Received Date : 30-Jan-2014

Revised Date : 11-Mar-2015

Accepted Date : 18-Mar-2015

Article type : Original Article

Genetic consequences of population expansions and contractions in the common hippopotamus (*Hippopotamus amphibius)* since the Late Pleistocene

Céline Stoffel[1,*], Christophe Dufresnes[1,*], John BA Okello[2,3], Christian Noirard[4,5], Pierre Joly[4], Silvester Nyakaana[6], Vincent B Muwanika[7], Nicolas Alcala[1], Séverine Vuilleumier[1], Hans R Siegismund[8] and Luca Fumagalli[1]

* equal contribution

[1] *Department of Ecology and Evolution, Laboratory for Conservation Biology, Biophore Building, University of Lausanne, 1015 Lausanne, Switzerland;*

[2] *Molecular Genetics Laboratory, Makerere University Institute of Environment and Natural Resources (formerly), College of Agricultural and Environmental Sciences, Makerere University, P.O. Box 7298, Kampala, Uganda;*

[3] *Department of Pathology & Molecular Medicine, and Division of Cancer Biology & Genetics, Cancer Research Institute, Queen's University Kingston, ON K7L 3N6, Canada;*

[4] *UMR5023, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, Université Claude Bernard Lyon 1, Bât. Darwin C, 69622 Villeurbanne, France;*

[5] *COECO, 16 rue Pître Chevalier, 44000 Nantes, France;*

[6] *Department of Biological Sciences, College of Natural Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda;*

[7] *Department of Environmental Management, College of Agricultural and Environmental Sciences, Makerere University, P. O. Box 7062/7298 Kampala, Uganda;*

[8] *Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark.*

Correspondence address: Luca Fumagalli, Department of Ecology and Evolution, Laboratory for Conservation Biology, Biophore Building, University of Lausanne, 1015 Lausanne, Switzerland;

Fax: +41 21 6924165;

E-mail: luca.fumagalli@unil.ch

Running title: Phylogeography of the Common Hippopotamus

Keywords: climate change, conservation genetics, phylogeography, population genetics, Sub-Saharan Africa

Abstract

Over the past two decades, an increasing amount of phylogeographic work has substantially improved our understanding of African biogeography, in particular the role played by Pleistocene pluvial-drought cycles on terrestrial vertebrates. However, still little is known on the evolutionary history of semi-aquatic animals, which faced tremendous challenges imposed by unpredictable availability of water resources. In this study, we investigate the Late Pleistocene history of the common hippopotamus (*Hippopotamus amphibius*), using mitochondrial and nuclear DNA sequence variation and range-wide sampling. We documented a global demographic and spatial expansion approximately 0.1-0.3 My ago, most likely associated with an episode of massive drainage overflow. These events presumably enabled a historical continent-wide gene flow among hippopotamus populations, and hence no clear continental-scale genetic structuring remains. Nevertheless, present-day hippopotamus populations are genetically disconnected, probably as a result of the mid-Holocene aridification and contemporary anthropogenic pressures. This unique pattern contrasts with the biogeographic paradigms established for savannah-adapted ungulate mammals and should be further investigated in other water-associated taxa. Our study has important consequences for the conservation of the hippo, an emblematic but threatened species that requires specific protection to curtail its long-term decline.

Introduction

The Quaternary climatic oscillations did not affect only the Palearctic and Nearctic ecozones, but it also drastically influenced the distribution and diversity of African biota (Hewitt 2000, 2004, Lorenzen *et al.* 2012). In the northern hemisphere, the ice sheets extended over variable portions of the continents, with cold and dry glacial interludes of warm and wet interglacial cycles. In the African continent, at least five Pleistocene wet-dry periods induced repeated shifts of vegetation types, driving the diversification of their associated fauna (Szabo *et al.* 1995, Bobe & Behrensmeyer 2004). In central Africa in particular, the interglacial cycles led to repeated extensions of tropical forests during pluvial periods, and reductions of grasslands during dry interpluvial periods, and hence molding the genetic structure of several savannah adapted species (Miller *et al.* 2010, Rohland *et al.* 2005, Barnett *et al.* 2006, Dehghani *et al.* 2008, Lorenzen *et al.* 2012). At the same time, local tectonic and volcanic activities (e.g. uplift of the East African Rift valley) exacerbated temporal and spatial environmental heterogeneity, promoting the emergence of regional diversity (Maslin 2007, Trauth *et al.* 2007, Lorenzen *et al.* 2012).

Comparative phylogeographies accordingly identified genetic differentiation shared across species in the western, eastern, southern and southwestern regions of Africa, and delimited the main suture zones (Hewitt 2004, Lorenzen *et al.* 2012). In ungulates, many sub-specific lineages follow this regional structure and meet south and west of the Rift valley in eastern Africa (Mathee & Robinson 1999, Van Hooft *et al.* 2002, Muwanika *et al.* 2003, Lorenzen *et al.* 2010). While recent genetic-based studies are unraveling the picture for terrestrial vertebrates, including water-dependent species like buffalos or elephants (Heller *et al.* 2008, 2012, Okello *et al.* 2008), little is still known on the evolutionary history of semi-aquatic taxa inhabiting African rivers and lakes. Water-dependent taxa had to cope up with the dramatic changes in hydrographic networks during Pleistocene climatic fluctuations. For

example, major lakes and rivers in the continent alternated drainage and flow directions during the climatic cycles, and likely caused frequent population bottlenecks, diversification and expansions, as comprehensively illustrated by the African cichlids radiation (Elmer *et al.* 2009).

The Hippopotamidae family (Mammalia, Artiodactyla) first appeared during the middle Miocene, and diversified in East Africa as intense Plio-Pleistocene tectonics-associated Rift extension disrupted established hydrographic networks. The resultant African paleo-basins gave rise to at least eight *Hippopotamus* species and a high level of endemism, such that a single drainage system could host up to four exclusive species (Boisserie 2005, Boisserie *et al.* 2011). The subsequent post-Pleistocene expansions of *Hippopotamus* appear tightly linked to global climate cooling marked with wet episodes and C_4 plant propagation (Boisserie *et al.* 2011).

Of the two extant hippopotamus species, the common hippopotamus (*Hippopotamus amphibius*, popularly, the hippo) is widespread across sub-Saharan Africa, where the daily and seasonal movements are fully reliant on the availability of standing water. As such, the species is an ideal candidate for insights into the genetic consequences of the Pleistocene hydrographic instabilities. Hippos spend most of their daytime submerged in water because of thermoregulatory requirements, and only come out at night to forage in proximate grasslands within 1-3 km (Lewison 2011). As predominantly grazers, hippos live in male-dominated, polygynous herds ranging from a dozen, during the rainy season, to several hundred animals during the dry season. The population genetics status of this species remained understudied, partly due to the challenges in obtaining DNA samples.

In the only genetic study published to date, Okello *et al.* (2005) suggested that the eastern African hippos experienced a population expansion probably associated with a Late Pleistocene drainage overflow, and showed that little regional genetic structure existed

among their studied populations. In this study, we extend this population genetic survey to the entire hippopotamus' range, analyzing both mitochondrial and nuclear DNA sequence markers to understand how the species was genetically affected by the Pleistocene recurrent remodeling of river systems. In addition, we investigate whether genetic variation is associated with the current taxonomic subdivision based on morphology (Lydekker 1915). As shown in other water-dependent ungulates, phylogeographic divergences may arise with dry cycles, when populations experienced genetic disconnection and bottlenecks (e.g. Heller *et al.* 2008, Okello *et al.* 2008, Lozenzen *et al.* 2012). Alternatively, because of interconnected hydrographic basins, large-scale demographic expansions and gene flow are also expected during pluvial cycles (e.g. Verheyen et al. 2003).

Although the reasons are not fully understood, the hippo's populations have been in a long-term decline, with the species consequently listed as Vulnerable in the IUCN Red List of threatened species (Lewison & Oliver 2008). Our range-wide phylogeographic study thus has considerable relevance for future conservation strategies of this emblematic species.

Methods

DNA sampling and extraction

A total of 238 samples of *H. amphibius* were included in this study, originating from three different sources. Firstly, biopsies and scats were sampled from 126 individuals in western and southern Africa, with the scats collected at different feces clusters to minimize the risk of re-sampling the same individuals (e.g. Marucco *et al.* 2011). Secondly, we analyzed 103 specimens from eastern and southeastern Africa included in Okello *et al.* (2005). Lastly, we obtained nine bone samples from eastern and central Africa from the Natural History Museum of Nantes (MHN) in France, and the Royal Museum for Central Africa in Brussels (RMCA), Belgium. Table 1 and Table S1 provide detailed information on sample origins,

types and collectors. Samples were stored in ethanol, silicagel or DMSO. A scat sample of the hippos' closest living relative, the pygmy hippopotamus (*Hexaprotodon liberiensis*), was collected at the Basel Zoo, Switzerland, for use as an outgroup in the molecular analyses.

Prior to extraction, ethanol-preserved samples were dried (30" at 60°C), bones pulverized (Retsch MM300 mixer mill, QIAGEN), and skin biopsies hydrated (500µl of sterile deionised water for 1h). We then extracted DNA following standard protocols for either scats (QIAAMP DNA Stool Mini Kit, QIAGEN) or skin biopsies and bone tissues (DNeasy Tissue Kit, QIAGEN). Bones and scats were processed in a separate laboratory dedicated to the extraction and preparation of pre-PCR reagents of low copy number DNA samples.

Mitochondrial markers amplification and sequencing

We analyzed variations in two mitochondrial markers. We first amplified the complete control region (CR) and portions of the neighboring tRNA-Pro and tRNA-Phe regions in all 238 specimens and the pygmy hippopotamus, using published and newly developed primers (Taberlet & Bouvet 1994, Fumagalli *et al.* 1996; Table S2). For fresh scat and skin biopsies samples, sequences were obtained by separate Polymerase Chain Reaction (PCR) of two parts with primers L15995/H16498 (amplicon CR-L) and primers L16517/H-583HIPPO (amplicon CR-R). For degraded DNA present in old scats and bones, each part was obtained from two separate PCRs (4 PCRs per individual in total) with primers L15995/H-HIPPO1, L-HIPPO1/H16498 (CR-L), and L16517/H-HIPPO3, L-HIPPO3/H-583HIPPO (CR-R). Figure S1 summarizes the CR amplification strategy, including primer locations and amplicon lengths. We carried PCR amplification in 25µl reaction volumes, each containing 1× PCR Buffer (Applied Biosystems), 2mM of $MgCl_2$, 0.2mM of dNTPs, 0.5µM of each primer, 0.2mg/mL of BSA, 1U of TaqGold polymerase (Applied Biosystems) and 2µL tissue-derived or 10µL of stool/bone-derived DNA. Amplification conditions of

tissue-derived DNA consisted of an initial denaturation step of 95°C for 3', followed by 40 cycles of denaturation (95°C for 30"), annealing (50°C for 30"), extension (72°C for 1'), and a final extension (72°C for 7'). For scat and bones, conditions only differed by the number of cycles (50) and annealing temperatures (50°C for CR-L, 56°C for CR-R).

Secondly, we used newly developed primers L5HippoCytB and H4HippoCytB (Table S2) to amplify 854bp of the mitochondrial *cytochrome-b* (*cyt-b*) in a subset of 34 individuals, chosen to represent most of the CR haplotype diversity. Reagents concentrations were identical to CR's, and conditions of PCR amplification consisted of 50 cycles of denaturation (95°C for 30"), annealing (50°C for 30") and extension (72°C for 1'). All mitochondrial sequences were obtained with a ABI3730 genetic analyzer (Applied Biosystem), and quality controlled, concatenated (CR) and aligned (CLUSTALW algorithm, Larkin *et al.* 2007) in MEGA 4.0 (Tamura *et al.* 2007). Sequences of primers, including internal (for CR), were removed off from further analyses.

Nuclear marker amplification, cloning and sequencing

We sequenced intron 4 (as well as a part of adjacent exon 4) of the nuclear high-mobility-group protein B2 (HMG-2, 520bp, Backström *et al.* 2008) in a subset of 35 individuals, representative of geographic regions and mitochondrial diversity. PCR amplifications were carried out in 25µl volumes, each containing 1× PCR Buffer (QIAGEN), 2.5mM of $MgCl_2$, 0.125mM of dNTPs, 0.5µM of each primer, 1U of Taq polymerase (QIAGEN) and 3µL of DNA, with a touch-down procedure: initial denaturation at 94°C for 5', 20 cycles including denaturation (95°C for 30"), annealing (decreasing from 60°C to 50°C, by 0.5°C per cycle), extension (72°C for 45"), and an additional 20 regular PCR cycles (94°C for 30", 50°C for 30" and 72°C for 1'). We then cloned PCR products (pGEM®-T Easy, PROMEGA) and sequenced 11 samples (≥ 8 clones each) to identify the main alleles. We directly sequenced

PCR products from the other 24 individuals, identified heterozygotes from the sequencing electropherograms and inferred their phases using the PHASE algorithm in DNASP (Librado & Rozas 2009), following Dufresnes *et al.* (2013). Only heterozygotes displaying new haplotypes (not present among the first 11 cloned samples) were subsequently cloned to ascertain their reconstructed genotypes. All PCR products were sequenced on an ABI3730 genetic analyzer; sequences were edited and aligned (CLUSTALW) in MEGA.

Molecular divergence of CR haplotypes

We estimated the mutation rate and diversification time (time to the most recent common ancestor, MRCA) of the CR sequences in BEAST 1.6.2 (Drummond & Rambaut 2007). We applied a strict molecular clock, (assumptions which could not be rejected) and a coalescent prior (appropriate for intraspecific radiations). For the latter, we selected the one that yielded the highest likelihood in preliminary runs (1 chain of 5 million iterations): coalescent - exponential growth. No specific prior distribution was set on the molecular clock rate. We chose a GTR+G+I model of sequence evolution (MRAIC.pl 1.4.4, Nylander 2004), and calibrated the tree to the splits with *H. liberiensis*, using a lognormal prior with a wide distribution centered on 10 Mya (95 % CI: 1.2-31.4). This calibration was based on fossil cladistic analyses (Boisserie 2005 and references therein), corroborating the early divergence of *H. liberiensis* prior to the diversification of the Archaepotamus group, which has no recorded fossil older than 8 Mya. Our wide prior thus covers the putative time since the last common ancestor of *H. liberiensis* and *H. amphibius*, around 10 Mya (Boisserie 2005). We ran two independent BEAST chains of 30 million iterations each, with a sampling every 1,000 steps, and then used TRACER 1.5 to visualize and combine the results, and to check for convergence (through Effective Sample Sizes, ESS). The first three millions iterations were discarded as burn-in.

Analyses of genetic structure and diversity

We explored the range-scale structure of the hippo through haplotype networks in HAPSTAR (minimum-spanning tree option, Teacher & Griffiths 2010) built for each of the three datasets (CR, *cyt-b*, HMG-2). These networks are based on pairwise differences calculated by ARLEQUIN 3.5 (Excoffier *et al.* 2005). For CR, we also computed nucleotide (π) and haplotype (hd) diversities (in ARLEQUIN, Excoffier *et al.* 2005), and the proportion of private haplotypes within populations (where n > 10) and for the main African regions (according to Lorenzen *et al.* 2012; see Figure 1). In this analysis, we grouped individuals from the geographically close western Tanzania localities to increase sample size (Table 1; loc. 14a-f). In populations where n > 10, we then estimated genetic differentiation by pairwise Φ_{st}, and carried out Mantel tests of isolation by distance (Φ_{st} vs continental distance), and exact tests of population differentiation in ARLEQUIN, each based on 10,000 permutations. Furthermore, we performed analyses of molecular variance (AMOVA with 10,000 permutations) in ARLEQUIN to test whether the hippo CR variation was associated with major river basins (when known, Rekacewicz 2007; see Table 1) and the taxonomic subdivision proposed by Lydekker (1915; see Table 1). Lydekker (1915) described five *H. amphibius* subspecies based on cranial anatomy and crude geographic distribution. Because the latter was approximated to national boundaries, we discarded populations 6 and 17 from the analyses, as they fell in supposedly parapatric areas (Table 1).

Demographic analyses

We retraced the demographic fluctuations of *H. amphibius* with the Extended Bayesian Skyline Plot (EBSP, Heled & Drummond 2008) implemented in BEAST 1.6.2, combining all available sequences for each gene (CR, *cyt-b*, HMG-2). We used GTR+G+I (CR), HKY+I (*cyt-b*) and HKY (HMG-2) models (MRAIC), and estimated the clock rates (μ parameters) of

cyt-b and nuclear sequences using the CR's μ estimate as a reference (which was obtained from the fossil calibration described above). Different μ parameters were thus used for each three datasets. Other parameters were either left as default or optimized for the EBSP (following Heled 2010), and the chain was run for 30 million iterations. We then used TRACER to assess burn-in and effective sample sizes (ESS) of parameters. To check whether combining mitochondrial and nuclear markers could have skewed our results (as possible e.g. in the case of sex-biased dispersal), we re-ran the EBSP based on the mitochondrial datasets only.

For mitochondrial markers (CR, *cyt-b*), we compared observed pairwise number of differences (mismatch distributions) to distributions simulated under models of demographic (Schneider & Excoffier 1999) and range expansions (Excoffier 2004), in ARLEQUIN, which computes tests of goodness-of-fit (sum of squared deviation (SSD) and Harpending's raggedness index) and confidence intervals of parameters (based on 10,000 bootstrap replicates). For CR, we estimated the time since expansion from the parameter τ ($\tau=2\mu t$, with μ the substitution rate per sequence, and t the time since expansion). Finally, we conducted tests of selective neutrality, including Fu's F_s, Tajima's D, and Ramos-Onsins and Rozas's R^2, and evaluated their significances based on 10,000 coalescent simulations (DNASP 5, Ramos-Onsins & Rozas 2002). Because of the low level of polymorphism at the nuclear HMG-2, we did not perform demographic analyses solely on this marker.

Demographic reconstructions using Bayesian skyline plots are known to be influenced by sampling scheme and population fragmentation, sometimes leading to the detection of spurious population bottlenecks (Chikhi *et al.* 2010, Heller *et al.* 2013). Since our EBSP analysis indicated a recent population bottleneck (see Results), we evaluated the likelihood of this demographic trend by simulating data from our three markers under four possible demographic scenarios. Markers histories were simulated using a coalescent simulator (MS,

Hudson 2002), considering our sampling scheme (n = 238, 34 and 35 individuals for CR, *cyt-b* and HMG-2 respectively). We considered complete linkage for mitochondrial markers and no linkage for the nuclear marker, with a rate of coalescent 4 times larger for the latter (diploid and bi-parentally inherited) compared to the formers (haploid and maternally transmitted)). We used SEQ-GEN (Rambaut & Grassly 1997) to generate sequences of similar size as the observed data (929bp for CR, 854bp for *cyt-b* and 521bp HMG-2) and considering their respective evolutionary model (GTR for CR, HKY for the two other markers) and their estimated mutation rate. Simulated sequences were processed in BEAST for EBSP inferences, following the same methodology as for the hippo data. Each scenario started with a population size of 25,000 individuals. Scenarios (A), (B) and (C) assumed a 20 fold population expansion starting 300,000 years ago and ending 20,000 years ago (corresponding to the observed pattern, see Results). This expansion was followed by either: (scenario A) population bottleneck and fragmentation; (scenario B) population bottleneck; (scenario C) population fragmentation. Scenario (D) assumed a fragmented population of constant size. Bottlenecks consisted of a 3 fold decrease of population size. Population fragmentation was simulated by a severe reduction in migration rates from 10 to 1.2 migrants per generation (reflecting the observed mean Φ_{st} value of 0.3, as $\Phi_{st} = 1/(1+2Nm)$, $Nm = (1/\Phi_{st} - 1)/2 \simeq 1.2$, with N the effective size of each population and m the migration rate).

Results

Marker variability and molecular dating

Sequences of the mitochondrial control region and flanking tRNAs (896bp + 33bp respectively, totaling 929bp) were the most variable (100 haplotypes, 89 polymorphic sites (9.6%), including 66 parsimony-informative (7.1%); n = 238 sequences). *Cyt-b* sequences (854bp) were also substantially polymorphic (23 haplotypes, 34 polymorphic sites (4.0%),

including 17 parsimony-informative sites (2.0%); n = 34 sequences). In contrast, we only found 7 closely-related haplotypes from the nuclear HMG-2 (520bp; 8 polymorphic sites (1.5%), including 1 parsimony-informative (0.2%); n = 70 sequences from 35 diploid individuals). We carefully checked for sequence incongruences (e.g. double peaks, stop-codons, sequence mismatch between overlapping amplicons) in order to avoid nuclear mitochondrial DNA (numt). The estimated time to the MRCA of *H. amphibius* CR sequences was approximately 0.7 Mya (95% HPD = 0.3-1.2), with a substitution rate μ of 0.052 substitutions/site/My (95 % HPD = 0.02-0.11; equivalent to 10% of divergence/My). Both chains yielded nearly identical results, suggesting they reached convergence.

Genetic structure and diversity

Both the mitochondrial and nuclear haplotype networks failed to reflect clear geographic structuring throughout Africa (Figure 1). Some branches were restricted to single geographic regions, notably in eastern Africa (Figure 1c), but haplotype distribution generally indicated weak continental structure: genetically distant haplotypes were found within the same regions (e.g. H02 and H03), and some genetically close haplotypes were also found more than 10,000 km apart (e.g. H26 and H31). Interestingly, only 2 out of 100 CR haplotypes were shared across the different biogeographic regions (H13, H47), and most populations featured high proportions of private haplotypes (Table 1). Accordingly, measures of population differentiation (Φ_{st}) were significantly high across all population pairs tested, with an average Φ_{st} of 0.33. The only exceptions were two pairs of proximate populations: loc. 5 and loc. 6 along the Niger River, and loc. 12a and14a-f in Eastern Africa; only the former pair was not significantly differentiated based on exact tests (Table S3). Mantel tests (Φ_{st} *versus* land distance) did not detect significant isolation by distance ($P = 0.08$).

Analyses of molecular variance revealed a significant effect of river basins, but not subspecific delineations, with most of the variance found within populations in both analyses (70.0 % and 70.6 % respectively; Table 2). Within population, nucleotide and haplotype diversities ranged from 4.6×10^{-4} to 1.9×10^{-2} and 0.167 to 0.857 respectively, being overall higher in eastern and southern Africa (Table 1, Figure 1).

Demographic analyses

The EBSP reconstruction revealed a 10 fold demographic expansion during the Late Pleistocene (300,000 – 30,000 years ago), followed by a drastic decrease since the past 15,000 years (Figure 2). Analyses based solely on mitochondrial markers yielded a similar pattern. In further support of past population expansion, the mismatch distributions of CR and *cyt-b* sequences were unimodal (Figure 2). Accordingly, both models of demographic and spatial expansion could not be significantly rejected and featured similar estimates (Table 3). Calculated from CR's τ using the substitution rate reported above, dates of expansions also point back to the Late Pleistocene (demographic: 150,000 years, spatial: 130,000 years; Table 3). Among the neutrality tests performed, only Fu's F_s was significant for CR (Table 3), but all three were significant for *cyt-b* (Table 3). The bottleneck inferred from the EBSP was supported by our simulation results (Figure 3). The observed hippo EBSP was compatible with simulated scenarios (A) and (B) –both accounting for population expansion followed by population bottleneck, with or without fragmentation (Figure 3A-B). Under scenarios (A) and (B), the observed mean EBSP was entirely contained within the 95% highest posterior density (HPD) intervals in 99.5% and 100% of simulations, respectively. On the contrary, the hippo EBSP was not compatible with scenarios (C) and (D) –without bottlenecks (Figure 3C-D). Under scenarios (C) and (D), the observed mean EBSP fell within the observed HPD interval in only 0.5% and 0% of simulations, respectively. Overall, simulations with

implemented bottlenecks generated sharp population declines in the EBSP (Figure 3A-B), which were not observed from bottleneck-free simulations (Figure 3C-D). Under fragmentation only, population sizes were fluctuating and decreasing in the recent past (Figure 3D) but did not match the pattern observed with hippo sequences.

Discussion

Our study aimed at understanding how the common hippo was genetically structured by historical fluctuations of hydrographic networks during the Quaternary climatic oscillations. Based on a wide geographical sampling and using three different polymorphic markers, we reconstructed the hippo's biogeographic history throughout Africa, which had been limited to the eastern part of the continent (Okello *et al.* 2005). The pattern obtained was different from that found in most savannah-adapted mammals, where regional genetic structuring testifies of isolation and subsequent population differentiation in Pleistocene grassland refugia. In contrast, the hippo displays no clearly delimited phylogeographic groups, but populations show evidence of contemporary genetic isolation. Our findings are particularly relevant for the conservation strategy of this vulnerable species.

Weak continental-scale structure and population expansion

Molecular dating analyses of the mitochondrial control region indicate that the hippo populations diversified at the beginning of the Late Pleistocene (time to MRCA = 0.7 Mya). At this time, Africa had just got out of several major humid episodes (with at least three moisture periods from 2.7 to 0.9 Mya in East Africa), and the long-term aridification of the continent intensified (Trauth *et al.* 2005). The dry cycles became more severe over the past 0.6 Mya, resulting in the replacement of tropical forests by savannah grasslands (deMenocal 1995, Hewitt 2004), as well as strong fluctuations in water levels or complete desiccation of

major lakes (Elmer *et al.* 2009). These processes triggered allopatric differentiations in many savannah animals like the African buffalos (Van Hooft *et al.* 2002), giraffe (Brown *et al.* 2007), bushbuck (Moodley & Bruford 2007), ostrich (Miller *et al.* 2010) and their associated predators (e.g. lion, Barnett *et al.* 2006; spotted hyena, Rohland *et al.* 2005). Hippos were presumably less affected by retreats and extensions of savannahs (although past Hippopotamidae expansions associate with C_4 plant propagation, Boisserie *et al.* 2011), but arid phases likely impacted the genetic diversity of this semi-aquatic species, for instance by inducing divergences between disconnected river basins.

Interestingly, our haplotype networks did not reflect strong geographic associations of such mitochondrial and nuclear divergences, although the control region appeared sufficiently polymorphic (comparable to other African ungulates, e.g. Van Hooft *et al.* 2002, Lorenzen *et al.* 2008). Several geographically-close populations share phylogenetically distant haplotypes, a pattern also shown by the slower-evolving *cyt-b* and HMG-2 networks (Figure 1). Such weak phylogeographic structure suggests that hippos experienced high levels of gene flow, which likely occurred during the period of population expansion depicted by our analyses (some 0.1-0.3 Mya). Continental-scale population growth was independently supported by the EBSP, mismatch distributions and most neutrality tests, with congruent time estimates (EBSP: 0.1-0.3 Mya; mismatch distribution: 0.1-0.2 Mya). The observed demographic explosion was probably associated with a Late Pleistocene drainage overflow, resulting in population increase and dispersal across the entire continent. Interconnection of hydrographic basins was corroborated by evolutionary studies on cichlid fishes in East Africa (Van Couvering 1982, Verheyen *et al.* 2003). Based on dating of lacustrine deposits, Szabo *et al.* (1995) described several pluvial episodes inducing the formation of African paleo-lakes since the past 320,000 years. Paleo-lakes and their associated networks like the Mega-Chad in central Africa and the Makgadikgadi in southern Africa used to be particularly extensive,

and probably connected wide parts of Africa (Drake & Bristow 2006, Podgorski *et al.* 2013), facilitating the population explosion of water-dependent species.

The hippo population expansion was not limited to Africa; according to fossil records, the species also spread throughout Europe between 350,000 and 120,000 years ago (Eltringham 1999, Schreve 2009). These events seem to correspond to Late Pleistocene expansions reported for eastern African hippos (Okello *et al.* 2005), and other water-associated mammals like the Cape buffalo (Van Hooft *et al.* 2002). This phylogeographic pattern remarkably differs from other African mammals living in open habitats, which usually feature spatially-explicit distributions of lineages, caused by dispersal barriers such as tropical forests (Lorenzen *et al.* 2012). In contrast, hippos seem to overcome such barriers, most likely thanks to their aquatic habits: tropical regions are crossed by large rivers and hippos are also known to migrate along coastal sea waters (Lewison 2011).

The distribution of the CR diversity suggests ancestral westward and southward colonization from eastern Africa, with associated drift resulting in lower variability across western and southern parts of the continent (Figure 1e). Accordingly, several CR lineages were found almost exclusively in eastern Africa (Figure 1). Although traces of ancestral populations were largely masked by subsequent genetic exchanges, eastern Africa is considered a cradle of diversity for many taxa (Hewitt 2000, Lorenzen *et al.* 2012), including humans (Elhassan *et al.* 2014). As such, this region is a plausible candidate refugium for the hippo, from where its genus diversified (Boisserie 2005, Boisserie *et al.* 2011). The complexity of the CR network (Figure 1) is also in support of an abrupt increase in migration as would occur with the reconnection of formerly isolated populations, which is expected to produce persistent high-levels of genetic diversity (Alcala *et al.* 2013).

Recent bottlenecks and limited contemporary gene-flow

Instead of past dispersion, one could argue that the absence of overall structure is still maintained by current migration. Translocations of living individuals as part of conservation programs (Eltringham 1999) would also have similar effects. However, multiple lines of evidence support limited contemporary gene flow across the range. First, we reported strong and significant Φ_{st} values and exact tests of differentiation between almost all pairs of populations. Secondly, the high proportion of private CR haplotypes is consistent with reduced gene flow between biogeographic areas (as also shown by region-specific lineages), a pattern mirrored even at local geographical scales (Table 1). Nevertheless, we could significantly relate a small proportion of the genetic variation to hydrographic basins, indicating that some migration events occur along regional water networks.

Present population disconnection can be considered in the light of recent demographic bottleneck inferred from our EBSP analyses (decrease of about 70% since the past 15,000 years, Figure 2). However, the inference of historical demographic changes with Bayesian skyline plots is known to be affected by population structure and sampling scheme, leaving under certain circumstances false signals of population decline (Chikhi *et al.* 2010, Heller *et al.* 2013). However, this issue had limited bias to our results. First, alternative demographic histories without expansion and/or bottleneck did not match the observed pattern (Figure 3C and 3D). Secondly, although fragmentation might impact the EBSP reconstruction, we could reject that it solely generated the population size decrease recovered from hippo's data (Figure 3). Thirdly, both population expansions and bottlenecks are detectable using our sampling scheme including numerous samples from many populations (Heller *et al.* 2013). Fourthly, this demographic scenario is consistent with the current fragmented distribution of hippo's populations and local genetic structure (Figure 1a). Sex-biased dispersal may also skew the coalescence patterns of mitochondrial *versus* nuclear markers, potentially biasing

our combined EBSP analyses, but analyses based solely on mtDNA depicted a similar demographic reconstruction. The timing of demographic events inferred from genetic data is also sensitive to the molecular calibration applied and additional points of calibrations would have increased the reliability of our results.

Our conclusions are also supported by the climatic history of the African continent and reported historic range distributions of other large African animals. These disconnection events are likely to result from the aridification of Africa during the Holocene. Our observations parallel the dramatic declines of the water-dependent African buffalo and savannah elephant (Heller *et al.* 2008, 2012, Okello *et al.* 2008). Like for the hippos, archeological evidences (cave carvings and paintings) suggest that these animals were abundant throughout the "green Sahara" within less than 10,000 years ago (deMenocal & Tierney 2012). At this epoch, most water bodies had much higher levels than today, with some paleo-lakes as inland seas. In fact, around 5,000 BCE, Lake Chad was larger than today's largest lake on Earth, the Caspian Sea, but is now severely threatened by desiccation (Drake & Bristow 2006). Reduction of hydrographic networks also triggered genetic differentiation in the Nile crocodile, structuring populations according to major river drainages (Hekkala *et al.* 2010). Altogether, these findings advocate for the disproportionate impact of aridification on water-dependent taxa, particularly semi-aquatic species, compared to other large African animals. Besides, recent anthropogenic pressures on hippo populations and habitats might have further intensified genetic bottlenecks and population fragmentation within the last 50 years (Lewison & Oliver 2008).

The marked genetic differentiation of hippos' populations, based on maternally-inherited mtDNA, could indicate that hippos, or at least females, are rather poor dispersers in the present context (aridification and reduced hydrographic networks, human disturbances). Whereas this might be surprising for such a large animal, it can be explained by its

movement-restrictive thermoregulatory constraints (Lewison 2011). Apart from a few isolated cases, very little is known on *H. amphibius* dispersal habits, as studies based on mark-recapture work or telemetry are perilous in this species (Eltringham 1999). Even the Plio-Pleistocene *Hippopotamus* relatives, also semiaquatic, probably had limited dispersal abilities as well (Boisserie *et al.* 2011). Faster-evolving markers, including Y-linked loci, will be useful to estimate sex-specific dispersal rates.

Implications for conservation

The Late Pleistocene history of the hippo can somehow be challenging for conservation biologists: despite population fragmentation and recent bottlenecks, no evident signal of continental-scale structuring was found. At first glance, phylogeographers would fail to identify Evolutionarily Significant Units (ESU, Moritz 1994) based on geographic distribution of genetic divergences or phenotypic variation. Indeed, neutral genetic variation was not associated with proposed taxonomic subdivision based on morphology (Table 2). Such results might lead to misinterpret the extant hippo populations as a single range-wide population unworthy of specific management plans. On the contrary, we clearly established that populations are nowadays fragmented and feature regional diversity hotspots (e.g. eastern Africa) and genetic endemism (i.e. high-proportion of private haplotypes) and thus deserve specific attention. Management units based on other estimators (e.g. demographic) might prove useful in accounting for the conservation value of closely related, but yet differentiated populations (Dufresnes *et al.* 2013).

The common hippopotamus is long-term declining, to such extent that it is now considered vulnerable by the IUCN. Recent census estimated a 7-20% decrease over the past 10 years, attributed to illegal hunting for ivory and meat, and habitat loss linked to human activities (particularly severe in western Africa (Lewison & Oliver 2008). Human mediated

interferences such as intensified grazing of livestock on river banks also considerably decreases the food resources available to hippos (Noirard *et al.* 2004). Consequent to these threats, the former highly-connected and widespread distribution range of hippos has been reduced to extremely scattered and fragmented populations that nowadays only subsist in protected areas (Figure 1a). From our results, it seems urgent to set up specific conservation strategies aiming at maintaining present wetland habitat connections, and encouraging regional-based management in order to protect a maximum amount of diversity, both genetic and phenotypic (i.e. proposed subspecies). Future regional studies using fast-evolving markers like microsatellites will be useful to identify the most vulnerable populations, i.e. harboring signs of reduced genetic diversity and/or inbreeding. In addition, screening for adaptive variation with population genomics tools could help distinguishing between locally-adapted populations for future management planning.

Conclusions

The demographic history of the common hippopotamus appears tightly linked to hydrographic fluctuations associated with wet-dry cycles, alternating population bottlenecks during arid phases and rapid recoveries during pluvial episodes, the latter purging traces of divergences generated by the former. This unique pattern contrasts with established biogeographic paradigms for most large African mammals, which usually feature range-scale phylogeographic structuring. Future comparative phylogeographies will be important to test whether this pattern is shared across other water-dependent or semi-aquatic taxa. Finally, the fact that the remaining hippo populations currently face genetic disconnection urges the need for specific protection and regional restoration programs.

Acknowledgments

We are thankful to W. Barklow and R.D. Beckwitt (Framingham State University, MA, USA), J. and F. Reinhard (BRinK, Kuzikus, Namibia), M. Dethier and A. Michez (Laboratoire de Foresterie tropicale et subtropicale, Gembloux, Belgium), M. Le Berre (Conservation et Développement Durable, France), A. Harouna (Ministry of Environment, Niger), M. Wali (W National Park, Niger), S. Hoby (Basel Zoo, Switzerland), A. Le Héritte (Biotropica Zoo, France), P. Gay (Doué La Fontaine Zoo, France), J.M. Lernould (CEPA association, France), as well as the Natural History Museum of Nantes (France) and the Royal Museum for Central Africa of Brussels (Belgium) for providing samples.
We also thank F. Biollaz, A. Horn, R. Burri, H. Collin, G. Grifoni, S. Biollay, S. Antoniazza, N. Salamin, C. Sépulcre, M. Stöck and N. Remollino for their help in the lab and useful discussions. CD was funded by a PhD fellowship from the Faculty of Biology and Medicine, University of Lausanne.

References

Alcala N, Streit D, Goudet J *et al.* (2013) Peak and persistent excess of genetic diversity following an abrupt migration increase. *Genetics*, **193**, 953-971.

Backström N, Fagerberg S, Ellegren H (2008) Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Molecular Ecology*, **17**, 964-980.

Barnett R, Yamaguchi N, Barnes I, Cooper A (2006) The origin, current diversity and future conservation of the modern lion (*Panthera leo*). *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2119-2125.

Bobe R, Behrensmeyer AK (2004) The expansion of grassland ecosystems in Africa in relation to mammalian evolution and the origin of the genus *Homo*. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **207**, 399-420.

Boisserie JR (2005) The phylogeny and taxonomy of Hippopotamidae (Mammalia: Artiodactyla): a review based on morphology and cladistic analysis. *Zoological Journal of the Linnean Society*, **143**, 1-26.

Boisserie JR, Lihoreau F, Brunet M (2005) The position of Hippopotamidae within Cetartiodactyla. *Proceedings of the National Academy of Sciences*, **102**, 1537-1541.

Boisserie JR, Fisher RE, Lihoreau F, Weston M (2011) Evolving between land and water: key questions on the emergence and history of the Hippopotamidae (Hippopotamoidea, Cetancodonta, Cetartiodactyla). *Biological Reviews*, **86**, 601-625.

Brown DM, Brenneman RA, Koepfli KP *et al.* (2007) Extensive population genetic structure in the giraffe. *BMC Biology*, **5**, 57.

Chikhi L, Sousa VC, Luisi P *et al.* (2010) The cofounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, **186**, 983-995.

Dehghani R, Wanntorp L, Pagani P *et al.* (2008) Phylogeography of the white-tailed mongoose (Herpestidae, Carnivora, Mammalia) based on partial sequences of the mtDNA control region. *Journal of Zoology*, **276**, 385-393.

Drake N, Bristow C (2006) Shorelines in the Sahara: geomorphological evidence for an enhanced monsoon from palaeolake Megachad. *The Holocene*, **16**, 901-911.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Dufresnes C, Wassef J, Ghali K *et al.* (2013) Conservation phylogeography: does historical diversity contribute to regional vulnerability in European tree frogs (*Hyla arborea*)? *Molecular Ecology*, **22**, 5669-5684.

Elhassan N, Gebremeskel EI, Elnour MA *et al.* (2014) The episode of genetic drift defining the migration of humans out of Africa as derived from a large East African population size. *PLoS ONE*, **9**, e97674.

Elmer KR, Reggio C, Wirth T (2009) Pleistocene desiccation in East Africa bottlenecked but did not extirpate the adaptive radiation of Lake Victoria haplochromine cichlid fishes. *Proceedings of the National Academy of Sciences*, **106**, 13404-13409.

Eltringham SK (1999) *The Hippos: Natural history and conservation*. Academic Press, London.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**, 853-864.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN ver 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Fumagalli L, Taberlet P, Favre L, Hausser J (1996) Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Molecular Biology and Evolution*, **13**, 31-46.

Hekkala ER, Amato G, DeSalle R, Blum MJ (2010) Molecular assessment of population differentiation and individual assignment potential of Nile crocodile (*Crocodylus niloticus*) populations. *Conservation Genetics*, **11**, 1435-1443.

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289.

Heled J (2010) Extended Bayesian skyline plot tutorial. Available from: tutorial.east.bio.ed.ac.uk/Tutorials

Heller R, Lorenzen ED, Okello JBA *et al.* (2008) Mid-Holocene decline in African buffalos inferred from Bayesian coalescent-based analyses of microsatellites and mitochondrial DNA. *Molecular Ecology*, **17**, 4845-4858.

Heller R, Brüniche-Olsen A, Siegismund HR (2012) Cape buffalo mitogenomics reveals a Holocene shift in the African human-megafauna dynamics. *Molecular Ecology*, **21**, 3947-3959.

Heller R, Chikhi L, Siegismund HR (2013) The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS ONE*, e62992.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Hewitt GM (2004) The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology*, **1**, 4.

Hudson RR (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337-338.

Kingdon J (1997) The Kingdon field guide to African Mammals. Academic Press, San Diego, California.

Larkin MA, Blackshields G, Brown NP *et al.* (2007) CLUSTAL W and CLUSTAL X version 2.0. *Bioinformatics*, **23**, 2947-2948.

Lewison RL (2011) Family Hippopotamidae (Hippopotamuses). In: *Handbook of the Mammals of the World. Vol. 2. Hoofed Mammals* (eds Wilson DE, Mittermeier RA), pp. 308-319. Lynx Ediciones, Barcelona, Spain.

Lewison, R, Oliver W (IUCN SSC Hippo Specialist Subgroup) (2008) *Hippopotamus amphibius*. In: IUCN 2013. IUCN Red List of Threatened Species. Version 2013.1. Downloaded on 16.07.2013.

Librado P, Rozas J (2009) DNASP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452.

Lorenzen ED, Arctander P, Siegismund HR (2008) High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): insights from mtDNA and microsatellites. *Molecular Ecology*, **17**, 2812-2824.

Lorenzen ED, Masembe C, Arctander P, Siegismund HR (2010) A long-standing Pleistocene refugium in southern Africa and a mosaic of refugia in East Africa: insights from mtDNA and the common eland antelope. *Journal of Biogeography*, **37**, 571-581.

Lorenzen ED, Heller R, Siegismund HR (2012) Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, **21**, 3656-3670.

Lydekker R (1915) *Catalogue of the Ungulate Mammals in the British Museum of Natural History*. Vol. 5 British Museum, London.

Marucco F, Boitani L, Pletscher DH, Schwartz MJ (2011) Bridging the gaps between non-invasive genetic sampling and population parameter estimation. *European Journal of Wildlife Research*, **57**:1-13.

Maslin M (2007) Tectonics, orbital forcing, global climate change and human evolution in Africa: introduction to the African paleoclimate special volume. *Journal of Human Evolution*, **53**, 2484-2498.

Mathee CA, Robinson TJ (1999) Mitochondrial DNA population structure of roan and sable antelope: implications for the translocation and conservation of the species. *Molecular Ecology*, **31**, 861-877.

deMenocal PB (1995) Plio-Pleistocene African climate. *Science*, **270**, 53-59.

deMenocal PB, Tierney JE (2012) Green Sahara: African humid periods paced by Earth's orbital changes. *Nature Education Knowledge*, **3**, 12.

Miller JM, Hallager S, Monfort SL *et al.* (2010) Phylogeographic analysis of nuclear and mtDNA supports subspecies designations in the ostrich (*Struthio camelus*). *Conservation Genetics*, **12**, 423-431.

Moodley Y, Bruford MW (2007) Molecular biogeography: towards an integrated framework for conserving pan-African biodiversity. *PLoS ONE*, **2**, e454.

Moritz C (1994) Defining 'evolutionarily significant units' for conservation. *Trends in Ecology and Evolution*, **9**, 373-375.

Muwanika VB, Nyakaana S, Siegismund HR, Arctander P (2003) Phylogeography and population structure of the common warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity*, **91**, 361-372.

Noirard C, Le Berre M, Ramousse R, Sépulcre C, Joly P (2004) Diets of sympatric hippopotamus (*Hippopotamus amphibius*) and zebus (*Bos indicus*) during the dry season in the "W" National Park (Niger Republic). *Game and Wildlife Science*, **21**, 423-431.

Nylander JAA (2004) MrAIC.pl Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Okello JBA, Nyakaana S, Masembe C *et al.* (2005). Mitochondrial DNA variation of the common hippopotamus: evidence for a recent population expansion. *Heredity*, **95**, 206-215.

Okello JBA, Wittemeyer G, Rasmussen HB *et al.* (2008) Effective population size dynamics reveal impacts of historic climate events and recent anthropogenic pressure in African elephants. *Molecular Ecology*, **17**, 3788-3799.

Podgorski JE, Green AG, Kgotlhang L *et al.* (2013) Paleo-megalake and paleo-megafan in southern Africa. *Geology*, 41: 1155-1158.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092-2100.

Rambaut A, Grass NC (1997) SEQ-GEN: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer applications in the biosciences*: *CABIOS*, **13**, 235-238.

Rekacewicz P (2007) Africa's rivers and lake basins cross many borders. Available at: http://www.grida.no/publications/vg/water2/page/3223.aspx

Rohland N, Pollack JL, Nagel D *et al.* (2005) The population history of extant and extinct hyenas. *Molecular Biology and Evolution*, **22**, 2435-2443.

Schneider S, Excoffier L (1999) Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079-1089.

Schreve DC (2009) A new record of Pleistocene hippopotamus from River Severn terrace deposits, Gloucester, UK-palaeoenvironmental setting and stratigraphical significance. *Proceedings of the Geologists' Association*, **120**, 58-64.

Szabo B, Haynes C, Maxwell TA (1995) Ages of Quaternary pluvial episodes determined by uranium-series and radiocarbon dating of lacustrine deposits of Eastern Sahara. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **113**, 227-242.

Taberlet P, Bouvet J (1994) Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown beat (*Ursus arctos*) in Europe. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 1293-1302.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599.

Teacher AGF, Griffiths DJ (2010) HAPSTAR: automated haplotype network layout and visualization. *Molecular Ecology Resources*, **11**, 151-153.

Trauth MH, Maslin MA, Deino A, Strecker MR (2005) Late Cenozoic Moisture History of East Africa. *Science* **309**, 2051-2053.

Trauth MH, Maslin MA, Deino AL *et al.* (2007) High- and low-altitude forcing of Plio-Pleistocene East African climate and human evolution. *Journal of Human Evolution*, **53**, 475-486.

Van Couvering JAH (1982) *Fossil Cichlid Fish of Africa*. Special Papers in Palaeontology, Vol. 29. Palaeontological Association, London.

Van Hooft WF, Groen AF, Prins HHT (2002) Phylogeography of the African buffalo based on mitochondrial and Y-chromosomal loci: Pleistocene origin and population expansion of the Cape buffalo subspecies. *Molecular Ecology*, **11**, 267-279.

Verheyen E, Salzburger W, Snoeks J, Meyer A (2003) Origin of the superflock Cichlid fishes from Lake Victoria, East Africa. *Science*, **200**, 325-329.

Data accessibility

Sequences are available on GENBANK (KR004948-KR005078) and accession numbers are listed in Table S1. Sequence alignments and codes used for analyses are archived on DRYAD, doi:10.5061/dryad.4cv02.

Author's contributions

LF designed and organized research. All authors performed research and CD, CS and LF drafted the manuscript.

Figure legends

Figure 1: Sampling map (a), minimum spanning tree haplotype networks (b: *cytochrome-b*, c: control-region, and d: nuclear HMG-2), and distribution of the control region (CR) diversity (e) in the common hippo. Current and former (Kingdon 1997) distribution ranges are represented in dashed and plain red respectively (data: Lewison & Oliver 2008). For localities (a) and haplotypes (b-d), circles are proportional to sample size (scale differs between networks) and colored according to the main biogeographic regions identified in Africa (following Lorenzen *et al.* 2012). Scales show branch lengths relative to pairwise number of differences. On map a), empty circles (i.e. loc. 2, 8, 10 and 15) represent localities where only the country of origin is known (see Table 1). Map e) represents CR's nucleotide diversity π (white bars) and haplotype diversity hd (black bars) in populations where $n > 10$.

Figure 2: Demographic analyses. The Extended Bayesian Skyline Plot (EBSP) shows a reconstruction of the demographic fluctuations since the Late Pleistocene, based on combined CR, *cyt-b* and HMG-2 sequence variation. For mitochondrial markers (CR, *cyt-b*), observed pairwise numbers of differences (white circles and plain lines) are plotted against models of sudden expansion (thick dash line) and their 95% confidence intervals (thin dash lines). Similar curves were obtained with models of spatial expansion (not shown, see Table 3 for statistics).

Figure 3: Correspondence between EBSP reconstruction of the hippo demographic history (black line) and 4 simulated demographic scenarios (colors). Scenarios are: (A) Pleistocene expansion followed by Holocene population bottleneck and fragmentation; (B) Pleistocene expansion followed by Holocene population bottleneck; (C) Pleistocene

expansion followed by Holocene population fragmentation; (D) Constant population size with structure. See the main text for scenarios parameters. In each panel, color gradients represent the percentage of simulations (out of 200 replicates) which led to the highest (95%) posterior density (HPD) regions. The mean hippo EBSP is contained within the majority of simulated HPD regions in panels (A) and (B), (99.5% and 100% of simulations, respectively), only in regions from 0.1 My ago to 0.6 My ago, but not from 0.1My ago to present (0.5% simulations) in panel (C) and is not contained within simulated HPD region (0% simulations) in panel (D). The Holocene period (last 20’000 years) is zoomed within inner panels.

Table 1: Detailed locality information. This includes river basins and subspecies as used in the AMOVA analyses (see Table 2). We grouped populations according to the main biogeographic regions, following Lorenzen *et al.* (2012, see also Figure 1). For regions and populations where n > 10, mtDNA control region nucleotide (π) and haplotypic (hd) diversity are provided along with the proportion of private haplotypes (pph). For some samples, only the country of origin is known. n: number of individuals, nph: number of individuals with private haplotypes.

loc. ID	Country	Locality	river basin	subspecies	n	nph	pph	π	hd	Collector/Origin
Western Africa					**59**	**59**	**1**	**0.012456**	**0.322**	
1	Senegal	Casamance	-	*H. a. amphibius*	1	1	-	-	-	MHN Nantes
2	Ivory Coast	Unknown	-	*H. a. amphibius*	1	1	-	-	-	MHN Nantes
3	Burkina Faso	Réserve de la Mare aux hippopotames	Volta	*H. a. amphibius*	17	17	1	0.000459	0.176	C. Noirard
-	Burkina Faso	Unknown	Volta	*H. a. amphibius*	4	-	-	-	-	C. Noirard
4	Ghana	Ntereso	Volta	*H. a. amphibius*	1	0	-	-	-	Okello *et al.* 2005
5	Niger	Ayorou	Niger	*H. a. tschadensis*	14	1	0.07	0.008138	0.214	C. Noirard
-	Niger	Unknown	Niger	*H. a. tschadensis*	5	-	-	-	-	C. Noirard
6	Benin/Niger	W National Park	Niger	*H. a. amphibius/tschadensis*	16	5	0.31	0.007729	0.375	C. Noirard
Central Africa					**16**	**15**	**0.94**	**0.009153**	**0.375**	
7	Chad	Bol	-	*H. a. tschadensis*	1	1	-	-	-	MHN Nantes
8	Cameroon	Unknown	-	*H. a. amphibius*	1	1	-	-	-	MHN Nantes
9	Gabon	Loango National Park	Ogooue	*H. a. amphibius*	12	12	1	0.000522	0.167	A. Michez
10	RD Congo	Unknown	Congo	-	2	-	-	-	-	RMCA, Brussels
Eastern Africa					**106**	**89**	**0.84**	**0.019307**	**0.509**	
11	Sudan	White Nile	Nile	*H. a. amphibius*	1	1	-	-	-	MHN Nantes
12a	Uganda	Murchinson Falls National Park	Nile	*H. a. amphibius*	23	20	0.87	0.01876	0.739	Okello *et al.* 2005
12b	Uganda	Queen Elizabeth National Park	Nile	*H. a. amphibius*	22	7	0.32	0.01694	0.318	Okello *et al.* 2005

12c	Uganda	L. Mburo National Park	Nile	*H. a. amphibius*	1	1	-	-	-	Okello *et al.* 2005
-	Uganda	Unknown	Nile	*H. a. amphibius*	7	-	-	-	-	C. Noirard
13a	Kenya	Masai Mara Game Reserve	Nile	*H. a. kiboko*	27	17	0.63	0.01628	0.593	Okello *et al.* 2005
13b	Kenya	Naivasha National Park	Nile	*H. a. kiboko*	9	0	-	-	-	Okello *et al.* 2005
14a	Tanzania	Ugallla	Congo	*H. a. amphibius*	6	3	-	-	-	Okello *et al.* 2005
14b	Tanzania	Moyowosi	Congo	*H. a. amphibius*	1	1	-	-	-	Okello *et al.* 2005
14c	Tanzania	Luganzo	Congo	*H. a. amphibius*	1	1	-	-	-	Okello *et al.* 2005
14d	Tanzania	Kigosi	Congo	*H. a. amphibius*	4	2	-	-	-	Okello *et al.* 2005
14e	Tanzania	Mlele	Congo	*H. a. amphibius*	1	1	-	-	-	Okello *et al.* 2005
14f	Tanzania	Niensi	Congo	*H. a. amphibius*	1	0	-	-	-	Okello *et al.* 2005
14a-f	Tanzania	pooled localities from Western-Tanzania	-	*H. a. amphibius*	14	9	0.64	0.01864	0.857	Okello *et al.* 2005
14g	Tanzania	Selous	-	*H. a. amphibius*	1	1	-	-	-	Okello *et al.* 2005
15	Burundi	Unknown	-	*H. a. amphibius*	1	1	-	-	-	MHN Nantes
Southern Africa					**57**	**56**	**0.98**	**0.017421**	**0.404**	
16a	Zambia	Luangwa National Park	Zambesi	*H. a. capensis*	4	4	-	-		Okello *et al.* 2005
16b	Zambia	Kafue S	Zambesi	*H. a. capensis*	1	0	-	-		Okello *et al.* 2005
16c	Zambia	Lake Bangweolo	Congo	*H. a. capensis*	1	1	-	-		Okello *et al.* 2005
17	Namibia	Caprivi National Park, Kwando River	Zambesi	*H. a. capensis/constrictus*	17	15	0.88	0.01342	0.412	F. Reinhard
18	South-Africa	Kruger National Park	Limpopo	*H. a. capensis*	34	34	1	0.01639	0.324	W. Barklow, C. Noirard

Table 2: Hierarchical analyses of molecular variance (AMOVA) based on mitochondrial control region (CR) sequences. Φ statistics (fixation indices) corresponds to Φ_{ST} (among populations among groups), Φ_{SC} (among populations within groups) and Φ_{CT} (among groups). *: p-value < 0.05.

Source of variation	df	Sum of squares	Variance components	Φ Statistic	% of variation	p-value
Hydrographic basins						
among basins	6	370.1	0.99	0.21	11.2	*
among populations within basins	19	316.0	1.66	0.30	16.8	*
within populations	206	1269.1	6.16	0.11	70.0	*
Taxonomic subdivision						
among subspecies	3	204.9	0.43	0.26	4.8	0.18
among populations within subspecies	25	434.6	2.21	0.29	24.6	*
within populations	174	1100.8	6.33	0.05	70.6	*

Table 3: Summary statistics for analyses of mismatch distribution under models of demographic (demo) and spatial expansion, and neutrality tests. Non-significant SSD and Harpending's raggedness statistics indicate that the expansion models cannot be rejected. τ is the population expansion parameter. Significantly negative (Fu's F_s, Tajima's D) or positive (Ramos-Onsis & Rozas's R^2) values for neutrality tests are expected in cases of population expansion. SSD: Sum of Square Deviations; CI: confidence interval; R^2: Ramos-Onsis and Rozas's R^2. *: p-value < 0.05, ^{NS}: non-significant.

		Control Region n = 238	*cyt-b* n = 38
demo	SSD	0.0014^{NS}	0.0025^{NS}
	Harpending's raggedness	0.0028^{NS}	0.0258^{NS}
	τ (95% CI)	15.7 (11.7-23.6)	4.08 (2.69-5.18)
	time (years) since expansion (95% CI)	150,000 (110,000-220,000)	-
spatial	SSD	0.0019^{NS}	0.0024^{NS}
	Harpending's raggedness	0.0028^{NS}	0.0258^{NS}
	τ (95% CI)	14.4 (11.0-19.8)	4.08 (2.69-5.18)
	time (years) since expansion (95% CI)	130,000 (100,000-180,000)	-
neutral tests	Fu's F_s	-58.5*	-15.9*
	Tajima's D	-0.23^{NS}	-1.86*
	Ramos-Onsis & Rozas's R^2	0.078^{NS}	0.052*

Supplementary material

Table S1: Individual-based sampling information, GENBANK accession numbers and identified haplotypes (numbers refer to Figure 1).

Table S2: PCR primers used in this study.

Table S3: Pairwise Φst matrix and exact tests of differentiation.

Figure S1: Amplification strategy for the mitochondrial control region.