

Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci

S Nyakaana¹, P Arctander² and HR Siegismund²

¹Institute of Environment and Natural Resources, Makerere University, P.O. Box 7298, Kampala, Uganda; ²Zoological Institute, Department of Evolutionary Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

Two hundred and thirty-six mitochondrial DNA nucleotide sequences were used in combination with polymorphism at four nuclear microsatellite loci to assess the amount and distribution of genetic variation within and between African savannah elephants. They were sampled from 11 localities in eastern, western and southern Africa. In the total sample, 43 haplotypes were identified and an overall nucleotide diversity of 2.0% was observed. High levels of polymorphism were also observed at the microsatellite loci both at the level of number of alleles and gene diversity. Nine to 14 alleles per locus across populations and 44 alleles in the total sample were found. The gene diversity ranged from 0.51 to 0.72

in the localities studied. An analysis of molecular variance showed significant genetic differentiation between populations within regions and also between regions. The extent of subdivision between populations at the mtDNA control region was approximately twice as high as shown by the microsatellite loci (mtDNA F_{ST} = 0.59; microsatellite R_{ST} = 0.31). We discuss our results in the light of Pleistocene refugia and attribute the observed pattern to population divergence in allopatry accompanied by a recent population admixture following a recent population expansion.

Heredity (2002) **89**, 90–98. doi:10.1038/sj.hdy.6800110

Keywords: savannah elephant; *Loxodonta africana*; control region; genetic differentiation; microsatellites; population structure

Introduction

The current distribution of elephants in Africa is restricted to, but extends all over the African continent south of the Sahara. Although elephants are capable of utilising all kinds of habitats ranging from dense forests to deserts, most of the surviving populations are currently found restricted within the boundaries of protected areas. They are highly mobile mammals capable of migrating over long distances in search of water and food and with little hindrance from ecological and geographical barriers. However, with their matrilineal social structure, which is characterised by maternal philopatry, their propensity to migrate and exchange individuals between populations is higher in males than in females (Nyakaana and Arctander, 1999). *A priori*, these characteristics are expected to maintain a reasonable level of homogenising gene flow between different populations, all other factors remaining constant.

Commercial poaching and encroachment of human settlements into former elephant ranges during the last 100 years has greatly threatened the survival and

viability of elephant populations. The current population size of the African elephant is estimated at only 400 000 to 500 000 individuals (Barnes *et al*, 1999). Consequently, the African elephant has been listed as an Appendix I species by the Convention for the International Trade of Endangered Species of wild flora and fauna (CITES), as a vulnerable species in the International Union for the Conservation of Nature (IUCN) red data book and as a threatened species by the United States Fish and Wildlife Service under the Endangered Species Act.

Many studies so far conducted on African elephants have concentrated mainly on their ecology, management and conservation, demography, biology and interaction with vegetation. Only a few studies to date have endeavoured to assess genetic variation in African elephant populations: Georgiadis *et al* (1994) and Siegismund and Arctander (1995) reported significant genetic differentiation between elephant populations in eastern and southern Africa based on molecular studies using restriction site analysis of mtDNA. Nyakaana and Arctander (1999) used nucleotide sequences of the rapidly evolving 5' end of the control region in conjunction with allele length variation at four microsatellite loci and found significant genetic subdivision between elephant populations in Uganda. Barriel *et al* (1999), using cytochrome *b* nucleotide variation, found high levels of divergence between the forest elephant and the savannah elephant (comparable to what is found between the savannah elephant and the Asian elephant). Roca *et al* (2001)

Correspondence: Dr S Nyakaana, Makerere University Institute of Environment and Natural Resources, PO Box 7298, Kampala, Uganda.
E-mail: SNyakaana@muenr.ac.ug

Received 27 November 2001; accepted 22 April 2002

analysed four nuclear genes and reported a very deep genetic division between forest and savannah populations of African elephants that led them to suggest that the two hitherto subspecies deserve species-level status. In this study we used nucleotide sequence variation in the highly variable mtDNA control region in conjunction with allele length variation at microsatellite loci to analyse population structure of the African savannah elephant sampled from eastern, southern and western Africa.

Materials and methods

Sampling and DNA extraction

Tissue and/or fresh faecal samples were collected from 245 elephants from 11 localities in eastern Africa (Kidepo Valley NP (KV) = 27, Queen Elizabeth NP (QE) = 68, Murchison Falls NP (MF) = 9, Masai Mara GR (MM) = 20, Amboseli NP = 27, Samburu GR (SA) = 9); western Africa (Ghana (WA) = 18); and southern Africa (Zimbabwe (ZBE) = 6, Botswana (BOT) = 25, Namibia (NAM) = 24 and Kruger NP (KG) = 12) as shown in Figure 1. Samples were preserved in 25% dimethylsulfoxide (DMSO) saturated with sodium chloride (Amos and Hoelzel, 1991) at room temperature in the field and at -80°C in the laboratory. Faecal samples constituted only 7.7% of all the samples analysed. Thirty-one percent of the sequences used in this study are derived from the previous study of Nyakaana and Arctander (1999). Total genomic DNA

was phenol-extracted using standard procedures (Sambrook *et al*, 1989).

Mitochondrial control region

Amplification and sequencing: We amplified 400 base pairs (bp) of the 5' hyper-variable segment of the control region using the polymerase chain reaction (PCR) with primers Laf CR1 and Laf CR2 (Nyakaana and Arctander, 1999). Symmetrical PCR amplifications (Saiki *et al*, 1988) were carried out in 20 μL reaction volumes containing 2–5 ng of total genomic DNA, 20 pmol of each primer, 200 μM dNTPs, 10 mM Tris-HCl, 1.5 mM MgCl_2 , 50 mM KCl, pH 8.3 and 0.4 units of *Taq* polymerase (Boehringer Mannheim GmbH). The cycling regime comprised of one cycle of initial denaturation at 94°C for 5 min followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 46°C for 2 min and extension at 72°C for 3 min. Asymmetrical amplification (Gyllensten and Erlich, 1988) to obtain single-strand DNA template for sequencing was carried out in 50 μL reaction volumes using 2–5 ng template DNA, 0.5 pmol of the sequencing primer, 50 pmol of the second primer, 500 μM dNTPs, 10 mM Tris-HCl, 1.5 mM MgCl_2 , 50 mM KCl, pH 8.3 and 0.4 units of *Taq* polymerase (Boehringer Mannheim GmbH). A cycling profile of 37 cycles of denaturation at 94°C for 1 min, annealing at 48°C and extension at 72°C for 1 min was used. All amplifications were carried out using a HYBAID thermocycler. No-template PCR controls were included in each set of amplification reactions to check

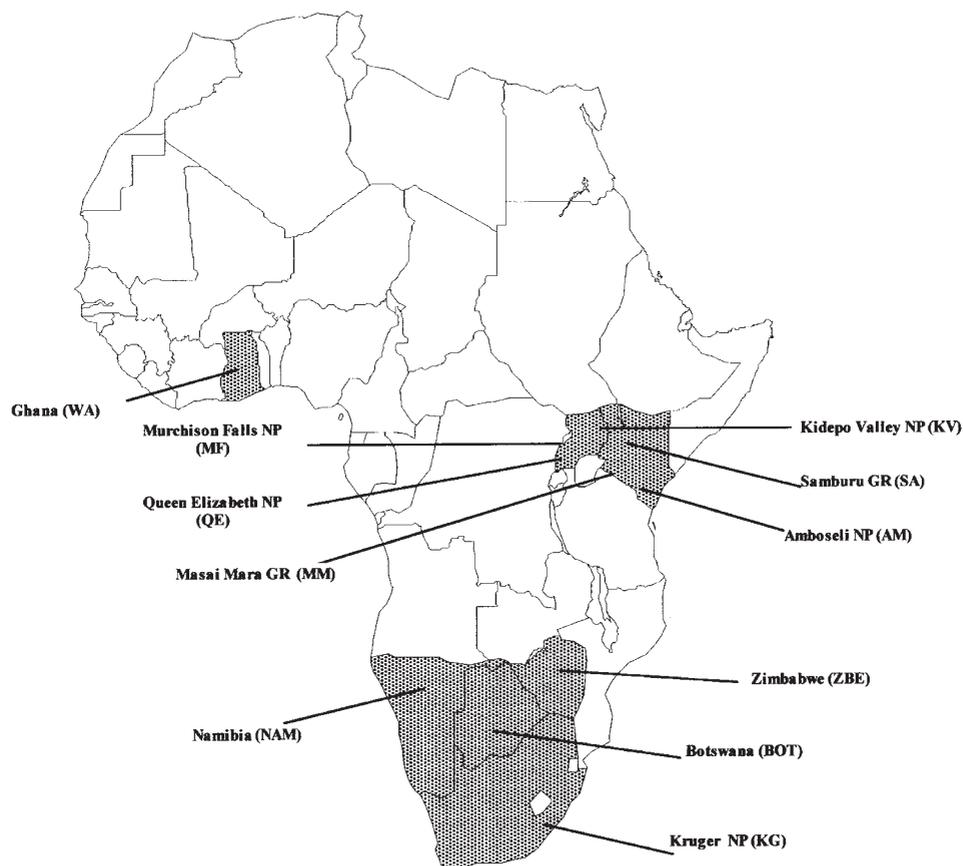


Figure 1 Map of Africa showing the geographical locations of the sampling localities.

for possible contamination. Primers Laf CR1 and Laf CR2 were used to sequence both strands using the dideoxy chain termination method (Sanger *et al*, 1977), [α - 35 S]-ATP (Amersham, UK) and a SEQUENASE version 2.0 kit (US Biochemical). The products were separated in a 6% polyacrylamide/7M urea gel and visualised by autoradiography on a Kodak film after exposure for 24–48 h. Some samples were sequenced for both strands using the dye terminator cycle sequencing technique on an ABI 377 automated sequencer (Perkin Elmer).

Sequence analyses: Sequences were aligned manually and edited using the program SeqApp version 1.9 (Gilbert, 1993). Within-population genetic diversity was estimated using the nucleotide diversity index (π) (Nei, 1987, equation 10.5). The extent of genetic differentiation between populations was quantified using the F_{ST} statistic computed from both the haplotype frequencies and the number of mutations between the haplotypes using the program ARLEQUIN version 2.0 (Schneider *et al*, 2000). The statistical significance levels of these two statistics were determined using 1000 Monte Carlo simulations.

The pattern of sequence evolution was portrayed using a minimum spanning network generated with the programs TCS version 1.13 (Clement *et al*, 2000) and ARLEQUIN (Schneider *et al*, 2000). In this network, the sequences can be at the nodes and the tips of the network. The net percentage sequence divergence between populations was used to estimate a tree among the populations using the neighbour-joining algorithm implemented in PHYLIP version 3.5c (Felsenstein, 1993).

The genetic structure of the populations was analysed in a hierarchical way where populations were *a priori* categorised into three regional groupings mentioned previously. An analysis of molecular variance approach (AMOVA) implemented in ARLEQUIN version 2.0 (Schneider *et al*, 2000), which uses both the frequency and sequence divergence between haplotypes, was used to estimate the extent of genetic differentiation between populations within each region and also between the three geographical regions. The hierarchical analysis was done by estimating the F_{ST} , F_{SC} and F_{CT} fixation indices that measure the proportion of genetic variation between populations in the total sample, between populations within regions and between different regions of populations, respectively. The significances of the F -statistics were evaluated with 1000 random permutations.

Microsatellite loci

Polymorphism was analysed at each of the four loci Laf MS01, Laf MS02, Laf MS03 and Laf MS04 (Nyakaana and Arctander, 1998) using the conditions as described in Nyakaana and Arctander (1999). The PCR products were electrophoresed on a 4% polyacrylamide gel on an ABI model 377 DNA automated sequencer. Tests for genotypic disequilibrium and deviations from the Hardy-Weinberg proportions were carried out using the Markov chain method as implemented in GENEPOP version 3.1 (Raymond and Rousset, 1997). The numbers of different alleles per locus plus observed and expected heterozygosities at each locus were used as indices of genetic diversity in each population. A modification of Fisher's exact test (Raymond and Rousset, 1995) was used to evaluate the significance of the differences in allele distri-

bution between populations using 1000 Monte Carlo permutations. Bonferroni corrections (Rice, 1989) were used to compensate for simultaneous tests in estimating significance levels.

We quantified the extent of genetic differentiation between populations using the R_{ST} statistic (Slatkin, 1995). This statistic is specifically adapted to microsatellite data because its computations are based on the stepwise mutation model (SMM) that explains the mutation processes believed to occur at microsatellite loci (Slatkin, 1995).

The microsatellite variation was analysed in a hierarchical manner similar to the control region variation with the program ARLEQUIN version 2.0 (Schneider *et al*, 2000). We tested for correlation between genetic distances (based on both mtDNA sequence and microsatellite loci data) and geographical distances between populations using a Mantel g -test (Mantel, 1967) both within and between regions. Two thousand permutations were used to estimate the significance of the correlations.

Results

Mitochondrial control region

Variation: Forty-three haplotypes defined by 47 polymorphic sites were found among the 236 individuals successfully sequenced (Figure 2). These have been deposited in GenBank with accession numbers AF106203 to AF106245. Of the 47 polymorphic sites, 44 were transitions, one involved a transversion, one involved both a transition and transversion, while one site represented a deletion/insertion event. Haplotype diversity ranged from 0.17 (Kruger NP) to 1.0 (Zimbabwe) (Table 1). Eight of the haplotypes were found in more than one population; four of which were restricted to the eastern African region (KV1, KV2, KV7 and KV8), two occurred in both eastern Africa and southern Africa populations (QE1 and QE13) and two were restricted to the southern Africa region (ZBE6 and BOT2) (see Figure 2).

Nucleotide diversity (π , Nei, 1987, equation 10.5) in the total sample was 2.0% but varied widely between populations, ranging from 0.08% in the Kruger population to 2.7% in Zimbabwe, representing approximately a 34-fold difference (Table 1). Haplotype and nucleotide diversities were significantly correlated ($r = 0.71$, $P < 0.05$).

Gene genealogy: A minimum spanning network for the haplotypes revealed three highly divergent clades: Clades A & B contained haplotypes from both the eastern and southern African regions and the two clades were separated by at least 10 mutational steps. Clade C was reciprocally monophyletic and contained four haplotypes all sampled from western Africa (Figure 3). This clade was distinguished from the rest by one fixed transition and one fixed transversion (positions 41 and, 209 respectively). All haplotypes in clade A had a deletion at position 39 while those in clades B & C had adenine in the same position. All haplotypes in clade A were inferred by the programme to be derived from one haplotype (QE1, rooting weight = 0.16) while those in clades B & C were derived from another haplotype (BOT16, rooting weight = 0.14). Eighty-two percent of all the individuals in clade A were sampled in eastern African

	10				20				30				40				POPULATION										
	1111111122				2222222222				2222222233				33333333				KV	MF	QE	MM	AM	SA	ZBE	BOT	NAM	KG	WA
	123344577	0135889900	1222333333	4555589900	1345558	1014901838	8912891569	6135123456	7014590507	1130186																	
KV1	AAAA-TCCCT	ACTTCTAGTA	TAGCCTAGCC	GCACACTGAA	TCCTGCC						9	3	1	7										20			
KV2TCG.TCA	C.....T						10													12			
KV7	.G.AC	.C.C.AC	.A.....TT	AT.T.A.G	.T.CA						5													11			
KV8											1	1		10										12			
KV17											1													1			
KV28											1													1			
MF1	.GG.AC	.C.C.AC	.A.....TT	AT.T.A.G	.T.CA						2													2			
MF5	GGG.AC	.C.C.AC	.A.....TT	AT.T.A.G	.T.CA						3													3			
QE1															64		15			3		2		84			
QE4	.G.AC	.C.C.AC	.GA.....TT	AT.T.A.G	.T.CA						2													2			
QE13	.AC	.C.C.AC	C.....TT	AT.T.A	.T.CA						1									3				4			
MM4																								1			
MM19																								1			
MM20																								1			
AM1																								5			
AM2																								1			
AM10	.G.AC	.C.C.AC	.A.....TT	AT.T.A	.T.CA																			1			
AM12																								5			
SA8																								1			
																								1			
ZBE1	...AC	...TC	C.....TT	AT.T.A	.T.CA						1													1			
ZBE2	...AC	...CCTC	C.....G	TT A.GT	.A...T.A						1													1			
ZBE3	...G	...T.G	C.....G	TT A.T	.C.....						1													1			
ZBE4	...G	...T.G	C.....TT	AT.T	.C.....						1													1			
ZBE5	...G	...T.G	C.....TT	AT.T	.C.....						1													1			
ZBE6	...AC	...C.TC	C.....G	TT A.T	.A...TCA						1													5			
BOT2	...AC	...C.TC	C.....TT	AT.T.A	.TCA										4									7			
BOT4	NNN																							2			
BOT6	.G.AC	.C.C.C	C.....TT	AT.T.A	.N.T.CA																			1			
BOT9	...AC	...C.TC	C.....G	TT A.GT	.A...T.A																			2			
BOT15	...AC	...C.TC	C.....TT	A.T	.A...T.CA																			1			
BOT16	...AC	...C.TC	C.....TT	AT.T.A	.T.CA																			1			
BOT21	...AC	...C.TC	C.....TT	AT.T.A	.CA																			1			
WC2																								1			
WC4																								5			
WC6	...AC	...T.C.TC	C.....TT	AT.T.A	.TTCA																			1			
WC13																								1			
KH2	N																							2			
KG1	...G	...T	C.....T	G.TT	AT.T																			1			
KG2	...G	...T	C.....G	TT AT.T	.C.....																			11			
WA3	...ACT.C	...C.....C	..A.....TT	A..T.CA	.T.CA																			5			
WA6	...ACT.C	...C.....C	..A.T.....TT	A..T.CA	.T.CA																			2			
WA11	...ACT.C	...C.T.....C	..A.T.....ATT	A..T.C	.CA																			1			
WA14	...ACT.C	...C.....C	..A.....TT	A..T.CA	.T.CA.T																			1			
	222222222	2222222232	2222222222	2222222222	22222222																			27			
																								9			
																								236			

Figure 2 Distribution of the 43 observed d-loop haplotypes from a sample of 236 elephants from 11 localities. The vertical numbers indicate the positions of polymorphic sites relative to haplotype KV1. A dash (-) represents a deletion introduced to optimise alignment. The bottom row of numbers below the sequences represents the number of different character states at each position.

Table 1 Summary statistics for control region variation in 11 elephant populations

	Population											
	KV	MF	Qe	MM	AM	SA	ZBE	BOT	NAM	KG	WA	Total
<i>n</i>	27	9	68	20	27	9	6	25	24	12	9	236
<i>A</i>	6	4	4	5	5	3	6	10	7	2	4	43
<i>H</i>	0.74	0.81	0.12	0.65	0.64	0.56	1.00	0.88	0.72	0.17	0.69	0.85
<i>p_s</i>	0.06	0.05	0.05	0.06	0.05	0.03	0.06	0.05	0.04	0.01	0.02	0.12
<i>π</i>	1.10	2.39	0.34	2.51	0.57	0.96	2.67	1.43	1.70	0.08	0.40	2.01

n = sample size; *A* = number of haplotypes in each population; *H* = haplotypes diversity; *p_s* = fraction of segregating sites; *π* = nucleotide diversity (in %).

localities and only 18% from the southern African localities. On the contrary, 60% of the individuals in clade B were sampled from southern Africa localities and 40% from the eastern Africa localities, mainly from Masai Mara game reserve.

Population differentiation: A hierarchical analysis of molecular variance of population structure revealed a highly significant subdivision between populations in the total sample ($F_{ST} = 0.591$, $P < 0.001$), between populations within each region ($F_{SC} = 0.420$, $P < 0.001$) and also between the eastern, southern and western regions of the continent ($F_{CT} = 0.294$, $P < 0.001$; see Table 2). All the pairwise comparisons among populations except one (BOT/ZBE) showed significant genetic differentiation

based on the F_{ST} statistic. An unrooted population tree based on F_{ST} distances cluster the populations into groups that are largely concordant with their geographical localities. For example, four of the six eastern African populations group together. The southern African populations are however divided into two groups and so are the eastern African populations (Figure 4a). A significant correlation between inter-population genetic distances and geographical distances at the continental level was observed ($g = 2.71$, $P = 0.005$). However, no correlation between genetic and geographical distances was revealed by the Mantel *g*-test at the regional level (eastern Africa; $g = -1.28$, $P = 0.91$; southern Africa; $g = 1.89$, $P = 0.08$).

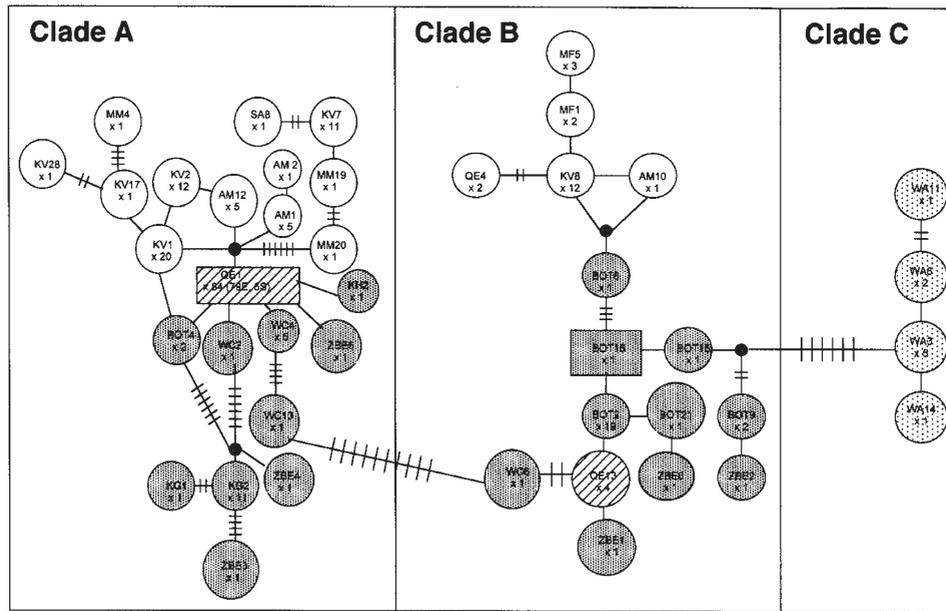


Figure 3 Minimum spanning network showing the phylogenetic relationship between the observed 43 mitochondrial control region haplotypes. Hatch marks along branches indicate the number of nucleotide differences separating each haplotype when they are more than one. Rectangles represent the inferred ancestral haplotypes. Haplotypes sampled from more than one region are represented by diagonal stripes. Haplotypes unique to each of the three regions are identified by differences in shading. The numbers represent the frequencies of each of the haplotypes.

Table 2 AMOVA results for hierarchical genetic subdivision of elephant populations

	Control region			Microsatellites		
	% of total variance	F-statistics	P-value	% of total variance	F-statistics	P-value
Among populations/ within groups	29.7	0.420	<0.001	3.4	0.047	<0.001
Among groups	29.4	0.294	<0.001	27.6	0.275	0.032
Among populations	40.9	0.591	<0.001	69.1	0.309	<0.001

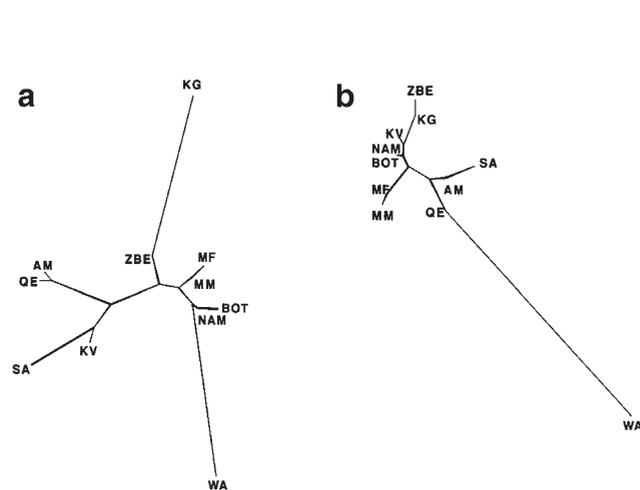


Figure 4 Distance population trees generated using F_{ST} distances based on mtDNA sequence variation (a) and R_{ST} distances based on microsatellite allele variation (b).

Microsatellite variation

Genotypic distribution and diversity: Fisher's pair-wise exact tests for genotypic disequilibria between the four loci were not significant in any of the populations ($P > 0.05$) (data not shown), implying independent assortment of alleles at these loci. The statistics describing genotypic distribution and diversity are summarised in Table 3. Observed genotypic proportions deviated significantly from Hardy-Weinberg expectations at two loci; Laf MS01 in QE, MM and NAM and Laf MS03 in KV, QE, AM and BOT. Only the QE population showed significant deviation from expected Hardy-Weinberg genotypic proportions across all loci after a Bonferroni correction (Rice, 1989).

Genetic diversity was measured in terms of number of alleles and observed and expected heterozygosities. All the loci were highly polymorphic, with the total number of alleles per locus ranging from nine (Laf MS04) to 14 (Laf MS02). A total of 44 different alleles were observed across all loci in all the populations. The number of

Table 3 Summary statistics of genetic variation at four microsatellite loci in 11 elephant populations

Locus	Population											Total
	KV	MF	QE	MM	AM	SA	ZBE	BOT	NAM	KG	WA	
Laf MS01												
<i>n</i>	22	9	55	17	17	6	6	22	21	12	16	203
<i>A</i>	6	7	8	8	8	6	4	8	7	6	4	11
H_o	0.91	1.00	0.80	0.59	0.82	1.00	0.83	0.82	1.00	1.00	0.50	
H_e	0.70	0.84	0.78	0.81	0.85	0.88	0.68	0.79	0.74	0.81	0.64	
F_{IS}	-0.31	-0.20	-0.02	0.28	0.03	-0.15	-0.25	-0.04	-0.35	0.25	0.22	
$P_{(HW)}$	0.08	0.14	0.001	0.002	0.34	0.75	0.08	0.15	<0.01	0.06	0.15	<0.01
Laf MS02												
<i>n</i>	22	9	55	16	17	6	6	22	21	12	18	204
<i>A</i>	7	6	11	6	6	3	3	6	6	4	12	14
H_o	0.91	1.00	0.87	0.86	1.00	0.67	1.00	0.91	0.71	0.67	0.94	
H_e	0.82	0.80	0.86	0.79	0.82	0.62	0.71	0.71	0.77	0.62	0.91	
F_{IS}	-0.12	-0.27	-0.02	-0.11	-0.23	-0.08	-0.46	-0.29	0.07	-0.07	0.04	
$P_{(HW)}$	0.87	1.00	0.38	0.32	0.07	1.00	0.58	0.38	0.79	0.16	0.07	0.44
Laf MS03												
<i>n</i>	22	9	52	17	17	6	6	22	20	12	17	200
<i>A</i>	5	3	6	2	4	3	3	3	2	3	9	10
H_o	0.73	0.56	0.87	0.53	0.88	1.00	1.00	0.82	0.50	0.42	0.88	
H_e	0.65	0.54	0.69	0.40	0.59	0.71	0.67	0.54	0.49	0.37	0.86	
F_{IS}	-0.12	-0.03	-0.27	-0.33	-0.51	-0.46	-0.58	-0.53	-0.02	-0.15	0.03	
$P_{(HW)}$	0.02	1.00	0.01	0.28	0.04	0.58	0.20	0.01	1.00	1.00	0.09	0.003
Laf MS04												
<i>n</i>	21	9	52	17	17	5	6	21	20	12	11	191
<i>A</i>	5	3	6	3	4	2	2	3	4	2	3	9
H_o	0.52	0.67	0.54	0.71	0.53	0.60	0.00	0.38	0.40	0.25	0.18	
H_e	0.56	0.62	0.54	0.58	0.48	0.47	0.30	0.41	0.38	0.23	0.18	
F_{IS}	0.06	-0.08	0.00	-0.23	-0.10	-0.33	1.00	0.07	-0.04	-0.10	0.03	
$P_{(HW)}$	0.87	1.00	0.83	0.32	1.00	1.00	0.09	0.28	0.49	1.00	1.00	0.96
$A_{(Total)}$	23	19	31	19	22	14	12	20	19	15	28	44
$H_e (Ave)$	0.68	0.70	0.72	0.65	0.69	0.67	0.59	0.61	0.60	0.51	0.65	
$F_{IS(Ave)}$	-0.12	-0.15	-0.08	-0.10	-0.20	-0.26	-0.30	-0.20	-0.09	-0.02	0.08	
$P_{(HW Total)}$	0.11	0.86	<0.01	0.01	0.08	1.00	0.08	0.02	0.02	0.32	0.08	<0.01

n = number of individuals genotyped; *A* = number of alleles detected in each population; $A_{(Total)}$ = total number of alleles summed over all loci; H_o = observed heterozygosity; H_e = expected heterozygosity; F_{IS} = Weir and Cockerham's (1984) analogue of Wright's fixation index; $P_{(HW)}$ = single locus *P*-values for Hardy-Weinberg tests; $P_{(HW Total)}$ = *P*-values for Hardy-Weinberg tests for each population across loci.

alleles per locus within populations ranged from two (Laf MS03, MM) to 12 (Laf MS02, WA) while the total number of alleles scored in each population at all loci ranged from 12 in ZBE to 31 in QE. Observed and expected heterozygosities ranged from 0.18 to 1.00 and 0.18 to 0.91, respectively. The average expected heterozygosities across all the loci for the eastern African populations were strikingly similar to each other with a mean of 0.68 but significantly greater than those of the southern African populations (mean = 0.58).

Population structure: A total of 14 private alleles distributed amongst four populations (QE, NAM, KG and WA) were observed across the four loci whereby eight of them were restricted to the Ghana population, which is the most divergent one. Two of the private alleles reached high frequencies in Ghana, 0.344 and 0.206 for the alleles Laf MS01 (186) and Laf MS03 (152) respectively. In addition, three alleles, Laf MS01 (198 and 200) and Laf MS04 (157) were relatively common in all the other populations but were uniquely absent from the Ghana popu-

lation. There is no clear-cut pattern that distinguishes the eastern and the southern African regions: these two regions mostly differ in allele frequencies rather than by having diagnostic alleles. This is also reflected by the fact that only 41 out of the 55 multi-locus R_{ST} estimates (Slatkin, 1995) (in contrast to the 54 out of 55 F_{ST} estimates for the mitochondrial locus) in pair-wise population comparisons were significant ($P < 0.05$). Four alleles Laf MS02 (146 and 148), and Laf MS03 (144 and 146) were ubiquitously distributed in all the populations. Interestingly, on a continental level the regions are differentiated at the same level ($R_{CT} = 0.28$) as for the variation of the control region ($F_{CT} = 0.29$) (see Table 2). Most of this differentiation is due to the diverged western African population. Unlike for the control region where an F_{SC} value of 0.42 is observed, a low differentiation is observed among populations within regions ($R_{SC} = 0.047$) for the microsatellites. On an overall level, the microsatellite loci show a differentiation among populations which is approximately half that observed for the control region ($R_{ST} = 0.31$ and $F_{ST} = 0.59$, Table 2).

A population tree generated using microsatellite loci based R_{ST} distances is shown in Figure 4b. Its topology differs from that generated using the mtDNA based F_{ST} distances in as far as the placement of the Kidepo Valley and Ghana populations is concerned: Kidepo Valley clusters with the southern Africa populations while the Ghana population clusters with the eastern African populations. Secondly, the branches connecting the populations are much shorter than in the population tree based on nucleotide differences.

A Mantel g -test revealed no significant correlation between inter-population Nei's genetic distance (Nei, 1978) calculated from allele frequency distributions and geographical distances among the eastern African populations ($g = -0.64$, $P = 0.7$) while a pattern of isolation-by-distance was evident among the southern African populations ($g = 1.87$, $P = 0.039$) and at the continental level ($g = 1.98$, $P = 0.035$).

Discussion

Mitochondrial control region

Although the occurrence of nuclear copies of the elephant mitochondrial region has been reported (Greenwood and Pääbo, 1999), we have no reason to suspect the presence of pseudogenes in our study. We found a high transition/transversion ratio (44:2), there were no double sequences on our gels, and an RT-PCR on two random samples resulted in identical sequences to the ones we originally obtained.

Variation: The average nucleotide diversity of 2.0% observed for the elephant control region is lower than what has been observed in mtDNA of most mammals examined to date, for example, Grant's gazelle (up to 6.2%, Arctander *et al*, 1996), water buck (4.1%, Simonsen, 1997) and the African buffalo (5.0%, Simonsen *et al*, 1998). However, our values of 2.0% nucleotide diversity and 85% haplotype diversity are comparable to the values of 1.8% and 87% respectively which was found in the Asian elephant by Fernando *et al* (2000). Our results also corroborate the findings of Drysdale and Florkiewicz (1989) based on allozyme studies, which revealed much lower genetic divergence between the genera, *Loxodonta* and *Elephas* compared with genera in other groups of mammals. The reduction in the rate of the molecular clock with the increase in body mass, longer generation time and lower metabolic rate has been suggested as a possible explanation for the reduced genetic diversity within the family Proboscidae (Martin *et al*, 1993). Low levels of nucleotide diversity are generally indicative of long-term small effective population sizes and could also be a result of other phenomena such as founder events and bottlenecks. Comparisons of the levels of variability of the elephant control region and the cytochrome *b* gene (which is expected to be much less variable than the control region in mammals) has revealed that the two loci have the same level of variability, suggesting that the control region as such has a reduced mutation rate in elephants (Nyakaana, unpublished). Tiedemann *et al* (1998) reported a cytochrome *b* mean sequence divergence of 2.1% from a study involving three populations from southern Africa: this figure is almost identical to

the nucleotide diversity we report in this study based on control region variation. Very low values of nucleotide diversity (0.06–0.07%) have also been observed in matrilineal whales with the same social structure as the elephants (Whitehead, 1998) whereby it was suggested that selection for matrilineally transmitted traits upon which neutral mitochondrial DNA alleles hitchhike could be a probable explanation.

Population structure: The fact that a highly significant portion of the variation (41%) is partitioned between populations clearly indicates that very little exchange of reproducing female individuals is taking place between the existing elephant populations. Even for those populations where significant genetic differentiation has not been found in pair-wise comparisons, the results need to be interpreted cautiously, given the fact that inferences about population genetic structure based on mtDNA variation tend to portray, to a large extent, the early coalescent events in the history of the populations and therefore reflect more of the ancient historical associations rather than the true contemporary patterns of gene flow (Barton and Wilson, 1986). Given the current state of fragmentation of the elephant range, our estimates of the indices of genetic differentiation may be an under estimate of the current level of genetic isolation of the existing populations.

Phylogenetic relationship between haplotypes: The minimum spanning network of the haplotypes reveals three genetically differentiated groups of the current elephant populations in Africa; one in the eastern, one in the southern, and the other in the western part of the continent. The ancestral haplotypes identified by the minimum spanning tree to be associated with these groups are QE1 and BOT16. Although no ancestral haplotype was sampled in western Africa, the population from this region is strongly subdivided from the rest of the populations. Our findings corroborate earlier studies on large mammals co-distributed with the savannah elephant whereby it has been argued that centres in these regions acted as refugia for a number of species during the drastic Pleistocene climatic changes (eg, Arctander *et al*, 1999; Flagstad *et al*, 2001). Isolation of populations into disjointed refugia has the effect of accentuating genetic divergence through genetic drift especially if the effective population size is small. The hierarchical analysis of molecular variance between continental regions clearly shows this significant genetic divergence between regions.

The co-occurrence of widely distributed haplotypes in the southern and eastern populations, in spite of the large sequence divergence separating the clades, can either be indicative of large long-term effective population sizes or it could be a result of recent secondary population admixture as a result of range expansion from refugia in the eastern and southern parts of the continent, where the populations have evolved in allopatry for a long time. It can be envisaged that populations restricted to refugia during unfavourable climatic cycles, would rapidly expand their range when conditions improved, thus favouring sympatric existence of highly divergent genotypes in areas where such expanding populations meet. A relatively recent expansion of the savannah elephant has also been suggested by Roca *et al* (2001) based on a

sequence of 1732 nucleotide base pairs from four nuclear genes. A comparison of the forest and savannah elephant showed a significantly lower level of variation in the savannah elephant (30-fold lower). This led them to suggest that this was due to a range expansion of the savannah elephant at the end of the Pleistocene when *Elephas iolensis* became extinct in Africa. In our study, this is clearly depicted in the star phylogeny of the southern African haplotypes emanating from the ancestral haplotype QE1 in clade A and the subsequent wide occurrence of this haplotype in populations both in eastern and southern Africa. The reciprocal monophyly of the western African clade is indicative of long-term absence of gene flow between this group and the rest of the regions. This finding is based on a relatively small sample from western Africa and might have to be revised when more individuals and/or locations are sampled. We do not favour the explanation of large long-term effective population sizes of females being responsible for incomplete lineage sorting as advanced by Georgiadis *et al* (1994). There is no evidence to indicate that very large elephant female census population sizes have existed in Africa for a long period of time to justify incomplete lineage sorting in a large, long-lived mammal like the elephant (see Kimura and Ohta, 1969; Avise *et al*, 1988).

Microsatellite loci

In general, the four microsatellite loci accorded to Hardy-Weinberg proportions. Only one population, QE, deviated when the significance values were combined across loci. The deviation is probably not caused by an excess of heterozygotes. The locus with the highest contribution to the overall significance value (Laf MS01) had an F_{IS} value of -0.02 , which is close to zero and the average value of F_{IS} is -0.10 , which is by far not extreme in comparison with other populations (see Table 3). We attribute this deviation to the breakdown of social structure in this population in response to the severe poaching this population has undergone (Nyakaana *et al*, 2001).

The microsatellite data depicted weaker population differentiation between populations in some of the pairwise comparisons both within and between regions. However, our data set was able to detect overall population subdivision between populations in the three regions sampled. Lack of strict concordance between nuclear and mitochondrial based genetic structures have frequently been observed in a number of studies (eg, Ferguson *et al*, 1991; Palumbi and Baker, 1994; Simonsen *et al*, 1998; Nyakaana and Arctander, 1999). Several explanations have been put forward to account for such observations. For example, the maternal inheritance of mtDNA loci increases their sensitivity to detect population structure. Such a mode of inheritance decreases its effective population size four times relative to that of the nuclear markers, thereby leading to faster genetic drift for the haploid and maternally inherited alleles (Wilson *et al*, 1985).

In addition, the matrilineal elephant social structure is characterised by female natal philopatry favouring male-biased gene flow (see Nyakaana and Arctander, 1999), which is usually manifested as a reduction in the level of genetic differentiation at nuclear loci. Depending on the extent of separation between populations, homoplasy resulting from saturation of mutations at some of the loci

can also lead to an underestimation of the extent of genetic differentiation between populations. The number of loci used in this study was relatively small; hence low subdivision indices may also be a reflection of the limited ability of our nuclear markers to detect fine scale differences between closely related populations. Future work involving use of additional recently isolated microsatellite loci (eg, Eggert, 2000) might greatly improve the resolution of the extent of genetic differentiation of these populations.

Conclusions

Our results from both mtDNA and nuclear loci clearly show significant levels of genetic subdivision between the elephant populations analysed in this study. In spite of the limited sampling carried out in the western African region, our results nevertheless clearly indicate that the savannah elephants have regionally differentiated groups that deserve to be protected as such (see Table 2). It is important that if an integrated conservation and management plan is to be designed, this observed significant genetic isolation between regional elephant groups should be taken into account, especially when decisions regarding reintroduction and/or translocations are to be considered. This would ensure the preservation of the genetic diversity and distinctiveness present in these populations of elephants.

Acknowledgements

This work was funded by the Danish International Development Agency (DANIDA) through the Wildlife Genetics Project. We thank the management of Uganda Wildlife authority and the chief wardens of Queen Elizabeth, Murchison Falls and Kidepo Valley national parks who granted us permission to conduct research in the Uganda national parks. Samples from outside Uganda were kindly provided through collaboration with the relevant bodies in the respective countries. We are greatly indebted to them.

References

- Amos W, Hoelzel AR (1991). Long term preservation of whale skin for DNA analysis. In: Hoelzel AR (ed) *Genetic Ecology of Whales and Dolphins*. Report of the International Whaling Commission, (Special Issue, 13). pp 99–104.
- Arctander P, Kat PW, Siegismund HR (1996). Extreme genetic differences among populations of Grant's gazelle, *Gazella granti*, in Kenya. *Heredity* 76: 465–475.
- Arctander P, Johansen C, Coutellec-Vreto M (1999). Phylogeography of three closely related African bovids (tribe Alcelaphini). *Mol Biol Evol* 16: 1724–1739.
- Avise JC, Ball RM, Arnold J (1988). Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. *Mol Biol Evol* 5: 331–344.
- Barnes RFW, Craig GC, Dublin HT, Simons W, Thouless CR (1999). *African Elephant Database 1998*. Occasional paper of the IUCN species survival commission No. 22. IUCN, Gland, Switzerland.
- Barriel V, Thuet E, Tassy P (1999). Molecular phylogeny of Elephantidae. Extreme divergence of the extant forest African elephant. *CR Acad Sci Paris Sciences de la vie* 322: 447–454.

- Barton NH, Wilson T (1986). Genealogies and geography. In: Harvey PH, Leigh Brown AJ, Maynard-Smith J, Nee S (eds) *New Uses for New Phylogenies*, Oxford University Press, New York. pp 23–56.
- Clement M, Posada D, Crandall KA (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657–1660.
- Drysdale T, Florkiewicz RF (1989). Electrophoretic variation within and between the two extant elephant species (Mammalia: Proboscidae). *J Mammal* **70**: 381–383.
- Eggert LS, Ramakrishnan U, Mundy NI, Woodruff DS (2000). Polymorphic microsatellite DNA markers in the African elephant (*Loxodonta africana*) and their use in the Asian elephant (*Elephas maximus*). *Mol Ecol* **9**: 2223–2225.
- Felsenstein J (1993). *PHYLIP (Phylogeny Inference Package) 3.5 edn* Department of Genetics, SK-50, University of Washington, Seattle, Washington.
- Ferguson MM, Danzmann RG, Hutchings JA (1991). Incongruent estimates of population differentiation among brook charr *Salvelinus fontinalis* from Capen Race, Newfoundland, Canada based upon allozyme and mitochondrial DNA variation. *J Fish Biol* **39**. (Suppl A): 79–85.
- Fernando P, Pfrender ME, Encalada SE, Lande R (2000). Mitochondrial DNA variation, phylogeography and population structure of the Asian elephant. *Heredity* **84**: 362–372.
- Flagstad O, Syvertsen PO, Stenseth NC, Jakobsen KS (2001). Environmental change and rates of evolution: the phylogeographic pattern within the hartebeest complex as related to climatic variation. *Proc R Soc Lond B* **268**: 667–677.
- Georgiadis N, Bischof L, Templeton A, Patton J, Karesh W, Western D (1994). Structure and history of African elephant populations: I eastern and southern Africa. *J Hered* **85**: 100–104.
- Gilbert D (1993). *SeqApp 1.9*. Biocomputing office, Biology Department, Indiana University, Bloomington.
- Greenwood AD, Pääbo S (1999). Nuclear insertion sequences of mitochondrial DNA predominate in hair but not in blood of elephants. *Mol Ecol* **8**: 133–137.
- Gyllensten UB, Erlich HA (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc Natl Acad Sci USA* **85**: 7652–7656.
- Kimura M, Ohta T (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**: 763–771.
- Mantel NA (1967). The detection of disease clustering and a generalised regression approach. *Cancer Res* **27**: 209–220.
- Martin AP, Palumbi SR (1993). Body size, metabolic rate, generation time and the molecular clock. *Proc Natl Acad Sci USA* **90**: 4087–4091.
- Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.
- Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.
- Nyakaana S, Arctander P (1998). Isolation and characterisation of microsatellite loci in the African elephant (*Loxodonta africana*). *Mol Ecol* **7**: 1436–1437.
- 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.
- Nyakaana S, Abe EL, Arctander P, Siegismund HR (2001). DNA evidence for elephant social behaviour breakdown in Queen Elizabeth National Park, Uganda. *Anim Cons* **4**: 231–237.
- Palumbi SR, Baker CS (1994). Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Mol Biol Evol* **11**: 426–435.
- Raymond M, Rousset F (1995). An exact test for population differentiation. *Evolution* **49**: 1280–1283.
- Raymond M, Rousset F (1997). GENEPOP version 3.1b: An updated version of GENEPOP (version 1.2), (originally published as) Population genetics software for exact tests and ecumenism. *J Hered* **86**: 248–249.
- Rice WR (1989). Analysing tables of statistical tests. *Evolution* **43**: 223–225.
- Roca AL, Georgiadis N, Pecon-Slattery J, O'Brien SJ (2001). Genetic evidence for two species of Elephant in Africa. *Science* **293**: 1473–1477.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT *et al* (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487–491.
- Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: New York. Second edition.
- Sanger F, Nicklen S., Coulson AR (1977). DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463–5467.
- Schneider S, Roessli D, Excoffier L (2000). *ARLEQUIN: A Software for Population Genetics Data Analysis, version 2.000*. Genetics and Biometry Laboratory, University of Geneva.
- Siegismund HR, Arctander P (1995). Structure of African elephant populations. *J Hered* **86**: 467–469.
- Simonsen BT (1997). Population Structure and History of African Bovids. PhD Thesis. Department of Population Biology, University of Copenhagen, Denmark.
- Simonsen BT, Siegismund HR, Arctander P (1998). Population structure of African Buffalo inferred from mtDNA sequences and microsatellite loci: high variation but low differentiation. *Mol Ecol* **7**: 225–237.
- Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.
- Tiedemann R, Kurt F, Haberhauer A, Hartl GB (1998). Differences in spatial genetic population between African and Asian elephant (*Loxodonta africana*, *Elephas maximus*) as revealed by sequence analysis of the mitochondrial cyt b gene. *Act Theriol Suppl* **5**: 123–134.
- Whitehead H (1998). Cultural selection and genetic diversity in matrilineal whales. *Science* **282**: 1708–1711.
- Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM *et al* (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol J Linn Soc* **26**: 375–400.