Data Descriptor: Bat trait, genetic and pathogen data from large-scale investigations of African fruit bats, *Eidolon helvum*

Alison J. Peel1,2,3, Kate S. Baker4,2,4,5, David T.S. Hayman1,2,6, Richard Suu-Ire7,8, Andrew C. Breed9, Guy-Crispin Gembu10, Tiziana Lembo11, Andrés Fernández-Lorasa12,22, David R. Sargan7, Anthony R. Fooks9, Andrew A. Cunningham2 & James L.N. Wood1

Bats, including African straw-coloured fruit bats (*Eidolon helvum*), have been highlighted as reservoirs of many recently emerged zoonotic viruses. This common, widespread and ecologically important species was the focus of longitudinal and continent-wide studies of the epidemiological and ecology of Lagos bat virus, henipaviruses and Achimota viruses. Here we present a spatial, morphological, demographic, genetic and serological dataset encompassing nine countries over an 8-year period. Genetic data comprises cytochrome *b* mitochondrial sequences (*n* = 608) and microsatellite genotypes from 18 loci (*n* = 544). Tooth-cementum analyses (*n* = 316) allowed derivation of rare age-specific serologic data for a lyssavirus, a henipavirus and two rubulaviruses. This dataset contributes a substantial volume of data on the ecology of *E. helvum* and its viruses and will be valuable for a wide range of studies, including viral transmission dynamic modelling in age-structured populations, investigation of seasonal reproductive asynchrony in wide-ranging species, ecological niche modelling, inference of island colonisation history, exploration of relationships between island and body size, and various spatial analyses of demographic, morphometric or serological data.

<table>
<thead>
<tr>
<th>Design Type(s)</th>
<th>epidemiological study ● genotyping design ● population genetics analysis objective ● individual genetic characteristics comparison design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement Type(s)</td>
<td>genetic characteristics information ● viral infection</td>
</tr>
<tr>
<td>Technology Type(s)</td>
<td>genotyping assay ● serum neutralization of viral infectivity assay ● DNA sequencing</td>
</tr>
<tr>
<td>Factor Type(s)</td>
<td>Eidolon helvum ● plagiopatagium ● blood ● Ilha de Sao Tome ● Accra ● Kisumu District ● Bondo District ● Madera District ● Kumasi ● Techiman Municipal District ● City of Kisangani ● Kasanka National Park ● Blantyre City ● Dar es Salaam Region ● Morogoro ● Municipality of Jinja ● City of Kampala ● Ilha da Principe ● Annobon Island ● Municipality of Malabo ● Bioko Island ● Rio Muni</td>
</tr>
<tr>
<td>Sample Characteristic(s)</td>
<td></td>
</tr>
</tbody>
</table>

1Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK. 2Institute of Zoology, Zoological Society of London, Regent’s Park, London NW1 4RY, UK. 3Environmental Futures Research Institute, Griffith University, Brisbane, Queensland 4111 Australia. 4Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. 5Institute for Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK. 6Molecular Epidemiology and Public Health Laboratory, Hopkirk Research Institute, Massey University, Private Bag, 11 222, Palmerston North 4442, New Zealand. 7Wildlife Division, Ghana Forestry Commission, Accra, Ghana. 8University of Ghana, Faculty of Animal Biology and Conservation Science, Box LG 571, Legon, Accra, Ghana. 9Animal and Plant Health Agency (APHA), Addlestone, Surrey KT15 3NB, UK. 10Faculté des Sciences, Université de Kisangani, 4, Avenue Kithima, commune Makibo, BP 2012, Kisangani, République Démocratique du Congo. 11Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Graham Kerr Building, Glasgow, G12 8QQ, Scotland. 12Museo Nacional de Ciencias Naturales, CSIC, José Gutiérrez Abascal 2, Madrid 28006, Spain. Correspondence and requests for materials should be addressed to A.J.P. (email: alisonpeel@gmail.com).
Background & Summary
The straw-coloured fruit bat (*Eidolon helvum*) is a common, widely distributed, migratory species, occurring across sub-Saharan Africa and some offshore islands (Fig. 1)\(^1,2\). Since 2007, investigations into the epidemiology and ecology of zoonotic viral infections in *E. helvum* have been undertaken via longitudinal sampling of wild populations in Ghana. Complementing this, between 2008–2011 and in 2014, cross-sectional sampling events were undertaken to determine the genetic population structure of *E. helvum*, and to assess whether the serological findings in Ghana were representative across the species’ range (Fig. 1).

Four viruses were the focus of our serological surveys in *E. helvum* bats: Lagos bat virus (LBV), African henipaviruses, Achimota virus 1 (AchPV1) and Achimota virus 2 (AchPV2). Lagos bat virus is one of at least 15 known species in the *Lyssavirus* genus\(^3\) and has been isolated from *E. helvum* on multiple occasions\(^4,5\). An African henipavirus is still yet to be isolated, however a full genome sequence has been obtained (putative name: African bat henipavirus Eid_hel/GH-M74a/GHA/2009 (M74))\(^6\). Achimota viruses 1 and 2 are closely related rubulaviruses for which serological evidence suggestive of spillover to humans in Africa exists\(^7\).
The specific aims of the data collection were to:

- Investigate whether antibodies to LBV, henipaviruses and Achimota viruses are present in *E. helvum* across its continental and island range, and to explore the antibody dynamics where possible.
- Describe the genetic metapopulation structure of *E. helvum* using a combination of mitochondrial (mtDNA) and microsatellite markers.
- Gather information on *E. helvum* distribution and seasonal patterns of reproduction.
- Combine results from these multidisciplinary studies to make inferences about virus transmission dynamics, and ultimately make inferences on the spillover risk to human populations.

Samples in this dataset are from 2827 bats from nine countries over an 8-year period (Fig. 2, Table 1 (available online only)). Raw data comprises spatial (roost location), seasonal (timing of sampling and seasonal birth pulses), morphological (forearm length, body weight), demographic (age, sex, reproductive status, mother-offspring relationships) and identification (individually numbered thumb-band) components. Data generated includes genetic characterisation (mtDNA sequencing and microsatellite genotyping) and serological assay results (for LBV, henipaviruses and Achimota viruses).

Multiple publications have arisen from these data, however many aspects remain unexplored. Demographic analyses have estimated birth and survival rates\(^8,9\), and explored the effect of hunting on the latter\(^7\). Variations in roost composition have suggested a fission-fusion social structure\(^7\). Serological analyses have identified: the presence of antibodies against LBV, henipaviruses and Achimota viruses in *E. helvum* in Ghana\(^7,10,11\), and more broadly across the species’ range, including isolated off-shore islands\(^7,12,13\); that these viruses circulate endemically in *E. helvum* in Ghana, with evidence of horizontal transmission\(^7,14,15\); and that *E. helvum* bats previously infected with LBV can have long-survival post infection\(^16\). The henipavirus dataset was used to develop a Bayesian method to determine appropriate cutoffs for serological assays\(^17\). Population genetic analyses identified that *E. helvum* are panmictic across their continental range, but that genetically isolated populations exist on isolated islands\(^12\).

Other publications arising from these samples, but based on analyses not included here, include the development of a universal real-time assay and a pseudotype neutralisation assay for Lyssaviruses\(^18,19\), microsatellite loci characterisation\(^20\), estimation of divergence times between *Eidolon* sister species\(^21\), inference of movement ecology based on stable isotope ratios\(^22\), demonstration of Ebola antibodies\(^16,23\), identification of multiple novel viruses\(^24\), and novel Bartonella species in bat flies collected from *E. helvum*\(^25,26\).

Figure 2. Sampling intensity per month, by country. Red represents high sampling intensity and the numbers of samples collected per month is recorded within each grid cell. Records in the database with unknown collection date are not represented here (nine from Annobon, seven from Bioko, 14 from Principe, 10 from Rio Muni and 13 from São Tomé).
This dataset contributes a substantial volume of data on the ecology of *E. helvum* and its viruses and will be valuable for a wide range of studies. In particular, an age-specific dataset such as this is rare and valuable for wildlife, especially bats. Further analyses could include viral transmission dynamic modelling in age-structured populations, including the use of cutting-edge Bayesian approaches to address complex epidemiological questions; time-series analyses on 5 years of wild henipavirus serological data from the same study site in Ghana (*n* = 1486 data points), investigation of seasonal reproductive asynchrony in wide-ranging species; ecological niche modelling; inference of island colonisation history, exploration of relationships between island and body size; and various spatial analyses of demographic, morphometric or serological data. Field samples (e.g. serum, blood cells, urine, skin samples) and extracted DNA from individual bats in this dataset exist in storage and the authors are open to collaborative requests to undertake further analyses.

**Methods**

These methods are expanded and modified versions of descriptions in our previous publications, as cited in each section below. All associated data can be found in *Eidolon helvum data 2007–2014.csv* [Data Citation 1].

**Capture and Data Collection**

Capture and sampling information has been described previously e.g. (refs 10,12). Sampling locations comprised 13 *E. helvum* roosting sites in continental Africa, and 14 in the four main islands in the Gulf of Guinea (Fig. 1, further detail in Table 1 (available online only)). In the majority of locations, data are from a single sampling event (sometimes comprising multiple sampling sessions within a one month period). Repeated sampling was conducted in Ghana (multiple sampling events per year over four years), Tanzania (one sampling event per year over two years) and Annobón (three sampling events over 4 years) as these locations were the focus of specific research studies. All fieldwork was undertaken under permits granted by national and local authorities (listed in Acknowledgements) and under ethics approval from the Zoological Society of London Ethics Committee (WLE/0489 and WLE/0467), using field protocols which followed ASM guidelines. Bats were captured at the roost with mist nets (6–18 m; 38 mm) as they departed the roost site at dusk, or returned at dawn. Except for a proportion of bats that were euthanased for virological studies (*n* = 238), bats were released following sampling. Additional samples and data were obtained from other research groups (*n* = 152) and in collaboration with local hunters in São Tomé (*n* = 102), where bats are hunted for human consumption.

Personal protective equipment (long clothing, face masks, eye protection and gloves) was worn during sample collection. Morphometric and demographic details were recorded from bats under manual restraint. Female reproductive status was assigned as non-reproductive, pregnant, or lactating, according to the descriptions provided in Table 2. The phase in the reproductive cycle (i.e. the time in months between the sampling date and the beginning of the last birthing season) was estimated based on published data and the pregnancy status of females (foetal size, assuming a true gestation period of 4 months (Mutere 1965)) or degree of juvenile development during sampling.

Age was assessed by morphological characteristics (Table 3) and all individuals were placed into one of four age classes: Neonate (N; < 2mths), Juvenile (J; 2– < 6 months), Sexually Immature (SI; 6– < 24 months) or Adult (A; ≥24 months). For a subset of samples, the timing of sampling in relation to the birthing season permitted further classification of SI individuals into 6-month age groups SL1, SL2 and SL3 (6 – < 12, 12– < 18, 18– < 24 months, respectively). Additionally, for bats that were hunted or euthanased following capture, upper canine teeth were extracted, air dried and shipped to the USA (Matson’s laboratory, USA) for histological examination to assess the number of tooth cementum annuli present. Following previous studies, it was assumed that each observed cementum layer represented

<table>
<thead>
<tr>
<th>Phase</th>
<th>Reproductive status</th>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–2</td>
<td>Lactating</td>
<td>L</td>
<td>Neonate attached and suckling, or milk able to be expressed from mammary glands of females</td>
</tr>
<tr>
<td>3–8</td>
<td>Non-reproductive</td>
<td>NR</td>
<td>Not detectably pregnant on abdominal palpation</td>
</tr>
<tr>
<td>9</td>
<td>Very early pregnant</td>
<td>VEP</td>
<td>Uterine bulge palpable on abdominal palpation, up to 1 cm diameter (detectable from ~ 0.5 cm diameter with careful palpation)</td>
</tr>
<tr>
<td>10</td>
<td>Early pregnant</td>
<td>EP</td>
<td>Uterine bulge palpable on abdominal palpation, 1–2 cm diameter</td>
</tr>
<tr>
<td>11</td>
<td>Mid pregnant</td>
<td>MP</td>
<td>Uterine bulge and foetus palpable on abdominal palpation, but no obvious abdominal distension of the female (2–3.5 cm)</td>
</tr>
<tr>
<td>12</td>
<td>Late pregnant</td>
<td>LP</td>
<td>Foetus palpable on abdominal palpation, with distension of the female’s abdomen so that it is wider than the base of the ribs</td>
</tr>
</tbody>
</table>

**Table 2.** Reproductive status classifications for female *E. helvum*. Phase represents the estimated time in months since the beginning of the last birthing season.
Each age estimation was scored with a certainty code: A: highest certainty of reported age (51% of samples, e.g. Fig. 3a,b), B: histological evidence supported a given age result ± 0.5–1.5 years (46% of samples, e.g. Fig. 3c), or C: tooth or section quality was too compromised to accurately age (3% of samples).

Genetic and blood samples were collected under manual restraint. Wing membrane biopsies (4-mm) were placed into 70% alcohol. Up to 1 ml blood was collected from the propatagial vein using a citrated 1 ml syringe and placed into a plain 1.5 ml eppendorf tube.

<table>
<thead>
<tr>
<th>Age class</th>
<th>Abbr.</th>
<th>Age (mths)</th>
<th>Timing of presence</th>
<th>Classification features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neonate</td>
<td>Neonate</td>
<td>&lt; 2</td>
<td>Up to 2 months after birth pulse</td>
<td>Suckling</td>
</tr>
<tr>
<td>Juvenile</td>
<td>Juvenile</td>
<td>2–&lt; 6</td>
<td>2-6 months after birth pulse</td>
<td>Small body size</td>
</tr>
<tr>
<td>Sexually Immature</td>
<td>SI</td>
<td>6 – &lt; 24</td>
<td>Adult size (or nearly adult size), but undeveloped nipples and genitils compared with adults</td>
<td>Early in this period, distinguishable from SL1 by smaller body size, and from Adults by lack of sexual development. In the last 3 months, females expected to be pregnant, but with undeveloped nipples.</td>
</tr>
<tr>
<td>Sexually Immature 1</td>
<td>SI.1</td>
<td>6 – &lt; 12</td>
<td>6–12 months after birth pulse</td>
<td>Early in this period, distinguishable from juveniles by size and adults by lack of sexual development.</td>
</tr>
<tr>
<td>Sexually Immature 2</td>
<td>SI.2</td>
<td>12 – &lt; 18</td>
<td>Up to 6 months after birth pulse</td>
<td>Early in this period, distinguishable from SI1 by larger body size, and from Adults by lack of sexual development. In the last 3 months this is less clear, however females expected to be pregnant with their first offspring (primiparous), but with undeveloped nipples.</td>
</tr>
<tr>
<td>Sexually Immature 3</td>
<td>SI.3</td>
<td>18 – &lt; 24</td>
<td>6–12 months after birth pulse</td>
<td>Early in this period, distinguishable from SI1 by larger body size, and from Adults by lack of sexual development. In the last 3 months this is less clear, however females expected to be pregnant with their first offspring (primiparous), but with undeveloped nipples.</td>
</tr>
<tr>
<td>Adult</td>
<td>Adult</td>
<td>≥ 24</td>
<td>All year</td>
<td>Full body size within ranges expected for an adult. Females have developed nipples, males with developed testicles.</td>
</tr>
</tbody>
</table>

Table 3. Age classification system used for *E. helvum* *°*SL.1 and SL.3 stages represent two different birth cohorts, overlapping in time. Where distinction between SL.1 and SL.3 was unclear during sampling and analysis of measurements, individuals were simply classified as ‘SI’ and omitted from extended age class analyses. °SI individuals become sexually active during the last 4 months of their second year, however the highly synchronous and seasonal nature of the birthing period means that it is simpler for individuals to be recognised as sexually mature only after the beginning of the birthing pulse. While pregnant SL.3 (primiparous) individuals are therefore not technically sexually ‘immature’, by classifying them as such, they can be differentiated from adult females (>2 years of age).

Figure 3. Histological sections of upper canine teeth from *E. helvum* for cementum age analysis (Giemsa stain). Photographs and captions courtesy of Gary Matson, Matson’s Laboratory, MT, USA. Each age estimation was scored with a certainty code: A: highest certainty of reported age, B: histological evidence supported a given age result ± 0.5–1.5 years, or C: tooth or section quality was too compromised to accurately age. (a) Bat ID 424. Cementum age 2, certainty code A. 100X. The tooth was in excellent histological condition, as indicated by the presence of periodontal membrane and good differential staining between anulli and light cementum. (b) Bat ID 62. Cementum age 6, certainty code A. 100X. Annuli are complex, with at least two components each year. A key feature of age analysis is resolving uncertainty about whether complex annulli or individual components are being used as age indicators. (c) Bat ID 44. Cementum age 13, certainty code B (13–15 yrs). 400X. The root tip of this tooth had been broken off during extraction. Missing cementum complicates age analysis, reducing the evidence available for evaluating whether anulli observed at one point may be clearly identifiable as components of complex annulli at another point.
**Molecular methods**

Molecular methods have been described previously\textsuperscript{12,20}. Genomic DNA was extracted from *E. helvum* tissues (predominantly wing membrane biopsies, but also liver and muscle samples, all stored in ethanol) using DNeasy Blood and Tissue Kits (QIAGEN Ltd., Crawley, West Sussex, UK). DNA was quantified using Quant-it PicoGreen dsDNA kits (Molecular probes, UK), and later using a Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, UK) and diluted to a standard concentration.

Twenty *E. helvum* loci developed in a previous study\textsuperscript{20} were quality-checked using a subset of samples. Loci E and Ae were discarded due to difficulty in scoring or high error rates and data from locus Ag were re-binned and re-scored, correcting earlier issues with allelic dropout. In total, 170 continental and 385 island samples were run as multiplex PCRs at 18 loci (TSY, FWB, MNQX, AgPK, AcAfAi, AdAh) in 10 μl PCRs, containing 4ng template DNA, 0.2 μM of each primer, and 5 μl Type-it Multiplex PCR Master Mix (QIAGEN Ltd.). Positive and negative controls were included on each plate and amplification was performed using the following conditions: 5 min at 94 °C; 30 cycles of 30 s at 95 °C, 90 s at 57 °C, and 30 s at 72 °C; then 30 s at 60 °C. Genotyping was performed by capillary electrophoresis using a Beckman CEQ 8000 (Beckman, UK). Allele sizes were scored automatically prior to manual verification. Genotyping data from 18 loci are provided in ‘Eidolon helvum data 2007–2014.csv’ [Data Citation 1].

Loci B has previously been identified as being X-linked\textsuperscript{20}.

Fragments of the mitochondrial DNA cytochrome b (cytb) gene were amplified from continental samples by PCR using the generic primers L14722 (5′-CGA AGC TTG ATA TGA AAA ACC ATC GTT G)\textsuperscript{32} and H15149 (5′-AAA CTG CAG CCC CTC AGA ATG ATA TTT GTC CTC A)\textsuperscript{33} in 20 μl reactions, containing 0.1–1 ng template DNA, 0.2 μM of each primer, 0.25 mM of each dNTP, 1.5 mM MgCl\textsubscript{2}, 0.25 μl of Taq polymerase (Invitrogen), and 0.2 μl 10X reaction buffer and with the following conditions: 5 min at 94 °C; 40 cycles of 1 min at 93 °C, 1 min at 54 °C, and 2 min at 72 °C; then 7 min at 72 °C. Although these generic primers were adequate with continental samples (8% PCR failure), amplification from isolated Gulf of Guinea island samples was less successful (48% PCR failure). Shortened primers (EhM2814 (5′-GCT TGA TAT GAA AAA CCA TCG TTG) and EhM2815 (5′-CAG CCC CTC AGA ATG ATA TTT GT) resulted in successful amplification when using Microzone MegaMix-Gold reagent (Microzone Ltd, UK). PCRs were performed in 20 μl reactions, containing 2 ng template DNA, 0.25 μM of each primer, and 10 μl MegaMix-Gold, using the following conditions: 5 min at 95 °C; 33 cycles of 30 sec at 95 °C, 30 sec at 53 °C, and 45 sec at 72 °C. PCR products were checked by gel electrophoresis on 1% agarose gels, purified using Exosap-IT clean-up (USB Europe, Germany) and sequenced in both directions on an ABI 3730xl DNA Analyser, (Applied Biosystems). Paired sequences were edited and aligned using the STADEN Package v1.6 (ref. 34). Multiple sequence alignment was performed using default settings in T-COFFEE\textsuperscript{35}. Sequences were checked manually and trimmed to a standard length (397 bp) in JALVIEW v2 (ref. 36). No sequence differences were detected in 38 samples sequenced using both primer pairs, so data were combined.

Data from 608 and 544 individuals is available for cytb and microsatellite analyses (at 18 loci), respectively (Table 1 (available online only)).

![Figure 4](SCIENTIFIC_DATA/3/160049/DOI: 10.1038/sdata.2016.49)

**Figure 4.** Correlation between ln(MFI) values pre- and post-repair of the Luminex machine used to run the assays, based on 293 samples. The linear regression line is in red. ($R^2$0.81, F-statistic: 1306 (1, 296), $P < 2.2e$–16).
Serological analyses

Serological methods have been described previously\textsuperscript{7,10–14}. A modified fluorescent antibody virus neutralization (mFAVN) assay using the LBVNig56 isolate was used to detect neutralising antibodies against LBV\textsuperscript{10,37}. Samples were tested in duplicate using threefold serial dilutions (representing reciprocal titres of 9, 27, 81, and 243–19,683). Human rabies immunoglobulin, LBV-positive rabbit serum, and rabies-vaccinated mouse serum were used as positive controls and negative rabbit and mouse serum were negative controls. Titres were considered positive at IC100 endpoint reciprocal dilutions >1: 9 (100% neutralisation of virus).

Henipavirus antibodies detected in African fruit bat samples using virus neutralisation assays, multiplexed microsphere assays and pseudotype assays developed to target other known henipaviruses (Hendra and Nipah viruses) and are presumed to represent cross-neutralisation or cross-reactivity\textsuperscript{12}. Here, Luminex multiplexed microsphere binding assays were used to detect antibodies against henipaviruses (HeV and NiV). In these assays, purified recombinant expressed henipavirus soluble G glycoproteins\textsuperscript{38} are conjugated to internally coloured and distinguishable microspheres, allowing
multiplexing. For African bat samples, stronger results were consistently observed in NiV binding assays and virus neutralisation tests\(^{13}\), so only NiV binding assay results are included in the dataset. Binding results are outputted as median fluorescence intensity (MFI) values of at least 100 microspheres for each virus type. In mid-December 2010, major repair work was undertaken on the Luminex machine being used for serological analysis. A subset of samples that had been analysed before the repairs were repeated to calibrate results (\(n = 293\)). MFI values pre- and post-repair work were significantly different, making the use of a single cutoff inappropriate\(^{17}\). Two approaches were taken to designate results as seropositive or seronegative. First, a Bayesian mixture model was applied as described in\(^{17}\). Cutoffs for pre- and post-repair work were determined so that samples above this cutoff were \(\geq 99\%\) likely to be in the seropositive distribution (MFI = 156.1 and 127.5, respectively). Second, linear regression of pre- and post-maintenance \(\ln\) (MFI) values demonstrated a significant linear relationship (Fig. 4, \(R^2 = 0.81\), F-statistic: 1306 (1, 296), \(P < 2.2e-16\)), and the variance decreases for higher MFI values (above the cutoff). Pre-maintenance MFI values were converted to post-maintenance values using the formula:

\[
\text{NEW\_MFI} = \exp(0.7795774 \times \ln(\text{OLD\_MFI}) + 0.4392832).
\]

The Bayesian mixture model was applied to this transformed and combined data using the same method. From this analysis, MFI values \(> 94.2\) were \(\geq 99\%\) likely to be in the seropositive distribution. Results from the two methods were compared and the second method resulted in the highest congruence between pre- and post-maintenance paired results (congruence in 266/298 samples versus 250/298 samples for the first method), and these data were therefore used in the final dataset. Raw MFI values are available on request.

Antibodies against Achimota viruses 1 and 2 were detected using virus neutralisation assays\(^{7}\), with all testing in duplicate. Samples were diluted to 1:20 and incubated with 200 TCID\(_{50}\) of virus for 30 min at 37°C prior to the addition of Vero cell suspension at an MOI equivalent to 0.01. Cell monolayers were assessed for evidence of virus neutralization 7 days post infection. Where sample volume permitted, positive samples were titrated in a 2-fold dilution series from 1:20 to 1:160 and retested using the same protocol.

### Data Records

The data are contained in a single comma-separated file (.csv format), entitled ‘Eidolon helvum data 2007–2014’ (Data Citation 1). Each row below the header represents an individual bat (\(n = 2,827\)), and the columns (\(n = 68\)) contain sample identifier information, demographic and morphometric data, and results of genetic and serological assays. Full descriptions of the column titles are included in the Table 4 (available online only).

### Technical Validation

#### Molecular analyses

Recommendations for minimisation and assessment of errors that may occur during the sampling, DNA extraction, amplification, sequencing, genotyping and data analysis processes were followed where possible\(^{39}\) (Table 5).

As previously described\(^{20}\), microsatellite loci were tested for evidence of departure from Hardy–Weinberg equilibrium (HWE) and genotypic disequilibrium using FSTAT 2.9 (ref. 40), with appropriate Bonferroni corrections for multiple testing. All loci were analysed in MICRO-

<table>
<thead>
<tr>
<th></th>
<th>cytb original primers</th>
<th>cytb new primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of samples</td>
<td>207</td>
<td>438</td>
</tr>
<tr>
<td>Number of samples repeat PCR and sequenced</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>Number of duplicates run per sample</td>
<td>1</td>
<td>70</td>
</tr>
<tr>
<td>Total number of repeats</td>
<td>28</td>
<td>70</td>
</tr>
<tr>
<td>% of sequences repeated</td>
<td>13.5%</td>
<td>16.0%</td>
</tr>
<tr>
<td>Number of base pairs in sequence</td>
<td>418</td>
<td>397</td>
</tr>
<tr>
<td>Number of repeated sequences</td>
<td>28</td>
<td>70</td>
</tr>
<tr>
<td>Number of base pairs compared</td>
<td>11704</td>
<td>27790</td>
</tr>
<tr>
<td>Number of sequence errors (mismatches)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>% PCR/sequence error (No of mismatches/no bp compared)</td>
<td>0.01%</td>
<td>0%</td>
</tr>
<tr>
<td>% samples extracted in duplicate, cytb PCR and sequenced</td>
<td>2.4%</td>
<td></td>
</tr>
<tr>
<td>Number of errors (mismatches)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>% error</td>
<td>0.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Mitochondrial DNA PCR repeat rates and sequencing error rates.

---

www.nature.com/sdata/
CHECKER\textsuperscript{41} to test for null alleles, stuttering and large allelic dropout as a cause of departure from HWE. Additionally, since Locus M displayed extremely low polymorphism (99.1% of individuals were homozygous for a particular allele), this locus was included in all PCR plates as a positive control and to determine inter-assay variability in allele fragment length. Error rates for microsatellite loci are reported in Peel et al.\textsuperscript{20} Inter-assay genotyping variability, measured by the variation in fragment length of the dominant allele of locus M on each plate, was low (range 134.32 – 134.66) across 27 runs and two control samples. Loci Y proved difficult to confidently bin due to alleles of single nucleotide difference and was therefore not included in the dataset.

Error rates for cyt b analyses were assessed by replicate extractions (performed on 2.4% of samples), replicate PCR and sequencing reactions (performed on 8–14% of extracted samples), and by inclusion of positive and negative controls for all extractions and PCRs. Poor quality mtDNA sequence traces were excluded. Background PCR and sequencing error rates of the new E. helvum cyt b primers EhM2814 and EhM2815 were assessed by running 70 replicates of a single sample. PCR and sequencing error rates were calculated at the base-pair level. Sequencing error rate was negligible (0–0.01%) across samples repeated in duplicate, and no substitutions were observed in the 70 replicate sequences obtained from a single sample (Table 6).

Serological analyses
All serological assays included positive and negative controls. Samples were tested in duplicate (LBV and Achimota viruses) or with 100 replicates (henipaviruses). Further validation procedures for multiplexed microsphere binding assays are presented as part of the methods, above.

Usage Notes
Users of these data are advised that importing the csv data file (Data Citation 1) into Microsoft Excel can result in formatting errors, particularly with the column ‘Teeth.Age.Range’. Rather than opening the file with Excel (by double-clicking, for example), it is suggested that users instead select ‘File>Import>csv file >Delimited’, then select the ‘Teeth.Age.Range’ column and set the column data format as ‘Text’. Alternatively, importing and processing the data into the software ‘R’\textsuperscript{42} may be preferable.

References
Acknowledgements

The authors thank the governments of Ghana, Tanzania, Malawi, Zambia, Uganda, Equatorial Guinea, and São Tomé and Príncipe for facilitating this research. For their invaluable support in planning and implementing the field work, we also thank the Wildlife Division of the Forestry Commission, and the Veterinary Services Directorate, Ghana; Zambian Wildlife Authority; Kasanka Trust; Malawi Ministry of Tourism, Wildlife and Culture, Malawi Ministry of Agriculture, Irrigation and Food Security; former Tanzania Ministry of Livestock and Fisheries Development (now Ministry of Agriculture, Livestock and Fisheries); Tanzania Wildlife Research Institute; Sokoine University of Agriculture, Tanzania; Makerere University, Uganda; Equatorial Guinea Ministerio de Agricultura y Bosques; Equatorial Guinea Instituto del Desarrollo Forestal y Gestión de las Áreas Protegidas; Universidad Nacional de Guinea Ecuatorial; São Tomé and Príncipe Ministério de Agricultura, Desenvolvimento Rural e Pesca; Ecosistémies Forestiers d’Afrique Centrale; Associação Monte Pico, São Tomé and Príncipe; Iñaki Rodriguez-Prieto; Alex Torrance; Lucrecia Bilé Osa Ahara; Heidi Ruffler; Ricardo Castro César de Sá; Ricardo Faustino de Lima; Mariana Carvalho and Meyir Ziekah. Additional genetic samples from the Gulf of Guinea islands and Rio Muni were supplied by Javier Juste (Museo Nacional de Ciencias Naturales, CSIC) and DJ Long (California Academy of Sciences). Additional sera were supplied by Ivan Kuzmin (previously Centers for Disease Control and Prevention). Christopher Broder (Unified Services University of the Health Sciences) supplied HeV and NiV soluble G glycoprotein for the henipavirus serological assays. The authors are indebted to Linfa Wang, Gary Crameri, Jennifer Barr and others at the CSIRO Australian Animal Health Laboratory, Geelong, Australia and Daniel Horton and David Seldon at the Animal Health and Veterinary Laboratories Agency (now Animal and Plant Health Agency), UK for providing invaluable support, constructive discussions and technical assistance throughout the project.

Funding for this study was provided by the Cambridge Infectious Diseases Consortium (grant VT0105) (A.J.P, D.T.S.H., K.S.B.), The Charles Slater Trust (A.J.P.), Zebra Foundation for Veterinary Zoological Education (A.J.P.), Isaac Newton Trust (A.J.P.), the Wellcome Trust (D.T.S.H., K.S.B.), a David H. Smith postdoctoral fellowship (D.T.S.H.), the RAPIDD program of the Science and Technology Directorate, Department of Homeland Security, Fogarty International Center, National Institutes of Health (D.T.S.H., J.L.N.W., A.R.F.), The Alborada Trust (J.L.N.W.), a European Union FP7 project ANTIGONE (Anticipating Global Onset of Novel Epidemics 278976) (A.C.B., J.L.N.W., A.A.C., A.R.F.), Royal Society Wolfson Research Merit Award (A.A.C.), Lincoln Park Zoo, Chicago, US (T.L.) and a Wellcome Trust Value in People award through the University of Glasgow (T.L.). KSB is in receipt of a Wellcome Trust Postdoctoral Training fellowship for clinicians (106690/Z/14/Z) and is supported by Wellcome Trust grant number 098051.

Data Citation

Author Contributions
A.J.P. conceived the research, conducted fieldwork, conducted genetic analyses, conducted serological assays, curated data file, and wrote the manuscript. J.L.N.W. conceived the research. K.S.B. conceived the research, conducted fieldwork, conducted serological assays, and edited the manuscript. A.C.B. conceived the research, and conducted fieldwork. G.C.G. conducted fieldwork. T.L. conducted fieldwork and edited the manuscript. R.S. conducted fieldwork. A.A.C. conceived the research. D.T.S.H. conceived the research, conducted fieldwork, conducted serological assays. A.F.L. conducted fieldwork. D.R.S. conceived the research and edited the manuscript. A.R.F. conceived the research and conducted serological assays.

Additional Information
Tables 1 and 4 are only available in the online version of this paper.

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


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

Metadata associated with this Data Descriptor is available at http://www.nature.com/sdata/ and is released under the CC0 waiver to maximize reuse.
Bat trait, genetic and pathogen data from large-scale investigations of African fruit bats, Eidolon helvum

Peel, AJ

2016-08-01