Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest

Authors: Constanze Hoffmann[1,†], Fee Zimmermann[1,2,†], Roman Biek[3], Hjalmar Kuehl[4], Kathrin Nowak[1], Roger Mundry[4], Anthony Agbor[4], Samuel Angedakin[4], Mimi Arandjelovic[4], Anja Blankenburg[1], Gregory Brazolla[4], Katherine Corogenes[4], Emmanuel Couacy-Hymann[5], Tobias Deschner[4], Paula Dieguez[4], Karsten Dierks[4], Ariane Düx[1], Susann Dupke[2], Henk Eshuis[4], Pierre Formenty[6], Yisa Ginath Yuh[4], Annemarie Goedmakers[7], Jan F. Gogarten[1,4,8], Anne-Céline Granjon[4], Scott McGraw[9], Roland Grunow[2], John Hart[10], Sorrel Jones[4], Jessica Junker[4], John Kiang[11], Kevin Langergraber[12], Juan Lapuente[4], Kevin Lee[4], Siv Aina Leendertz[1], Floraine Léguillon[1], Vera Leinert[13], Therese Löhrich[1,4], Sergio Marrocoli[4], Kerstin Mätz-Rensing[14], Amelia Meier[4], Kevin Merkel[1], Sonja Metzger[1], Mizuki Murai[4], Svenja Niedorf[1], Hélène De Nys[1,4], Andreas Sachse[1], Joost van Schijndel[4], Ulla Thiesen[1], Els Ton[7], Doris Wu[1,4], Lothar H. Wieler[15], Christophe Boesch[4], Silke R. Klee[2], Roman M. Wittig[4], Sébastien Calvignac-Spencer[1], Fabian H. Leendertz[1,*]

[†]Equal contributions

*Corresponding author. E-mail: LeendertzF@rki.de

Affiliations:

[1]Robert Koch Institute, P3: "Epidemiology of Highly Pathogenic Microorganisms", Seestraße 10-11, 13353 Berlin, Germany.

[2]Robert Koch Institute, ZBS 2: Centre for Biological Threats and Special Pathogens, Highly Pathogenic Microorganisms, Seestraße 10-11, 13353 Berlin, Germany.

[3]Institute of Biodiversity, Animal Health and Comparative Medicine, Boyd Orr Centre for Population and Ecosystem Health, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom.

[4]Max Planck Institute for Evolutionary Anthropology (MPI EVAN), Deutscher Platz 6, 04103 Leipzig, Germany.

[5]LANADA/LCVB, Bingerville, 206, Côte d'Ivoire.

[6]World Health Organization, 1211 Geneva 27, Switzerland.

[7]Chimbo Foundation, Amstel 49, 1011 PW Amsterdam, Netherlands.

[8]McGill University, Department of Biology, 855 Sherbrooke Street West Montreal, QC, Canada, H3A 2T7.

[9]The Ohio State University, Department of Anthropology, 4034 Smith Laboratory, 174 W. 18th Avenue, Columbus, OH, USA.

[10]Lukuru Foundation, 1235 Avenue des Poids Lourds / Quartier de Kingabois, Kinshasa, DRC.

[11]Limbe Wildlife Centre, Limbe, Cameroon.

[12]Arizona State University, PO Box 872402, Tempe, AZ 85287-2402 USA.

[13]Wild Chimpanzee Foundation (WCF), Deutscher Platz 6, 04103 Leipzig, Germany.

[14]German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany.

[15]Robert Koch Institute, Seestraße 10-11, 13353 Berlin, Germany.

Anthrax is a globally significant animal disease and zoonosis. Despite this, current knowledge of anthrax ecology is largely limited to arid ecosystems, where outbreaks are most commonly reported[1-3]. We reveal the dynamics of an anthrax causing agent, *Bacillus cereus* biovar *anthracis*, in a tropical rainforest with severe consequences for local wildlife communities. Using data and samples collected over three decades we find that rainforest anthrax is a persistent and widespread cause of death for a broad range of mammalian hosts. We predict that this pathogen will accelerate the decline and possibly result in the extirpation of local chimpanzee (*Pan troglodytes verus*) populations. Our findings illuminate the epidemiology of a cryptic pathogen and have important implications for conservation.

Anthrax is a disease of wildlife, livestock and humans predominantly affecting low and middle-income countries[2,4,5]. Although widely distributed, including some temperate regions, anthrax is most commonly associated with arid ecosystems, particularly African

savannas[1,3,6-11]. In these systems, major outbreaks typically cause high mortality in a few wild and domestic ungulate species at a time and usually exhibit strong seasonal and inter-annual variation[2,3,5,11,12]. For example, in Krüger National Park, South Africa, die-offs in kudus (*Tragelaphus strepsiceros*) and impalas (*Aepyceros melampus*) occur in the dry season with a ten year periodicity coinciding with rainfall cycles[11]. In Etosha National Park, Namibia, mortality in elephants (*Loxodonta Africana*) peaks at the start of the wet season, while plains ungulates (*Equus quagga, Conochaetes taurineus, Antidorcas marsupialis*) are most affected at the end of the wet season[3,13]. Such varying dynamics underline the importance of investigating the pathogen in close relation with its ecosystem, but so far anthrax research in Africa has been biased towards well-studied savanna regions.

In 2001, lethal anthrax-like cases in wild chimpanzees were reported in a rainforest habitat: Taï National Park (TNP), Côte d'Ivoire (Fig. S1)[14]. The causative agent was a bacterium combining the chromosomal background of *Bacillus cereus* with the virulence plasmids of *B. anthracis* (*Bacillus cereus* biovar *anthracis; Bcbva)*[15]. Pathology and histopathology of *Bcbva* cases were clearly suggestive of anthrax and in small animal models *Bcbva* was as virulent as *B. anthracis*[14-16]. *Bcbva* cases have since been described in animals in Cameroon (CM), Central African Republic (CAR) and the Democratic Republic of Congo[17,18], suggesting a broad sub-Saharan distribution (Fig. 1). However, the epidemiology of anthrax-like disease caused by *Bcbva* (hereafter anthrax), and to what extent it matches that of classical anthrax, remain poorly understood.

We address this knowledge gap by testing a unique set of samples collected in TNP over 26 years. We started collecting bones in 1989 resulting in bones from 75 individual

mammals (Table S7, Supplementary information S4). From 1996 on, we investigated 204 fresh carcasses (Table S2, Supplementary information S2). Since bone and carcass discovery was linked to the collection of chimpanzee behavioral data, we expected detection of *Bcbva* to be biased towards chimpanzees and other easily detectable medium to large-bodied mammals. We therefore tested whether carrion flies, which are relatively unbiased samplers of mammalian DNA[19], might also collect *Bcbva* or its genetic material while feeding and ovipositing on carcasses. Starting in 2008, we applied different horizontal and vertical sampling schemes to collect 1,634 flies (Table S1 and S4, Supplementary information S3). We retrieved *Bcbva* isolates from all three sample types (bones, carcasses, flies). These allowed us to generate 178 whole genome sequences spanning from 1996 to 2014 (Table S8). To clarify the distribution of *Bcbva* on a larger scale, we sampled 1089 flies and 136 bones from 16 other sites in 11 sub-Saharan countries from 2012 to 2014 (Fig. 1, Table S1).

In TNP we detected *Bcbva* DNA in 81 carcasses (40%; Fig. 2A, Extended Data Fig. S1, Extended Data Fig. S2, Table S2), 26 bones (35%, Table S7) and 80 flies (5%; Fig. 2B, Extended Data Fig. S3, Table S4). We could perform histopathological examinations on 15 positive carcasses and in all cases pathology was consistent with a lethal anthrax infection (Table S2). Overall, 38% of observed local wildlife mortality was associated with *Bc*bv*a* (Tables S2 and S4), meeting the highest levels of mortality reported for classical anthrax outbreaks in savanna ecosystems [12,20]. We observed no obvious seasonal variation in *Bcbva* carcass incidence, suggesting ongoing anthrax activity in the area (Generalized Linear Mixed Model (GLMM), $\chi 2=6.3$, df=10, P=0.789, Supplementary information S8a). However, *Bcbva* detection in flies peaked from December to March,

117 coinciding with the only distinct dry period in the park (GLMM, $\chi 2$=6.9, df=2, P=0.032,
118 Extended Data Fig. S4, Supplementary information S8b). This suggests climatic
119 conditions may influence *Bcbva* ecology in TNP, similar to observations from *B.*
120 *anthracis* in savannas[1], though seasonal mortality appears less pronounced.

121 *Bcbva* differed dramatically from *B. anthracis* in terms of host range. Ungulates
122 constitute the vast majority (> 99 %) of anthrax cases in savanna ecosystems[11,12,20]. In
123 contrast, and in line with the more diverse fauna found in rainforests, we observed *Bcbva*
124 fatalities in a broader range of species in TNP, including chimpanzees (31/55), six
125 monkey species (21/81), duikers (26/40), mongooses (2/2) and porcupines (1/26 other
126 mammals) (Table S2). To further explore the host range of *Bcbva*, we analyzed the gut
127 content of all mammal and *Bcbva* positive flies (n=28, Table S1) using amplicon deep
128 sequencing. We detected sequences from most of the aforementioned species, and from
129 species belonging to 11 further mammalian genera, including carnivores, rodents and bats
130 (Table S5, Supplementary information S3e). This suggests that *Bcbva* may affect an even
131 broader range of mammals than inferred from carcass monitoring alone. Further, meal
132 compositions of mammal positive *Bcbva* positive flies (n=28) and mammal positive
133 *Bcbva* negative flies (n=29) did not differ significantly (GLMMs, Supplementary
134 information S8c), which may support the notion that there is no substantial difference in
135 *Bcbva* susceptibility among species.

136 To gain further insight into the ecology of *Bcbva*, we investigated 178 genomes derived
137 from isolates obtained from necropsy samples, bones and flies, collected between 1996
138 and 2014 (Table S8). Considering 126 chromosomal sequences originating from separate
139 hosts (mammals and flies) we detected 298 single nucleotide polymorphisms (SNP).

140 Plasmids contained negligible amounts of variation (Supplementary information S7a).
141 The maximum distance observed between isolates was 69 SNPs (median: 26 SNPs); the
142 most distant isolates originated in flies caught in two consecutive years only 6 km apart.
143 In comparison, a maximum distance of only 20 SNPs was observed in *B. anthracis*
144 isolates derived from cattle samples collected in the French Alps between 1997 and
145 2009[21]. The high genetic diversity observed in TNP is consistent with extensive *Bcbva*
146 activity in the area and suggests that this pathogen did not emerge recently (Fig.3,
147 Extended Data Fig. S5). In addition, considerably more divergence was seen compared to
148 isolates from other countries[17,18], supporting the notion that *Bcbva* has been circulating in
149 sub-Saharan Africa for an even much longer period than what we determined in TNP
150 (Extended Data Fig. S6, Supplementary information S7).To assess within-host diversity
151 we sequenced the genomes of two to six independent isolates for a subset of carcasses
152 and flies (Table S9). Two strains differing by 42 chromosomal SNPs were isolated from a
153 single fly, likely reflecting multiple carcass meals[19], which further highlights the
154 commonness of *Bcbva* in TNP. Otherwise, the maximum distance observed within one
155 host was two chromosomal SNPs (mean: 0.35 SNPs). Within-host heterogeneity thus
156 seems negligible compared to the overall diversity observed for *Bcbva* suggesting strains
157 differing by more than two SNPs originate from separate carcasses.

158 *Bcbva* positive carcasses were broadly distributed throughout the TNP research area,
159 without the kind of geographic clustering described for anthrax in savanna
160 ecosystems[12,22] (Fig. 2A). We determined *Bcbva* prevalence within and outside the
161 research area using a subset of 908 flies caught systematically according a grid system
162 within 19 days (Extended Data Fig. S7). We detected *Bcbva* positive flies in 16/83 traps

(Additional Data Table S1). Prevalence was higher in the research area (8/21 traps *Bcbva* positive) than in the surrounding forest belt (8/62 traps *Bcbva* positive) (Fisher´s Exact Test, $P = 0.02$). Long-term research activity within the TNP research area has had a protective effect on wildlife and led to an increased density of mammals[23], which might explain higher *Bcbva* activity. Genome data revealed multiple contemporaneous transmission chains caused by co-circulating strains (2 to 48 SNPs distance, median: 25 SNPs) in different areas of the park over the short time period of the fly snapshot (Extended Data Fig. S8). For low genomic distances (≤ 35 SNPs), genomic and geographic distances of all TNP isolates were positively correlated ($R^2 = 0.72$), providing further indication of spatially restricted transmission (Extended Data Fig. S9), which might reflect carcass-mediated spread of *Bcbva*. Since wildlife cases included exclusively arboreal monkeys (Table S2), we explored the vertical distribution of *Bcbva* by catching flies simultaneously on the ground and up to 30m into the canopy. We detected *Bcbva* in 12 of 103 canopy flies (11.7%) and retrieved isolates from five of these (Table S4, Additional Data Table S1). While on the ground carcass deposition sites are likely to be the source of *Bcbva* infections, flies may contribute to *Bcbva* transmission in the upper strata of the rainforest[24].

Fly samples indicated a large proportion of undetected anthrax mortality. During 19 days of focused fly sampling, we retrieved *Bcbva* isolates from 17 flies, with 13 strains being more than two SNPs different from any other strain. Since two SNPs appear to be the upper level of within-host diversity (Table S9), this implies the presence of at least 13 different *Bcbva* positive carcasses. Yet, during the same sampling period, only three *Bcbva* positive carcasses were discovered and their isolates all corresponded to one of the

186 fly *Bcbva* lineages (≤ 2 SNPs difference). This suggests carcass monitoring alone
187 underestimates mortality by at least an order of magnitude.

188 We investigated the consequences of *Bcbva*-induced mortality on the species best studied
189 in this ecosystem, chimpanzees. Chimpanzees have a low reproduction rate[25] and are thus
190 particularly sensitive towards external changes to their environment. Based on
191 demographic data collected from habituated groups in TNP, we simulated population
192 viability at a 150 years horizon across a broad range of demographic models including
193 and excluding anthrax induced mortality (Fig. S7 and S8). Our simulations showed that,
194 with *Bcbva*, the TNP chimpanzee population would only have high chances to persist in
195 the case of an overall annual per capita mortality rate due to other causes of 1% (Fig. S7
196 and S8). Such a low mortality rate is, however, not even observed in captive
197 chimpanzees. In wild chimpanzees the lowest annual per capita mortality rate is 4% (in
198 early adults)[25]. Under such a survival probability (0.96), the simulated presence of
199 anthrax invariably led to a clearly reduced survival probability of communities (Fig. 4).
200 For example, 76/84 models resulted in extirpation probability higher than 50%, while the
201 model which we consider the most realistic (community size 60, maximum age 46 years
202 and inter birth interval 6 years) resulted in an extirpation probability of 89% (Fig. 4). Our
203 simulations therefore suggest that anthrax induced mortality will result in deterministic
204 population declines and possible extirpation of TNP chimpanzees over the next 150
205 years. The risk of extirpation will increase if chimpanzee mortality due to hunting and
206 human-borne infectious diseases continues to rise [23,26,27].

207 To determine whether similar unrecognized effects on wildlife might be occurring
208 elsewhere, we tested 784 flies collected at eight different sites, as well as 136 bones from

twelve sites in five and nine sub-Saharan countries, respectively (Fig. S3, Table S1). All sites had chimpanzee populations but none (nor the country) had previously reported *Bcbva* cases. We only detected *Bcbva* genetic material in 2 of 105 flies and 1 of 8 bones collected in the Grebo National Forest (GNF) in Liberia, about 40 km from TNP (Fig. S1). The genome sequences of isolates from the two fly samples nested within the diversity of *Bcbva* in TNP which may indicate an epidemiological link (Fig. 3). We did not detect *Bcbva* in 305 flies from two sites where *Bcbva* cases have been previously reported (Dja Reserve, CM, and Dzanga Sangha Protected Areas, CAR, Table S4). While the lack of detection at other sites needs to be interpreted with caution due to variable fly species composition (Extended Data Fig. S10, Supplementary information S3f), these data suggest that *Bcbva* dynamics may also vary across rainforest ecosystems. It will be important to further uncover the scale and environmental drivers behind *Bcbva* prevalence. Such knowledge will be critical for mitigating against the detrimental effects of *Bcbva* on wildlife and for better assessing human infection risk, which for anthrax in rainforest ecosystems has, to date, been considered very low.

References (1-27)

1 Hampson, K. *et al.* Predictability of anthrax infection in the Serengeti, Tanzania. *J Appl Ecol* **48**, 1333-1344, doi:10.1111/j.1365-2664.2011.02030.x (2011).

2 Hugh-Jones, M. & De Vos, V. Anthrax and wildlife. *Revue Scientifique et Technique-Office International des Epizooties* **21**, 359-384 (2002).

3 Lindeque, P. M. & Turnbull, P. C. Ecology and epidemiology of anthrax in the Etosha National Park, Namibia. *Onderstepoort J Vet Res* **61**, 71-83 (1994).

4 Beyer, W. & Turnbull, P. C. Anthrax in animals. *Mol Aspects Med* **30**, 481-489, doi:10.1016/j.mam.2009.08.004 (2009).

5 Turnbull, P. C. B. *WHO Guidelines Approved by the Guidelines Review Committee* (World Health Organization, Department of Communicable Diseases Surveillance and Response, 2008).

6 Good, K. M., Houser, A., Arntzen, L. & Turnbull, P. C. Naturally acquired anthrax antibodies in a cheetah (*Acinonyx jubatus*) in Botswana. *Journal of wildlife diseases* **44**, 721-723 (2008).
7 Wafula, M. M., Patrick, A. & Charles, T. Managing the 2004/05 anthrax outbreak in Queen Elizabeth and Lake Mburo National Parks, Uganda. *African Journal of Ecology* **46**, 24-31 (2008).
8 Clegg, S., Turnbull, P., Foggin, C. & Lindeque, P. Massive outbreak of anthrax in wildlife in the Malilangwe Wildlife Reserve, Zimbabwe. *The Veterinary Record* **160**, 113-118 (2007).
9 Muoria, P. K. *et al.* Anthrax outbreak among Grevy's zebra (*Equus grevyi*) in Samburu, Kenya. *African Journal of Ecology* **45**, 483-489 (2007).
10 Turnbull, P. *et al.* Anthrax in wildlife in the Luangwa Valley, Zambia. *The Veterinary Record* **128**, 399-403 (1991).
11 de Vos, V. The ecology of anthrax in the Kruger National Park, South Africa. *Salisbury Med Bull Suppl* **68**, 19-23 (1990).
12 Lembo, T. *et al.* Serologic surveillance of anthrax in the Serengeti ecosystem, Tanzania, 1996-2009. *Emerg Infect Dis* **17**, 387-394, doi:10.3201/eid1703.101290 (2011).
13 Beyer, W. *et al.* Distribution and molecular evolution of *Bacillus anthracis* genotypes in Namibia. *PLoS neglected tropical diseases* **6**, e1534 (2012).
14 Leendertz, F. H. *et al.* Anthrax kills wild chimpanzees in a tropical rainforest. *Nature* **430**, 451-452, doi:10.1038/nature02722 (2004).
15 Klee, S. R. *et al.* The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids. *PLoS One* **5**, 0010986 (2010).
16 Brézillon, C. *et al.* Capsules, toxins and AtxA as virulence factors of emerging *Bacillus cereus* biovar *anthracis*. *PLoS Negl Trop Dis* **9**, e0003455, doi:10.1371/journal.pntd.0003455 (2015).
17 Antonation, K. S. *et al. Bacillus cereus* biovar *anthracis* causing anthrax in sub-Saharan Africa - Chromosomal monophyly and broad geographic distribution. *PLOS Neglected Tropical Diseases* (2016).
18 Leendertz, F. H. *et al.* Anthrax in Western and Central African great apes. *American journal of primatology* **68**, 928-933, doi:10.1002/ajp.20298 (2006).
19 Calvignac-Spencer, S. *et al.* Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular ecology* **22**, 915-924, doi:10.1111/mec.12183 (2013).
20 Berry, H. H. Surveillance and control of anthrax and rabies in wild herbivores and carnivores in Namibia. *Rev Sci Tech* **12**, 137-146 (1993).
21 Vergnaud, G. *et al.* Comparison of French and worldwide *Bacillus anthracis* strains favors a recent, Post-Columbian origin of the predominant North-American Clade. *PLoS One* **11**, e0146216, doi:10.1371/journal.pone.0146216 (2016).
22 Smith, K. L. *et al. Bacillus anthracis* diversity in Kruger National Park. *Journal of clinical microbiology* **38**, 3780-3784 (2000).

23 Campbell, G., Kuehl, H., Diarrassouba, A., N'Goran, P. K. & Boesch, C. Long-term research sites as refugia for threatened and over-harvested species. *Biology letters* **7**, 723-726, doi:10.1098/rsbl.2011.0155 (2011).
24 Blackburn, J. K., Van Ert, M., Mullins, J. C., Hadfield, T. L. & Hugh-Jones, M. E. The necrophagous fly anthrax transmission pathway: empirical and genetic evidence from wildlife epizootics. *Vector-Borne and Zoonotic Diseases* **14**, 576-583 (2014).
25 Hill, K. *et al.* Mortality rates among wild chimpanzees. *Journal of human evolution* **40**, 437-450, doi:10.1006/jhev.2001.0469 (2001).
26 Köndgen, S. *et al.* Pandemic human viruses cause decline of endangered great apes. *Current biology : CB* **18**, 260-264, doi:10.1016/j.cub.2008.01.012 (2008).
27 Boesch, C. & Boesch-Achermann, H. *The chimpanzees of the Taï Forest: Behavioural ecology and evolution*. (Oxford University Press, USA, 2000).

Supplementary information

Supplementary information: this file contains a more detailed method section as well as additional tables (Tables S1-10) and figures (Fig. S1-8).

Additional Data Table S1: Results derived from the analyses of flies caught in TNP analyzed in this study. This file includes results from PCR and culture as well as flymeal analysis results for a selection of flies.

Additional Data Table S2: Results of fly meal analysis with taxonomic assignment at genus level. This file provides the number of sequences per amplicon assigned at genus level.

Additional Data Table S3: Results of fly meal analysis with taxonomic assignment at order level. This file provides the number of sequences per amplicon assigned at order level.

Acknowledgements

We thank the authorities in Côte d'Ivoire for long-term support,especially the Ministry of the Environment and Forests, the Ministry of Research, the directorship of the Taï

National Park, and the CSRS in Abidjan. We also thank the national authorities from all other countries for providing permissions for our research (MINFoF, MINRESI, the Service de la Conservation de la Réserve du Dja, Cameron, in CAR the Ministère d'Eaux et Fôret, Chasse et Peche and the Ministère de l'Education Nationale, de l'Alphabetisation, de l'Enseignement Superieur, et de la Recherche, the Agence Nationale des Parcs Nationaux, Gabon, Centre National de la Recherche Scientifique et Technologique, Gabon, Direction des Eaux, Forêts et Chasses, Senegal, Forestry Development Authority, Liberia, Institut Congolais pour la Conservation de la Nature, DR-Congo, Ministère de l'Agriculture de l'Elevage et des Eaux et Forêts, Guinea, Instituto da Biodiversidade e das Áreas Protegidas (IBAP), Guinea-Bissau, Ministère de la Recherche Scientifique, DR-Congo, Ministère de le Recherche Scientifique et Technologique, R-Congo, Nigeria National Park Service, Nigeria, Uganda National Council for Science and Technology, Ugandan Wildlife Authority, Uganda).

We thank the WWF CAR for their logistical support. For the collection of samples in TNP we thank the field assistants, Arthur Henlin, Katerina Albrechtova and Alexander Lang. We are also grateful to the field assistants from all other sites for their support. For laboratory we are grateful to Silke Becker, Tatjana Franz, Sabine Howaldt, Angelika Lander, Petra Lochau, Herbert Nattermann and Andy Schneider, for sequencing we thank the RKI Central Sequencing Laboratory, particularly Julia Hinzmann, Andreas Nitsche and Julia Tesch and for bioinformatic support to Piotr Wojciech Dabrowski and Torsten Semmler from RKI, as well as Graham Hamilton at Glasgow Polyomics. We would further like to thank Teresa Börding, Thurston Hicks, Yasmin Moebius, Volker Sommer, Klaus Zuberbühler and Martine Peeters, and for administrative support Maja Kovacev-

334 Wegener. For funding we thank the German Research Council DFG KL 2521/1-1 and the
335 Sonnenfeld-Stiftung. We would like to thank the Max-Planck-Society and Krekeler
336 Foundation for funding of the *Pan African Programme*.
337

338 **Author contributions**

339 CH, FZ, AA, SA, MA, GB, KC, PD, KD, HE, PF, YG, AG, AG, SMG, JH, SJ, JJ, JK,
340 KL, JL, KL, VL, TL, SM, AM, SM, MM, JvS and ET collected flies, bones and
341 according field data. Necropsies on wildlife found dead were performed by FZ, KN, AB,
342 ECH, AD, PF, SAL, TL, SM, SN, HDN and FHL and laboratory analyses performed by
343 CH, FZ, KN, SD, KMR, KM, SM, HDN, AS, UT, SK, SC and FHL. The data were
344 analyzed by CH, FZ, RB, HK, RM and SC and the manuscript prepared by CH, FZ, RB,
345 HK, RM, JG, SC and FHL. The manuscript was revised and approved by all authors. The
346 study was supervised by CB, RW, SC and FHL.

347

348 **Author information**

349 Reprints and permissions information is available at www.nature.com/reprints. The
350 authors declare no competing financial interests. Correspondence and requests for
351 materials should be addressed to leendertzf@rki.de.

352

353 **Figure legends**

354 **Fig. 1. *Bcbva* occurrence and study sampling sites in sub-Saharan Africa.** Sites with
355 known *Bcbva* occurrence are indicated in red. Detection of *Bcbva* in Taï National Park,
356 Dja Reserve, Dzanga-Sangha Protected Areas and Luebo has been described in previous

studies. For all *Bcbva* sites, except Luebo, samples were available. Within this study we could identify Grebo as a new site of *Bcbva* occurrence. *Bcbva* was not detected at the other tested sub-Saharan sites (indicated in black).

Fig. 2. ***Bcbva* cases in Taï National Park.** (**A**) *Bc*bv*a* positive and negative carcasses. 38% of the observed wildlife mortality in Taï National Park is due to *Bcbva*. *Bcbva* positive carcasses were broadly distributed throughout the research area with no obvious pattern identifiable. GPS data was available for 113 of 204 detected carcasses and not for those detected before 2001. (**B**) *Bcbva* positive and negative fly traps. Five percent of all analyzed flies contained *Bcbva* genetic material. Flies were also caught outside the research area. A systematic snapshot sampling revealed higher prevalence of *Bcbva* positive fly traps within the research area.

Fig. 3. Phylogenomic tree of *Bcbva* isolates. Maximum likelihood tree based on chromosomal sequences of *Bcbva* isolates from TNP (Côte d'Ivoire, n=124) and Grebo (Liberia, n=2). One sequence per host (mammals/flies, two divergent isolates for fly 600) was included and the final alignment of variant sites measured 298bp.Internal branches with bootstrap values lower than 90 are colored in grey. The colored strip represents different host species. The tree was rooted using the heuristic residual mean squared function in *TempEst* v 1.5. The scale bar is in substitution per chromosomal site.

Fig. 4. Proportions of simulated chimpanzee communities surviving 150 years with and without presence of anthrax. Shown are results for different community sizes and anthrax being absent (a, blue boxes) or present (p, red boxes). Bars represent median

estimates and boxes quartiles across a range of simulation models assuming different inter birth intervals and maximum ages. All models summarized here assumed an annual per capita survival rate of 0.96.

Methods

Study sites

TNP covers an area of 3,300 km^2 and an additional 200 km^2 buffer zone. Since 2001 a veterinary program conducts outbreak investigations in wildlife. We defined the research area as the habitat ranges of the three habituated chimpanzee groups plus a 500 m buffer zone (103 km^2; Fig. S2).

Samples belonging to the large-scale data set were collected at 16 sites in 11 sub-Saharan countries stretching from Senegal to Uganda (Fig. S3, Table S1). Most sites (14 out of 16) were temporary research sites of the *Pan African Programme* (www.panafrican.eva.mpg.de) where *Bcbva* has not been described. Additional samples were obtained from Dja Faunal Reserve (DJR), Cameroon[18] and Dzanga-Sangha Protected Areas (DSPA), Central African Republic[17], where *Bcbva* cases have been previously described. Study sites are described in detail in Supplementary information S1.

Necropsies

Carcass monitoring was performed in TNP by a veterinarian, performing necropsies on every carcass reported by researchers working in the forest (n=173). Samples of all inner organs were collected, as far as carcass decomposition allowed. Necropsies followed a

403 standardized protocol, including use of full personal protective equipment. Carcass sites
404 were decontaminated according to World Health Organization (WHO) guidelines[5,28]. For
405 each sample aliquots were stored in liquid nitrogen and formalin in the field. Frozen
406 samples were transported on dry ice and subsequently stored at -80°C. We received
407 additional tissue samples from carcasses sampled by the WHO in TNP between 1996 and
408 2000 (n=31) (Table S2).

409 Rather than using serology, which would also detect animals that survived non-lethal
410 infections, we used PCRs to detect the presence of anthrax in internal organs to confirm
411 that anthrax was the likely cause of death. DNA was extracted from various tissues per
412 animal (liver, spleen and lung when available) using the DNeasy Blood and Tissue Kit
413 (Qiagen, Hilden, Germany); extracts were quantified using a Nanodrop (Thermo Fisher
414 Scientific, Waltham, MA, USA) and stored at -20°C. Two hundred ng DNA or 5 µl of
415 DNA extract (if DNA concentration was below 40 ng/µl) were tested for anthrax in
416 duplicate real-time PCR reactions (details in Supplementary Methods S2c). The full
417 anthrax assay used includes three real-time PCRs, each targeting one of the following
418 gene markers: *pag* (gene for protective antigen) located on the pXO1 plasmid[29], *capB*
419 (gene for capsule synthesis) located on pXO2 and *Island IV*, a chromosomal marker
420 specific for *Bcbva*[15,17] (Table S3). Samples were first tested for *pag* and samples positive
421 in duplicate for *pag* were tested for *capB* and *Island IV* (Extended Data Fig. S1 and
422 Extended Data Fig. S2).

423 Culture under BSL3 conditions was attempted for all PCR positive necropsy samples
424 collected until the end of 2013 (June 2014 for duikers) (Table S2). A native and heat-
425 treated (65°C for 30 min, to assess presence of spores) aliquot were plated onto the

following agar plates: Columbia blood agar (Oxoid, Wesel, Germany), blood-trimethoprim agar (1.6 mg trimethoprim, 6.4 mg sulfamethoxazole, 20 mg polymyxin B per liter agar medium) and Cereus Ident agar (Heipha Diagnostica, Eppelheim, Germany) with the chromogenic substrate 5-bromo-4-chloro-3-indoxyl-myoinositol-1-phosphate[30]. Cultures were incubated at 37°C and monitored daily. Morphologically suspicious colonies were sub-cultured and tested in real-time PCR. *Bcbva* was cultured from native and heat-treated samples indicating the presence of heat-resistant spores. Isolates were frozen in Microbank tubes (Mast Diagnostica, Reinfeld, Germany) at -80°C.

Histopathology was performed on a subset of necropsy samples, including 15 *Bcbva* PCR positive necropsy samples (Table S2). No signs of anthrax infection were detected in carcasses that were PCR negative for anthrax, while for PCR positive carcasses the most consistent histopathologic finding was per-acute to acute anthrax related pneumonia characterized by mild lymphohistiocytic infiltrates and intraalveolar eosinophilic and proteinaceous or fibrinous material. Numerous bacilli were found intravascular and intraalveolar. Multifocal alveolar and peribronchiolar hemorrhages were present in all animals. Lymph node changes consisted of sinus histiocytosis, cortical hemorrages and edema especially in the mediastinal, tracheobronchiolar and mesenteric lymph nodes. Huge amounts of bacilli were demonstrable in the sinusoids. Within the abdominal cavity the spleen was the organ most affected, with myriads of bacilli visible in the splenic sinusoids, partly embedded in fibrin deposition. There was moderate lymphoid depletion, lymphocytolysis and histiocytosis. The liver parenchyma was severely congested with masses of bacilli within the hepatic sinusoids. All anthrax PCR positive carcasses were

448 also tested for filoviruses[31] and respiratory diseases [26] to rule out co-infection with other
449 common causes of death in this ecosystem.

450 ***Blow flies***

451 Flies were caught on the ground and in the canopy using custom-made traps (Fig. S4 and
452 S5, Supplementary information S3a). Trapping was done for 60 min or until 20 flies were
453 collected. Flies were euthanized with ether and stored at -20°C in 2 ml Cryotubes (Carl
454 Roth) containing up to 10 flies or at ambient temperature on silica in 50 ml Falcon tubes
455 (Thermo Fisher Scientific) containing up to 20 flies. In TNP, 726 flies were randomly
456 collected within the research area in 2008, 2009, 2012 and 2013 (Table S4). Another 908
457 flies were collected over 19 days in May and June 2014 according to a 2x2 km grid
458 system covering the research area and 225 km^2 surrounding the research area (referred to
459 as "snapshot flies"; Extended Data Fig. S7, Table S4). At a larger scale, 784 flies were
460 collected at 8 sites within 5 sub-Saharan countries (*Pan African Programme*) from 2012
461 to 2014 (Table S4) and 305 flies were analyzed from two sub-Saharan sites, DJR (n=105)
462 and DSPA[17,18] (n=200) (Table S4). In total, 2,723 flies were analyzed (Table S4).

463 DNA extraction of individual flies was performed using the GeneMATRIX Stool DNA
464 Purification Kit (Roboklon, Berlin, Germany). We followed manufacturer's instructions
465 except that each fly was first cut into small pieces using sterilized scissors before being
466 homogenized using a Fast Prep® (MP Biomedicals, Santa Ana, CA, USA). DNA
467 concentration measurements and anthrax testing by real-time PCR were performed as
468 described for necropsy samples (Table S4).

469 A subset of 50 flies containing high *pag* copy numbers underwent bacterial culture (Table
470 S4, Additional Data Table S1). Half of the fly mush remaining after DNA extraction was

plated directly onto the same culture media described for necropsy samples. Additionally, a 10 µl aliquot of the mush was diluted 1:10 in sterile NaCl, heat treated for 30 min at 65°C and plated. *Bcbva* was retrieved from native and heat-treated samples, indicating the presence of heat-resistant spores in flies. An on-site study in TNP also used direct culture of 204 flies without preceding PCR testing. Flies were homogenized and plated directly onto Cereus Ident agar. Suspicious colonies were sub-cultured on blood-trimethoprim agar and tested in real-time PCR. This approach yielded another 21 *Bcbva* isolates.

To examine whether certain mammals were preferentially affected by *Bcbva*, we tested for differences in fly meal composition of anthrax positive and negative flies. We screened a subset of 750 TNP flies for mammalian DNA using a real time PCR targeting a 130 bp fragment of mammalian 16S mitochondrial DNA (described in Calvignac-Spencer et al.[19]). We chose a subset of mammal and anthrax positive (n=28) and the according number of mammal positive but anthrax negative flies (n=29) from the same traps (Additional Data Table S1). To dissect fly meal composition, we used a metabarcoding approach, whereby 16S amplicons were deep-sequenced, adapting the amplicon preparation protocol provided by Illumina (San Diego, CA, US) (Supplementary information S3e). We used a custom pipeline to determine taxonomic assessment of each read to the genus and order level described in the Supplementary information S3e (Table S5, Additional Data Tables S1 and S2). Sequences assigned to domestic animals were regarded as contamination as it was shown that even stringent anti-contamination procedures do not prevent the amplification of human and domestic animal sequences present in the environment and reagents[32].

494 Details on blow fly analyses and results are in Supplementary information S3.

495 ***Bones***

496 Bones were collected in TNP and 12 *Pan African Programme* sites in 9 countries (Fig.
497 S3, Table S1 and S7). Bones were transported and stored at ambient temperature. DNA
498 was extracted using a silica-based method[33,34] (Suppementary information S4b). Bone
499 extracts were tested by real-time PCR as described for necropsy samples (Table S1 and
500 S7). Powder from PCR positive bones was also used for bacterial culture attempts after
501 homogenization in sterile NaCl (Table S7). We processed the homogenates as described
502 above for necropsy samples with one native aliquot and one heat- treated aliquot. Details
503 on bone analyses and additional results are in Supplementary information S4.

504 ***Whole-genome sequencing of Bcbva isolates and SNP calling***

505 Table S8 contains a complete list of all *Bcbva* isolates sequenced (Fig. S6). Isolate
506 preparation and extraction is described in the Supplementary Methods S6a. Libraries for
507 whole-genome sequencing were prepared with the Nextera XT DNA Library Preparation
508 Kit (Illumina). Libraries were pooled and sequenced on the HiSeq 1500 platform
509 (Illumina) in rapid run mode using either v1 (2x150 bp) or v2 (2x250 bp) chemistry.

510 Illumina adapters were removed using *scythe* v0.993[35] and trimmed with *sickle* v1.33[36]
511 applying a quality threshold of 25. Quality trimmed reads were aligned to the reference
512 genome (*Bcbva* strain CI, Accession numbers CP001746-749) with the BWA-MEM
513 algorithm implemented in *bwa* v0.7.12-r1039[37]. For conversion to bam format, sorting,
514 deduplication and indexing of aligned reads, we used the *picard tools* 1.136[31] software
515 package applying the commands *SortSam*, *MarkDuplicates* and *BuildBamIndex*.
516 Subsequent variant calling was performed using the *Genome Analysis Toolkit* (*GATK*)

v3.4[38-40]. We realigned bam files with the tools *RealignerTargetCreator* and *IndelRealigner*. Variants were called with *UnifiedGenotyper* with a minimum phred scaled confidence threshold of 30 for SNPs to be called. Hard filtering of SNP sites was done with the *VariantFiltration* command using recommended filter settings. With the *SelectVariants* command, only SNP sites that passed the filter were selected for further processing. *SelectVariants* was also used to exclude all SNPs with a coverage < 5x, a minor allele frequency of > 0.1 and a *GATK* Genotype Quality value < 99. Final consensus sequences were composed with *FastaAlternateReferenceMaker*. We assessed coverages of all samples with the *GATK* tools *DepthOfCoverage* and *CoveredByNSamplesSites*. Details and further analysis of whole-genome sequencing of *Bcbva* isolates and SNP calling is Supplementary information S6.

Phylogenetic analyses

126 genome sequences (one isolate per mammal/fly) from TNP and GNF (Table S8) were aligned and stripped of non-variant sites with *Geneious Pro* v8.1.3 (Biomatters ltd.)[41]. Resulting alignments of variable sites were 298, 18 and 11 bp long for the chromosome, pXO1 and pXO2 respectively. Given the low number of variable sites in pXO1 and pXO2, we only performed phylogenetic analyses on the chromosome alignment. *jModelTest* v2.1.4[42] was used for determination of the best nucleotide substitution model in a maximum likelihood (ML) framework, resulting in the choice of TVMef[43].

ML analysis was performed with *PhyML* v20131022[44] using a combination of subtree-pruning-regrafting (SPR) and nearest-neighbor-interchange (NNI) tree search algorithms. Branch support was estimated using non-parametric bootstrapping with 100 pseudo-

replicates. The tree was rooted using the heuristic residual mean squared function in *TempEst v 1.5*[45], placing the root at the position resulting in the most clock-like structure of the data(Fig. 3) .

We also performed phylogenetic analyses using the Bayesian Markov Chain Monte Carlo (BMCMC) sampling approach implemented in *BEAST* v1.8.2[46] specifying a constant population coalescent tree prior and assuming an uncorrelated lognormal relaxed molecular clock[47] (Supplementary information S7c). The maximum clade credibility tree derived from this analysis was very similar to the ML tree (Fig. 3).

Another data set was assembled to compare *Bcbva* from TNP to other strains from sub-Saharan Africa. It included the chromosomal sequences from a representative TNP genome, GNF ones, as well as previously published genomes determined from isolates derived from *Bcbva* cases in CAR and CM[15,17] (Extended Data Fig. S6). The alignment was compiled as described above and contained 1,016 variable positions. Model selection with *jModelTest* v2.1.4[42] selected a TPM1[48] nucleotide substitution model. We performed ML analyses as described above.

Statistical analyses

To test the effect of season on the probability of a carcass or fly, respectively, being anthrax positive, we used a Generalized Linear Mixed Model (GLMM)[49] with binomial error structure and logit link function[50]. As predictors we included the species (monkeys, chimps, duikers, others, blow flies), season and their interaction. 'Season' was modelled by first turning the sampling date into a circular variable and including its sine and cosine into the model. As random intercept effects we included trap id (i.e., GPS location) and the combination of sampling date and GPS location, the latter accounting for potential

563 non-independence of flies sampled on the same day from the same trap. We further
564 included random slopes of season within trap id [51,52]. To test the effect of season we
565 compared the full model with a null model lacking the fixed effects of season and its
566 interaction with species[53], using a likelihood ratio test[54]. Sample size for this model was
567 1803 samples (carcasses and flies), collected at 352 locations and 328 combinations of
568 sampling date and location including necropsy samples and flies.

569 In a second model we tested whether the probability of a fly to be tested anthrax positive
570 was influenced by season and the amount of mammalian DNA within in the fly. We used
571 a Generalized Linear Mixed Model (GLMM)[49] with binomial error structure and logit
572 link function[50]. Into this we included the amount of mammalian DNA found within the
573 fly (determined with real time PCR described above) and season as fixed effects. 'Season'
574 we modeled by first turning the sampling date into a circular variable and then including
575 it sine and cosine into the model. Since the amount of mammalian DNA within the fly
576 was highly skewed, we log transformed it before fitting the model. As random effects
577 (random intercepts) we included the ID of the trap and the date of sampling. To avoid
578 overconfident estimates we included random slopes of the amount of mammal DNA
579 within trap ID and trapping date[51,52]. As an overall test of the effects of the amount of
580 mammal DNA and season we compared the full model with a null model lacking these
581 effects[53] using a likelihood ratio test[54]. We also used likelihood ratio tests to test for the
582 individual predictors (comparing the full model with a respective reduced model lacking
583 the predictor to be tested[51]). We fitted the model in *R*[55] using the function *glmer* of the *R*
584 package *lme4* (version 1.1-10[56]). To estimate model stability we excluded levels of the
585 random effects one at a time which did not indicate influential levels to exist. The total

586 sample size for this model was a total of 474 flies caught on 43 days in 33 traps.
587 (Extended Data Fig. S4, Table S10, Supplementary information S8b).

588 To evaluate the reproducibility of fly meal identification for each fly we correlated the
589 proportion of sequence counts per amplicon (two amplicons per fly) that was assigned to
590 different mammalian genera using Spearman correlation. To test whether there were
591 differences in fly meal composition of anthrax positive and anthrax negative flies, we
592 tested whether detection of a given mammal taxon in a fly sample was associated with
593 anthrax positivity. We used GLMMs[49] applied separately for each mammal genus
594 identified in the flies. The response was whether the fly was anthrax positive and the key
595 predictor with fixed effect was mammal presence. We considered a mammal to be
596 present when it was detected in at least one of the two amplicons per fly. We included
597 only those mammal genera in the model that were detected in at least five of all generated
598 amplicons (two per fly). In addition to mammal presence, we included tid and the factor
599 sampling date as random effects (random intercepts)[51,52]. Models were fitted with
600 binomial error structure and logit link function[50]. Sample size for all models was 57 flies,
601 caught in 22 different traps on 13 days. To test whether mammal presence had an impact
602 on anthrax positivity, we dropped mammal presence from the model[53] and compared the
603 models using a likelihood ratio test[54]. Model stability was accessed as above. We fitted
604 models at two different taxonomic resolutions: one with taxonomic assignment at genus
605 level and the other at order level. GLMMs were fit in *R*[55] using the function *glmer* of the
606 *R* package *lme4* v1.1-10[56].

607 To evaluate geographic distribution of *Bcbva* in TNP we checked whether, due to higher
608 mammal density[23], *Bcbva* was more likely to occur inside the research area. To test this

609 hypothesis we analyzed 908 flies from 83 different traps (Extended Data Fig. S7, Table
610 S4, Additional Data Table S1). 21 traps were located within the research area and 62
611 traps in the adjoining forest belt. 8/21 traps within the research area were anthrax positive
612 and 8/62 outside the research area. We compared the two groups using Fisher Exact´s
613 Test (Supplementary information S8d).

614 To learn more about the spatial dynamics of *Bcbva* in TNP, we investigated the
615 correlation between genetic and geographic distances. To correct for genetic and spatial
616 autocorrelation, we excluded strains from the data set that originated from the same fly
617 catching point (in a 1 km^2 radius) on the same day or from the same followed-up outbreak
618 in mammals. Only one strain was kept per outbreak or fly catching point, the selection
619 criterion being high average coverage of the genome (Table S8, Supplementary
620 information S8e). Geographic distances (in km) were derived from GPS data using
621 *GeographicDistanceMatrixGenerator* v1.2.3[57]. Genetic distances were approximated
622 using the relative distances drawn from a Maximum Likelihood Tree built in *PhyML*
623 *v20131022*[44] with the *R* package *ape*[58] using the *cophenetic* function. Multiple regression
624 on distance matrices (MRM) as implemented in the R *ecodist* package[59] using 1000
625 permutations and Spearman correlations was performed on genetic and geographic
626 distance matrices. To examine variation within genetic lineages, we binned our data by
627 genetic distance (bin size=relative genetic distance of 0.03, approx. 2.5 SNPs) and
628 focused on groups with low genetic distance (max relative genetic distance <0.5) and
629 their mean geographic distance (Extended Data Fig. S9). Homogeneity of variance
630 between groups was assured with the Fligner Killeen test ($p=0.07$; >0.05 as requested).

To evaluate the impact *Bcbva* could have on the TNP chimpanzee population, we conducted a simulation (Supplementary information S8f). We first defined a series of population parameters for the simulation[27]. We simulated the survival prospects of chimpanzee communities of a given size, with individuals reproducing at certain regular intervals after maturation, having a maximum age, and an annual survival probability. Since most of these parameters are associated with considerable uncertainty and since we wanted to assess to what extent the simulation results depend on the particular parameters chosen we parameterized the simulations as follows: Initial community size: 20 to 80 individuals (increment: 10); inter-birth interval: 4 to 7 years (increment: 1); interval after death of infant: 1 year; maximum age: 40 to 50 years (increment: 2); age of first reproduction of males and females: 10 and 14 years, respectively. Since per capita annual survival probability without the influence of anthrax is unknown (mortality cases due to anthrax may not be detected in all cases, in particular before necropsies were made systematically), we simulated per capita annual survival probabilities from 0.93 to 0.99 (increment: 0.03). In addition, we made survival probability density dependent, as this is a common characteristic observed in many species including chimpanzees[60]. For this we introduced a logistic function (1/(1+exp(-(20-0.08*community size)))) that increased or reduced mortality rate as a function of chimpanzee community size. At the beginning of each simulation run we generated a community of the simulated size by randomly allocating sexes (proportion of females: 0.7) and ages (uniformly distributed between 10 and the simulated maximum age) to individuals. To avoid stochastic effects of the initially generated community, we let the simulation run for 50 time steps (i.e., 'years') without anthrax presence before the evaluated time period began.

We estimated the risk of annual anthrax outbreak probability, its dependence on community size and the number of individuals affected as exp (-1.83+0.039*community size) from a Poisson regression (null, full model comparison, χ^2= 7.89, df=1, p <0.01). We simulated both an anthrax and a non-anthrax scenarios for 150 time steps (i.e., 'years') with 100 replications each and for each possible combination of the simulated parameters. Communities were considered to be extinct, when no reproducing females were present.

All *R* scripts are available upon request. Details on statistical analyses and additional results are in Supplementary information S8.

Data availability

Raw reads of 16S amplicons are available in the European Nucleotide Archive (ENA) under project accession number PRJEB14554, sample accession numbers ERS1217219-336. Raw reads for all 178 *Bcbva* isolates from TNP and GNF are available in the ENA under project accession number PRJEB14616, sample accession numbers ERS1222903-3080. Variable position alignments are available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.v8bn7.

Extended data figure legends

Extended Data Fig. S1. Necropsies performed since 1996. Shown is the total amount of necropsies performed per year in TNP from 1996 to 2015. Grey bars indicate the number and according proportion of *Bcbva* positive necropsies. In the years 2003 and 2010 only limited veterinary service was available at TNP due to political insecurity in the region.

Extended Data Fig. S2. Geographic location of *Bcbva* positive carcasses in Taï National Park. Shown are *Bcbva* positive tested necropsies in TNP since 2001. GPS data was available for 70 of all positive tested (n=81) necropsies.

Extended Data Fig. S3. Effect of mammalian DNA content on anthrax positivity in flies. . Shown is the probability of *Bcbva* positivity (PA, pag respectively) as a function of the amount of mammalian DNA (copies) found in a fly. The amount of mammal DNA was binned (bin width 0.25) and the area of the points depicts the number of flies (range: 1 to 206) in the respective bins. The dashed line indicates the fitted model and the dotted lines the 95% confidence interval.

Extended Data Fig. S4. Effect of season on anthrax positivity in flies. Shown is the probability of *Bcbva* (PA) positivity over the course of a year (binned in 10 day periods). The area of the points depicts the number of flies in the respective ten days period. The dashed line indicates the fitted model and the dotted lines the 95% confidence interval.

Extended Data Fig. S 5. Maximum clade credibility tree based on chromosomal sequences of Bcbva isolates from TNP (Côte d'Ivoire, n=124) and Grebo (Liberia, n=2). One sequence per host hosts (mammals/flies, two divergent isolates for fly 600) was included and the final alignment of variant sites measured 298bp. Size of nodes represents posterior probability values. The location of the root received a posterior probability of 1.

Extended Data Fig. S 6. Maximum likelihood tree for sub-Saharan Bcbva strains. ML tree based on chromosomal sequences of *Bcbva* strains from Côte d'Ivoire, Cameroon, Central African Republic and Liberia. The alignment of variant sites measured 1016bp. Bootstrap values are shown above branches and the scale bar reflects the genome-wide substitution rate. The tree was rooted using TempEst v 1.5.

Extended Data Fig. S 7. Fly snapshot sampling scheme. For the fly snapshot flies were caught following a 2x2 km grid system within and outside the research area within 19 days. In total 908 snapshot flies were analyzed.

Extended Data Fig. S 8. Genetic and geographic distances of *Bcbva* isolates from the fly snapshot. (A) Maximum Likelihood Tree based on chromosomal sequences of *Bcbva* isolates from the 19 day fly snapshot. Each dot represents one fly isolate. Colors were chosen to illustrate the distribution of genetically clustering isolates on the map presented in panel B. The final alignment of variant sites measured 123bp. Bootstrap values are shown above all internal branches. The tree was rooted using the "best-fit" option in

Path-O-Gen v1.2. The scale bar is in substitution per site. (B) Geographic origin of *Bcbva* isolates collected during the fly snapshot. Colors correspond to ML tree (A). Big circles represent two isolates.

Extended Data Fig. S 9. Boxplot of genetic and mean geographic distances. Bcbva isolates from TNP were binned by relative genetic distance (bin size = 0.03, approx. 2.5 SNPs).The two most genetically distant isolates received a value of 1 and all other distances were scaled accordingly. Diamonds indicate the geographic distance means of the groups. To examine variation within genetic lineages, we analyzed isolates with low genetic distance (max relative genetic distance < 0.5, marked in blue) and their mean geographic distance. For low genomic distances, genetic and mean geographic distances are correlated (p= 4x10-5, R2= 0.72).

Extended Data Fig. S10. Fly species composition based on GMYC analysis. Fly species composition for three sites with known Bcbva occurrence: TNP (Côte d´Ivoire) (A), DJR (Cameroon) (B), DSPA (Central African Republic) (C). Shown are the proportions of flies per site in % belonging to one single fly species identified with GMYC models. Different colors indicate different taxonomic fly families.

References (28-60)

28 Turnbull, P. C. B. (World Health Organization, Department of Communicable Diseases Surveillance and Response, 1998).

29 Ellerbrok, H. *et al.* Rapid and sensitive identification of pathogenic and apathogenic *Bacillus anthracis* by real-time PCR. *FEMS Microbiol Lett* **214**, 51-59 (2002).

30 Klee, S. R. *et al.* Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. *J Bacteriol* **188**, 5333-5344 (2006).

31 Panning, M. *et al.* Diagnostic reverse-transcription polymerase chain reaction kit for filoviruses based on the strain collections of all European biosafety level 4 laboratories. *Journal of Infectious Diseases* **196**, S199-S204 (2007).

32 Leonard, J. A. *et al.* Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archaeological Science* **34**, 1361-1366 (2007).

33 Gamba, C. *et al.* Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Molecular ecology resources* **16**, 459-469, doi:10.1111/1755-0998.12470 (2016).

34 Rohland, N. & Hofreiter, M. Ancient DNA extraction from bones and teeth. *Nature protocols* **2**, 1756-1762, doi:10.1038/nprot.2007.247 (2007).

35 Buffalo, V. Scythe: A 3'-end adapter contaminant trimmer [Software]. Available at https://github.com/vsbuffalo/scythe.

36 Joshi, N. A. & Fass, J. N. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files [Software]. Available at https://github.com/najoshi/sickle. (2011).

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

38 Auwera, G. A. *et al.* From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 11.10. 11-11.10. 33 (2013).

39 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491-498 (2011).

40 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

41 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).

42 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772, doi:10.1038/nmeth.2109 (2012).

43 Posada, D. Using MODELTEST and PAUP* to select a model of nucleotide substitution. *Current protocols in bioinformatics*, 6.5. 1-6.5. 14 (2003).

44 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, doi:10.1093/sysbio/syq010 (2010).

45 Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution* **2** (2016).
46 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, doi:10.1093/molbev/mss075 (2012).
47 Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS biology* **4**, e88, doi:10.1371/journal.pbio.0040088 (2006).
48 Kimura, M. Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences* **78**, 454-458 (1981).
49 Baayen, R. H. *Analyzing linguistic data: A practical introduction to statistics using R*. (Cambridge University Press, 2008).
50 McCullagh, P. & Nelder, J. A. *Generalized linear models*. Vol. 37 (CRC press, 1989).
51 Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Mem Lang* **68**, doi:10.1016/j.jml.2012.11.001 (2013).
52 Schielzeth, H. & Forstmeier, W. Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology* **20**, 416-420 (2009).
53 Forstmeier, W. & Schielzeth, H. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology* **65**, 47-55 (2011).
54 Dobson, A. J. & Barnett, A. *An introduction to generalized linear models*. (CRC press, 2008).
55 Team, R. C. *R: A language and environment for statistical computing.*, (R Foundation for Statistical Computing, Vienna, Austria, 2015).
56 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1-48 (2014).
57 Ersts, P. Geographic distance matrix generator (version 1.2. 3). American Museum of Natural History. Available at http://biodiversityinformatics.amnh.org/open_source/gdmg. *Center for Biodiversity and Conservation* (2011).
58 Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289-290 (2004).
59 Goslee, S. C. & Urban, D. L. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**, 1-19 (2007).
60 Kuehl, H., Elzner, C., Möbius, Y., Boesch, C. & Walsh, P. D. The price of play: Self-organized infant mortality cycles in chimpanzees. *PLoS One* **3**, doi:10.1371/journal.pone.0002440 (2008).