MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 6134–6147 doi: 10.1111/mec.13472

New evidence for hybrid zones of forest and savanna elephants in Central and West Africa

SAMRAT MONDOL,*[a,1] IDA MOLTKE,†[a,2] JOHN HART,‡ MICHAEL KEIGWIN,§ LISA BROWN,¶ MATTHEW STEPHENS† and SAMUEL K. WASSER*

**Center for Conservation Biology, Department of Biology, University of Washington, Box 351800, Seattle, WA 98195-1800, USA, †Department of Statistics and Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA, ‡Lukuru Wildlife Research Foundation, 1235 Poids Lourds, Kinshasa, Democratic Republic of Congo, §Uganda Conservation Foundation, Plot 4, Henlon Close, Luthuli Avenue, Bugolobi, P.O. Box 34020, Kampala, Uganda, ¶Department of Biostatistics, University of Washington, Box 357232, Seattle, WA 98195-7232, USA*

Abstract

The African elephant consists of forest and savanna subspecies. Both subspecies are highly endangered due to severe poaching and habitat loss, and knowledge of their population structure is vital to their conservation. Previous studies have demonstrated marked genetic and morphological differences between forest and savanna elephants, and despite extensive sampling, genetic evidence of hybridization between them has been restricted largely to a few hybrids in the Garamba region of northeastern Democratic Republic of Congo (DRC). Here, we present new genetic data on hybridization from previously unsampled areas of Africa. Novel statistical methods applied to these data identify 46 hybrid samples – many more than have been previously identified – only two of which are from the Garamba region. The remaining 44 are from three other geographically distinct locations: a major hybrid zone along the border of the DRC and Uganda, a second potential hybrid zone in Central African Republic and a smaller fraction of hybrids in the Pendjari–Arli complex of West Africa. Most of the hybrids show evidence of interbreeding over more than one generation, demonstrating that hybrids are fertile. Mitochondrial and Y chromosome data demonstrate that the hybridization is bidirectional, involving males and females from both subspecies. We hypothesize that the hybrid zones may have been facilitated by poaching and habitat modification. The localized geography and rarity of hybrid zones, their possible facilitation from human pressures, and the high divergence and genetic distinctness of forest and savanna elephants throughout their ranges, are consistent with calls for separate species classification.

Keywords: conservation, Elephant hybridization, forest elephant, genetic population structure, savanna elephant

Received 23 July 2015; revision received 7 November 2015; accepted 10 November 2015

Correspondence: Samuel K. Wasser, Fax: +1 206-221-5253; E-mail: wassers@u.washington.edu

[1]Present address: Wildlife Institute of India, Chandrabani, Dehradun, Uttarakhand, India

[2]Present address: Department of Biology, University of Copenhagen, Copenhagen, 2200, Denmark

[a]Equal contributions.

Introduction

The African elephant consists of two subspecies: forest elephants (*Loxodonta africana cyclotis*) and savanna elephants (*L. a. africana*). They are collectively listed by CITES as a single Appendix I species in most countries, classifying them among the world's most

endangered species (CITES 2014). Current rates of population decline, coupled with accelerating loss and fragmentation of range, could eliminate the majority of Africa's elephants in the next decade, with considerable ecological and economic impacts for the continent (UNEP *et al.* 2013).

There has been extensive debate over reclassification of forest and savanna elephants as separate species (Groves & Grubb 2000; Grubb *et al.* 2000; Roca *et al.* 2001, 2015; Comstock *et al.* 2002) for morphological, genetic, ecological and conservation reasons. On the one hand, forest and savanna elephants are easily distinguished both morphologically (Groves & Grubb 2000; Grubb *et al.* 2000) and genetically (Roca *et al.* 2001, 2015; Comstock *et al.* 2002). They show high levels of genetic differentiation (FST = 0.94) (Roca *et al.* 2001), live in different habitats and diverged >2.5 million year ago (Roca *et al.* 2001). Furthermore, although forest and savanna elephants occasionally interbreed where their ranges meet (Backhaus 1958; Groves & Grubb 2000), a variety of genetic evidence (Roca *et al.* 2015), including the rarity of hybrids reported to date (Roca *et al.* 2001; Comstock *et al.* 2002), support the reclassification into two species. On the other hand, conservation concerns surrounding their reclassification are less clear-cut in the current severe poaching climate. The Center for Biological Diversity (2015) petitioned the US Fish and Wildlife Service for separate species listings under the Endangered Species Act, while the International Union for Conservation of Nature is calling for further research on the matter (IUCN 2015).

Here, we present new genetic data and analyses that address several previously unanswered questions about hybridization between forest and savanna elephants: Are hybrids fertile, and do they involve genetic contributions from both sexes of both subspecies? We also discuss how the occurrence of hybrids impacts the classification of forest and savanna elephants into separate species, and possible reasons for the restricted geographic distribution of hybrids.

Results are based on combining genetic data on newly collected samples from previously unsampled locations, with a large existing genetic database of elephant samples from across Africa used for DNA-based assignment of poached ivory (Wasser *et al.* 2015). We collected elephant DNA samples from three previously unsampled forest–savanna ecotones in Central and West Africa: along the border of DRC and Uganda, which is in the northern part of the Albertine Rift, extending south to the northern tip of Rwanda; the Manovo-Gounda-St Floris complex in the northern part of the Central African Republic (CAR); and the Pendjari–Arli complex on the border of Benin and Burkina Faso. The collection of new samples was driven by the hypothesis that the combination of severe poaching and habitat loss in contact zones between forest and savanna habitats in Africa will push elephants from more disturbed areas into less disturbed areas with different habitat, leading to zones of hybridization. Hybridization and introgression are common in response to human disturbance in plants and vertebrates, and particularly where there is limited gene flow between populations (Rhymer & Simberloff 1996). All three regions above have mixed forest–savanna ecotones and have been heavily affected by poaching and habitat loss. In particular, both sides of the DRC–Uganda border experienced severe poaching in the 1970s up to the early 1980s, after which poaching pressures became asymmetric, remaining severe only on the DRC side (Kayanja & Douglas-Hamilton 1983; Keigwin 2001; Mubalama & Mushenzi 2004; Keigwin *et al.* 2016; Wasser *et al.* 2015). Many areas along the DRC–Uganda border including the Albertine Rift have also had a long history of high-density human occupation, likely to have impacted elephant distributions, and continue to experience high human population growth (The Worldwatch Institute 2014; Uganda historical demography data 2014). Northern CAR's elephants suffered severe poaching in the 1970s and 1980s followed by further pressure on remnant populations over the past decade (Bouche *et al.* 2010, 2011). The neighbouring population to the east in southern Sudan has experienced persistent poaching pressure associated with warfare since the early 1970s, whereas poaching pressures to the north in Chad have been most intensive over the past two decades (Hammer 2014). By contrast, West African populations experienced their heaviest losses from poaching over a century ago, with perpetual habitat fragmentation since that time (Barnes 1999).

Using a panel of 16 microsatellite loci, mtDNA and Y chromosome markers, and novel statistical analyses, we identified varying degrees of forest–savanna hybridization across these three newly sampled regions. Through additional analyses of the identified hybrids, we find that most hybrids are at least second generation, implying that hybrids are fertile. Moreover, analyses of Y chromosome and mtDNA data demonstrate that hybridization has been bidirectional, involving fathers and mothers of both subspecies. Our discussion addresses the consequences of these findings for conservation of the African elephant, and for the ecosystems they inhabit.

Materials and methods

Sample collection

Dung samples used in this study were collected between 2000 and 2014 as part of African elephant DNA forensic database by the Center for Conservation

Biology, University of Washington (Wasser *et al.* 2015). Targeted dung sampling in the mixed forest–savanna habitat zones along the DRC–Uganda border (209 samples), northern CAR (8 samples) and the Pendjari–Arli complex along the Benin–Burkina Faso border (74 samples) was conducted between 2010 and 2014. All consecutive samples were collected ≥1 km apart to minimize chances of obtaining multiple samples from the same family group. Dung samples were stored in 20% DMSO-TNE buffer in sealed vials in the field and shipped to our laboratory at the University of Washington, where they were stored at −20 °C until extracted. All samples in this study were transported in full compliance with USDA and CITES regulations. A total of 1369 elephant samples, genotyped at 10 or more loci, from 79 locations were used in this study.

DNA extraction and amplification

DNA extraction protocols for dung and tissue are described in Wasser *et al.* (2004). In brief, each dung sample was extracted for DNA in duplicate with a Qiagen Stool DNA kit (Qiagen, Valencia, CA, USA) followed by a Gene Clean III nucleic acid isolation kit (MP Biomedicals, Irvine, CA, USA). In all cases, a negative control was included for every set of 10 extractions.

Each DNA extract was amplified twice for a panel of 16 dinucleotide microsatellite loci described in Comstock *et al.* and others (Nyakaana & Arctander 1998; Comstock *et al.* 2000, 2002). All loci exhibited linkage equilibrium. As allelic dropout and/or null alleles can cause deviations from Hardy–Weinberg equilibrium, we explicitly accounted for this in our analyses. The amplification protocol was slightly modified from Wasser *et al.* (2004). PCRs were performed in 10 μL reaction volumes with 5 μL of Qiagen multiplex mix (Qiagen, Valencia, CA, USA), 0.2 μM of primer mix, 1 μL of BSA and 2 μL of faecal DNA extract. Positive and negative PCR controls were included in all analyses. Two microlitre of the PCR product was subjected to fragment analysis in an ABI 3100 capillary array genetic analyser. Following our ongoing forensic analysis protocol (Wasser *et al.* 2015), allele sizes were scored using GENEMARKER v.2.4 with categories defined for each allele size bin with a 0.5-bp tolerance. In brief, an allele had to be observed by itself at least three times to be confirmed as a homozygote. Both alleles had to be observed a minimum of two times to be confirmed as a heterozygote. Otherwise, both alleles were scored as missing.

We used molecular sexing tools, as detailed in Ahlering *et al.* (2011; see also Mondol *et al.* 2014) to sex representative samples from pure forest (13 samples with posterior probability of being a pure forest elephant >0.95), pure savanna (18 samples with posterior probability of being a pure forest elephant >0.95) and all potential hybrid samples that still had sufficient DNA (35 samples with hybrid probability (HP) >0.95 and one with HP = 0.74) (calculations are presented below in the section: *Methods for identifying and classifying hybrids*). Two labelled Y chromosome markers (SRY and AMELY) and one labelled X chromosome marker (PLP) were multiplexed together in a single PCR. Each faecal sample was amplified in 10 μL volumes with 5 μL of Qiagen multiplex mix (Qiagen, Valencia, CA, USA), 4 μg of BSA, 0.2 μM of primers and 3 μL of faecal DNA extract. The PCR conditions included an initial denaturation at 95 °C for 15 min, followed by 55 cycles of denaturation at 95 °C for 30 s, annealing at 58 °C for 30 s and extension at 72 °C for 45 s, and a final extension of 72 °C for 30 min. Negative control and one male and female positive control were included in each set of reactions. Post-PCR, 1 μL of product was mixed with formamide and run on an ABI 3100 genetic analyser with ROX400 size standard (Applied Biosystems). PCR product sizes were visualized using GENEMARKER v.2.4 and analysed using an allelic bin created specifically for elephant sexing markers. For unambiguous sex identification, the procedure was repeated a total of three times to ensure data quality.

We then sequenced the mitochondrial control region (526 bp) (Debruyne 2005; Roca *et al.* 2005; Johnson *et al.* 2007) for all of the molecular sexed samples and the Y-chromosomal amelogenin gene (719 bp) (Roca *et al.* 2005) of the subset of the sexed samples that were determined to be males. For the amelogenin gene, we designed two new sets of primers to amplify smaller overlapping fragments from dung.

(AmelY-F1-CGCTTGATTCTAAGACCTTTGCGCT;
AmelY-R1-CAGACCATATCTGAATCTATGAGGC;
AmelY-F2- GCATGAAGGAAACTCTGAATCTGAG;
AmelY-R2- TAGAATCACCATCACTTTTACCGAC).

Amplification was performed in 20 μL reaction volumes with 7.5 μL of multiplex mix (Qiagen, Valencia, CA, USA), 4 μg of BSA, 0.2 μM of primer mix and 3 μL of faecal DNA extract. The temperature regime included an initial denaturation (94 °C for 15 min); 40 cycles of denaturation (94 °C for 30 s), annealing (58 °C for 45 s) and extension (72 °C for 45 s); followed by a final extension (72 °C for 30 min). Both positive and negative PCR controls were included in all amplifications to monitor contamination. All mtDNA and Y-chromosomal fragments were amplified once from both DNA elutions and then sequenced bidirectionally. Sequences were aligned using MEGA v5.0 (Tamura *et al.* 2011) and analysed for frame-shift mutations prior to conducting any further analyses. All sequence data were submitted to GenBank (Accession no: KF534810–KF534908).

Data filtering

We first removed all samples that did not amplify at 10 or more loci, as per our forensic protocol (Wasser *et al.* 2015). Then, to reduce the number of potential duplicate samples, we used an iterative procedure to remove samples that differed by less than three alleles across all loci. This procedure removed 83 samples, leaving 1369 samples that differ from one another by at least three alleles. The threshold of three alleles was chosen after examining histograms of the pairwise differences between each sample and its closest match, and noting an excess of samples with two or fewer differences with their closest match.

STRUCTURE *analyses*

We performed three analyses using the program STRUCTURE (Pritchard *et al.* 2000): one analysis of all 1369 samples depicted in Fig. 1(B) (data set I), one including the 219 samples from the NE part of DRC, Uganda and Rwanda, depicted in Fig. 1(C) (data set II) and one including the 260 samples from western DRC, Central Africa Republic (CAR), Congo, Gabon, Cameroon and eastern Nigeria, depicted in Fig. 1(D) (data set III). The first of these three analyses was performed to get an initial overview of the population structure in the data set, with results shown in Fig. 1(A). Further, all three analyses were used to identify sets of putative pure forest and putative pure savanna elephants that were used as reference samples in the hybrid analyses described below. Both forest and savanna reference sets included all samples with a respective ancestry proportion larger than 0.95. Unlike the areas depicted in Fig. 1(C) and 1(D), the area depicted in Fig. 1(E) (Ghana, Togo, Benin, Burkina Faso) was not analysed separately because our analyses of the full data set (data set 1) revealed that there were no savanna elephants from this area in our data set. A separate hybrid analysis of this region was thus not possible, as it requires reference samples from both forest and savanna populations.

All the analyses were performed using STRUCTURE version 2.3.2.1 with the number of populations, K, set to 2, and the parameters INFERALPHA and POPALPHAS set to TRUE. To ensure convergence, we ran 5 MCMC chains for each approach, each with a burn-in of 200 000 followed by 4 000 000 iterations.

Methods for identifying and classifying hybrids

We implemented an 'empirical Bayes' method to quantify the support for each sample being a pure forest elephant, a pure savanna elephant or a hybrid, given a set of pure forest or savanna 'reference' samples (software available at https://github.com/stephenslab/EBhybrids). In brief, we first extended the likelihood from Wasser *et al.* (2004) to distinguish first- and second-generation hybrids. We then embedded this likelihood in a hierarchical model that combined information across samples to estimate the proportion of each type of hybrid, and used these estimated proportions to compute a posterior probability that each sample is either of pure or hybrid type. Importantly, this method allows for both genotyping error and allelic dropout. Details of each step follow.

The likelihood. Let g denote the genotypes of a given sample with alleles at locus l (=1, …, L) being (g_{l1}, g_{l2}) and let f denote the frequencies of all alleles with f_{jlk} being the frequency of allele j at locus l in subspecies k. Here, k takes the values 'forest' and 'savanna', but for generality we denote the values s_1 and s_2. Following Wasser *et al.* (2004), we estimated the allele frequencies from the reference samples using $f_{jlk} = \frac{n_{jlk}+1}{n_{+lk}+m_l}$ where n_{jlk} is the number of times allele j is observed at locus l in reference samples for subspecies k, $n_{+lk} = \sum_j n_{jlk}$ and m_l is the number of observed alleles across both subspecies at locus l. When computing likelihoods, we obtained the estimates of f from reference data sets consisting of pure forest and pure savanna elephants, respectively, identified using STRUCTURE (see above section on STRUCTURE analyses). Because likelihoods from all samples from an area are needed when estimating the parameters of the hierarchical model described later, we computed likelihoods, not only for the potential hybrids, but also for the samples in the reference data sets. When computing likelihoods for a sample from the reference data sets, we excluded that particular sample before estimating f.

Each hybrid type is characterized by a probability vector $q = (q_0, q_1, q_2)$, where q_w is the probability of having w alleles from subspecies s_2 at any locus. Specifically:

- A pure s_1 sample is characterized by q = (1,0,0)
- A pure s_2 sample is characterized by q = (0,0,1)
- An F1 sample (first-generation hybrid) is characterized by q = (0,1,0)
- An F2 sample (offspring of two F1s) characterized by q = (0.25,0.5,0.25)
- A backcross to s_1 (offspring of an F1 and a pure s_1) is characterized by q = (0.5,0.5,0)
- A backcross to s_2 (offspring of an F1 and a pure s_2) is characterized by q = (0,0.5,0.5)

These are all the hybrid types we considered here, but other values for q (e.g. third generation hybrids) could also be considered.

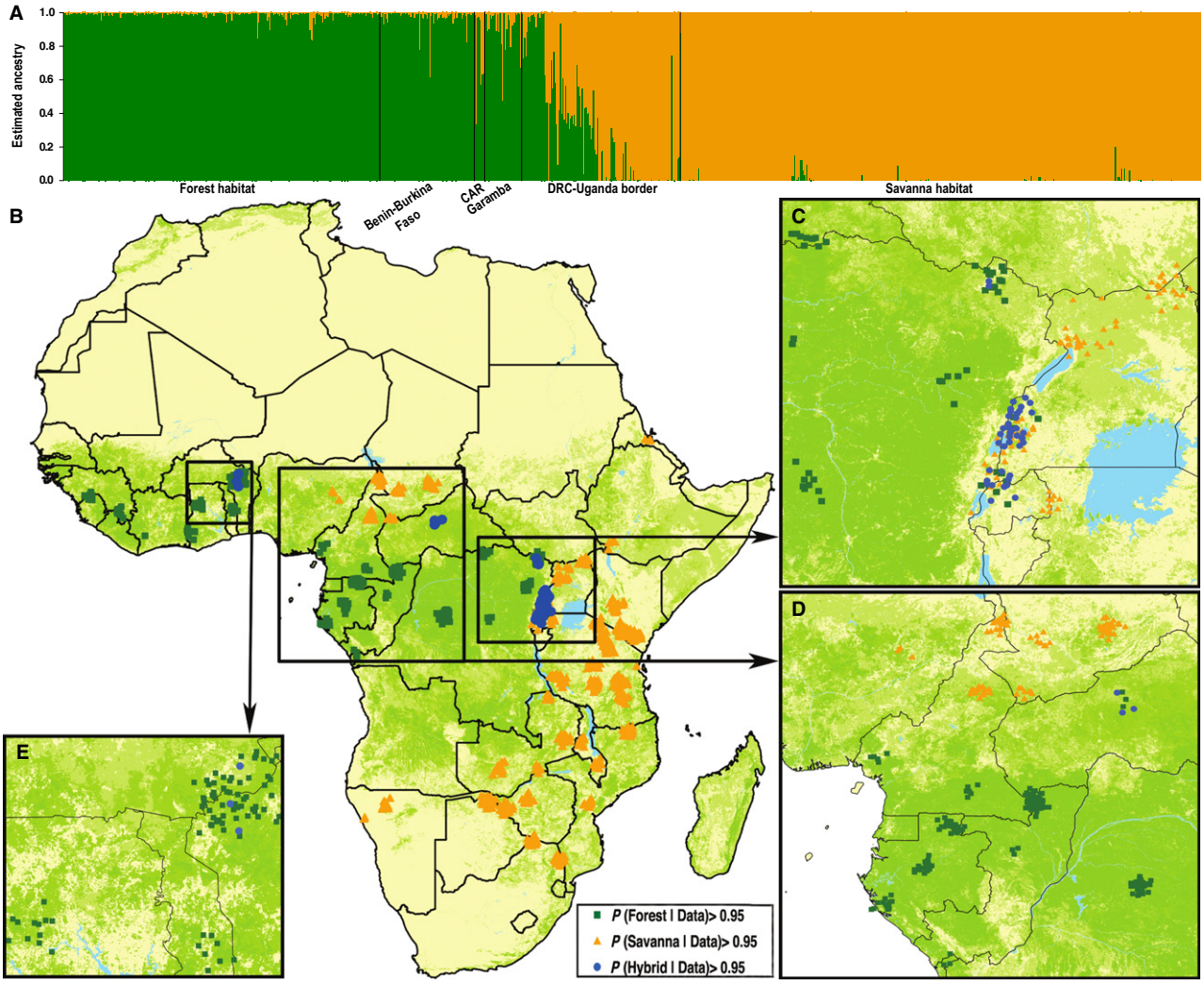

Fig. 1 Elephant STRUCTURE and hybrid analysis results. (A) Plot of ancestry proportions for all the samples estimated using STRUCTURE. The samples are ordered according to sampling location. The highlighted sampling locations in CAR, Garamba National Park, DRC–Uganda border and Benin-Burkina Faso contain numerous samples with estimated ancestry proportions between 0.2 and 0.8. Other sampling locations contain almost exclusively 'pure' forest or savanna samples. (B) Estimated hybrid probabilities (HP) for all samples from across Africa. (C) Estimated hybrid probabilities (HP) for the Uganda–DRC border (as well as Garamba National Park in northeast part of DRC and the Rwanda-DRC border to the south). (D) Estimated hybrid probabilities (HP) for CAR, extending through the western part of Central Africa. (E) Estimated hybrid probabilities (HP) for West Africa. Note that the HPs in B and E are from our 'global' hybrid analysis, that is they are computed using reference samples from across Africa, whereas HPs in C and D are from our local hybrid analyses, that is computed using reference samples only from those areas (see main text). In all of B–E, only samples whose hybrid/pure status could be confidently inferred are shown (HP > 0.95 or HP < 0.05).

Assuming linkage equilibrium within each subspecies, the likelihood for q (i.e. the probability of observing g, given q) can be written as:

$$L(q; g, f, \delta, \gamma_l) = \prod_{l=1}^{L} p(g_{l1}, g_{l2} | q, f, \delta, \gamma_l)$$

where δ denotes an overall genotyping error probability and $\gamma_l = (\gamma_{ls1}, \gamma_{ls2})$ denotes the allelic dropout (and/or null allele) rate in each subspecies in locus l. We assume Hardy–Weinberg equilibrium in each subspecies, hence

$$\begin{aligned} p(i,j|q,f,\delta,\gamma_l) = {} & q_0 \cdot (1-\gamma_{ls1}) \cdot p(i|k=s_1,f,\delta) \\ & \cdot p(j|k=s_1,f,\delta) \cdot 2 + \\ & q_2 \cdot (1-\gamma_{ls2}) \cdot p(i|k=s_2,f,\delta) \\ & \cdot p(j|k=s_2,f,\delta) \cdot 2 + \\ & q_1 \cdot (1-\gamma_{ls12}) \cdot [p(i|k=s_1,f,\delta) \cdot \\ & p(j|k=s_2,f,\delta) + p(i|k=s_2,f,\delta) \cdot \\ & p(j|k=s_1,f,\delta)] \end{aligned}$$

when $i \neq j$ and

$$\begin{aligned} p(i,j|q,f,\delta,\gamma_l) = q_0 \cdot [&(1-\gamma_{ls1}) \cdot p(j|k=s_1,f,\delta) \cdot p(j|k=s_1,f,\delta) \\ &+\gamma_{ls1} \cdot p(j|k=s_1,f,\delta)]+ \\ q_2 \cdot [&(1-\gamma_{ls2}) \cdot p(j|k=s_2,f,\delta) \cdot p(j|k=s_2,f,\delta) \\ &+\gamma_{ls2} \cdot p(j|k=s_2,f,\delta)]+ \\ q_1 \cdot [&(1-\gamma_{ls12}) \cdot p(j|k=s_1,f,\delta) \cdot p(j|k=s_2,f,\delta) \\ &+\gamma_{ls12} \cdot 0.5 \cdot (p(j|k=s_1,f,\delta)+p(j|k=s_2,f,\delta))] \end{aligned}$$

when $i = j$ with

$$p(j|k,f,\delta) = (1-\delta)f_{jlk} + \frac{\delta}{m_l} \quad (1)$$

Here, m_l is the number of observed alleles across both subspecies at locus l, *and* γ_{ls12} is the average of γ_{ls1} and γ_{ls2}.

Parameter estimation. We estimated the genotyping error probability, δ, and the locus- and subspecies-specific allelic dropout rates, γ_{ls1} and γ_{ls2}, from the reference samples identified by a STRUCTURE analyses of the entire data set except for the 74 samples from the Pendjari–Arli complex along the Benin–Burkina Faso border region (as we collected those samples much later than all other samples and after performing the parameter estimation). We first estimated δ by maximizing the likelihood on a grid of δ values ranging from 0 to 0.15 and γ values ranging from 0 to 0.5, using a step size of 0.005, and with all γ values fixed to the same value across loci. For both forest and savanna reference samples, the maximum-likelihood estimate was δ = 0. We therefore fixed δ to 0. We then performed maximum-likelihood estimation of γ_{lk} for each locus separately for the forest and savanna reference samples. The inferred values are shown in Figure S1 (Supporting information) and are highly correlated with subspecies and marker-specific estimates of the inbreeding coefficient F (Figure S2, Supporting information).

Hierarchical model. To estimate the proportion of samples of each hybrid type, we used a hierarchical model to combine information across all samples. Specifically, let $\pi(q)$ denote the proportion of samples of type q, then the likelihood for π, assuming samples $i = 1, \ldots, n$ to be independent, is given by

$$L(\pi; g_1, \ldots, g_n) = p(g_1, \ldots, g_n|\pi) = \prod_{i=1}^{n} \sum_{q} \pi(q) p(g_i|q)$$

We estimated π by maximum likelihood, using the EM algorithm.

Posterior probabilities. Given π, the posterior probability that a sample with genotypes g is of type $q = (q_0, q_1, q_2)$ is given by

$$p(q|g) = \frac{p(g|q)\pi(q)}{p(g)}$$

We used the six possible values for q described above, corresponding to pure forest, pure savanna, F1, F2, backcross to forest and backcross to savanna. Hence, the output for each sample is six posterior probabilities that sum to 1. Further, the overall probability that the sample is hybrid, p(hybrid | g), which we denote 'HP', is given by the sum of the posterior probabilities for the four hybrid types (F1, F2 and the two backcross types).

Application of hierarchical model. We applied the hierarchical model described above and computed posterior probabilities for three data sets defined in the section on STRUCTURE analyses: data set I (= full data set, Fig. 1B), data set II (= all samples from the area in Fig. 1C) and data set III (= all samples from the area in Fig. 1D). In each case, we first estimated f from reference samples, as described above, where the reference samples were obtained by applying STRUCTURE to the data set under consideration (see section about STRUCTURE analyses). Then, we applied the hierarchical model to estimate π, and finally computed the hybrid posterior probabilities. Assuming that the proportion of samples of each hybrid type (π) differ across Africa, restricting the hierarchical model analyses to local areas should lead to better estimates of allele frequencies, local estimates for π and estimates of the hybrid posterior probabilities.

Visualization of autosomal data

To visualize the autosomal data of each sample, we define the likelihood ratio (LR) for allele j in locus l having forest origin as

$$\mathrm{LR}_{jl\mathrm{forest}} = \frac{p(j|k=\mathrm{forest}, f, \delta)}{p(j|k=\mathrm{savanna}, f, \delta)}$$

where $p(j|k,f,\delta)$ is given by equation (1). Similarly, we define the LR for allele j having savanna origin as $\mathrm{LR}_{jl\mathrm{savanna}} = 1/\mathrm{LR}_{jl\mathrm{forest}}$. In Fig. 2, and in Figures S3 and S4 (Supporting information), we plotted the alleles at each locus, with the top allele at each locus in each sample being the one with the larger value of $\mathrm{LR}_{jl\mathrm{savanna}}$, and with each allele coloured to indicate its LR value.

Analyses of MtDNA and Y-chromosomal DNA

All sequences (both mtDNA and amelogenin) were analysed separately to estimate trees relating the

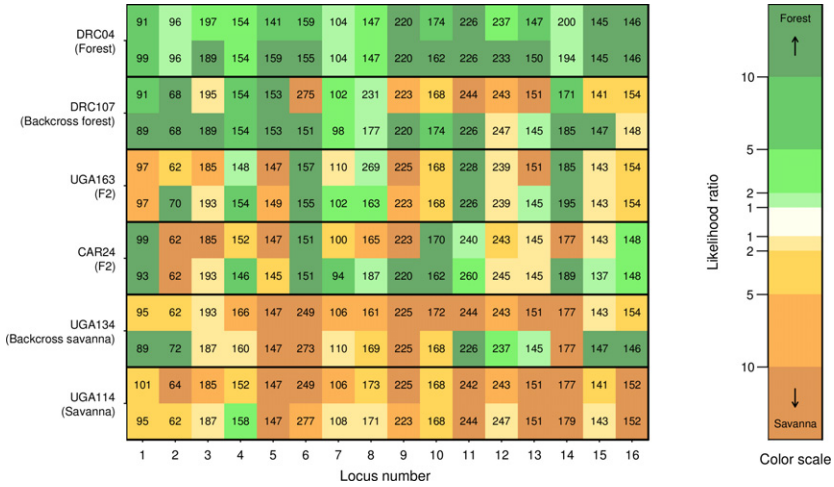

Fig. 2 Visualization of genotypes for six illustrative samples. For each sample, the 16 columns show genotypes (arranged in 2 rows) at 16 autosomal loci. Each allele is coloured according to its likelihood ratio (LR) for having forest (or savanna) origin; see methods for details. The inferred (most probable) hybrid category for each sample is indicated in parentheses. 'Backcross forest' = backcross between an F1 elephant and a forest elephant; 'Backcross savanna' = backcross between an 'F1' elephant and a savanna elephant. (Our model considered only hybrids up to second generation; it is possible that some or all of the hybrid samples involve hybridization beyond the second generation). Abbreviations same as Fig. 3.

samples using a Bayesian approach implemented in program MRBAYES v3.2 (Ronquist *et al.* 2012). Each of the gene fragments was analysed using jModelTest (Posada 2008) to determine its appropriate nucleotide substitution model. We used a two-parameter substitution model and a gamma distribution of evolutionary rates across sites, with the value of the shape parameter estimated from the data. The MCMC analysis used 2 runs of 4 chains each for a run length of 5×10^7 generations. No molecular clock was enforced in any analysis. So, tree and branch lengths depict topology rather than distance.

Results

Genetic analysis identifies a hybrid zone

We collected elephant dung samples in northern CAR, along Democratic Republic of Congo (DRC)–Uganda border and along the Pendjari–Arli complex on the Benin–Burkina Faso border, genotyped them at 16 autosomal microsatellite loci (Wasser *et al.* 2004), and combined these with data on the same loci from forest and savanna elephants across sub-Saharan Africa (Wasser *et al.* 2015). After quality control to filter out poorly amplifying samples and potential duplicates, our data set contained 291 samples from the targeted regions (209 from the DRC–Uganda border region, 8 from northern CAR and 74 from the Benin–Burkina Faso border region) and a total of 1369 unique elephant samples from 79 locations.

To obtain an initial overview of population structure, we used the program STRUCTURE (Pritchard *et al.* 2000) to perform a clustering analysis of all 1369 samples. This showed that the majority of samples were estimated to have group-specific ancestry (Fig. 1A), verifying the known division of these samples into forest and savanna subspecies. Nonetheless, numerous samples have substantial ancestry from both groups. For example, 54 samples were estimated to have >0.2 estimated ancestry in both groups, representing potential hybrids. Four of these potential hybrids are from Garamba National Park, which is where almost all the few previously reported hybrids are from. All but one of the remainder are from the three targeted regions: 43 from the DRC–Uganda border region, with most occurring on the Uganda side, 4 from northern CAR and 2 from the Benin–Burkina Faso border (Table 1). One sample from Namibia, with estimated forest ancestry 0.20, seems unlikely to be a true hybrid based on both its location and the more detailed analyses below.

To further assess the evidence for hybridization, we developed a new statistical method to explicitly quantify the support for each sample being hybrid (a 'hybrid probability', HP). As such inference can be sensitive to genotyping error due to allelic dropout or null alleles, our methods accounted for this (see Materials and Methods). Applying this method to the entire data set identified 55 samples with HP > 0.95 (Fig. 1B). These hybrid samples all originated from one of the three targeted regions, or from Garamba National Park (Table 1), and the results were highly concordant with

Table 1 Number of hybrids identified in three different analyses. For each of the three analysis approaches, the numbers of hybrids identified in each targeted region, as well as Garamba National Park, is listed. For the STRUCTURE analysis, the listed numbers are the numbers of samples with >0.2 estimated ancestries from both forest and savanna elephants. For the hybrid analyses, the listed numbers are the numbers of samples with HP > 0.95. NA indicates 'not applicable' and is used for the regions that were not included in our local analyses

Analysis	Garamba National Park	DRC–Uganda border	Northern CAR	Benin–Burkina Faso border	Other regions	In total
STRUCTURE	4	43	4	2	1	54
Global hybrid analysis	3	45	4	3	0	55
Local hybrid analysis	2	38	3	NA	NA	43

the STRUCTURE results (of the 54 putative hybrids identified by our STRUCTURE analysis, 51 had HP > 0.95).

Our hybrid analysis methods depend on specification of reference samples to estimate allele frequencies in the source forest and savanna populations. In the above 'global' analysis of the entire data set, we used reference samples drawn from across the entire continent. However, these global reference samples may not accurately capture the allele frequencies of the most relevant source populations: those that are geographically proximal to the putative hybrids being examined. To assess the robustness of our results to this issue, we therefore also conducted a 'local' analysis of each of the two targeted regions in Central Africa (along with Garamba National Park) using only geographically proximal samples as the references. We did not conduct a similar analysis in West Africa due to the scarcity of geographically proximal savanna reference samples. The results of these 'local' analyses were largely consistent with the global results (Table 1; Figs 1C and 1D), identifying 43 samples with HP > 0.95: 38 from the southern half of the DRC–Uganda border region, two from Garamba National Park in northeastern DRC (Fig. 1C) and three from northern CAR (Fig. 1D). One additional sample from northern CAR (sample ID CAR40) showed moderately high HP (HP = 0.74), and its hybrid status was subsequently supported by mtDNA and Y chromosome data. Further inspection of the three CAR samples with HP > 0.95 revealed two genetically very similar samples of same gender (differing from one another by only four alleles); to be conservative, only one of the two was included in subsequent analyses, and in the hybrid counts, we report in the rest of this paper. Accordingly, we also decreased the sample size for northern CAR by one, making it seven.

In summary, our three analyses (STRUCTURE, 'global' and 'local' analyses) produce largely consistent results (Table 1), and identify more hybrids than all previous genetic studies combined. All analyses indicate that the northern Albertine Rift along the southern half of the DRC–Uganda border region contains a large proportion of hybrid elephants and could be considered a hybrid zone. For example, the local analysis identifies 38 hybrids of 209 samples from the entire region along the DRC–Uganda border, and all of these hybrids are among the 102 samples we have from the southern half of this region. The previously reported hybrid zone from Garamba National Park in northeastern DRC was also reconfirmed (Roca *et al.* 2001; Comstock *et al.* 2002). The northern CAR region represents another potential hybrid zone with three hybrids of a total of seven elephants sampled from that region. Finally, we identified three hybrids of 74 individuals from West Africa in the Pendjari–Arli protected area complex on the border of Benin and Burkina Faso. As far as we are aware, these are the first reports of hybrids in West Africa based on genetic data.

Because the local analyses identified slightly fewer hybrids than the other analyses, we focus on the local analyses for the remainder of the paper to be conservative.

Hybrids are fertile

A key question is whether forest and savanna hybrids are fertile. If hybrids are infertile, then all hybrids will be first-generation (F1) hybrids, with one parent from each subspecies. Such hybrids will display a distinctive genetic signature, carrying exactly one allele from a savanna parent and one allele from a forest parent at each locus. Second-generation hybrids (F2 or backcross) would display different distinctive genetic signatures. We implemented a statistical hybrid classification method that exploits this to quantify the probability for each sample being either a 'pure' sample, an F1 hybrid or a second-generation hybrid (see Materials and Methods).

Applying this classification method, all hybrids identified in the 'local' analyses above (HP > 0.95) had probability >0.95 for being a second-generation hybrid.

Figure 2 shows examples of F2 and backcross hybrids, with a 'pure' sample of each subspecies for comparison (Figures S3 and S4, Supporting information). For example, sample UGA163 is clearly hybrid, with many alleles from each subspecies. Further, at least three of its loci (6, 11 and 14) are heterozygous for forest alleles only, indicating that UGA163 is not an F1. Importantly, this conclusion could not be explained by allelic dropout.

In summary, our data show that hybridization has occurred over multiple generations, and we conclude that hybrids are fertile.

The hybridization is bidirectional

Hybridization between a forest and savanna elephant could result from either a male forest elephant mating a savanna female, or a male savanna elephant mating a forest female. Hybridization is said to be 'bidirectional' if both types of matings occur. To assess whether hybridization is bidirectional, we sequenced Y chromosome and mtDNA from 36 hybrid samples, 13 pure forest samples and 18 pure savanna samples, and used MrBAYES (Ronquist *et al.* 2012) to estimate trees relating the samples (Fig. 3). These samples were selected based on our 'Hybrid probability' analyses, where each selected pure forest, savanna or hybrid samples showed posterior probability of >0.95. For both markers, the pure forest and pure savanna samples fall into different clades, but some hybrids join the forest clade and others join the savanna clade (Table 2). Thus, hybrid elephants contain, in aggregate, alleles characteristic of forest mtDNA, savanna mtDNA, forest Y chromosomes and savanna Y chromosomes, implying that hybridization has occurred bidirectionally.

Discussion

Using a large genetic data set, including samples from previously unsampled regions, our study provides new insights into African elephant hybridization. We found genetic evidence for hybridization between forest and savanna elephants in three distinct regions: the northern Albertine Rift along the DRC–Uganda border, northern CAR and the Pendjari–Arli complex along the Benin–Burkina Faso border. In the northern Albertine Rift, 38 of 102 samples showed strong evidence for being hybrids between forest and savanna elephants (HP > 0.95) establishing this as a major hybrid zone. In northern CAR, three of seven samples showed strong evidence for being hybrid, suggesting that this region is also a potential hybrid zone. In the Benin–Burkina Faso border region, 3 of 74 samples showed evidence for being hybrids, suggesting some level of genetic mixing in this region as well. Few, if any, of the hybrids we examined were first-generation hybrids, indicating that hybrids are fertile and that they have multigenerational histories of genetic mixing. Hybridization is also 'bidirectional' – that is involves males and females of both subspecies. This demonstrates that hybridization is not just a result of savanna males outcompeting the smaller forest males, as hypothesized by Roca *et al.* (2015). Finally, we have presented a new methodology to identify and classify hybrids, taking account of potential genotyping errors and allelic dropout, which can be applied to other species.

Our data represent by far the largest set of hybrids between forest and savanna elephants identified by genetic data to date. Previous studies (Roca *et al.* 2001; Comstock *et al.* 2002; Wasser *et al.* 2004; Wasser *et al.* 2007; Eggert *et al.* 2002) have identified only a few hybrids, mainly from the Garamba National Park and vicinity in the northeastern DRC (Roca *et al.* 2001; Comstock *et al.* 2002). No hybrids have previously been identified in the Uganda–DRC border region, in northern CAR or West Africa using genetic measures, although Groves & Grubb (2000) identified hybrids in the Uganda–DRC border region based on morphological metrics. Importantly, the hybrids we have identified appear in relatively geographically restricted locations. We have found no evidence of hybridization between forest and savanna elephants from any other sampled sites across the savanna–forest juncture. This raises the question: What is special about these places?

By far the largest number of hybrids occurred in the DRC–Uganda border in the Albertine Rift. While the northern CAR site had comparable proportion of hybrids, the sample size for this region is small. One question is therefore whether particular features of the DRC–Uganda region may have facilitated the establishment of a major hybrid zone. The ecotones in this region are a fine mosaic of forest and woodland–savanna habitat gradients with a very dynamic palaeohistory (Hamilton *et al.* 2001; McGlynn *et al.* 2013). This

Fig. 3 Trees relating samples based on mtDNA (A) and Y chromosome (B) data. Each sample is coloured according to its hybrid/pure savanna/pure forest probability estimated from autosomal loci. (CAR40 is designated as hybrid based on both its hybrid probability from autosomal data (HP = 0.74), and because its mtDNA clusters with forest mtDNA whereas its Y data clusters with savanna Y chromosomes.). CAR = Central Africa Republic, DRC = Democratic Republic of Congo, GR = Garamba, DRC, RWA = Akagera Rwanda, RWV = Virunga Rwanda, UGA = Uganda.

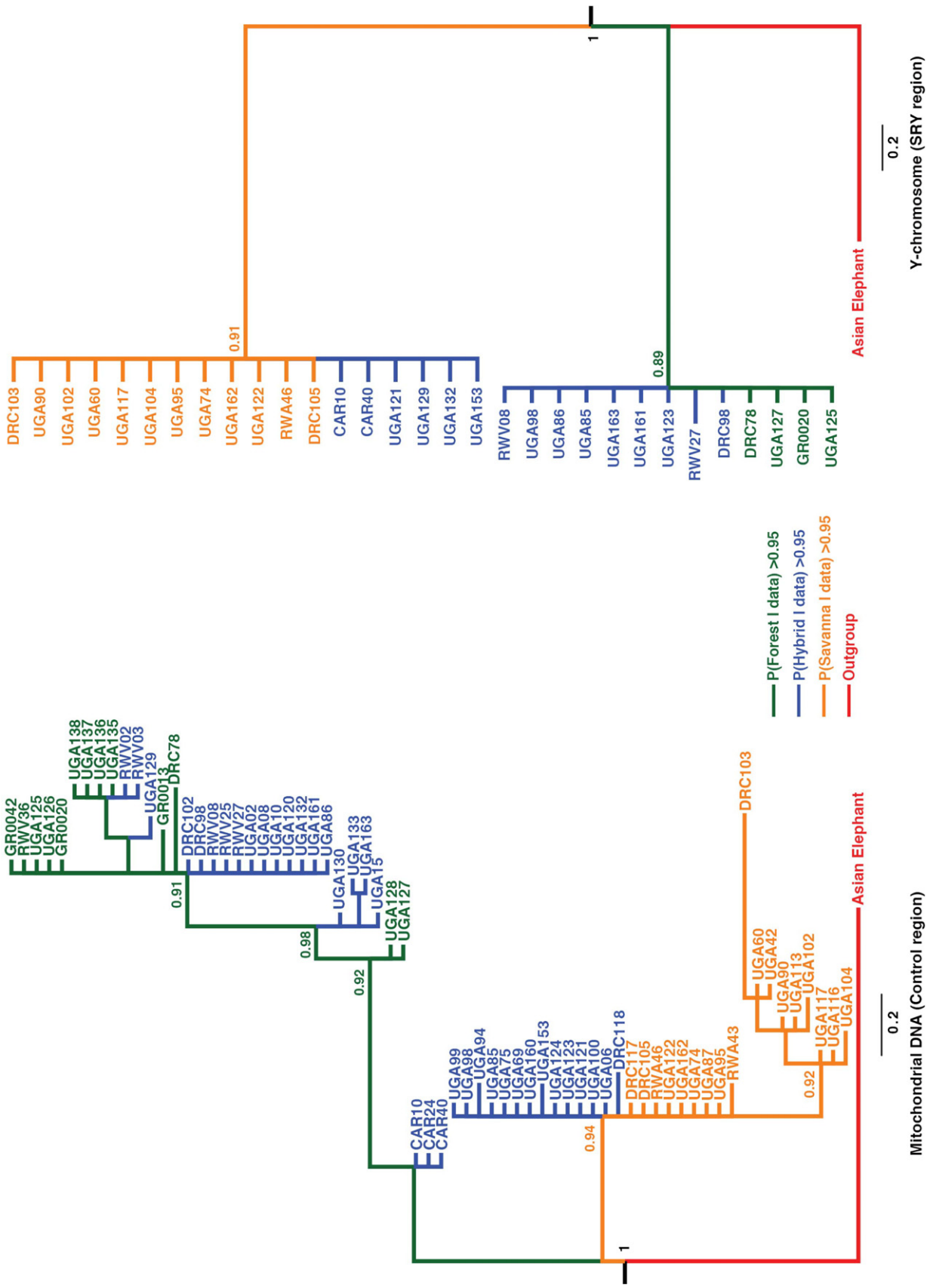

Table 2 Cluster membership counts for the all elephants present in Fig. 3. NA indicates 'not applicable' and is used for the female samples in which Y-chromosomal DNA was not available

	mtDNA cluster/YDNA cluster					
Species classification	Forest/Forest	Forest/Savanna	Savanna/Forest	Savanna/Savanna	Forest/NA	Savanna/NA
Hybrid	6	4	3	2	12	9
Pure savanna	0	0	0	12	0	6
Pure forest	4	0	0	0	9	0

contrasts with the more abrupt rainforest–savanna transition in the Pendjari–Arli site in West Africa, and the intermediate gradients in the Manovo-Gounda-St Floris complex in northern CAR (Fay 1988; East 1999; Barriere *et al.* 2005; IUCN 2015).

The DRC–Uganda region has also experienced extreme recent habitat modification coupled with cycles of elephant population loss and partial recovery over the past 50 years or more suggesting that hybridization may partly result from human pressures. Poaching was high on both sides of the border up until 1981, after which poaching greatly diminished on the Uganda side (Kayanja & Douglas-Hamilton 1983). Keigwin (2004; Keigwin *et al.* 2016) found a strong unidirectional bias in movements of elephants moving from DRC into Uganda between 2001 and 2003, but the reverse may well have occurred 40 years earlier when Ugandan parks were emptied of elephants during the Idi Amin era. Consistently, the elephant hybrids we identified were most common in woodland savanna habitat on the Uganda side of the DRC–Uganda border. The lack of F1 hybrids in our study remains a mystery, although this may be partly explained by the massive elephant population decline that occurred in eastern DRC by 2005. Eastern DRC ivory went from the most to least common ivory in forest elephant ivory seizures around that same time (Wasser *et al.* 2015). The relative decline in large-scale poaching on the DRC side around or prior to 2005 may also explain the absence of F1 hybrids in that area as the number of forest elephants moving into Uganda from DRC should have been greatly diminished by then.

Consistent with the idea that hybridization may, at least partly, be caused by human activity, we also found a high proportion of hybrids in northern CAR. This region borders Chad and South Sudan, where recent population declines from poaching have been equally extreme. (Bouche *et al.* 2010, 2011) By contrast, we found a relatively small proportion of hybrids in West Africa, where poaching-related population declines probably occurred much earlier, 1–2 centuries ago. Overall, it seems likely that ecological and human factors have worked together in favouring hybridization in all hybrid zones identified to date. Habitat gradients in eastern Central Africa may have allowed low levels of hybridization over time, while providing the right conditions for human–elephant interactions to precipitate a recent surge in hybridizations.

One way to assess the potential role of human activities would be to attempt to date the hybridization events. If hybridization is very recent, then it seems likely to have been driven, or at least exacerbated, by human influences. The high proportion of hybrids found in the southern DRC–Uganda border region, in combination with the genetically distinct forest and savanna elephants we identified in that same region, suggest that these hybridization events are likely recent on evolutionary timescales: long-term hybridization events would have made such strong separation between forest and savanna elephants difficult to retain. A more precise timeframe for hybridization could be obtained by analysing longer genomic fragments of forest vs. savanna origin in hybrid samples (Falush *et al.* 2003), using, for example, genomewide SNP data. Recent hybridization would be characterized by long consecutive sequences ('tracts') of forest and savanna nucleotides remaining intact in hybrid individuals, due to the small amount of recombination that occurs each generation. In contrast, older hybridization would be characterized by interspersing of shorter tracts of forest and savanna nucleotides. If hybridization has been both ancient and ongoing in the populations, considerable variation in tract length would be expected among individuals.

While our current knowledge of elephant hybrid behaviour is negligible, hybrids can have different behaviour than both parental species (Stratton & Uetz 1986; Helbig 1991; Hobel & Gerhardt 2003), and the introduction of species with new behaviours into complex interdependent ecosystems can have dramatic consequences (Rhymer & Simberloff 1996; Ryan *et al.* 2009; Peters & Kleindorfer 2015). This raises the question: Is recent occurrence of elephant hybrids a potential problem for ecology and human livelihoods? Impacts on ecology could be a particular concern in the Albertine Rift, which includes several World Heritage Sites characterized by high biodiversity and numerous endemic species (Plumptre *et al.* 2007; UNESCO 2013). Increased human–

elephant conflict is another potential concern, particularly if elephants from DRC have a weak affinity for the savanna habitat in Uganda, increasing their tendency to wander and encounter crops. Studies are needed to determine whether hybrid elephant behaviour in savanna habitats differs from that of savanna elephants.

Besides having potential implications for ecology and human livelihoods, some may argue that our results also have implications for the reclassification of the African Elephant species (*Loxodonta*) as two separate species (forest- *L. cyclotis* and savanna- *L. africana*) (Roca *et al.* 2001; Comstock *et al.* 2002; Bachhaus 1958; Frade 1955). But our identification of hybridization in several distinct sites does not contradict calls for reclassification, especially if this hybridization is recent. Although reproductive isolation has been historically seen as primary to the definition of species, hybrid zones between distinct species have been commonly reported across a wide range of taxa (Gray 1972; Rhymer & Simberloff 1996; Jolly *et al.* 1997). Our STRUCTURE analysis showed clear divisions of forest and savanna populations, consistent with previous genetic and morphological studies supporting reclassification into separate species (Frade 1955; Groves & Grubb 2000; Grubb *et al.* 2000; Roca *et al.* 2001, 2015; Comstock *et al.* 2002). However, ultimately, the question of reclassification must consider not only the population genetic structure, but also the potential impact of classification on conservation of both forest and savanna elephants (Allendorf *et al.* 2001; Fitzpatrick *et al.* 2015).

Acknowledgements

Patrick Agaba, Nick Geogiadis, Teresa Hart, Trip Jennings, Andrew Maser, Hoerst Oebel, Al Roca, US Embassy-Benin and many others contributed to sample collections. Permission to collect samples in Uganda and DRC was, respectively, provided by Uganda Wildlife Authority and Institut Congolais pour la Conservation de la Nature (ICCN) of Democratic Republic of Congo. Celia Mailand, Amy Torkelson and Rebecca Booth conducted the DNA analyses. Tim O'Brien and Cathy Laurie helped with initial analysis. Funding was provided by Wiatt Foundation of National Geographic Society, USFWS African Elephant Conservation Fund, Bosack and Kruger Charitable Foundation, Fulbright Doctoral and Professional Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Ahlering MA, Hailer F, Roberts MT, Foley C (2011) A simple and accurate method to sex savannah, forest and Asian elephants using noninvasive sampling techniques. *Molecular Ecology Resources*, **11**, 831–834.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613–622.

Backhaus D (1958) Zur Variabilitat der ausseren systematischen Merkmale des afrikanischen Elefanten (*Loxodonta Cuvier*, 1825). *Saugetierk. Mitt*, **6**, 166–173.

Barnes RFW (1999) Is there a future for elephants in West Africa? *Mammal Review*, **29**, 175–199.

Barriere P, Hutterer R, Nicolas V *et al.* (2005) Investigating the role of natural gallery forests outside the Congolese rainforest as a refuge for African forest shrews. *Belgian Journal of Zoology*, **135**, 21–29.

Bouche P, Renaud PÄ, Lejeune P *et al.* (2010) Has the final countdown to wildlife extinction in Northern Central African Republic begun? *African Journal of Ecology*, **48**, 994–1003.

Bouche P, Douglas-Hamilton I, Wittemyer G *et al.* (2011) Will elephants soon disappear from West African savannahs? *PLoS One*, **6**, e20619.

Center for Biological Diversity (2015) www.biologicaldiversity.org/news/press_releasess/2015/african-elephant-06-10-2015.html, last assessed October 8, 2015.

CITES (2014) Endangered species list. http://www.cites.org/eng/app/appendices.php, last accessed October 8, 2015.

Comstock KE, Wasser SK, Ostrander EA (2000) Polymorphic microsatellite DNA loci identified in the African elephant (*Loxodonta africana*). *Molecular Ecology*, **9**, 1004–1006.

Comstock KE, Georgiadis N, Pecon-Slattery J *et al.* (2002) Patterns of molecular genetic variation among African elephant populations. *Molecular Ecology*, **11**, 2489–2498.

Debruyne R (2005) A case study of apparent conflict between molecular phylogenies: the interrelationships of African elephants. *Cladistics*, **21**, 31–50.

East R (1999) African antelope database 1998. IUCN.

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

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Fay JM (1988) Forest monkey populations in the Central African Republic: the northern limits. A census in Manovo-Gounda-St Floris National park. *Mammalia*, **52**, 57–74.

Fitzpatrick BM, Maureen E, Johnson JR *et al.* (2015) Hybridization and the species problem in conservation. *Current Zoology*, **61**, 206–216.

Frade F (1955) Ordre des proboscidiens (*Proboscidea illiger*, 1811). *Traite de zoologie*, **17**, 715–875.

Gray AP (1972) Mammalian hybrids. A check-list with bibliography, 2nd Edition. Commonwealth Bureau of Animal Breeding and Genetics, Edinburgh. Technical Communication.

Groves CP, Grubb P (2000) Do *Loxodonta cyclotis* and *L. africana* interbreed. *Elephant*, **2**, 4–7.

Grubb P, Groves CP, Dudley JP, Shoshani J (2000) Living African elephants belong to two species: *Loxodonta africana* (Blumenbach, 1797) and *Loxodonta cyclotis* (Matschie, 1900). *Elephant*, **2**, 1–4.

Hamilton A, Taylor D, Howard P (2001) Hotspots in African forests as Quaternary refugia. *African Rain Forest Ecology and Conservation: An Interdisciplinary Perspective*, pp. 57–67. Yale University Press, New Haven, Connecticut.

Hammer J (2014) The race to stop Africa's elephant poachers. Smithsonian Magazine, July. (http://www.smithsonianmag.com/science-nature/race-stop-africas-elephant-poachers-180951853/?no-ist)

Helbig AJ (1991) Inheritance of migratory direction in a bird species: a cross-breeding experiment with SE-and SW-migrating blackcaps (*Sylvia atricapilla*). *Behavioral Ecology and Sociobiology*, **28**, 9–12.

Hobel G, Gerhardt HC (2003) Reproductive character displacement in the acoustic communication system of green tree frogs (*Hyla cinerea*). *Evolution*, **57**, 894–904.

IUCN (2015) IUCN Red List website. www.iucnredlist.org/details/12392/0, last accessed October 8, 2015.

Johnson M, Clifford S, Goossens B *et al.* (2007) Complex phylogeographic history of central African forest elephants and its implications for taxonomy. *BMC Evolutionary Biology*, **7**, 244.

Jolly CJ, Woolley-Barker T, Beyene S *et al.* (1997) Intergeneric hybrid baboons. *International Journal of Primatology*, **18**, 597–627.

Kayanja FIB, Douglas-Hamilton I (1983) Impact of the unexpected: a case history of the Uganda National Parks. *Swara*, **6**, 8–14.

Keigwin MD (2001) Elephants, Crops and People project in Ishasha Sector Southern Queen Elizabeth National Park, Uganda. *Pachyderm*, **31**, 73–78.

Keigwin MD (2004) Threat to the recovery of the Ishasha elephant recovery in southern Queen Elizabeth Conservation Area. Proceedings of the International Elephant Foundation 148–159: In International Elephant Research Symposium Proceedings. Fort Worth, Texas, Fort Worth Zoo. Pp. 1–163.

Keigwin M, Wabukawo V, Wasser SK (2016) Impacts on transboundary elephant movements between Queen Elizabeth National Park, Uganda and the Park National des Virunga, Democratic Republic of Congo and associated crop raiding. *Pachyderm*, (in press).

McGlynn G, Mooney S, Taylor D (2013) Palaeoecological evidence for Holocene environmental change from the Virunga volcanoes in the Albertine Rift, central Africa. *Quaternary Science Reviews*, **61**, 32–46.

Mondol S, Mailand CR, Wasser SK (2014) Male biased sex ratio of poached elephants is negatively related to poaching intensity over time. *Conservation Genetics*, **15**, 1259–1263.

Mubalama L, Mushenzi N (2004) Monitoring law enforcement and illegal activities in the northern sector of the Parc National des Virunga, Democratic Republic of Congo. *Pachyderm*, **36**, 16–29.

Nyakaana S, Arctander P (1998) Isolation and characterization of microsatellite loci in the African elephant, Loxodonta africana. *Molecular Ecology*, **7**, 1436–1437.

Peters KJ, Kleindorfer S (2015) Divergent foraging behavior in a hybrid zone: Darwin's tree finches (*Camarhynchus spp.*) on Floreana Island. *Current Zoology*, **61**, 181–190.

Plumptre AJ, Davenport TRB, Behangana M *et al.* (2007) The biodiversity of the Albertine Rift. *Biological Conservation*, **134**, 178–194.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review for Ecology, Evolution and Systematics*, **27**, 83–109.

Roca AL, Georgiadis N, Pecon-Slattery J, O'Brien SJ (2001) Genetic evidence for two species of elephant in Africa. *Science*, **293**, 1473–1477.

Roca AL, Georgiadis N, O'Brien SJ (2005) Cytonuclear genomic dissociation in African elephant species. *Nature Genetics*, **37**, 96–100.

Roca AL, Ishida Y, Brandt AL *et al.* (2015) Elephant natural history: a genomic perspective. *Annual Review of Animal Biosciences*, **3**, 139–167.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Ryan ME, Johnson JR, Fitzpatrick BM (2009) Invasive hybrid tiger salamander genotypes impact native amphibians. *Proceedings of the National Academy of Sciences of the USA*, **106**, 11166–11171.

Stratton GE, Uetz GW (1986) The inheritance of courtship behavior in Schizocosa wolf spiders (*Araneae lycosidae*). *Evolution*, **40**, 129–141.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

The Worldwatch Institute (2014) website (http://www.worldwatch.org/node/4525)

Uganda historical demography data (2014) (http://www.populstat.info/Africa/ugandac.htm).

UNEP, CITES, IUCN TRAFFIC (2013) Elephants in the dust: the African elephant crisis. A rapid response assessment. United Nations Environment Programme, GRID-Arendal. (http://www.grida.no/publications/rr/elephants/)

UNESCO (2013) World Heritage Site List. http://whc.unesco.org/en/list.

Wasser SK, Shedlock AM, Comstock K *et al.* (2004) Assigning African elephant DNA to geographic region of origin: applications to the ivory trade. *Proceedings of the National Academy of Sciences of the USA*, **101**, 14847–14852.

Wasser SK, Mailand C, Booth R, *et al.* (2007) Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban. *Proceedings of the National Academy of Sciences of the USA*, **102**, 4228–4233.

Wasser SK, Brown L, Mailand C *et al.* (2015) Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. *Science*, **349**, 84–87.

The idea for this paper was conceived by S.K.W. S.K.W., M.K. and J.H. coordinated sample collections. Data were generated by S.K.W., S.M. and L.B. The statistical methods for estimating hybrid probabilities were developed by I.M. and M.S. Statistical analyses were conducted by I.M., S.M. and L.B. The manuscript was written by S.K.W., S.M., I.M. and M.S. All co-authors contributed final edits to the ms.

Data accessibility

All genotype data, sample locations, hybrid probabilities, structure ancestries and sequence alignment files (of 526-bp mitochondrial and 719-bp Y chromosome partial control region sequences of both forest, savannah and hybrid elephants) have been accessioned on DRYAD (doi:10.5061/dryad.d4 m07); hybrid probability computation codes are posted on GitHub (https://github.com/stephenslab/EBhybrids); all sequence data were submitted to GenBank (Accession no.: KF534810–KF534908).

Supporting information

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

Fig. S1 Marker and subspecies specific maximum likelihood (ML) estimates of the allelic dropout rates, γ. The dotted lines show the intervals of parameter values with likelihoods <1.92 log-likelihood units (ll) from the log-likelihoods of the ML estimates.

Fig. S2 Marker and subspecies species-specific maximum likelihood (ML) estimates of the allelic dropout rates, γ compared to marker and subspecies specific estimates of the inbreeding coefficient F.

Fig. S3 Genetic data from the samples shown in Fig. 1(C). For each sample the 16 genotyped autosomal alleles are arranged in two rows and 16 columns (one column for each marker) and each allele is coloured according its likelihood ratio (LR) for having forest (or savannah) origin, see methods for details.

Fig. S4 Genetic data from the samples shown in Fig. 1(D). Each sample is depicted by its autosomal alleles arranged in two rows and 16 columns (one column for each marker) and each allele is coloured according its likelihood ratio (LR) for having forest (or savannah) origin, see methods for details.