

SCIENTIFIC REPORTS



Assessing the feasibility of fly based surveillance of wildlife infectious diseases

Constanze Hoffmann¹, Melanie Stockhausen^{1,2}, Kevin Merkel¹, Sébastien Calvignac-Spencer¹ & Fabian H. Leendertz¹

Received: 08 August 2016
Accepted: 02 November 2016
Published: 30 November 2016

Monitoring wildlife infectious agents requires acquiring samples suitable for analyses, which is often logistically demanding. A possible alternative to invasive or non-invasive sampling of wild-living vertebrates is the use of vertebrate material contained in invertebrates feeding on them, their feces, or their remains. Carrion flies have been shown to contain vertebrate DNA; here we investigate whether they might also be suitable for wildlife pathogen detection. We collected 498 flies in Taï National Park, Côte d'Ivoire, a tropical rainforest and examined them for adenoviruses (family *Adenoviridae*), whose DNA is frequently shed in feces of local mammals. Adenoviral DNA was detected in 6/142 mammal-positive flies. Phylogenetic analyses revealed that five of these sequences were closely related to sequences obtained from local non-human primates, while the sixth sequence was closely related to a murine adenovirus. Next-generation sequencing-based DNA-profiling of the meals of the respective flies identified putative hosts that were a good fit to those suggested by adenoviral sequence affinities. We conclude that, while characterizing the genetic diversity of wildlife infectious agents through fly-based monitoring may not be cost-efficient, this method could probably be used to detect the genetic material of wildlife infectious agents causing wildlife mass mortality in pristine areas.

The recent epidemic of Ebola virus in West Africa has again shown the relevance of emerging infectious diseases (EID) for global public health and economies^{1,2}. Like the Ebola virus, most EIDs are of zoonotic origin and involve wildlife reservoir hosts³. Emergence is likely facilitated by increased contact of humans and domestic animals with wildlife combined with insufficient access to health care⁴. Therefore, it is most intense in resource-poor tropical regions, such as those in Central and West Africa where the Ebola virus repeatedly emerged^{3,5}. To prevent disease emergence, monitoring of wildlife infectious agents in these hotspot regions is essential. Some prior knowledge on these microorganisms' genetic diversity may accelerate the identification of early-stage spillover events and help mitigate their outcome.

Surveillance of wildlife infectious agents faces many hurdles. Poor infrastructure and limited resources in remote areas hinder sample collection, storage and transport in such a way that setting up the logistic framework necessary for acquiring wildlife samples is often challenging. This is even more complicated when endangered species are to be monitored. For instance, non-human primate species, which are promising candidates for detecting potential zoonotic agents⁶, are generally strictly protected.

This means that only non-invasive sampling methods can be used, e.g. fecal sampling⁷. Fortunately, nucleic acids of vertebrate-infecting microorganisms exhibiting a variety of tissue tropisms have already been recovered from fecal samples^{8–10}. These methods come with some disadvantages: they incur additional organizational costs, e.g. extra staff often needs to be recruited and specifically trained, they are labor-intensive and they exclude candidate reservoir species for which fecal sampling is impractical or impossible, e.g. wild rodents. This highlights the need for the development of complementary and cost-efficient monitoring techniques.

Blood-sucking invertebrates might represent interesting alternative sources of vertebrate-pathogen nucleic acids. Blood meals of many hematophagous arthropods have been demonstrated to contain DNA from their vertebrate hosts and the pathogens for which they act as vectors, i.e. malaria in birds¹¹. Besides vector-borne pathogens, it is also possible that non-vector-borne pathogens could be contained in blood meals. Interesting examples include the detection of RNA from H5N1 avian influenza A viruses in a mosquitoes' blood meal^{12–14}.

¹Epidemiology of highly pathogenic microorganisms, Robert Koch-Institute, Berlin, Germany. ²Albert-Ludwigs-University, Faculty of Biology, Freiburg, Germany. Correspondence and requests for materials should be addressed to C.H. (email: con.hoffmann@gmail.com) or F.H.L. (email: LeendertzF@rki.de)

Grubaugh and colleagues went a step further and proposed to use blood meal analysis as a tool to survey human pathogens in remote tropical locales, which they refer to as *xenosurveillance*¹⁵. Blood-sucking arthropods however often exhibit strong host preferences, which may be suboptimal when the objective is to survey infectious agent diversity in complex ecosystems with high biodiversity.

Non blood-sucking invertebrates feeding on vertebrate fecal matter and/or carrion, such as blow and flesh flies (here referred to simply as flies), might also be suitable for the surveillance of wildlife infectious agents. Flies are abundant and ubiquitous, have little host preference and are easy to trap^{16,17}. We also recently showed that flies often contain DNA fragments of their mammalian hosts^{16,17}. Finally, the genetic material of a number of hitchhiked microorganisms was already detected in flies, including food borne bacteria, e.g. *Salmonella* spp., and enteric viruses^{18–21}. For example, Newcastle disease virus (NDV) RNA was detected in, and even virions were isolated from, flies collected in the vicinity of infected chickens. Similarly, H5N1 RNA was found in flies collected in the surroundings of a poultry farm with infected birds^{22,23}.

These studies however focused on flies caught near high-density vertebrate populations, which raises questions about the broad applicability of this method. In this study, we investigate whether flies are suitable for vertebrate-infecting microorganism surveillance in complex ecosystems with high species richness. We analyzed flies collected in a remote tropical rainforest, Tai National Park (TNP), Côte d'Ivoire, and focused on an *a priori* favorable target: adenoviruses (AdV; family *Adenoviridae*). AdV are shed massively in feces, are usually host-specific and have already been detected in many vertebrates in TNP^{24,25}.

Results

Out of 498 flies, 156 (31%) contained mammalian 16S mt DNA. We considered all mammal positive flies as suitable for AdV screening, but due to shortage of material we could only test 142 of these flies. From eight flies AdV DNA could be amplified and sequenced once and six AdV sequences could be confirmed a second time. A BLAST search revealed that four of the six sequences were $\geq 98\%$ identical to a simian AdV sequence determined from a mona monkey (*Cercopithecus campbelli*, KP274048) in Côte d'Ivoire (Fly 92, Fly 101, Fly 740, Fly 1355)²⁶. The remaining two sequences (Fly 381, Fly 1375) showed 100% identity with simian AdV sequences obtained from captive chimpanzees in the US (FJ025905, FJ025926) and from a wild chimpanzee in TNP (JN163974) and 98% identity with a murine AdV 2 sequence (NC014899), respectively (Table 1).

We also performed phylogenetic analyses in both maximum likelihood and Bayesian frameworks to better determine the position of fly-derived AdV sequences within the AdV family tree (Fig. 1). In line with the BLAST search, sequences from Fly 92, Fly 101, Fly 740 and Fly 1355 formed a well-supported clade with the mona monkey AdV sequence (aLRT 0.99, pp 1; Fig. 2C). The sequence of Fly 381 clustered with AdV sequences from captive and wild chimpanzees (FJ025905, FJ025926, JN163974, FJ025906, FJ025904, FJ0295899; aLRT 0.98, pp 1; Fig. 2B)^{24,27}. These sequences nested within the clade corresponding to species *Human mastadenovirus C*, albeit with a much lower statistical support (HAdV-C; aLRT 0.93, pp 0.85). The sequence of Fly 1375 was most closely related to the murine AdV B (Murine AdV 2) sequence (aLRT 0.99, pp 1; Fig. 2A).

Meal analyses based on Sanger sequencing identified plausible hosts in 3 of the 6 AdV positive flies (Fly 92, Fly 101, Fly 1355; Table 1). To further investigate potential hosts, we performed an in-depth fly meal analysis of all six AdV positive flies using a metabarcoding approach. After quality trimming, 65,552 reads—8,683 to 12,278 reads per fly - were used for taxonomic assignment. Overall, we identified hosts from 9 mammal families and 10 genera/species. Five flies contained DNA from multiple hosts (Fly 92, Fly 101, Fly 381, Fly 1355, Fly 1375; Table 1). The 3 plausible hosts identified by Sanger sequencing were confirmed by this approach. Fly 740, which harbored one of the simian AdV sequences, only contained rodent DNA fragments. For the last two flies, the metabarcoding approach revealed the presence of DNA fragments belonging to plausible hosts, i.e. rodents in Fly 1375 and a hominid in Fly 381. As the hominid family contains the two closely related genera *Homo* and *Pan*, we manually checked the according sequences and were able to refine the assignment to *Pan troglodytes*.

Discussion

We investigated the feasibility of using DNA derived from flies for the surveillance of wildlife infectious diseases. We were able to detect short AdV sequences in 6 flies, that is 4.2% of all mammal positive flies. We used these sequences for phylogenetic analyses and found that most represented AdVs known to infect monkeys and great apes in the region^{24,25}. The close relationship of four sequences with an AdV sequence obtained from a single mona monkey supports the notion that this AdV may be relatively abundant in the region²⁶. The fifth fly-based simian AdV sequence clustered with HAdV-C sequences and clearly belonged to the chimpanzee clade. HAdV-C viruses are very host-specific and seem to have co-diverged with their hominid hosts²⁵. We also detected what is likely a new rodent AdV, thereby underlining the potential of flies to also monitor small-bodied species. Finally, our high-throughput fly meal analyses identified multiple hosts, including plausible ones, in 5 of 6 AdV positive flies. These results demonstrate that fly-based analyses allow for the simultaneous characterization of microorganism genetic diversity and their distribution in local mammalian hosts.

In comparison with detection rates in fecal samples (11 to 58%), the AdV detection rate in flies appears low^{25,28}. This might result from the extreme dilution of vertebrate-infecting microorganisms in carrion flies, which itself results from the interplay of meal quantity, quality, frequency and the speed of digestive processes²⁹. Given this low detection rate, systematic screenings would probably only make sense where fly collections established for other purposes, e.g. mammal diversity assessment, are available. Sample pooling combined with deep sequencing of PCR products may help decrease the workload and costs of such a screening approach.

Further investigations are needed to determine the extent to which the approach described here is applicable to other microorganisms. The low detection rate of AdV sequences in flies suggests that surveillance of non-enteric microorganisms might be complicated. However, in the case of outbreaks with massive production of microorganisms, e.g. Ebola virus outbreaks³⁰, there might be a good chance that pathogen nucleic acids

AdV Fly ID	closest AdV BLAST hit (identity/host/accession)	Mammal 16 S - fly meal analysis			
		closest BLAST hit	Sanger	NGS MiSeq	
			species assignment (assigned raw reads)	genus assignment (assigned raw reads)	family assignment (assigned raw reads)
92	Simian AdV (98%/Cercopithecus campbelli/KP274048)	Cercopithecus nictitans 96%	Cercopithecus nictitans (9289)	Cercopithecus (9859)	Cercopithecidae (11194)
			Colobus guereza (1335)	Colobus (1335)	Suidae (123)
			Cercopithecus campbelli (559)	Sus (123)	not assigned (38)
			Sus scrofa (123)	not assigned (38)	
			not assigned (49)		
101	Simian AdV (99%/Cercopithecus campbelli/KP274048)	Cercopithecus diana 90%	Cercopithecus diana (9327)	Cercopithecus (10266)	Cercopithecidae (10266)
			Cercopithecus campbelli (825)	Sus (395)	Suidae (395)
			Sus scrofa (352)	not assigned (24)	not assigned (24)
			not assigned (181)		
381	Simian AdV (100%/Pan troglodytes/JN163974, FJ025905, FJ025926)	dirty sequence	not assigned (10526)	Felis (6657)	Felidae (6657)
			Felis catus (178)	not assigned (4047)	Hominidae ² (4047)
740	Simian AdV (98%/Cercopithecus campbelli/KP274048)	unknown rodent 100% ¹	Cricetomys sp. 1 PG-2014 (6767)	Cricetomys (7050)	Nesomyidae (7050)
			not assigned (5511)	not assigned (5228)	not assigned (5228)
1355	Simian AdV (100%/Cercopithecus campbelli/KP274048)	Cercopithecus campbelli 99%	Cercopithecus campbelli (7529)	Cercopithecus (7571)	Cercopithecidae (11805)
			not assigned (4318)	not assigned (4259)	not assigned (42)
1375	Murine AdV 2 (98%/Mus musculus/NC014899)	Cephalophini 97%	not assigned (5252)	not assigned (3196)	
			Civettictis civetta (2598)	Civettictis (2598)	Viverridae (2598)
			Rattus rattus (620)	Cephalophus (1916)	Bovidae (1916)
			Neoromicia somalicus (213)	Rattus (760)	Muridae (760)
				Neoromicia (213)	Vespertilionidae (213)
					not assigned (57)

Table 1. Results of AdV screening and in depth fly meal analysis. ¹Assignment based on a local database. ²Manual BLAST search revealed 100% identity with *Pan troglodytes*.

are detectable in flies. Of course, the detection probability will depend on the biology of the microorganism of interest. Here also, high throughput sequencing approaches (including shotgun sequencing) could open up new perspectives, as recently shown with mosquitoes^{13,14}. It was recently demonstrated that portable sequencing devices such as the MinION (Oxford Nanopore Technologies, Oxford, United Kingdom) can be used for on-site sequencing in outbreak situations^{31,32}. These technologies only require a basic molecular laboratory in the field. Such laboratories currently allow users to perform sequencing, though the high error rate and relatively low throughput of MinIONs currently limit them to amplicon sequencing based approaches. The limited needs in the present study suggest it could be feasible to conduct fly/amplicon-based wildlife surveillance during major outbreaks, including in resource-poor countries.

Invertebrates other than carrion flies and blood sucking arthropods might also constitute a valuable source of information on vertebrate-infecting microorganisms. For example, leeches can ingest several times their weight host blood in a single meal and could therefore be seen as long-term blood tanks. Most recently, a number of viruses (with DNA or RNA genomes) were shown to persist up to four months in experimentally fed aquatic leeches, with bovine parvovirus being detectable for up to six months³³. Terrestrial leeches, whose lifestyle might be more compatible with broad, undirected wildlife molecular epidemiology, were also recently shown to allow retrieval of their host's DNA³⁴. Both aquatic and terrestrial leeches warrant a careful examination of their potential as tools for wildlife microorganism sampling.

Finally, an alternative to microorganism nucleic acid detection might be the detection of antibodies reactive to these microorganisms. If this is feasible, it would open the potential to examine wildlife exposure to microorganisms. Detection of trypanosome-reactive antibodies from a number of haematophagous dipterans was reported as early as 1962³⁵. This potential tool then fell into a long-lasting oblivion until its recent rediscovery by Barbazan and colleagues, who showed that blood-fed mosquitoes contain detectable levels of various virus-reactive antibodies³⁶. Determining whether invertebrate-based serological surveys can be conducted in the wild promises to be an exciting area of future research.

Material and Methods

Sample collection. Sample collection was performed with the permission of the Ivorian national parks authorities (OIPR) and the ministry of research of Côte d'Ivoire. Flies used in this study were captured in Taï National Park, Côte d'Ivoire, a tropical rainforest with remarkable mammal biodiversity. Overall, 498 flies were captured using customized fly traps consisting of a pyramidal mosquito net over a plastic bowl containing a

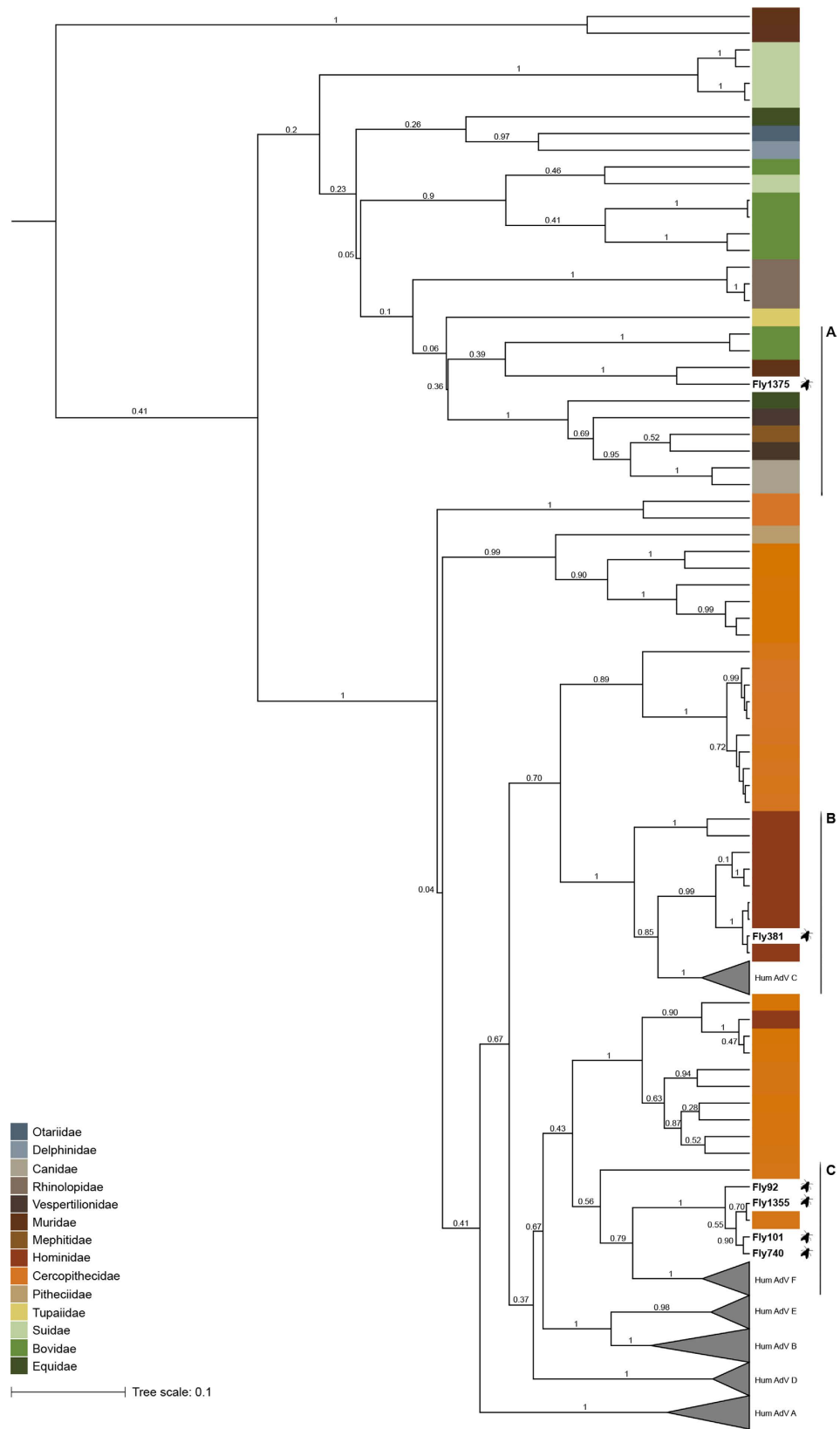


Figure 1. Maximum clade credibility (MCC) tree of mastadenoviruses. This MCC tree is based on the Bayesian analysis of a 370 bp long alignment of hexon gene sequences. Posterior probabilities are plotted above branches. The tree was built under a clock model and thus is rooted. Enlarged section A, B and C are shown in Fig. 2.

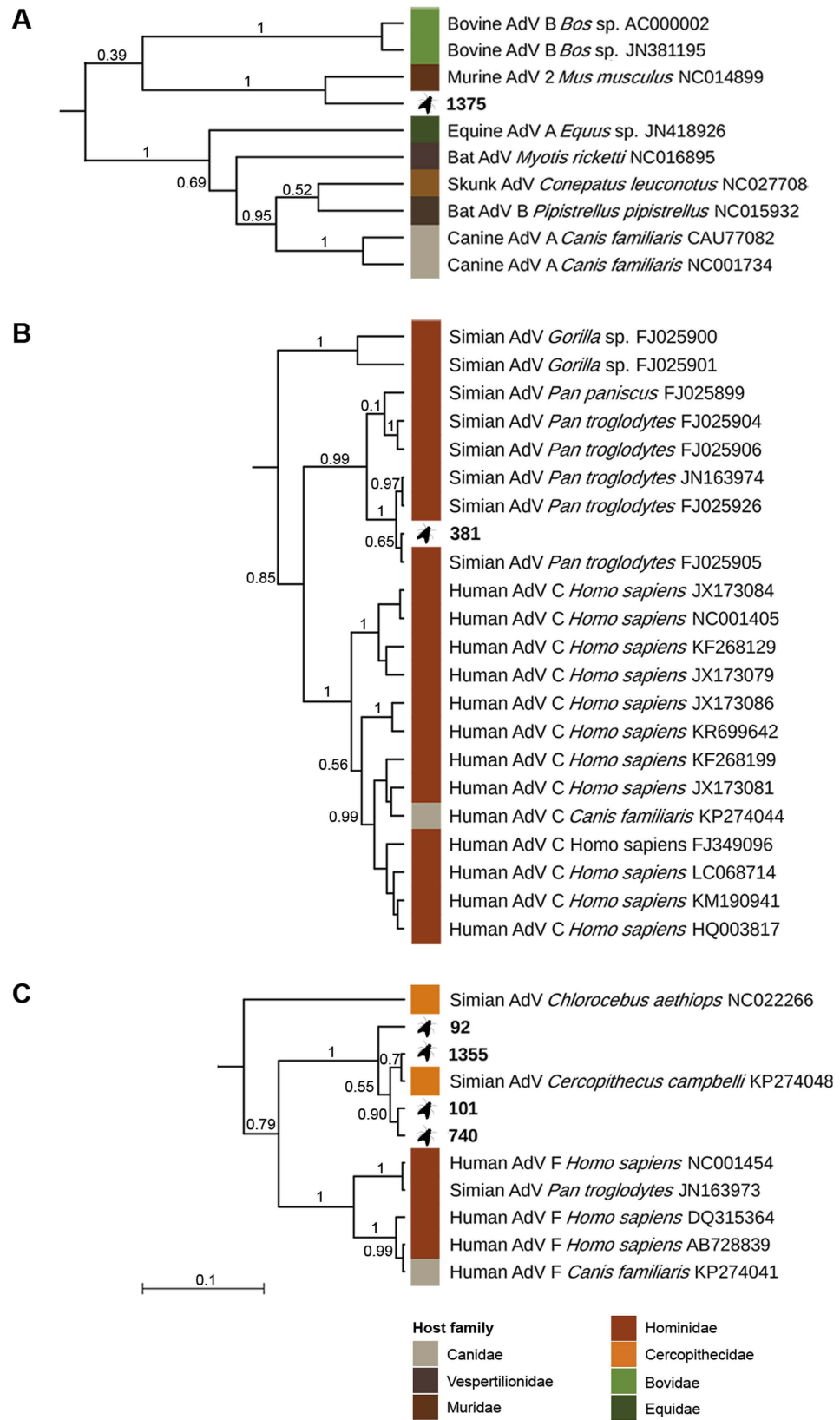


Figure 2. Enlarged sections of the maximum clade credibility of mastadenoviruses. The original MCC tree (Fig. 1) was based on the Bayesian analysis of a 370 bp long alignment of hexon gene sequences. Posterior probabilities are plotted above branches. Study sequences are in bold, reference sequences are represented by host name and accession number.

commercial bait (Unkonventionelle Produkte Feldner, Waldsee, Germany) or a piece of meat¹⁶. After collection, flies were either placed in Cryotubes (ThermoFischer, Waltham, MA, USA), and stored in liquid nitrogen tanks, or in 50 ml Falcon tubes (Carl Roth, Karlsruhe, Germany) containing silica and stored either at ambient temperature or 4 °C.

Nucleic acid extraction. Single flies were cut with sterile scissors and crushed in 250 µL phosphate buffer saline using the FastPrep[®]-24 Instrument (MP Biomedicals, Illkirch, France). One hundred µL of the resulting mix were then used to extract carrion fly DNA using the EURx GeneMatrix Stool DNA Purification Kit (Roboklon, Berlin, Germany).

Identification of flies suitable for adenovirus screening. We considered that flies containing mammal DNA would be suitable for adenovirus screening. To show the presence of mammal DNA, we first ran a real time PCR targeting a short 130 bp fragment of mitochondrial 16 S rRNA (16Smam1 5'-CggTTggggTgACCTCggA-3' and 16Smam2 5'-gCTgTTATCCCTAgggTAACT-3')³⁷. Reaction volume was 12.5 µL and contained 6.25 µL GoTaq[®] qPCR Master Mix (Promega, Fitchburg, WI, USA), 0.2 µM of each primer, 1 µM of blocking primer (human blocking primer 16Smam_blkhum3 5'-CggTTggggCgACCTCggAgCAGAACCC—spacerC3³⁸), 1 µL DNA extract or 1 µL of a diluted African palm civet (*Nandinia binotata*) PCR product (10¹–10⁵ molecules per µL) for standard reactions. Cycling conditions were: 95 °C 5 min, 40 cycles [95 °C 15 sec, 64 °C 1 min], and 95 °C 1 min, 55 °C 30 sec, 95 °C 30 sec. For all positive flies, we attempted to generate PCR products for sequencing using the same amplification primer pair as for the real-time PCR^{16,37}. PCRs were carried out in a total volume of 25 µL and seeded with 200 ng of DNA (DNA concentration > 40 ng/µL) or 5 µL DNA (DNA concentration < 40 ng/µL). Reactions contained 0.2 mM dNTP (with dUTP replacing dTTP), 4 mM MgCl₂, 0.2 µM of each primer, 1 µM of two different blocking primer (human blocking primer 16Smam_blkhum3 5'-CggTTggggCgACCTCggAgCAGAACCC—spacerC3³⁸, pig blocking primer 16Smam_blksl1 5'-CggTTggggTgACCTCggAgTACAAAAAAC—spacerC3¹⁶), 0.3 U Amperase[®] uracil N-glycosylase (Invitrogen), 1.25U Platinum[®] Taq Polymerase (Invitrogen) and 2.5 µL 10x PCR Buffer (Invitrogen). Cycling condition were: 45 °C 7 min, 95 °C 15 min, 42 cycles [95 °C 30 sec, 64 °C 30 sec, 72 °C 1 min], and 72 °C 10 min. PCR products were cleaned up with ExoSAP-IT[®] (Affymetrix, Santa Clara, CA, USA) and sequenced in both directions according to Sanger's method using the BigDye Terminator kit v3.1 (ThermoFischer). All chromatograms were evaluated using the software Geneious Pro v9.1.3 (Biomatters Ltd., Auckland, New Zealand)³⁹. Sequences were assigned to species or higher taxa using BLAST⁴⁰ and following the rationale depicted in Calvignac-Spencer and colleague's study¹⁶. Most of these assignments were made in course of the study of Schubert & Stockhausen *et al.* 2014. Flies that produced a band of the expected size but did not yield interpretable sequences were also used for AdV screening.

Adenovirus screening. We implemented various countermeasures to minimize contamination. To avoid cross-contamination with native AdV DNA, DNA extraction was never performed simultaneously with other sample types (fecal and tissue samples). To minimize contamination with PCR products, PCR setup and post-PCR analysis steps were performed in separate, dedicated rooms. In addition, a glovebox exclusively dedicated to fly analysis was used to set up all PCR performed for this study. We also used dUTP instead of dTTP and cleaned our confirmatory reactions with uracil n-glycosylase (UNG) to further reduce the likelihood of carry over contamination with PCR products (see below). It should also be noted that before this study, AdV sequences had never been amplified in the laboratory where AdV screening was performed.

A semi-nested PCR system described by Pauly *et al.*²⁶ was used for detection of adenoviruses²⁶. Primers had been designed for the generic detection of mastadenoviruses and targeted a short 160 bp fragment of the hexon gene (6500 s 5'-CgCAGTggKCNTWCATgCACAT-3', 6500 s 5'-ACCCACgAYgTSACNACNgA-3', 6500as 5'-gTgCCggTgTANggYTTRAA-3'). All PCRs were carried out in a 25 µL mix containing 0.2 mM dNTP (with dUTPs replacing dTTPs), 4 mM MgCl₂, 0.2 µM of each primer, 1.25U Platinum[®] Taq Polymerase (Invitrogen) and 2.5 µL 10x PCR Buffer (Invitrogen). Reactions of the first round were seeded with 200 ng DNA extract or 5 µL if DNA concentration was below 40 ng/µL and the second round with 1.5 µL of 1:40 diluted PCR product of the first round. Cycling conditions were: 95 °C 5 min, 40 cycles [95 °C 30 sec, 56 °C 30 sec, 72 °C 60 sec], and 72 °C 10 min. PCR products were Sanger sequenced as described above. Flies apparently containing amplifiable AdV nucleic acids were confirmed using the same assay but including 0.3 U Amperase[®] UNG (Invitrogen) in the first round reaction so as to minimize the risk of contamination with PCR products. Again, PCR products that yielded a band were sequenced using the Sanger method. All chromatograms were evaluated using Geneious Pro v9.1.3³⁹ and the respective sequences were confirmed to be adenoviral by a BLAST search⁴⁰.

Phylogenetic analysis. The dataset used for phylogenetic analysis comprised AdV sequences generated in this study (n = 6) and hexon gene sequences extracted from all available complete genomes of the genera *Mastadenovirus* and *Atadenovirus* (n = 506). This set of sequences was reduced to only contain unique sequences using FaBox v1.41⁴¹. All remaining 363 sequences were aligned at the nucleotide level in SeaView v4⁴², using the MUSCLE algorithm⁴³. Conserved blocks were selected using Gblocks⁴⁴ (as implemented in SeaView v4) resulting in an alignment of 370 positions. After block selection, sequences were de-replicated again using FaBox v1.41 (218 unique sequences). The best-fit model of nucleotide substitution was selected using JModelTest v2.14 and the Bayesian information criterion⁴⁵ (SYM + I + G). Maximum likelihood (ML) as well as Bayesian frameworks were used for tree reconstruction. The ML tree was reconstructed using PhyML v3.0⁴⁶. Branch support was estimated using SH-like approximate likelihood-ratio tests. The Bayesian phylogeny was estimated using BEAST v1.8.2⁴⁷ under the assumption of a relaxed clock (lognormal) and a Yule tree prior. Multiple Markov chain Monte

Carlo analyses were run; convergence and effective sample sizes were checked in Tracer v1.6 (combined effective sampling size was >200). Tree files of all runs were combined using LogCombiner v1.8.2 and the maximum clade credibility tree was extracted using TreeAnnotator v1.8.2.

Identification of fly meals. We performed in depth meal analysis of the 6 AdV-positive flies using a metabarcoding approach. Primary 16S amplicons were generated with the same primers and under the same conditions as mentioned above^{16,37}. Preparation of the generated amplicons for the Illumina MiSeq (San Diego, CA, US) sequencing platform included a first PCR in which Illumina specific overhang adapters are added to the fragment and a second PCR in which sequencing adapters and sample specific indexes were added. The first PCR reaction contained 25 µL of 1:50 diluted PCR-product, 0.2 mM dNTP, 4 mM MgCl₂, 0.2 µM of each fusion primer (16S-mam primer plus Illumina specific adapter sequence), 1 µM of each blocking primer, 1.25 U Platinum[®] Taq Polymerase (Invitrogen) and 10x PCR Buffer (Invitrogen) in a total volume of 50 µL. Cycling conditions were: 95 °C 5 min, 5 cycles [95 °C 30 sec, 64 °C 30 sec, 72 °C 1 min], and 72 °C 10 min. The second PCR mix contained 5 µL of cleaned PCR product from the first PCR, 25 µL KAPA HiFi[™] Hot Start ReadyMix (PeqLab, Erlangen, Germany), 10 µM of each Nextera Index Primer (Illumina) in a total volume of 50 µL. Cycling conditions were 95 °C 3 min, 15 cycles of 95 °C 30 sec, 55 °C 30 sec, 72 °C 30 sec followed by elongation at 72 °C for 5 min. Both the first and second PCR products were cleaned using the Agencourt[®] AMPure[®] XP PCR1 Purification system (Beckman Coulter, CA, USA). The final dual-indexed amplicon libraries were quantified using Quant-iT[™] dsDNA Broad-Range Assay Kit (Invitrogen) and pooled equimolarly. The pool was sequenced on an Illumina MiSeq platform using the MiSeq Reagent Kit v2 (2 × 150 bp; Illumina). Raw reads were analyzed using a custom bioinformatic pipeline: paired-end reads were first merged with the program *illumina-paired-end* of the software package *OBITools* v1.1.18 setting the minimum alignment score to 40. Primer sequences were then removed using the program *Cutadapt* v1.2.1^{48,49} before quality trimming was conducted with the program *Trimmomatic* v0.35⁵⁰ setting the quality score to 30 over a sliding window of four bases. We then de-replicated identical sequences and filtered out those that occurred less than 10 times using the *obiuniq* and *obigrep* commands of *OBITools*. For taxonomic assignment a reference database was built by performing an *in silico* PCR on all mammalian and vertebrate sequences available at Genbank (<http://www.ncbi.nlm.nih.gov/genbank/>) using the program *ecoPCR* v0.2^{51,52}. This not only contained the reference sequences themselves, but also a unique *taxid* that links each sequence to a taxonomy database where the taxonomic information was stored. For the assignment itself we used the *ecotag* command of *OBITools*. *Ecotag* uses the global alignment algorithm Needleman-Wunsch to find the most similar sequence to the query sequence in the reference database with a minimum identity level of 0.95 (primary reference sequence). The query sequence is then assigned to the most common recent ancestor of the primary reference sequence and the most similar reference sequence to the primary reference (secondary reference sequence).

References

- Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
- Bartsch, S. M., Gorham, K. & Lee, B. Y. The cost of an Ebola case. *Pathog. Glob. Health* **109**, 4–9 (2015).
- Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nat. Lett.* **451**, 990–994 (2008).
- Daszak, P. Emerging Infectious Diseases of Wildlife - Threats to Biodiversity and Human Health. *Science* **287**, 443–449 (2000).
- Keesing, F. *et al.* Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468**, 647–652 (2010).
- Calvignac-Spencer, S., Leendertz, S. A. J., Gillespie, T. R. & Leendertz, F. H. Wild great apes as sentinels and sources of infectious disease. *Clin. Microbiol. Infect.* **18**, 521–527 (2012).
- Leendertz, F. H. *et al.* Pathogens as drivers of population declines: The importance of systematic monitoring in great apes and other threatened mammals. *Biol. Conserv.* **131**, 325–337 (2006).
- Liu, W. M. *et al.* Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* **467**, 420–467 (2010).
- Keele, B. F. *et al.* Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523–526 (2006).
- Köndgen, S. *et al.* Pandemic Human Viruses Cause Decline of Endangered Great Apes. *Curr. Biol.* **18**, 260–264 (2008).
- Mehus, J. O. & Vaughan, J. A. Molecular identification of vertebrate and hemoparasite DNA within mosquito blood meals from eastern North Dakota. *Vector Borne Zoonotic Dis.* **13**, 818–24 (2013).
- Barbazan, P. *et al.* Detection of H5N1 avian influenza virus from mosquitoes collected in an infected poultry farm in Thailand. *Vector Borne Zoonotic Dis.* **8**, 105–109 (2008).
- Ng, T. F. F. *et al.* Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. *PLoS One* **6**, e20579 (2011).
- Shi, C. *et al.* A metagenomic survey of viral abundance and diversity in mosquitoes from Hubei province. *PLoS One* **10**, e0129845 (2015).
- Grubaugh, N. D. *et al.* Xenosurveillance: A novel mosquito-based approach for examining the human-pathogen landscape. *PLoS Negl. Trop. Dis.* **9**, e0003628 (2015).
- Calvignac-Spencer, S. *et al.* Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Mol. Ecol.* **22**, 915–924 (2013).
- Schubert, G. *et al.* Targeted detection of mammalian species using carrion fly-derived DNA. *Mol. Ecol. Resour.* **15**, 285–294 (2015).
- Collinet-Adler, S. *et al.* Environmental factors associated with high fly densities and Diarrhea in Vellore, India. *Appl. Environ. Microbiol.* **81**, 6053–6058 (2015).
- Chaiwong, T. *et al.* The blow fly, *Chrysomya megacephala*, and the house fly, *Musca domestica*, as mechanical vectors of pathogenic bacteria in northeast Thailand. *Trop. Biomed.* **31**, 336–346 (2014).
- Bagshaw, C., Isdell, A. E., Thiruvaiyaru, D. S., Brisbin, I. L. & Sanchez, S. Molecular detection of canine parvovirus in flies (Diptera) at open and closed canine facilities in the eastern United States. *Prev. Vet. Med.* **114**, 276–284 (2014).
- Riordan, J. T., Paul, J. R., Yoshioka, I. & Horstmann, D. M. The detection of poliovirus and other enteric viruses in flies. Results of tests carried out during an oral poliovirus vaccine trial. *Am. J. Hyg.* **74**, 123–36 (1961).
- Chakrabarti, S. *et al.* Detection and isolation of exotic Newcastle disease virus from field-collected flies. *J. Med. Entomol.* **44**, 840–4 (2007).
- Sawabe, K. *et al.* Detection and isolation of highly pathogenic H5N1 avian influenza A viruses from blow flies collected in the vicinity of an infected poultry farm in Kyoto, Japan, 2004. *Am. J. Trop. Med. Hyg.* **75**, 327–332 (2006).

24. Wevers, D. *et al.* Novel adenoviruses in wild primates: a high level of genetic diversity and evidence of zoonotic transmissions. *J. Virol.* **85**, 10774–84 (2011).
25. Hoppe, E. *et al.* Multiple cross-species transmission events of human adenoviruses (HAdV) during hominine evolution. *Mol. Biol. Evol.* **32**, 2072–2084 (2015).
26. Pauly, M. *et al.* Adenovirus in rural Côte d'Ivoire: High diversity and cross-species detection. *Ecohealth* **12**, 441–452 (2015).
27. Roy, S. *et al.* Isolation and characterization of adenoviruses persistently shed from the gastrointestinal tract of non-human primates. *PLoS Pathog.* **5**, e1000503 (2009).
28. Zheng, X. *et al.* Detection of novel adenoviruses in fecal specimens from rodents and shrews in southern China. *Virus Genes* **52**, 417–421 (2016).
29. Calvignac-Spencer, S., Leendertz, F. H., Gilbert, M. T. P. & Schubert, G. 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**, 1004–1013 (2013).
30. Formenty, P. *et al.* Ebola virus outbreak among wild chimpanzees living in a rain forest of Cote d'Ivoire. *J. Infect. Dis.* **179**, 120–S126 (1999).
31. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
32. Hoenen, T. *et al.* Nanopore sequencing as a rapidly deployable Ebola outbreak tool. *Emerg. Infect. Dis.* **22**, 331–334 (2016).
33. Al-Khleif, A. *et al.* Tenacity of mammalian viruses in the gut of leeches fed with porcine blood. *J. Med. Microbiol.* **60**, 787–792 (2011).
34. Schnell, I. B. *et al.* Erratum: Screening mammal biodiversity using DNA from leeches. *Curr. Biol.* **22**, 1980 (2012).
35. Cunningham, M. P., Harley, J. M., Southon, H. A. & Lumsden, W. H. Detection of antibodies in blood meals of hematophagous Diptera. *Science* **138**, 32–33 (1962).
36. Barbazan, P., Palabodeewat, S., Nitatpattana, N. & Gonzalez, J.-P. Detection of host virus-reactive antibodies in blood meals of naturally engorged mosquitoes. *Vector Borne Zoonotic Dis.* **9**, 103–8 (2009).
37. Taylor, P. G. Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Mol. Biol. Evol.* **13**, 283–285 (1996).
38. Boessenkool, S. *et al.* Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Mol. Ecol.* **21**, 1806–1815 (2012).
39. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
40. Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
41. Villesen, P. FaBox: An online toolbox for FASTA sequences. *Mol. Ecol. Notes* **7**, 965–968 (2007).
42. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–4 (2010).
43. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
44. Talavera, G., Dincă, V. & Vila, R. Factors affecting species delimitations with the GMYC model: Insights from a butterfly survey. *Methods Ecol. Evol.* **4**, 1101–1110 (2013).
45. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772–772 (2012).
46. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
47. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
48. Boyer, F. *et al.* obitools: A unix-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* **16**, 176–182 (2016).
49. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
50. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
51. Bellemain, E. *et al.* ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiol.* **10**, 189 (2010).
52. Ficetola, G. F. *et al.* An in silico approach for the evaluation of DNA barcodes. *BMC Genomics* **11**, 434 (2010).

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 are also especially indebted to Sonja Metzger, Kathrin Nowak, Floraine Léguillon and Katerina Albrechtova who collected many of the carrion flies used in this study. For editing the language of the manuscript we thank Jan Frederik Gogarten. We are thankful to the Sonnenfeld-Stiftung for funding. Finally, we thank all members of the Epidemiology of highly pathogenic microorganisms group for stimulating discussions.

Author Contributions

S.C., M.S., C.H. and F.L. designed the study. K.M., M.S. and C.H. performed the experiments. C.H. and S.C. analyzed the data and C.H., M. S. and S.C. wrote the manuscript. F.L. contributed to the interpretation of the data and critically read the manuscript.

Additional Information

Accession codes: AdV sequences and raw reads of fly meal analysis are available at the European Nucleotide Archive under the accession numbers (Fly 92: LT617635; Fly 101: LT617636; Fly 381: LT617637; Fly 740: LT617638; Fly 1355: LT617639; Fly 1375: LT617640) and study accession number PRJEB14554 (sample accession numbers ERS1272921- ERS1272926).

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Hoffmann, C. *et al.* Assessing the feasibility of fly based surveillance of wildlife infectious diseases. *Sci. Rep.* **6**, 37952; doi: 10.1038/srep37952 (2016).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016