

# Mitochondrial DNA variation of the common hippopotamus: evidence for a recent population expansion

JBA Okello<sup>1</sup>, S Nyakaana<sup>1</sup>, C Masembe<sup>1</sup>, HR Siegismund<sup>2</sup> and P Arctander<sup>2</sup>

<sup>1</sup>Molecular Biology Laboratory, Makerere University Institute of Environment & Natural Resources, PO Box 7298, Kampala, Uganda;

<sup>2</sup>Department of Evolutionary Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

Mitochondrial DNA control region sequence variation was obtained and the population history of the common hippopotamus was inferred from 109 individuals from 13 localities covering six populations in sub-Saharan Africa. In all, 100 haplotypes were defined, of which 98 were locality specific. A relatively low overall nucleotide diversity was observed ( $\pi = 1.9\%$ ), as compared to other large mammals so far studied from the same region. Within populations, nucleotide diversity varied from 1.52% in Zambia to 1.92% in Queen Elizabeth and Masai Mara. Overall, low but significant genetic differentiation was observed in the total data set ( $F_{ST} = 0.138$ ;  $P = 0.001$ ), and at the population level, patterns of differentiation support previously suggested hippopotamus subspecies designations ( $F_{CT} = 0.103$ ;  $P = 0.015$ ). Evidence that the common hippopotamus recently expanded were

revealed by: (i) lack of clear geographical structure among haplotypes, (ii) mismatch distributions of pairwise differences ( $r = 0.0053$ ;  $P = 0.012$ ) and site-frequency spectra, (iii) Fu's neutrality statistics ( $F_S = -155.409$ ;  $P < 0.00001$ ) and (iv) Fu and Li's statistical tests ( $D^* = -3.191$ ;  $P < 0.01$ ,  $F^* = -2.668$ ;  $P = 0.01$ ). Mismatch distributions, site-frequency spectra and neutrality statistics performed at subspecies level also supported expansion of *Hippopotamus amphibius* across Africa. We interpret observed common hippopotamus population history in terms of Pleistocene drainage overflow and suggest recognising the three subspecies that were sampled in this study as separate management units in future conservation planning.

*Heredity* (2005) **95**, 206–215. doi:10.1038/sj.hdy.6800711; published online 20 July 2005

**Keywords:** mitochondrial DNA; common hippopotamus; control region; variation; population history; recent expansion

## Introduction

The common hippopotamus (*Hippopotamus amphibius*), commonly referred to as the hippo, is a semiaquatic artiodactyl found in sub-Saharan Africa (Kingdon, 1989; Eltringham, 1999). Hippos have become much more restricted in their distribution over recent decades as a result of rampant hunting by humans for their meat and ivory and continued harassment by farmers as agricultural pest. Only about 157 000 individuals now occur, in fragmented populations in rivers and lakes within some protected areas of eastern, western and southern Africa (Eltringham, 1993). Five subspecies of hippos have been proposed (Lydekker, 1915; Grubb, 1993), but this classification has not been generally accepted (Eltringham, 1999). Above the species level, recent molecular data have supported a sister-group relationship between hippopotomids and cetaceans (Ursing and Arnason, 1998; Gatesy *et al*, 1999; Matthee *et al*, 2001; Amrine-Madsen *et al*, 2003).

The earliest origin of hippopotamuses remains unclear (Boisserie, 2005), although the first fossil Hippopotamidae

appeared in the East African Lower Miocene (Kingdon, 1989). The genus *Hippopotamus* itself appeared about 2.5 million years ago (Turner and Wood, 1993) and was established by the Plio-Pleistocene, evidenced by their fossil records (Coryndon, 1970, 1973). This was a period of major climatic changes and regional uplift that influenced large-mammalian evolution in Africa (deMenocal, 1995; Patridge *et al*, 1995).

Despite the hippo's popularity, little is known about its general biology, dispersal phase of its life history and its population genetics (Eltringham, 1999). This has partly been due to the aquatic and nocturnal habits, and difficulty in obtaining samples. Hippos are territorial in water, where males hold a linear territory consisting of a shoreline and a narrow strip of the bank that is heavily defended against challenging bulls (Klingel, 1991). The territorial male has an absolute mating right to the females attracted to his territory, and this depends on the male's quality and the topographical features of his territory (Eltringham, 1999). The average age at puberty of the common hippopotamus is 7.5 and 9.5 years for males and females, respectively (Laws and Clough, 1966). Since there is currently an interconnected network of water system in Africa and hippos rely heavily on water in their movement and habitation, a relatively homogenised genetic pattern among hippopotamus populations is expected, albeit with quantifiable historical signatures.

Mitochondrial DNA sequence variation has been used intensively in phylogeographic studies, where the

Correspondence: JBA Okello, Molecular Biology Laboratory, Makerere University Institute of Environment & Natural Resources, PO Box 7298, Kampala, Uganda. E-mails: jbaokello@muienr.mak.ac.ug, okellojba@hotmail.com

Received 23 April 2005; accepted 7 June 2005; published online 20 July 2005

distribution of its haplotypes across the range of a species can infer its population history and structure. Phylogeographical studies on many mammals in Africa have revealed that their genetic structures were influenced by geological and climatic events of the Pleistocene, which resulted in repeated isolation of populations into refugia, and subsequent expansion of the surviving individuals when environmental conditions became favourable (Arctander *et al*, 1999; Flagstad *et al*, 2001; Nersting and Arctander, 2001; Muwanika *et al*, 2003a). These climatic events of the Pleistocene are also expected to have influenced the genetic history of the common hippopotamus, in terms of population reductions and subsequent expansion when climate became favourable.

In this study, we present the first population data on hippos in Africa based on patterns of mitochondrial control region sequence variation and infer the population processes that have led to the current genetic pattern in the species. The above information is necessary in formulating an effective management strategy for this highly fragmented but intrinsically charismatic species.

## Methods

### Distribution and sampling

The former range of the hippopotamus is shown in Figure 1, together with sampling localities. The demarcation of the subspecies is not clear and based on national boundaries. *Hippopotamus amphibius amphibius*: the nominate race is said to occur in southern Sudan, Ethiopia and



**Figure 1** Schematic map showing 13 sampling localities of hippopotamus in eastern and southern Africa. Sampling sites, population and subspecies groupings are: 1, Murchison Falls National Park ( $n=28$ ) of MFN population; 2, Queen Elizabeth National Park ( $n=25$ ) of QEN population; 3, Ugalla ( $n=8$ ); 4, Moyowosi ( $n=1$ ); 5, Luganzo ( $n=1$ ); 6, Kigosi ( $n=5$ ); 7, Mlele ( $n=1$ ); 8, Niensi ( $n=1$ ); and 9, Selous ( $n=1$ ) of Tanzania (TNZ) population, and above populations constitute *H. a. amphibius* subspecies. 10, Masai Mara Game Reserve ( $n=25$ ) of MAM population and 11, Naivasha National Park ( $n=9$ ) of NAV population, cover *H. a. kiboko* subspecies. 12, Luangwa National Park ( $n=4$ ) and 13, Kafue National Park ( $n=2$ ) of Zambia (ZAM) population, represent *H. a. capensis* subspecies. The current distribution is not mapped, and the shading depicts the former range across Africa, according to Kingdon (1997).

northern Democratic Republic of Congo, extending West to Gambia and further south to Tanzania and Mozambique. *Hippopotamus amphibius tshadensis*: populations of this subspecies occur in Chad and Niger. *Hippopotamus amphibius kiboko*: occur in Kenya and Somalia. *Hippopotamus amphibius capensis*: distributed from Zambia south to South Africa. *Hippopotamus amphibius constrictus*: occur in Angola, southern Democratic Republic of Congo and Namibia.

Skin biopsy samples were collected from hippos using a modified remote biopsy sampling technique of Karesh *et al* (1987). A total of 109 samples were collected from hippopotamuses in national parks and game reserves in eastern and southern Africa covering three of the proposed subspecies above (details in Figure 1). Samples were preserved in 25% dimethylsulphoxide saturated with sodium chloride (Amos and Hoelzel, 1991), stored at ambient temperature in the field and at  $-80^{\circ}\text{C}$  in the laboratory.

### DNA extraction, mtDNA amplification and sequencing

Total genomic DNA was extracted from samples using DNeasy<sup>®</sup> Tissue Kit (Qiagen) according to the manufacturer's instructions. Approximately 530 bp of the 5'-end segment of the mitochondrial DNA control region was PCR amplified using primers *Ham-1* (5'-CCACCATCAG CACCCAAA-3') and *Ham-2* (5'-GAGATGTCTTATT TAAGGGGAA-3'). One primer was 5' biotinylated for each strand, and both strands were amplified and sequenced. Amplifications were carried out in 50  $\mu\text{l}$  reaction volumes containing 2–5 ng of genomic DNA, 50 mM dNTPs, 20 pmol of each of the primers, 1 U of *Taq* DNA polymerase (Boehringer Mannheim GmbH) and 50 mM  $\text{MgCl}_2$  (Boehringer Mannheim GmbH). The following cycling parameters were used: one cycle of initial denaturation at  $94^{\circ}\text{C}$  for 5 min, followed by 32 cycles of denaturation at  $94^{\circ}\text{C}$  for 1 min, annealing at  $46^{\circ}\text{C}$  for 1 min and extension at  $72^{\circ}\text{C}$  for 2 min. The double-stranded PCR products were cleaned using QIAquick<sup>™</sup> PCR Purification Kit (Qiagen) and separated into single strands using M-280 streptavidin-coated paramagnetic beads (Dyna<sup>™</sup>). The single-stranded DNA template was directly sequenced by the dideoxy chain-termination method (Sanger *et al*, 1977) using Sequenase<sup>™</sup> 2.0 Kit (United States Biochemical),  $\alpha^{35}\text{S}$ -dATP (Amersham) and primers complimentary to the template. The sequencing reaction products were electrophoresed in 6% polyacrylamide/7M urea gel, fixed, dried and visualised after autoradiography.

### Genetic diversity and structure

Sequences were aligned by eye in the program *BIOEDIT* 5.0.9 (Hall, 1999). The nucleotide diversity,  $\pi$  (Nei, 1987; 10.5) for the total data, each individual population and the number of net nucleotide substitutions per site among populations,  $D_a$  (Nei, 1987; 10.21), were estimated using the program *DnaSP* 4.0 (Rozas *et al*, 2003). Population pairwise net genetic distances based on Slatkin's linearised  $F_{ST}$  (Slatkin, 1995), and hierarchical analysis of molecular variance (AMOVA) (Excoffier *et al*, 1992) were estimated using the computer program *ARLEQUIN* 2.000 (Schneider *et al*, 2000) and their significance evaluated based on 1023 random permutations. A correlation between genetic and interpopulation

land distances was tested with a Mantel test (Mantel, 1967; Manly, 1985) with significance based on 1000 random permutations. Only distances across land could be used in this test because of no direct river systems connecting the hippopotamus populations sampled.

The neighbour-joining relationship among populations based on net nucleotide substitutions per site,  $D_a$ , was estimated using *NEIGHBOUR* algorithm in *PHYLIP* 3.5c (Felsenstein, 1993). Among haplotypes, phylogenetic relationships were estimated by the neighbour-joining method using maximum-likelihood settings and 1000 bootstrap replicates, incorporating a gamma-corrected (GTR + I + G) model of evolution (Rodriguez *et al*, 1990), using *PAUP\** 4.0b8 (Swofford, 1998). This model corrects for multiple hits and site rate heterogeneity, and was chosen by both the hierarchical likelihood ratio tests and Akaike information criterion implemented in the computer program *MRMODELTEST* 1.1b (Nylander JAA, Uppsala University, Sweden), a simplified version of *MODELTEST* 3.06 (Posada and Crandall, 1998). Intraspecific gene genealogies were inferred using statistical parsimony method (Templeton *et al*, 1992) in the software *TCS* 1.13 (Clement *et al*, 2000).

### Demographic analysis

Past demographic parameters were assessed using the distribution of pairwise sequence differences (mismatch distribution) of Rogers and Harpending (1992) and site-frequency spectra (distribution of the allelic frequency at a site) of Tajima (1989) using the program *DnaSP* 4.0 (Rozas *et al*, 2003). In both cases, the observed and expected pattern in an expanding population scenario was assessed in the total sample and within each of the three subspecies covered in this study. A recently expanded population shows a unimodal mismatch distribution without large differences in the frequency of the ranked pairwise differences (a low raggedness statistics). The smoothness of the observed distribution was quantified by the raggedness statistic,  $r$  (Harpending, 1994), and the confidence intervals were provided by computer simulations using the coalescent algorithm in *DnaSP*. In the case of site-frequency spectra, populations that have undergone expansion show characteristic features in the frequency of mutation classes (Donnelly *et al*, 2001), where a recently expanded population displays an excess of singleton mutation classes compared with a stationary population (Ewens, 1972).

It has been argued that methods based on mismatch distributions use little of the information accumulated in the data (Felsenstein, 1992), and they have been shown to

be of low statistical power in detecting population expansion (Ramos-Onsins and Rozas, 2002). We used Fu's  $F_s$  statistics (Fu, 1997) to substantiate inference of population growth and range expansion as revealed by mismatch and site-frequency distributions, and test for the presence of genetic hitchhiking – which produces a similar pattern. This powerful population expansion test takes into account haplotype frequencies under neutrality, stationarity and panmixis, as derived by Ewens (1972), but is sensitive to background selection. Fu and Li (1993)'s  $D^*$  and  $F^*$  tests were, therefore, used to detect the presence of background selection in the whole hippopotamus sample analysed. Coalescence simulations based on a neutral infinite-sites model and assuming a large constant population size (Hudson, 1990) were applied to determine the confidence intervals of  $F_s$ ,  $D^*$  and  $F^*$  statistics observed.

## Results

### Genetic variation

A 400 bp segment of the 5'-end control region was bidirectionally sequenced for 109 individuals. A total of 55 segregating sites were observed, defining 100 haplotypes and sequences of these haplotypes were submitted to GenBank (accession numbers DQ103911–DQ104010). Haplotypes MFN5, MFN14 and MAM16 occurred in more than one population, while haplotypes MFN8, QEN2, TNZ1 and TNZ9 occurred more than once but were locality specific.

Preliminary analyses revealed high levels of homogeneity among samples obtained from localities in Tanzania and Zambia. Consequently, the samples were pooled in each case and subsequently designated as TNZ and ZAM populations, respectively (Figure 1). The highest fraction of segregating sites was found in MAM population (7.7%), while the least was in ZAM population (4.0%). Nucleotide diversity,  $\pi$ , in the total sample was 1.94%. For each individual population,  $\pi$  varied from 1.52% in ZAM to 1.92% in QEN and MAM (Table 1).

### Population structure

Overall, significant but low population subdivisions were observed within the total sample ( $F_{ST}=0.138$ ;  $P=0.001$ ) and among designated subspecies ( $F_{CT}=0.103$ ;  $P=0.015$ ). Comparisons of pairs of population samples based on Slatkin's linearised  $F_{ST}$  revealed significant genetic differentiation in 10 out of 15 pairwise population comparisons, with values ranging from 2.1%

**Table 1** Slatkin's  $F_{ST}$  values (Slatkin, 1995) from pairwise population comparisons (above the diagonal) and the number of net nucleotide substitutions per site among populations,  $D_a$  (Nei, 1987; 10.21) (below the diagonal)

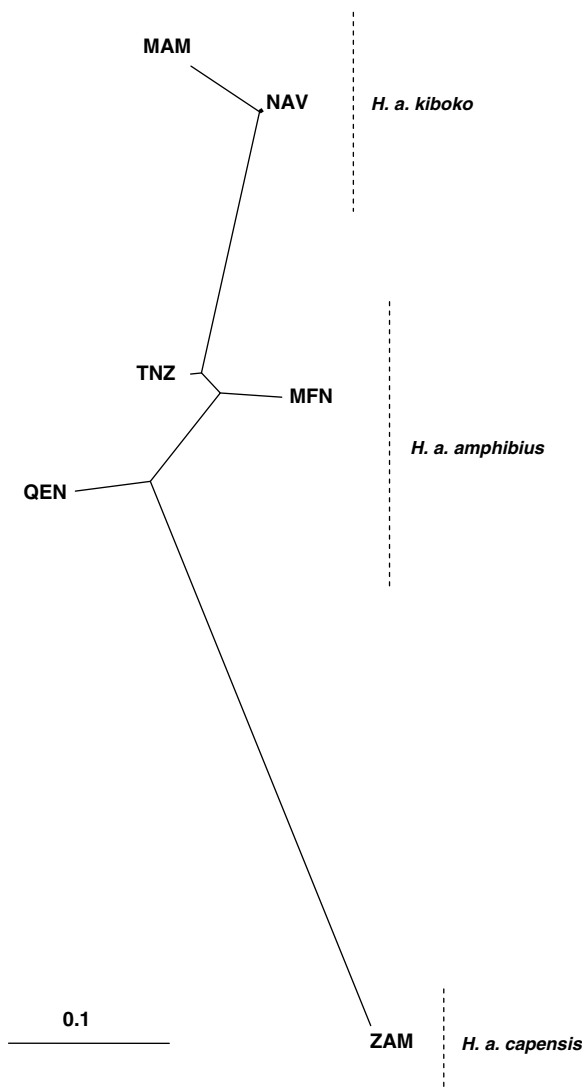
	MFN	QEN	TNZ	MAM	NAV	ZAM
MFN	<b>1.71</b>	0.064 <sup>a</sup>	0.030	0.120 <sup>a</sup>	0.131 <sup>a</sup>	0.205 <sup>a</sup>
QEN	0.120	<b>1.92</b>	0.025	0.102 <sup>a</sup>	0.064	0.148 <sup>a</sup>
TNZ	0.054	0.047	<b>1.70</b>	0.115 <sup>a</sup>	0.081	0.207 <sup>a</sup>
MAM	0.244	0.216	0.236	<b>1.92</b>	0.021	0.295 <sup>a</sup>
NAV	0.255	0.134	0.148	0.050	<b>1.65</b>	0.309 <sup>a</sup>
ZAM	0.614	0.512	0.626	0.957	0.938	<b>1.52</b>

Percentage nucleotide diversity for each individual population,  $\pi$  (Nei, 1987; 10.5) is shown in bold along the diagonal.

<sup>a</sup>Denotes significant  $F_{ST}$  values at 5% significance level.

between MAM and NAV to 30.9% between NAV and ZAM (see Table 1). A significant correlation between net interpopulation genetic distances,  $D_a$  (Nei, 1987; 10.21), and distances across land was observed ( $g=1.61$ ;  $P=0.05$ ).

An unrooted neighbour-joining population tree based on net genetic distances shows relationships among the six hippopotamus populations sampled (Figure 2). Populations belonging to *H. a. kiboko* and *H. a. capensis* subspecies are portrayed as distinct groups radiating from *H. a. amphibius*, the nominate subspecies. The neighbour-joining phylogenetic tree (Figure 3), however, reveals no major geographical partitioning of haplotypes, except for a few instances where haplotypes from the same locality clustered together, for example, MFN and MAM haplotypes. Most of the branches in this tree could not be statistically resolved with more than 50% bootstrap support, due to lack of deep structuring of haplotypes. The statistical parsimony network (Figure 4) shows no clear geographical structuring, except for small



**Figure 2** Neighbour-joining population relationships based on Nei's (1987) net genetic distances among the six hippopotamus populations.

star-like clusters of haplotypes from MAM, MFN and NAV. Despite lack of clear distinct lineages, the most ancestral haplotypes are inferred to be from MFN population of *H. a. amphibius* subspecies.

### Past population expansion

The distribution of pairwise nucleotide differences for the total hippopotamus sample, *H. a. amphibius* and *H. a. kiboko* subspecies, revealed generally smooth unimodal patterns (Figure 5), characteristic of a population that has undergone a large expansion. The statistical fit of observed pairwise distributions as quantified by raggedness tests and simulations based on coalescent estimates was significant in all the cases, except for subspecies *H. a. capensis*, which shows a highly ragged mismatch distribution (Figure 5). With the exception of *H. a. capensis*, our data support recent population growth and range expansion in the hippos. The distributions of mutation classes revealed by site-frequency spectra of mitochondrial DNA control region analysed in this study show an excess of singleton mutations relative to the frequency expected under stationarity and neutrality (Figure 6). The large differences between the observed and expected frequencies of mutant classes in both total sample and the individual subspecies are consistent with a recent population growth.

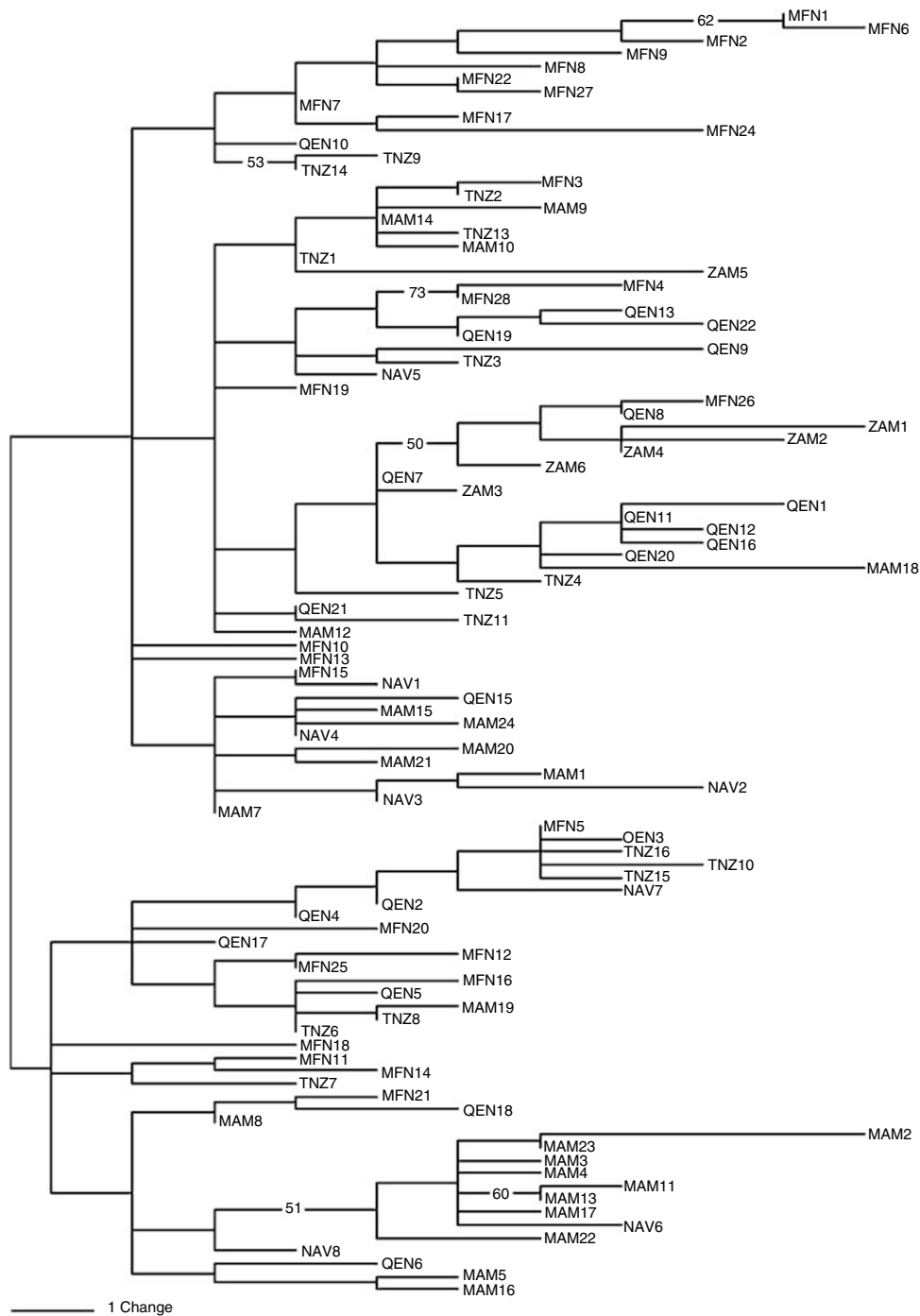
Fu's  $F_s$  statistic is more powerful for detecting deviations from neutrality and thereby testing population expansion and genetic hitchhiking in the data which produces a similar pattern like that of recent population expansion. Genetic hitchhiking was ruled out by Fu's  $F_s$ , all the values were negative, large and highly significant except for *H. a. capensis* (Figure 6). The presence of background selection in the total DNA sample analysed was also rejected by Fu and Li's tests ( $D^*=-3.191$ ,  $P<0.01$ ;  $F^*=-2.668$ ;  $P=0.01$ ). In concert, the above parameters provide support for a recent population expansion of the common hippopotamus in Africa.

## Discussion

### Genetic variability and differentiation

The common hippopotamus has a relatively low variation at the mitochondrial DNA control region, low but significant differentiation among most populations, and show evidence of a recent population expansion. Overall nucleotide diversity of 1.9% observed in this study is relatively low on comparison with diversity indices of other large African mammals from the same region (Table 2). Observed nucleotide diversity in hippos is only comparable to that in elephants (2.0%), the latter were drastically reduced in numbers in the last few decades due to poaching for ivory (Muwanika *et al*, 2003b). Although there has been no recorded bottleneck in the history of the common hippopotamus in Africa, the low sequence variability and nucleotide diversity observed in hippopotamus may be due to the relatively short coalescence time since the population expansion event.

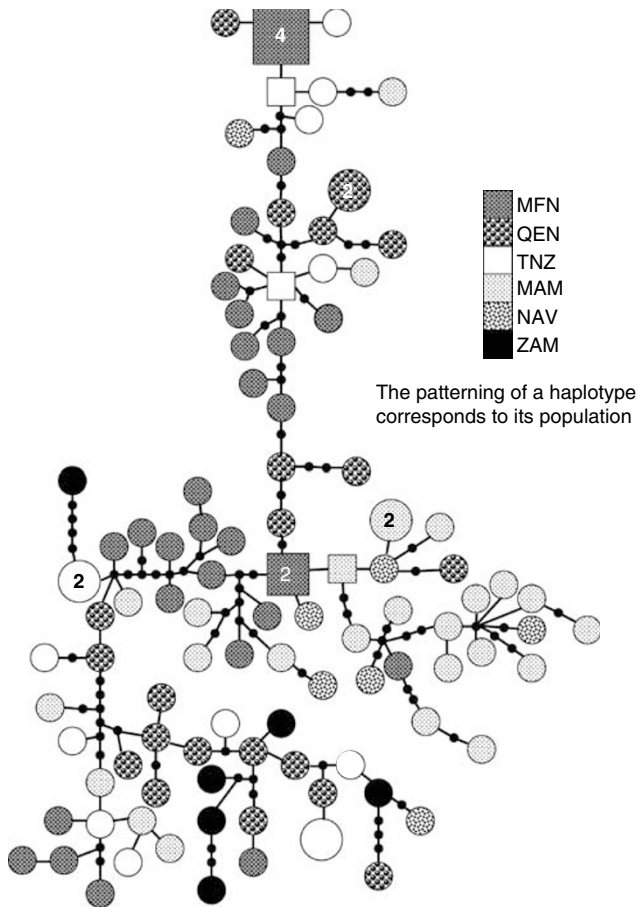
Hippos are heavily dependent on rivers, and since the river systems in Africa are to a large extent interconnected, one would expect to find little or no genetic differentiation among these populations. However, this study shows a low but significant overall genetic differentiation ( $P=0.001$ ) among most populations



**Figure 3** A midpoint rooted neighbor-joining phylogenetic tree based on maximum-likelihood distance settings (GTR+G+I model of evolution) and 1000 bootstrap replicates among mtDNA haplotypes of *Hippopotamus amphibius*. Zero branches were collapsed creating polytomies and statistical support for nodes with  $\geq 50\%$  bootstrap replicates are indicated.

compared (Table 1). Even though the pairwise  $F_{ST}$  values among populations are somewhat low, the relationships based on net interpopulation genetic distances (Figure 2) portray *H. a. kiboko* and *H. a. capensis* as distinct subspecies that radiated from *H. a. amphibius*, the nominate subspecies. Observed subspecies differentiation is further supported by results from AMOVA ( $F_{CT} = 0.103$ ;  $P = 0.015$ ), which show that 10.34% of the

total variation is attributed to differences among subspecies. This might reflect effects of isolation by distance ( $P = 0.05$ ) and regional historical divergence that could have arisen during or after the Pleistocene population expansion. Evolution of many other large African mammals was also influenced by these climatic changes (Arctander *et al*, 1999; Flagstad *et al*, 2001; Nersting and Arctander, 2001; Muwanika *et al*, 2003a).



**Figure 4** A statistical parsimony network showing relationships between hippopotamus mtDNA haplotypes from eastern and southern Africa. Haplotypes are represented as circles and squares for ancestral ones. The minimum number of steps connecting parsimoniously two haplotypes is indicated as a thick-black square, and a filled-small circle represents an extinct or missing haplotype that might have not been sampled. The size of the square or circle and pattern assigned correspond to the haplotype frequency and population, respectively, and numbers in haplotypes correspond to the number of individual sharing the haplotypes.

### Population expansion

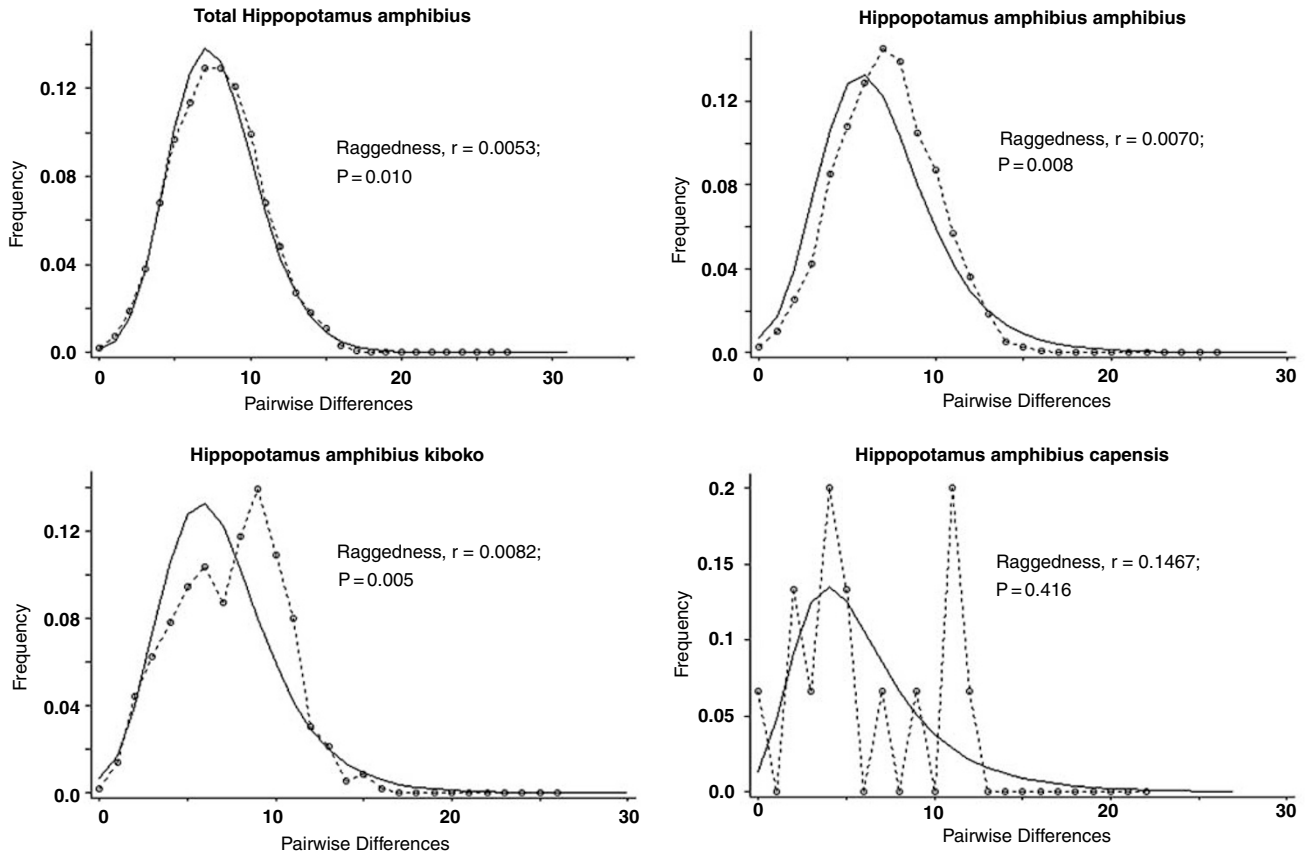
Population size changes leave particular footprints that may eventually be detected in DNA sequence data (Tajima, 1989; Slatkin and Hudson, 1991; Rogers and Harpending, 1992). The haplotypic relationships portrayed in both the neighbour-joining tree (Figure 3) and statistical parsimony network (Figure 4) showed no clear geographical structuring, except a few star-like clusters. Lack of geographical structuring among haplotypes in both the phylogenetic tree and statistical parsimony network in Figures 3 and 4, despite significant subdivision among populations, is interpreted as the signature of a young population. A relatively recent demographic event, such as a population growth, causes most of the coalescent events to occur before the expansion and, consequently, samples of these populations have gene genealogies stretched near the external nodes and compressed near the root (star genealogies). This arises because there is much less stochastic elimination of lineages in a rapidly growing population (Avise *et al*, 1984).

Further support for hippopotamus population expansion was provided by the smooth mismatch distributions of mtDNA haplotypes observed in the whole population and individual subspecies (Figure 5). Computer simulations based on coalescent process provided statistical support for the smoothness of the observed distributions (as quantified by the raggedness indices). Although the mismatch distributions of *H. a. kiboko* and *H. a. capensis* are not very smooth on initial inspection, simulation of the raggedness index confirmed that the observed distribution for *H. a. kiboko* is essentially unimodal. For *H. a. capensis*, the apparent lack of smoothness is nonsignificant, which might be attributed mainly to the small sample size as an unknown palaeohydrological scenario could, quite probably, have shaped the southern Africa region. The aquatic fossil record from this part of Africa is still incomplete (Stewart, 2001), which therefore makes it hard to clearly decipher the hydropalaeobiology of this region. In general, it is known that a smooth or bell-shaped frequency distribution of pairwise differences between haplotypes is expected in a recently expanded population because most alleles are descended from one or a few ancestral types (Rogers and Harpending, 1992). If the mean and mode of the mismatch distribution is low, the expansion occurred recently, and if high, the expansion is more ancient (Rogers, 1995). The distribution of mutation classes for the total sample and individual subspecies reveal an excess of singleton mutations relative to the frequencies expected under stationarity and neutrality (Figure 6). Given the large differences between the observed and the expected frequencies of mutant classes, these observations are consistent with recent population growth. Rapid population growth leaves an excess of recent mutations or rare alleles (Fu, 1997), and therefore is the most likely explanation for the observation.

Past population expansion signatures, detected in the effects on Fu's  $F_s$  statistic, based on gene frequency distribution of Ewens (1972), confirmed these inferences about the population dynamics of the common hippopotamus. Negative and significant  $F_s$  statistical values in the total sample and individual subspecies (Figure 6) gave strong evidence of past population expansion and ruled out genetic hitchhiking and background selection, an evolutionary force producing patterns similar to population expansion (Fu and Li's  $D^* = -3.191$ ,  $P < 0.01$ ; and  $F^* = -2.668$ ;  $P = 0.01$ ). Positive values of these statistics would have suggested that the sequences were evolving non-neutrally in a pattern consistent with background selection (Fu and Li, 1993).

### Evolutionary and conservation implications

The first hippopotamus appeared about 2.5 million years ago (Kingdon, 1989; Turner and Wood, 1993) and became established by the Pleistocene, where fossil records show their abundance (Coryndon, 1970, 1973). This period, characterised by major climatic fluctuations and tectonic movements, influenced the evolution of many large mammals in Africa (deMenocal, 1995; Patridge *et al*, 1995). The common hippopotamus might have been conscripted into a refugium somewhere in the range of the current MFN population as indicated by the presence of the most ancestral haplotypes in the statistical parsimony network (Figure 4). Although it is difficult



**Figure 5** Mismatch distributions of mitochondrial DNA sequences of the common hippopotamus based on pairwise nucleotide differences in the total population and subspecies. Solid lines in the curves indicate the expected distributions under expansion and dotted lines indicate the observed distributions under population expansion. The raggedness statistics and corresponding *P*-values are given.

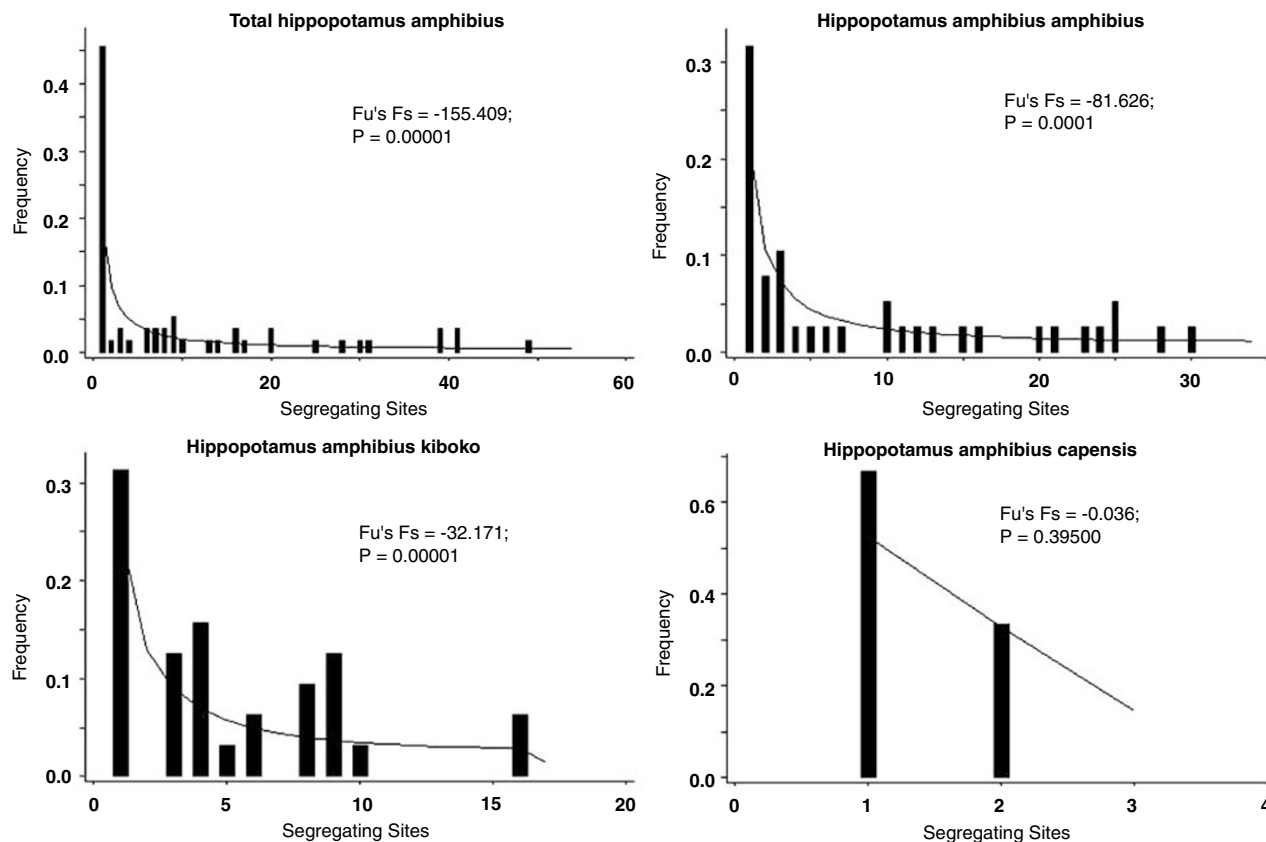
to date accurately the expansion time from mitochondrial DNA data, we hypothesise the recent hippopotamus population expansion reported here was facilitated by the Pleistocene overflow that caused an interconnected network of lake and river systems in sub-Saharan Africa (VanCouvering, 1982).

Geological evidence suggested considerable tectonic and volcanic activity occurred in middle Pleistocene, affecting climate and hydrology of this region (Stewart, 2001). Dramatic climatic changes created rain shadows in most parts of Africa, drying major water bodies such as Lake Victoria (Johnson *et al*, 2000; Beuning *et al*, 2002). Many aquatic animals that inhabited Africa by then were decimated, and a few however survived in river refugia, for example, Cichlids (Seehausen, 2002; Verheyen *et al*, 2003). Terrestrial mammals such as impala (*Aepyceros melampus*), greater kudu (*Tragelaphus strepsiceros*) and wildebeest (*Connochaetes taurinus*) recolonised their current range from southern Africa refugia (Arctander *et al*, 1999; Nersting and Arctander, 2001). The topi (*Damaliscus lunatus*), hartebeest (*Alcelaphus buselaphus*) and common warthog (*Phacochoerus africanus*) also had a southern, eastern and for the two latter, additional western regional refugia from where they expanded when environmental conditions improved (Arctander *et al*, 1999; Flagstad *et al*, 2001; Muwanika *et al*, 2003a).

Prior to the existence of modern day Lake Victoria drainage system and as cooling down from mid-Pleistocene violent climatic cycles was occurring, rivers

were flowing east to west across Uganda, and possibly into the Congo basin and the Atlantic Ocean (Cooke, 1958; deHeinzelin, 1964; Pickford *et al*, 1993). Upwarp occurred in the Western Rift Valley, reversing the east-west flowing rivers, flooding and hence creating Lake Victoria. Overflowing of Lake Victoria occurred into the Nile River System in the north (Pickford *et al*, 1993) and Malagarasi river system in the south (Stager *et al*, 1986). As in the case of Western Africa, there are many gaps in palaeohydrology for the southern part of the continent (Gasse, 2000; Holmgren *et al*, 2003) and very few reported aquatic fossil records (Stewart, 2001). Although no direct evidence suggests the southern part of Africa was affected dramatically in the same way during the Pleistocene as other parts of the continent (Stewart, 2001), the Zambezi, Okavango and Kafue rivers system is thought to have formed an ancient major southern drainage that connected the Congo system via Lake Tanganyika and the east coast rivers (Bell-Cross, 1972; Banister and Clarke, 1980). Overall, this proposed late-Pleistocene hydrology network in sub-Saharan Africa might have facilitated the rapid proliferation and wide distribution of aquatic faunas like hippopotamuses as suggested in this study.

When a population grows rapidly, genetic variation will be accumulated and maintained and in the long run will be beneficial to the success of the species (Su *et al*, 2001). Recent population expansion and considerable amount of genetic variation within and among



**Figure 6** Site-frequency spectra of the mitochondrial DNA d-loop sequences in the total hippopotamus population and the three subspecies covered in this study. Solid lines in the site-frequency spectra indicate the expected distributions under neutrality and at equilibrium. Fu's  $F_s$  statistics and corresponding  $P$ -values are given.

**Table 2** Comparisons of diversity indices of the hippopotamus with those of other large mammals studied from the same region based on nucleotide sequence of the 5'-end of the mitochondrial control region

Mammal	Bases	% Var.	( $\pi$ )	Reference
Hippo	400	12.8	1.9	This study
Warthog	370	7.0	4.0	Muwanika <i>et al</i> (2003a)
Elephants	400	8.5	2.0	Nyakaana <i>et al</i> (2002)
Impala	400	2.0	3.6	Nersting and Arctander (2001)
Kobs	354	29.1	4.9	Birungi and Arctander (2000)

fragmented hippo populations observed are encouraging for the general genetic diversity status of this threatened species. However, the positive effects of the past population expansion inferred from this study are being undermined by current demographic trends in hippopotamus populations, since their numbers are dramatically declining and populations are becoming more fragmented (Klingel, 1991; Lamprey and Michelmore, 1996; Eltringham, 1993, 1999). Discrepancy between genetic and census data is because genetic patterns from deep past events are robust to recent population changes (Rogers, 1997). The significant differentiation observed among most of our study populations indicate low levels of gene flow among these remnants, which appears to be a direct consequence of recent human-driven habitat

isolation. More conservation efforts should therefore be put on these remnant populations to protect them from further degradation. And since our findings corroborate morphological subspecific differentiation as in Lydekker (1915) and Grubb (1993), we support the recognition of three hippopotamus subspecies analysed here as separate management units with fragmented subpopulations.

## Acknowledgements

This work is part of an ongoing research program on the genetics and evolution of large African mammals (The African Wildlife Genetics Project), sponsored by the Danish International Development Agency and the Danish Natural Science Research Council. It was conducted at the Institute of Environment and Natural Resources, Makerere University, Uganda in collaboration with the Department of Evolutionary Biology, University of Copenhagen, Denmark. We thank wildlife authorities in Uganda, Kenya, Tanzania and Zambia for allowing us to conduct research in their protected areas. Special thanks go to David Moyer for providing samples from Tanzania, and to Vincent B Muwanika and Stephen Buah for their various supports.

## 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*. International Whaling Commission: Cambridge. pp 99–104.
- Amrine-Madsen H, Koep KP, Wayne C, Mark S, Springer MS (2003). A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. *Mol Phylogenet Evol* **28**: 225–240.
- Arctander P, Johansen C, Coutellic-Vreto MA (1999). Phylogeography of three closely related African bovids (Tribe Alcelaphini). *Mol Biol Evol* **16**: 1724–1739.
- Avise JC, Neigel JE, Arnold J (1984). Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J Mol Evol* **156**: 99–105.
- Banister KE, Clarke MA (1980). A revision of the large *Barbus* (Pisces, Cyprinidae) of Lake Malawi with a reconstruction of the history of the southern African Rift Valley Lakes. *J Nat Hist* **14**: 483–542.
- Bell-Cross G (1972). The fish fauna of the Zambezi River system. *Arnoldia* **29**: 1–19.
- Beuning KRM, Russell KK, Wolfe J, Brent BB (2002). Reassessment of Lake Victoria–Upper Nile River paleohydrology from oxygen isotope records of lake-sediment cellulose. *Geology* **30**: 559–562.
- Birungi J, Arctander P (2000). Large sequence divergence of mitochondrial DNA genotypes of the control region within populations of the African antelope, Kob (*Kobus kob*). *Mol Ecol* **9**: 1997–2008.
- Boisserie JR (2005). The phylogeny and taxonomy of Hippopotamidae (Mammalia: Artiodactyla): a review based on morphology and cladistic analysis. *Zool J Linn Soc* **143**: 1–26.
- Clement M, Posada D, Crandall K (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657–1660.
- Cooke HBS (1958). Observations relating to Quaternary environments in East Africa. *Trans Geo Soc S Afr* **60**: 1–74.
- Coryndon SC (1970). The extent of variation in Fossil Hippopotamus from Africa. *Symp Zool Soc Lond* **26**: 135–147.
- Coryndon SC (1973). *Fossil Hippopotamidae from Pliocene/Pleistocene Successions of the Rudolph Basin*. Paper for Conference on Stratigraphy, Paleontology and Evolution. Wenner Grenn Foundation: New York.
- deHeinzelin J (1964). Addendum. In: Howell FC, Bourlière F (eds) *African Ecology and Human Evolution*. Viking Foundation Publishers: London. pp 648–654.
- deMenocal PB (1995). Plio-Pleistocene African climate. *Science* **270**: 53–58.
- Donnelly MJ, Licht MC, Lehmann T (2001). Evidence for a recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*. *Mol Biol Evol* **18**: 1353–1364.
- Eltringham SK (1993). The common hippopotamus (*Hippopotamus amphibius*). In: Oliver WRL (ed) *Pigs, Peccaries and Hippos*. IUCN: Gland. pp 43–55.
- Eltringham SK (1999). *The Hippos*. Academic Press: London.
- Ewens WJ (1972). The sampling theory of selectively neutral alleles. *Theor Popul Biol* **3**: 87–112.
- Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.
- Felsenstein J (1992). Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genet Res* **59**: 139–147.
- Felsenstein J (1993). *PHYLIP (Phylogeny Inference Package)*, Version 3.5C. Distributed by the author. University of Washington: Seattle, USA.
- Flagstad Ø, Syvertsen OP, Stenseth CN, Jakobsen SK (2001). Environmental change and rates of evolution: the phylogeographic pattern within the hartebeest complex as related to climatic variation. *Proc R Soc Lond Ser B* **268**: 667–677.
- Fu Y-X (1997). Statistical tests of neutrality against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.
- Fu Y-X, Li W-H (1993). Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- Gasse F (2000). Hydrological changes in the African tropics since the Last Glacial Maximum. *Quatern Sci Rev* **19**: 189–211.
- Gatesy J, Milinkovitch M, Waddell V, Stanhope M (1999). Stability of cladistic relationships between Cetacea and higher-level artiodactyl taxa. *Syst Biol* **48**: 6–20.
- Grubb P (1993). The Afrotropical Hippopotamus and Hexaprotodon: taxonomy and description. In: Oliver WRL (ed) *Pigs, Peccaries and Hippos*. IUCN: Gland. pp 41–43.
- Hall TA (1999). *BIOEDIT*: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp* **41**: 95–98.
- Harpending HC (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* **66**: 591–600.
- Holmgren K, Lee-Thorp JA, Cooper GRJ, Lundblad K, Partridge TC, Scott L et al (2003). Persistent millennial-scale climatic variability over the past 25,000 years in Southern Africa. *Quatern Sci Rev* **22**: 2311–2326.
- Hudson RR (1990). Gene genealogies and the coalescent process. *Oxf Surv Evol Biol* **7**: 1–44.
- Johnson TC, Kelts K, Odada E (2000). The holocene history of Lake Victoria. *Ambio* **29**: 2–11.
- Karesh WB, Smith F, Frazer-Taylor H (1987). A remote method for obtaining skin samples. *Conserv Biol* **1**: 261–262.
- Kingdon J (1989). *East African Mammals. An Atlas of Evolution in Africa (Large Mammals)*, Vol IIIB. The University of Chicago Press: Chicago.
- Kingdon J (1997). *The Kingdon Field Guide to African Mammals*. Academic Press: London.
- Klingel H (1991). The social organization and behavior of the *Hippopotamus amphibius*. In: Kayanja FIB, Edroma EL (eds) *African Wildlife Research and Management*. International Council of Scientific Unions: Paris. pp 73–75.
- Lamprey RH, Michelmore F (1996). *The Wildlife Protected Areas of Uganda*. Ministry of Tourism, Wildlife and Antiquities: Kampala, Uganda.
- Laws RM, Clough G (1966). Observations on reproduction in the hippopotamus *Hippopotamus amphibius*. *Linn Symp Zoo Soc Lond* **15**: 117–140.
- Lydekker R (1915). *Catalogue of the Ungulate Mammals in the British Museum of Natural History*, Vol. 5. British Museum: London.
- Manly BFJ (1985). *The Statistics of Natural Selection on Animal Populations*. Chapman & Hall: London.
- Mantel N (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.
- Mathee CA, Burzla JD, Taylor JF, Davis SK (2001). Mining the mammalian genome for artiodactyl systematics. *Syst Biol* **50**: 367–390.
- Muwanika VB, Nyakaana S, Siegismund HR, Arctander P (2003a). Phylogeography and population structure of the common warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity* **91**: 361–372.
- Muwanika VB, Siegismund HR, Okello JBA, Masembe C, Arctander P, Nyakaana S (2003b). A recent bottleneck in the warthog and elephant populations of Queen Elizabeth National Park, revealed by a comparative study of four mammalian species in Uganda national parks. *Anim Conserv* **6**: 237–245.
- Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.
- Nersting LG, Arctander P (2001). Phylogeography and conservation of impala and greater kudu. *Mol Ecol* **10**: 711–719.

- Nyakaana S, Arctander P, Siegismund H (2002). Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci. *Heredity* **89**: 90–98.
- Patridge TC, Wood BA, deMenocal PB (1995). The influence of global and regional uplift on large-mammalian evolution in East and Southern Africa. In: Vrba ES, Denton GH, Patridge TC, Burckle LH (eds) *Paleoclimate and Evolution, with Emphasis on Human Origins*. Yale University Press: New Haven. pp 330–355.
- Pickford M, Senut B, Hadoto D (1993). Geology and palaeobiology of the Albertine Rift Valley Uganda-Zaire. *Publ Occ CIFEG* **24**: 1–180.
- Posada D, Crandall KA (1998). MODELTEST 3.06: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.
- Ramos-Onsins SE, Rozas J (2002). Statistical properties of the new neutrality test against population growth. *Mol Biol Evol* **19**: 2092–2100.
- Rodriguez F, Oliver JF, Marin A, Medina JR (1990). The general stochastic model of nucleotide substitutions. *J Theor Biol* **142**: 485–501.
- Rogers AR (1995). Genetic evidence for a Pleistocene population explosion. *Evolution* **49**: 608–615.
- Rogers AR (1997). Population structure and modern human origins. In: Donnelly P, Tavar S (eds) *Progress in Population Genetics and Human Evolution*. Institute for Mathematics and its Applications: University of Minnesota. pp 55–79.
- Rogers AR, Harpending H (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* **9**: 552–569.
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463–5467.
- Schneider S, Jean-Mark K, Roessli D, Excoffier L (2000). ARLEQUIN 2.000: A Software for Population Genetic Data Analysis. Genetics and Biometry Laboratory, University of Geneva: Geneva.
- Seehausen O (2002). Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14,600 yrs history for its cichlid species flock. *Proc R Soc Lond Ser B* **269**: 491–497.
- Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.
- Slatkin M, Hudson RR (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**: 555–562.
- Stager JC, Reinthal PN, Livingstone DA (1986). A 25 000-year history for Lake Victoria, East Africa, and some comments on its significance for the evolution of cichlid fishes. *Freshwater Biol* **16**: 15–19.
- Stewart KM (2001). The freshwater fish of Neogene Africa (Miocene–Pleistocene): systematics and biogeography. *Fish Fish* **2**: 177–230.
- Su B, Fu Y, Wang Y, Jin L, Chakraborty R (2001). Genetic diversity and population history in Red Panda (*Ailurus fulgens*) as inferred from mitochondrial DNA sequence variations. *Mol Biol Evol* **18**: 1070–1076.
- Swofford DL (1998). PAUP\*, Version 4.0b. *Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. Sinauer Associates: Sunderland, MA.
- Tajima F (1989). The effect of change in population size on DNA polymorphism. *Genetics* **123**: 597–601.
- Templeton AR, Crandall KA, Sing CF (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619–633.
- Turner A, Wood BA (1993). Comparative palaeontological context for the evolution of the early hominid masticatory system. *J Hum Evol* **24**: 229–236.
- Ursing BM, Arnason U (1998). Analyses of mitochondrial genomes strongly support a hippopotamus – whale clade. *Proc R Soc Lond Ser B* **265**: 2251–2255.
- VanCouvering JAH (1982). *Fossil Cichlid Fish of Africa*, Special Papers in Palaeontology, Vol. 29. Palaeontological Association: London.
- Verheyen E, Salzburger W, Snoeks J, Meyer A (2003). Origin of the superflock Cichlid fishes from Lake Victoria, East Africa. *Science* **300**: 325–329.