Received: 4 November 2022 | Revised: 30 January 2023 | Accepted: 17 February 2023

DOI: 10.1002/edn3.401

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

WILEY

SPECIAL ISSUE ARTICLE

Fly iDNA suggests strict reliance of the causative agent of sylvatic anthrax on rainforest ecosystems

Mueena Jahan[1,2,3] | Lorenzo Lagostina[1,2] | Tobias Gräßle[1,2] | Emmanuel Couacy-Hymann[4] | Leonce Kouadio[4,5] | Valère K. Kouakou[4] | Hermann A. Krou[4] | Arsène M. Mossoun[4] | Livia V. Patrono[1,2,6] | Kamilla Pléh[1,2] | Jonas A. Steiner[1] | Noma Yves[7] | Fabian H. Leendertz[1,2] | Sébastien Calvignac-Spencer[2,6] | Jan F. Gogarten[1,2,6,8]

[1]Helmholtz Institute for One Health, Greifswald, Germany

[2]Epidemiology of Highly Pathogenic Organisms, Robert Koch Institute, Berlin, Germany

[3]Department of Microbiology and Public Health, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur, Bangladesh

[4]Laboratory for Innovative Research for Emerging Diseases and Diagnostics, Centre National de Recherche Agronomique, NGO One Health for All, Abidjan, Côte d'Ivoire

[5]Université Peleforo Gon Coulibaly, Korhogo, Côte d'Ivoire

[6]Viral Evolution, Robert Koch Institute, Berlin, Germany

[7]Laboratoire d'écologie moléculaire et de génétique evolutive, Department of Genetics, University Felix Houphouët Boigny, Abidjan, Côte d'Ivoire

[8]Applied Zoology and Nature Conservation, University of Greifswald, Greifswald, Germany

Correspondence
Sébastien Calvignac-Spencer and Jan F. Gogarten, Epidemiology of Highly Pathogenic Organisms, Robert Koch Institute, Berlin, Germany and the Helmholtz Institute for One Health, Greifswald, Germany.
Email: calvignacs@gmail.com and jan.gogarten@gmail.com

Funding information
Deutsche Forschungsgemeinschaft; European Commission; Ministry of Science and Technology, Government of the People's Republic of Bangladesh

Abstract

Metabarcoding of invertebrate-derived DNA (iDNA) is increasingly used to describe vertebrate diversity in terrestrial ecosystems. Fly iDNA has also shown potential as a tool for detecting pathogens. Combining these approaches makes fly iDNA a promising tool for understanding the ecology and distribution of novel pathogens or emerging infectious diseases. Here, we use fly iDNA to explore the geographic distribution of *Bacillus cereus* biovar *anthracis* (Bcbva) along a gradient from the forest within Taï National Park, Côte d'Ivoire, out to surrounding villages. We tested fly pools ($N = 100$ pools of 5 flies) collected in the forest ($N = 25$ pools), along the forest edge ($N = 50$ pools), and near surrounding villages ($N = 25$ pools) for Bcbva. Using the same iDNA, we sought to reconstruct fly and mammal communities with metabarcoding, with the aim of investigating potential links with Bcbva detection. We detected Bcbva in 5/100 fly pools and positivity varied significantly across the habitat types (forest = 4/25, edge = 1/50, village = 0/25). It was possible to culture Bcbva from all positive fly pools, confirming their positivity, while sequencing of their whole genomes revealed a considerable portion of known genomic diversity for this pathogen. iDNA generated data about the mammal and fly communities in these habitats, revealing the highest

Sébastien Calvignac-Spencer and Jan F. Gogarten should be considered joint senior author.

mammal diversity in the forest and considerable changes in fly community composition along the gradient. Bcbva host range estimates from fly iDNA were largely identical to the results of long-term carcass monitoring efforts in the region. We show that fly iDNA can generate data on the geographic distribution and host range of a pathogen at kilometer scales, as well as reveal the pathogen's phylogenetic diversity. Our results highlight the power of fly iDNA for mammal biomonitoring and pathogen surveillance.

KEYWORDS
Anthropized environments, *Bacillus cereus* biovar *anthracis*, biodiversity, environmental DNA (eDNA), invertebrate-derived DNA (iDNA), mammals, sylvatic anthrax, wildlife infectious diseases

1 | INTRODUCTION

Anthropogenic disturbance is causing major declines in local and global biodiversity. These changes have been linked to the emergence of pathogens, not just in human populations, but among wildlife as well (Hale et al., 2022; Stegen et al., 2017). Systematic monitoring of both wildlife populations and their infectious diseases is needed to help inform conservation efforts and design disease emergence mitigation strategies (Butchart et al., 2010).

Long-term wildlife monitoring can be an effective tool for understanding mammal populations and their disease dynamics (Hoffmann et al., 2017; Kuisma et al., 2019). Unfortunately, long-term behavioral observations, that can lead to the recording of potential ill-health manifestations, and the detection and sampling of carcasses in the wild, are extremely resource intensive and, therefore, not particularly scalable. Camera trapping represents a means to detect a diversity of mammal species in an area, potentially even providing data on abundance (Gilbert et al., 2021), but can only detect extremely visible disease symptoms (Hockings et al., 2021) and cannot result in firm diagnosis, that is, the identification of the causative agent. Non-invasive sampling of mammal excreta (e.g., of feces, urine) has proven useful for detection of both mammals, the antibodies they produce when exposed to a pathogen, and the direct detection of pathogen genetic material (Mombo et al., 2020; Morin et al., 2016; Santiago et al., 2002), but fresh fecal sample collection can be challenging. This is particularly true in habitats like tropical rainforests, where feces quickly decompose and mammals that are not habituated to humans are typically hard to observe even briefly, let alone long enough to observe them defecating.

In contrast, trace amounts of animal DNA shed into the environment (environmental DNA, eDNA) can be detected with metabarcoding techniques and represents an effective and scalable tool for tracking the distribution of wildlife and at least some pathogens, particularly in aquatic ecosystems (Andruszkiewicz et al., 2017; Bohmann et al., 2014; Ficetola et al., 2008; Lynggaard et al., 2022; Stat et al., 2017). In terrestrial ecosystems, the animal DNA found in invertebrates that come into contact with animals or their by-products as part of their life cycle (e.g., leeches, sand flies, carrion flies) have shown their ability to rapidly assess mammal biodiversity in tropical and temperate ecosystems (Calvignac-Spencer et al., 2013; Gogarten et al., 2020; Hoffmann et al., 2018; Rodgers et al., 2017; Schnell et al., 2012). Crucially, invertebrate-derived DNA (iDNA) has also demonstrated its utility as a tool for detecting pathogens. Thus, iDNA can reveal information about the ecology of a pathogen, even where the invertebrate sampler is not a vector of the pathogen under scrutiny (Bitome-Essono et al., 2017; Hoffmann et al., 2017).

Such iDNA-based pathogen surveillance can be particularly useful for novel pathogens or emerging infectious diseases, for which the host range and geographic distribution are initially poorly understood or rapidly changing. For example, a new form of anthrax caused by *Bacillus cereus* biovar *anthracis* (Bcbva) was identified in dead wild chimpanzees in a tropical rainforest ecosystem in Taï National Park (TNP), Côte d'Ivoire, in 2001 (Leendertz et al., 2004). Genomic studies revealed that while chromosomally distinct from the causative agent of savannah anthrax, *Bacillus anthracis*, Bcbva shares two virulence plasmids with it, PX01 and PX02 (Klee et al., 2010). Both bacteria show vegetative and endospore stages in their life cycle, with clear evidence for replication inside mammal carcasses. Yet, it is still unclear how similar their ecologies are. Indeed, even the apparent restriction of Bcbva to forested areas is uncertain—a recent seroprevalence study detected anti-Bcbva antibodies in the human population living around TNP but failed at linking seropositivity to forest use or contact to wildlife (Dupke et al., 2020).

Fly iDNA has clear potential to contribute to understanding the restriction of Bcbva to forest ecosystems and aspects of this pathogen's ecology. For classic anthrax, flies have been implicated as a potential mechanical vector and the high proportion of flies carrying viable Bcbva (~5% contained Bcbva), particularly in flies associated with monkey groups, supports the notion that flies play a role in spreading Bcbva as well (Fasanella et al., 2010; Gogarten et al., 2019; Hoffmann et al., 2017). Fly iDNA also contributed to understanding the ecology of the pathogen by expanding the putative host range of Bcbva, through detection of the bacterium and the DNA of a diversity of mammal species in specific flies (Hoffmann et al., 2017). Fly iDNA also helped to understand the distribution of the pathogen at

various spatial scales, from the prevalence of Bcbva across TNP, to an expansion of its known geographic distribution to include Liberia (Hoffmann et al., 2017).

Here, we use fly iDNA to explore the distribution of Bcbva along a gradient from the TNP forest to the surrounding villages. Using the same iDNA, we also aimed to reconstruct fly and mammal communities along this gradient with the aim of investigating potential links with Bcbva detection.

2 | MATERIALS AND METHODS

2.1 | Study site and sample collections

Flies were trapped on a gradient spanning from pristine forest within TNP to the surrounding villages on alternate days from the end of July to the beginning of September 2021, during the rainy season. A total of nine trapping areas were located along three parallel transects, covering three different habitats. Specifically, three trapping areas were selected in villages' periphery, with traps located ~100 m away from the last house of the three villages (hereafter referred to as village habitat; Figure 1). Three areas were selected within the transition area between cultivated land and TNP, where degraded forest encountered cocoa fields or manioc/banana fields (hereafter referred to as the edge habitat; Figure 1). Lastly, three areas were selected well into the pristine forest within TNP, in a mix of swampy and dry primary forest (hereafter referred to as forest habitat; Figure 1).

Fly trapping was carried out for 20–30 min, depending on the trapping success, with the aim of collecting 40 flies per sampling event. Flies were trapped on the ground using custom-made nets (for a detailed description, see Calvignac-Spencer et al., 2013; Hoffmann et al., 2018). Internal organs of cow and pig were purchased from the local market and used as bait and a piece of net minimized direct contact between flies and the bait. Previous analyses in TNP of the fly species attracted with a trapping approach using a piece of a decaying animal or a commercially available bait based on animal proteins that mimic a decaying carcass, described a diversity of dipterans from three families (Calliphoridae, Sarcophagidae, and Muscidae; Calvignac-Spencer et al., 2013; Hoffmann et al., 2017, Gogarten et al., 2019). Flies were euthanized with ether and stored on silica at room temperature in 15 mL falcon tubes containing up to 10–12 flies from a single trap ($N = 2350$). We selected 100 fly pools (5 random flies were selected from a single trap/pool) for this study; 25 pools were from the forest, 50 pools from the edge, and 25 fly pools from the village. We elected to concentrate sampling on the edge, as this was the habitat that was the least well characterized in terms of Bcbva prevalence and the mammal community present.

2.2 | DNA extraction

The DNA was extracted from these fly pools using the GeneMatrix Stool DNA purification kit (Roboklon, Berlin, Germany). In brief, five flies were transferred to a bead tube and sliced into smaller pieces with autoclaved scissors and then homogenized using the tissue lyser II (Qiagen). After centrifugation, supernatants were subsequently processed according to the manufacturer's protocol. DNA concentrations were then measured using the Qubit 3 fluorometer and the high sensitivity dsDNA assay kit (Invitrogen by Thermo Fisher Scientific).

2.3 | Bcbva analyses

2.3.1 | Detection using real time PCR

To determine Bcbva presence, fly pool extracts were screened with three quantitative PCR assays targeting three different gene markers (Hoffmann et al., 2017). All extracts were tested for the first gene marker, *pag* (protective antigen gene, located on the pXO1 plasmid; Ellerbrok et al., 2002). Samples that tested positive for the *pag* gene were subsequently screened for the presence of the second gene marker *capB* (gene for capsule synthesis; located on the

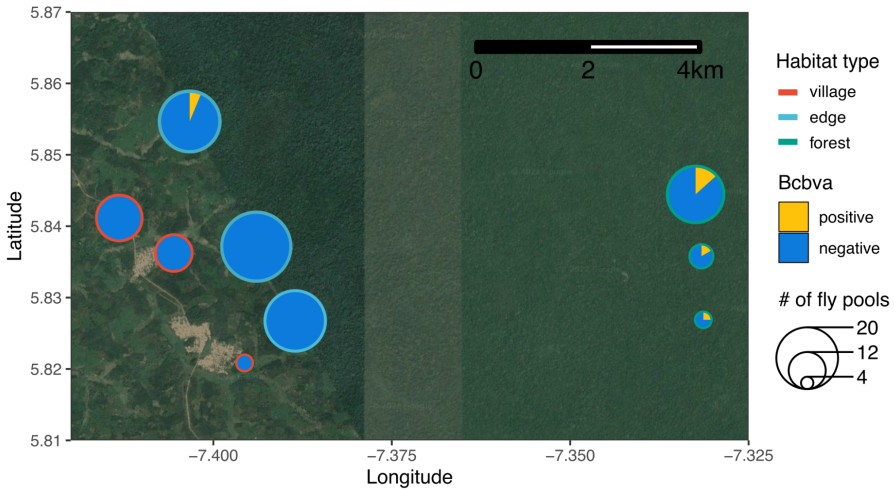

FIGURE 1 Locations of fly trapping along the park boundary of Taï National Park. The area of the circles corresponds to the number of fly pools examined at each location, while the pie chart indicates the number of fly pools positive or negative for Bcbva. The color of the edges of the pie charts corresponds to the habitat type where sampling occurred.

pXO2 plasmid) and the third marker, *Island IV* (chromosomal marker) that is specific for Bcbva, allowing to differentiate Bcbva and *Bacillus anthracis* (Klee et al., 2010). Like Hoffmann et al. (2017), we elected to use three quantitative PCR assays, rather than just a single assay for *Island IV*, to reduce false positives and increase certainty that we were detecting Bcbva and not Bcbva (or another hypothetical and yet unknown *Bacillus cereus* harboring *Island IV*) lacking virulence plasmids (a possibility that cannot be excluded but that would likely affect the bacterium's pathogenic potential). All assays were carried out with two PCR replicates per extract and only considered positive if both replicates were positive. Positive results were further confirmed by culture and whole genome sequencing (see below).

PCRs were carried out in a total volume of 25 μL, consisting of 2.5 mM dUTPs, 50 mM $MgCl_2$, 10 μM of each primer, 10 μM of the analogous probe, 1.25 U Platinum® Taq Polymerase (Invitrogen), 2.5 μL 10× Rxn buffer (Invitrogen), and molecular grade water. The whole reaction was seeded with 200 ng DNA ($N = 16$) or 5 μL of DNA extracts when DNA concentration was below 40 ng/μl ($N = 84$). The following cycling conditions were used: 95°C for 10 min, followed by 45 cycles at 95°C for 15 s and at 60°C for 34 s. Quantitative PCR was carried out using the AriaMX Real-Time PCR system (Agilent Technologies) and analyzed using the Agilent AriaMx software system.

2.3.2 | Culture from qPCR positive isolates

Bacterial culture was conducted via dilution streaking or spread plating of 10 μL fly pool-PBS supernatant onto R & F® *Bacillus cereus/Bacillus thuringiensis* Chromogenic Plating Medium (R & F Products) and incubation at 37°C overnight. Bcbva suspicious, phospholipase-C deficient (as indicated by white colony growth) single colonies were picked.

2.3.3 | Whole genome sequencing of Bcbva isolates

DNA extraction

DNA extraction was performed using the DNeasy Blood and Tissue kit, following the manufacturer's protocol for Gram-positive bacteria (Qiagen). Briefly, Bcbva suspicious colonies were individually picked for DNA extraction. Each colony was transferred into an extraction tube containing 162 μL of lysis buffer. Then 18 μL of lysozyme was added to each extraction tube and the colony was resuspended. This bacterial suspension was then extracted following the manufacturer's protocol for Gram-positive bacteria.

Bacterial DNA concentrations were measured with a Qubit 3 fluorometer and the high sensitivity dsDNA assay kit (Invitrogen by Thermo Fisher Scientific).

Library preparation and sequencing

Bcbva libraries were prepared for whole genome sequencing using the Nextera XT DNA library prep kit (Illumina). Following the manufacturer's protocol, 1 ng of genomic DNA was tagmented, indexed with the Nextera XT index primers and PCR amplified. The cycling conditions were as follows: 72°C for 3 min, 95°C for 30 s, 12 cycles of 95°C for 10 s, 55°C for 30 s, 72°C for 30 s, and a final step of elongation of 72°C for 5 min. Amplified products were then cleaned up using MagSi NGSPREP Plus beads (Steinbrenner Laborsysteme GmbH) to purify the library DNA and short library fragments were removed. Libraries were then quantified using a Qubit, normalized and pooled for sequencing on an Illumina NextSeq 2000 with P2 reagents and 2 × 150 cycles.

2.3.4 | Mammal and fly analyses

DNA extracts were also used to reconstruct the mammals and fly species diversity in and across the TNP using a metabarcoding approach. For mammal metabarcoding, a three-step PCR assay was used to reduce the amplification biases as described in Hoffmann et al. (2017). For the first PCR assay, 130 bp of 16 S mitochondrial DNA was PCR amplified using two universal primers 16 S mam1(5′-CGGTTGGGGTGACCTCGGA-3′), 16 S mam2 (5′-GCT GTT ATC CCT AGG GTA ACT-3′) along with the two sets of blocking primers to reduce the amplification potential laboratory contamination from human and pig, 16 S mam_blk hum 3 (5′-CGGTTGGGGCGACC TCGGAGCAGAACCC-3′) and 16Smam_blkpig (5′-CGGTTGGGGT GACCTCGGAGTACAAAAAAC-3′), respectively. PCRs were carried out with four PCR replicates for all samples and negative template controls. Each PCR was carried out in a total volume of 25 μL, where the reaction was seeded with 200 ng of DNA ($N = 16$) or 5 μL DNA if the DNA concentration < 40 ng/μL ($N = 84$). The reaction mixture included 2.5 mM dNTP (replaced by dUTP), 50 mM $MgCl_2$, 10 μM of each primer, 10 μM of two blocking primers (human and pig), 0.3 U Amperase® uracil N-glycosylase (Invitrogen), 1.25 U Platinum® Taq Polymerase (Invitrogen), 2.5 μL 10× PCR Buffer (Invitrogen), and molecular grade water. Cycling conditions were as follows: 45°C for 7 min, 95°C for 15 min, 42 cycles of 95°C for 30 s, 64°C for 30 s, 72°C for 1 min, followed by elongation at 72°C for 10 min. PCR-amplified products were then visualized on 1.5% agarose gel. All negative template controls were negative on the gel, but were further processed as positive samples. For the latter, products of the expected size were gel-excised under the UV transilluminator and purified using the quick gel purification kit (purelink™; Invitrogen).

Gel-purified products were then submitted to a second PCR. For the PCR (total volume 25 μL), it was seeded with 5 μL of 16 S PCR-product, 2.5 mM dNTP, 4 mM $MgCl_2$, 10 μM of each fusion primer (16Smam primer appended with overhang Illumina specific adapter sequence), 1.25 U Platinum® Taq Polymerase (Invitrogen), and 2.5 10× PCR Buffer (Invitrogen). Cycling conditions were as follows: 95°C for 5 min, 15 cycles of 95°C for 30 s, 64°C for 30 s, 72°C for 1 min, and a final elongation of 72°C for 10 min. These amplified products were cleaned using paramagnetic beads with 1.8 μL × MagSi NGSPREP Plus beads to 1 μL amplicon ratio and eluted in 17.5 μL TET buffer.

In parallel, we also proceeded with fly species metabarcoding, adapting the protocol of Gogarten et al. (2022). Briefly, a 180 bp mitochondrial fragment of cytochrome oxidase C subunit 1 was directly PCR amplified using the ANML primers (Jusino et al., 2019) in fusion with Illumina adapters (total length ~ 250 bp). The primers were LC1490_adapter (5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTCAACAAATCATAAAGATATTGG-3′) and CO1-CFMRa_Adapter (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGWACTAATCAATTTCCAAATCC-3′). Reaction mixture included 2 μL of DNA extract, 2.5 mM dNTP, 4 mM $MgCl_2$, 10 μM of each fusion primer, 1.25 U Platinum® Taq Polymerase (Invitrogen), and 2.5 μL of 10× PCR Buffer (Invitrogen) in a total volume of 15 μL. Cycling conditions were: 95°C for 5 min, 5 cycles of 94°C for 60 s, 45°C for 90 s, 72°C for 90 s followed by 35 cycles of 94°C for 60 s, 50°C for 90 s, 72°C for 60 s, and a final elongation of 72°C for 7 min. The PCR amplicons were then cleaned up using paramagnetic beads (MagSi NGSPREP Plus beads).

Amplicons from both the fly and mammal metabarcoding from each sample were then pooled and dual-indexed using a Nextera XT Index kit. Briefly, each of the 12.5 μL PCR mixtures contained 1.25 μL of bead-purified PCR amplicons, 6.25 μL 2× KAPA HiFi HotStart ReadyMix (peqLab), 1.25 μL of each Nextera XT index primer (Illumina) and 2.5 μL of nuclease-free water. Cycling conditions were 95°C for 3 min, 8 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and a final elongation step of 72°C for 5 min. The amplicons were further purified using 1.8 μL × MagSi NGSPREP Plus beads to 1 μL amplicon ratio and eluted in a 17.5 μL TET buffer. The indexed amplicon libraries were then quantified using a Qubit 3 and the high sensitivity dsDNA assay kit (Invitrogen) and sequenced on an Illumina Nextseq 500 with a mid-output kit v.2 and 2 × 150 cycles.

2.4 | In silico analyses

2.4.1 | Bcbva genomic analysis

Previously published whole genome sequencing data from Bcbva isolates collected in TNP were downloaded from the Sequence Read Archive (SRA) and converted to fastq files using the sra-toolkit (v. 2.22.0; Leinonen et al., 2011). Fastp (v.0.12.4) was used for read adapter trimming and quality assessment (Chen et al., 2018). Genomic variants were called via the snippy tool (v. 4.6.0; https://github.com/tseemann/snippy), comparing genomes to the Bcbva CI typestrain (GCA_000143605.1) and a variant-site multi fasta alignment was generated using SNP-sites (v.2.2.3; Page et al., 2016). The DNA evolution model was selected via modeltest-ng (v. 0.1.7) and the tree was inferred in a maximum likelihood framework with the help of raxml-ng (v.1.1) starting from 20 random and 20 parsimony trees (Darriba et al., 2020; Kozlov et al., 2019). Branch support values of the best ML tree were evaluated by performing 1000 bootstraps and calculating Transfer Bootstrap Expectation values (Lemoine et al., 2018).

2.4.2 | Mammal metabarcoding

We joined paired-end raw reads using the *illuminapairedend* command of the OBITools package (v1.2.13), setting the minimum alignment score to 40 and removing any non-overlapping reads (Boyer et al., 2016). Primer sequences were removed using the *ngsfilter* command in OBITools and then quality-trimmed with Trimmomatic (v0.36), using a minimum quality score of 30 over a sliding window of four bases, as well as a leading and trailing minimum quality score of 30 and a minimum surviving read length of 80 bp. We then de-replicated the surviving reads using the OBITools *obiuniq* command. We built a reference database using the OBITools *ecoPCR* (v0.2) command to run an in silico PCR on all mammal sequences in GenBank, allowing three mismatches between primers and reference sequences and a synthetic product length between 50 and 800 bp. We then used this database to assign a taxonomy to surviving reads from our experiment using the OBITools *ecotag* command, with a minimum identity level of 0.97. Downstream analyses were based on sequences that were assignable to a mammal species.

Negative controls contained only a single read that could be assigned to the species level, but to conservatively avoid false positives, we only considered a species present when it was detected with at least 10 reads and the species represented at least 0.1% of the total assignable reads for the pool. In addition, the pipeline dropped reads assigned to domestic genera and frequent laboratory contaminants (i.e., those assigned to the genera *Bos*, *Canis*, *Sus*, *Equus*, and *Homo*), as well as reads that were assigned to the genus *Macaca*, for which a closer investigation with blast revealed that these represented a hit to a bacterial artificial chromosome and not the targeted 16 S barcoding region.

2.4.3 | Fly metabarcoding

Primers were removed with cutadapt v2.1 (Martin, 2011) and assigned to amplicon sequence variants (ASVs) using the DADA2 pipeline (Callahan et al., 2016). Sequences trimmed for quality using the filtering parameters (maxN = 0, truncQ = 2, rm.phix = TRUE and maxEE = 2, minLen = 20, trimRight = 1) with the *filterAndTrim* function and then denoised, merged, and chimeras removed with the DADA2 pipeline implemented in R (Callahan et al., 2016). Reads were then assigned to taxa using the eukaryote CO1 reference set v4.0 with the RDP classifier (Porter & Hajibabaei, 2018; Wang et al., 2007). We followed the suggestion of Porter and Hajibabaei (2018) and used bootstrap support cutoff value of 0.6 as this was shown to produce at least 99% correct assignments with barcodes of this length (Porter & Hajibabaei, 2018). To focus on flies, we then considered only ASVs that were assigned to the family Diptera at this threshold. Negative controls contained only two reads that were assigned to the family Diptera, but to conservatively avoid false positives, we considered a ASV present only when it represented at least 5% of the Dipteran reads for a given sample and at least 10 reads. This higher threshold compared to the mammal detections was selected as the fly pools

were known to contain five flies, rather than the trace amounts of mammalian DNA found in and on these flies.

Surviving ASVs were then assigned to molecular operational taxonomic units (MOTUs) by first aligning ASVs with Geneious Prime (v2021.2.2) and creating a phylogeny of the COI barcode using IQTREE (V2.2.0.3), testing for the best fitting model with ModelFinder (N = 55 sequences). Bootstrap support was estimated using the ultra-fast bootstrap procedure with 1000 pseudo-replicates. Subsequent analyses focused on the best fitting model according to BIC (TIM2+F+G4). This tree was then used as the input for a general-mixed Yule/coalescent (GMYC) analysis implemented in the R package splits (Fujisawa & Barraclough, 2013), with the aim of delineating putative species (Pons et al., 2006). This approach required the tree be rooted and ultrametric, which was achieved using the *chronos* function in the R package ape (Paradis & Schliep, 2019) and *midpoint.root* function in the R package phytools (Revell, 2012). The GMYC approach resulted in the delineation of 34 fly MOTUs, which formed the basis of subsequent analyses.

2.4.4 | Statistical analysis and visualizations

We used a Fisher's exact to compare the Bcbva positivity in the three habitat types. To test for a difference in the number of mammal or fly species detected per fly pool in the three habitat types, we used a Generalized Linear Model with a Poisson error structure and log link function (McCullagh & Nelder, 1989). We fit a model including habitat as a fixed effect and fit the model in R using the function glm. The sample for this model comprised 100 fly pools. To establish the significance of the full model, we used a likelihood ratio test (Dobson & Barnett, 2002), comparing its deviance with that of a null model comprising only the intercept.

A satellite image of the sampling region was downloaded from the google api and pie charts were plotted on this image using the *ggmap*, *scatterpie*, and *ggplot2* R packages (Kahle & Wickham, 2013; Wickham, 2016). We constructed species accumulation curves of mammal and fly species diversity using the BiodiversityR, R package to calculate the expected mean species richness for different sample sizes (Kindt & Coe, 2005). We constructed Venn diagrams of species overlap between habitat types using the ggvenn R package (Yan, 2021). To account for differences in the number of fly pools tested in the different habitat types, for the Venn diagrams we downsampled the data to the minimum number of fly pools in a habitat type (N = 25). The maximum likelihood phylogeny was plotted with the help of the ggtree package (Yu, 2020).

All statistical analyses were performed in R v4.2.1 (R Core Team, 2021).

3 | RESULTS

We detected Bcbva in 5 of the 100 fly pools tested. Bcbva positivity of fly pools varied significantly across the three habitat types (forest = 4/25, edge = 1/50, and village = 0/25: Fisher's exact test; P = 0.026; Figure 1). It was possible to culture Bcbva from all five fly pools, confirming their positivity. We generated whole genomes from each of these pools with a chromosomal depth of coverage of at least 114 X.

We generated a total of 2,530,239 paired reads from the pooled 16S mammal metabarcoding and COI fly metabarcoding experiment. Of these paired reads, 1,593,484 survived the obitools processing pipeline for mammals and were assigned a taxonomic rank ($\bar{x}_{\text{per fly pool}} \pm \text{SD} = 15{,}934 \pm 9242$ paired reads). The number of mammalian species per fly pool varied by habitat type ($\chi^2 = 62.2$, df = 2, $p < 0.001$; Figure 2a), with more mammal species detected per fly pool in the forest than either the edge or village ($z = 6.47$, $p < 0.001$); similarly, there were more mammal species in the edge than the village ($z = 3.85$, $p < 0.001$). Species accumulation curves revealed that mammal diversity in the village plateaued well below the species diversity observed in the forest and edge (Figure 2b). In both the edge and the forest, the species accumulation curves had not yet plateaued, suggesting more species are yet to be described in these habitats. The initial rate of accumulation of novel species detections was higher in the forest than the edge habitat. There was considerable overlap in the species detected in the edge and village habitats, while none of the mammal species detected in the village were unique to the village. We detected five species in all habitat types (*Cercopithecus campbelli*, Campbell's monkey; *Cercopithecus diana*, Diana monkey; *Cricetomys* sp., giant pouched rat; *Piliocolobus badius*, western red colobus; *Procolobus verus*, olive colobus) and a number of species were detected in both the edge and the forest, but not the village (*Atherurus africanus*, African brush-tailed porcupine; *Crossarchus* sp.; *Crossarchus obscurus*, common kusimanse; *Nandinia binotata*, African palm civet; *Philantomba maxwellii*, Maxwell's duiker; *Potamochoerus porcus*, red river hog), while *Protoxerus stangeri*, the forest giant squirrel, was the only species detected in both the forest and village, but not the edge. Prior to downsampling, there were eight species detected only in the forest (*Anomalurus beecrofti*, Beecroft's flying squirrel; *Anomalurus derbianus*, Lord Derby's scaly-tailed squirrel; *Cephalophus jentinki*, Jentink's duiker; *Cercopithecus nictitans*, greater spot-nosed monkey; Crossarchus sp.; *Dendrohyrax dorsalis*, western tree hyrax; *Hybomys trivirgatus*, Temminck's striped mouse; *Hyemoschus aquaticus*, water chevrotain), while nine species were only detected in the edge (*Cephalophus* sp.; *Cephalophus dorsalis*, bay duiker; *Cercopithecus petaurista*, lesser spot-nosed monkey; *Colobus polykomos*, king colobus; *Crocidura grandiceps*, long-headed shrew; *Eidolon helvum*, straw-colored fruit bat; *Grammomys* sp.; *Lemniscomys striatus*, Typical striped grass mouse; *Mastomys natalensis*, Natal multimammate mouse; *Mops condylurus*, Angolan free-tailed bat; Figure 2c).

All of the nine mammal species detected in Bcbva-positive fly pools (African brush-tailed porcupine, Campbell's monkey, Diana monkey, forest giant squirrel, giant pouched rat, greater spot-nosed monkey, Maxwell's duiker, olive colobus, western red colobus) were also detected in Bcbva-negative fly pools. Most of these species were not specific to the edge or forest, with the exception of greater

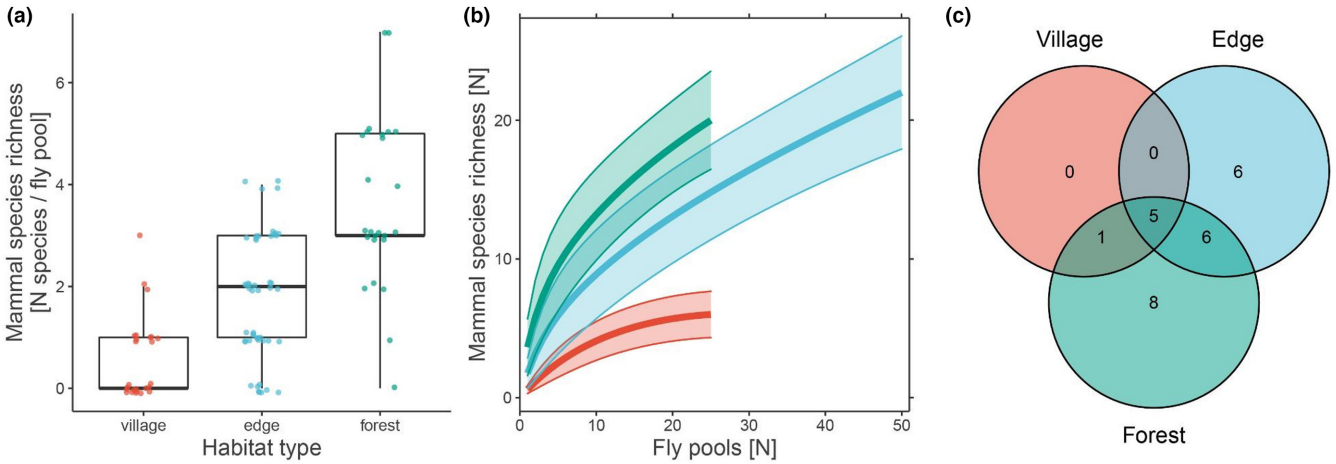

FIGURE 2 (a) A box and whisker plot of the number of mammal species detections per fly pool in the different habitat types. The horizontal line in the whisker plots represents the mean, while the lower and upper bounds of the boxes indicate the first and third quartiles, respectively. The upper and lower whiskers extend to the largest and lowest values no more than 1.5 times the interquartile range away from the hinge. Raw data are plotted as points, with the color corresponding to the habitat type. (b) Mammal species accumulation curves for the three habitat types, with the colors indicating the different habitats as in a and c, with the shade areas indicating the 95% confidence interval. (c) Venn diagrams showing the overlap of the mammal species detected in each of the different habitat types. Data were downsampled for the Venn diagram so that the mammal detections are from the same number of fly pools ($N = 25$) in each habitat type.

spot-nosed monkey and Maxwell's duiker, that were found only in the forest, while the forest giant squirrel was detected in the forest and village. Decades of carcass monitoring in TNP have detected Bcbva in carcasses of *Cephalophus* sp., West African chimpanzees (*Pan troglodytes verus*), Hystricidae, Herpestidae, king colobus, western red colobus, Diana monkeys, lesser spot-nosed monkeys, Campbell's monkeys, and sooty mangabeys (*Cercocebus atys*), which suggest considerable overlap in carcass- and fly-based association of Bcbva and mammal species. Here, the forest giant squirrel, greater spot-nosed monkey, Maxwell's duiker, and giant pouched rat represent potential hosts detected for the first time in Bcbva-positive flies, though are in complete accordance with previous genus-level analyses of mammal detections in Bcbva-positive flies from the forest interior of TNP.

We generated a total of 488,546 paired fly COI metabarcoding reads, of which 445,420 could be assigned a taxonomy with the RDP classifier ($\bar{x}_{\text{per fly pool}} \pm \text{SD} = 4454.18 \pm 2920$ paired reads). Of these, 442,034 were assigned to one of the 34 Dipteran MOTUs ($\bar{x}_{\text{per fly pool}} = 4420 \pm 2915$ paired reads) and 432,065 met the minimum proportion of reads in a sample thresholds ($\bar{x}_{\text{per fly pool}} = 4320 \pm 2869$ paired reads). Fly MOTU detection per fly pool varied significantly by habitat type ($\chi^2 = 7.18$, df = 2, $P = 0.028$; Figure 3a), with more fly MOTUs detected per fly pool in the edge than the village ($z = 2.57$, $P = 0.01$), with no significant difference in the fly MOTUs detected per fly pool in the forest or village. The fly species accumulation curve plateaued in the village habitat suggesting much of the fly MOTU diversity had been described, but in both the edge and forest, the curves were still increasing at the maximum sample size, suggesting that further sampling would continue to reveal novel fly MOTUs in these habitats (Figure 3b). There were four generalist MOTUs that were detected in all habitats; these could be assigned to the families Calliphoridae and Sarcophagidae by the RDP classifier. In addition, there were many novel MOTUs detected only in one habitat (Figure 3b); the 10 MOTUs detected only in the edge belonged to the family Calliphoridae, while the 3 MOTUs detected only in the forest belonged to the family Calliphoridae, and the one MOTU detected only in the village belonged to the family Muscidae (Figure 3c). All six of the fly MOTUs detected in Bcbva-positive fly pools were also detected in Bcbva-negative fly pools and could be assigned to the families Calliphoridae and Sarcophagidae.

The five Bcbva genomes generated here span a considerable portion of known Bcbva diversity in this ecosystem (Figure 4). The one Bcbva detection in a fly captured at the edge revealed a genome identical to one of those sampled in the forest.

4 | DISCUSSION

We used fly iDNA to explore the distribution of Bcbva along a gradient from villages to the forest. At the same time, iDNA generated data about the mammal and fly communities in these habitats, providing insights into the ecology of this pathogen. For both Bcbva and mammal species DNA, the detection rate varied across habitat types. The highest Bcbva detection rate was in the forest, where we also saw the highest mammal species richness and highest mammal detection rate in flies. We did not detect Bcbva in the village habitat, which was also the most mammal and fly species-poor environment. The Bcbva that we detected with iDNA spanned much of the known genomic diversity of this pathogen generated with decades of carcass monitoring (Hoffmann et al., 2017). Similarly, the mammal hosts detected in Bcbva-positive fly pools showed considerable overlap with the mammal species' whose carcasses contained Bcbva

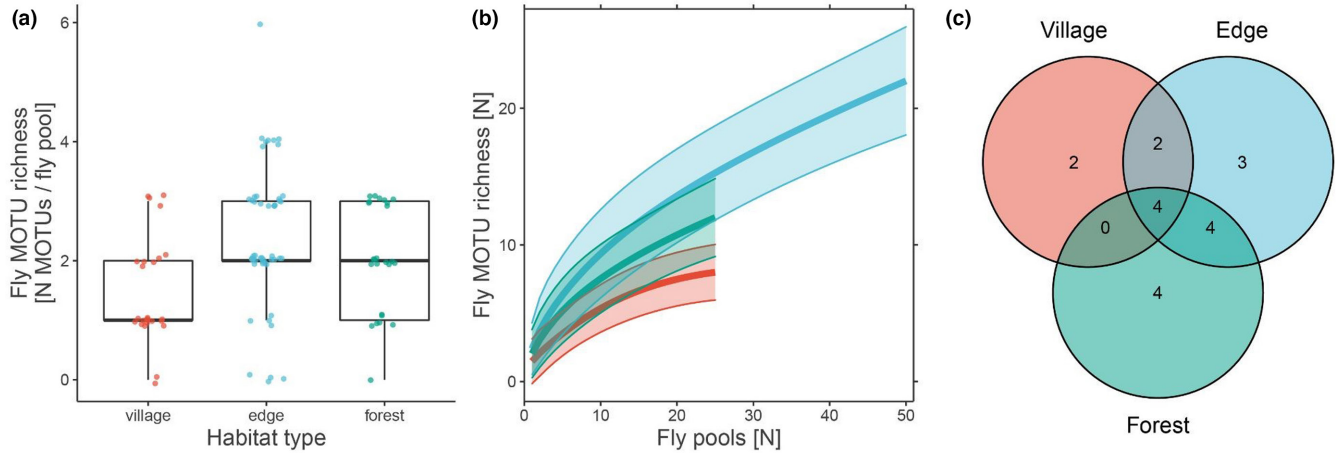

FIGURE 3 (a) A box and whisker plot of the number of fly MOTU detections per fly pool in the different habitat types. The horizontal line in the whisker plots represents the mean, while the lower and upper bounds of the boxes indicate the first and third quartiles, respectively. The upper and lower whiskers extend to the largest and lowest values no more than 1.5 times the interquartile range away from the hinge. Raw data are plotted as points, with the color corresponding to the habitat type. (b) Fly MOTU accumulation curves for the three habitat types, with the colors indicating the different habitats as in a and c, with the shade areas indicating the 95% confidence interval. (c) Venn diagrams showing the overlap of the fly MOTUs detected in each of the different habitat types. Data were downsampled for the Venn diagram so that the fly MOTU detections are from the same number of fly pools ($N = 25$) in each habitat type.

(Hoffmann et al., 2017). Our results highlight the power of iDNA for biomonitoring and pathogen surveillance of emerging and newly described pathogens.

The high Bcbva detection rate in forest flies described here is not unprecedented (4/25 fly pools); indeed previous work in TNP on individual flies, found high Bcbva detection rates in flies forming a long-term association with a group of sooty mangabeys in TNP (~7% of individual flies positive for Bcbva; Gogarten et al., 2019). A broader survey of flies across TNP forest also suggested a high positivity rate in the research area where hunting is rare (~5% of individual flies; Hoffmann et al., 2017). Interestingly, Bcbva positivity varied across the forest habitat and it was more likely to detect Bcbva in flies within the research area than adjoining forest where mammal diversity and biomass was lower because of hunting. In these TNP fly iDNA studies, Bcbva positivity was also more likely in flies that contained more mammal DNA, perhaps indicating recent contact with an animal carcass (Hoffmann et al., 2017). Our findings echo these results, in that Bcbva detection was highest in the forest habitat, where mammal detection rates in fly pools and overall mammal species diversity were highest. We hypothesize that a larger number of mammal species and a higher mammal biomass support more Bcbva mortality in an area, which, in turn, results in more flies coming into contact with Bcbva.

Traditional survey techniques have been used in TNP to monitor populations of certain mammal species and found mammal declines that correlate with signs of hunting (Hoppe-Dominik et al., 2011; Köndgen et al., 2008). While it was hypothesized that the forest edge represents a hostile environment for wild mammals due to ease of access to hunters, this was difficult to quantify in part because animals might also modify their behavior to reduce detection in these hunting areas (Benhaiem et al., 2008). Our iDNA results suggest that mammal biodiversity is considerably lower at the forest edge than a few kilometers away in the forest. Some caution is warranted when interpreting iDNA for biodiversity monitoring, in that detecting the DNA of an animal need not mean the animal was alive in that habitat (i.e., DNA could originate from wildlife carcasses hunted in the forest but processed in the village). Indeed, the detection of non-human primate DNA in village flies might be indicative of the regular consumption of these species in the region (Refisch & Koné, 2005). While camera traps, acoustic monitoring, or dung counts can also be useful for detection of certain species, our findings suggest that iDNA represents a useful tool for describing biodiversity in edge and village ecosystems where mammal monitoring with traditional surveys can be more challenging.

To reduce costs and increase our sample sizes, we used a fly pooling strategy. Pooling has drawbacks when studying pathogen ecology, in particular when looking for mammal host–pathogen detection correlations to assign a pathogen to a particular host or vector species, since it is not possible to make sure that host and pathogen DNA stem from a same individual fly (Alfano et al., 2021; Mwakasungula et al., 2022). This problem is even complicated by potential DNA movements in the collection tube before the analyses, which might increase the frequency of spurious codetection events. These limitations can be compensated with larger sample sizes that enable more detections and allow the use of modeling approaches to identify codetections across multiple pools; this was not feasible here as we only had five Bcbva detections. Despite these challenges, the high degree of overlap between iDNA host species estimates and those generated from carcass monitoring suggests this pooling strategy can generate biologically meaningful results in keeping with the extremely broad host range known for this pathogen (Hoffmann et al., 2017).

In contrast to mammal diversity, fly MOTU diversity was higher at the forest edge than the forest or villages. A number of factors

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

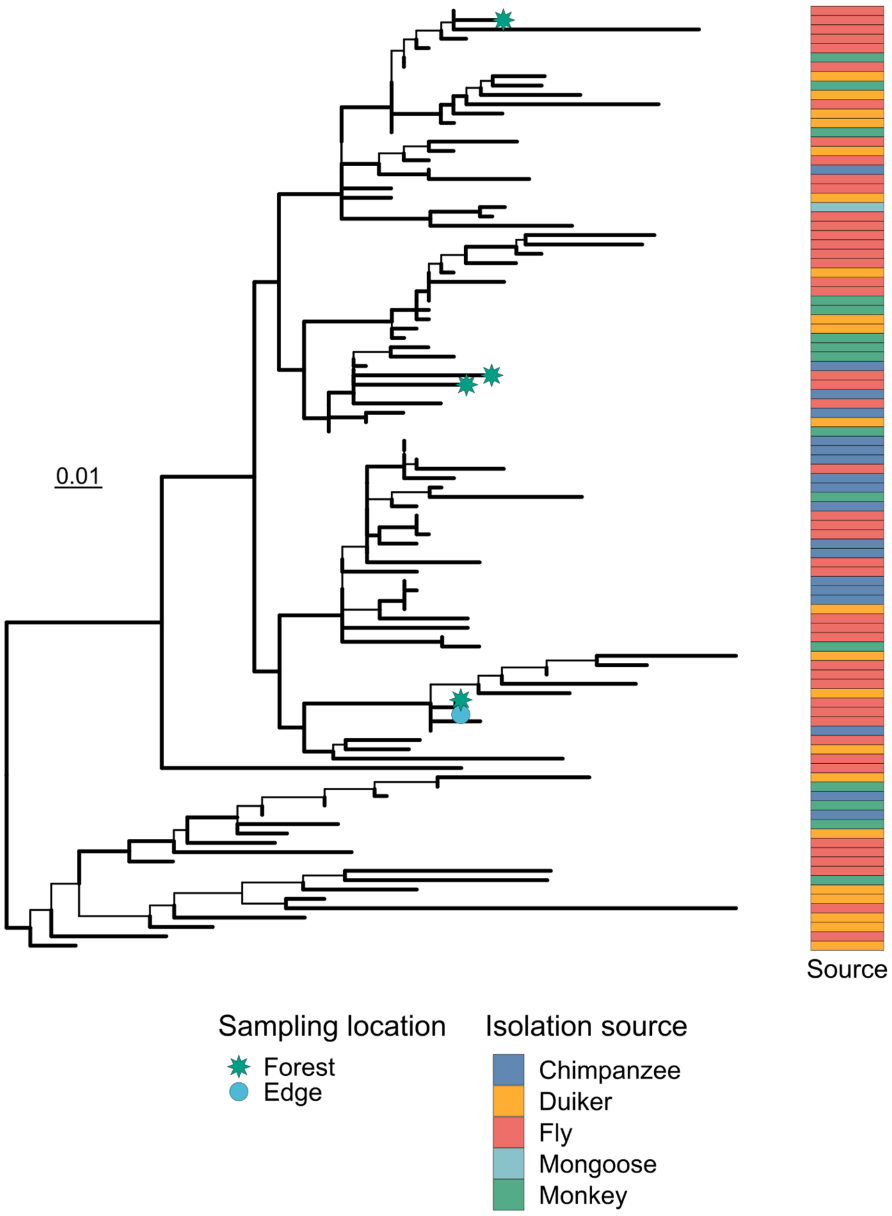

FIGURE 4 Maximum likelihood phylogeny of Bcbva chromosomes collected in TNP. The tree is midpoint rooted and internal branches with transfer bootstrap values below 0.7 are displayed in non-bold. Tip points are displayed for the sequences generated within this study, whereas the points color and shape represent the fly sampling location. The isolation sources of all sequences are illustrated by a colored strip. Substitutions per site are represented by the scale bar.

might be involved in driving this pattern, but the habitat heterogeneity at the edge may support the presence of both more anthropogenic flies, those that thrive in more savannah like ecosystems, as well as forest specialists. This is in keeping with broader studies of invertebrate biodiversity at edges, that suggest even sharp habitat edges as defined by major changes in soils and plants do not serve as hard edges but broad transition zones for many invertebrate taxa (Dangerfield et al., 2003). Indeed, forest edges in a tropical forest in Indonesia showed higher levels of insect diversity than the forest interior (Darsono et al., 2020). Whether particular fly species are more likely to carry Bcbva is unclear, but the variation in fly species community composition might be linked to the variation in Bcbva detection rates across habitats.

Flies have been implicated as vectors for a number of pathogens, including *Bacillus anthracis*, the causative agent of classic anthrax (Blackburn et al., 2010; Greenberg, 1971; Turell & Knudson, 1987). Our finding of a Bcbva-positive fly pool at the forest edge, in areas frequently used by people, may represent a route of exposure to Bcbva. Flies can travel large distances, at the scales examined here; for example, a mark–recapture experiment in TNP showed that flies moved at least 1.3 km in the forest with a monkey group (Gogarten et al., 2019). Similarly, flies can move at least a few 100 m across the forest edge into village areas around Kibale National Park, Uganda suggesting they may serve as mechanical vector between these ecosystems (Jahan et al., 2023). While the potential mobility of flies may impact the scale at which pathogen and mammal iDNA-based monitoring is meaningful, the biological signal extracted from the mammal and fly diversity estimates across these habitats suggests these fly populations are not a homogeneous population mixture at this spatial scale.

We detected an identical Bcbva isolate in a fly captured in the forest and edge. We previously described Bcbva diversity within carcasses and found that isolates differed by a maximum of two chromosomal single nucleotide polymorphisms (Hoffmann et al., 2017), suggesting these flies were likely exposed at the same carcass or

epidemiologically linked carcasses. The lack of Bcbva detection in village areas suggests that Bcbva is not causing mammal mortality in this habitat at the scale seen in neighboring forests. While the seropositivity of people to Bcbva shows that exposure happens in this region, how this exposure happens is unclear and determining the impact of Bcbva on people and their livestock represents an important area of research.

Our results demonstrate the power of iDNA for biomonitoring and pathogen surveillance along gradients of anthropogenic disturbance. Fly iDNA revealed the spatial scope and host range of Bcbva at the forest edge and surrounding villages and supported a strict reliance of this bacterium on the rainforest ecosystem.

AUTHOR CONTRIBUTIONS

MJ, LL, TG, ECH, FHL, SCS, and JFG contributed to the conception or design of the study and MJ, LL, TG, SCS, and JFG prepared the first draft of the manuscript. All authors contributed to the acquisition, analysis, or interpretation of the data, as well as contributed to editing and shaping the final manuscript.

ACKNOWLEDGMENTS

For their support and permission to conduct this study, we would like to thank the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, the Centre de Recherche en Ecologie, and the Ministère des Eaux et Fôrets in Côte d'Ivoire, as well as the Office Ivoirien des Parcs et Réserves. We also thank the Sequencing Unit of the Robert Koch Institute for their assistance with this project. We are grateful to the Centre Suisse de Recherches Scientifiques en Côte d'Ivoire for their support in organizing and actualizing this project. Work in the Côte d'Ivoire was supported by the EU-Biodiversa BIODIV-AFREID project (LE 1813/17-1: Project number: 428839112). MJ was funded by the Bangabandhu Science and Technology Fellowship Trust, Ministry of Science and Technology, Government of the People's Republic of Bangladesh. JFG was supported by the DFG grant "The ecology and evolution of primate phageomes" (GO 3443/1-1: Project number: 453352748). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Next-generation sequencing reads used for the reconstruction of Bcbva genomes are available through the short read archive (project accession number: PRJNA940060). All additional metabarcoding sequences generated for this study have been uploaded to Zenodo, with the logic that assignment to a particular organism is always uncertain and should not supersede existing sequences linked to a particular host in available sequence databases (doi: 10.5281/zenodo.7688127).

ORCID

Mueena Jahan https://orcid.org/0000-0003-1224-9615
Lorenzo Lagostina https://orcid.org/0000-0002-6661-705X
Tobias Gräßle https://orcid.org/0000-0001-9103-3139
Emmanuel Couacy-Hymann https://orcid.org/0000-0003-2593-6932
Valère K. Kouakou https://orcid.org/0000-0002-5024-1721
Livia V. Patrono https://orcid.org/0000-0003-3978-3607
Kamilla Pléh https://orcid.org/0000-0002-2528-5482
Fabian H. Leendertz https://orcid.org/0000-0002-2169-7375
Sébastien Calvignac-Spencer https://orcid.org/0000-0003-4834-0509
Jan F. Gogarten https://orcid.org/0000-0003-1889-4113

REFERENCES

Alfano, N., Dayaram, A., Axtner, J., Tsangaras, K., Kampmann, M. L., Mohamed, A., Wong, S. T., Gilbert, M. T. P., Wilting, A., & Greenwood, A. D. (2021). Non-invasive surveys of mammalian viruses using environmental DNA. *Methods in Ecology and Evolution*, *12*(10), 1941–1952. https://doi.org/10.1111/2041-210X.13661

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), 1–20. https://doi.org/10.1371/journal.pone.0176343

Benhaiem, S., Delon, M., Lourtet, B., Cargnelutti, B., Aulagnier, S., Hewison, A. J. M., Morellet, N., & Verheyden, H. (2008). Hunting increases vigilance levels in roe deer and modifies feeding site selection. *Animal Behaviour*, *76*(3), 611–618. https://doi.org/10.1016/j.anbehav.2008.03.012

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*, 1–21. https://doi.org/10.7554/eLife.22069

Blackburn, J. K., Curtis, A., Hadfield, T. L., O'Shea, B., Mitchell, M. A., & Hugh-Jones, M. E. (2010). Confirmation of *bacillus anthracis* from flesh-eating flies collected during a West Texas anthrax season. *Journal of Wildlife Diseases*, *46*(3), 918–922. https://doi.org/10.7589/0090-3558-46.3.918

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, *29*(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). OBITOOLS: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176–182. https://doi.org/10.1111/1755-0998.12428

Butchart, S. H. M., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J. P. W., Almond, R. E. A., Baillie, J. E. M., 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

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

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. https://doi.org/10.1111/mec.12183

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Dangerfield, J. M., Pik, A. J., Britton, D., Holmes, A., Gillings, M., Oliver, I., Briscoe, D., & Beattie, A. J. (2003). Patterns of invertebrate biodiversity across a natural edge. *Austral Ecology*, *28*(3), 227–236. https://doi.org/10.1046/j.1442-9993.2003.01240.x

Darsono, R., E, Santoso, S., Sudiana, E., Nasution, E. K., Aprilliana, H., & Chasanah, T. (2020). Insect diversity in various distances to forest edge in small nature reserve: A case study of BantarbolangNature Reserve, Central Java, Indonesia. *Biodiversitas*, *21*(10), 4821–4828.

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution*, *37*(1), 291–294. https://doi.org/10.1093/molbev/msz189

Dobson, A. J., & Barnett, A. G. (2002). *An introduction to generalized linear models* (Second ed.). CRC Press LLC.

Dupke, S., Schubert, G., Beudjé, F., Barduhn, A., Pauly, M., Couacy-Hymann, E., Grunow, R., Akoua-Koffi, C., Leendertz, F., & Klee, S. (2020). Serological evidence for human exposure to *Bacillus cereus* biovar *anthracis* in the villages around Taï National Park, Côte D'ivoire. *PLoS Neglected Tropical Diseases*, *14*(5), e0008292. https://doi.org/10.1371/journal.pntd.0008292.t001

Ellerbrok, H., Nattermann, H., Özel, M., Beutin, L., Appel, B., & Pauli, G. (2002). Rapid and sensitive identification of pathogenic and apathogenic *bacillus anthracis* by real-time PCR. *FEMS Microbiology Letters*, *214*(1), 51–59. https://doi.org/10.1016/S0378-1097(02)00837-6

Fasanella, A., Scasciamacchia, S., Garofolo, G., Giangaspero, A., Tarsitano, E., & Adone, R. (2010). Evaluation of the house fly *Musca domestica* as a mechanical vector for an anthrax. *PLoS One*, *5*(8), 4–8. https://doi.org/10.1371/journal.pone.0012219

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. *Systematic Biology*, *62*(5), 707–724. https://doi.org/10.1093/sysbio/syt033

Gilbert, N. A., Clare, J. D. J., Stenglein, J. L., & Zuckerberg, B. (2021). Abundance estimation of unmarked animals based on camera-trap data. *Conservation Biology*, *35*(1), 88–100. https://doi.org/10.1111/cobi.13517

Gogarten, J. F., Calvignac-Spencer, S., Nunn, C. L., Ulrich, M., Saiepour, N., Nielsen, H. V., Deschner, T., Fichtel, C., Kappeler, P. M., Knauf, S., Müller-Klein, N., Ostner, J., Robbins, M. M., Sangmaneedet, S., Schülke, O., Surbeck, M., Wittig, R. M., Sliwa, A., Strube, C., … Noll, A. (2020). Metabarcoding of eukaryotic parasite communities describes diverse parasite assemblages spanning the primate phylogeny. *Molecular Ecology Resources*, *20*(1), 204–215. https://doi.org/10.1111/1755-0998.13101

Gogarten, J. F., Düx, A., Mubemba, B., Pléh, K., Hoffmann, C., Mielke, A., Müller-Tiburtius, J., Sachse, A., Wittig, R. M., Calvignac-Spencer, S., & Leendertz, F. H. (2019). Tropical rainforest flies carrying pathogens form stable associations with social nonhuman primates. *Molecular Ecology*, *28*(18), 4242–4258. https://doi.org/10.1111/mec.15145

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

Greenberg, B. (1971). *Flies and disease, Vol. 1: Ecology, classification and biotic associations; flies and disease. Volume II. Biology and disease transmission*. Princeton University Press.

Hale, V. L., Dennis, P. M., McBride, D. S., Nolting, J. M., Madden, C., Huey, D., Ehrlich, M., Grieser, J., Winston, J., Lombardi, D., Gibson, S., Saif, L., Killian, M. L., Lantz, K., Tell, R. M., Torchetti, M., Robbe-Austerman, S., Nelson, M. I., Faith, S. A., & Bowman, A. S. (2022). SARS-CoV-2 infection in free-ranging white-tailed deer. *Nature*, *602*(7897), 481–486. https://doi.org/10.1038/s41586-021-04353-x

Hockings, K. J., Mubemba, B., Avanzi, C., Pleh, K., Düx, A., Bersacola, E., Bessa, J., Ramon, M., Metzger, S., Patrono, L. V., Jaffe, J. E., Benjak, A., Bonneaud, C., Busso, P., Couacy-Hymann, E., Gado, M., Gagneux, S., Johnson, R. C., Kodio, M., … Leendertz, F. H. (2021). Leprosy in wild chimpanzees. *Nature*, *598*(7882), 652–656. https://doi.org/10.1038/s41586-021-03968-4

Hoffmann, C., Merkel, K., Sachse, A., Rodríguez, P., Leendertz, F. H., & Calvignac-Spencer, S. (2018). Blow flies as urban wildlife sensors. *Molecular Ecology Resources*, *18*(3), 502–510. https://doi.org/10.1111/1755-0998.12754

Hoffmann, C., 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), 82–85. https://doi.org/10.1038/nature23309

Hoppe-Dominik, B., Kühl, H. S., Radl, G., & Fischer, F. (2011). Long-term monitoring of large rainforest mammals in the biosphere Reserve of Taï National Park, Côte d'Ivoire. *African Journal of Ecology*, *49*(4), 450–458. https://doi.org/10.1111/j.1365-2028.2011.01277.x

Jahan, M., Calvignac-Spencer, S., Chapman, C. A., Kalbitzer, U., Leendertz, F. H., Omeja, P. A., Sarkar, D., Ulrich, M., & Gogarten, J. F. (2023). The movement of pathogen carrying flies at the human–wildlife interface. *EcoHealth*, *19*, 450–457. https://doi.org/10.1007/s10393-022-01621-8

Jusino, M. A., Banik, M. T., Palmer, J. M., Wray, A. K., Xiao, L., Pelton, E., Barber, J. R., Kawahara, A. Y., Gratton, C., Peery, M. Z., & Lindner, D. L. (2019). An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Molecular Ecology Resources*, *19*(1), 176–190. https://doi.org/10.1111/1755-0998.12951

Kahle, D., & Wickham, H. (2013). Ggmap: Spatial visualization with ggplot2. *R Journal*, *5*(1), 144–161. https://doi.org/10.32614/rj-2013-014

Kindt, R., & Coe, R. (2005). Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. In *World Agroforestry Centre (ICRAF)*. World Agroforestry Centre (ICRAF).

Klee, S. R., Brzuszkiewicz, E. B., Nattermann, H., Brüggemann, H., Dupke, S., Wollherr, A., Franz, T., Pauli, G., Appel, B., Liebl, W., Couacy-Hymann, E., Boesch, C., Meyer, F. D., Leendertz, F. H., Ellerbrok, H., Gottschalk, G., Grunow, R., & Liesegang, H. (2010). The genome of a *bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids. *PLoS One*, *5*(7), e10986. https://doi.org/10.1371/journal.pone.0010986

Köndgen, S., Kühl, H., N'Goran, P. K., Walsh, P. D., Schenk, S., Ernst, N., Biek, R., Formenty, P., Mätz-Rensing, K., Schweiger, B., Junglen, S., Ellerbrok, H., Nitsche, A., Briese, T., Lipkin, W. I., Pauli, G., Boesch, C., & Leendertz, F. H. (2008). Pandemic human viruses cause decline of endangered great apes. *Current Biology*, *18*(4), 260–264. https://doi.org/10.1016/j.cub.2008.01.012

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Kuisma, E., Olson, S. H., Cameron, K. N., Reed, P. E., Karesh, W. B., Ondzie, A. I., Akongo, M. J., Kaba, S. D., Fischer, R. J., Seifert, S.

N., Munõz-Fontela, C., Becker-Ziaja, B., Escudero-Pérez, B., Goma-Nkoua, C., Munster, V. J., & Mombouli, J. V. (2019). Long-term wildlife mortality surveillance in northern Congo: A model for the detection of Ebola virus disease epizootics. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *374*(1782), 1–8. https://doi.org/10.1098/rstb.2018.0339

Leendertz, F. H., Ellerbrok, H., Boesch, C., Couacy-Hymann, E., Mätz-Rensing, K., Hakenbeck, R., Bergmann, C., Abaza, P., Junglen, S., Moebius, Y., Vigilant, L., Formenty, P., & Pauli, G. (2004). Anthrax kills wild chimpanzees in a tropical rainforest. *Nature*, *430*(6998), 451–452. https://doi.org/10.1038/nature02722

Leinonen, R., Sugawara, H., & Shumway, M. (2011). The sequence read archive. *Nucleic Acids Research*, *39*(SUPPL. 1), 2010–2012. https://doi.org/10.1093/nar/gkq1019

Lemoine, F., Domelevo Entfellner, J. B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., & Gascuel, O. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*, *556*(7702), 452–456. https://doi.org/10.1038/s41586-018-0043-0

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*(3), 701–707.e5. https://doi.org/10.1016/j.cub.2021.12.014

Martin, M. (2011). CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*, 10–12. https://doi.org/10.14806/ej.17.1.200

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed.). Chapman and Hall. https://doi.org/10.1007/978-1-4899-3242-6

Mombo, I. M., Fritz, M., Becquart, P., Liegeois, F., Elguero, E., Boundenga, L., Mebaley, T. N., Prugnolle, F., Maganga, G. D., & Leroy, E. M. (2020). Detection of ebola virus antibodies in fecal samples of great apes in Gabon. *Viruses*, *12*(12), 1–10. https://doi.org/10.3390/v12121347

Morin, D. J., Kelly, M. J., & Waits, L. P. (2016). Monitoring coyote population dynamics with fecal DNA and spatial capture–recapture. *Journal of Wildlife Management*, *80*(5), 824–836. https://doi.org/10.1002/jwmg.21080

Mwakasungula, S., Rougeron, V., Arnathau, C., Boundenga, L., Miguel, E., Boissière, A., Jiolle, D., Durand, P., Msigwa, A., Mswata, S., Olotu, A., Sterkers, Y., Roche, B., Killeen, G., Cerqueira, F., Bitome-Essono, P. Y., Bretagnolle, F., Masanja, H., Paupy, C., … Prugnolle, F. (2022). Using haematophagous fly blood meals to study the diversity of blood-borne pathogens infecting wild mammals. *Molecular Ecology Resources*, *0–13*, 2915–2927. https://doi.org/10.1111/1755-0998.13670

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., & Harris, S. R. (2016). SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*, *2*(4), e000056. https://doi.org/10.1099/mgen.0.000056

Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., & Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, *55*(4), 595–609. https://doi.org/10.1080/10635150600852011

Porter, T. M., & Hajibabaei, M. (2018). Automated high throughput animal CO1 metabarcode classification. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-22505-4

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Refisch, J., & Koné, I. (2005). Impact of commercial hunting on monkey populations in the Taï region, Côte D'ivoire. *Biotropica*, *37*(1), 136–144. https://doi.org/10.1111/j.1744-7429.2005.03174.x

Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rodgers, T. W., Xu, C. C. Y., Giacalone, J., Kapheim, K. M., Saltonstall, K., Vargas, M., Yu, D. W., 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. https://doi.org/10.1111/1755-0998.12701

Santiago, M. L., Rodenburg, C. M., Kamenya, S., Bibollet-Ruche, F., Gao, F., Bailes, E., Meleth, S., Soong, S. J., Kilby, J. M., Moldoveanu, Z., Fahey, B., Muller, M. N., Ayouba, A., Nerrienet, E., McClure, H. M., Heeney, J. L., Pusey, A. E., Collins, D. A., Boesch, C., … Hahn, B. H. (2002). SIVcpz in wild chimpanzees. *Science*, *295*(5554), 465. https://doi.org/10.1126/science.295.5554.465

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

Stat, M., Huggett, M. J., Bernasconi, R., Dibattista, J. D., Berry, T. E., Newman, S. J., Harvey, E. S., & Bunce, M. (2017). Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-12501-5

Stegen, G., Pasmans, F., Schmidt, B. R., Rouffaer, L. O., Van Praet, S., Schaub, M., Canessa, S., Laudelout, A., Kinet, T., Adriaensen, C., Haesebrouck, F., Bert, W., Bossuyt, F., & Martel, A. (2017). Drivers of salamander extirpation mediated by *Batrachochytrium salamandrivorans*. *Nature*, *544*(7650), 353–356. https://doi.org/10.1038/nature22059

Turell, M. J., & Knudson, G. B. (1987). Mechanical transmission of *bacillus anthracis* by stable flies (*Stomoxys calcitrans*) and mosquitoes (*Aedes aegypti* and *Aedes taeniorhynchus*). *Infection and Immunity*, *55*(8), 1859–1861. https://doi.org/10.1128/iai.55.8.1859-1861.1987

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-98141-3

Yan, L. (2021). ggvenn: Draw Venn Diagram by "ggplot2". R Package Version 0.1.9. https://CRAN.R-project.org/package=ggvenn

Yu, G. (2020). Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics*, *69*(1), 1–18. https://doi.org/10.1002/cpbi.96

How to cite this article: Jahan, M., Lagostina, L., Gräßle, T., Couacy-Hymann, E., Kouadio, L., Kouakou, V. K., Krou, H. A., Mossoun, A. M., Patrono, L. V., Pléh, K., Steiner, J. A., Yves, N., Leendertz, F. H., Calvignac-Spencer, S., & Gogarten, J. F. (2024). Fly iDNA suggests strict reliance of the causative agent of sylvatic anthrax on rainforest ecosystems. *Environmental DNA*, *6*, e401. https://doi.org/10.1002/edn3.401