bowl containing a piece of meat or bottle-based traps in which a piece of meat had been placed (Fig. S1, Supporting information). In Taï, all flies were placed in tubes and stored in liquid nitrogen containers immediately after collection. In Kirindy, flies were placed in tubes filled with 96% ethanol and later stored at −20 °C.

### *DNA extraction*

Entire flies were placed in FastPrep® lysing matrix tubes (MP Biomedicals, Illkirch, France) filled with 250 μL phosphate buffer saline and cut several times with sterilized scissors. Tubes were then placed in a FastPrep® system and agitated three times for five-seconds. For individual fly analyses, nucleic acids were extracted from 100 μL fly mixture using the GeneMATRIX Stool DNA Purification Kit (Roboklon, Berlin, Germany) according to manufacturer's instructions. Except for 30 Taï flies, which were extracted by a service provider (GenExpress, Berlin, Germany) to provide an independent proof of feasibility, individual extractions were carried out at the Robert Koch Institute. For the fly pool analysis, 10 μL of each selected mixture was added to the pool and the resulting 300 μL were extracted at once using the same kit as for individual extractions. To assess the efficiency of this protocol in removing PCR inhibitors, the first 40 DNA extracts were used in a quantitative inhibition test. In the latter, quantitative PCR assays aimed at amplifying 1 million copies of a short chicken DNA fragment (a portion of the CXCR4 gene) were led in the presence of 5 μL fly DNA extract. The expected number of initial chicken templates was always derived from the experiments, suggesting fly DNA extracts did not exhibit significant inhibitory effect (data not shown).

Nucleic acids were also extracted from 34 tissue samples which had been obtained from Taï specimens, either from carcasses or darted specimens (bats, carnivores, artiodactyls, primates and rodents) using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) according to manufacturer's instructions.

DNA concentrations were measured for all extracts using a Nanodrop device (Thermo Scientific, Waltham, MA, USA).

### *Mammal-targeting PCR assays for carrion fly screening*

Three assays targeting taxa of different breadth (chimpanzee, mammals and vertebrates) were used to screen fly DNA extracts. So as to decrease the sensitivity to classical laboratory contaminants such as human or porcine DNA, PCR were performed in the presence of

appropriate blocking primers (Vestheim & Jarman 2008). All assays were optimized using a quantitative PCR approach similar to (Boessenkool *et al.* 2011), which we applied to a panel of 29 DNA extracts obtained from four humans, four other hominoids, nine cercopithecids, one lorisid, five artiodactyls, two rodents, two bats and one pangolin.

The chimpanzee-specific assay targeted a 114 base pair (bp) long fragment of the mitochondrial 16S gene (16S), using the following primers: 16Sape_f 5′-ATTCTCCTCCGCATAAGCCT-3′, 16Sape_r 5′-GCCTGTGTTGGGTTAACAG-3′ and 16ape_blkhum 5′-TCCGCATAAGCCTGCGTCAGATTAAAACAC–spacerC3. The pan-mammal assay targeted a *c.* 130 bp fragment of 16S. It was derived from the literature (Boessenkool *et al.* 2011) but was supplemented with a pig blocking primer: 16Smam1 5′-CGGTTGGGGTGACCTCGGA-3′, 16Smam2 5′-GCTGTTATCCCTAGGGTAACT-3′, 16Smam_blkhum3 5′-CGGTTGGGGCGACCTCGGAGCAGAACCC—spacerC3 and 16Smam_blkpig 5′-CGGTTGGGGTGACCTCGGAGTACAAAAAAC—spacerC3. PCR were also performed using a variant of this system, to determine success rates for a longer fragment. In that case, 16Smam2 was substituted with 16Smam4 5′-AGATAGAAACCGACCTGGAT-3′, resulting in an amplicon length of about 300 bp. Finally, the pan-vertebrate assay targeted a *c.* 140-bp fragment of 12S. It was also derived from the literature (Riaz *et al.* 2011; Shehzad *et al.* 2012), but supplemented with human and pig blocking primers: 12S_V5f 5′-ACTGGGATTAGATACCCC-3′, 12S_V5r 5′-TAGAACAGGCTCCTCTAG-3′, 12S_V5_blkhum 5′-TACCCCACTATGCTTAGCCCTAAACCTCAACAGTTAAATC–spacerC3 and 12S_V5_blkpig 5′-TACCCCACTATGCCTAGCCCTAAACCCAAATAGTTACAT–spacerC3.

PCRs were performed in a total volume of 25 μL and were started with 200 ng DNA (DNA concentration >40 ng/μL) or 5 μL DNA extract (DNA concentration <40 ng/μL). PCR mixes were set-up so as to allow reducing as much as possible carry over contamination with PCR products. They contained 0.2 μM of each primer, 1 μM of each blocking primer, 200 μM dNTP (with dUTP replacing dTTP), 0.3 U AmpErase® uracil *N*-glycosylase (Invitrogen, Carlsbad, CA, USA), 4 mM $MgCl_2$, 1X PCR buffer and 1.25 U Platinum® *Taq* polymerase (Invitrogen). Cycling conditions were: 7 min at 45 °C, 10 min at 95 °C, 42 (short and long pan-mammal and ape-specific 16S) or 47 (pan-vertebrate 12S) cycles of 30 s at 95 °C, 30 s at 64 °C (short and long pan-mammal 16S), 57 °C (ape-specific 16S) or 60 °C (pan-vertebrate 12S), 1 min at 72 °C and 10 min at 72 °C. PCR products were visualized on an agarose gel and cleaned up using ExoSAP-IT® (Affymetrix, Santa Clara, CA, USA) before being sequenced in both directions using the Sanger method. All chromatograms were examined using Geneious Pro v5.4.3 (http://www.geneious.com/). A detailed depiction of all experiments performed on all flies used in this study is available as Table S1 (Supporting information).

To provide a further control for contamination, the 30 DNA extracts generated externally were also PCR-amplified by the same service provider (GenExpress). All other PCRs were performed at the Robert Koch Institute.

### *Next-generation sequencing of carrion fly DNA amplified with pan-mammal PCR systems*

The above-mentioned short and long 16S assays were used to generate amplicons for deep sequencing. For this, 200 ng DNA from the 30-fly-pool DNA extract was used to seed reactions. Two different protocols were tested. Two amplicons (one with each primer pair) were generated according to a modified protocol in which normal primers were replaced with fusion primers and normal dNTP were used. Two additional amplicons were generated using the normal screening protocol (i.e. using normal primers and dUTP-containing dNTP). Those were diluted 50 times and used as template for a further five cycle-long PCR using fusion primers and normal dNTP. This last strategy was implemented as it has recently been shown to limit the biases introduced by fusion primer-based amplification (Berry *et al.* 2011). The four PCR products were then cleaned using MinElute columns (Qiagen) and their DNA concentration assessed using a Qubit® fluorometer (Invitrogen). Sequencing of individual amplicons was performed on one-eighth of a pico titre plate using a GS FLX platform, according to manufacturer's instructions (Roche Applied Sciences, Penzberg, Germany). Reads were filtered using the PRINSEQ v0.19.2 online interface (Schmieder & Edwards 2011). Reads were discarded if <84 and >109 bp (excluding primers) for the short fragment or <220 and >72 bp (excluding primers) for the long fragment and/or if containing one bp with a quality score below ten or exhibiting an average quality score below 15. After filtering, a minimum of 16 470 reads per PCR product was retained for further analysis. Reads were finally clustered using the online tool cd-hit-est (Li & Godzik 2006; Huang *et al.* 2010) applying two successive similarity thresholds set-up at 97% and 94%. No noticeable difference between PCR protocols could be identified, but the number of PCR products analysed (2 and 2) was obviously way too low to detect anything else than major shifts. For each PCR product, a FASTQ file comprising all raw reads is archived in DRYAD under doi: 10.5061/dryad.57vg4.

### *Taxonomic assignment of mammal sequences*

Taxonomic assignment was performed using BLAST (Altschul *et al.* 1990). Sequences were compared with NCBI's nonredundant database through an online BLAST search and, for sequences generated with the pan-mammal 16S assays, to our own collection of 16S sequences from Taï mammals through a local BLAST search.

Even taking into account the sequences determined for this study, very few sequences have been determined from mammals living in Taï. Actually, <60% of the 232 mammal species reported to be present in the entire Côte d'Ivoire have been examined for 16S diversity (as deduced from a search in NCBI's Nucleotide database of all 232 species names). In most cases, these sequences were not even determined from individuals from Côte d'Ivoire. It was therefore impossible to apply stringent identity level thresholds for species or genus identification, as sometimes done for eDNA analyses (e.g. 100% threshold for species identification in Thomsen *et al.* (2012)). Most species identifications (86%) were performed based on 16S sequences. Among those, 76% exhibited identity levels $\geq$99% (84% considering identity levels $\geq$98%). As a comparison, chimpanzees and bonobos (*Pan paniscus*), which are thought to have diverged from each other 2 million years ago (Bjork *et al.* 2011), exhibit <97% identity for the same region. In these cases where lower identity levels were observed (typically between 94% and 97%), identification was made on a case-per-case basis combining BLAST information (e.g. second hit exhibiting a marked drop in *e*-value) and available biological information (e.g. phylogenetic relationships or reported presence in the forest). As a test of the validity of our species assignment, we further applied a clustering approach to short 16S sequences using cd-hit-est and a threshold of 95%. This equated all operational taxonomic units with individual species found through BLAST analysis, except for *Myonycteris torquata* and *Hypsignathus monstrosus* (which are recently diverged fruit bat species). We are therefore confident that most assignments are correct and, at the very least, that the extent of mammal diversity presented in this article reflects the mammal diversity captured by our sample of flies. All sequences are available as Appendix S1 (Supporting information).

### *Species accumulation curve estimation*

Because always several flies were captured in a trap, the detection of species cannot be considered as independent events. We therefore did not fit curves to species detected in individual flies. Instead, we used the 15 'batches' that contained flies captured in one trap. Batch size ranged from 2 to 35 flies and number of detected species ranged from 0 to 10 (the input file used in further analyses is provided as Table S2, Supporting information). We first tried to fit an asymptotic curve (Eggert *et al.* 2002) to the cumulative number of species $S_i$ detected in $n$ flies

$$\sum_{i=1}^{n} S_i = a \times (1 - \exp(b \times x_n))$$

where $a$ and $b$ are the parameters to be estimated and $x$ is the cumulative number of flies. We did this by resampling the data set 1000 times without replacement to derive mean values and confidence limits. However, for many resampled data sets, parameters could not be estimated reliably because the fitting algorithm did not converge. This indicated that the asymptotic trend in the resampled data sets was not consistent enough and thus the data set was too small to derive such statistics. We therefore decided to only fit the curve to the mean values of the bootstrapped data ($y = 29.38 \times (1 - \exp(-0.00899 \times$ number of flies))). All analyses were performed in R (R Development Core Team 2008).

### *Carrion fly family/species determination*

PCR were prepared as above mentioned using a primer pair widely used to target a *c.* 700-bp COI fragment in invertebrates (Folmer *et al.* 1994). All reactions were seeded with 50 ng DNA. Cycling conditions were: 5 min at 95 °C, 35 cycles of 30 s at 95 °C, 30 s at 50 °C, 1 min at 72 °C and 10 min at 72 °C. PCR products were visualized and sequenced as above mentioned. For assignment to Calliphoridae or Sarcophagidae, we used BLAST (Altschul *et al.* 1990) and a threshold of 94% identity to published sequences from species belonging to one or the other family (intersubfamilies pairwise distance over this fragment ranged between 7% and 15%). Applying this criterion, 95% sequences could be assigned confidently at the family level. A subset of flies could be determined down to the species level, using a threshold of 99% identity to published sequences.

## Results

### *Amplifying mammalian DNA from carrion flies with known exposure to carrion*

We captured 86 flies (under dissection nets) during necropsies of carcasses from five different mammalian species (Table 1). Using ape-specific, pan-mammal and pan-vertebrate PCR systems targeting short (<170 bp) fragments of the mitochondrial 16S and 12S genes, 60/86 flies (70%) were shown to contain amplifiable DNA (Table 1). The corresponding sequences were assignable to the expected species for 54/86 flies (63%), with

**Table 1** Analysis of flies with known exposure to carrion

| Dissected specimen | No. of analysed flies | No. of dissected mammal positive (%) | No. of all mammal positive (%) | Other mammals* (no. of flies) |
|---|---|---|---|---|
| African palm civet *Nandinia binotata* | 10 | 7 (70) | 7 (70) | — |
| Shrew (unknown species) *Crocidura* sp. | 15 | 7 (47) | 11 (73) | Diana monkey (2) *Cercopithecus diana*<br>Lesser spot-nosed monkey (1) *Cercopithecus petaurista*<br>Olive colobus† (1) *Procolobus verus*<br>Sooty mangabey (1) *Cercocebus atys*<br>Swamp rat† (1) *Malacomys* sp.<br>Unknown murinae 3† (1)<br>Western chimpanzee† (1) *Pan troglodytes verus* |
| Sooty mangabey *Cercocebus atys* | 35 | 31 (89) | 31 (89) | Duiker (1) *Cephalophus* sp. (unknown species 2‡) |
| Western chimpanzee *Pan troglodytes verus* | 11 | 4 (36) | 6 (54) | Diana monkey† (1) *Cercopithecus diana*<br>Western red colobus (1) *Piliocolobus badius badius* |
| Western red colobus *Piliocolobus badius badius* | 15 | 5 (33) | 6 (40) | Duiker (1) *Cephalophus* sp. (unknown species 1‡) |

*Where species identification was not possible, higher taxa to which DNA sequences could be related are shown. Number of flies found positive for a given taxon is given in brackets.
†Co-detection with another mammal species.
‡These duiker DNA sequences could not be attributed to species but the observed differences (4/107 bp; 12S fragment) were considered incompatible with them belonging to different individuals of a single species.
See also Table S3 (Supporting information).

success rates ranging between 33% and 89% per net (Table 1). In addition to the expected species, DNA of other mammals was detected in 11 flies (13%; Table 1).

### *Assessing mammalian biodiversity using DNA from randomly collected carrion flies*

We then analysed 115 flies without known exposure to carrion, captured at random locations in Taï ($n = 75$) and Kirindy ($n = 40$). Twenty-nine of 75 flies from Taï (39%) were found positive for mammalian DNA (Table 2). Mitochondrial 16S and 12S sequences from Taï identified 16 taxa belonging to the orders Artiodactyla, Chiroptera, Eulipotyphla, Primates and Rodentia (Table 2). Twelve of them were identified to the species level, all of which were known to occur in Taï. The four remaining taxa were two murinae, one hystricid (porcupine) and one crocidurinae (white-toothed shrew), of which several species occur in Taï (only two hystricids, many more murinae and shrews). A species accumulation curve derived from all flies captured in Taï (including flies captured under nets; $n = 161$) evidenced that increasing sampling size would have resulted in detecting additional mammal species (the small size of the sampling prevented any further prediction; Fig. 1). Seventeen of 40 flies from Kirindy (42%) were found positive for mammalian DNA. Four mammal taxa belonging to the orders Afrosoricidae, Carnivora and Primates were identified; three of them at the species level (Table 2). All were known Malagasy endemics. Two bird species, one in Taï and one in Kirindy, were also detected with the pan-vertebrate PCR system targeting 12S (Table 2). One amphibian species was detected in Taï using the pan-mammal PCR system targeting 16S (Table 2). Seven flies (five in Taï and two in Kirindy) contained DNA from multiple mammal/vertebrate taxa.

**Table 2** Taxa identified from randomly collected flies

| Field site | Identified taxon | |
|---|---|---|
| | Class/mammal order | Genus/species |
| Taï National Park Côte d'Ivoire | Artiodactyla | Jentink's duiker (1) *Cephalophus jentinki* |
| | | Maxwell's duiker (1) *Philantomba maxwellii* |
| | | Pygmy hippo (1) *Hexaprotodon liberiensis* |
| | | Water chevrotain (1) *Hyemoschus aquaticus* |
| | Chiroptera | Little collared fruit bat (3) *Myonycteris torquata* |
| | | Hammer-headed bat (3) *Hypsignathus monstrosus* |
| | Eulipotyphla | Shrew (1) *Crocidura sp.* |
| | Primates | Sooty mangabey (5) *Cercocebus atys* |
| | | Diana monkey (4) *Cercopithecus diana* |
| | | Campbell's monkey (3) *Cercopithecus campbelli* |
| | | Greater spot-nosed monkey (2) *Cercopithecus nictitans* |
| | | Black and white colobus (1) *Colobus polykomos* |
| | | Western red colobus (1) *Piliocolobus badius badius* |
| | Rodentia | Porcupine (4) *Atherurus africanus* or *Hystrix cristata* |
| | | Unknown murinae 1 (1) |
| | | Unknown murinae 2 (1) |
| | Amphibia | Screeching frog/ squeaker† (1) *Arthroleptis* sp. |
| | Aves | Hornbill‡ (1) *Bycanistes sp.* |
| Kirindy Reserve Madagascar | Afrosoricida | Tailless tenrec (6) *Tenrec ecaudatus* |
| | Carnivora | Euplerid (7) C*ryptoprocta ferox* or *Mungotictis decemlineata* |
| | Primates | Dwarf lemur (2) *Cheirogaleus medius* |
| | | Grey mouse lemur (1) *Microcebus murinus* |
| | Aves | Water rail‡ (3) *Rallus aquaticus* |

The number of flies found positive for a given taxon is given in brackets. Within sites and mammalian order/classes, taxa are ordered in function of their detection frequency, except for those detected once, which are ordered alphabetically.
†Amphibian species.
‡Bird species.
See also Table S3 (Supporting information).

### *Targeting longer DNA fragments*

We retested 150 flies using a modified pan-mammal 16S assay, with an amplicon length of 300 bp. Success rates did not change dramatically when compared with results obtained with short fragments: 52/59 flies captured under dissection nets (88%) and 23/91 randomly collected flies (25%) were positive for mammalian DNA (vs. 63% and 40% with short fragments).

### *Assessing pooling as a high-throughput strategy*

We pooled and extracted at once remnants of 30 flies already analysed individually. Two short and two long 16S amplicons were generated from the extract using fusion primers and directly sequenced on a 454 GS FLX. This allowed for the identification of, respectively, 8/12 (first short PCR product), 10/12 (second), 8/9 (first long PCR product) and 8/9 (second) of the taxa identified through individual analyses and Sanger's sequencing (Table 3). A single taxon was detected only through individual sequencing whereas four others were only detected through the pooled fly NGS approach (Table 3).

### *Identifying suitable carrion fly species*

Finally, to determine the fly species captured and analysed, we generated COI sequences from 181 flies. In both Taï and Kirindy, blow flies (family Calliphoridae) were abundant, accounting for over 90% of captured individuals and being found in all traps. A minimum of three species were identified in Taï (blow fly species identification was not possible for Kirindy flies): *Chrysomya inclinata* (60 individuals captured in 6 traps), *Chrysomya putoria* (52 in 14 traps) and *Chrysomya albiceps* (13 in six traps). Flesh flies (family Sarcophagidae) accounted for the remaining 10% of individuals and were only found in 9/21 traps. A minimum of one species was identified in Taï (*Sarcophaga africana*, three in two traps) and in Kirindy (*Sarcophaga tibialis*, two in two traps). In Taï, 49% of randomly collected blow flies comprised amplifiable mammal DNA (27/55 caught in 8/11 blow fly containing traps), while no flesh fly was positive for mammal DNA (0/14 caught in five traps). Mammal DNA recovery appeared to be less frequent from *C. putoria* (39%; $n = 38$) than from the two other blow fly species (*C. inclinata* + *C. albiceps*; 73%; $n = 15$). In Kirindy, 40% blow flies (15/37; 2/4 traps) and 2/2 flesh flies (2/2 traps) were positive. Finally, fly families or species did not seem to exhibit clear host preference (small sample sizes precluded any statistical assessment).

## Discussion

Here, we establish that carrion flies, and more particularly blow flies, can be used as indicators of mammalian

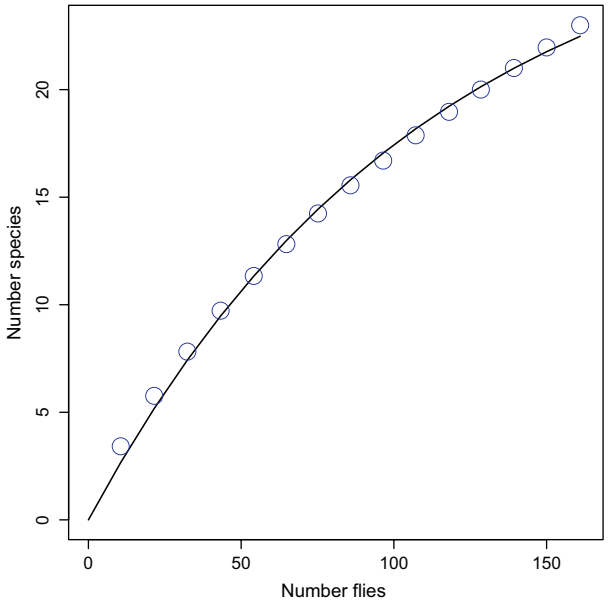


**Fig. 1** Mammal species accumulation curve for flies captured in Taï.

biodiversity. For this, we first demonstrate that under controlled conditions (dissection nets), mammalian DNA is retrievable from flies that have recently been in contact with a mammal carcass. We then show that a substantial proportion of carrion flies (40%) caught randomly in two tropical forests contains amplifiable fragments of mammalian DNA assignable to a broad range of mammals, with some flies harbouring mammal DNA from multiple species (thus from multiple meals). More precisely, the random collection of 75 and 40 flies at Taï and Kirindy allows for the recovery of DNA from 16 and four mammal species, respectively. For Kirindy, this corresponds to 13% of the known local mammalian community (31 species). Mammal biodiversity is not yet fully characterized in Taï, so the fraction sampled here cannot be estimated. However, striking outcomes were the sampling of six of the nine species of the local primate community (Boesch & Boesch-Achermann 2001; McGraw *et al.* 2007) and the detection of Jentink's duiker (*Cephalophus jentinki*), a very rare and endangered species whose entire population is estimated to be <3500 individuals (IUCN SSC Antelope Specialist Group 2008). A species' accumulation curve analysis from Taï flies further suggests that continued sampling would have resulted in the recovery of additional mammal species' DNA. This performance compares well with that of other invertebrate-based studies. Analysing 25 leeches collected in Vietnam, Schnell *et al.* (2012) for instance detected six mammal species, including two rare species only recently described. Taken together, our results indicate that blow fly-derived DNA, like haematophagous insects or leech-derived DNA, can capture local mammal biodiversity.

In a comparative perspective, an important question is that of invertebrate feeding strategies, which might introduce biases in the mammal spectrum detected. This also applies to carrion flies species, which have sometimes been reported to exhibit host preferences (Norris 1965). In spite of a relatively moderate sample size, our data do not indicate a strong sampling bias related to either average body size or forest strata use of the host species. Both small (<0.1 kg) and large (up to 235 kg) as well as terrestrial and arboreal mammals were identified (Table S3, Supporting information). This might contrast with other invertebrates presenting well-known host preferences (e.g. mosquitoes; Lyimo & Ferguson 2009). It is also worth noting that tsetse fly-based studies have not yet detected small-bodied mammals (Konnai *et al.* 2008; Muturi *et al.* 2011) and that the presence of only terrestrial mammals was revealed through leeches (Schnell *et al.* 2012). Although further work is needed to provide a proper comparison, it may be that carrion flies feed on a larger spectrum of mammals than blood-sucking invertebrates.

In the end, it is, however, clear that blow flies, haematophagous insects and leeches should be considered as parts of the same vertebrate sampler toolbox, from which one tool will occasionally supersede the others, depending on the ecosystem investigated or the conceived application. For example, while leeches have a much more restricted distribution than blow flies, it can be expected that they will contain more mammal DNA, for longer periods of time than blow flies, which points at them as ideal candidates for DNA-demanding applications (Schnell *et al.* 2012).

Carrion fly-derived DNA might be used to address a number of questions related to mammal biodiversity. We show here an example of high-throughput mammal biodiversity assessment using NGS of fly pool DNA. Our experiment provides a picture of mammalian biodiversity that matches closely the one obtained using individual-based methods. This suggests that no strong bias arose from the competition between different species' DNA, as sometimes observed in bacterial metagenomics (Hong *et al.* 2009). Hence, NGS from fly pools represents a very effective sequencing strategy in terms of laboratory costs and sample usage. The fact that amphibian and bird sequences were also detected suggests that similar strategies may be developed to assess the biodiversity of other animal groups. Non-mammal taxa, however, only formed a small fraction of the vertebrate sequences found here, even when a pan-vertebrate system was used, so this aspect warrants more investigation.

**Table 3** Taxon identification from 30 unpooled (Sanger's sequencing) and pooled (454 sequencing) flies

| | Short 16S | | | Long 16S | | |
|---|---|---|---|---|---|---|
| | Sanger's | 454 1st product | 454 2nd product | Sanger's | 454 1st product | 454 2nd product |
| Identified taxon | | | | | | |
| Campbell's monkey *Cercopithecus campbelli* | X | X | X | | X | X |
| Diana monkey *Cercopithecus diana* | X | X | X | X | X | X |
| Greater spot-nosed monkey *Cercopithecus nictitans* | X | X | X | X | X | X |
| Hammer-headed bat *Hypsignathus monstrosus* | X | X | X | X | X | X |
| Lesser spot-nosed monkey *Cercopithecus petaurista* | X | | X | X | | X |
| Olive colobus *Procolobus verus* | X | X | | X | X | X |
| Porcupine *Atherurus africanus* or *Hystrix cristata* | X | X | X | X | X | X |
| Screeching frog/ squeaker *Arthroleptis* sp. | X | | X | X | X | X |
| Sooty mangabey *Cercocebus atys* | X | X | X | X | X | X |
| Unknown murinae 1 | X | | X | | | |
| Unknown murinae 2 | X | X | X | X | X | |
| Unknown murinae 3 | X | | | | | |
| Black and white colobus[454] *Colobus polykomos* | | X | X | | | X |
| Red river hog[454] *Potamochoerus porcus* | | | X | | | |
| Unknown anomaluridae 1[454] | | X | | | | |
| Unknown muroidae 1[454] | | X | | | | |
| Species count | 12 | 11 | 12 | 9 | 9 | 10 |
| Overlap with Sanger's | na | 8 | 10 | na | 8 | 8 |

The first and second PCR products were obtained using the same PCR systems but different protocols (see Materials and methods). [454]Sequences of this taxon were only found in PCR products analysed with NGS.

Targeted species monitoring might also benefit from the use of blow fly-derived DNA. For example, longitudinal fly sampling may be used to monitor mortality dynamics like those caused by massive die-offs of infectious origin, like the gorilla (*Gorilla gorilla*) die-off caused by Ebola in 2002–2003 (Bermejo *et al.* 2006). For such studies to be accurate, further information on the origin of mammal DNA transported by flies will, however, be required. That is, while mammal DNA from the gut of flies might represent actual feeding on cadavers, mammal DNA amplifiable from the exoskeleton might only reflect the landing on an animal.

Noninvasively collected samples have been widely used to study the size and dynamics of various mammal populations through microsatellite analyses (e.g. Douadi *et al.* 2007). Microsatellites usually vary in size between 100 and 300 bp (Selkoe & Toonen 2006), matching the maximum fragment length of 300 bp recovered here. Hence, it is conceivable that carrion fly analysis might be extended to the identification of individuals and a more in-depth investigation of single mammal populations.

Taken together, carrion fly DNA analysis might offer a broad range of potential applications linked with the

assessment and monitoring of mammalian biodiversity. We are therefore convinced that this new tool will quickly come into use and contribute to the advancement of the DNA-based evaluation of the diversity of wild mammals and other organisms.

## Acknowledgements

We thank the Ivorian authorities for long-term support, especially the Ministry of the Environment and Forests as well as the Ministry of Research, the directorship of the Taï National Park, and the Swiss Research Center in Abidjan. We also thank the Ministry of the Environment of Madagascar, the Département Biologie Animale of the University of Antananarivo and the CNFEREF/Morondava for their support and authorization of this study. We also thank L. Vigilant and T. R. Gillespie who provided interesting comments and proofread this manuscript; J. Tesch, W. Dabrowsky and A. Radonic from the sequencing laboratory of the Robert Koch Institute for their excellent practical help; all members of the Emerging Zoonoses group for stimulating discussions.

## References

Ahumada JA, Silva CE, Gajapersad K *et al.* (2011) Community structure and diversity of tropical forest mammals: data from a global camera trap network. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2703–2711.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Bermejo M, Rodriguez-Teijeiro JD, Illera G, Barroso A, Vilá C, Walsh PD (2006) Ebola outbreak killed 5000 gorillas. *Science*, **314**, 1564.

Berry D, Ben Mahfoudh K, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, **77**, 7846–7849.

Bjork A, Liu W, Wertheim JO, Hahn BH, Worobey M (2011) Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. *Molecular Biology and Evolution*, **28**, 615–623.

Boesch C, Boesch-Achermann H (2001) *The Chimpanzees of the Taï Forest*. Oxford University Press, New York.

Boessenkool S, Epp LS, Haile J *et al.* (2011) Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology*, **21**, 1806–1815.

Butchart SH, Walpole M, Collen B *et al.* (2010) Global biodiversity: indicators of recent declines. *Science*, **328**, 1164–1168.

Campbell G, Kuehl H, Diarrassouba A, N'Goran PK, Boesch C (2011) Long-term research sites as refugia for threatened and over-harvested species. *Biology Letters*, **7**, 723–726.

Douadi MI, Gatti S, Levrero F *et al.* (2007) Sex-biased dispersal in western lowland gorillas (*Gorilla gorilla gorilla*). *Molecular Ecology*, **16**, 2247–2259.

Eggert LS, Rasner CA, Woodruff DS (2002) The evolution and phylogeography of the African elephant inferred from mitochondrial DNA sequence and nuclear microsatellite markers. *Proceedings of the Royal Society B: Biological Sciences*, **269**, 1993–2006.

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**, 294–299.

Hong S, Bunge J, Leslin C, Jeon S, Epstein SS (2009) Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME Journal*, **3**, 1365–1373.

Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, **26**, 680–682.

IUCN SSC Antelope Specialist Group (2008) *Cephalophus jentinki*. In: IUCN 2012.IUCN Red List of Threatened Species. Version 2012.2. www.iucnredlist.org, Downloaded on 12 December 2012.

Jones KE, Safi K (2011) Ecology and evolution of mammalian biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2451–2461.

Junglen S, Kurth A, Kuehl H *et al.* (2009) Examining landscape factors influencing relative distribution of mosquito genera and frequency of virus infection. *EcoHealth*, **6**, 239–249.

Kappeler PM, Fichtel C (2012) A 15-year perspective on the social organization and life history of Sifaka in Kirindy Forest. In: *Long-Term Field Studies of Primates* (eds Kappeler PM, Watts DP), pp. 101–124. Springer, Berlin.

Kent RJ (2009) Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. *Molecular Ecology Resources*, **9**, 4–18.

Konnai S, Mekata H, Odbileg R *et al.* (2008) Detection of *Trypanosoma brucei* in field-captured tsetse flies and identification of host species fed on by the infected flies. *Vector-Borne and Zoonotic Diseases*, **8**, 565–573.

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

Lyimo IN, Ferguson HM (2009) Ecological and evolutionary determinants of host species choice in mosquito vectors. *Trends in Parasitology*, **25**, 189–196.

McGraw S, Zuberbühler K, Noe R (2007) Monkeys of the Tai Forest. Cambridge University Press, Cambridge, UK.

Muturi CN, Ouma JO, Malele II *et al.* (2011) Tracking the feeding patterns of tsetse flies (*Glossina* genus) by analysis of bloodmeals using mitochondrial cytochromes genes. *PLoS ONE*, **6**, e17284.

Norris KR (1965) The bionomics of blow flies. *Annual Review of Entomology*, **10**, 47–68.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing* (ed. Computing RFfS). Vienna, Austria.

Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, **39**, e145.

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**, 863–864.

Schnell IB, Thomsen PF, Wilkinson N *et al.* (2012) Screening mammal biodiversity using DNA from leeches. *Current Biology*, **22**, R262–R263.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–629.

Shehzad W, Riaz T, Nawaz MA *et al.* (2012) Carnivore diet analysis based on next-generation sequencing: application to

the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, **21**, 1951–1965.
Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution*, **14**, 323–327.
Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA. *Molecular Ecology*, **21**, 1789–1793.
Thomsen PF, Kielgast J, Iversen LL *et al.* (2012) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, **21**, 2565–2573.
Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, **5**, 12.
Vigilant L, Guschanski K (2009) Using genetics to understand the dynamics of wild primate populations. *Primates*, **50**, 105–120.

S.C.S. and F.H.L. designed and supervised the experiments; S.M., C.B. and P.M.K. contributed samples; K. M., N.K. and G.S. performed the experiments; S.C.S. and H.K. analysed the results; S.C.S., G.S. and F.H.L. drafted the article; all authors revised the article.

## Data accessibility

Sanger's sequences generated in this study are all provided as Appendix S1 (Supporting information). Raw data file of NGS as well as Appendix S1 (Supporting information) have been deposited to DRYAD and are accessible under doi: 10.5061/dryad.57vg4.

## Supporting information

Additional supporting information may be found in the online version of this article.

**Figure S1** Fly traps used for this study.

**Table S1** Complete summary of results.

**Table S2** Input table used for analyses with R.

**Table S3** Complete list of taxa identified in this study.

**Appendix S1** Alignment comprising all sequences generated for this study.