Origin and phylogeography of African savannah elephants (*Loxodonta africana*) in Kruger and nearby parks in southern Africa

Alida de Flamingh[1,2], Alfred L. Roca[2,3,4], Rudi J. van Aarde[1]

Corresponding author: Alida de Flamingh, adeflamingh@gmail.com, (+1) 217 333 4679

[1] Conservation Ecology Research Unit, Department of Zoology and Entomology, University of Pretoria, Pretoria 0002, South Africa
[2] Program in Ecology, Evolution and Conservation Biology, School of Integrative Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[3] Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[4] Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Conflict of Interest: The authors declare that they have no conflict of interest.

Abstract:

African savannah elephants (*Loxodonta africana*) occur in fragmented and isolated populations across southern Africa. Transfrontier conservation efforts aim at preventing the negative effects of population fragmentation by maintaining and restoring linkages between protected areas. We sought to identify genetic linkages by comparing the elephants in Kruger National Park (South Africa) to populations in nearby countries (Botswana, Mozambique, Zambia and Zimbabwe). We used a 446 base pair mitochondrial DNA (mtDNA) control region fragment (141 individuals) and 9 nuclear DNA (nDNA) microsatellite markers (69 individuals) to investigate phylogenetic relationships and gene flow among elephant populations. The mtDNA and nDNA phylogeographic patterns were incongruent, with mtDNA patterns likely reflecting the effects of ancient female migrations, with patterns persisting due to female philopatry, and nDNA patterns likely reflecting male-mediated dispersal. Kruger elephant heterozygosity and differentiation were examined, and were not consistent with genetic isolation, a depleted gene pool or a strong founder effect. Mitochondrial DNA geographic patterns suggested that the Kruger population was founded by elephants from areas both north and south of Kruger, or has been augmented through migration from more than one geographic source. We discuss our findings in light of the need for conservation initiatives that aim at maintaining or restoring connectivity among populations. Such initiatives may provide a sustainable, self-regulating management approach for elephants in southern Africa while maintaining genetic diversity within and gene flow between Kruger Park and nearby regions.

Keywords: African elephant, *Loxodonta africana*, phylogeography, population structure, gene flow, Kruger National Park

Introduction:

Landscape fragmentation and habitat loss are globally important drivers of population declines among species (Fischer and Lindenmayer, 2007). Protected areas may serve as species strongholds, but they often consist of discontinuous patches that represent fragments of the once continuous historic ranges of many mega-herbivores and carnivores (Ripple et al., 2015). The fragmentation and resulting isolation of populations may have deleterious demographic and genetic consequences, because restricted gene flow between populations may lead to genetic drift, inbreeding, the expression of deleterious alleles, and reduced fitness; it can also increase the risk of local extirpation (Allendorf et al., 2013). For example, Fitzpatrick and Evans (2009) showed that reduced heterozygosity impaired sperm quality in endangered mammals, and Miller et al. (2011) showed that island tammar wallaby populations with low genetic diversity and high levels of inbreeding had an increased frequency of morphological abnormalities.

Recent conservation initiatives in Africa have sought to counter the effects of population fragmentation and isolation. These initiatives rely on the restoration and maintenance of distributional ranges to link formerly isolated populations through the creation of “megaparks” (van Aarde and Jackson, 2007), and transfrontier conservation areas (Hanks, 2003). Such initiatives can be considered functional if they disrupt population isolation by allowing dispersal or migration between nearby populations. Linking isolated populations may have beneficial genetic consequences if isolation has been a driving force of genetic degradation associated with bottlenecks and the subsequent loss of genetic diversity due to drift.

Information on the genetic structure and characteristics of elephant populations may help establish whether isolation and founder effects have resulted in genetically depleted, homogenous populations, and may provide support to the proposed benefits of enabling dispersal between isolated populations. African savannah elephants are no longer distributed across a continuous geographic

range, but instead have contracted ranges that largely overlap with protected areas (Ripple et al., 2015). Elephants in Kruger National Park in South Africa are confined by fencing, while in parts of Angola, Botswana, Namibia, Mozambique and Zimbabwe elephant movements are restricted by the abundance of people or by resource limitations (Roever et al., 2013). Gene flow between Kruger and nearby parks has not been quantified, but is thought to be limited due to park fences and other barriers (e.g., areas of high human density) that obstruct dispersal between populations. Furthermore, the Kruger elephant population is believed to have derived from a single source population. Elephants were widely distributed across southern Africa, including the present-day location of Kruger, and their movements were largely unrestricted prior to European colonization (Whyte, 2001a). After colonization hunting decimated the population in Kruger to the point of extirpation or near extirpation (Hall-Martin, 1992; Pienaar, 1963; Whyte, 2001a). It is believed that the population was re-established by elephants that migrated from Mozambique into Kruger at an entry point approximately midway between the Park's northern and southern boundaries after the proclamation of the Sabie Game Reserve in 1898 and the arrival of the first warden in 1903. Elephants are believed to have migrated from this entry point into the northern and southern regions of the park (Whyte, 2001). The influx of elephants from Mozambique increased the population in Kruger to about 25 individuals in 1912 and about 100 individuals in 1926. By 1963 the population was estimated to include 1750 individuals and by the end of the 20th century the number of elephants in Kruger was fast approaching the 10 000 mark (van Aarde et al., 1999; Young et al., 2009).

Our present study examined the genetic diversity and connectivity of elephants in South Africa's Kruger National Park, comparing them to elephants in nearby parks in Botswana, Mozambique, Zambia and Zimbabwe. We amplified a 446 base pair (bp) mitochondrial DNA (mtDNA) control region fragment from 141 elephants, and 9 nuclear DNA (nDNA) microsatellite loci from 69 elephants.

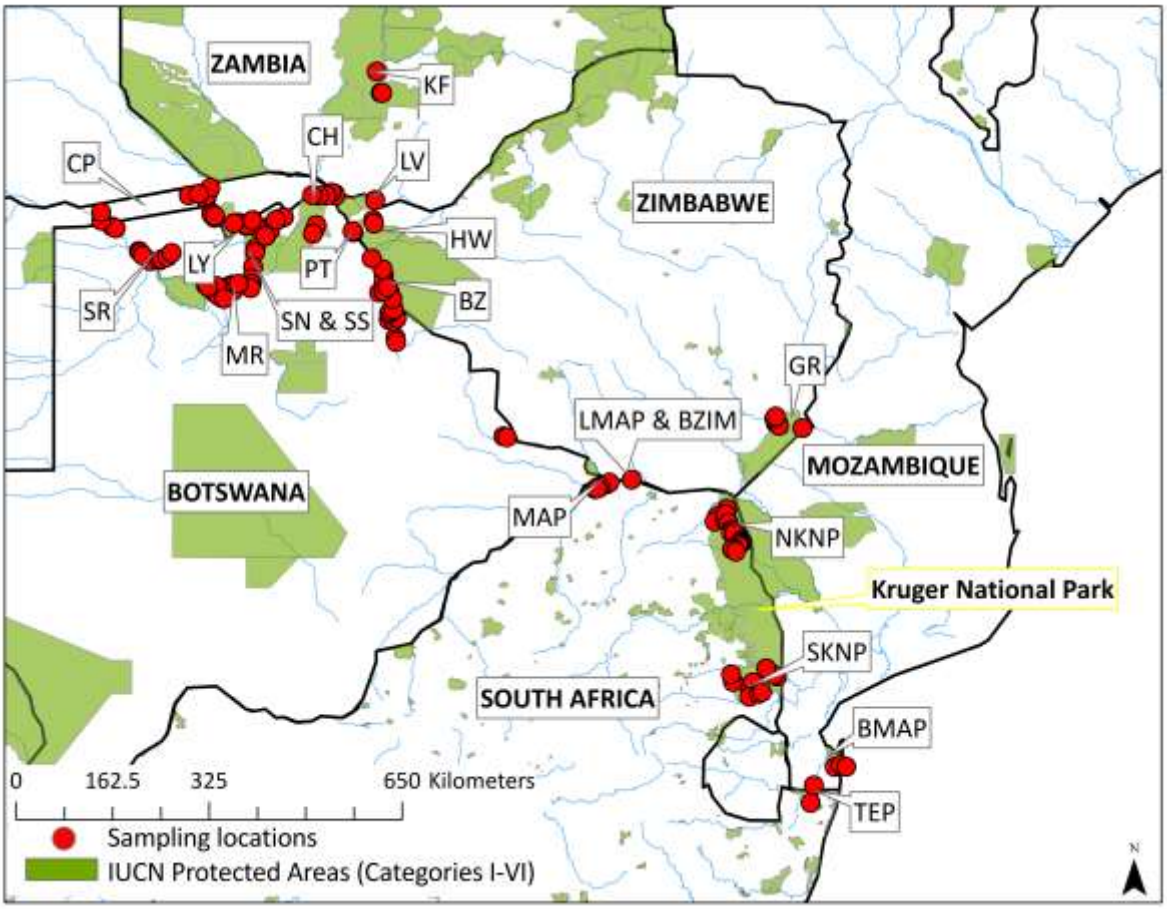

Fig. 1 Study sites and locations where savannah elephant fecal samples were collected (circles, n=141). Study site abbreviations include SKNP and NKNP = southern and northern Kruger National Park, respectively, South Africa; MAP, LMAP and BZIM = Mapungubwe National Park (MAP) and along the Limpopo River that connects Mapungubwe and Kruger National Park (LMAP and BZIM), South Africa; TEP and BMAP = Tembe Elephant Park, South Africa and Maputo Elephant Reserve, Mozambique; KF = collected from Kafue National Park, Zambia; LV = Livingstone Town, Zambia; PT = Pandamatenga, Botswana; MR = Moremi Game Reserve, Botswana; SS and SN = south and north Savuti Game reserve, respectively, Botswana; CH = Chobe National Park, Botswana; SR = Seronga, Botswana; LY = Linyanti River, Botswana; BZ = the border of Botswana and Zimbabwe; HW = Hwange National Park, Zimbabwe; GR = Gonarezhou National Park, Zimbabwe; and CP = Caprivi region, Namibia.

The elephant dung samples were collected across various national parks and conservation areas in southern Africa (Figure 1). Translocations were used to limit the increasing elephant population in Kruger from 1970 to 1994 (Scholes and Mennell, 2008 (Ch5); Whyte, 2001a), and translocations from Kruger to other areas continue to the present day. However, translocations have mainly been to areas

(parks, conservation areas and privately owned land) in South Africa that did not form part of this study (Dublin and Niskanen, 2003; Scholes and Mennell, 2008 (Ch5)).

We expected to detect in Kruger the negative genetic consequences that may follow a founder effect and population fragmentation and isolation. Compared with other southern African elephant populations, we expected Kruger's elephant population to show reduced genetic variability, and to find evidence of limited gene flow between Kruger and nearby parks, and the signatures of a single founding population.

Materials and Methods:

Sample collection

The collection of samples of elephant dung was sanctioned by appropriate authorities prior to collection (see Supplementary Table 8 for a list of permits). Sampling locality abbreviations include SKNP and NKNP = southern and northern Kruger National Park, respectively, South Africa; MAP, LMAP and BZIM = Mapungubwe National Park (MAP) and along the Limpopo River that connects Mapungubwe and Kruger National Park (LMAP and BZIM), South Africa; TEP and BMAP = Tembe Elephant Park, South Africa and Maputo Elephant Reserve, Mozambique; KF = collected from Kafue National Park, Zambia; LV = Livingstone Town, Zambia; PT = Pandamatenga, Botswana; MR = Moremi Game Reserve, Botswana; SS and SN = south and north Savuti Game Reserve, respectively, Botswana; CH = Chobe National Park, Botswana; SR = Seronga, Botswana; LY = Linyanti River, Botswana; BZ = the border of Botswana and Zimbabwe; HW = Hwange National Park, Zimbabwe; GR = Gonarezhou National Park, Zimbabwe; and CP = Caprivi region, Namibia.

For a comparison of nDNA heterozygosity by geographic region, the samples were grouped geographically into 7 regions including Chobe Region (CR) = CH, MR, CP, SS, SN, LY; Hwange Region (HR)

= HW, PT; Livingstone Region (LR) = LV; Botswana-Zimbabwe border Region (BZR) = BZ; Gonarezhou National Park (GNP) = GR; Mapungubwe Region (MR) = MAP, LMAP, BZIM; Kruger Region (KR) = NKNP, SKNP. Regional heterozygosity was not calculated for samples from TEP, BMAP, KF, SR and LV since samples from those areas were not genotyped for nDNA. Samples were grouped to represent contiguous geographic areas: samples that were from contiguous national parks or reserves, e.g. Chobe, Savuti, Linyanti and Moremi were grouped together and assigned the name of the largest national park; samples that were in close proximity to a national park but fell outside its borders were grouped as part of that park (e.g. the grouping of LMAP and BZIM as part of MAP); and samples that were collected along the Botswana-Zimbabwe border were grouped together.

Sample grouping

Samples were grouped and analyzed as follows: 1) groups determined by Bayesian clustering analysis in Geneland (Guillot et al., 2005), 2) seven regions based on geography and connectivity for nDNA heterozygosity comparisons, namely CR, HR, LR, BZ, GR, MR, KR and 3) four regions (all samples north of Kruger, samples in northern Kruger, samples in southern Kruger, and all samples south of Kruger) to compare mtDNA haplotype distribution analysis between north Kruger, south Kruger and nearby parks.

Data generation

Dung sample collection, DNA extraction and DNA amplification procedures for mtDNA sequences and nDNA microsatellites are described by de Flamingh et al. (2015). However, microsatellite genotypes determined for this study were replicated 4 times for homozygotes and 3 times for heterozygotes. Mitochondrial DNA sequence data (446 base pairs) for 141 individuals and nDNA data for 69 individuals (Fig. 1, Supplementary Table 1) were generated at the Conservation Ecology Research Unit (CERU), University of Pretoria.

The difference in sample size for mtDNA and nDNA data can be attributed to the use of degraded fecal DNA - see de Flamingh et al. (2015) for a description of sample collection procedures. Mitochondrial DNA is often better preserved than nDNA in degraded DNA samples (Schwarz et al., 2009). Therefore, DNA extraction yield for mtDNA is often much higher than nDNA, leading to a higher amplification success of mtDNA markers compared to nDNA markers. This is especially apparent in degraded DNA samples (Andreasson et al., 2002). In addition, microsatellite genotype errors are often associated with degraded DNA amplification, e.g. null allele or false allele amplification (de Flamingh et al., 2014; Taberlet et al., 1999). Combined these factors resulted in fewer individuals being genotyped for nDNA than mtDNA. However, the nDNA dataset sample size (69 individuals) was larger than sample sizes used by other studies to represent the same geographic area; e.g., Comstock et al. (2002) used 19 individuals, Eggert et al. (2002) used 54 individuals, and Roca et al. (2001) used 40 samples to represent the same geographic region.

Outgroup taxa for mtDNA phylogenetic analyses included sequences from the Asian elephant (*Elephas maximus*) (Maikaeu et al., 2007; Rogaev et al., 2006) and the woolly mammoth (*Mammuthus primigenius*) (Enk et al., 2011; Gilbert et al., 2008; Krause et al., 2006). GenBank accession numbers for outgroup taxa are listed in Supplementary Table 2.

Phylogenetic analysis

Sequence data were aligned using the program MUSCLE (Edgar, 2004). We visually inspected the aligned sequences, excluded all oligonucleotide primer sequences, and verified each polymorphic site by viewing chromatograms in the program CLC Bio Genomics workbench (CLC Bio, Cambridge, MA).

The substitution model that best describes the variation in our data was identified using the program jModelTest2 (Darriba et al., 2012). Based on the lowest likelihood scores for models that fall within the 95% confidence interval for the Akaike and Bayesian Information Criterion, we used the HKY+I+G substitution model because it was consistently ranked highly by both criteria. Genescan

fragments for nDNA microsatellite loci were viewed and scored using GeneMapper ® Software Version 3.7 (Applied Biosystems, Foster City, USA). Scored data were then exported in table format to Excel spreadsheets for further analysis.

Distinct haplotypes present among the 147 individuals (141 African elephants and 6 individuals from outgroup taxa) were identified using DnaSP 5.10.01 (Librado and Rozas, 2009), and used to construct a median-joining network using the software Network 4.6.1.0 (Fluxux Technology Limited). Haplotypes were compared to previously published control region sequences and assigned to one of eight mtDNA subclades using the methods described by Ishida et al., 2013. Haplotypes assigned to the same subclade were verified as grouping together in the mtDNA median-joining network. Haplotypes 1 and 15, which were separated from other haplotypes by long branches, were queried using BlastN to find matches to the sequences of Ishida et al. (2013).

To establish the origins of Kruger elephants, mitochondrial DNA haplotype frequencies were compared across elephants in parks north of Kruger, northern Kruger, southern Kruger, and parks south of Kruger. For this comparison the haplotypes were assigned to four regions: all parks north of Kruger, northern Kruger, southern Kruger and all parks south of Kruger. The proportion of elephants in each region that carried each distinct haplotype was visualized by a stacked bar graph produced in Excel 15.12.3 (©Microsoft) that showed the most common haplotypes at the top of each bar, with rare haplotypes at the base of the bar.

Bayesian analyses were performed using the program MrBayes 3.2.6 (Ronquist et al., 2012). We performed three independent runs. The first run included 1 million generations with a sampling frequency of 10 000, the second run included 1 million generations a sampling frequency of 100, and our third run included 1 000 generation sampling all possible trees. The latter two runs were used to investigate convergence and stationarity of log likelihood values. For both runs the parameters converged within the first ca. 1 000 generations, confirming that convergence and stationarity had been

reached within the initial run. Subsequent Bayesian analyses included a generation time of 1 million and a sample frequency of 10 000. Runs incorporated four chains and the HKY+I+G substitution model. Run length sufficiency was determined using the program Tracer 1.6.0 (Rambaut et al., 2014) by confirming adequate chain mixture and convergence and stationarity upon a likelihood score. The ESS value indicated that a burn-in of 10% was sufficient (convergence was reached quickly), and we therefore discarded 10% of the initial trees. As an alternative visualization of topology support, Bayesian output trees were plotted as a cloudogram using the program DensiTree 2.2 (Bouckaert 2010).

Phylogeographic analysis

We used Geneland (Guillot et al., 2005) to determine how many sub-populations (K) of genetically similar individuals were present in our study, for both the mtDNA and nDNA datasets. We tested the possibility of K = 1 through K = 10, using 500 000 iterations with a thinning factor of 100. We employed the correlated allele frequency and spatial models to map probabilities of population membership across the landscape, using a spatial uncertainty parameter of 0.01 that was calculated for the same data by de Flamingh et al., (2015). We verified Geneland clustering results with the program BAPS (Corander et al., 2003) using the "spatial clustering of individuals" algorithm for mtDNA sequence data (Kmax 15) and nDNA microsatellites (Kmax range of 5, 10 and 15 with each value replicated 5 times).

Geneland uses geo-referenced multi-locus data, we therefore converted the sequence data to haploid data based on binary code, where all the polymorphic sites (n=47) were considered as individual loci, with the possibility of 4 alleles (ACTG) at each of the loci (Guillot et al. 2009; Wheeler et al., 2010). We used an analysis of molecular variance (AMOVA) to calculate inter-population genetic variation (F_{st} – fixation index) for both nDNA and mtDNA. The AMOVA and exact test of population differentiation for mtDNA were implemented in Arlequin (Excoffier and Lischer, 2010a), and the AMOVA for nDNA was implemented using GenAlEx (Peakall and Smouse, 2006). Observed nDNA heterozygosity was calculated

for the seven regions (CNP, HNP, LNP, BZB, GNP, MNP, and KR) defined above using the program GenAlEx (Peakall and Smouse, 2006). We used an AMOVA in GenAlEx (Peakall and Smouse, 2006) and exact tests of population differentiation in Arlequin (Excoffier and Lischer, 2010a) to determine the degree of genetic differentiation between these seven regions. We used a Kruskal-Wallis test to determine if there was a significant difference in heterozygosity across geographic groups (McDonald, 2009). We compared the overall population (all populations combined) with only individuals found in Kruger. We calculated the number of alleles and the expected heterozygosity for each of these groupings using the program Arlequin (Excoffier and Lischer, 2010b), and included an exact test of population differentiation to determine whether the overall population differs significantly from the Kruger population. We used a Mann-Whitney U test to determine whether heterozygosity was significantly lower in Kruger compared to the overall population.

We used the program Bottleneck v. 1.2.02 (Cornuet and Luikart, 1996) to investigate recent effective population size reductions from allele data frequencies for the nine nDNA loci. For populations that underwent a recent bottleneck, this program finds a signature of excess of heterozygosity relative to expected heterozygosity (based on the observed number of alleles) since during a bottleneck, a population loses allelic diversity more rapidly than it loses heterozygosity (Cornuet and Luikart, 1996). We determined the probability of a recent bottleneck using the sign test and Wilcoxon test assuming a stepwise mutation model (SMM) and a two phase model (TPM) (proportion of SMM in TPM = 0.00, variance of the geometric distribution for TPM = 0.36) for 100,000 iterations across all 9 loci. Using the same program, we determined whether there was a deficit of rarer alleles and a modal shift in allele proportions in different allele frequency classes since these may also be indicators of a recent bottleneck (Luikart et al., 1998).

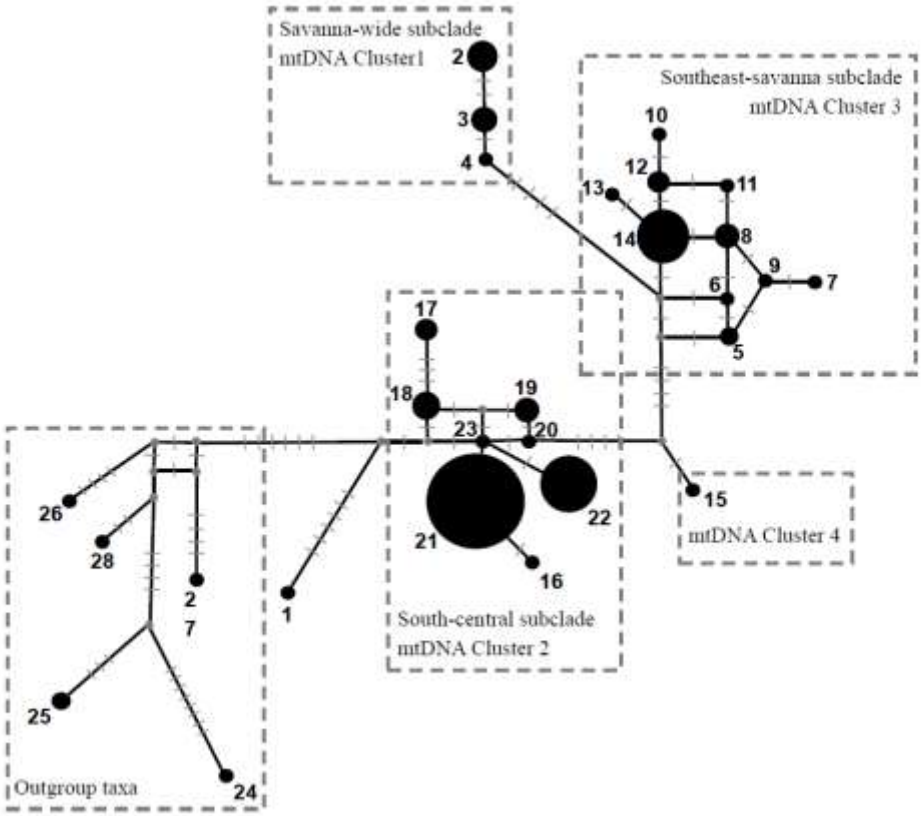

Fig. 2 A median-joining haplotype network (based on pairwise distances) for all mtDNA sequences show separation into groups that correspond to the mtDNA clusters identified using Geneland. Outgroup sequences, groups corresponding to Geneland mtDNA Clusters 1 through 4, and previously identified mtDNA sublades (Ishida et al., 2013) to which haplotypes in this study were assigned are indicated in the dashed rectangles. Haplotypes H1 and H15, each found in Kafue in a single individual each, were distinctive from the other sequences and not categorized to a subclade. Branch length is proportional to the number of mutational differences (indicated as cross-hatches) between haplotypes, circle size is proportionate to the number of individuals carrying a haplotype (see Supplementary Figure 3 for an inset of the sub-clades reported by Ishida et al. 2013).

Results:

Mitochondrial DNA analyses

Mitochondrial DNA sequences were generated for 141 elephants. The alignment consisted of 446 nucleotide sites of which 47 were polymorphic. There were 28 distinct mtDNA haplotypes, with the individuals that carried each haplotype listed in Supplementary Table 2 (GenBank accession numbers are also listed in Supplementary Table 2). African elephant haplotypes clustered into four groups in a

median-joining network (Fig. 2), exclusive of haplotypes carried by the outgroup species (Haplotypes 24-28).

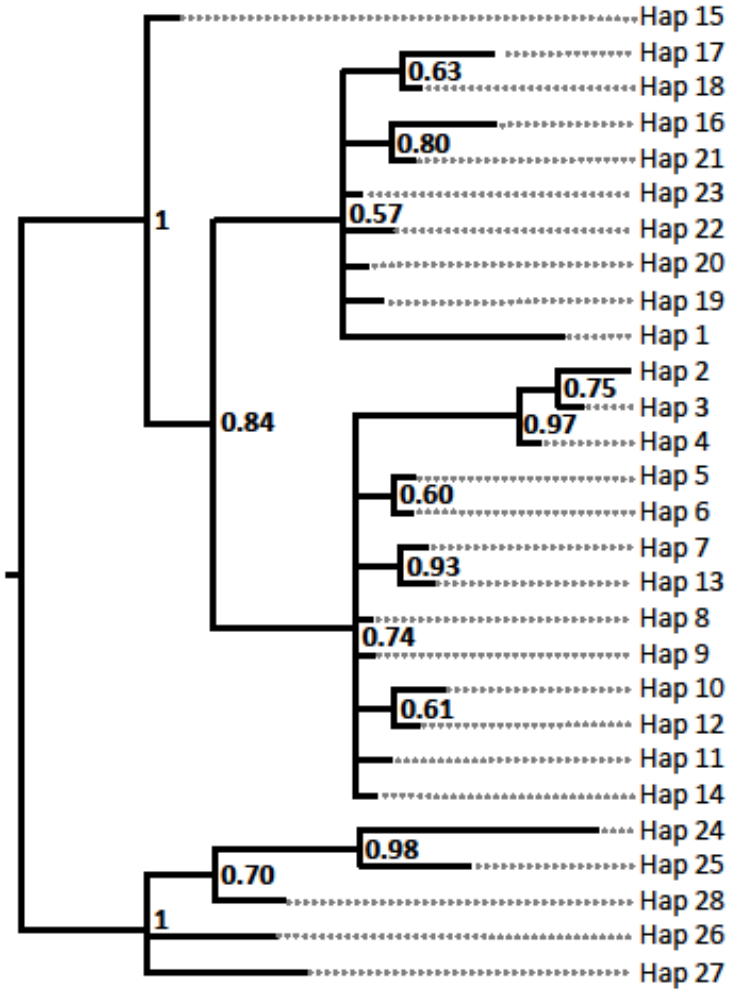

Fig. 3 Bayesian analysis showed limited resolution of relationships between mtDNA haplotypes, where polytomies were present at several nodes. There was high support for the two distinct mtDNA clades. This topology represents a 50% majority-rule consensus tree, with node support values (posterior probabilities) shown next to their corresponding nodes.

The phylogeny inferred using Bayesian analysis showed limited resolution of the relationships among haplotypes with several polytomies present (Fig. 3). There was high support for the recovery of two distinct savannah elephant mtDNA clades, with Haplotype 15 outside of the two clades. The

Bayesian analysis grouped Haplotypes 2, 3 and 4 into a separate clade, while the relationships among other haplotypes were not well resolved.

An alternative visualization of the Bayesian analysis through a cloudogram plot illustrated the uncertainty in terminal taxon relationships (Supplementary Fig. 1). The cloudogram of all trees (indicated in green) and consensus trees (indicated in blue) did show Haplotype 15 as distinct from other haplotypes, but did not group the remaining savannah elephant mtDNA haplotypes into two well supported distinct clades. Other clades of high support in Bayesian analysis (e.g., Haplotypes 2, 3 and 4) are illustrated by the darker blue bands in the cloudogram.

Almost all haplotypes in the study area were identified as belonging to three of the previously identified continental mtDNA subclades described by Ishida et al., (2013) (Savannah-wide, Southeast-savannah and South-central subclades; Fig. 2; Supplementary Fig. 2). However, haplotypes 1 and 15, each found in Kafue and each limited to a single individual sampled in that park (Fig. 4), were distant from the other elephant haplotypes in the network. Haplotype 15 in particular was phylogenetically distinct in the Bayesian and cloudogram phylogenies. The most similar sequences in GenBank to H1 belonged to the North-central subclade, while H15 was closest to haplotypes within the South-Central subclade, although we did not assign these to subclades in Figure 2 given the available sequence information and lack of definitive clustering. The South-central and North-central subclades are subdivisions of the “F-clade” of African elephant mtDNA, which are believed to have originally derived from the African forest elephant (*Loxodonta cyclotis*) before being transferred to the savannah species by hybridization and backcrossing to savannah elephant males. The inter-species transfer of mtDNA is common among closely related species of elephantids and other taxa in which males and not females disperse (Li et al., 2015; Petit and Excoffier, 2009; Roca et al., 2015).

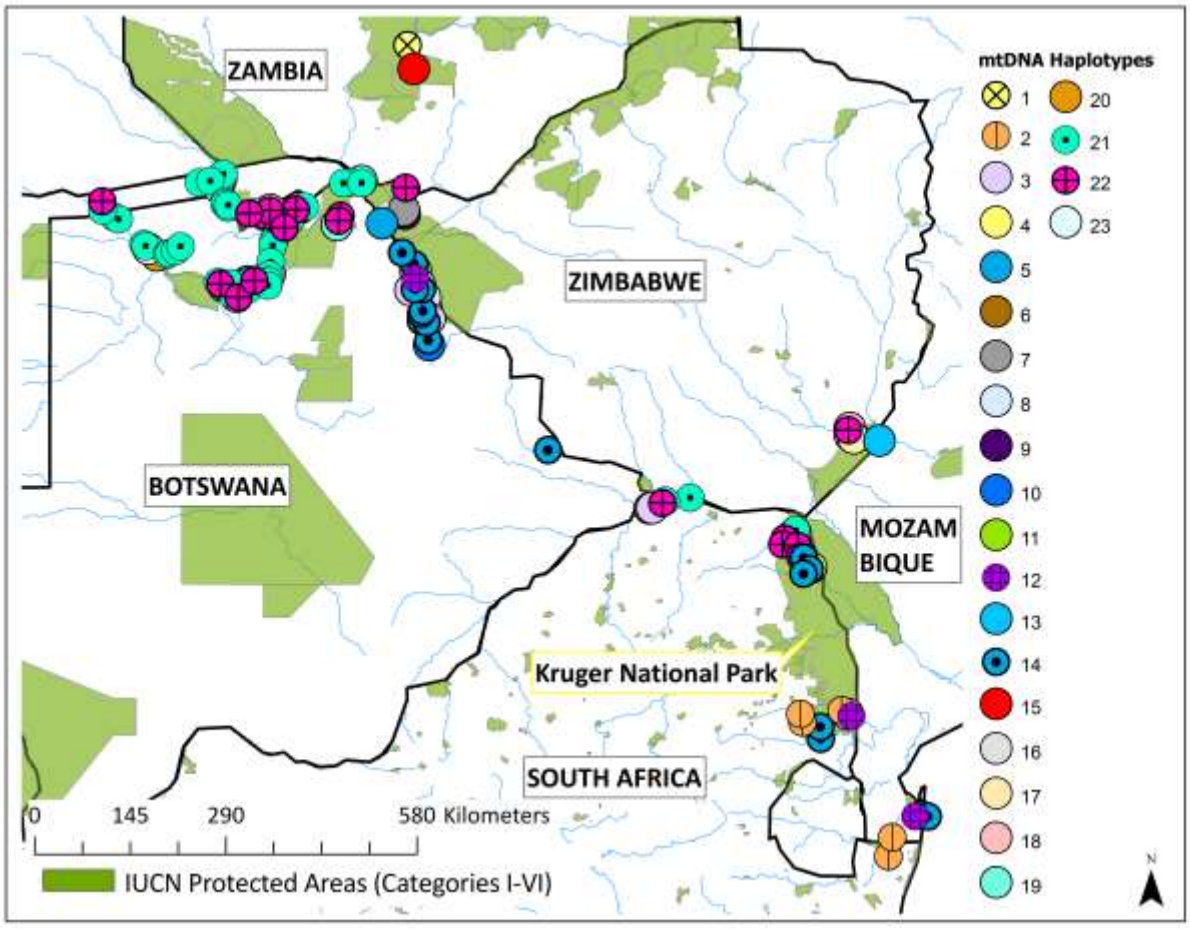

Fig. 4 Geographic distribution of mitochondrial DNA haplotypes (color figure available online).

Phylogeographic analyses of mitochondrial and nuclear DNA datasets

A Bayesian clustering algorithm implemented in Geneland identified four (K=4) genetically distinct clusters of mtDNA. Three of these mtDNA clusters were present in many national parks throughout the sampling area (see the maps of the posterior probability of belonging to a cluster in Fig. 5), but also were not geographically contiguous. Cluster 4, which included only the single individual that carried Haplotype 1, was limited to the extreme northern part of our study area (see Supplementary Table 3 for haplotypes and number of individuals within each of the clusters). Since Geneland clusters

may reflect both geographic and genetic differences, we examined the degree to which these clusters would show genetic differences by treating them as “populations”. AMOVA found that most (86.06%) of the genetic variation could be attributed to variation among the “populations,” with little (13.92%) variation within each of the “populations” (Table 1), resulting in a high Fst value of 0.86. An exact test for population differentiation showed that cluster 4 (Haplotype 1) was not significantly different from other clusters (presumably since it represented a single sample) but all other clusters (1, 2 and 3) differed significantly, and also showed significant Fst values. The clusters identified by Geneland corresponded to clades with high support in the phylogeny (e.g. cluster 1 contains individuals with haplotype 2, 3 and 4), suggesting that the mtDNA patterns were being driven by the genetic distinctiveness among the phylogenetic clades of mitochondrial DNA. Similar to Geneland, BAPS clustering analysis of mtDNA sequence data found the best partition of data to be four populations (K=4; Log(ml)=-720.110) that were discontinuously distributed across the landscape (Supplementary Fig. 3). The geographically discontinuous patterns of mtDNA (Fig. 4) reflect the complex history of mitochondrial lineages that have persisted through time and are evident across the landscape as high mtDNA haplotype diversities in relatively small geographic space. These patterns may also reflect historical migrations that occurred when elephant distribution was continuous; such patterns for mtDNA would then have persisted due to low female dispersal among elephants. Such patterns call into question the suitability of using mtDNA to study population structure in African savannah elephants.

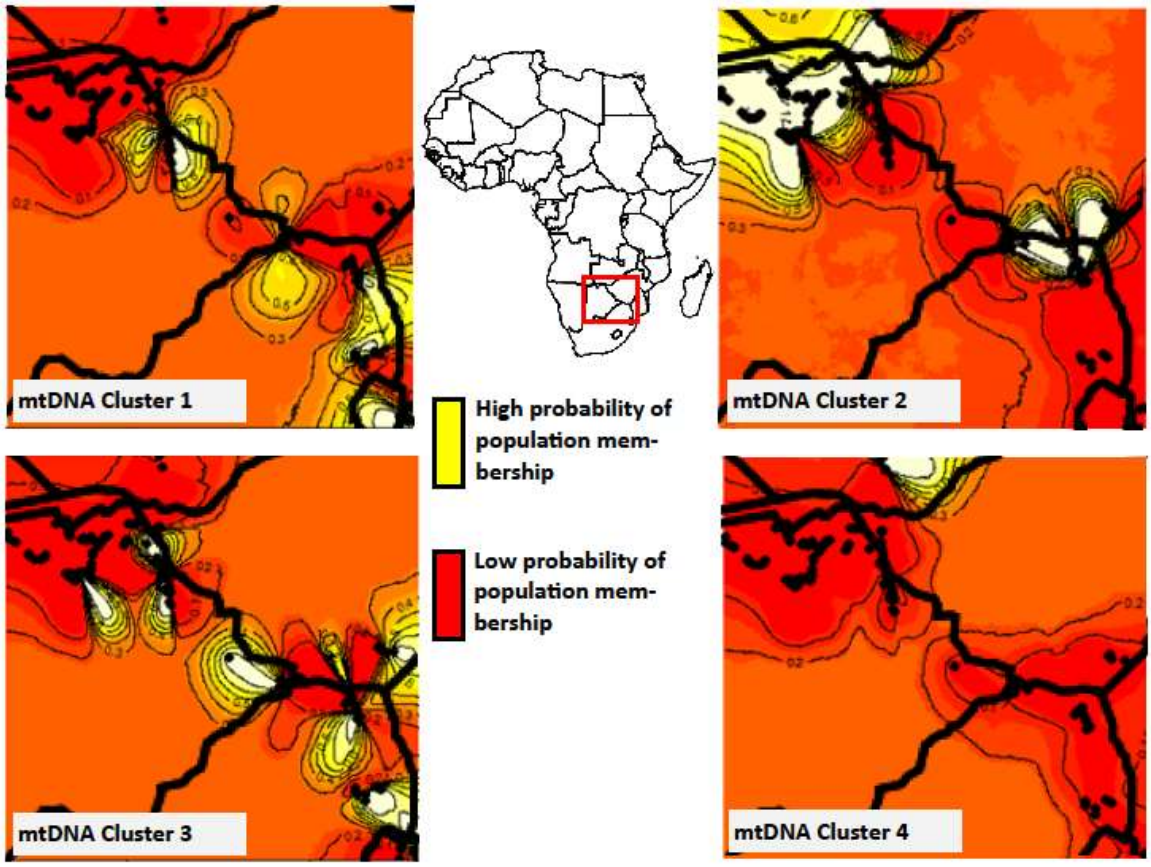

Fig. 5 Four mtDNA clusters were identified based on a Bayesian clustering algorithm implemented in the program Geneland. Most of the "clusters" are discontinuous across geographic space, while Cluster 4 was limited to a single geographic location in the extreme north of the study region. Cluster details can be found in Supplementary Table 3. White and yellow represent areas of high probability of population membership (color figure available online).

Table 1 Analysis of molecular variance (AMOVA) of mtDNA clusters identified by the program Geneland showed high differentiation between clusters (Fst), with most of the variation attributed to differences among clusters

Source of variation	df	Sum of squares	Variance components	Percentage of variation
Among clusters	3	350.274	5.281	86.08
Within clusters	137	117.059	0.854	13.92
Total	140	467.333	6.136	100
Fixation index (Fst):	0.86076			

Geneland identified four (K=4) distinct clusters of nDNA. Each nuclear DNA microsatellite cluster was geographically contiguous and did not overlap geographically with other clusters (Fig. 6). Cluster 1 represented a central population with individuals from Mapungubwe National Park and along the southern part of the Botswana-Zimbabwe border. Cluster 2 was found in the northern part of the study region in Chobe National Park and along the northern section of the Botswana-Zimbabwe border. Clusters 3 and 4, respectively, were centered in the northern and southern part of Kruger National Park, South Africa. However, AMOVA (Table 2) of the nDNA clusters found extremely low differentiation between clusters (Fst = 0.022), with most of the variation (98%) found within clusters and little (2%) variation attributed to among cluster differences. None of the clusters was significantly different ($p>0.05$) genetically in pairwise Fst comparisons between clusters. Since Geneland clusters may reflect both geographic and genetic differences, the lack of genetic differentiation suggests that the clustering was an artifact of the geographic patterns present in the dataset, and did not reflect strong nuclear genetic differentiation among clusters. This is supported by the BAPS clustering analysis that found this nDNA dataset to conform to a single population (K=1; Log(ml)=-1889.757) (Supplementary Fig. 3).

Table 2 Analysis of molecular variance (AMOVA) of nDNA clusters identified by the program Geneland showed low differentiation between nDNA clusters (Fst), with most of the variation attributed to differences within clusters

Source of variation	df	Sum of squares	Estimated variance	Percentage of variation
Among clusters	3	17.267	0.075	2
Within clusters	134	453.64	3.385	98
Total	137	470.913	3.461	100
Fixation index (Fst):	0.022			

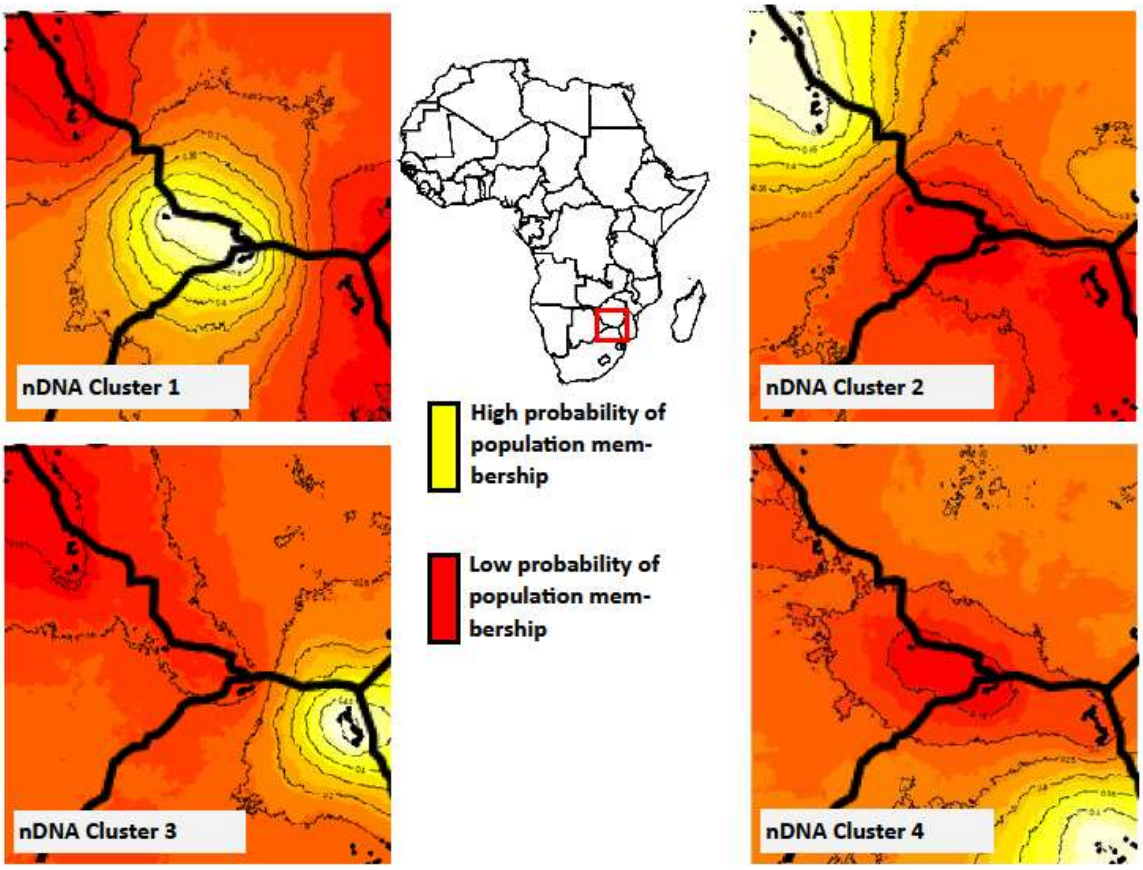

Fig. 6 Four nDNA clusters were identified by a Bayesian clustering algorithm implemented in the program Geneland. Cluster 1 (8 individuals) falls within the center of the study area, cluster 2 (28 individuals) is limited to the extreme north, cluster 3 (24 individuals) is centered in the northern part of Kruger National Park in South Africa, and cluster 4 (9 individuals) is centered in the southern part of Kruger National Park. White and yellow represent areas of high probability of population membership (color figure available online).

Comparison of Kruger elephants to other populations in southern Africa

To establish the origins of Kruger elephants, mitochondrial DNA haplotype frequencies were compared across elephants in parks north of Kruger, northern Kruger, southern Kruger, and parks south of Kruger (Fig. 7). Although all four regions shared some haplotypes (e.g., Haplotype 14 – Fig. 7), this comparison revealed that common haplotypes in northern Kruger National Park were also common in parks north of Kruger but not common in parks south of Kruger. By contrast, haplotypes that were common in southern Kruger National Park were also common in parks south of Kruger, but not common

in parks north of Kruger (Fig. 7). For example, haplotypes 21 and 22 were both common in north Kruger and northern parks, while being uncommon in south Kruger and southern parks. By contrast, haplotypes 12 and 2 were common in south Kruger and southern Parks, but uncommon in north Kruger and northern parks. This would seem to call into question the hypothesis that all elephants in Kruger derived from a single source population.

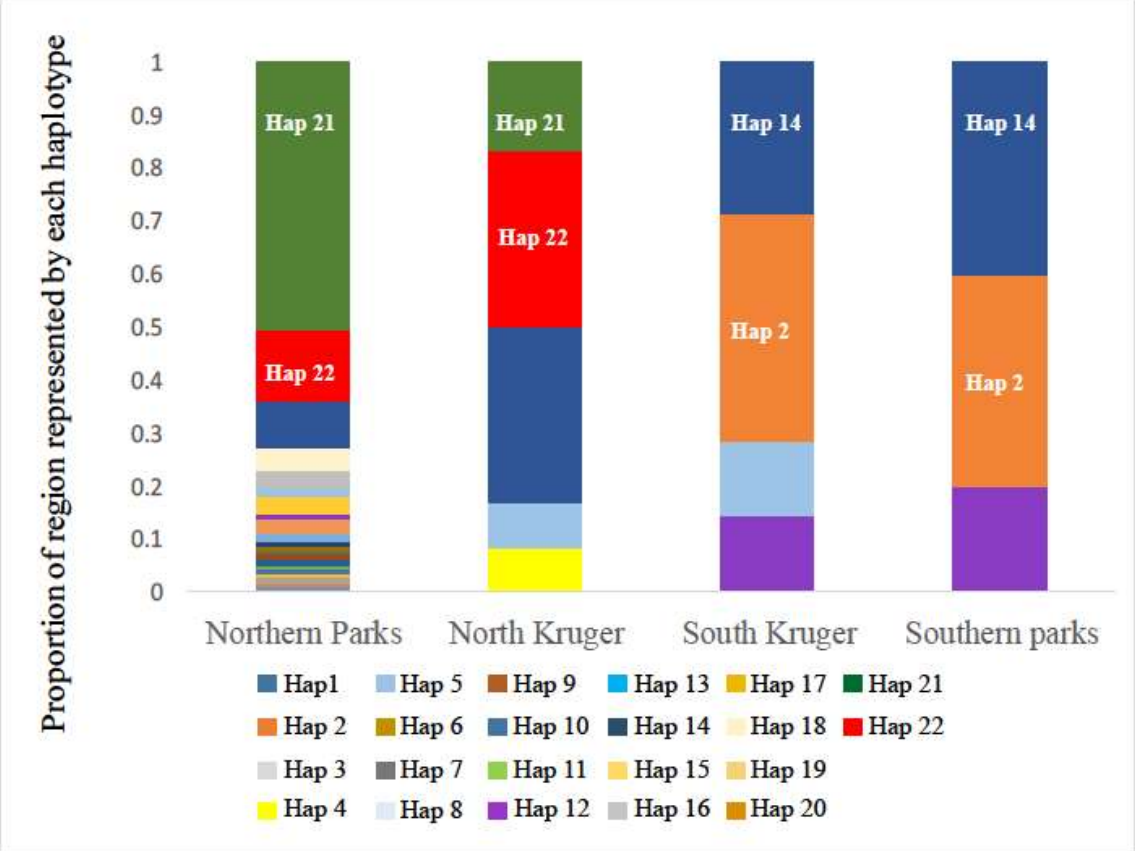

Fig. 7 Haplotype frequencies are shown for elephants placed into four geographic groupings: north of Kruger (N=117), northern Kruger (N=12), southern Kruger (N=7), and south of Kruger (N=5). Haplotypes are color coded and shown from common (top) to rare (bottom). The haplotype color legend is shown below the bar chart, and common haplotypes in each region are indicated on the bar graph. Note the similarity in frequencies of common haplotypes, both between north Kruger and northern parks (especially haplotypes 21 and 22) and between south Kruger and southern parks (especially haplotypes 12 and 2).

Finally, we would expect that Kruger National Park, if it had been subject to a founder effect followed by isolation, could show reduced nuclear genetic diversity. The elephants in our study area

were grouped into seven geographic regions and heterozygosity was determined for each of the regions. The observed nDNA heterozygosity for elephants in Kruger was higher than the value for four of the seven regions, and Kruger overall ranked fourth highest among the seven regions (Fig. 8, Supplementary Table 4). Elephants from Hwange Region, the Botswana-Zimbabwe border Region, and Gonarezhou Region had observed heterozygosity higher than Kruger, while those in Chobe Region, Livingstone Region, and Mapungubwe Region had observed heterozygosities lower than the Kruger Region. There was low differentiation between these regions (Fst = 0.027, Supplementary Table 5) with no significant differentiation between populations (exact test of population differentiation p>0.05), and heterozygosity did not differ significantly between these regions (Kruskal-Wallis K=5.62; p>0.5). The

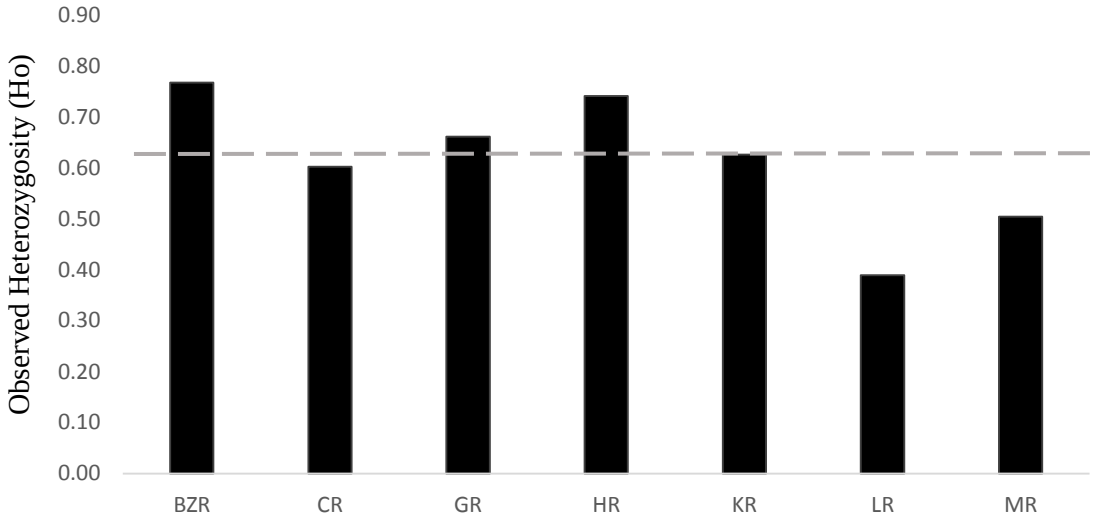

Fig. 8 Observed heterozygosity for seven regions defined by geographic proximity and connectivity. The regions are the Botswana-Zimbabwe border Region (BZR) = BZ; Chobe Region (CR) = CH, MR, CP, SS, SN, LY; Gonarezhou Region (GR) = GR; Hwange Region (HR) = HW, PT; Kruger Region (KR) = NKNP, SKNP; Livingstone Region (LR) = LV; Mapungubwe Region (MR) = MAP, LMAP, BZIM. The dashed horizontal line indicates the observed heterozygosity of elephants in Kruger National Park, and was included to aid comparisons between regions. Locality abbreviations are defined in the main text.

Kruger population did not significantly differ genetically from the overall population (exact test of population differentiation $p>0.05$), and heterozygosity in Kruger did not differ significantly from overall heterozygosity (Mann-Whitney U test $U=33$; $p>0.5$).

We did not find evidence of a recent reduction in effective population size when considering allele data frequencies (Supplementary Table 7). Prob(H>He) is the probability that the heterozygosity (H) is larger than the average (He) under the null hypothesis, if Prob(H>He) is lower than 0.05, the null hypothesis (mutation drift equilibrium) is rejected in favor of the hypothesis of a recent genetic bottleneck. Prob(H>He) for the sign test under the SSM and TPM models was >0.5, and Prob(H>He) for both 1-tailed and 2-tailed Wilcoxon tests under the SMM and TPM models was >0.5. The mode shift test showed a normal L-shaped distribution as expected under mutation drift equilibrium.

Discussion:

Contrary to our hypothesis, the elephant population in Kruger did not show evidence of genetic isolation, a depleted gene pool or a strong founder effect. Genetic patterns from nDNA showed little differentiation between Kruger and nearby parks, with Kruger elephants showing levels of heterozygosity similar to those of nearby regions. Observed heterozygosity in Kruger ranked fourth highest when compared to six other populations across the study area, and heterozygosity in Kruger did not differ significantly from the overall heterozygosity of all other populations combined, and therefore did not show evidence of a depleted gene pool that typically results from genetic isolation and/or a small founder population. This is corroborated by the lack of evidence of a detectable population bottleneck, suggesting that the Kruger population did not experience a recent reduction in effective population size. Although the present study did not detect a population bottleneck, previous studies have shown that known population bottlenecks are not always detected (Peery et al., 2012; Whitehouse and Harley, 2001). The relatively recent population extirpation followed by a rapid increase in

population size (along with potential subsequent migration) may have precluded the genetic signature of a bottleneck in Kruger.

Our results suggest that the Kruger elephant population was founded by elephants from more than one geographic source, and not from a single source population as has been previously suggested (Whyte, 2001b). Mitochondrial DNA frequency comparisons (Fig. 7) indicated that elephants in southern Kruger may have originated from populations south of Kruger, while elephants in northern Kruger may have migrated into Kruger from regions north of Kruger. At the least, our findings suggest that if Kruger was founded by a single source population, it has since been augmented through gene flow and migration from nearby parks either to the north or to the south of Kruger.

Phylogenetic relationships among most mtDNA haplotypes were poorly resolved by the Bayesian analyses, which is not unexpected when using a single relatively short sequence for phylogenetic inference. Future studies that aim to better determine the relationships among haplotypes should incorporate longer sequences (e.g., complete mitogenomes).

One relationship that was evident in the Bayesian analysis was the placement of Haplotype 15 outside of a clade containing the other savannah elephant haplotypes. Haplotype 15 along with Haplotypes 16-23 (South-central subclade) and Haplotype 1 (North-central subclade) (Supplementary Fig. 2) are part of the F-clade, a grouping of mitochondrial DNA that is widely distributed among savannah elephants across Africa, but is believed to have originated in ancient inter-species transfer from the African forest elephant (Roca et al. 2015). Our findings suggest that ancient hybridization and backcrossing led to the transfer of forest mtDNA haplotypes that spread through female migration to Kafue National Park which contains Haplotype 15 as part of the South-central subclade, and also to other northern parks in the study area (e.g., Chobe National Park) that also contain haplotypes from the South-central subclade.

The discordance between mtDNA and nDNA clusters is consistent with previous reports of differences in the African elephant mitochondrial and nuclear phylogeographic patterns (Archie et al., 2008, Okello et al. (2008), Nyakaana and Arctander (1999), Ishida et al. (2011). Ishida et al. (2011) showed that genetic patterns based on nDNA and mtDNA can be quite different across the African continent. Mitochondrial DNA patterns reflect ancient female migration, with the patterns persisting due to lack of dispersal of females from core social groups (Archie et al., 2007; Fishlock and Lee, 2013). By comparison, the male-biased dispersal typical of elephants (Nyakaana and Arctander, 1999) would tend to erase geographic differences in nDNA patterns, leading to nDNA phylogeographic patterns that may differ greatly from those present among maternally inherited mtDNA.

Management strategies for elephants in Kruger National Park have encompassed a myriad of approaches (Whyte, 2001b). Early management strived to increase elephant numbers in the park, later strategies aimed at decreasing elephant numbers in the park, while current strategies aim at sustainable self-regulation of population numbers through ecological processes such as resource dependent or induced dispersal and density dependent population growth stabilization (van Aarde et al., 1999). Our findings suggest that Kruger's elephant population has not suffered from the negative genetic consequences that may be associated with population isolation. This is likely due to historical founding events involving more than one source population, along with a rapid increase to a large population size, and possible subsequent immigration into Kruger that augmented the gene pool. Similar genetic patterns may be found in parks with similar histories to Kruger. Ultimately this study highlights the importance of migration and gene flow for maintaining genetic diversity in populations. Our findings suggest that Kruger forms part of a functional entity in which migration helped to maintain a relatively diverse gene pool. Our findings therefore emphasize the need for conservation initiatives such as transfrontier conservation areas that aim at maintaining connectivity between populations. Initiatives that maintain connectivity may provide a sustainable, self-regulating management approach for

elephants in southern Africa that simultaneously upholds genetic diversity and gene flow across Kruger and nearby parks.

Acknowledgements

The projected was funded through grants to RJvA from the International Fund for Animal Welfare and the Conservation Foundation (Zambia). Elephants Without Borders (Botswana) kindly facilitated sample collection. The Zambian Wildlife Authority, the Department of Wildlife and National Parks (Botswana), South African National Parks (SANParks) and the Department of Agriculture, Forestry and Fisheries (South Africa) sanctioned our research activities. We acknowledge the support of a team of post-graduate students that assisted with the collection of study material, and the support provided by the University of Pretoria's Sequencing Facility. Travel and research support for AdeF was provided through the Francis M. and Harlie M. Clark Research Support Grant (2016), the Harley J. Van Cleave Research Award, and the University of Illinois Graduate College Dissertation Project Travel Grant. AdeF was also supported by the Cooperative State Research, Education, and Extension Service, US Department of Agriculture, under project number ILLU 875-952. ALR was supported by the US Fish and Wildlife Service African Elephant Conservation Fund.

Data Accessibility

Mitochondrial sequences have been deposited in GenBank (MF062095- MF062117).

References:

Allendorf, F.W., Luikart, G., Aitken, S.N., 2013. Conservation and the genetics of populations, 2nd ed. Blackwell Publishing, Wiley and Sons. LTD, West Sussex, UK.

Andreasson, H., Gyllensten, U., Allen, M., 2002. Real-time DNA quantification of nuclear and mitochondrial DNA in forensic analysis. Biotechniques 33, 402–411.

Archie, E.A., HOLLISTER-SMITH, J.A., Poole, J.H., Lee, P.C., Moss, C.J., Maldonado, J.E., Fleischer, R.C., Alberts, S.C., 2007. Behavioural inbreeding avoidance in wild African elephants. Mol. Ecol. 16, 4138–4148.

Archie, E.A., Maldonado, J.E., HOLLISTER-SMITH, J.A., Poole, J.H., Moss, C.J., Fleischer, R.C., Alberts, S.C., 2008. Fine-scale population genetic structure in a fission–fusion society. Mol. Ecol. 17, 2666–2679.

Comstock, K.E., Georgiadis, N., Pecon-Slattery, J., Roca, A.L., Ostrander, E.A., O'Brien, S.J., Wasser, S.K., 2002. Patterns of molecular genetic variation among African elephant populations. Mol. Ecol. 11, 2489–2498. doi:10.1046/j.1365-294X.2002.01615.x

Corander, J., Waldmann, P., Sillanpää, M.J., 2003. Bayesian analysis of genetic differentiation between populations. Genetics 163, 367–374.

Cornuet, J.M., Luikart, G., 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144, 2001–2014.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772–772.

de Flamingh, A., Sole, C.L., Van Aarde, R.J., 2015. Genetic evidence for spatial structuring in a continuous African elephant (Loxodonta africana) population. Conserv. Genet. 16, 613–623.

de Flamingh, A., Sole, C.L., van Aarde, R.J., 2014. Microsatellite repeat motif and amplicon length affect amplification success of degraded faecal DNA. Conserv. Genet. Resour. 6, 503–505. doi:10.1007/s12686-014-0160-5

Dublin, H.T., Niskanen, L., 2003. Guidelines for the in situ translocation of the African elephant for conservation purposes. IUCN.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Eggert, L.S., Rasner, C.A., Woodruff, D.S., 2002. The evolution and phylogeography of the African elephant inferred from mitochondrial DNA sequence and nuclear microsatellite markers. Proc. R. Soc. Lond. B Biol. Sci. 269, 1993–2006.

Enk, J., Devault, A., Debruyne, R., King, C.E., Treangen, T., O'Rourke, D., Salzberg, S.L., Fisher, D., MacPhee, R., Poinar, H., 2011. Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. Genome Biol. 12, 1.

Excoffier, L., Lischer, H.E., 2010a. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567.

Excoffier, L., Lischer, H.E.L., 2010b. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x

Fischer, J., Lindenmayer, D.B., 2007. Landscape modification and habitat fragmentation: a synthesis. Glob. Ecol. Biogeogr. 16, 265–280. doi:10.1111/j.1466-8238.2007.00287.x

Fishlock, V., Lee, P.C., 2013. Forest elephants: fission–fusion and social arenas. Anim. Behav. 85, 357–363.

Fitzpatrick, J.L., Evans, J.P., 2009. Reduced heterozygosity impairs sperm quality in endangered mammals. Biol. Lett. 5, 320–323. doi:10.1098/rsbl.2008.0734

Gilbert, M.T.P., Drautz, D.I., Lesk, A.M., Ho, S.Y., Qi, J., Ratan, A., Hsu, C.-H., Sher, A., Dalén, L., Götherström, A., others, 2008. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. Proc. Natl. Acad. Sci. 105, 8327–8332.

Guillot, G., Mortier, F., Estoup, A., 2005. Geneland: a computer package for landscape genetics. Mol. Ecol. Notes 5, 712–715. doi:10.1111/j.1471-8286.2005.01031.x

Hall-Martin, A.J., 1992. Distribution and status of the African elephant Loxodonta africana in South Africa. Koedoe 35, 65–88.

Hanks, J., 2003. Transfrontier Conservation Areas (TFCAs) in Southern Africa: their role in conserving biodiversity, socioeconomic development and promoting a culture of peace. J. Sustain. For. 17, 127–148.

Ishida, Y., Georgiadis, N.J., Hondo, T., Roca, A.L., 2013. Triangulating the provenance of African elephants using mitochondrial DNA. Evol. Appl. 6, 253–265. doi:10.1111/j.1752-4571.2012.00286.x

Ishida, Y., Oleksyk, T.K., Georgiadis, N.J., David, V.A., Zhao, K., Stephens, R.M., Kolokotronis, S.-O., Roca, A.L., 2011. Reconciling apparent conflicts between mitochondrial and nuclear phylogenies in African elephants. PloS One 6, e20642.

Krause, J., Dear, P.H., Pollack, J.L., Slatkin, M., Spriggs, H., Barnes, I., Lister, A.M., Ebersberger, I., Pääbo, S., Hofreiter, M., 2006. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature 439, 724–727.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Luikart, G., Allendorf, F.W., Cornuet, J.M., Sherwin, W.B., 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J. Hered. 89, 238–247.

Maikaeu, U., Wajjwalku, W., Siripholwat, V., Mahasawankul, S., 2007. The study of nucleotides arrangement of mitochondrial DNA in Thai elephants. KKU Vet. J. 17, 11–21.

McDonald, J.H., 2009. Handbook of biological statistics. Sparky House Publishing Baltimore, MD.

Miller, E.J., Eldridge, M.D.B., Morris, K.D., Zenger, K.R., Herbert, C.A., 2011. Genetic consequences of isolation: island tammar wallaby (Macropus eugenii) populations and the conservation of threatened species. Conserv. Genet. 12, 1619–1631. doi:10.1007/s10592-011-0265-2

Nyakaana, S., Arctander, P., 1999. Population genetic structure of the African elephant in Uganda based on variation at mitochondrial and nuclear loci: evidence for male-biased gene flow. Mol. Ecol. 8, 1105–1115.

Okello, J.B., Masembe, C., Rasmussen, H.B., Wittemyer, G., Omondi, P., Kahindi, O., Muwanika, V.B., Arctander, P., Douglas-Hamilton, I., Nyakaana, S., others, 2008. Population genetic structure of savannah elephants in Kenya: conservation and management implications. J. Hered. 99, 443–452.

Peakall, R., Smouse, P.E., 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6, 288–295.

Peery, M.Z., Kirby, R., Reid, B.N., Stoelting, R., Doucet-BëEr, E., Robinson, S., VáSquez-Carrillo, C., Pauli, J.N., PalsbøLl, P.J., 2012. Reliability of genetic bottleneck tests for detecting recent population declines: RELIABILITY OF GENETIC BOTTLENECK TESTS. Mol. Ecol. 21, 3403–3418. doi:10.1111/j.1365-294X.2012.05635.x

Pienaar, U. de V., 1963. Large mammals of the Kruger National Park - their distribution and present-day status. Koedoe 6, 1–137.

Rambaut, A., Suchard, M., Xie, D., Drummond, A., 2014. Tracer v1.6. Tracer V16 Available HttpbeastbioedacukTracer 6.

Ripple, W.J., Newsome, T.M., Wolf, C., Dirzo, R., Everatt, K.T., Galetti, M., Hayward, M.W., Kerley, G.I.H., Levi, T., Lindsey, P.A., Macdonald, D.W., Malhi, Y., Painter, L.E., Sandom, C.J., Terborgh, J., Van Valkenburgh, B., 2015. Collapse of the world's largest herbivores. Sci. Adv. 1. doi:10.1126/sciadv.1400103

Roca, A.L., Georgiadis, N., Pecon-Slattery, J., O'Brien, S.J., 2001. Genetic evidence for two species of elephant in Africa. Science 293, 1473–1477.

Roca, A.L., Ishida, Y., Brandt, A.L., Benjamin, N.R., Zhao, K., Georgiadis, N.J., 2015. Elephant natural history: a genomic perspective. Annu Rev Anim Biosci 3, 139–167.

Roever, C.L., van Aarde, R.J., Leggett, K., 2013. Functional connectivity within conservation networks: Delineating corridors for African elephants. Biol. Conserv. 157, 128–135. doi:10.1016/j.biocon.2012.06.025

Rogaev, E.I., Moliaka, Y.K., Malyarchuk, B.A., Kondrashov, F.A., Derenko, M.V., Chumakov, I., Grigorenko, A.P., 2006. Complete mitochondrial genome and phylogeny of Pleistocene mammoth Mammuthus primigenius. PLoS Biol 4, e73.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Scholes, R.J., Mennell, K.G. (Eds.), 2008. Elephant management: a scientific assessment for South Africa. Wits University Press, Johannesburg, South Africa.

Schwarz, C., Debruyne, R., Kuch, M., McNally, E., Schwarcz, H., Aubrey, A.D., Bada, J., Poinar, H., 2009. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic Acids Res. 37, 3215–3229. doi:10.1093/nar/gkp159

Taberlet, P., Waits, L.P., Luikart, G., 1999. Noninvasive genetic sampling: look before you leap. Trends Ecol. Evol. 14, 323–327.

van Aarde, R., Whyte, I., Pimm, S., 1999. Culling and the dynamics of the Kruger National Park African elephant population. Anim. Conserv. 2, 287–294.

van Aarde, R.J., Jackson, T.P., 2007. Megaparks for metapopulations: addressing the causes of locally high elephant numbers in southern Africa. Biol. Conserv. 134, 289–297.

van Aarde, R.J., Whyte, I., Pimm, S., 1999. Culling and the dynamics of the Kruger National Park African elephant population. Anim. Conserv. 2, 287–294.

Whitehouse, A.M., Harley, E.H., 2001. Post-bottleneck genetic diversity of elephant populations in South Africa, revealed using microsatellite analysis. Mol. Ecol. 10, 2139–2149.

Whyte, I., 2001a. Conservation management of the Kruger National Park elephant population. University of Pretoria, Pretoria, South Africa.

Whyte, I., 2001b. Conservation management of the Kruger National Park elephant population. Pretoria: UP. Ph. D. thesis: Zoology.

Young, K.D., Ferreira, S.M., van AARDE, R.J., 2009. The influence of increasing population size and vegetation productivity on elephant distribution in the Kruger National Park. Austral Ecol. 34, 329–342. doi:10.1111/j.1442-9993.2009.01934.x