The maternal history of the sable antelope (*Hippotragus niger*) inferred from the genomic analysis of complete mitochondrial sequences

Joana Maria Laranjeira Rocha
Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia
2014

Orientador
Raquel Godinho
Assistant Researcher at CIBIO/InBIO
Invited Assistant Professor at University of Porto

Co-orientador
Pedro Vaz Pinto
Ph.D. student at CIBIO/InBIO
Researcher at ISCED – Huíla, Angola

Todas as correcções determinadas pelo júri, e só essas, foram efectuadas.
O Presidente do Júri,

Porto, _____/_____/_____

“It is a cursed evil to any man to become as absorbed in any subject as I am in mine”

Charles Darwin

Acknowledgments

First I would like to thank to my supervisor Dr. Raquel Godinho for accepting and trusting me with such an outstanding project, and also for her extraordinary guidance. Thank you, Raquel, for dealing with my overachieving paranoia, my fears and insecurities. Thank you for being patient with me and trusting me so fiercely. I could not have found a better supervisor, or a better friend.

Secondly I would like to thank to Pedro Vaz Pinto, for giving me the chance to study a species so close to his heart and for all the support given as my co-supervisor. Thank you Pedro for all the discussions, extensive email exchanging and strong enthusiasm.

It is imperative to thank to Professor Nuno Ferrand, for the warm welcoming to CIBIO and for his wise advices during the most difficult and undecided times since the beginning of this master thesis. In particular, I would like to thank him for the life-changing opportunity to work abroad at the Max Planck Institute for Evolutionary Anthropology (MPI-EVA) in Leipzig.

I would also like to thank to Dr. Matthias Meyer for receiving me at his group at the MPI-EVA. In particular, for providing all the means necessary for me being able to do my work in the most comfortable and friendly working environment a student could possibly wish for. I would also like to thank him for his guidance, suggestions, and for taking the time to sit with me and discuss my project. I never felt treated differently than any of his PhD students and for that I will always feel grateful.

I would also like to thank to the technical staff of CTM-CIBIO. In particular, to Susana Lopes for making me fell at ease to come to her in any circumstances, to Diana Castro for ordering and putting together all the museum samples I needed to take to Leipzig, and Patricia Ribeiro, who helped me a lot during my bachelor's thesis, introducing me to all basic laboratorial techniques, and for teaching me skills that proved to be most valuable during the time I spent in the lab at the MPI-EVA. I would also like to thank to Diana Lobo for all the help given from Portugal while I was in Germany, and for being always available.

I would like to thank to the Smithsonian institute (Washington, DC) and to the natural history museum of Vila Viçosa (Portugal) for providing most of the historical samples that integrated the genetic dataset used in this study. I would also like to thank to Banco BES (Portugal) and to ExxonMobil Foundation (USA) for financing all the projects related to the sable antelope and, therefore, allowing me to be a part of a small fraction of these projects.

I would also like to thank to my Max Planck friends and colleagues. Thank you Viviane and Mateja for making me feel comfortable enough to come to you with any questions about the lab, in particular Viviane for supervising me during the final step of lab proceedings. Thank you, Alex, for being a nice and quiet desk mate. Thank you Amin for reminding me that Thursday is climbing day. Thank you Philip for not letting me stand alone as the only masters' student in our group, and for going for a swim with me once in a while. Thank you Manjusha and Rohit for letting me taste the most delicious Indian food ever, and Manjusha for letting me try your sari.

I would like to particularly thank to Ana T. Duggan for all the advices given during my staying at the MPI-EVA, and for our email exchanges during the summer to discuss my project, talk about analyses and for motivating me during the final steps of writing. Thank you Ana for your major contribution and suggestions. I could not ask for a better colleague.

I would also like to particularly thank Gabriel for all the support given during the bioinformatics part, for our discussions about the best way of assembling the data, for letting me be the first person to use schmutzi (which I have no doubt it will be a tremendous success in the field of bioinformatics), and for our late night conversations at the institute. Thank you for not letting me stand alone as the most workaholic person in that department. Thank you for being an amazing colleague, and even better friend.

If there is someone that has contributed enormously for the success of this project, even without being directly involved in it, was Homa Papoli. Thank you, Homa, for receiving me so well in our home at Ostraβe, during a time of my life where everything seemed lost. Thank you for inviting me for a tea in your Persian red carpet, for teaching me your beautiful language, for all the intellectual discussions and Poirot movies, for the diners and happy moments at the city center, for the opera, for the classic concerts, and walking on the park. You became my new family while my old one was falling apart. Thank you.

I would also like to thank Alice for dealing with my stress during the final countdown of the thesis' delivery. Thank you for the help given with minor, yet extremely important, details. Most importantly, thank you for being my friend and never let me down.

Finally I would like to thank to my friends in Porto, master colleagues and family for all the support given during this year. In particular to Pedro, Diogo and Maria João for being at the airport with wide open arms to receive me back in Porto, and for helping me in re-adapting to my life in Porto once again.

Abstract

The sable antelope is a savanna woodland species with a distribution confined to the central, southern and eastern part of the African continent. Previous studies focused on assessing the patterns of genetic variation within *H. niger* have been limited by the scant resolution associated to analyzing short fragments of the mitochondrial genome, and by the lack of representative specimens covering the species whole geographic range in Africa. This has foreclosed the accurate estimation of divergence times between the different sable antelope subspecies and, thus, the explicit testing of phylogeographic hypotheses regarding the species evolutionary patterns. With the current development of target-enrichment methods for next-generation sequencing, whole mitochondrial genomes are increasingly being used for population genetics and phylogeography. The ability to provide higher levels of variability and resolution than those obtained when using mtDNA fragments, has been having dramatic impact in solving long-standing questions about the evolutionary history of iconic species. In this study, we made use of next-generation sequencing methods to capture 215 complete mitochondrial genomes from a representative set of modern samples, and samples from African museum collections, covering the species whole geographic range. This allowed us to generate the most comprehensive dataset of *H. niger* to date and, therefore, to re-assess the phylogeographic history of the species in Africa with an unprecedented accuracy. In contrast to previous studies that failed to explain the deep divergence found between a clade of sables from western Tanzanian populations and the remainder *H. niger*, our study suggests that this highly divergent clade is the result of an old mitochondrial introgression from a currently extint *Hippotragus* species that diverged from *H. niger* around 1.7 mya. The complete absence of nuclear introgression, in contrast to mtDNA, suggests that strong gender-related asymmetries were involved. Our results also point to climatic shifts and geomorphological changes occurring during the Pleistocene as responsible for most of the species phylogeographic history in central, southern and eastern Africa. This allowed us to formulate an explicit hypothesis to explain the close phylogenetic relationship between the critically endangered giant sable antelope of Angola and sables from western Tanzania and Malawi. We believe that the ancestral population(s) to these sables might have been expanded in savanna habitat regions in between their current distribution, presumably in the Congo basin or northern Zambia, during a glacial period around 150 thousand years ago, followed by dispersal into northern Angola in the subsequent warm period, in which rainforests would have displaced savanna habitats. Ultimately, our study could integrate future comparative phylogeographic studies that would allow the reconstruction of the major trends in the histories of vicariance and dispersal in central, southern and eastern Africa, which are still poorly understood. We believe this is a

compelling contribution to phylogeography in Africa as most of available literature is focused on study systems from the Northern Hemisphere. Overall, the results obtained in this study can be used as empirical evidence that the use of complete mitochondrial genomes allows achieving higher phylogenetic resolution, and more accurate divergence estimates to assess intraspecific relationships. Therefore, we strongly advise the use of complete mitochondrial genomes in studies that wish to address complex intraspecific relationships or simply to provide higher phylogenetic resolution, and more precise divergence estimates, than those obtained using traditional shorter fragments.

Keywords: sable antelope, complete mitochondrial genomes, next-generation sequencing, phylogeography, introgression, Pleistocene.

Resumo

A palanca negra (*Hippotragus niger*) é uma espécie de antílope de savana cuja distribuição geográfica está confinada ao centro, sul e leste do continente Africano. Entre as principais limitações dos estudos anteriores que analisaram os padrões de diversidade genética em *H. niger* contam-se a fraca resolução associada à sequenciação de pequenos fragmentos do genoma mitocondrial e a falta de amostras representativas da distribuição geográfica da espécie em Africa. Estas limitações têm impedido a estimativa de tempos de divergência entre as diferentes subespécies de palanca negra e, consequentemente, a avaliação de diferentes hipóteses filogeográficas que procuram explicar os padrões evolutivos da espécie. Com o crescente desenvolvimento de estratégias laboratoriais de captura para sequenciação de DNA com métodos de nova geração, o uso de genomas mitocondriais completos em estudos de genética populacional e filogeografia tem sido cada vez mais frequente. A capacidade de obter níveis cada vez maiores de variabilidade genética e resolução, em comparação com o uso de pequenos fragmentos mitocondriais, tem sido fulcral para resolver questões importantes acerca da história evolutiva de espécies icónicas. Neste estudo, usaram-se métodos de sequenciação de nova geração para capturar um total de 215 genomas mitocondriais de *H. niger* a partir de um conjunto representativo de amostras modernas e de colecções de museu que abrangem toda a distribuição geográfica da espécie. Com esta abordagem, foi possível gerar a mais completa base de dados de *H. niger* obtida até à data, e reavaliar a história filogeográfica da espécie em Africa com um nível de resolução sem precedentes. Ao contrário de estudos anteriores que não conseguiram explicar a elevada divergência entre um clado de palancas provenientes de populações a oeste da Tanzania (*west Tanzanian clade*) e as restantes populações de *H. niger*, o nosso estudo sugere que que este clado resulta de uma introgressão antiga que terá envolvido uma espécie de *Hippotragus* actualmente extinta, que divergiu de *H. niger* há cerca de 1.7 milhões de anos. A ausência de sinais de miscigenação no genoma nuclear sugere que fortes assimetrias associadas ao sexo terão estado envolvidas no processo de introgressão. O nosso estudo sugere ainda que a maior parte da história filogeográfica da espécie no centro, sul e leste de Africa pode ser explicada por mudanças climáticas e alterações geomorfológicas drásticas que terão ocorrido durante o Pleistoceno. Neste contexto, os nossos resultados permitem traçar uma hipótese explicativa da proximidade filogenética entre a palanca negra gigante de Angola, uma das espécies de antílopes mais ameaçadas de África, e plancas localizadas a oeste da Tanzania e Malawi. Segundo esta hipótese, uma população ancestral e comum a estas palancas terá estado expandida em regiões intermediárias à sua distribuição actual, presumivelmente Congo ou norte da Zambia, durante um período glacial de há 150 mil anos, em que os habitats de

savana teriam substituído florestas tropicais, com uma dispersão adicional para Angola no período quente subsequente, durante o qual as savanas terão dado lugar às florestas tropicais. No futuro, o nosso trabalho poderá ser integrado em estudos mais gerais de filogeografia comparada que procurem reconstruir os grandes padrões de vicariância e dispersão nas regiões pouco estudadas do centro, sul e leste de Africa. Este tipo de estudos representaria uma grande contribuição para a filogeografia em Africa, uma vez que a maior parte da literatura disponível está focada em espécies do Hemisfério Norte. Em conjunto, os resultados agora obtidos mostram que o aumento da resolução filogenética, e da exactidão dos tempos de divergência, obtidos com o uso de genomas mitocondriais completos, terão um papel decisivo no esclarecimento de relações intra-específicas complexas.

Palavras-chave: palanca negra, genomas mitocondriais completos, sequenciação de nova geração, filogeografia, introgressão, Pleistoceno.

Table of contents

List of Tables

Table 5_SI - Prior distributions used for Bayesian analyses of divergence times. Posterior distributions from Bibi (2013) were used as priors in the first phase of analyses (Phase I). The priors used for the second phase of analyses (Phase II) came from Phase I posteriors on the genus *Hippotragus* and on the highly divergent lineage from Tanzania/remainder *H. niger*. Mya=million years ago.

Table 6_SI – DNA polymorphism measures calculated for the complete mtDNA dataset, excluding a total of 429 bp of the hypervariable regions, and for the nuclear intron PGK1 dataset (*H. equinus* sequences included in both datasets).

Table 7_SI - Molecular diversity indexes for the complete mtDNA dataset, excluding a total of 429 bp of the hypervariable regions.

Table 8_SI - Genetic divergence between different *H. niger* haplogroups, in comparison to the outgroup (*H. equinus*), and the highly divergent haplogroup from Tanzania (*Hippotragus sp*.).

Table 9_SI - Molecular diversity indexes for the nuclear intron PGK1.

Table 10_SI – DNA polymorphism measures calculated for the dataset including complete mtDNA sequences from the highly divergent lineage from Tanzania (*Hippotragus sp.*).

Table 11_SI- Molecular diversity indexes for the dataset including complete mtDNA sequences from the highly divergent lineage from Tanzania (*Hippotragus sp*.).

Table 12_SI - Genetic Diversity summary statistics, and respective standard deviations, for the highly divergent complete mtDNA sequences from Tanzania.

Table 13_SI- DNA polymorphism measures for *H. niger* complete mitochondrial sequences.

Table 14_SI - Molecular diversity indexes for complete mtDNA sequences of *H. niger*.

Table 15_SI - Genetic diversity summary statistics, and respective standard deviations, for the different populations sampled, based on complete mtDNA genome sequences of *H. niger*.

Table 16_SI- Neutrality tests results for each haplogroup of *H. niger*. All tests were non-significant in all haplogroups (p-value > 0.05).

Table 17_SI- AMOVA results for a hierarchical arrangement in two groups: populations located easterly to the Rift Valley and populations located westerly to the eastern branch of the Rift Valley.

List of Figures

Fig. 2- Neighbor-Net network based on uncorrected patristic distances implemented in SplitsTree, excluding sites with insertion/deletions and missing data (N). *Hippotragus equinus* was used as outgroup. A highly divergent haplogroup of sables from Tanzania was labeled as *Hippotragus sp*. No concrete subspecific category was attributed to this haplogroup, given its high divergence from all other haplogroups (labeled as *niger*-like). Numbers next to branches indicate bootstrap values. Branch lengths are proportional to genetic distance. Scale bars represent sequence divergence. The different *H. niger* haplogroups and sub-haplogroups are highlighted in bold.

Fig. 3- Median time to the most recent common ancestor (TMRCA) – green dots - and 95% highest posterior density intervals (HPDI) –upper and lower bars - for the different *niger*-like haplogroups, and Temperature oscillations during the mid-to-late Pleistocene. The x-axis is represented by the splits between *niger*-like haplogroups (left-to right from the oldest to the more recent), and temperature (increasing left-to-right). Glacial periods, with colder and generally more arid conditions, are highlighted in blue. The y-axis represents time of divergence expressed in thousands of years (kya). Temperature oscillations graphic adapted from Berger and Bickert (1997) *in* Flagstad et al. (2001).

Fig. 4- Median-joining networks based on complete mtDNA sequences for each of the different *H. niger* haplogroups, and for the previously identified *Hippotragus sp.* All the networks were generated using PopART. Sites with insertion/deletions were not considered by the software, preventing the discrimination of some of the previously identified haplotypes, namely the second haplotype belonging to haplogroup E (Table 2). Haplogroup names are highlighted in bold. The numbers of mutations separating the different haplotypes are presented as hatch marks on top of the branches.

Fig. 5- Haplotype sharing among *H. niger* populations. The number of sequences is represented in curved brackets preceded by the population name. Cells are colored according to the different haplogroups. Cells with color gradients reflect haplogroup sharing at a given population. Numbers inside the cells represent the number of haplotypes. Empty cells reflect absence of haplotype sharing. Individuals from undetermined populations (*Zambia*) are not represented. For the five haplotypes found in Shimba-Hills, two belong to haplogroup E and three to ESe. For the three haplotypes found in Matesi, two belong to S1 and one to S2. For the three haplotypes found in Triangle two belong to S1 and three to S2.

III-General Discussion and Conclusions

Fig.1- Median-joining network of the PGK1 intron sequences, including all haplotypes of *H. niger* and one haplotype belonging to the outgroup, *H. equinus*. Given that the same dataset

than mtDNA was used, individuals were labeled accordingly to the respective mtDNA haplogroup. The individuals carrying the highly divergent haplogroup from Tanzania in mtDNA (*Hippotragus sp.*) share the same haplotypes with individuals belonging to the reminder *niger*-like mtDNA haplogroups, with no extreme intra-specific divergence being observed in the nuclear data.

Fig.2- Neighbor-Net network based on uncorrected patristic distances implemented in SplitsTree. Numbers next to branches indicate bootstrap values. Branch lengths are proportional to genetic distance. Scale bar represents sequence divergence. The different *H. niger* haplogroups and sub-haplogroups are highlighted in bold with their respective names and codes.

Fig.3- Genetic divergence between different *H. niger* haplogroups (*niger*-like): A-Average number of nucleotide substitutions per site between populations (Dxy), with Jukes and Cantor correction (JC), estimated for the control region and complete mitochondrial genomes (mitogenomes); B- Number of fixed differences estimated for the control region and complete mitochondrial genomes (mitogenomes). For abbreviations of haplogroup names, see Figure 2 from chapter II.

Supplementary Information

Fig 1_SI - Ratio of transitions per samples for a subset of both modern and historic samples. Historic samples are underlined in blue, while modern samples are underlined in red. The plots show that there are no significant differences between the transitions/transversions ratio of historic and modern samples. From these observations it is possible to infer that there is not a bias towards transitions in historic samples as a result of deamination.

Fig 2_SI - Divergence (left) and coverage (right) with respect to the position in the mitochondrial genome, using as example a sample with an average coverage of 432.242 (HN332 from Ugalla, Tanzania). These plots illustrate a typical case in which it is evident the extreme divergence at the control region (15421-16507 bp), which is coincident with a drop in coverage. Note that, apart from the control region, all positions in the remainder sections of the mitochondrial genome have a very high coverage (from 200 to 800).

Fig 3_SI - Median-joining network of the PGK1 intron sequences including all haplotypes of *H. niger* and one haplotype belonging to the outgroup, *H. equinus*. Given that the same dataset than mtDNA was used, individuals were labeled accordingly to the respective mtDNA haplogroup. The individuals carrying the highly divergent haplogroup from Tanzania in mtDNA (*Hippotragus sp.*) share the same haplotypes with individuals belonging to the

reminder *niger*-like mtDNA haplogroups, with no extreme intraspecific divergence being observed in the nuclear data.

Fig. 4_SI - Bayesian phylogenetic tree of Hippotraginae phylogeny of whole mitochondrial genomes (excepting the control region) from public databases (TreeBase repository of the Bovidae dataset published by Bibi (2013)) and new sequences from this study. Posterior probabilities for nodes of interest are indicated on top of the tree branches (note the these are not the same as branch nodes). Clades corresponding to *niger*-like haplogroups are highlighted in blue, while the highly divergent clade of sables from Tanzania (described by Pitra et al. (2002)) named as *Hippotragus sp.* is highlighted in grey.

Fig. 5_SI - Bayesian phylogenetic tree of 215 *H. niger* and 2 *H. equinus* whole mitochondrial genomes (excluding 429 bp of the HVR). Clades corresponding to *niger*-like haplogroups are highlighted in different colors, followed by the respective label. Posterior probabilities are indicated for nodes of interest.

Fig. 6_SI - Haplotype Frequency for *Hippotragus sp*. and haplogroup Tz in the seven sampled populations from Tanzania.

Fig. 7_SI - Multidimensional Scaling analysis for all *H. niger* populations. Only populations with more than 2 individuals were considered. Stress value was 0.2473, which is considered fit according to the multidimensional scaling stress evaluation table proposed by Sturrock and Rocha (2000).

Fig.8_SI - Potential distribution of the major vegetation zones of Africa in the present day (A), during the maximum of the penultimate glaciation (thought to be characterized by the same climatic conditions as the Last Glacial Maximum) (B) and during the moistest phase of the Eemian Interglacial (thought to be characterized by the same climatic conditions of the early Holocene, around 8,000 14C years ago) (C).

List of Abbreviations

A - Haplogroup including haplotypes from Angola

AMOVA - Analysis of Molecular Variance

BWA - Burrows-Wheeler Aligner

CR - Control Region

Cytb - cytochrome b

Da - Number of net of nucleotide substitutions per site between populations

DNA - Deoxyribonucleic acid

DnaSP - DNA Sequence Polymorphism

DRC - Democratic Republic of Congo

Dxy - Average number of nucleotides substitutions per site between populations

E - Haplogroup including haplotypes from Kenya

ESe - Haplogroup including haplotypes from Kenya and northern Mozambique

ESS - Effective Sample Size

Fd. - Number of fixed differences

h - Number of haplotypes

Hd - Haplotype diversity

HPDI - Highest Posterior Density Interval

HPRT - Hypoxanthine phosphoribosyltransferase 1

HVR - Hypervariable Regions

JC - Jukes and Cantor

Ma - sub-haplogroup including haplotypes from Malawi

MCC - Maximum Clade Credibility

MPD - Mean Pairwise Distance

MT - Mitochondria

mtDNA - Mitochondrial DNA

NGS - Next Generation Sequencing

NP – National Park

P1 - Primer pair Hn1377F-Hn6895R

P2 - Primer pair Hn6538F-Hn12024R

P3 - Primer pair Hn1058F-14756R

P4 - Primer pair Hn14637F-Hn3305R

PCR - Polymerase Chain Reaction

PDHA1 - pyruvate dehydrogenase alpha 1

PGK1 - phosphoglycerate kinase 1

RNA - Ribonucleic acid

S - Number of polymorphic/segregating sites

S1 - Haplogroup including haplotypes from Congo, Zambia, Zimbabwe and southern Mozambique

S1a - sub-haplogroup including haplotypes from Congo and Zambia

S1b - sub-haplogroup including haplotypes from Zimbabwe and southern Mozambique

S2 - Haplogroup including haplotypes from Zambia, Zimbabwe, southern Mozambique, Namibia and Botswana

S2a - sub-haplogroup including haplotypes from Zambia

S2b - sub-haplogroup including haplotypes from Zimbabwe, southern Mozambique, Namibia, Botswana

Sd - Standard deviation

SNP - Single Nucleotide Polymorphism

TMRCA - Time to the Most Recent Common Ancestor

Tz - sub-haplogroup including haplotypes from Tanzania

π - Nucleotide diversity

I- General Introduction

General Introduction

1. Sequencing technologies: the Next Generation

Next Generation Sequencing (NGS), or massive parallel sequencing technologies, are driving the transformation of population genetics into population genomics. These technologies have led to the development of robust laboratorial protocols for obtaining highly informative genome-wide massive sequencing data, and created the need for increasingly sophisticated tools for data analysis. This has allowed full understanding of the evolutionary history of many organisms at a cost-effective genome-wide scale, and is dramatically changing the way we see other species and biodiversity as a whole.

1.1-Next Generation sequences technologies: principles, advantages and current applications

In 1977, Sanger and Coulson published a method for DNA sequence determination (Sanger et al., 1977; Sanger, 1977) that remained the most commonly used approach for DNA sequencing until the past few years (Morozova and Marra, 2008). The advances in the Sanger sequencing technology, based on capillary electrophoresis of individual fluorescently labelled sequencing reaction products (Hunkapiller et al., 1991; Swerdlow and Gesteland, 1990), were instrumental for the completion of the human whole genome sequencing initiatives (Lander et al., 2001; Venter et al., 2001). However, using such methods for sequencing hundreds of complete genomes is extremely costly and inaccessible to most researchers.

Recently, the Sanger technology has been outperformed by alternative Next Generation Sequencing (NGS) technologies that require a specialized infrastructure of robotics, bioinformatics, computer databases and instrumentation (Mardis, 2008). These new sequencing technologies have brought an unprecedented speed to DNA sequencing, leading the way to new applications in molecular biology. They also have overcome the limited scalability, throughput, speed and resolution of the Sanger capillary sequencing, by using very different approaches that enable millions of sequencing reactions to be performed in parallel (Reis-Filho, 2009).

The most important difference that sets apart NGS methods from the Sanger method is that NGS provides millions of short reads (35-250 bp) in parallel, instead of the typical Sanger longer reads (650-800 bp). While in the capillary-based electrophoresis method the

sequencing reaction and nucleotide detection are decoupled, in NGS both sequencing and imaging occur at the same time in a step-by-step fashion (Mardis, 2011). This is followed by the assembly, in which reads are reassembled using a known reference genome as a scaffold (resequencing), or in the absence of a reference genome (*de novo* sequencing).

The implementation of NGS technologies has revolutionized many fields, including phylogenetics, population genetics and phylogeography, by allowing the transition from an analysis that was conventionally limited to few small fragments to whole molecule sequences. The ability to sequence the whole genome of related organisms has allowed, for the first time, large scale comparative and evolutionary studies that did not seem feasible before (Metzker, 2010). Examples of great achievements only possible after the implementation of NGS are the analysis of the nuclear genome from a 28000 year old mammoth (Poinar et al., 2006) or the Neanderthal genome (Green et al., 2006; Noonan et al., 2006).

Currently, the major applications of NGS platforms include: variant discovery by resequencing targeted regions of interest or whole genome *de novo* assemblies; discovery of transcription factor binding sites and gene regulation analysis; noncoding RNA expression profiling; SNP discovery; structural variation analysis; cataloguing the transcriptomes of cells, tissues and organisms; genome-wide profiling of epigenetic marks and chromatin structure; species classification and/or gene discovery by metagenomics studies (Petrosino et al., 2009); DNA-protein interaction analysis (Chip-Seq); and other applications never before seen by the Sanger sequencing (Morozova and Marra, 2008; Shendure and Ji, 2008). NGS technologies are also being used to characterize the evolutionary relationships of museum specimens and fossil remains, thereby providing a very important temporal perspective on the genetic diversity of past populations or species (Briggs et al., 2009; Cooper et al., 2001; Green et al., 2006; Meyer et al., 2013; Noonan et al., 2006; Poinar et al., 2006; Rogaev et al., 2006; Rohland et al., 2007, 2010; Stiller et al., 2009).

Although NGS enables the comprehensive analysis of genomes to be inexpensive, routine, and widespread, it has some limitations. These include the massive scale of the data to be analyzed, the significant decrease in the read length, and in the dramatically different error profiles of each read type, when compared to the Sanger technology (Mardis, 2011). So far, the major problem is the assembly of short reads, especially in areas of the genome associated with sequence repeats and insertion/deletions (indels) (Shendure and Ji, 2008). Nevertheless, there has been a tremendous effort on the development of sequencing approaches for the improvement of mapping efficiency, and on bioinformatic pipelines to

overcome most of these problems (Mardis, 2006, 2011; Miller et al., 2010; Pop and Salzberg, 2008).

Ultimately, there is no question that, despite its challenges, the continuous evolution of NGS technologies will allow, in a not so distant future, the study of any ecosystem of the earth at the sequence level. This will provide a much deeper understanding of the full spectrum of genetic variation in natural populations. Moreover, it will be possible to understand past genetic diversity through sequenced flora and fauna of at least the last 100,000 years (Schuster, 2008). For instance, Meyer et al., (2013) sequenced the oldest hominin DNA, from the fossilized leg bone of an early human from Spain who died 400,000 years ago. A short while ago, this would have been beyond any scientist's dream.

1.2-Target enrichment strategies for Next Generation Sequencing

Despite the lower costs provided by NGS technologies, whole-genome sequencing is still expensive, and the sequencing of large numbers of complex genomes is not yet entirely feasible (Metzker, 2010; Schloss, 2008). Alternatively, massively parallel sequencing can be combined with DNA capturing methods, also known as “target-enrichment” methods, for focused analysis of specific genomic regions of interest [e.g. exomes (Ng et al., 2009; Tang et al., 2012); and mitochondrial genomes (Briggs et al., 2009; Maricic et al., 2010; Stiller et al., 2009)]. This targeted sequencing is not only more time- and cost-effective, but also the resulting data are much easier to assemble, thereby enabling researches to focus time, expenses and data storage on the genome region of interest (Mamanova et al., 2010). This justifies the considerable effort to develop target-enrichment strategies in the past years, from both high-quality and ancient DNA samples (e.g. Briggs et al., 2009; Fu et al., 2013; Hodges et al., 2009; Lemmon et al., 2012; Maricic et al., 2010; McCormack et al., 2012; Stiller et al., 2009; Tang et al., 2012). Despite increasing sample preparation cost and time, this is an ingenious way of increasing the coverage of selected targets, reducing sequencing costs and saving sequencing time.

2. Mitogenomics

2.1- Mitochondrial DNA as a desirable genetic marker to infer the matrilineal history of mammals.

The animal mtDNA has been the genetic marker of election in most phylogeographic studies over the past 30 years, and has been essential to provide a much deeper understanding of evolution in natural populations (Avise, 2000; Wan et al., 2004). Phylogeographic studies traditionally use mitochondrial DNA because of the well-known advantages of this molecule, such as its ubiquitous presence in almost all animals, high copy number, lack of recombination, effective haploidity in DNA sequences, maternal inheritance, high mutation rate, extensive intraspecific polymorphism, and relatively small simple structured genome without the complicating features that characterize nuclear genomes, like repetitive DNA, transposable elements, pseudogenes and introns (Avise et al., 1987).

The fact the animal mtDNA has a high copy number and extranuclear cytoplasmic location makes this molecule much easier to isolate and analyze. This is particular relevant for analyzing ancient DNA and DNA extracted from museum specimens, making mtDNA the molecule of choice in studies dealing with highly degraded samples (Pakendorf & Stoneking, 2005). The maternal transmission mode of mtDNA is one of the greatest advantages of using this molecule as a genetic marker. This property of the mitochondria allows tracing lineages in time and space, and to estimate the matrilineal histories of populations without the misleading effects of biparental inheritance, and recombination, associated to the nuclear DNA (Pakendorf and Stoneking, 2005). The maternal inheritance also reduces the mtDNA effective population size, which means that variants are fixed much faster between speciation events (Curole and Kocher, 1999). In other words, the high mutation rate that characterizes the mtDNA allows for new character states to emerge within the lifespan of a species (Avise et al., 1987). This facilitates the inference of evolutionary patterns, especially in social species with a matrilineal group structure (Vilstrup et al., 2011).

The animal mtDNA is composed by 37 genes disposed in a closed circular molecule about 16 000 to 17 000 nucleotide pairs in length. The molecule also includes a noncoding region, the control region (CR), associated to mtDNA replication (Avise, 2009). The CR is likely to be the most appropriate region of the mitochondria for phylogeographic studies, because of its fast evolution rate and hypervariable properties (Wan et al., 2004). The rapid evolution rate of mtDNA control region mirrors the presence of high nucleotide sequence variation within most animal species, essential for phylogeographic analyses (Avise, 2009).

Many species have proved to exhibit strong geographically correlated patterns in the arrangements of mtDNA lineages (Avise et al., 1987). Nevertheless, results can be misleading for species with poor phylogeographic structure, unless large amounts of sequence data are provided (Hoelzel et al., 1998, 2002; Lippold et al., 2011). In many mammalian species, the variation seen in limited segments of the mitochondrial genome is often associated to high levels of recurrent mutations, which blurs the structure of phylogenetic trees and leads to the discovery of virtually impossible ancient branches within the trees (Achilli et al., 2012). Moreover, mutations accumulate at different rates in the different sections of the mitochondria, which limits accurate and precise estimations of divergence times in studies focused on a small portion of the genome (Meadows et al., 2011; Rohland et al., 2007).

2.2- The important role of mitogenomics in mammalian phylogenetics, population genetics and phylogeography

The analysis of nucleotide sequences from complete mitochondrial genomes, commonly referred as mitogenomics, has provided useful information on mammalian evolution (Arnason et al., 2008). To date, hundreds of mammalian species are represented by complete mitochondrial genomes. However, most of the analyses that have made use of these available mitochondrial genomes have been mainly focused in answering deep phylogenetic questions. For instance, the complete sequencing of mitochondrial genomes has been used to improve the resolution and statistical robustness of maternal phylogenies when single-gene estimates are poorly resolved between species (Alexander et al., 2013; Arnason et al., 1993, 2004, 2007, 2008; Chan et al., 2010; Chiou et al., 2011; Cooper et al., 2001; Duchêne et al., 2011; Finstermeier et al., 2013; Liedigk et al., 2012; Rogaev et al., 2006; Rohland et al., 2007; Vilstrup et al., 2011; Yu et al., 2007; Zinner et al., 2013). This has been particularly important for taxa with rapid evolutionary radiation like the Bovidae, Hylobatidae and Ursidae (Bibi, 2013; Chan et al., 2010; Yu et al., 2007). Furthermore, mitogenomic sequencing has been widely used in phylogenetics to produce more credible and robust estimates of divergence times (Arnason et al., 2008; Duchêne et al., 2011; Finstermeier et al., 2013; Zinner et al., 2013).

Although it has become increasingly common to sequence the whole mitochondrial genome for human population genetics studies, fewer are the cases that have made the same for population genetics and phylogeography of other animal species. With the current target-enrichment strategies for NGS, animal mitogenomes can be rapidly sequenced at lower costs, and are good candidates for population genetics and phylogeographic studies.

This has been made in several species and proved particularly relevant for phylogeographic studies that focused their analysis on shorter mitochondrial sequences, failing to provide a clear pattern of mtDNA variation. For example, the complete sequencing of mitogenomes has been useful to solve the low levels of mtDNA genetic diversity found in many Cetacean species, while phylogeographic studies based on the analysis of shorter mitochondrial fragments found no consistent geographical pattern of diversity [e.g. sperm whale (Alexander et al., 2013) and killer whale (Foote et al., 2011; Morin et al., 2010)]. The complete sequencing of mitochondrial genomes has also been proved useful in solving the low mtDNA variation of the rare carnivore species *Martes pennati* (Knaus et al., 2011), and the complex within-species relationships of cave bears (Stiller et al., 2009).

These examples illustrate the power of mitogenomics to improve genetic signal in intraspecific studies, thus solving challenging phylogeographic patterns in species with low mtDNA diversity. They also highlight the potential of complete mitochondrial DNA sequencing to address questions of divergence and diversity at a population scale, and not just at higher taxonomic levels. The power of mitogenomics can also be very useful for population analyses in species with low levels of nucleotide diversity within haplogroups (as shown in Shamblin et al., 2012). Furthermore, these examples show that particular regions of the mtDNA, such as the control region D-loop, or cytochrome b gene, not always accurately reflect the intraspecific relationships and variation of the complete mitochondrial genomes.

The study of complete mitochondrial genomes has also been proved useful to detect ancient substructure in early mtDNA lineages. This was the case of a study by Barbieri et al. (2013) with full mitochondrial genomes of southern African human populations. In this study parts of the haplogroup composition of ancestral populations that went extinct were incorporated in the gene pool of an extant population. This was also the case of another study using whole mitochondrial genome sequences of domestic horses, in which modern horses breeds were found to present a wide sample of mtDNA diversity found in ancestral, and currently extinct, wild horse populations (Lippold et al., 2011). Similarly, according to Achilli et al., (2008) the extinct wild aurochsen from Europe transmitted their mtDNA to domesticated taurine breeds and, thus, mitochondrial genomes of extinct aurochs survived in domestic cattle. In fact, mitochondrial genomes have been proved useful to reveal the major haplogroups that underwent domestication in modern horses and sheep, and in clarifying the relationships with their already extinct wild counterparts (Achilli et al., 2012; Lippold et al., 2011; Meadows et al., 2011). These examples highlight the relevance of studying complete mitochondrial genomes to decipher possible interactions between already extinct and

currently extant populations, or species, as well as to provide genetic evidence on domestication events.

Finally, the complete sequencing of mitochondrial genomes has also revolutionized the field of historical and ancient DNA, providing population genomics information from museum and paleontological specimens (Briggs et al., 2009; Cooper et al., 2001; Ho and Gilbert, 2010; Meyer et al., 2013; Rogaev et al., 2006; Rohland et al., 2007; Stiller et al., 2009). This has enabled scientists to answer long-standing questions of human evolution, and to uncover the matrilineal history of some iconic already extinct animal species, with an unprecedented power of analysis.

3. Introducing the Sable Antelope (*Hippotragus niger*)

The sable antelope (*Hippotragus niger*; Harris, 1838; Fig.1- A) is a member of the bovid subfamily Hippotraginae, along with its sister species, the roan antelope (*Hippotragus equinus*; Desmarest, 1804; Fig.1- B), and the gemsbok clade (*Oryx* sp.; Fig 1- C). The subfamily of Hippotraginae is one of the seven monophyletic subfamilies belonging to the Bovidae family (Order Artiodactyla), which additionally includes the earliest diverging Bovinae and the more recent Cephalophinae, Antilopinae, Caprinae, Reducinae and Alcelaphinae (Matthee and Davis, 2001). Both roan and sable antelope belong to the African endemic genus *Hippotragus*, and are widely distributed along the continent, living sometimes in sympatry (Kingdon, 1982; Skinner and Chimimba, 2005). The genus *Hippotragus* also includes a third species, the now extinct Blue Buck of South Africa (*Hippotragus leucophaeus*; Pallas, 1766; Fig. 1- D), which was the first large African antelope to disappear in historic times, in the end of the 18th century (Broom, 1949).

The sable antelope is amongst the largest of the African plains antelope, and is found widely scattered through the savanna woodlands of central, southern and eastern Africa, from east Africa and Mozambique across to the southern Congo Basin, with an isolated population in central Angola (Ansell, 1971; Kingdon, 1982). Although *H. niger* can occur in a variety of habitat types in the savanna biome, sables usually prefer open woodland with adjacent vleis of grassland, avoiding areas with high tree density and short grass (Skinner and Chimimba, 2005). The species' social organization is essentially matrilocal. While females remain with the herd for life, the males are evicted around the age of three by the territorial bull (Thompson, 1993).

Fig.1-Members of the bovid subfamily Hippotraginae and Genus *Hippotragus*: A-*Hippotragus niger* (Sable antelope); B-*Hippotragus equinus* (Roan antelope); C-*Oryx gazella* (gemsbok); D- Scientific illustration of the extinct *Hippotragus leucophaeus* (Blue buck), (A) http://www.panoramio.com/photo/83443393; (B) http://www.dianibeach.co.uk/Roan.jpg; (C) http://www.primack.net/animals/animal-pictures/oryx.jpg; (D) http://www.animalpicturesarchive.com/ArchOLD-6/1138370665.jpg

This species has been eliminated from large areas of its former range by intensive poaching, and loss of habitat due to the expansion of agricultural settlements (East, 1999). The horns and body stature of *H. niger* make this species a prized trophy animal to many big-game hunters. In order to continue to serve its purpose as a hunting trophy, the sable antelope has been reintroduced not only in many parts of its former range, but also in areas where it never occurred naturally (East, 1999). The translocation of *H. niger* has been particularly common within the subregion of southern Africa. The potential loss of genetic variability as a result of isolation on game farms poses an upcoming threat.

3.1- Sable antelope geographic distribution and recognized subspecies

The sable antelope is a savanna woodland species with a distribution confined to the central, southern and eastern part of the African continent (Skinner and Chimimba, 2005). Ansell (1971) recognized four subspecies of *H. niger*, each one representing different regions throughout the African woodlands: *H. n. variani* (Thomas, 1916), distinctively known as Giant Sable antelope (restricted area in northern Angola, between the Luando and Cuanza rivers), *H. n. kirkii* (Gray, 1872; Zambia, Malawi, Katanga province in the Democratic Republic of Congo and eastern Angola), *H. n. rooseveltii* (Heller 1910; southeastern Kenya, eastern Tanzania) and *H. n. niger* (Harris, 1838; southern African regions to the south of the Zambezi river). This recognition was based only on morphological characters, such as skull morphology, horn length and shape, presence or absence of a white stripe along the muzzle, body length and differences on the coat pigmentation (Ansell, 1971).

Apart from the isolated and very distinct *H. n. variani* subspecies, the exact limits and transitions between the recognized subspecies are extremely uncertain (Ansell, 1972). Groves (1983) suggested the existence of a fifth subspecies of sable antelope in eastern Zambia and Malawi, which he proposed to be recognized as *H. n. anselli* (Groves, 1983). This subspecies has been less accepted, and is usually not recognized in the current literature, going against the classification proposed by Ansell (1971) as *H. n. kirkii* (see also Ansell, 1978).

Despite several attempts to geographically delineate the different *H. niger* subspecies, the validity and geographical limits of such subspecies have not yet been strictly defined (Ansell, 1971, 1972, 1978; Kingdon, 1982; Roberts, 1951; and references therein). This is most likely due to inadequate sample size, to the failure to rigorously test variation in size and phenotypic markings, and to the lack of congruence and poor geographical representation in the original descriptions (Matthee and Robinson, 1999). Still, it is possible

to delineate the following approximate geographical range of *H. niger* in Africa according to the consensus distribution provided by most authors (Figure 2): *H. n. variani* in northern Angola; *H. n. rooseveltii* in southeastern Kenya and Tanzanian populations located easterly to the Great Rift Valley (Figure 3) and northern Mozambique; *H. n. kirkii* in Zambia, adjacent parts of DRC, eastern Angola, Malawi, and in western Tanzania between the eastern arc and the Albertine Rift (Figure 3); and *H. n. niger* in Zimbabwe, southern Mozambique, South Africa, Botswana, Namibia and southern Angola. Although this consensus distribution provides an overall accepted geographic range of *H. niger* in Africa, some uncertainties remain about the presence of some subspecies in poorly studied regions, namely of *H. n. kirkii* in western Tanzania and *H. n. rooseveltii* in northern Mozambique.

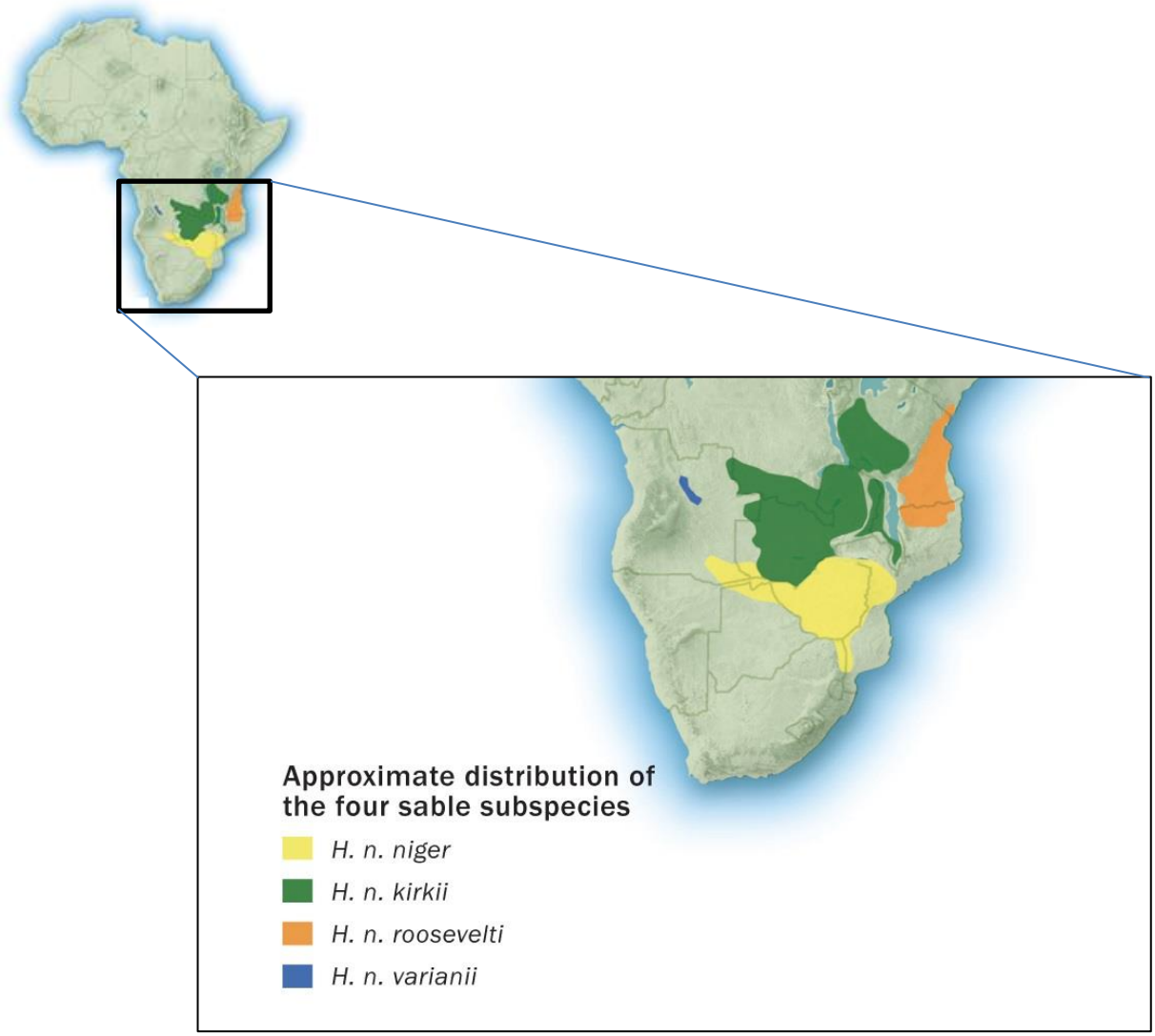

Fig. 2- Approximate geographic range of the four subspecies of *H. niger* in Africa according to the consensus distribution of most authors, including Roberts (1951), Ansell (1971, 1972, 1978), Kingdon (1982), and references therein. The *H. n. anselli* subspecies proposed by Groves (1983) is not represented, as it is not considered by most authors.

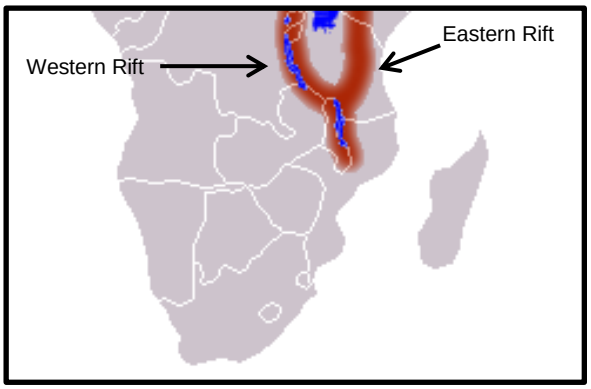

Fig. 3- East African Rift Valley. Western Rift is used to designate the western branch corresponding to the Albertine Rift. Eastern Rift is used to designate the eastern branch of the Great Rift Valley, known as eastern arc (adapted from http://upload.wikimedia.org/wikipedia/commons/5/55/Great_Rift_Valley.png).

3.2- A Giant among sable antelopes: the *Hippotragus niger variani* subspecies (Thomas, 1916)

The most iconic and distinct of the sable antelopes is the *Hippotragus niger variani* subspecies, known as the Giant Sable Antelope of Angola (Figure 4). This unique sable with breathtaking curved horns that can reach over 165 cm long was first discovered by Frank Varian in 1909, but it was only until seven years later that it was recognized as an entirely new subspecies of sable antelope (Thomas, 1916) (Walker, 2002).

The giant sable antelope was thought to be extinct after almost three decades of Angolan civil war. Nevertheless intensive field work, camera trapping and genetic evidence led to the rediscovery of this sable in 2005 (Pitra et al., 2006). Currently, this subspecies is classified by IUCN as *Critically Endangered* due to severe decline in population size and several other threats, including recent occurrences of hybridization with its sister species, the roan antelope (*H. equinus*) (IUCN, 2014). Among the major obstacles to the conservation of this subspecies are the current lack of protected area management, loss of habitat, demography and poaching (Kingdon et al., 2013). This giant among sables occurs in only two populations in northern Angola: one between the Luando and Kwanza rivers, and another in Cangandala (Figure 5) (Estes, 1971). Past estimations put the population at somewhere over 250 animals, which is around 10% of the 2000-3000 estimated to have survived in 1969/1970 (Skinner and Chimimba, 2005). Currently no more than 100 individuals survive in these populations (Kingdon et al., 2013). When the population of Cangandala was rediscovered in 2005 it was reduced to nine females, and conservation efforts had to be taken to translocate males from the Luando Strict National Reserve into the Cangandala National Park (Estes RD (2013) *Hippotragus niger* Sable Antelope in Kingdon et

al., 2013). In spite of recent successes in the conservation of the giant sable, its future remains uncertain. The evolutionary history and origin of the Giant Sable antelope remain to elucidate to date.

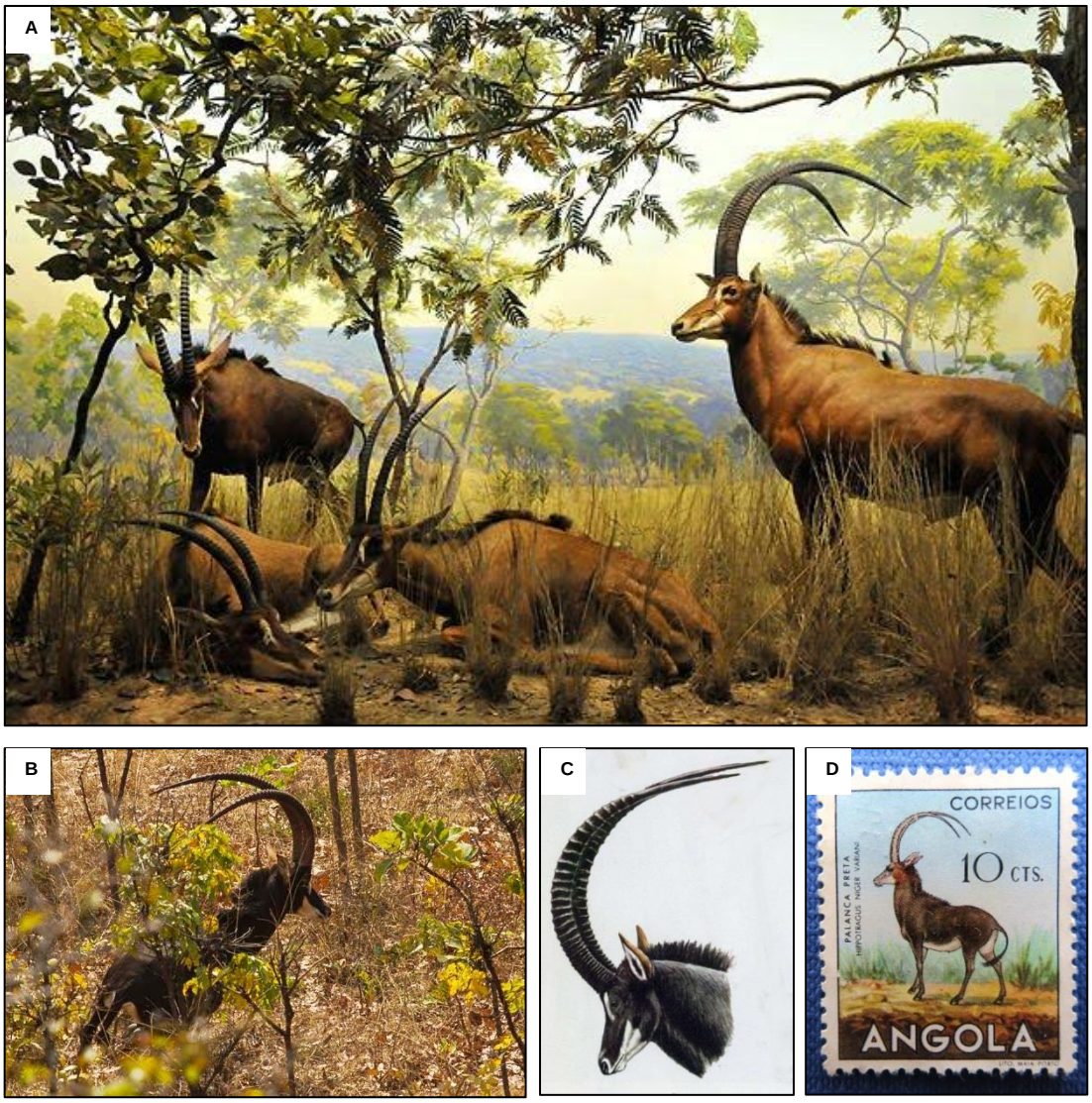

Fig.4- *Hippotragus niger variani* subspecies (giant sable antelope): A- Diorama showing a male and females in the American History Natural Museum, New York; B- Adult male in in the Angolan woodland savanna; C- Side view of an adult male (schematic representation); D- Angolan postal stamp. (A) Photo taken by Joana Rocha at the American History Natural Museum, New York; (B) Photo by Pedro Vaz Pinto in the Luando Natural Integral Reserve, Angola; (C) http://www.planet-mammiferes.org/Photos/Ongule/Antilope/HippNigT.jpg (D) http://www.arcadedarwin.com/wp-content/uploads/2013/01/palanca-preta-10cts.jpg;

Fig.5- Geographical location of the *Hippotragus niger varia*ni subspecies (giant gable antelope) in a restricted area between the Cuanza and Luando rivers in northern Angola - the Luando strict Natural Reserve - and in Cangandala National Park (obtained from the Africa Geographic Magazine, Part I, June 2010).

3.3– MtDNA Population Genetics and Phylogeography of the sable antelope

Ansell (1971) recognized four subspecies of *H. niger* and, later, several studies analyzed mitochondrial DNA (mtDNA) sequence variation, in order to further understand the phylogenetic and geographical relationships of *H. niger* in Africa. However, a clear picture of the evolutionary history of *H. niger* subspecies from different geographic areas still remains unclear.

In 1999, Matthee and Robinson analyzed mtDNA control region sequences in order to assess the relationships between the geographical dispersion and the genetic variation of *H. niger*. Their results showed that *H. niger* maternal lineages could be grouped in two major clades (Figure 6, a)): one represented by *H. niger* from west Tanzanian populations, and a second, geographically more diverse group including sables from Zambia, Malawi, southern Africa (Zimbabwe and South African samples) and Angola (Matthee and Robinson, 1999). The authors assigned sables from Zambia and Malawi to the *H. n. kirkii* subspecies and the southern African sables to *H. n. niger*, in accordance with Ansell's (1971) subspecies classification. Nonetheless, sables from west Tanzanian populations were doubtfully attributed to *H. n. rooseveltii.* This later subspecies had never been recognized westerly to

eastern branch of the Rift Valley, being traditionally known from southeastern Kenya and eastern Tanzania. Within the second and more geographically diverse clade (clade II; Fig. 6-a)), three less supported sub clades (<50% bootstrap) were additionally found: one grouping sables from Malawi and Angola together (II-1; Fig. 6-a)), another clade exclusively represented by sables from Zambia (II-2; Fig. 6-a)) and a third clade including sables from Zambia and southern Africa (II-3; Fig.6-a)). The authors attributed the failure to group sables from the second clade along subspecies boundaries, i.e. the lack of reciprocal monophyly in *H. n. variani*, *H. n. kirkii* and *H. n. niger*, to low sampling size that restricted phylogenetic resolution within this clade.

In order to better understand the phylogeographic patterns associated to *H. niger,* Pitra et al. (2002) used both control region and cytochrome b gene mtDNA sequences from a wider sample, including individuals from east Tanzania and southern Kenya, geographic regions that were not previously used, but did not include samples from Malawi, and of the critically endangered giant gable antelope subspecies (*H. n. variani*) from Angola (Figure 6, b)). According to this study, the sable antelope is apparently split into three major distinctive clades. One of these clades, represented by some west Tanzanian individuals, exhibited an extremely divergent lineage (Clade I; Fig. 6-b)). The two remaining clades, more related to each other, were represented by individuals from i) western Tanzania, Zambia and southern Africa (Clade II; Fig.6-b), and ii) Kenya and east Tanzania (Clade III; Fig.6-b). The authors found an extreme intraspecific differentiation of 18.2% between some mitochondrial haplotypes from the western Tanzania population (Clade I; Fig. 6-b)) and all other *H. niger* haplotypes (Clade II and III; Fig.6-b), including other haplotypes found in morphologically monotypic sable antelopes also from west Tanzania (Clade II; Fig.6-b)). This observation was interpreted as the result of an initial allopatric fragmentation between *West Tanzania (I)* and the remainder populations studied, followed by an unidirectional long-distance colonization from southern Africa (presumably Zambia) to west Tanzania, and was classified as “an exceptional case of historical outbreeding in African sable antelope populations” (Pitra et al., 2002).

In the end, the authors correctly ascribed the *H. n. roosevelti* subspecies to the eastern African sables. They also proposed to classify the highly divergent sable from west Tanzania as *H. n. kirkii*, and the remainder sables as *H. n. niger*. The latter included individuals from Zambia, the sables from southern African regions to the south of the Zambezi river, and the sables from *West Tanzania (II)* (as a result from introgressive hybridization that followed an unidirectional colonization of Zambian sables to west Tanzania). This new subspecies classification, not only goes against Ansell (1971) subspecific categories, but also is based in the assumption of an improbable fact such as the

outbreeding of *H. n. niger* into western Tanzania. Moreover, the authors could not explain how the extremely divergent maternal ancestry of the west Tanzanian sables attributed to *H. n. kirkii* is not reflected in any appropriate morphological feature (i.e. why *West Tanzania I*, classified as *H. n. kirkii*, and *West Tanzania II*, classified as *H. n. niger*, are morphologically monotypic).

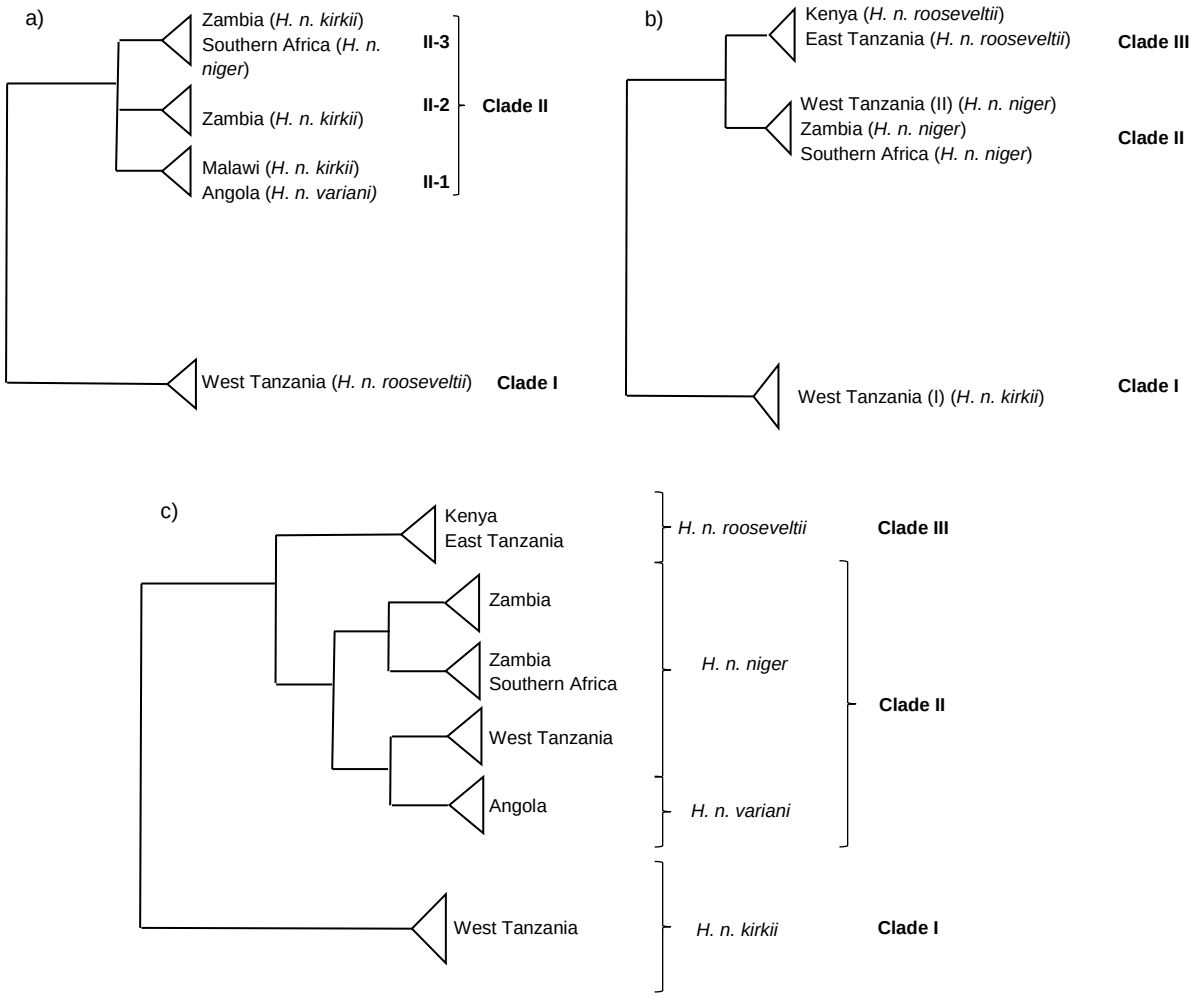

Fig. 6 – Schematic representation of the phylogenetic tree presented by Mathee and Robinson (1999), in which the west Tanzanian clade was wrongly attributed to *H. n. rooseveltii* (a); schematic representation of the phylogenetic tree presented by and Pitra et al. (2002) (b); and consensus schematic representation of the phylogenetic trees presented by and Pitra et al. (2006) and van Vuuren et al. (2010) (c).

The two subsequent studies that followed Pitra et al. (2002) work used mtDNA control region to infer the phylogenetic relationships of *H. niger*, this time including the previously unrepresented *H. n. variani* subspecies (Pitra et al., 2006; van Vuuren et al., 2010). Nonetheless they failed to include samples from Malawi, which so far has been only

documented in Matthee and Robinson (1999). Both studies adopted the same subspecies classification proposed by Pitra et al. (2002), and found the same tripartite pattern of genetic subdivision (Figure 2-c). The authors additionally found two sub-clades within clade II: one represented by sables from Zambia and southern Africa (*H. n. niger* subspecies), which they consider to be reciprocally monophyletic; and another not reciprocally monophyletic sub-clade represented by sables from western Tanzania (also ascribed to *H. n. niger*) and Angola (*H. n. variani*). The explanation given for the lack of reciprocal monophyly in this last sub-clade was that peripheral populations of *H. n. niger* in Angola and south Tanzania were possibly founded by long-distance colonizations from a common source population (Pitra et al., 2006). They supported this hypothesis with the argument that similar patterns were found in other savanna antelope species, probably shaped by climate fluctuations during the Pleistocene. Nonetheless they did not estimated divergence times that could validate, or not, such hypothesis.

Although the analyses of mtDNA fragments provided in these studies have contributed significantly to our understanding of the patterns of genetic variation and population structure of *H. niger*, conflicting and unresolved relationships remain. Among the main inconsistencies across studies are: the lack of representative samples from the species whole geographic range, with all failing to represent Congo and Mozambique, the proposed subspecies classification, and the intriguing intraspecific difference among the morphologically monotypic sable antelopes from west Tanzania. The origin of the giant sable antelope of Angola remains to elucidate to date.

The advances in DNA sequencing are making it possible to obtain much longer and diverse whole mitochondrial genome sequences, from increasingly larger numbers of individuals for population genetics and phylogeography. The advantages of using fully sequenced mitochondrial genomes include more accurate tree topologies and estimates of divergence times. In such context, the use of complete mitochondrial genomes, as opposed to shorter fragments, may help to resolve the presently unresolved phylogeographic structure of *H. niger* in Africa. It can also provide more data on causes and timing of divergence between lineages and intraspecific diversity.

4. Objectives and thesis structure

This study aims to make use of Illumina next generation sequencing to generate fully sequenced mitochondrial genomes from a large number of modern and historic samples, covering the whole geographic range of *H. niger* in Africa. The first major goal of using this approach is to reconstruct the complete maternal history of *H. niger* with an outstanding power of analysis, thereby providing a much deeper understanding of the phylogeographic structure of the species in Africa. The second goal is to hopefully contribute to solve the unknown origin of the critically endangered Angolan giant sable antelope, and also to solve the complex intraspecific relationships of *H. niger*.

In such context, it is intended:

1) to assess genetic diversity and associate it to a geographical context;
2) to assess population structure, past population dynamics and demography;
3) to assess phylogenetic relationships;
4) to estimate accurate divergence times between the major *H. niger* haplogroups or clades.

Additionally, this study aims to compare the information gained from the analysis of complete mitochondrial genome sequences with that obtained from shorter mtDNA fragments, to address questions of divergence and diversity at the population level.

This work is fully accomplished in Chapter II (The maternal history of the sable antelope inferred from the genomic analysis of complete mitochondrial sequences) in the form of a manuscript to be submitted to a specialized journal.

References

Achilli, A., Olivieri, A., Pellecchia, M., and Uboldi, C. (2008). Mitochondrial genomes of extinct aurochs survive in domestic cattle. Curr. Biol. 157–158.

Achilli, A., Olivieri, A., Soares, P., Lancioni, H., Hooshiar Kashani, B., Perego, U. a, Nergadze, S.G., Carossa, V., Santagostino, M., Capomaccio, S., et al. (2012). Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. Proc. Natl. Acad. Sci. U. S. A. *109*, 2449–2454.

Alexander, A., Steel, D., Slikas, B., Hoekzema, K., Carraher, C., Parks, M., Cronn, R., and Baker, C.S. (2013). Low diversity in the mitogenome of sperm whales revealed by next-generation sequencing. Genome Biol. Evol. *5*, 113–129.

Ansell, W. (1971). Order Artiodactyla. Mamm. Africa an Identif. Manual, Part 15 1–84.

Ansell, W. (1972). Order Artiodactyla. Part 15 of Meester, A.; Setzer, H.(Editors). The mammals of Africa: an identification manual. Washington DC. Smithson. Inst.

Ansell, W. (1978). The mammals of Zambia (National Parks & Wildlife Service.).

Arnason, U., Gullberg, a, and Widegren, B. (1993). Cetacean mitochondrial DNA control region: sequences of all extant baleen whales and two sperm whale species. Mol. Biol. Evol. *10*, 960–970.

Arnason, U., Gullberg, A., and Janke, A. (2004). Mitogenomic analyses provide new insights into cetacean origin and evolution. Gene *333*, 27–34.

Arnason, U., Gullberg, A., Janke, A., and Kullberg, M. (2007). Mitogenomic analyses of caniform relationships. Mol. Phylogenet. Evol. *45*, 863–874.

Arnason, U., Adegoke, J. a, Gullberg, A., Harley, E.H., Janke, A., and Kullberg, M. (2008). Mitogenomic relationships of placental mammals and molecular estimates of their divergences. Gene *421*, 37–51.

Avise, J.C. (2000). Phylogeography: The History and Formation of Species.

Avise, J.C. (2009). Phylogeography: retrospect and prospect. J. Biogeogr. *36*, 3–15.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., and Saunders, N.C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. *v. 18 p. 4*.

Barbieri, C., Vicente, M., Rocha, J., Mpoloka, S.W., Stoneking, M., and Pakendorf, B. (2013). Ancient substructure in early mtDNA lineages of southern Africa. Am. J. Hum. Genet. *92*, 285–292.

Berger, W., and Bickert, T. (1997). Reconstruction of atmospheric CO 2 from ice-core data and the deep-sea record of Ontong Java Plateau: the Milankovitch chron. Oceanogr. Lit. … 466–495.

Bibi, F. (2013). A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. BMC Evol. Biol. *13*, 166.

Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., et al. (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science *325*, 318–321.

Broom, R. (1949). The Extinct Blue Buck of South Africa : Abstract : Nature.

Chan, Y.C., Roos, C., Inoue-Murayama, M., Inoue, E., Shih, C.-C., Pei, K.J.-C., and Vigilant, L. (2010). Mitochondrial genome sequences effectively reveal the phylogeny of Hylobates gibbons. PLoS One *5*, e14419.

Chiou, K.L., Pozzi, L., Lynch Alfaro, J.W., and Di Fiore, A. (2011). Pleistocene diversification of living squirrel monkeys (Saimiri spp.) inferred from complete mitochondrial genome sequences. Mol. Phylogenet. Evol. *59*, 736–745.

Cooper, a, Lalueza-Fox, C., Anderson, S., Rambaut, a, Austin, J., and Ward, R. (2001). Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature *409*, 704–707.

Curole, J.P., and Kocher, T.D. (1999). Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol. Evol. *14*, 394–398.

Duchêne, S., Archer, F.I., Vilstrup, J., Caballero, S., and Morin, P. a (2011). Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. PLoS One *6*, e27138.

East, R. (1999). African antelope database 1998. IUCN/SSC Antelope Spec. Gr. 209–214.

Estes, R. (1971). The Giant Sable Antelope.

Finstermeier, K., Zinner, D., Brameier, M., Meyer, M., Kreuz, E., Hofreiter, M., and Roos, C. (2013). A mitogenomic phylogeny of living primates. PLoS One *8*, e69504.

Foote, A.D., Morin, P. a, Durban, J.W., Willerslev, E., Orlando, L., and Gilbert, M.T.P. (2011). Out of the Pacific and back again: insights into the matrilineal history of Pacific killer whale ecotypes. PLoS One *6*, e24980.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. a, Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. U. S. A. *110*, 2223–2227.

Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F., Du, L., Egholm, M., Rothberg, J.M., Paunovic, M., et al. (2006). Analysis of one million base pairs of Neanderthal DNA. Nature *444*, 330–336.

Groves, C. (1983). A new subspecies of sable antelope, Hippotragus niger (Harris 1838). Rev. Zool. Africaine 821–828.

Ho, S.Y.W., and Gilbert, M.T.P. (2010). Ancient mitogenomics. Mitochondrion *10*, 1–11.

Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Benjamin Gordon, D., Brizuela, L., Richard McCombie, W., and Hannon, G.J. (2009). Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nat. Protoc. *4*, 960–974.

Hoelzel, a R., Dahlheim, M., and Stern, S.J. (1998). Low genetic variation among killer whales (Orcinus orca) in the eastern north Pacific and genetic differentiation between foraging specialists. J. Hered. *89*, 121–128.

Hoelzel, a R., Natoli, A., Dahlheim, M.E., Olavarria, C., Baird, R.W., and Black, N. a (2002). Low worldwide genetic diversity in the killer whale (Orcinus orca): implications for demographic history. Proc. Biol. Sci. *269*, 1467–1473.

Hunkapiller, T., Kaiser, R., Koop, B., and Hood, L. (1991). Large-scale and automated DNA sequence determination. Science (80-.). *254*, 59–67.

Kingdon, J. (1982). Horse-like antelopes, Hippotragines. East African Mamm. an Atlas Evol. Africa 539–577.

Kingdon, J., Happold, D., Butynski, T., Hoffmann, M., Happold, M., and Kalina, J. (2013). Mammals of Africa, Volumes 1-6 (A&C Black).

Knaus, B.J., Cronn, R., Liston, A., Pilgrim, K., and Schwartz, M.K. (2011). Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. BMC Ecol. *11*, 10.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Lemmon, A.R., Emme, S.A., and Lemmon, E.M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. *61*, 727–744.

Liedigk, R., Yang, M., Jablonski, N.G., Momberg, F., Geissmann, T., Lwin, N., Hla, T.H., Liu, Z., Wong, B., Ming, L., et al. (2012). Evolutionary history of the odd-nosed monkeys and the phylogenetic position of the newly described Myanmar snub-nosed monkey Rhinopithecus strykeri. PLoS One *7*, e37418.

Lippold, S., Matzke, N.J., Reissmann, M., and Hofreiter, M. (2011). Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. BMC Evol. Biol. *11*, 328.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., and Turner, D.J. (2010). Target-enrichment strategies for next-generation sequencing. Nat. Methods *7*, 111–118.

Mardis, E.R. (2006). Anticipating the 1,000 dollar genome. Genome Biol. *7*, 112.

Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. Trends Genet. *24*, 133–141.

Mardis, E.R. (2011). A decade's perspective on DNA sequencing technology. Nature *470*, 198–203.

Maricic, T., Whitten, M., and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One *5*, e14004.

Matthee, C. a, and Davis, S.K. (2001). Molecular insights into the evolution of the family Bovidae: a nuclear DNA perspective. Mol. Biol. Evol. *18*, 1220–1230.

Matthee, C. a, and Robinson, T.J. (1999). Mitochondrial DNA population structure of roan and sable antelope: implications for the translocation and conservation of the species. Mol. Ecol. *8*, 227–238.

McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T., and Glenn, T.C. (2012). Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. *22*, 746–754.

Meadows, J.R.S., Hiendleder, S., and Kijas, J.W. (2011). Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. Heredity (Edinb). *106*, 700–706.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat. Rev. Genet. *11*, 31–46.

Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.-L., Martínez, I., Gracia, A., de Castro, J.M.B., Carbonell, E., et al. (2013). A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature *000*.

Miller, J.R., Koren, S., and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. Genomics *95*, 315–327.

Morin, P.A., Archer, F.I., Foote, A.D., Vilstrup, J., Allen, E.E., Wade, P., Durban, J., Parsons, K., Pitman, R., Li, L., et al. (2010). Complete mitochondrial genome phylogeographic analysis of killer whales (Orcinus orca) indicates multiple species. Genome Res. *20*, 908–916.

Morozova, O., and Marra, M. a (2008). Applications of next-generation sequencing technologies in functional genomics. Genomics *92*, 255–264.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. Nature *461*, 272–276.

Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J.K., et al. (2006). Sequencing and analysis of Neanderthal genomic DNA. Science *314*, 1113–1118.

Pakendorf, B., and Stoneking, M. (2005). Mitochondrial DNA and human evolution. Annu. Rev. Genomics Hum. Genet. *6*, 165–183.

Petrosino, J.F., Highlander, S., Luna, R.A., Gibbs, R.A., and Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. Clin. Chem. *55*, 856–866.

Pitra, C., Hansen, A.J., Lieckfeldt, D., and Arctander, P. (2002). An exceptional case of historical outbreeding in African sable antelope populations. Mol. Ecol. *11*, 1197–1208.

Pitra, C., VazPinto, P., O'Keeffe, B.W.J., Willows-Munro, S., Jansen van Vuuren, B., and Robinson, T.J. (2006). DNA-led rediscovery of the giant sable antelope in Angola. Eur. J. Wildl. Res. *52*, 145–152.

Poinar, H.N., Schwarz, C., Qi, J., Shapiro, B., Macphee, R.D.E., Buigues, B., Tikhonov, A., Huson, D.H., Tomsho, L.P., Auch, A., et al. (2006). Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science *311*, 392–394.

Pop, M., and Salzberg, S.L. (2008). Bioinformatics challenges of new sequencing technology. Trends Genet. *24*, 142–149.

Reis-Filho, J.S. (2009). Next-generation sequencing. Breast Cancer Res. *11 Suppl 3*, S12.

Roberts, A. (1951). The Mammals of South Africa. Mamm. South Africa.

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

Rohland, N., Malaspinas, A.-S., Pollack, J.L., Slatkin, M., Matheus, P., and Hofreiter, M. (2007). Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. PLoS Biol. *5*, e207.

Rohland, N., Reich, D., Mallick, S., Meyer, M., Green, R.E., Georgiadis, N.J., Roca, A.L., and Hofreiter, M. (2010). Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. PLoS Biol. *8*, e1000564.

Rokas, A., and Carroll, S.B. (2005). More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol. Biol. Evol. *22*, 1337–1344.

Sanger, F. (1977). DNA Sequencing with Chain-Terminating Inhibitors. Proc. Natl. Acad. Sci. *74*, 5463–5467.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M., and Smith, M. (1977). Nucleotide sequence of bacteriophage φX174 DNA. Nature *265*, 687–695.

Schloss, J.A. (2008). c o m m e n ta r y How to get genomes at one ten-thousandth the cost. *26*, 1113–1115.

Schuster, S.C. (2008). Next-generation sequencing transforms today ' s biology. *5*, 16–18.

Shamblin, B.M., Bjorndal, K. a, Bolten, A.B., Hillis-Starr, Z.M., Lundgren, I., Naro-Maciel, E., and Nairn, C.J. (2012). Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. Mol. Ecol. *21*, 2330–2340.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nat. Biotechnol. *26*, 1135–1145.

Skinner, J., and Chimimba, C. (2005). The mammals of the southern African sub-region (Cambridge University Press).

Stiller, M., Knapp, M., and Stenzel, U. (2009). Direct multiplex sequencing (DMPS)—a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. Genome Res. 1843–1848.

Sturrock, K., and Rocha, J. (2000). A Multidimensional Scaling Stress Evaluation Table. Field Methods *12*, 49–60.

Swerdlow, H., and Gesteland, R. (1990). Capillary gel electrophoresis for rapid, high resolution DNA sequencing. Nucleic Acids Res. *18*, 1415–1419.

Tang, W., Qian, D., Ahmad, S., Mattox, D., Todd, N.W., Han, H., Huang, S., Li, Y., Wang, Y., Li, H., et al. (2012). A low-cost exon capture method suitable for large-scale screening of genetic deafness by the massively-parallel sequencing approach. Genet. Test. Mol. Biomarkers *16*, 536–542.

Thompson, K. V. (1993). Aggressive behavior and dominance hierarchies in female sable antelope,Hippotragus niger: Implications for captive management. Zoo Biol. *12*, 189–202.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The Sequence of the Human Genome. *291*.

Vilstrup, J.T., Ho, S.Y., Foote, A.D., Morin, P. a, Kreb, D., Krützen, M., Parra, G.J., Robertson, K.M., de Stephanis, R., Verborgh, P., et al. (2011). Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. BMC Evol. Biol. *11*, 65.

vanVuuren, B.J., Robinson, T.J., VazPinto, P., Estes, R., and Matthee, C.A. (2010). Western Zambian sable: Are they a Geographic Extension of the Giant sable Antelope? South African J. Wildl. Res. *40*, 35–42.

Walker, J. (2002). A certain curve of horn. Atl. Mon. Press 477.

Wan, Q.-H., Wu, H., Fujihara, T., and Fang, S.-G. (2004). Which genetic marker for which conservation genetics issue? Electrophoresis *25*, 2165–2176.

Yu, L., Li, Y.W., Ryder, O. a, and Zhang, Y.-P. (2007). Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. BMC Evol. Biol. *7*, 198.

Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L.F., and Roos, C. (2013). Baboon phylogeny as inferred from complete mitochondrial genomes. Am. J. Phys. Anthropol. *150*, 133–140.

II- The maternal history of the sable antelope inferred from the genomic analysis of complete mitochondrial sequences

The maternal history of the sable antelope (*Hippotragus niger*) inferred from the genomic analysis of complete mitochondrial sequences

Abstract

The sable antelope (*Hippotragus niger*) is amongst the largest of the African plains antelopes, and is found widely scattered throughout the savanna woodlands of central, southern and eastern Africa. Although previous phylogeographic analyses using mtDNA fragments have contributed significantly to our understanding of the patterns of genetic variation of *H. niger* throughout the continent, conflicting and unresolved relationships remain. In this study we provide the most comprehensive dataset of *H. niger* to date covering the species whole geographic range in Africa. To this end, we made use of currently available target-enrichment strategies for next-generation sequencing to generate 215 complete mitochondrial genomes from living antelopes and from historic samples from African museum collections. With our larger sample size and wide geographic coverage, we have been able to elucidate the phylogenetic relationships, the geographic patterns of diversity, and the timing of divergence of the different *H. niger* haplogroups. This allowed explaining the presence of a previously described highly divergent lineage of west Tanzanian sables. We propose the hypothesis that the sables carrying this lineage experienced gene flow from a currently extint species that diverged from *H. niger* around 1.7 mya. We also hypothesize that major climatic shifts and geomorphological changes occurring during the Pleistocene might be responsible for most of the species phylogeographic history. This allowed us to formulate an explicit hypothesis to explain the close phylogenetic relationship between the critically endangered giant sable antelope of Angola and sables from western Tanzania and Malawi. We believe that the ancestral population(s) to these sables might have been expanded in savanna habitat regions in between their current distribution, presumably in the Congo basin or northern Zambia, during a glacial period around 150 thousand years ago, followed by dispersal into northern Angola. Ultimately, the detailed phylogeographic scenario provided for the species as a whole in a region can be used to illuminate the evolutionary history of other species, and to allow the explicit testing of phylogeographic hypotheses in the regions of central, southern and eastern Africa, which are still poorly understood. Ultimately, this study adds evidence that complete mitochondrial genomes are crucial to provide improved statistical phylogeography and more precise estimates of divergence times.

Introduction

The African continent has been subjected to major geographic and climatic changes in the deep past. These include mountain formation, rifting, volcanic eruptions and changes in vegetation caused by temperature oscillations, all of which led to profound variation in plant and animal communities (Steele, 2007) . Most authors suggest that during cold periods, arid savanna habitats were expanding, being replaced by tropical forests during warm and humid periods, only to expand again in subsequent cooler and drier periods (Cohen et al., 2007; Dupont, 2011). These marked oscillations between cool, dry periods with humid and warm phases have resulted in the repeated fragmentation of habitats, isolating plants and animals, yet increasing biodiversity (Hewitt, 2004a, 2004b). Although past temperature oscillations are well documented, particularly during the Pleistocene glacial cycles (Van Andel, 1996; Avise et al., 1998; deMenocal, 2004; Trauth et al., 2009), the impact of such changes on the fauna of central, southern and eastern Africa is still poorly understood (Beheregaray, 2008).

The sable antelope (*Hippotragus niger,* Harris, 1838) is a savanna woodland species with a distribution confined to the central, southern and eastern African sub regions (Ansell, 1971; Kingdon, 1982). Although the species can occur in different savanna habitats, it usually prefers open woodlands and grasslands, avoiding areas with high tree density and short grass (Skinner and Chimimba, 2005). There have been several attempts to geographically delineate different morphological subspecies within *H. niger.* Ansell (1971) recognized four subspecies with the following geographic distribution: *H. n. variani* (Thomas, 1916; restricted area in northern Angola, between the Luando and Cuanza rivers, and immediately north of Luando), *H. n. kirkii* (Gray, 1872; Zambia, Malawi and province of Katanga in the Democratic Republic of Congo), *H. n. roosevelti* (Heller 1910; southeastern Kenya, eastern Tanzania) and *H. n. niger* (Harris, 1838; southern African populations to the south of the Zambezi river). A fifth less accepted subspecies, *H. n. anselli*, was also recognized by Groves (1983) in Malawi. The most iconic and distinct of the sable antelopes is the *H. n. variani* subspecies, famously known as the *Critically Endangered* giant sable antelope of Angola (IUCN 2006-2014). In spite of recent success on the conservation of this giant among sables, its future remains uncertain and its past origins, so far, remain a mystery.

Previous studies analyzed mtDNA control region and cytochrome b sequence variation with limited success to determine the phylogenetic and phylogeographic relationships of *H. niger* in Africa (Matthee and Robinson, 1999; Pitra et al., 2002, 2006; van Vuuren et al., 2010). Although the analyses of mtDNA fragments provided in these studies

have contributed significantly to our understanding of the combined patterns of gene trees with geography, conflicting and unresolved relationships remain. Among the problems and inconsistencies across studies are: the lack of geographic representative samples, with all falling to represent Malawi, Democratic Republic of Congo, northern and southern Mozambique, and the proposed subspecies classification. Previous studies uncovered a highly divergent clade of west Tanzanian sables, which was either attributed to the *H. n. roosevelti* (Matthee and Robinson, 1999) or *H. n. kirkii* subspecies (Pitra et al., 2002, 2006; van Vuuren et al., 2010). In fact, one of these studies interpreted this extreme intraspecific differentiation as the result of an initial allopatric fragmentation between west Tanzanian sables and the remainder populations studied, followed by an unidirectional long-distance colonization from southern Africa to west Tanzania (Pitra et al., 2002). However, the authors could not explain how this extremely divergent maternal ancestry of the west Tanzanian sables, attributed to *H. n. kirkii,* is not reflected in any obvious morphological feature. Also, one study ascribed the *H. n. kirkii* subspecies to Zambia (Matthee and Robinson, 1999), in accordance with past classifications, while others named it *H. n. niger* (Pitra et al., 2002, 2006; van Vuuren et al., 2010). Another limitation in such studies is the scant resolution associated to analyzing only one or few mtDNA fragments. For example, none of the studies used the available mtDNA fragments to estimate divergence times within *H. niger* and, if that would be the case, they would likely be biased (as shown by Duchêne et al., 2011). This has foreclosed the reconstruction of the histories of vicariance and dispersal of *H. niger* across its geographic range that could, for example, help to elucidate the past origins of the iconic giant sable antelope. Finally, no nuclear data was included in the analyses, which is essential to provide and unbiased picture of the evolutionary history of any species, from both male and female genetic components. For instance, mtDNA patterns in species with sex-biased dispersal or a matrilocal group structure, such as the sable antelope, are usually misinterpreted in the absence of independent nuclear loci.

The matrilineal history of animals, and particularly of mammals, has started recently to be illuminated by the analysis of complete mitochondrial genomes (mitogenomics) (Arnason et al., 2008). The advantages of using fully sequenced mitochondrial genomes, as opposed to short mtDNA fragments, include more accurate tree topologies and more credible estimates of divergence times (Duchêne et al., 2011; Finstermeier et al., 2013; Liedigk et al., 2012; Zinner et al., 2013). Despite complete mitochondrial genomes (mitogenomes) are now available for hundreds of mammalian species, they have been mostly generated for deep phylogenetic analysis. To date, with the exception of humans, not many studies have made use of mitogenomics to provide insights on the spatial patterns of genetic variation within other animal species. The current development of target-enrichment strategies for next

generation sequencing is making it possible to obtain whole mitochondrial genome sequences, from increasingly larger numbers of individuals for population genetics and phylogeography (Morin et al., 2010). This has allowed solving complex within-species relationships (Stiller et al., 2009), and challenging phylogeographic patterns in species with very rapid radiations, or very low mtDNA variation (Shamblin et al., 2012). The use of whole mitochondrial genomes for phylogeography has also been proved relevant to decipher possible interactions between already extinct and currently extant species (Lippold et al., 2011), as well as to provide genetic evidence on domestication events (Achilli et al., 2012; Lippold et al., 2011; Meadows et al., 2011). Finally, the analyses of fully sequenced mitogenomes obtained from ancient and historical samples have been crucial to understand the matrilineal histories of currently extinct mammal species, such as the wholly mammoth (Krause et al., 2006), auroch (Edwards et al., 2010) and Tasmanian tiger (Miller et al., 2009).

In such context, the use of complete mitochondrial genomes may help to better clarify the presently unresolved phylogeographic structure of *H. niger* in Africa, and to provide insights on the causes and timing of divergence between lineages. Additionally, the inclusion of historic samples from African museum collections that could fill past sampling gaps would enable a more comprehensive assessment of the species phylogeographic patterns from a larger sampling size and wider geographic coverage. Finally, the use of independent nuclear genes, such as X- or Y- linked intron fragments, could also provide a valuable control to the often sex-biased mtDNA patterns that result from the maternally inherited nature of this genetic marker (Palumbi and Baker, 1994 Avise, 2009).

So far, a global scenario of the matrilineal history of *H. niger* is still unclear, and the origin of the critically endangered Angolan giant sable antelope remains a mystery. In this study, we intend to generate the most comprehensive dataset of sable antelope to date, based on hundreds of complete mitochondrial genomes, extracted from a large number of modern and historic samples covering the species full geographic range in Africa. Specifically, we want 1) to assess genetic diversity in a geographical explicit context; 2) to assess population structure, past population dynamics and demography; 3) to resolve the phylogenetic relationships within *H. niger*; and 4) to estimate accurate divergence times between haplogroups with a precision of temporal discrimination not previously possible. We also intend to use an X-linked nuclear intron fragment as a control to the mtDNA patterns. Ultimately, we believe that the detailed phylogeographic scenario here provided as a whole in a region – central, southern and eastern Africa – could be used to illuminate the evolutionary history of other species, and to allow the explicit testing of phylogeographic hypotheses under a scenario of past climatic oscillations and geomorphological changes.

Material and Methods

Sample collection and DNA extraction

A total of 233 modern samples were collected from several populations covering the whole geographic range of *H. niger* in Africa, with the exception of Congo and southern Mozambique (Figure 1; Table 1_SI in Suppl. Info.). Among these 233 modern samples are included two samples of *H. equinus* to outgroup *H. niger* in data analysis. Given the phylogeographic nature of this study, all samples were collected from native populations. Additionally, 33 historic samples were collected from different museums, two of which belonged to *H. equinus*. Such historic samples provide additional representativeness to the species geographic range, including Congo and southern Mozambique (Figure 1). Sample size, sample identification, associated museum and country of origin are described in detail in Table 2 of Supplementary Information (Table 2_SI). The museum specimens whose exact population of origin in Zambia was undetermined were labeled as *Zambia*.

For modern samples, total genomic DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit following manufacturer's instructions. Blood and tissue samples were preserved refrigerated before and after genomic DNA extraction. Historic samples were subjected to the ancient DNA extraction method proposed by Dabney et al., (2013), in a sterilized laboratory hood, using four extraction blanks. This was performed after grinding each sample with a mortar and pestle until 50 mg of fine-grained powder was obtained, as described in Rohland & Hofreiter (2007).

Production of double stranded indexed libraries

Genomic DNA extracted from 233 modern samples was used for double stranded library preparation, as described in Maricic et al. (2010) (Text 1_S1 in Suppl. Info.). For the 33 historic/museum samples, DNA libraries preparation followed Meyer and Kircher (2010), with the modifications described in Kircher et al. (2012), using the regular Ilumina multiplex adaptors. The double stranded libraries were then amplified and purified as described in Dabney and Meyer (2012), but with increased primer concentrations to obtain higher yields of end product (Text 2_SI in Suppl. Info.). These were not pooled for capture, as described in Fu et al. (2013).

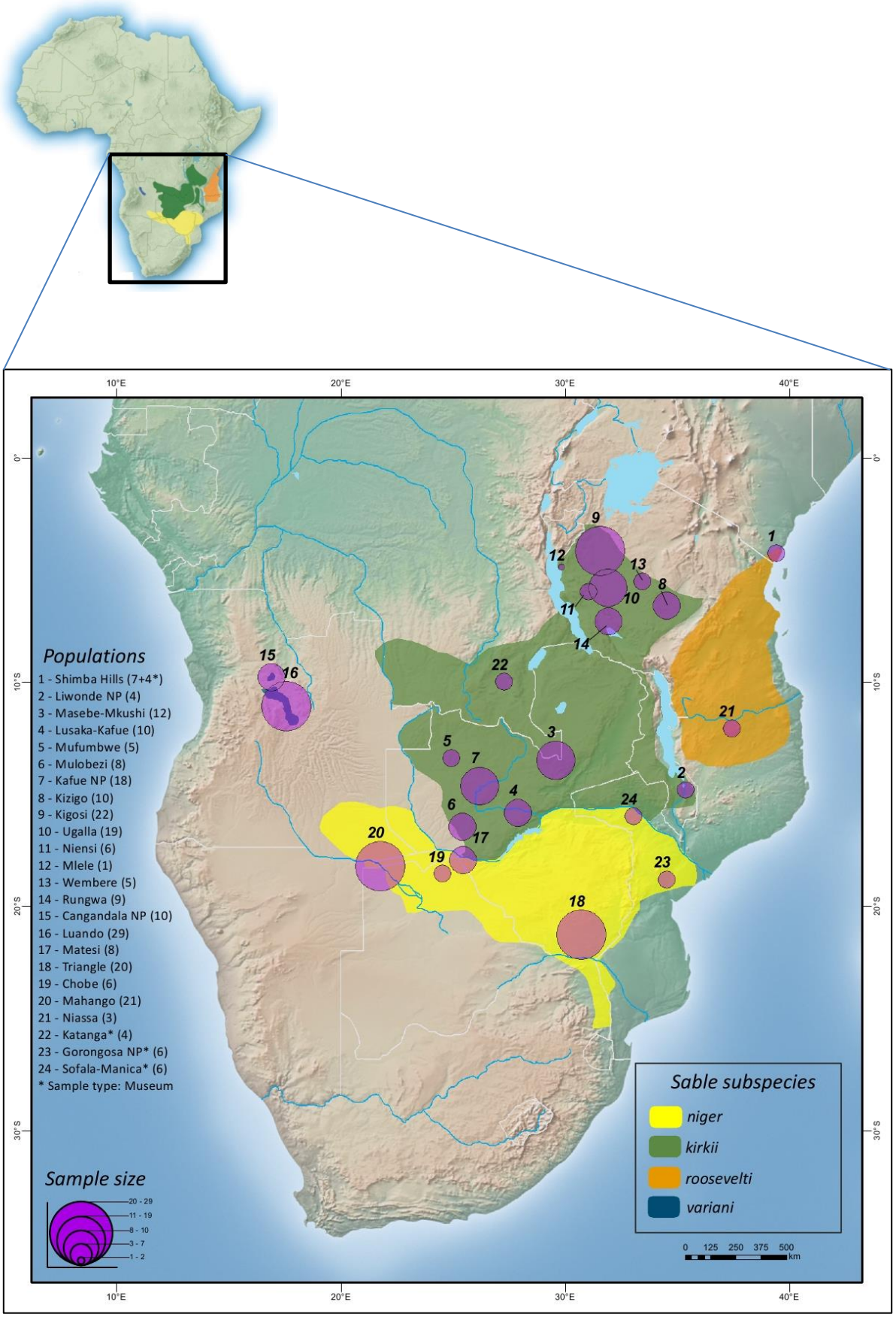

Fig.1 – Populations of *H. niger* sampled according to the different subspecies' distribution in Africa. Museum specimens from Zambia (additional 10 samples) with undetermined population are not represented in the map (see Table 2_SI in Suppl. Info.).

Bait Production

Four overlapping long-range PCR products encircling the sable antelope whole mitochondrial genome were produced, using the Expand Long Range dNTPack kit (Roche) in a 25 uL PCR reaction (Table 3_SI in Suppl. Info.), with the following primer pairs: Hn1377F (CTATGTGGCAAAATAGTGAG) and Hn6895R (TATGGTGTTGGCTTG AAACC), Hn6538F (ATAACATGAGCCAAAATCC) and Hn12024R (TATGATGGATCATGTGACG), Hn10583F (AACGTCTAAACGCCGGTCT) and Hn14756R (CCTGTGGGGTTGTTGGAG), Hn14637F (TGAGGGGGATTCTCCGTA) and Hn3305R (TGCTCGGTTTGTTTCTGC). These primers allowed the amplification of four long range PCR fragments of 5537 bp, 5504 bp, 4191 bp and 5193 bp, respectively (see Text 3_SI and Table 4_SI in Suppl. Info. for PCR conditions). The PCR fragments were purified using carboxyl-coated magnetic beads [0.1% Sera-Mag Speed Beads, ThermoScientific; 18% PEG-8000 (w/v) for modern samples and 12% PEG-8000 (w/v) for historic samples, Promega; 1M NaCl; 10 mM Tris-HCl pH 8.0; 1mM EDTA; 0.05% Tween-20], NanoDrop quantified (NanoDrop ND-1000, Thermo Scientific) and pooled in equimolar amounts to a total of 3 ug for modern samples, and to a total of 100 ug for historical samples. The pooled products were sheared in M220 Focused-ultrasonicator (Covaris) to produce fragments around 350 base-pairs in size. For modern samples, the products were ligated with Bio-T/B adapters, purified with the Qiaquick PCR purification Kit (Qiagen), NanoDrop quantified, made single stranded and immobilized on streptavidin-coated magnetic beads, exactly as described in Maricic et al. (2010). For historic samples, the products were ligated with Bio-T/B adapters, purified using carboxyl-coated magnetic beads, NanoDrop quantified, made single stranded and immobilized on streptavidin-coated magnetic beads using an updated protocol to Fu et al. (2013) (Text 4_SI in Suppl. Info.).

MtDNA capture and Sequencing

For modern samples, multiplexed capture of mtDNA sequences was performed as in Maricic et al. (2010). For historic samples, mtDNA capture was performed using a protocol updated from Fu et al. (2013), which allows for only one round of capture (Text 4_SI in Suppl. Info.). Modern DNA enriched libraries were sequenced on a single lane of an Ilumina flow cell with 76+7 cycles, following the manufacturer's instructions for Illumina Multiplex sequencing on the MiSeq platform, and using recipes for double-indexed paired-end sequencing (Kircher et al., 2012). The same sequencing strategy was performed on MiSeq for historic DNA enriched libraries, but with 130+7+130+7 cycles instead. Base calling was performed with Bustard (Illumina Inc.). Raw sequences were demultiplexed separated by sample using their index read. For both modern and historic samples, adapter trimming and merging of overlapping sequence stretches were processed using leehom (Renaud et al.,

2014). For historic samples, leehom was used with the "–ancientDNA" option and unmerged reads were discarded from downstream analyses. For paired-end reads merged into a single sequence, to reduce potential noise, only those with length greater than 35bp were retained.

MtDNA Sequence Assembly

The resulting reads from 233 modern and 33 historic samples were then mapped against the available *H. niger* complete mitochondrial genome (Genbank JN632648.1) using the mapper BWA v.0.5.10 (Li and Durbin, 2009). As the current version of the software does not allow for mapping to a circular reference, a customized mapping workflow was used (see Text 5_SI in Suppl. Info. for further details). Reads of outgroup samples were mapped against *H. equinus* complete mitochondrial genome (Genbank NC_020712.1).

For each sample, the consensus sequence was subsequently called using a Bayesian maximum *a posteriori* approach that incorporates mapping quality, base quality and, in the case of historic samples, deamination into the consensus call (see Text 6_SI and Figure 1_SI in Suppl. Info.). For each base, a probability of error was also produced. Due to the hyper-variability of the mtDNA control region D-loop, next generation sequencing (NGS) based approaches are likely to have reduced coverage due to the inability to map highly divergent reads. As expected, our results showed a dip in coverage for such region (see Figure 2_SI in Suppl. Info.). To account for this, primers L15910 (Hoelzel et al., 1991) and H16498 (Shields and Kocher, 1991) were used to amplify the complete mitochondrial control region (CR), and to sequence modern samples using Sanger technology (Text 7_SI in Suppl. Info.). This approach was not used for historical samples due to low quality DNA. Briefly, the unfiltered consensus and corresponding probabilities for all modern samples were used along their complementary Sanger sequence to produce the final mitochondrial sequence used for downstream analyses. Further details about the incorporation of the Sanger bases into the Illumina consensus call are found in Text 8_SI in Suppl. Info.

PCR amplification and Sanger sequencing of an X-linked intron fragment

The intron correspondent to the region between Exon 5 and 6 of the X-linked PGK1 gene (phosphoglycerate kinase 1) was PCR amplified, and the resulting fragments of 809 bp were sequenced using Sanger technology in a 3130XL genetic analyzer, following the BigDye Terminator v3.1 Cycle sequencing protocol. Amplifications and sequencing were made with primers newly designed by us: Exon5F-TAGAAGCCTTCCGAGCTTCA and Exon6R-GGCCTTGGCAAAGTAGTTCA (see Text 9_SI in Suppl. Info.). As this is unfeasible for museum samples due to the discontinuity of DNA fragments this approach was used for modern samples only.

Data analysis

The full set of 266 complete mitochondrial genomes were aligned using MUSCLE v3.8.31 (Edgar, 2004) and Mafft v.7.017b (Katoh and Standley, 2013) separately. The alignments were tested for the maximum likelihood-tree yielding higher log-likelihood scores, and for the most parsimonious tree using Phangorn R-package (Schliep, 2011), to select the best multiple sequence alignment for all downstream analyses. Individuals with masked bases (N) or poor quality were removed from downstream analyses. We used the following elimination criteria: 1) individuals with masked bases in more than 25% of the complete mitochondrial genome length were removed; 2) individuals with masked bases at polymorphic sites were removed, as long as it did not compromise the representativeness of the species geographic distribution. The final dataset was imported to DnaSP v.5.10 (Librado and Rozas, 2009) and translated to proteins. No stop codons were found in unexpected positions of the mitochondrial genome.

As a first step, we reconstructed the phylogenetic relationships within *H. niger*, using *H. equinus* as outgroup. We constructed a Neighbor-Net network (Bryant and Moulton, 2004) based on uncorrected patristic distances and bootstrap analysis with 1000 replicates using Split Tree v.4.13.1 (Huson and Bryant, 2006). For this, and for all downstream analyses using *H. equinus*, gaps were excluded and a total of 429bp from the hypervariable regions (HVR) I and II were removed, due to unresolved alignment between *H. niger* and *H. equinus* sequences. Descriptive summary statistics of DNA polymorphism were also estimated to better characterize our dataset. Genetic distance parameters were estimated with DnaSP, to assess divergence within *H. niger*, and divergence to the outgroup. The same analyses were performed on the nuclear intron dataset to compare with the mtDNA dataset. This was done after solving the phase ambiguity of the sequences using the algorithms provided by PHASE 2.1 (Stephens and Donnelly, 2003; Stephens et al., 2001) in DnaSP. The minimum number of recombinant events was evaluated by Hudson and Kaplan (1985) test.

Bayesian phylogenetic trees and estimates of the time of divergence were performed using BEAST v1.8.0 (Drummond and Rambaut, 2007). Best-fit substitution models were estimated with JMODELTEST v.2.1.5 (Darriba et al., 2012). Two sequential analyses were performed: first to estimate divergence times for the genus *Hippotragus*, and secondly for haplogroups within *H. niger*. Posterior distributions for divergence times of members of the tribes Hippotragini and Alcelaphini (Bibi, 2013) were used to estimate the time to the most recent common (TMRCA) ancestor of *Hippotragus*, using a set of samples representative of each of the different *H. niger* haplogroups (Text 10_SI in Suppl. Info.). The posterior distributions from this first analysis were then used as priors on a second phase to estimate

divergence times for *H. niger* using our generated dataset (excluding 429 bp of HVR I and II; Table 5_SI in Suppl. Info.). In the first phase of analysis the mtDNA dataset was analyzed as a single partition, using the GTR+G substitution model, as determined by Akaike's information criterion (AIC) in JMODELTEST, with strict molecular clock and a constant size coalescent tree prior. We used a strict molecular clock as preliminary analyses suggested it fits the data better than relaxed clock models (Text 11_SI in Suppl. Info.). We used 100 million total MCMC steps, with samples taken every 1000 steps. The same parameters were used for the second phase of analysis, but with the GTR+G+I substitution model instead. Acceptable mixing and convergence to the stationary distribution were checked with Tracer v.1.6.0. Effective sample sizes were all above 1000 in all clades for the two sequential analyses. After inspection with Tracer, we discarded appropriate number of steps as burn-in (20%), and combined the resulting tree samples for subsequent estimation of posteriors. Finally, maximum clade credibility (MCC) trees were estimated for each phase of analysis, using median-heights and a posterior probability limit of 0.5, with TreeAnnotator v.1.8.0.

In the second step of our analysis, we explored a dataset made exclusively of *H. niger* complete mitochondrial sequences (including the HVR). We estimated DNA polymorphism and genetic diversity parameters for the whole dataset, and for each of the haplogroups with ARLEQUIN v.3.11 (Excoffier et al., 2005). We next constructed a median-joining network for each haplogroup using PopART v.1 software (http://popart.otago.ac.nz/links.shtml) to assess the relationships between the different haplotypes. Haplogroup frequency and haplotype sharing in the different *H. niger* populations were also estimated. In order to test for geographic structure of the mtDNA diversity, an analysis focused on population relationships (multidimensional scaling analysis) was performed in Statistica version 10 (StatSoft Inc., 2011). For the same purpose, we conducted an analysis of molecular variance (AMOVA) in ARLEQUIN, with 10000 random permutations. We performed two types of hierarchical arrangements: one considering the population subdivision revealed by the median-joining networks, and another considering plausible geographic barriers to dispersal. The hierarchical arrangements that maximize the values of Φ_{CT} and are statistically significant indicate the most parsimonious geographical subdivisions. We also used ARLEQUIN to test for signatures of demographic expansion in *H. niger*, under the infinite-site model -by estimating Tajima's D (Tajima, 1989) and Fu's Fs (Fu, 1997) - and under the infinite-allele model, by performing the Ewens-Watterson test (Slatkin, 1996; Watterson, 1978). Mismatch distribution analyses within mitochondrial haplogroups were also carried on DnaSP. Finally, direct estimations of population parameters using the coalescent, such as the exponential growth parameter (g), were conducted using FLUCTUATE

v.1.4 program (http://evolution.gs.washington.edu/lamarc/fluctuate/) provided by LAMARC (Kuhner, 2006).

Results

I. Phylogenetic relationships of *H. niger*

Full-length mitochondrial genomes were sequenced and assembled for 262 *H. niger* samples, and for 4 samples of the outgroup species (*H. equinus)*. After filtering sequences with masked bases, we obtained a final dataset of 217 sequences (Text 12_SI in Suppl. Info.), from which a total of 429 bp of the HVR were excluded due to saturation and alignment ambiguities. This resulted in a 16114 bp dataset, with 1937 polymorphic sites when indels (insertion/deletions) and missing data (N) are excluded (see Table 6 and 7_SI in Suppl. Info.).

The most striking feature of the Neighbor-Net network generated from this dataset is the deep divergence between a group of Tanzanian sables and all other *H. niger* groups, including sables from Tanzania (Figure 2). Given the highly divergent nature of this lineage we named it *Hippotragus sp.* The Neighbor-Net network can be additionally divided in six major haplogroups, with high support by bootstrap values. These haplogroups can be easily distinguished by their geographic correlation with the eastern, central and southern African sub regions. Two haplogroups from east Africa, including haplotypes from Kenya (haplogroup E1, East Africa 1), and from Kenya and Northern Mozambique (haplogroup E2, East Africa 2), appear more closely related to each other, while highly divergent from the remainder haplogroups. Haplotypes from Angola (A) appear as one haplogroup, more closely related to a central-east African haplogroup (CE). This last haplogroup can be divided in two divergent sub-haplogroups: one composed by haplotypes originating from western Tanzania (Tz), and another from Malawi (Ma). Finally, the southern African haplogroup encompassing haplotypes from Congo, Zambia, Zimbabwe and southern Mozambique (S1) appears to be closely related to another southern African haplogroup englobing haplotypes from Zambia, Zimbabwe, southern Mozambique, and additionally found in Namibia and Botswana (S2). It is also possible to delineate sub-haplogroups within S1 and S2: two including haplotypes from Zambian populations (S1a and S2a), and the other two including haplotypes from populations located south to the Zambezi river (S1b and S2b).

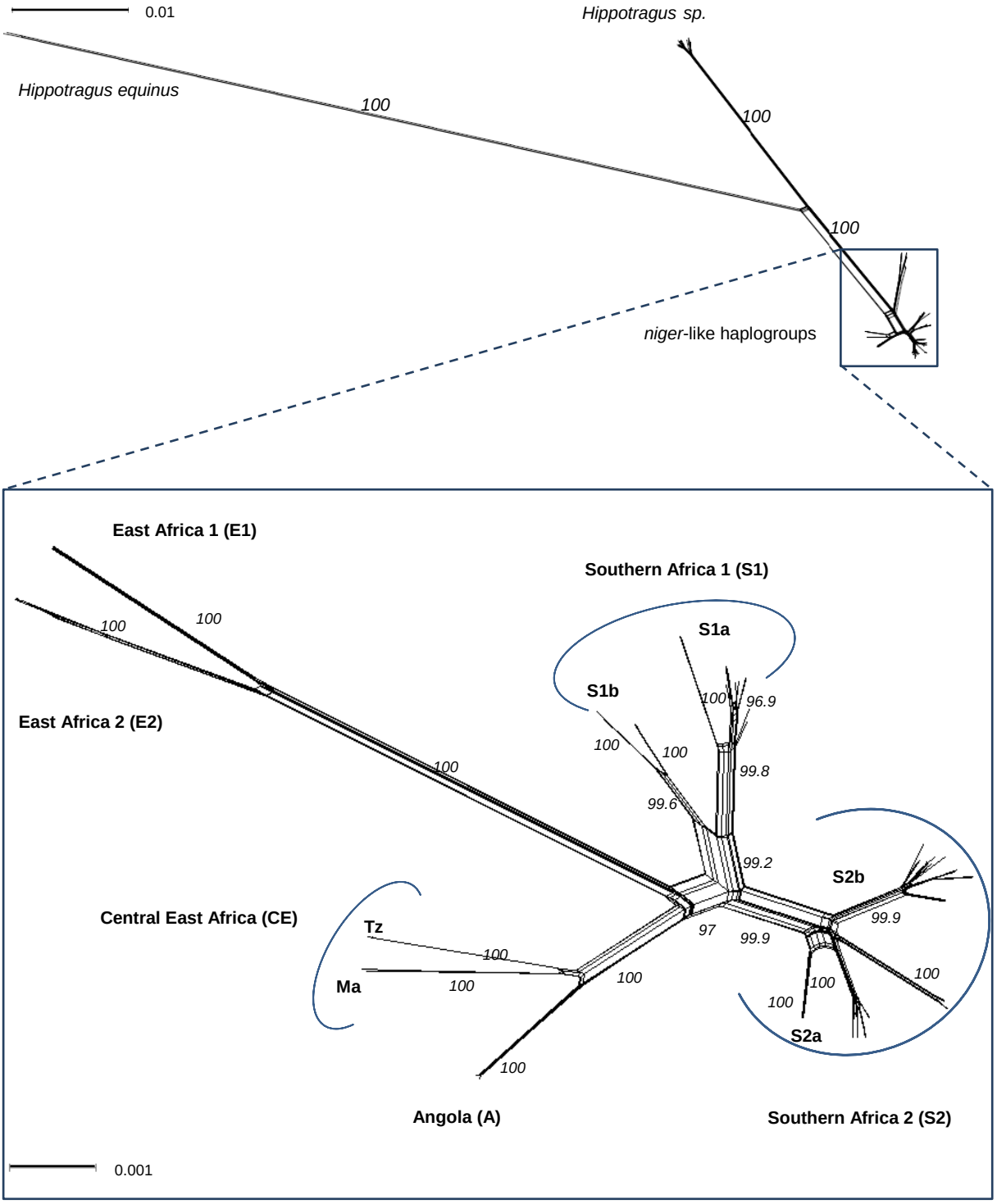

Fig. 2- Neighbor-Net network based on uncorrected patristic distances implemented in SplitsTree, excluding sites with insertion/deletions and missing data (N). *Hippotragus equinus* was used as outgroup. A highly divergent haplogroup of sables from Tanzania was labeled as *Hippotragus sp*. No concrete subspecific category was attributed to this haplogroup, given its high divergence from all other haplogroups (labeled as *niger*-like). Numbers next to branches indicate bootstrap values. Branch lengths are proportional to genetic distance. Scale bars represent sequence divergence. The different *H. niger* haplogroups and sub-haplogroups are highlighted in bold. Each of the *niger*-like haplogroups was labeled accordingly to the respective African region sub structuring: East Africa 1, East Africa 2, Central East Africa (containing sub-haplogroups from Tanzania -Tz- and Malawi - Ma), Southern Africa 1 and Southern Africa 2 (each containing sub-haplogroups a and b).

To evaluate the extent of the divergence between the different *H. niger* haplogroups, and from *Hippotragus sp*. and *H. equinus*, genetic distance measures and the number of fixed differences were calculated (Table 1; Table 8_SI in Suppl. Info.). According to the results from Table 1, the mean divergence between the *niger*-like haplogroups is, approximately, 0.7 %, while the divergence between *Hippotragus sp.* and the rest of *H. niger* is, approximately, 3 %.This clearly illustrates that there is no divergence among *niger*-like haplogroups that can be comparable to the highly divergent *Hippotragus sp*.

Table 1- Genetic divergences between the different *H. niger* mtDNA haplogroups, including *H. equinus* as outgroup.

Comparison	**Dxy (JC)**	**Da (JC)**	**Fd.**
Mean for *niger*-like haplogroups	0.00729 ± 0.00253	0.00631 ± 0.00272	86
niger-like haplogroups vs *Hippotragus sp.*	0.03436 ± 0.00140	0.03157 ± 0.00141	406
*nige*r-like haplogroups vs *H. equinus*	0.09134 ± 0.01550	0.08764 ± 0.01552	1211
Hippotragus sp. vs *H. equinus*	0.09342 ± 0.01987	0.09121 ± 0.01988	1350

Abbreviations are as follows: Average number of nucleotide substitutions (Dxy) and number of net nucleotide substitutions (Da) per site between populations, with Jukes and Cantor correction (JC); Number of fixed differences (Fd.)

The X-linked PGK1 intron fragment was selected as a control to the mtDNA phylogeographic patterns, and for possible gender-related asymmetries associated this genetic marker. We selected an X-linked gene because of the mode of inheritance of this chromosome, being single-copy in males and two-copy in females, allowing the definition of haplotypes in males and from that to resolve the phase ambiguity of females. The final dataset, trimmed for 660bp to standardize size for all samples, showed low genetic diversity, exhibiting only five haplotypes with seven polymorphic sites among the analyzed sequences, one of which belonged to *H. equinus* (Table 6 and 9_SI in Suppl. Info.). No recombinant haplotypes were observed among sequences. Despite most haplotypes grouped together, it is still expectable to observe clearly divergent haplotypes: one for the outgroup (*H. equinus*) and another for the highly divergent group of west Tanzanian sables (*Hippotragus sp*.). Although, a clear divergent haplotype was found for the outgroup, no nuclear haplotype reflecting the highly divergent mitochondrial haplogroup *Hippotragus sp.* was observed (Figure 3_SI in Suppl. Info.).

II. Divergence Times

Our estimated divergence times were in line with the fossil calibration points available in the literature (Bibi et al., 2013) (Table 5_SI and Text 11_SI in Suppl. Info.). Given the absence of conflicting calibrations in our analyses, we can confidently rely on the estimations obtained for the different *H. niger* haplogroups. Bayesian phylogenetic analyses also confirmed the same intraspecific relationships found in Figure 2 with high support by posterior values (Figure 4 and 5_SI in Suppl. Info.). The estimated divergence time between *H. niger* and *H. equinus* was around 6.7 million years ago (95% HPDI from 6.2 to 7.2 mya).

One of the most remarkable results from Figure 2 is the deep divergence of the *Hippotragus sp.* lineage from all other sables. The estimated divergence time between *Hippotragus sp.* and *niger*-like haplogroups was 1.73 mya (95% HPDI from 1.6 to 1.9 mya), back in the early Pleistocene, which contrasts with the time to the most recent common ancestor (TMRCA) of all *niger*-like haplogroups, around 422.9 thousand years ago (95% HPDI from 367.3 to 477.6 kya), later on the mid-Pleistocene. This last date also represents the split time between eastern haplogroups (E1 and E2) from the remainder haplogroups. Figure 3 summarizes the divergence times of the different *niger*-like haplogroups, with 95% highest posterior density interval (HPDI), and temperature oscillations between the last 50 and 500 kya.

Haplogroups from central east Africa (CE) and Angola (A) diverged from the southern African haplogroups (S1 and S2) around 244.8 kya (95% HPDI from 210.7 to 281.9 kya). Later, haplogroups E1 and E2 split from each other around 188.6 kya (95% HPDI from 148.6 to 230.7 kya). This is contemporary to the split between haplogroups S1 and S2 around 181.7 kya (95% HPDI from 152.6 to 211.9 kya), followed by the split between the Angolan haplogroup and CE around 145.5 kya (95% HPDI from 117.2 to 175.6 kya). The divergence time between sub-haplogroups S1a and S1b, around 128.1 kya (95% HPDI from 101.0 to 154.5 kya) is contemporary to the split between sub-haplogroups S2a and S2b, around 101.1 kya (95% HPDI from 81.0 to 125.3), and to the split time between sub-haplogroups Ma and Tz around 116.1 kya (95% HPDI from 90.3 to 142.8 kya). With the exception of the split time between Angola and CE, all haplogroups seem to have diverged during phases of warmer climate (pluvials or interglacials) that followed phases of cooler climate (interpluvials or glacial phases) during the mid-Pleistocene (Figure 3).

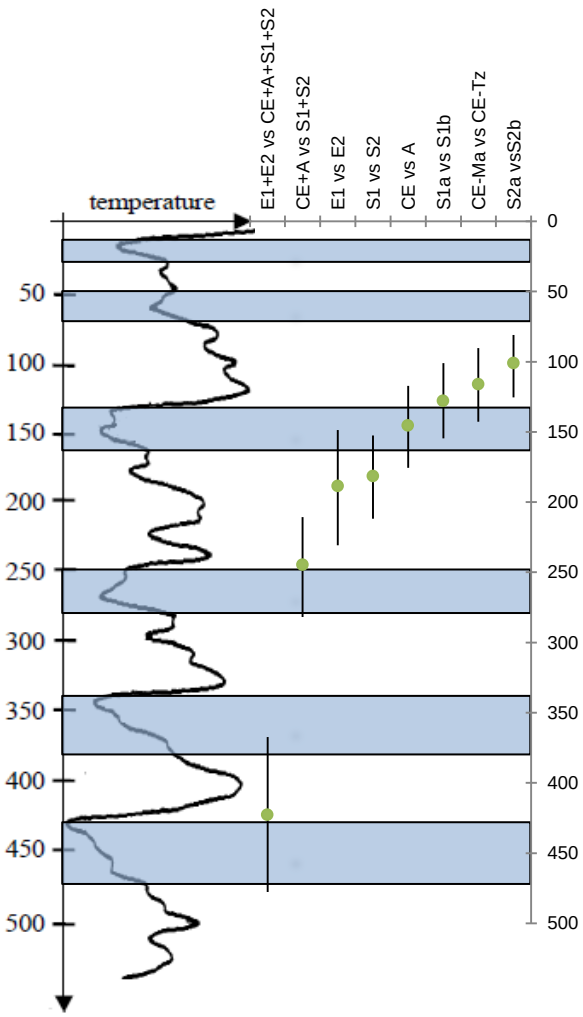

Fig. 3- Median time to the most recent common ancestor (TMRCA) – green dots - and 95% highest posterior density intervals (HPDI) –upper and lower bars - for the different *niger*-like haplogroups, and Temperature oscillations during the mid-to-late Pleistocene. The x-axis is represented by the splits between *niger*-like haplogroups (left-to right from the oldest to the more recent), and temperature (increasing left-to-right). Glacial periods, with colder and generally more arid conditions, are highlighted in blue. The y-axis represents time of divergence expressed in thousands of years (kya). Temperature oscillations graphic adapted from Berger and Bickert (1997) *in* Flagstad et al. (2001).

III. Phylogeographic and mtDNA population genetics analyses

Among 46 sequences from *Hippotragus sp.*, 29 haplotypes were found, presenting higher haplotype diversity than any of the *niger*-like haplogroups, including the west Tanzanian sub-haplogroup (Tz) from Central East Africa (Table 2). The number of sequences achieved in Tanzania for *Hippotragus sp.*, though, was higher than the *niger*-like Tanzanian sub-haplogroup (Tz). Sequences from the *Hippotragus sp.* dataset were further explored for their polymorphism, diversity as a whole and population diversity (Table 10, 11 and 12_SI in Suppl. Info.). The results show high genetic diversity in the seven sampled Tanzanian populations and consistently predominant frequencies of *Hippotragus sp.* relative to *niger*-like sub-haplogroup Tz, the latter restricted to Kizigo, Niensi and Rungwa (Figure 6_SI in Suppl. Info.).

Fifty seven haplotypes were observed among the 169 analyzed complete mitochondrial sequences belonging to *niger*-like haplogroups, having a total length of 16514 bp with 535 polymorphic sites, when indels are considered (Table 13 and 14_SI in Suppl. Info.). We can observe an increasing of genetic diversity towards the south of the African continent (Table 2): the haplogroup with lowest genetic diversity is found in east Africa (E1), while most of the genetic diversity of *H. niger* lies within southern Africa clades (S1 and S2). When considering the number of haplotypes relatively to the number of individuals, it is clear that the less diverse haplogroup is the one present in Angola, with only three haplotypes and two polymorphic sites out of 36 sequences. For the exact same number of sequences, we observed 19 haplotypes with 124 polymorphic sites within haplogroup S1. Genetic diversity statistics were also greater in this haplogroup. Additionally, haplogroups such as E2 and CE, which have 17% and 22% of the sequences of haplogroup A, present higher genetic diversity than Angola (Table 2).

Table 2- Genetic diversity summary statistics based on complete mtDNA genome sequences

Haplogroups	N	S	h	Hd	π	MPD
***Hippotragus* sp.**	46	136	29	0.966 ± 0.013	0.00166 ± 0.00082	27.370 ± 12.207
***niger*-like haplogroups**	169	535	57	0.952 ± 0.008	0.00547 ± 0.00261	90.268 ± 38.999
A	36	2	3	0.408 ± 0.086	0.00003 ± 0.00003	0.494 ± 0.433
E1	6	1	2	0.333 ± 0.215	0.00002 ± 0.00003	0.333 ± 0.380
E2	6	25	5	0.933 ± 0.122	0.00064 ± 0.00039	10.533 ± 5.606
CE	8	83	5	0.857 ± 0.108	0.00275 ± 0.00152	45.357 ± 22.064
Ma	4	1	2	0.500 ± 0.265	0.00003 ± 0.00004	0.500 ± 0.519
Tz	4	11	3	0.833 ± 0.222	0.00043 ± 0.00031	7.167 ± 4.258
S1	36	124	19	0.922 ± 0.027	0.00236 ± 0.00116	38.890 ± 17.301
S1a	23	70	13	0.893 ± 0.050	0.00119 ± 0.00061	19.636 ± 9.017
S1b	13	32	6	0.718 ± 0.128	0.00082 ± 0.00044	13.487 ± 6.488
S2	77	147	23	0.914 ± 0.016	0.00204 ± 0.00010	33.737 ± 14.862
S2a	38	76	11	0.836 ± 0.041	0.00149 ± 0.00074	24.543 ± 11.022
S2b	39	64	12	0.816 ± 0.045	0.00087 ± 0.00044	14.310 ± 6.554

Abbreviations as follows: N, number of samples; S, number of polymorphic (segregating) sites; h, number of haplotypes; Hd, Haplotype diversity (and respective standard deviation); π, nucleotide diversity (and respective standard deviation); MPD, mean pairwise distance (and respective standard deviation). All parameters were estimated accounting for insertion/deletions and allowing 5% of missing data. Haplogroups are labeled as in Figure 2.

In order to analyze how genetic diversity is geographically structured at a population-scale, median-joining networks were generated for each of the previously identified lineages (Figure 4). The *Hippotragus sp.* median-joining network shows that all haplotypes are connected sparsely, generally in similar frequencies, with no apparent geographic structure within Tanzania. Moreover, haplotypes are considerably divergent from each other. The

median-joining networks for haplogroups E1 and A show very strong geographic specificity. The clustering of haplotypes from Kenya and northern Mozambique in the same haplogroup (E2) suggests a possible connection between sables located easterly to the Rift Valley. In turn, the close relationship between haplotypes from Tanzania and Malawi evidences a possible connection between sables located westerly to the eastern arc of the Rift Valley (CE). Haplogroups S1 and S2 show less clear geographic structure, being widespread among southern populations (Figure 4). Although haplotypes from museum specimens from Zambia do not add anything to the phylogeographic structure of *H. niger* at a population-scale, as the exact population of origin is unknown, they do provide additional representativeness for haplogroups S1 and S2.

Generally, populations do not exhibit more than one haplogroup, but Shimba Hills (Kenya), Mufumbwe and Mulobezi (Zambia), Matesi and Triangle (Zimbabwe) share two haplogroups each. The population of Shimba Hills shows 86% of haplogroup E1 and 14 % of haplogroup E2. Mufumbwe and Matesi exhibit 75% and 67% of haplogroup S2, respectively. Finally, Triangle and Mulobezi are represented by 50% of each haplogroup S1 and S2. The lack of geographic structure observed for the two haplogroups in southern Africa (S1 and S2) clearly contrasts with the strong geographic structure found within each of them. Both S1a and S2a are present only in Zambian populations, while S1b and S2b were only observed to the south of the Zambezi river. As expected, more diverse populations correspond to those sharing divergent haplogroups (Table 15_SI in Suppl. Info.).

An analysis of haplotype sharing among populations (Figure 5) also shows the strong geographic specificity of haplogroups E1 and A, where no interpopulation sharing was found. Although the population of Niassa shares one of the haplogroups present in Shimba Hills (E2), no haplotype sharing was found. Similarly, no haplotype sharing was found among populations of haplogroup CE. The lack of haplotype sharing within haplogroups indicates a comparative lack of recent interpopulation contact. Figure 5 also provides a sense of how widespread haplogroups S1 and S2 are in southern Africa, even though haplotype sharing is mostly restricted to members of the same population. Briefly, haplogroup S1 is more northerly concentrated, reaching Triangle in its southernmost distribution, while haplogroup S2 is more concentrated in the southern part of the continent with its northernmost distribution in Mufumbwe. As expected, analyses focused solely on populations, such as multidimensional scaling analysis, did not show any signs of geographic correlation for haplogroups S1 and S2 (Text 13_SI; Figure 7_SI in Suppl. Info.).

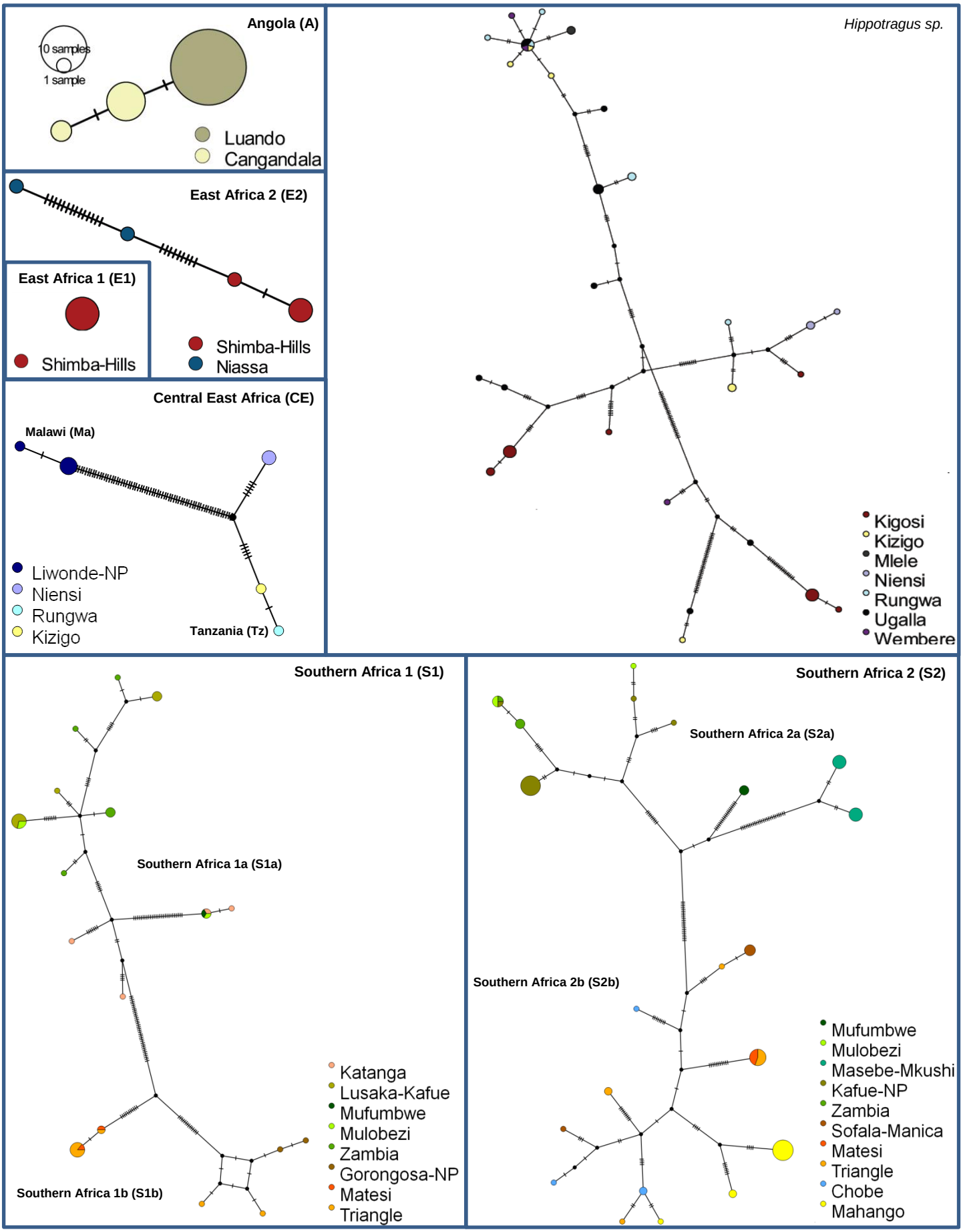

Fig. 4- Median-joining networks based on complete mtDNA sequences for each of the different *H. niger* haplogroups, and for the previously identified *Hippotragus* sp. All the networks were generated using PopART. Sites with insertion/deletions were not considered by the software, preventing the discrimination of some of the previously identified haplotypes, namely the second haplotype belonging to haplogroup E1 (Table 2). Haplogroup names are highlighted in bold, according to Figure 2. The numbers of mutations separating the different haplotypes are presented as hatch marks on top of the branches.

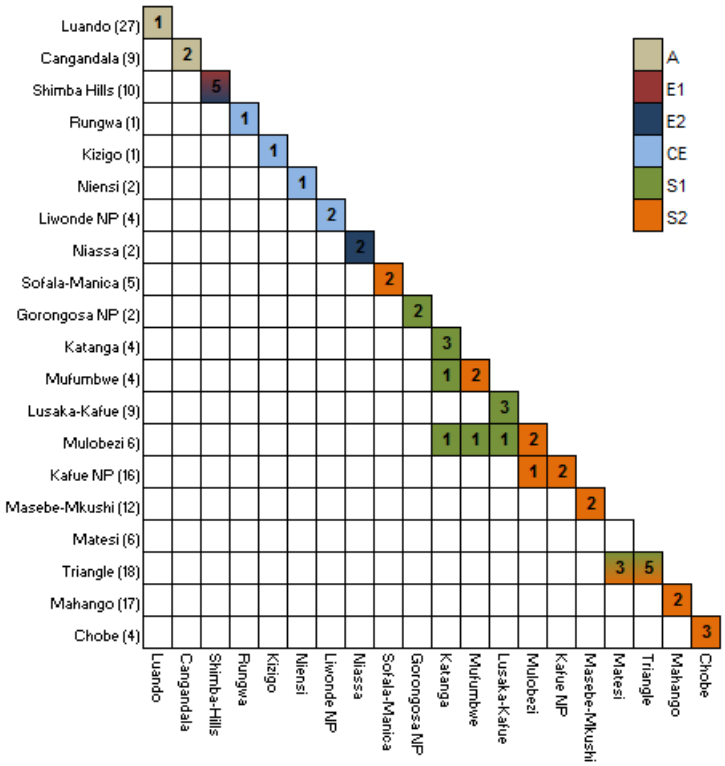

Fig. 5- Haplotype sharing among *H. niger* populations. The number of sequences is represented in curved brackets preceded by the population name. Cells are colored according to the different haplogroups. Cells with color gradients reflect haplogroup sharing at a given population. Numbers inside the cells represent the number of haplotypes. Empty cells reflect absence of haplotype sharing. Individuals from undetermined populations (*Zambia*) are not represented. For the five haplotypes found in Shimba-Hills, two belong to haplogroup E1 and three to E2. For the three haplotypes found in Matesi, two belong to S1 and one to S2. For the three haplotypes found in Triangle two belong to S1 and three to S2.

In accordance with low interpopulation sharing, values for all neutrality tests were non-significant in all haplogroups, and therefore are not indicative of a strong departure from mutation-drift equilibrium (Table 16_SI in Suppl. Info.). Mismatch analysis also reflected the stable demographic history of *H. niger*, with multimodal distributions in most haplogroups (data not shown). Estimations of growth parameter were non-significant in all haplogroups with the exception of haplogroups CE and S2 (data not shown). The growth values were extremely negative and significant (-936.114 with 95% CI [-1668.1, -204.1] for CE; and -636.6 with a 95% CI [-1235.9, -37.3] for S2), which could be indicative of population shrinkage within these haplogroups. However, these significant values are more likely to be misled by the strong genetic structure associated to these haplogroups, which can be divided into clearly distinct sub-haplogroups (Tanzania and Malawi for CE; and north and south the Zambezi river for S2).

The AMOVA analysis revealed that the partition of mtDNA variation that better represents the phylogeographic structure of *H. niger* in Africa is when individuals from

different populations are grouped according to the identified haplogroups and sub-haplogroups (Table 3). Approximately 81.2% of the mtDNA variation was attributed to differences among haplogroups, while 13.7 % and 4.7 % are attributed to differences among populations within haplogroups and within populations, respectively. Another analysis also showed significant ϕ-CT values when tested the hypothesis of a separation between populations located easterly and westerly to the eastern branch of the Rift Valley, suggesting an allopatric origin and complete geographic isolation of haplogroups from eastern (E1 and E2) and central-southeastern (CE, S1 and S2) Africa (Table 17_SI in Suppl. Info.).

Table 3-AMOVA results for a hierarchical arrangement according to the following partition of genetic diversity: among haplogroups, populations within haplogroups, within haplogroups.

Type of variation	Among haplogroups	Among populations within haplogroups	Within populations
d.f.	8	16	134
% of variation	81.59	13.66	4.75
ϕ-Statistics	ϕCT	ϕSC	ϕST
Fixation indices	0.81589	0.74210	0.95252

All values represented are significant: p-value=0.00000

Discussion

To our knowledge, this study provides the first ungulate phylogeographic assessment comprising central, southern and eastern Africa, based on the complete mitochondrial genome and using both modern and historic samples. So far, only one study using mitogenomics of ungulates in Africa has been reported, but with a more restricted geographic coverage, not fully comprising central, southern and eastern Africa (Heller et al., 2012). Therefore, we consider that this study represents a milestone on the comprehensive assessment of ungulates in Africa.

The sable antelope maternal history has been limited by the scant resolution associated to analyses focused on shorter mtDNA segments. This has hampered the estimation of credible divergence times between haplogroups and subsequent formulation of plausible hypotheses that could explain the observed phylogeographic patterns. Our ability to fully sequence and assemble hundreds of mitochondrial genomes from geographic representative samples, allowed us to reconstruct the evolutionary patterns of *H. niger* in Africa to an extent not previously possible. Not only we were able to provide, for the very first time, accurate and precise estimates of divergence times within *H. niger*, but we were also able to increase sampling size and fill gaps in the species geographic range by adding historic samples to our extensive dataset.

Previous studies that used complete mitochondrial genomes to overcome the limitations associated to analyzing small mtDNA fragments have strongly emphasized the relevance in using whole genome sequencing to address questions of divergence and diversity at a population scale. As expected, the genetic diversity observed for each *H. niger* haplogroup was higher using complete mitochondrial genome sequences (data not shown). Still, the number of haplotypes gained from sequencing the complete genome was not substantially higher than those observed for the CR. Therefore, using complete genomes did not seem to have added any significantly valuable information to the intraspecific diversity of *H. niger*. On the other hand, the number of polymorphic sites and number of fixed differences gained from the analysis of complete mitogenomes was substantially higher than those observed for the CR (data not shown). The comparison between *H. niger* fully sequenced genomes with the CR, therefore, confirms previous findings that using mitogenomes improves phylogenetic resolution.

MtDNA introgression from an extinct species

The deep and ancestral divergence between the west Tanzanian haplogroup and the remainder *niger*-like haplogroups evidenced in our study was already described by Pitra et al. (2002), and was interpreted as the result of an initial allopatric fragmentation between west Tanzanian sables and the remainder populations studied, followed by an unidirectional long-distance colonization from southern Africa to Tanzania. Nonetheless, this hypothesis failed to explain how the extremely divergent maternal ancestry of the Tanzanian sables is not reflected in any appropriate morphological feature (Pitra et al., 2002).

The estimated split time between *niger*-like haplogroups and the *Hippotragus sp.* lineage of 1.7 mya lies within a very well documented speciation period, around 1.6 to 1.8 mya, associated to shifts in African climate conditions (deMenocal, 2004). Because sables carrying this lineage do not appear to differ morphologically from other sables, and no other *Hippotragus* species seems to be found in the region around 1.7 mya, as suggested by fossil data, we could be facing a distinct *Hippotragus* species that has become extinct. This seems to be supported by the median-joining network and genetic diversity summary statistics uncovered in this study for *Hippotragus sp.* (Figure 4)**.** Not only this lineage is divergent in relation to all other sables but is highly divergent and diverse within itself, with the highest levels of nucleotide and haplotype diversity (Table 2). Our results, coupled with paleontological and previous genetic evidence, therefore suggest that the most likely explanation for the presence of this highly divergent lineage in west Tanzanian sables is through contact with a now extinct *Hippotragus* species, representing an event of mitochondrial introgression.

The replacement of a species mtDNA due to interspecific hybridization events has been very well described in mammals (Alves et al., 2008; Hailer et al., 2012; Roca et al., 2005). The introgression of mitochondrial sequences from currently extinct populations/species into the gene pool of extant populations/species, in particular, has also been documented (Achilli et al., 2008; Barbieri et al., 2013; Lippold et al., 2011). For example, hares in Iberian Peninsula represent a classic case of extensive interspecific introgression, harboring high frequencies of mtDNA from an arctic/boreal species that is currently extinct in the region (Melo-Ferreira et al., 2009).

Despite being the election marker in phylogeography during the past thirty years, mtDNA has the limitation of being a single maternally inherited locus. Nuclear markers from multiple loci would be needed to provide a fuller picture of the evolutionary history of *H. niger*, from both male and female genetic components. Unfortunately, studies of autosomal DNA variation in *H. niger* are not yet available. In this study we have used an X-linked intron and, although no signs of introgression were found, this might be due to the absence of enough genetic diversity in the studied fragment (Figure 3_SI). Nonetheless, unpublished data using microsatellites on the same dataset (Vaz Pinto et al., unpublished) do not either show any evidence of this introgression for Tanzanian sables. The contrast between nuclear and mtDNA suggests that strong gender-related asymmetries are likely to be involved (Palumbi and Baker, 1994), which would be congruent with the matrilocal social organization of *H. niger* (Skinner and Chimimba, 2005; Thompson, 1993). Therefore, we believe that *H. niger* males would have occasionally bred with females of the presumably extint *Hippotragus* species, leading to an unidirectional interspecific hybridization event that would have culminated with the replacement of sables mtDNA by *Hippotragus sp*. Ultimately, this hypothesis could explain why the extremely divergent clade from Tanzanian sables does not reflect any obviously distinctive morphological feature, since they are, in fact, *H. niger* from central east Africa carrying the mitogenome from a presumably extint species.

Nonetheless, other alternatives should be carefully considered to avoid an established tendency in phylogeography of explaining unexpected patterns in mtDNA variation based on common standard hypotheses, without ruling out other possibilities (Funk and Omland, 2003). We consider that, together with our main hypothesis of introgression, a second hypothesis that could be formulated is the existence of incomplete lineage sorting. Although these two hypotheses are difficult to distinguish, there are some patterns associated to introgression that can be identified and used to dismiss incomplete lineage sorting as a possible explanation for *Hippotragus sp*. First, incomplete lineage sorting is less likely to involve highly divergent lineages. Recently introgressed haplotypes usually assume a highly derived position in phylogenetic clades (Funk and Omland, 2003), as it is the case of

Hippotragus sp. Secondly, incomplete lineage sorting has no apparent relation with geography (Hare and Avise (1998) in Funk and Omland (2003)). Finally, given the effective population size of the mitochondrial genome, which is one quarter of the nuclear genome, stochastic lineage sorting is expected to progress more rapidly for mitochondrial haplotypes (Funk and Omland, 2003). Therefore, incomplete lineage sorting would be much more likely in the nuclear genome, than in the mitochondrial genome. The absence of *Hippotragus sp.* in nuclear DNA, points to mtDNA introgression as the most plausible explanation for the existence of such a highly divergent clade. In the end, the combination of a nuclear background with strong geographic specificity proved to be good diagnostic tests for mtDNA introgression.

Interesting complex patterns revealed by complete mitochondrial genomes during the mid-Pleistocene

The branching structure of the mtDNA phylogeny and complex phylogeographic patterns found in *H. niger* may have been shaped by events of climate change occurring at different periods in eastern, central and southern Africa during the Pleistocene (Avise et al., 1998; deMenocal, 2004; Schefuss et al., 2003). Although little is known about the impact of climate oscillations in the Southern Hemisphere, in particular in central and southern Africa (Beheregaray, 2008), previous studies evidenced that savanna-adapted species tend to retract during warm periods and expand during glacial periods (Flagstad et al., 2001; Lorenzen et al., 2012). This tendency seems to be the opposite of the well-studied Boreal and Temperate species of the Northern Hemisphere, which retract during glacial cycles and expand in periods of warmer climate (Hewitt, 2004a, 2004b). Nonetheless, all share a pattern in which periods of retraction give rise to differentiation in spatial and temporal refugia, which are thought to be responsible for high inter- and intraspecific diversity, while periods of range expansions lead to the establishment of secondary contacts.

In accordance with previous ungulates studies, our results revealed a clear pattern of population subdivision during warmer periods that followed glacial cycles (Figure 3). This is particularly evident for the first and second major splits between eastern and southeastern haplogroups from the remainder *niger*-like haplogroups (E1+E2 from CE+A+S1+S2), and between central and southern African haplogroups (CE+A from S1+S2). The same can be said for the most recent splits between sub-haplogroups from southern African (S1a from S1b; S2a from S2b), and central east African (Tz from Ma). Nonetheless, some exceptions were found, such as the split of haplogroups A and CE during a glacial period around 150 to 130 kya. Furthermore, the contemporary split times between haplogroups CE and A from S1 and S2, and the split of haplogroup E1 from E2, do not seem to have occurred immediately

after glacial cycles, but rather in the end of the warm period. These incongruences, coupled with the fact that confidence intervals for some splits encompass both glacials and interglacials (Figure 3), require a more cautious interpretation of divergence times before speculating trends in phylogeographic patterns solely based on climate changes. It is possible that other factors than climate oscillations, and derived changes in vegetation, might have led to population fragmentation during warmer periods in the mid-Pleistocene. These include major geomorphological changes associated to tectonic movements, such as rifting (Hamilton, 1982), mountain formation, and river barriers (Cotterill, 2003, 2006; Voelker et al., 2013).

The interplay between climatically determined habitat fluctuations and geological features is still poorly understood. Several authors point to geological barriers such as rifting and mountain formation as the responsible factors for the high levels of diversity and divergence in east Africa (Gibernau and Montuire, 1996; Hamilton, 1982). The current sharp disjunctions in geology, relief and distribution of the miombo woodlands associated to the Great Rift Valley, seem to mirror past geomorphological changes that resulted in population fragmentation. The Great Rift Valley consists in a Y-shaped circle of mountains composed by the Albertine Rift Mountains, in the western branch, and the East Arc Rift Mountains, in the eastern branch. The deep split between eastern haplogroups (E1 and E2) from the remainder *niger*-like haplogroups, reflects a clear geographic barrier between sables located westerly and easterly to the Eastern Arc Rift (also supported by the AMOVA: Table 17_SI). Although this barrier could have resulted from differences in vegetation associated to Pleistocene climate shifts that restricted gene flow between sables from east Africa and sables from central and southern Africa, it is more likely to be the result of mountain barriers. Such east African-central and southern African phylogeographic break has also been detected in other ungulates, rodents and carnivores (Flagstad et al., 2001; Girman et al., 1993; Lorenzen et al., 2012), suggesting that the Eastern Arc Mountains might represent a boundary for previously widespread taxa. The fact that some haplotypes from Kenya are more closely related to haplotypes from northern Mozambique than to haplotypes of the same population indicates that once sables were connected through the eastern side of the Rift Valley, and then split. As this seems to have happened during the end of the warm period, major geomorphological changes might have been involved as well (Hamilton, 1982). This is consistent with previous studies that group sables from Tanzanian populations located easterly to the Rift Valley with sables from Kenya (Pitra et al., 2002, 2006; van Vuuren, 2010), linking all *H. niger* from east Africa.

The fact that southern and central east African haplogroups (CE, S1 and S2) are composed of long separate branches and unique haplotypes that can be easily clustered in

geographic specific sub-haplogroups, as so notably evidenced by the median-joining networks generated for each haplogroup (Figure 4), indicates lack of recent interpopulation contact and might represent the remains of an ancient and richer diversity in sables from central and southern Africa. We believe that, during the glacial period that preceded the split between central (CE+A) and southern (S1+S2) haplogroups, sables from ancestral populations to these haplogroups could have been widespread in the western side of the Riff Valley, in the savannah habitats that occupied most these regions (see Figure 6 in Suppl. Info.). Then, with the beginning of the subsequent glacial period, they probably got fragmented in two refugia: one in central east Africa and another to the south. Although is speculative to geographically place these refugia in Africa, due to lack of both fossil and genetic data in this specific time frame, other studies with ungulates point to refugia in the south and central east regions (Lorenzen et al., 2012).

The split between eastern haplogroups (E1 and E2) seems to be related to changes in vegetation in the interglacial period (Figure 3). However, given the low sampling size in eastern populations (Figure 1), together with the fact that sables are known to be currently widespread in eastern Africa, we don't have data to assume that the divergence between the two is related to habitat fragmentation. What we do know is that these eastern haplogroups have been able to remain divergent and separate until the present day, despite apparent lack of contemporary barriers to gene flow. The split between eastern haplogroups was accompanied by the southern split into S1 and S2 (Figure 3). The pattern and depth of genetic diversity among S1 and S2 support a complex phylogeographic history, probably stemming from both climatic oscillations and geomorphological changes during the Pleistocene. Members of the same populations cluster in the two different haplogroups named as S1 and S2, and members from geographically distant populations are clustered in the same haplogroup (either S1 or S2; see Figure 2 and 4). This indicates possible common ancestry of widely separated populations. The fact that the split between S1 and S2 seem to have occurred in the end of the warmer climate suggests that they were able to survive together during most of this period, and only later started to diverge, possibly as a result of river barriers (Cotterill, 2003, 2006). Despite being nearly impossible to associate this split event to either geological barriers or climate oscillations, or both, due to lack of data, the fact that this split is contemporary to another in a different African subregion (E1 from E2) could support better the effect of climate shifts as major drivers of fragmentation. This also seems to be supported by the split of two hartebeest lineages around the exact same period, also in different geographic regions (Flagstad et al., 2001).

The most complex phylogeographic history so far evidenced in our results is the split between the Angolan haplogroup from central east Africa around 145.5 kya (Figure 3). In

contrast to all other *niger*-like haplogroups, this split event seems to have occurred during a glacial cycle. During this period, known as the maximum of the penultimate glaciation (150 to 130 kya), the African continent is thought to have suffered similar conditions as those observed for the last glacial maxima, in which rainforests were contracted and savanna habitats occupied most of the territory (Figure 8_SI in Suppl. Info.), leading to the appearance of savanna adapted species in habitats no longer suitable in the present African continent (Van Andel, 1996; Cohen et al., 2007). This period is also the center stage for the origin of the iconic giant sable antelope of Angola. The close phylogenetic relationship between Angolan and Central East African sables is hard to explain as no present population that could historically link these haplogroups together seems to be found today.

Pleistocene major climatic and geomorphological changes created unique habitats, many of which are now completely fragmented, or no longer exist today (Steele, 2007). We believe that Angolan and Central East haplogroups descend from ancestral populations that could have been previously found scattered in between the current distribution of the two haplogroups, possibly in the Congo basin or northern Zambia, during the glacial period (Figure 3). It is possible that some sables were more restricted to the west and others to the east, probably as a consequence of river barriers imposed by the Congo drainage system. Examples of cases in which population substructure was caused by the Congo drainage system include the bonobo, birds and the okapi (Hvilsom et al., 2014; Stankiewicz and de Wit, 2006; Stanton et al., 2014; Voelker et al., 2013). This would explain population fragmentation in a glacial cycle, even though savannah habitats still existed. The beginning of the subsequent interglacial (Figure 3) would have led to the displacement of savannah habitats by rainforests, fragmenting populations that no longer exist today (Figure 8_SI in Suppl. Info.). It is possible that, then, some sables would have found dispersal corridors that allowed them to colonize other areas, including those of the present haplogroup distribution in northern Angola, western Tanzania and Malawi. In that sense, the Angolan and central eastern haplogroups might be the remnants of an ancient and richer genetic diversity. The beginning of the interglacial period also seems to have resulted in the splits of southern African central east African haplogroups into sub-haplogroups (a and b for southern haplogroups and Tz and Ma for central east haplogroups).

After the split between S1 and S2 these haplogroups might have established secondary contact with each other during the glacial cycle that preceded the split into the four sub-haplogroups (see Figure 3). The fragmentation of southern haplogroups into sub-haplogroups a and b, located to the north and south of the Zambezi river, respectively, in the subsequent interglacial, strongly suggest that the Zambezi may have been an effective geographic barrier to gene flow until the present day. This could be the result of the high

levels of humidity and heavy rains characteristic of this warmer period, which may have turned the Zambezi impassable, restricting gene flow between sub-haplogroups located on either side of the river (Van Andel, 1996; Schefuss et al., 2003; Figure 8_SI in Suppl. Info.). On the other hand, the geomorphology of the Zambezi river has been quite dynamic (Cotterill, 2003, 2006; Moore et al., 2007) and this could also be the result of major geomorphological changes in the river drainage system. The lack of haplotype sharing between sub-haplogroups (Figure 5) suggests that they were each fragmented in different refugia. The contemporary split between central African sub-haplogroups (Ma from Tz; Figure 3) also seems to be the result of transitions from cool-arid to warm-humid periods. These sub-haplogroups were probably connected in the western side of the East Arc Mountains scattered through the savanna grasslands, and were then fragmented with the beginning of the interglacial. However, just as the Zambezi for southern haplogroups and the east Arc mountains for eastern haplogroups, the Muchinga escarpments, together with lake Malawi, are more likely to have acted as impassible geological barriers, fragmenting the central east haplogroup into sub-haplogroups Tz and Malawi until today (Hamilton, 1982).

Overall, despite range expansions and periods of retraction, sable populations seem to have been able to remain stable and large. Such stable demographic history is also reflected in the high levels of genetic diversity observed in sable antelope haplogroups, particularly the southern ones. Haplogroups S1 and S2 have similar and high levels of genetic diversity, as shown by the number of haplotypes and respective nucleotide and haplotype diversities. The stable demographic history is also reflected in the structure of the median-joining network observed for these haplogroups, with long separate branches and unique haplotypes (Figure 4). Although in S2 the fragmentation in a and b could be associated to shrinkage, as evidenced by the growth parameter, this is more likely to be a false signal of population decline associated to the strong sub-structure evidenced. Population structure and migration can give rise to false signals of population size changes (Hvilsom et al., 2014). The same can be said regarding the central east African haplogroup which is highly structured in two distinct sub-haplogroups, Ma and Tz (Figure 4). Overall, there is a decrease in genetic diversity from the south to central-east and east Africa, with the Angolan and eastern haplogroups having the lowest genetic diversity. This pattern has also been observed in other ungulates such as the common eland, wildebeest, kudu and impala (Lorenzen et al., 2012).

Although shifts in African climate and vegetation, and geomorphological changes during the Pleistocene seem to be roughly contemporary to the phylogeographic patterns of *H. niger* uncovered in this study, detailed non-speculative comparisons are hampered by

sampling gaps, dating uncertainties and preservational biases in the current fossil record in central, southern and eastern Africa (deMenocal, 2004).

Subspecies nomenclature

Despite the difficult challenge in setting criteria that work for defining subspecies and species, all reviews that have focused so far in this problem agree that data should support discreetness and biological significance of the subspecies, and that species are separately evolving metapopulation lineages (Haig et al., 2006; De Queiroz, 2007). The genetic data generated in this study demonstrates that populations situated easterly to the Rift Valley are monophyletic and form sister taxa substantially divergent from the remainder *H. niger* haplogroups. This is consistent with the subspecies classification proposed by Ansell (1971) as *Hippotragus niger roosevelti*. Our results also confirm previous reports of the presence of *H. n. roosevelti* in northern Mozambique (Booth, 2012). With the exception of sables from Malawi, which are clearly distinct from those found in Zambia, other haplogroups seem to be consistent with previous classifications. One possibility would be to accept the subspecies classification proposed by Grooves (1983) for Malawi (*H. n. anselli*), and extend it to western Tanzania, even though this would invalidate Ansell's (1971) classification as the same subspecies found in Zambia (*H. n. kirkii*). The Angolan haplogroup also seems to be congruent with the critically endangered giant sable antelope subspecies. Despite both Zambian and remainder southern African populations share two genetically distinct haplogroups (S1 and S2), there substructure into sub-haplogroups located to the north and south of the Zambezi river seems to be congruent with past classifications in *H. n. kirkii* and *H. n. niger*, respectively. Nonetheless, we recommend that further analysis of nuclear data and more detailed morphological work should be performed for subspecies validation.

Conclusions

To conclude, we have investigated the maternal history of several *H. niger* populations, from complete mitochondrial genomes covering the species full geographic range in Africa. Using museum specimens from never before sampled populations has allowed us to re-evaluate the phylogeography of the species. With this extensive dataset, we considerably increased our knowledge of the mtDNA variation within *H. niger*, which has been associated to complex phylogeographic patterns that date back to the Pleistocene. We propose the hypothesis that Tanzanian sables carrying a previously described highly divergent lineage experienced gene flow from a currently extint species that diverged from *H. niger* around 1.7 mya. Our results point to what appears to be the combined effects of changes in vegetation caused by Pleistocene climatic shifts and geomorphological changes caused by tectonics as responsible for most of the species phylogeographic history in Africa.

Our attempts to obtain a valid geographical representation of the species were often taken at the expense of larger individual population sizes. The limited sampling in some populations, such as Congo, Mozambique and Malawi, may mean we have missed haplotype sharing. Nonetheless, this does not seem to have had major repercussions in the phylogeographic patterns retrieved from this study. We believe that this study could integrate future comparative phylogeographic studies uncover the major trends in the histories of vicariance and dispersal in central, southern and eastern Africa. Despite the existence of comparative reviews of ungulates in Africa (e.g. Lorenzen et al., 2012), they should be updated using a consistent multidisciplinary approach that would include both genetic and environmental data, with reliable mutation rates and applying the same methodology for all species. This is extremely relevant for the field of phylogeography as most of publications so far have focused on study systems from the Northern Hemisphere (Beheregaray, 2008). Finally, this study adds evidence that deep diversity and divergence within species can be captured using target enrichment methods for next generation sequencing to generate large numbers of mitogenomes, and provides a useful tool for better understanding of population structure and evolutionary patterns in *H. niger*.

References

Achilli, A., Olivieri, A., Pellecchia, M., and Uboldi, C. (2008). Mitochondrial genomes of extinct aurochs survive in domestic cattle. Curr. Biol. 157–158.

Achilli, A., Olivieri, A., Soares, P., Lancioni, H., Hooshiar Kashani, B., Perego, U. a, Nergadze, S.G., Carossa, V., Santagostino, M., Capomaccio, S., et al. (2012). Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. Proc. Natl. Acad. Sci. U. S. A. *109*, 2449–2454.

Alexander, A., Steel, D., Slikas, B., Hoekzema, K., Carraher, C., Parks, M., Cronn, R., and Baker, C.S. (2013). Low diversity in the mitogenome of sperm whales revealed by next-generation sequencing. Genome Biol. Evol. *5*, 113–129.

Alves, P.C., Melo-Ferreira, J., Freitas, H., and Boursot, P. (2008). The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus Lepus. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *363*, 2831–2839.

Van Andel, T. (1996). Palaeolithic landscapes of Europe and environs, 150,000-25,000 years ago: An overview. Quat. Sci. Rev. *15*, 481–500.

Ansell, W. (1971). Order Artiodactyla. Mamm. Africa an Identif. Manual, Part 15 1–84.

Arnason, U., Gullberg, and Widegren, B. (1993). Cetacean mitochondrial DNA control region: sequences of all extant baleen whales and two sperm whale species. Mol. Biol. Evol. *10*, 960–970.

Arnason, U., Gullberg, A., and Janke, A. (2004). Mitogenomic analyses provide new insights into cetacean origin and evolution. Gene *333*, 27–34.

Arnason, U., Gullberg, A., Janke, A., and Kullberg, M. (2007). Mitogenomic analyses of caniform relationships. Mol. Phylogenet. Evol. *45*, 863–874.

Arnason, U., Adegoke, J. Gullberg, A., Harley, E.H., Janke, A., and Kullberg, M. (2008). Mitogenomic relationships of placental mammals and molecular estimates of their divergences. Gene *421*, 37–51.

Avise, J.C., Walker, D., and Johns, G.C. (1998). Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc. Biol. Sci. *265*, 1707–1712.

Barbieri, C., Vicente, M., Rocha, J., Mpoloka, S.W., Stoneking, M., and Pakendorf, B. (2013). Ancient substructure in early mtDNA lineages of southern Africa. Am. J. Hum. Genet. *92*, 285–292.

Beheregaray, L.B. (2008). Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. Mol. Ecol. *17*, 3754–3774.

Berger, W., and Bickert, T. (1997). Reconstruction of atmospheric CO 2 from ice-core data and the deep-sea record of Ontong Java Plateau: the Milankovitch chron. Oceanogr. Lit. … 466–495.

Bibi, F. (2013). A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. BMC Evol. Biol. *13*, 166.

Booth, V. (2012). Current Status of Sable Antelope Classification in Niassa.

Bryant, D., and Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. *21*, 255–265.

Chan, Y.C., Roos, C., Inoue-Murayama, M., Inoue, E., Shih, C.C., Pei, K.J.C., and Vigilant, L. (2010). Mitochondrial genome sequences effectively reveal the phylogeny of Hylobates gibbons. PLoS One *5*, e14419.

Chiou, K.L., Pozzi, L., Lynch Alfaro, J.W., and Di Fiore, A. (2011). Pleistocene diversification of living squirrel monkeys (Saimiri spp.) inferred from complete mitochondrial genome sequences. Mol. Phylogenet. Evol. *59*, 736–745.

Cohen, A.S., Stone, J.R., Beuning, K.R.M., Park, L.E., Reinthal, P.N., Dettman, D., Scholz, C.A., Johnson, T.C., King, J.W., Talbot, M.R., et al. (2007). Ecological consequences of early Late Pleistocene megadroughts in tropical Africa. Proc. Natl. Acad. Sci. U. S. A. *104*, 16422–16427.

Cooper, a, Lalueza-Fox, C., Anderson, S., Rambaut, a, Austin, J., and Ward, R. (2001). Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature *409*, 704–707.

Cotterill, F.P.D. (2003). Geomorphological influences on vicariant evolution in some African mammals in the Zambezi basin: some lessons for conservation. Ecol. Conserv. Small Antelope. Proc. an Int. Symp. Duiker Dwarf Antelope Africa. Filander Verlag, Fürth. 11 – 58.

Cotterill, F.P.D. (2006). The evolutionary history and taxonomy of the Kobus leche species complex of South-Central Africa in the context of palaeo-drainage dynamics. Stellenbosch : Stellenbosch University.

Curole, J.P., and Kocher, T.D. (1999). Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol. Evol. *14*, 394–398.

Dabney, J., and Meyer, M. (2012). Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. Biotechniques *52*, 87–94.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., et al. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. U. S. A. *110*, 15758–15763.

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

deMenocal, P.B. (2004). African climate change and faunal evolution during the Pliocene–Pleistocene. Earth Planet. Sci. Lett. *220*, 3–24.

Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. *7*, 214.

Duchêne, S., Archer, F.I., Vilstrup, J., Caballero, S., and Morin, P. a (2011). Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. PLoS One *6*, e27138.

Dupont, L. (2011). Orbital scale vegetation change in Africa. Quat. Sci. Rev. *30*, 3589–3602.

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

Edwards, C.J., Magee, D.A., Park, S.D.E., McGettigan, P.A., Lohan, A.J., Murphy, A., Finlay, E.K., Shapiro, B., Chamberlain, A.T., Richards, M.B., et al. (2010). A complete mitochondrial genome sequence from a mesolithic wild aurochs (Bos primigenius). PLoS One *5*, e9255.

Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. Online *1*, 47–50.

Finstermeier, K., Zinner, D., Brameier, M., Meyer, M., Kreuz, E., Hofreiter, M., and Roos, C. (2013). A mitogenomic phylogeny of living primates. PLoS One *8*, e69504.

Flagstad, O., Syvertsen, P.O., Stenseth, N.C., and Jakobsen, K.S. (2001). Environmental change and rates of evolution: the phylogeographic pattern within the hartebeest complex as related to climatic variation. Proc. Biol. Sci. *268*, 667–677.

Fu, Y.X. (1997). Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. Genetics *147*, 915–925.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. a, Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. U. S. A. *110*, 2223–2227.

Funk, D.J., and Omland, K.E. (2003). S PECIES -L EVEL P ARAPHYLY AND P OLYPHYLY : Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. *34*, 397–423.

Gibernau, M., and Montuire, S. (1996). Mammal diversity and environment evolution during the Plio-Pleistocene in East Africa. *4*.

Girman, D.J., Kat, P.W., Mills, M.G.L., Ginsberg, J.R., Borner, M., Wilson, V., Fanshawe, J.H., Fitzgibbon, C., Lau, L.M., and Wayne, R.K. (1993). Molecular Genetic and Morphological Analyses of the African Wild Dog (Lycaon pictus). J. Hered. *84*, 450–459.

Groves, C. (1983). A new subspecies of sable antelope, Hippotragus niger (Harris 1838). Rev. Zool. Africaine 821–828.

Haig, S.M., Beever, E.A., Chambers, S.M., Draheim, H.M., Dugger, B.D., Dunham, S., Elliott-Smith, E., Fontaine, J.B., Kesler, D.C., Knaus, B.J., et al. (2006). Taxonomic considerations in listing subspecies under the U.S. Endangered Species Act. Conserv. Biol. *20*, 1584–1594.

Hailer, F., Kutschera, V.E., Hallström, B.M., Klassert, D., Fain, S.R., Leonard, J. a, Arnason, U., and Janke, A. (2012). Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. Science *336*, 344–347.

Hamilton, A.C. (1982). Environmental history of East Africa: a study of the Quaternary.

Hare, M.P., and Avise, J.C. (1998). Population structure in the American oyster as inferred by nuclear gene genealogies. Mol. Biol. Evol. *15*, 119–128.

Heller, R., Brüniche-Olsen, A., and Siegismund, H.R. (2012). Cape buffalo mitogenomics reveals a Holocene shift in the African human-megafauna dynamics. Mol. Ecol. *21*, 3947–3959.

Hewitt, G.M. (2004a). Genetic consequences of climatic oscillations in the Quaternary. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *359*, 183–95; discussion 195.

Hewitt, G.M. (2004b). The structure of biodiversity - insights from molecular phylogeography. Front. Zool. *1*, 4.

Hoelzel, a R., Hancock, J.M., and Dover, G. a (1991). Evolution of the cetacean mitochondrial D-loop region. Mol. Biol. Evol. *8*, 475–493.

Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. *23*, 254–267.

Hvilsom, C., Carlsen, F., Heller, R., Jaffré, N., and Siegismund, H.R. (2014). Contrasting demographic histories of the neighboring bonobo and chimpanzee. Primates. *55*, 101–112.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kingdon, J. (1982). Horse-like antelopes, Hippotragines. East African Mamm. an Atlas Evol. Africa 539–577.

Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, e3.

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

Kuhner, M.K. (2006). LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics *22*, 768–770.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

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

Liedigk, R., Yang, M., Jablonski, N.G., Momberg, F., Geissmann, T., Lwin, N., Hla, T.H., Liu, Z., Wong, B., Ming, L., et al. (2012). Evolutionary history of the odd-nosed monkeys and the phylogenetic position of the newly described Myanmar snub-nosed monkey Rhinopithecus strykeri. PLoS One *7*, e37418.

Lippold, S., Matzke, N.J., Reissmann, M., and Hofreiter, M. (2011). Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. BMC Evol. Biol. *11*, 328.

Lorenzen, E.D., Heller, R., and Siegismund, H.R. (2012). Comparative phylogeography of African savannah ungulates. Mol. Ecol. *21*, 3656–3670.

Maricic, T., Whitten, M., and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One *5*, e14004.

Matthee, C. a, and Robinson, T.J. (1999). Mitochondrial DNA population structure of roan and sable antelope: implications for the translocation and conservation of the species. Mol. Ecol. *8*, 227–238.

Meadows, J.R.S., Hiendleder, S., and Kijas, J.W. (2011). Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. Heredity (Edinb). *106*, 700–706.

Melo-Ferreira, J., Alves, P.C., Freitas, H., Ferrand, N., and Boursot, P. (2009). The genomic legacy from the extinct Lepus timidus to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. Mol. Ecol. *18*, 2643–2658.

Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448.

Miller, W., Drautz, D.I., Janecka, J.E., Lesk, A.M., Ratan, A., Tomsho, L.P., Packard, M., Zhang, Y., McClellan, L.R., Qi, J., et al. (2009). The mitochondrial genome sequence of the Tasmanian tiger (Thylacinus cynocephalus). Genome Res. *19*, 213–220.

Moore, A.E., Cotterill, F.P.D.W., Main, M.P.L., and Williams, H.B. (2007). The Zambezi River.

Morin, P.A., Archer, F.I., Foote, A.D., Vilstrup, J., Allen, E.E., Wade, P., Durban, J., Parsons, K., Pitman, R., Li, L., et al. (2010). Complete mitochondrial genome phylogeographic analysis of killer whales (Orcinus orca) indicates multiple species. Genome Res. *20*, 908–916.

Palumbi, S., and Baker, C. (1994). Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol. Biol. Evol. *11*, 426–435.

Pitra, C., Hansen, A.J., Lieckfeldt, D., and Arctander, P. (2002). An exceptional case of historical outbreeding in African sable antelope populations. Mol. Ecol. *11*, 1197–1208.

Pitra, C., VazPinto, P., O'Keeffe, B.W.J., Willows-Munro, S., Jansen van Vuuren, B., and Robinson, T.J. (2006). DNA-led rediscovery of the giant sable antelope in Angola. Eur. J. Wildl. Res. *52*, 145–152.

De Queiroz, K. (2007). Species concepts and species delimitation. Syst. Biol. *56*, 879–886.

Renaud, G., Stenzel, U., and Kelso, J. (2014). leeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Res. 1–7.

Roca, A.L., Georgiadis, N., and O'Brien, S.J. (2005). Cytonuclear genomic dissociation in African elephant species. Nat. Genet. *37*, 96–100.

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

Rohland, N., and Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. Nat. Protoc. *2*, 1756–1762.

Rohland, N., Malaspinas, A.-S., Pollack, J.L., Slatkin, M., Matheus, P., and Hofreiter, M. (2007). Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. PLoS Biol. *5*, e207.

Rokas, A., and Carroll, S.B. (2005). More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol. Biol. Evol. *22*, 1337–1344.

Schefuss, E., Schouten, S., Jansen, J.H.F., and Sinninghe Damsté, J.S. (2003). African vegetation controlled by tropical sea surface temperatures in the mid-Pleistocene period. Nature *422*, 418–421.

Schliep, K.P. (2011). phangorn: phylogenetic analysis in R. Bioinformatics *27*, 592–593.

Shamblin, B.M., Bjorndal, K. a, Bolten, A.B., Hillis-Starr, Z.M., Lundgren, I., Naro-Maciel, E., and Nairn, C.J. (2012). Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. Mol. Ecol. *21*, 2330–2340.

Shields, G.F., and Kocher, T.D. (1991). Phylogenetic relationships of north American ursids based on analysis of mitochondrial DNA. Evolution (N. Y). *45*, 218–221.

Skinner, J., and Chimimba, C. (2005). The mammals of the southern African sub-region (Cambridge University Press).

Slatkin, M. (1996). A correction to the exact test based on the Ewens sampling distribution. Genet. Res. *68*, 259–260.

Stankiewicz, J., and de Wit, M.J. (2006). A proposed drainage evolution model for Central Africa—Did the Congo flow east? J. African Earth Sci. *44*, 75–84.

Stanton, D.W.G., Hart, J., Galbusera, P., Helsen, P., Shephard, J., Kümpel, N.F., Wang, J., Ewen, J.G., and Bruford, M.W. (2014). Distinct and diverse: range-wide phylogeography reveals ancient lineages and high genetic variation in the endangered okapi (Okapia johnstoni). PLoS One *9*, e101081.

StatSoft Inc. (2011). STATISTICA (data analysis software system), version 10, www.statsoft.com

Steele, T.E. (2007). Late Pleistocene of Africa. Elsevier B.V 3139–3150.

Stephens, M., and Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. *73*, 1162–1169.

Stephens, M., Smith, N.J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. *68*, 978–989.

Stiller, M., Knapp, M., and Stenzel, U. (2009). Direct multiplex sequencing (DMPS)—a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. Genome Res. 1843–1848.

Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics *123*, 585–595.

Thompson, K. V. (1993). Aggressive behavior and dominance hierarchies in female sable antelope,Hippotragus niger: Implications for captive management. Zoo Biol. *12*, 189–202.

Trauth, M.H., Larrasoaña, J.C., and Mudelsee, M. (2009). Trends, rhythms and events in Plio-Pleistocene African climate. Quat. Sci. Rev. *28*, 399–411.

Vilstrup, J.T., Ho, S.Y., Foote, A.D., Morin, P. a, Kreb, D., Krützen, M., Parra, G.J., Robertson, K.M., de Stephanis, R., Verborgh, P., et al. (2011). Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. BMC Evol. Biol. *11*, 65.

Voelker, G., Marks, B.D., Kahindo, C., A'genonga, U., Bapeamoni, F., Duffie, L.E., Huntley, J.W., Mulotwa, E., Rosenbaum, S. a, and Light, J.E. (2013). River barriers and cryptic biodiversity in an evolutionary museum. Ecol. Evol. *3*, 536–545.

vanVuuren, B.J., Robinson, T.J., VazPinto, P., Estes, R., and Matthee, C.A. (2010). Western Zambian sable: Are they a Geographic Extension of the Giant sable Antelope? South African J. Wildl. Res. *40*, 35–42.

Watterson, G.A. (1978). The Homozygosity Test of Neutrality. Genetics *88*, 405–417.

Yu, L., Li, Y.W., Ryder, O., and Zhang, Y.-P. (2007). Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. BMC Evol. Biol. *7*, 198.

Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L.F., and Roos, C. (2013). Baboon phylogeny as inferred from complete mitochondrial genomes. Am. J. Phys. Anthropol. *150*, 133–140.

III-General Discussion and Conclusions

General Discussion

The use of hundreds of fully sequenced mitogenomes made possible to reconstruct the maternal history of the sable antelope with a great power of analysis. With our larger sample size and wide geographic coverage, we have been able to elucidate the phylogenetic relationships, the patterns of diversity, and the distribution of the different *H. niger* haplogroups. We were also able to explain the presence of the intriguing Tanzanian clade described by Pitra et al. (2002) as a result of mtDNA introgression from a now extinct species. More importantly, we were able to formulate an explicit hypothesis concerning the origin of the iconic giant sable antelope of Angola. Ultimately, the detailed phylogeographic scenario that we provided for the species as a whole in a region – central, southern and eastern Africa – which is still poorly understood, can be used to illuminate the evolutionary history of other species and to allow the explicit testing of phylogeographic hypotheses in this region.

I- Nuclear data as a control to the mtDNA patterns

Despite being the marker of choice in most phylogeographic studies, mtDNA has the limitation of being a single maternally inherited locus. As a consequence of being maternally inherited, evolutionary patterns inferred from mtDNA genetic variation can be sex-biased (Palumbi and Baker, 1994), particularly in species that are matrilocal such as the sable antelope. In such context, mtDNA should not be singly used to infer evolutionary patterns of species as it only reflects their maternal history. Nuclear markers are needed to obtain data from multiple loci and from both male and female genetic components (as shown by Hailer et al., 2012).

The use of nuclear genes can be a valuable and successful tool for phylogeographic assessments at the intraspecific level, in particular to be used as a control for gender related asymmetries in studies focused on mtDNA. Nonetheless, its utility for phylogeographic analyses is usually constrained by low evolutionary rate, difficult isolation due to low copy number, and intragenic recombination (Avise, 2009). These constraints can be somehow circumvented through the identification of nuclear genes that evolve more rapidly than average, which is the case of intron sequences of protein-coding genes. For mammal species, these can also be somewhat overcome through the choice of genes located on the X- or Y-chromosomes (Avise, 2009). As a single-copy gene, the Y-chromosome shares the same property of mtDNA of being non-recombinant. Since the X chromosome is present with one copy in males and two copies in females, it is easy to define haplotypes in males, and from that to resolve the phase ambiguity of females.

Comparing the degree of mtDNA genetic diversity with the nuclear genome can be a difficult challenge, especially in the case of *H. niger* whose nuclear genome is not yet characterized. Once in our study the nuclear DNA was established as a control to the mtDNA phylogeographic patterns, it became necessary to define genes that could be comparable to the extensive degree of genetic polymorphism observed in whole mitochondrial genomes. For that reason, six nuclear intron sequences of three protein-coding X-linked genes were selected and tested: HPRT (hypoxanthine phosphoribosyltransferase 1), PGK1 (phosphoglycerate kinase 1) and PDHA1 (pyruvate dehydrogenase alpha 1) (Table 1). Because there were no sequences of sable antelope nuclear DNA available in the literature and in Genbank database, primers were design from *Bos taurus* (cattle/ domestic cow) chromosome X sequence. The primers were design, using the Primer 3 software (http://frodo.wi.mit.edu/primer3/input.htm), in order to anchor exons, which are conservative, and flank introns, which are much more variable (Table 1).

Table 1- Primer sequences used to amplify several fragments of the HPRT, PGK1 and PDHA1 genes in sable antelope.

Nuclear Genes	Region in the Genome amplified	Fragment length (bp)	Primer Foward (F)	Primer Reverse (R)
HPRT	Exon2F-Exon3R	1635	5' TAGCGATGATGAACCAGG	5´ CATCTCCTTCATCACATC
	Exon7F-Exon9R	801	5'CCCTCGAAGTGTTGGATA	5´TCGCTAATGACACACACG
PGK1	Exon5F-Exon6R	809	5'TAGAAGCCT TCCGAGCTTCA	5'GGCCTTGGCAAAGTAGTTCA
	Exon9F-Exon11R	794	5'GGGCTGCATCACCATCATAG	5'CATCCACTCCAGGAAGCACT
PDHA1	Exon6F-Exon7R	974	5'GGATTGCTCTGGCCTGTAAG	5'TGCTGCTCT TTCCACAGATG
	Exon3F-Exon4R	995	5'GGGCTCAAATACTACAGGA	5'GACAGAGAGTCCACGAGT

With the exception of a 660 bp fragment of the PGK1 intron sequences (Exon5F-Exon6R), no genetic diversity was found in the tested nuclear introns, which were discarded from this work (data not shown). PGK1 intron was then analyzed and we could prove that it did not present evidences of intragenic recombination, thus overcoming one of the main limitations associated to single-copy nuclear genes, and providing almost the same historical clarity characteristic of non-recombining mitochondrial genomes. Nonetheless, the studied fragment has only 4% of the mitochondrial genome length (Table 2; also found in supplementary information: Table 6_SI). Furthermore, only four *H. niger* haplotypes were found in the PGK1 intron fragment, which represent less than 10 % of those found in complete mitochondrial genomes. Ideally several independent nuclear introns should have been included in our analysis. The technical and biological hurdles of extracting explicit genetic data from nuclear DNA sequences turned out to be impassable for this work, as

reflected by the low genetic polymorphism of PGK1 and complete lack of variation in the five other introns. This could be overcome by the whole genome sequencing of the sable antelope, which is currently on-going and will provide access to multiple independent loci- dozens or even hundreds of nuclear intron sequences - across the nuclear genome. As highlighted by Hailer et al., (2012), multilocus genomic analyses are crucial for an accurate understanding of evolutionary history. These authors were able to sequence and analyze 9116 nucleotides from 14 independent nuclear loci (introns) of polar, brown and black bears. Using this extensive dataset, and comparable to the mitochondrial genomes of the same species, they were able to confirm previous suggestions that polar bears carry introgressed brown bear mitochondrial DNA due to past hybridization. Therefore, using at least the same amount of introns we hope to provide the same clarity for the sable antelope's evolutionary history.

Despite the nuclear intron used in this study did not present enough resolution to retrieve any conclusions regarding the phylogeographic structure of *H. niger* in Africa, it ended up being a valuable control to the mtDNA pattern, reflecting the absence of the mtDNA introgressed haplogroup at the nuclear level (Figure 1; also found in supplementary information: Figure 3_SI). In fact, the level of divergence exhibited by the mitochondrial genome should leave a signal at the nuclear genome. Also, evidences from unpublished microsatellite data (Vaz Pinto et al., unpublished) gave us the support to corroborate our hypothesis for the presence of the highly divergent haplogroup in morphologically monotypic sables from Tanzania as the result from an introgression driven by an extinct species. Moreover, such confirmation of the absence of *Hippotragus sp* in the nuclear genome allowed us to exclude other hypotheses, such as incomplete lineage sorting.

Table 2 – DNA polymorphism measures calculated for the complete mtDNA dataset, excluding a total of 429 bp of the hypervariable regions, and for the nuclear intron PGK1 dataset (*H. equinus* sequences included in both datasets)

	MtDNA	PGK1
Number of sequences used	217	273
Total number of sites	16114	660
Total number of sites (excluding sites with gaps / missing data)	16084	659
Sites with alignment gaps or missing data	30 (29 gaps+ 1 N at position 4751)	1
Invariable (monomorphic) sites *	14147	652
Variable (polymorphic) sites*	1937	7
Singleton variable sites*	49	0
Parsimony informative sites*	1888	7

Note: Although the X-linked intron PGK1 was not sequenced in historic samples, the number of sequences for PGK1 is higher than those of mtDNA because each female specimen gives the information for two sequences. The * stands for parameters that do not account for insertion/deletions (gaps) and missing data (N).

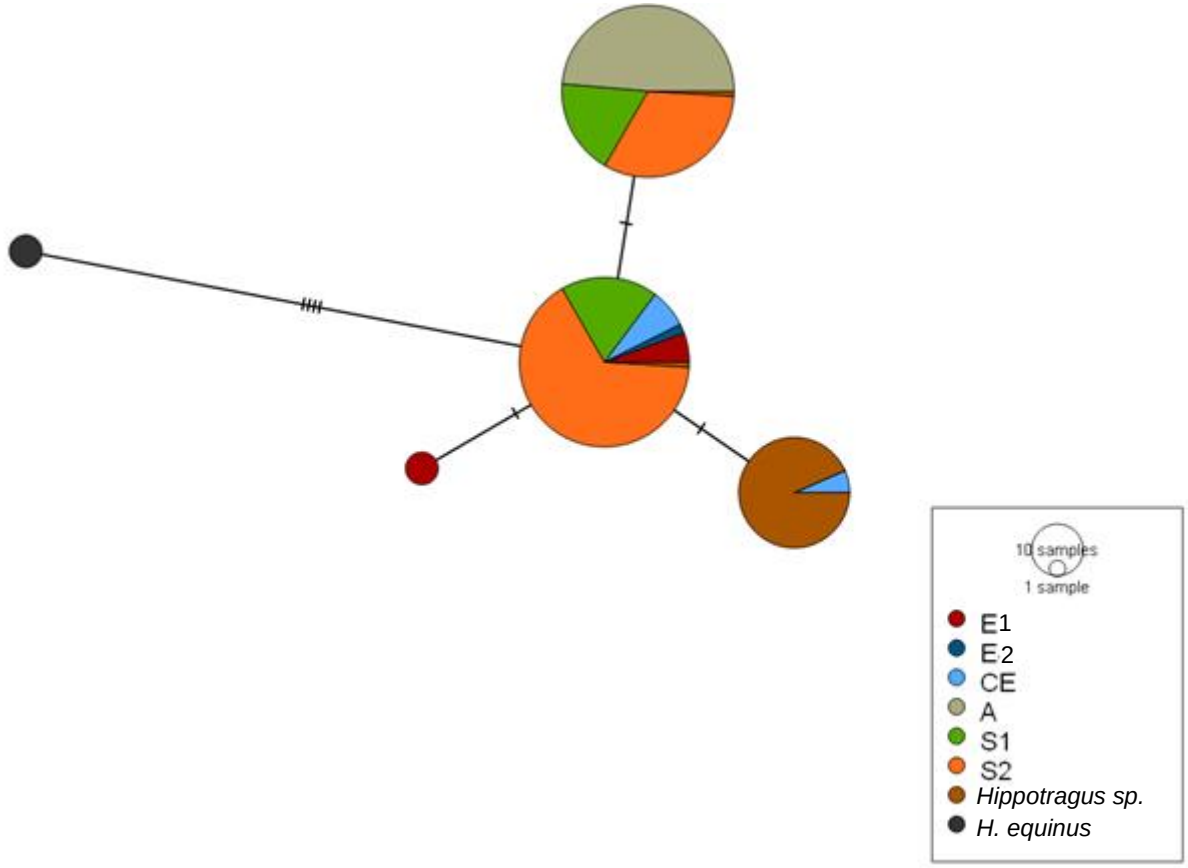

Fig. 1- Median-joining network of the PGK1 intron sequences, including all haplotypes of *H. niger* and one haplotype belonging to the outgroup, *H. equinus*. Given that the same dataset than mtDNA was used, individuals were labeled accordingly to the respective mtDNA haplogroup. The individuals carrying the highly divergent haplogroup from Tanzania in mtDNA (*Hippotragus sp.*) share the same haplotypes with individuals belonging to the reminder *niger*-like mtDNA haplogroups, with no extreme intraspecific divergence being observed in the nuclear data.

II- Implications of whole mitochondrial genome sequencing for population genetics and phylogeography: whole mitochondrial genomes vs Control Region

In order to compare the genetic information gained from the use of complete mitochondrial genomes with that from the mtDNA control region (CR), some preliminary analyses focused just on this shorter region (1094 bp when insertion/deletions are excluded) were performed for the same dataset used in mitogenomic analyses. We estimated the same genetic divergence and diversity measures used for analyses on complete mitochondrial genomes (Table 3 and 4), and constructed a neighbor-net to assess the phylogenetic relationships between *H. niger* CR sequences (Figure 2).

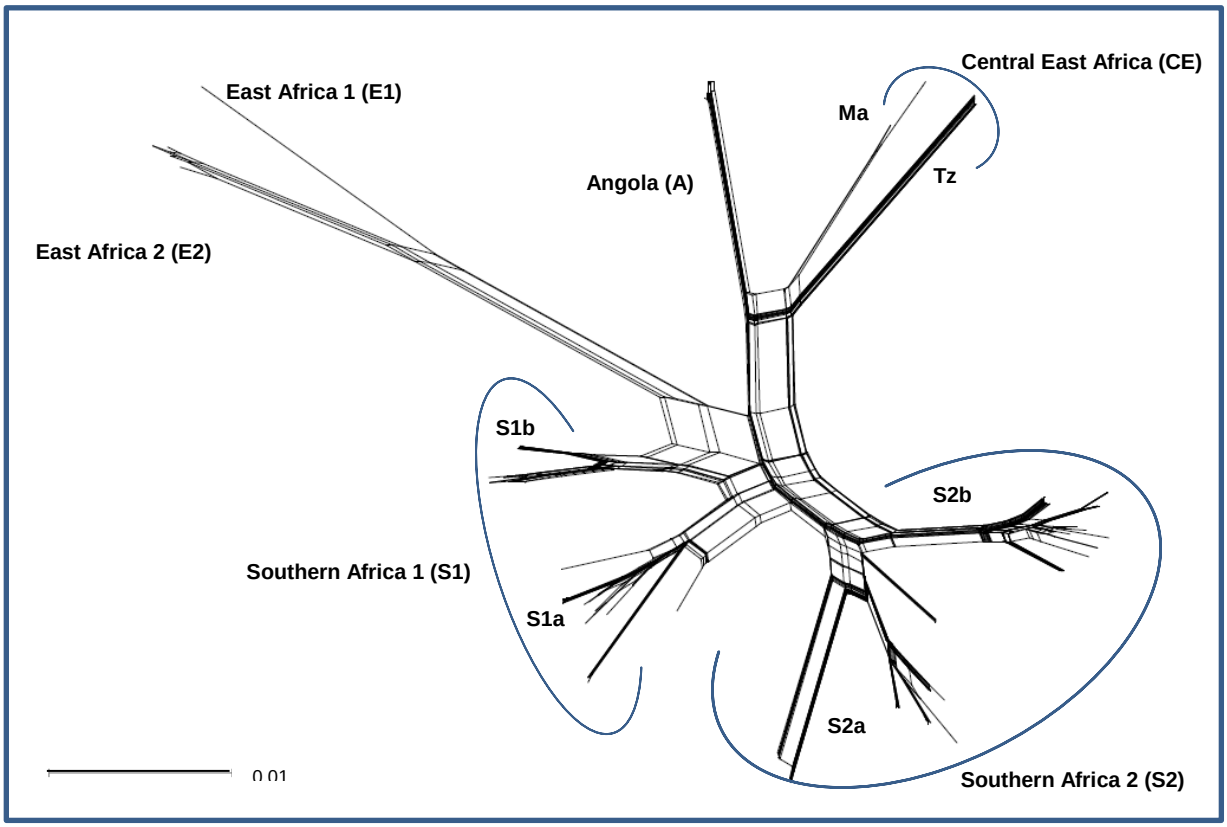

Fig.2- Neighbor-Net network based on uncorrected patristic distances implemented in SplitsTree. Branch lengths are proportional to genetic distance. Scale bar represents sequence divergence. The different *H. niger* haplogroups and sub-haplogroups are highlighted in bold with their respective names and codes.

As expected, the phylogenetic relationships of the different *H. niger* haplogroups using the mtDNA CR did not differ from those inferred from compete genome sequences (Figure 2). However, divergences between haplogroups are much higher when using the mtDNA CR (Table 3; comparable to Table 8_SI in Suppl. Info.). For instance the average divergence between *niger*-like haplogroups was approximately 0.7 % when using complete sequences, which is about six times lower than that estimated for the CR (see also Figure 3-A). The divergence of 12.7 % between *niger*-like haplogroups and the highly divergent *Hippotragus sp.* haplogroup is consistent with that found by Pitra et al. (2002), also using mtDNA CR. This clearly contrasts with the 3% divergence estimated from complete mitochondrial genomes. Given that the 3% divergence between *Hippotragus sp.* and *niger*-like haplogroups was estimated to be around 1.7 mya, then a 12.7% divergence means that these two clades would have split considerably earlier. Similarly, estimates of divergence times between *niger*-like haplogroups would be biased towards earlier dates. This is of utmost importance, as explaining phylogeographic patterns based on time-related processes, such as the climate and geomorphological changes that occurred during the Pleistocene, is dependent on the accurate estimation of divergence times. Therefore, conclusions regarding divergence times should be carefully drawn in analyses focused on short mtDNA fragments.

Such contrasts in divergence estimates can be easily explained by the high mutation rate associated to the control region, in particular to the hypervariable regions (Galtier et al., 2006), as well as by the shorter size of this mtDNA fragment. Briefly, the divergence between two sequences can be roughly estimated by dividing the number of polymorphic sites by the total number of sites (sequence length). Given the highly polymorphic nature of the control region, in contrast to its shorter length, which represents about 7% of the complete mitogenome, it is expected that divergence estimates are substantially higher than over the complete mitogenome. Therefore, our results provide evidence for an overestimation of genetic distances in previous studies of *H. niger* using the mtDNA control region (e.g. Pitra et al., 2002, 2006).

Table 3- Genetic divergence between different *H. niger* haplogroups, in comparison to the highly divergent clade from Tanzania (*Hippotragus sp.),* using control region (CR) sequences.

Comparisons	Dxy (JC)	Da (JC)	Fd.
A vs E1	0.05739 ± 0.02024	0.05721 ± 0.00204	60
A vs E2	0.05417 ± 0.01794	0.05294 ± 0.01794	53
A vs CE	0.02883 ± 0.00852	0.01971 ± 0.00864	15
A vs S1	0.03566 ± 0.00521	0.02756 ± 0.00514	17
A vs S2	0.04042 ± 0.00424	0.03210 ± 0.00425	18
E1 vs E2	0.03131 ± 0.01352	0.03027± 0.01353	31
E1 vs CE	0.05880 ± 0.02343	0.04986 ± 0.02347	45
E1 vs S1	0.05241 ± 0.01099	0.04450 ± 0.01100	35
E1 vs S2	0.05696 ± 0.00956	0.04883 ± 0.00957	33
E2 vs CE	0.05806 ± 0.01757	0.04807 ± 0.01764	42
E2 vs S1	0.05055 ± 0.00848	0.04159 ± 0.00851	31
E2 vs S2	0.05920 ± 0.00865	0.05003 ± 0.00867	33
CE vs S1	0.03945 ± 0.00629	0.02258 ± 0.00646	9
CE vs S2	0.04217± 0.00562	0.02509 ± 0.00580	11
S1 vs S2	0.03117 ± 0.00216	0.01513 ± 0.00219	3
Mean all *H. niger*	0.04644 ± 0.01129	0.03770 ± 0.01367	29
H. niger vs *Hippotragus sp.*	0.12689 ± 0.00536	0.10534 ± 0.00539	64

Abbreviations as follows: Average number of nucleotide substitutions (Dxy) and Number of net nucleotide substitutions (Da) per site between populations, with Jukes and Cantor correction (JC); Fd. Number of fixed differences. For abbreviations of haplogroup names, see Figure 2 from chapter II.

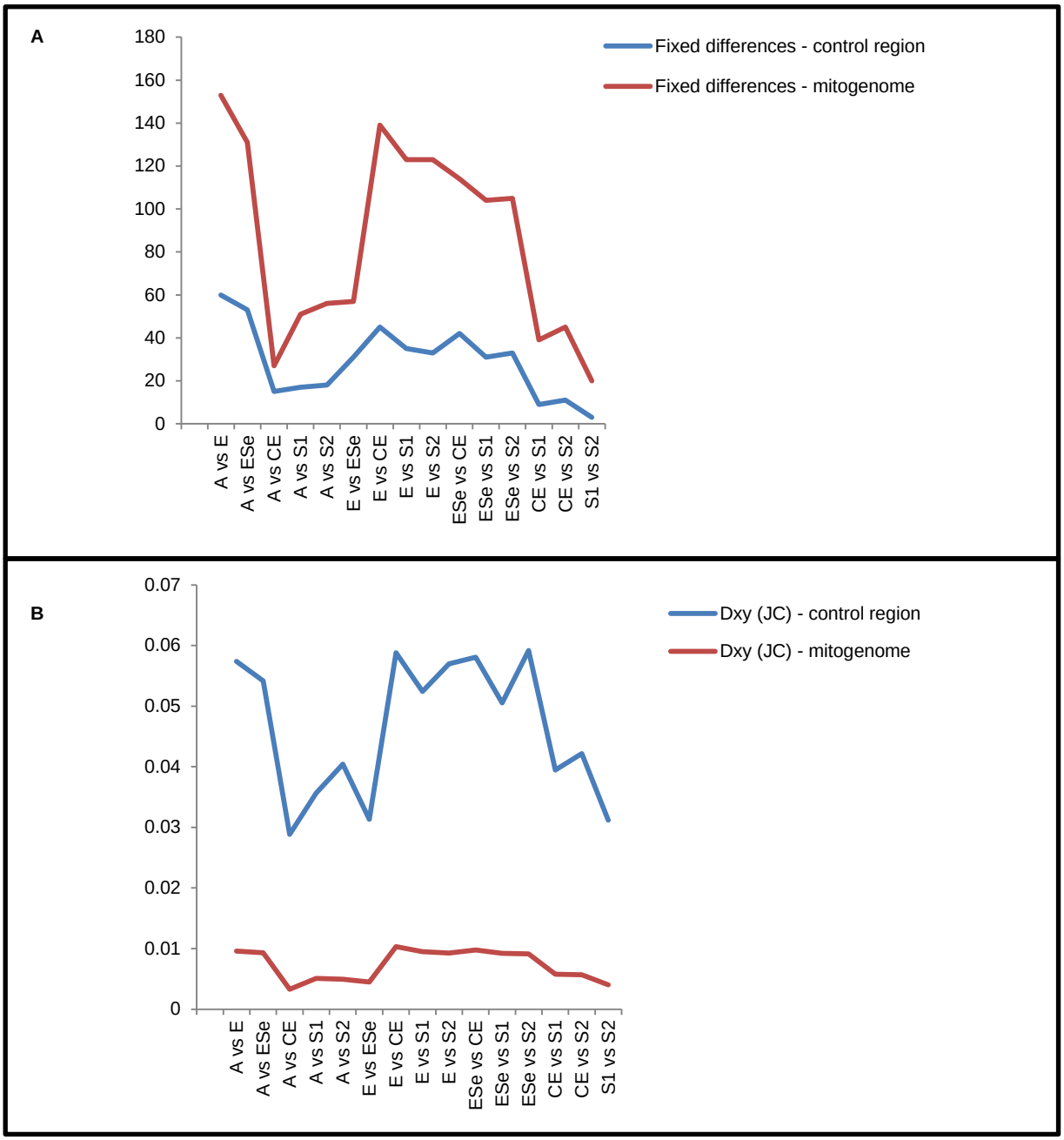

Fig.3- Genetic divergence between different *H. niger* haplogroups (*niger*-like): A-Average number of nucleotide substitutions per site between populations (Dxy), with Jukes and Cantor correction (JC), estimated for the control region and complete mitochondrial genomes (mitogenomes); B- Number of fixed differences estimated for the control region and complete mitochondrial genomes (mitogenomes). For abbreviations of haplogroup names, see Figure 2 from chapter II.

While divergence estimates were higher for the CR, the number of fixed differences among haplogroups was substantially lower (Table 3 Figure 3-B). For example, while 406 fixed differences were found between haplogroup *Hippotragus sp.* and *niger*-like for complete mtDNA sequences, only 64 fixed differences were found for the CR. This can be easily associated to the shorter length of the CR in comparison to the whole mitogenome, showing that adding more segments to the analyses can significantly increase the number of fixed differences between sequences. The same pattern was found regarding the number of

segregating sites (Table 4; comparable to Table 2 in chapter II). According to Table 4 only 159 polymorphic sites were found for the entire CR dataset based on *niger*-like sequences, which clearly contrasts to the 535 polymorphic sites found in complete mtDNA sequences. Similarly, only 59 polymorphic sites were found among *Hippotragus sp.* CR sequences, as opposed to the 136 found in the complete mitogenome. This shows that increasing sequence length allowed for, approximately, 4 times more polymorphic sites than those obtained using shorter mtDNA fragments. This increment of information provided higher phylogenetic resolution to accurately resolve the intraspecific relationships of *H. niger* in Africa.

While using complete genomes seems to have been fundamental to accurately estimate divergence between haplogroups, it did not seem to have a major impact in assessing intraspecific diversity. Most of the *niger*-like haplogroups revealed the same number of haplotypes and levels of haplotype diversity (Table 4). Nucleotide diversity and mean pairwise differences were underestimated, though, as a consequence of the lower number of polymorphic sites found in the CR. Of the 57 haplotypes found in complete mitogenomes from all *niger*-like sequences, 50 were discriminated using the CR. This means that only 7 haplotypes were gained from the analysis of complete sequences. These included one haplotype in Angola, two haplotypes in the central east African haplogroup (haplogroup including haplotypes from Tanzania and Malawi), one haplotype in the southern African haplogroup classified as S1 (haplogroup including haplotypes from Congo, Zambia, Zimbabwe and southern Mozambique) and three haplotypes in the southern African haplogroup classified as S2 (haplogroup including haplotypes from Zambia, Zimbabwe, southern Mozambique, Namibia and Botswana). Similarly, of the 29 haplotypes estimated for *Hippotragus sp.* using complete mitochondrial sequences, 23 were discriminated for the CR, which means that only 6 haplotypes would be disregarded in analyses using CR sequences.

To conclude, this study can be used as evidence of the potential impact of mitogenomes-scale information in addressing questions of divergence and diversity at a population scale. With the development of target-enrichment strategies for next generation sequencing, generating hundreds of fully sequenced mitochondrial genome is becoming an easy and relatively inexpensive task. According to our results, longer sequences provide greater power of phylogenetic inference. Therefore, we strongly advise the use of complete mitochondrial genomes in population studies that wish to address complex intraspecific relationships, or simply to provide higher phylogenetic resolution, and more precise divergence estimates than those obtained using traditional short fragments, such as cytochrome b (cytb) and the control region (CR).

Table 4- Genetic diversity summary statistics for control region mtDNA sequences

Haplogroups	N	S	h	Hd	π	MPD
***Hippotragus* sp**	46	59	23	0.942 ± 0.018	0.01085 ± 0.00555	11.828 ± 5.454
***niger*-like haplogroups**	169	159	50	0.948± 0.008	0.03254 ± 0.01576	35.506 ± 15.532
A	36	1	2	0.386 ± 0.074	0.00035 ± 0.00038	0.386 ± 0.372
E1	6	1	2	0.333 ± 0.215	0.00002 ± 0.00003	0.333 ± 0.380
E2	6	7	5	0.933 ± 0.122	0.00288 ± 0.00199	3.133 ± 1.884
CE	8	38	3	0.714 ± 0.123	0.01907 ± 0.01076	20.714 ± 10.259
Ma	4	0	1	0.000 ± 0.000	0.00000 ± 0.00000	0.000 ± 0.000
Tz	4	7	2	0.667 ± 0.204	0.00043 ± 0.00032	4.667 ± 2.885
S1	36	56	18	0.900 ± 0.033	0.01618 ± 0.00818	17.584 ± 7.996
S1a	23	36	13	0.893 ± 0.050	0.00923 ± 0.00488	10.032 ± 4.760
S1b	13	15	5	0.538 ± 0.161	0.00541 ± 0.00310	5.872 ± 2.999
S2	77	69	20	0.908 ± 0.015	0.01633 ± 0.00812	17.770 ± 7.976
S2a	38	38	9	0.815 ± 0.038	0.01148 ± 0.00589	12.492 ± 11.5.764
S2b	39	30	11	0.811 ± 0.044	0.00629 ± 0.00336	6.842 3.291

Abbreviations as follows: number of samples; N, Number of Samples; S, number of polymorphic (segregating) sites; h, number of haplotypes; Hd, Haplotype diversity; π, nucleotide diversity; MPD, mean pairwise distance. All parameters were estimated accounting for insertion/deletions and allowing 5% of missing data. For abbreviations of haplogroup names, see Figure 2 from chapter II.

Conclusions and Future work

In conclusion, we have investigated the maternal history of several *H. niger* populations. Our results, using complete mitochondrial genomes covering the species full geographic range in Africa, demonstrate that it is very unlikely the hypothesis proposed by Pitra et al. (2002) for the presence of a highly divergent clade from Tanzania explained only within *H. niger*. Instead, we propose the hypothesis that the sables from Tanzanian populations carrying the highly divergent haplogroup experienced gene flow from a currently extint species that diverged from *H. niger* around 1.7 mya. We also provide, for the very first time, insights towards the origin of the giant sable antelope of Angola. Ultimately, we demonstrate that using complete mitochondrial genomes can provide higher phylogenetic resolution, and more precise divergence estimates to assess intraspecific relationships. In such context, studies based on single genes that so far have not been able to provide a fully resolved phylogeographic pattern for their study case would substantially benefit from the alternative use of fully sequenced genomes. More importantly, studies that wish to associate divergence to time-related phenomena will benefit from more precise estimates of divergence times.

The great majority of phylogeographic studies focused their analysis on the mitochondrial genome, and mtDNA still stands as the most commonly used marker in

phylogeography. Nonetheless, it is undeniable that this study will benefit from the integration of mtDNA with nuclear or multilocus DNA data. Until recently, isolating nuclear haplotypes from diploid individuals and dealing with the intragenic recombination associated to the nuclear genome was a very difficult and laborious task (Avise, 2009). Now, with the current high-throughput sequencing technologies and bioinformatic pipelines, it is possible to recover gene genealogies from multiple unlinked nuclear loci fast and at a relatively low cost. In such context, a much fuller perspective on *H. niger* evolutionary history could be gained by adding nuclear data. This is currently being driven by the on-going project of the whole nuclear genome sequencing of the sable antelope. The access to multilocus genomic data will allow one to search for patterns of inter-locus genealogical concordance that could, ultimately, provide richer and more accurate interpretations of the evolutionary history of *H. niger* in the woodland savannas of central, southern and eastern Africa.

In other words, the mitochondrial DNA is an excellent marker to understand evolutionary patterns, but evolutionary processes are better explained by multilocus markers. In such context the work provided by this study should be further complemented with multilocus nuclear data. This does not compromise mtDNA status as election marker in phylogeography, but rather that it should be used in an integrative approach with multiple loci to even further understand the evolutionary history of the sable antelope in Africa.

References

Avise, J.C. (2009). Phylogeography: retrospect and prospect. J. Biogeogr. *36*, 3–15.

Galtier, N., Enard, D., Radondy, Y., Bazin, E., and Belkhir, K. (2006). Mutation hot spots in mammalian mitochondrial DNA. Genome Res. *16*, 215–222.

Hailer, F., Kutschera, V.E., Hallström, B.M., Klassert, D., Fain, S.R., Leonard, J. a, Arnason, U., and Janke, A. (2012). Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. Science *336*, 344–347.

Palumbi, S., and Baker, C. (1994). Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol. Biol. Evol. *11*, 426–435.

Pitra, C., Hansen, A.J., Lieckfeldt, D., and Arctander, P. (2002). An exceptional case of historical outbreeding in African sable antelope populations. Mol. Ecol. *11*, 1197–1208.

Pitra, C., VazPinto, P., O'Keeffe, B.W.J., Willows-Munro, S., Jansen van Vuuren, B., and Robinson, T.J. (2006). DNA-led rediscovery of the giant sable antelope in Angola. Eur. J. Wildl. Res. *52*, 145–152.

Supplementary Information

I-Supplementary Information for Material and Methods

A-Texts

Text 1_SI: About 800 ng of modern DNA was sheared (Biorruptor, Diogenode, Liege, Beligum) 7 times for 5 minutes, fragmenting the DNA to a range of 200 to 400 base-pairs. Of the 800 ng of sheared DNA, 200 ng were used to produce the indexed libraries as described in Meyer and Kircher (2010). Then, the indexed libraries were purified using a MinElute PCR purification kit (Qiagen, Hilden, Germany), NanoDrop quantified (NanoDrop ND-1000, Thermo Scientific), and pooled in equimolar amounts to a total of 2 ug for capture.

Text 2_SI: The DNA double stranded libraries were amplified using the AccuPrime Pfx Polymerase (Invitrogen) with the following reaction set up: 1x AccuPrime Pfx buffer, 1 uM P5 indexing primer, 1uM of P7 indexing Primer (primers synthesized by Sigma-Aldrich, RPC-purified, dissolved in TE – 10 mM Tris-HCL, 1mM EDTA, pH 8.0 - at 100uM and diluted 10-fold in water), and 1.25 units AccuPrime Pfx polymerase. The thermocylcing conditions were the following: 2 min at 95ºC, 20 cycles of 20 sec at 95ºC, 30s at 60ºC, and 1 min at 68ºC, with a final extension of 5 min at 68ºC. Reactions were purified using MinElute PCR purification kit (Qiagen) and eluted in 15uL TE buffer.

Text 3_SI: PCR cyclic conditions were different for each pair of primers. The thermal cycler program for primer pair Hn1377F-Hn6895R consisted in an initial denaturation step of 2 min at 92°C, followed by a touchdown program with 9 cycles of a denaturation step at 92°C for 10s, annealing at 60°C for 20s, decreasing 0.5°C in each cycle, and extension at 68°C for 6min, followed by 21 cycles of 92°C for 10s, 56°C for 20s and 68ºC for 6min. The same cycling conditions were applied to the primer pair Hn6538F-Hn12024R, but using a total number of 37 cycles instead (9+28). PCR conditions for primer pair Hn10583F-Hn14756R consisted in an initial denaturation step of 2min at 92°C, followed by 9 cycles of a denaturation step at 92°C for 10s, annealing at 66°C for 20s, and extension at 68°C for 6min, followed by 21 cycles of 92°C for 10s, 62ºC for 20s and 68ºC for 6min. PCR conditions for primer pair Hn14637F-Hn3305R consisted in an initial denaturation step of 2min at 92°C, followed by 9 cycles of a denaturation step at 92°C for 10s, annealing at 63°C for 20s, and extension at 68°C for 6min, followed by 21 cycles of 92°C for 10s, 59ºC for 20s and 68ºC for 6min. The final extension was at 68°C for 6 min for all primers (see also Table 4_SI).

Text 4_SI: Bait and library pool preparation for capture were performed differently than Fu et al. (2013). First, 800 uL (the equivalent of 20 uL per capture reaction) of Dynabeads

MyOneC1 (Life Technologies) were pelleted in a magnet rack, and washed twice in 1000 uL of BWT buffer (1M NaCl, 10mM Tris-HCL, 1mM EDTA, pH 8.0, 0.05 % Tween-20). Then the beads were resuspended in 800 uL of BWT (the correspondent volume required for each capture reaction), and 47 uL of bait were added to the tube containing the beads (volume of bait necessary to obtain 500 ng of bait per capture reaction). The tube containing the beads/bait mixture was rotated for 15 minutes at room temperature. The beads were pelleted in a magnetic rack and washed with 1 mL TT buffer (1mM Tris-HCl pH 8.0, 0.01 % Tween-20). Then the beads were twice resuspended in 1 mL of melt solution (prepared fresh: 125 mM NaOH, 0.01% Tween-20), rotated for 5 min at room temperature and pelleted again using a magnetic rack. Finally, the beads were washed with 1mL TT buffer, resuspended in 800 uL of BWT buffer (volume correspondent to 20 uL times the number of capture reactions), dispensed in aliquots of 20 uL to wells of a 96-well semi-skirted PCR plate (ABI) and stored in the fridge. For each capture/hybridization reaction a sample library pool was created by combining 3.42 uL of sample library (~2 ug), 14.58 uL of nuclease-free water, 0.5 uL 500 uM BO4 (GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-Phosphate), 0.5 uL 500 uM BO6 (CAAGCAGAAGACGGCATACGAGAT-Phosphate), 0.5 uL 500 uM BO8 (GTGTAGATCTCGGTGGTCGCCGTATCATT-Phosphate), 0.5 uL 500 uM BO10 (AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-Phosphate), 5 uL of Agilent blocking agent and 25 uL of 2x HI-RPM hybridization buffer (Agilent). The sample library pool was incubated for 3 min at 95 ºC and held at 37ºC for 5 min. Meanwhile, the beads were pelleted on a magnetic rack, discarding the supernatant. Then, the entire library pool was added to the beads and incubated for 2 days at 65 ºC in a hybridization oven, with a rotation of 12 rpm. Post hybridization washes and elution (equivalent to the third day of capture) were performed exactly as described in Fu et al. (2013). In order to check the successful retrieval of library molecules, 1 uL of each capture eluate was quantified by qPCR. The remaining eluate (29 uL) was then amplified as described by Fu et al. (2013) but using the primers IS5 (10uM) and IS6 (10uM) instead of the genomic primers R1 and multiplex R2, and using 30 cycles instead of the 29 cycles described in the paper for the thermal cycling conditions. Amplified libraries were purified exactly as described in Fu et al. (2013). The amplified capture eluates were eluted in 20 uL EB (10 mM Tris-HCl, pH 8) and their concentrations measured by NanoDrop. After the capture round, libraries were prepared for sequencing also as described by Fu et al. (2013). No second round of capture was required.

Text 5_SI: All reads were mapped using a customized version (see https://github.com/udo-stenzel/network-aware-bwa) of the BWA v.0.5.10 (Li and Durbin, 2009) mapper. The reads of modern samples were mapped with the seeding turned off, allowing a 0.00001 missing probability and 2 gap opening events (options "-n 0.00001 -o 2 -l 16500"). The reads of

historic samples were also mapped with the same parameters but with a 0.01 missing probability instead as reads were shorter and the probability of incorrectly mapping molecules was higher. As BWA does not recognize that the mitochondrion is a circular molecule, the first 1200 base pairs of the reference genome were copied to the end of the sequence to obtain an even coverage across the mitochondrial genome. Reads landing within the junction where then split in two and reads falling entirely within the last copied segment were repositioned within the first 1200 bp. Putative PCR duplicates were removed and a consensus was called for each set of duplicated reads. Both aforementioned tasks were performed using in-house scripts (bam-rewrap and bam-rmdup respectively from https://github.com/udo-stenzel/biohazard).

Text 6_SI: Given the alignment to the mitochondrial reference, a consensus call was required for downstream analyses. Using a Bayesian maximum *a posteriori* approach (https://github.com/grenaud/schmutzi[g1]), the more likely base along with its error probability was produced for both modern and historic samples. Briefly, the probability of observing the data given a specific nucleotide was computed by treating each read as an independent observation. The probability of observing a certain base in a given read was computed using the quality scores for that base and the probability of mismappings provided by the mapping quality. To minimize the contribution of deaminated bases in historic samples, the probability of certain nucleotide substitutions at a given position in the sequence was incorporated. The frequency of nucleotide substitution at given positions in the sequence was computed. This frequency was turned into a position specific substitution matrix which in turn, was used as input for the mitochondrial consensus caller. As residual deamination enriches for transitions rather than transversions, the ratio of transitions over total mutations was plotted for both historic and modern samples (Figure 1_SI). No significant enrichment for historic samples was evidenced by these plots. A threshold on the base error probability of 3.16e-13 (approximately 125 on the PHRED scale) was used to mask low quality or low coverage bases. Such a confidence threshold is usually achieved after having at least 3 high-quality reads supporting a position. The resulting consensus was used in downstream analyses.

Text 7_SI: We used the primers L15910-5′GAATTCCCCGGTCTTGTAAACC (Hoelzel et al., 1991) and H16498-5′CCTGAAGTAGGAACCAGATG (Shields and Kocher, 1991) to amplify an 1100 bp fragment, correspondent to the 5’ end of the control region. PCR amplifications were performed following the Bioline MyTaq HS DNA Polymerase protocol on a T100- Cycler (Bio-Rad). Amplifications were performed in a final volume of 10 uL using approximately 10-20 ng of DNA (1uL of DNA extraction), 5 uL of Bioline PCR MyTaq/Qiagen PCR MasterMix, and 0.4 uL for each primer (0.4uM). PCR conditions consisted of an initial denaturation step

of 10min at 95°C, followed by a touchdown program with 5 cycles of a denaturation step at 95°C for 30s, annealing at 58°C for 30s, decreasing 0.5°C in each cycle, and extension at 72°C for 1min, followed by 35 cycles of 95°C for 30s, 56°C for 30s and 72°C for 1min. The final extension was at 72°C for 10min. The PCR products were cleaned-up by adding 1 uL of ExoSAP mix to each PCR sample [0.025 uL Exonuclease I; 0.250 uL SAP; 9.725 uL Milli Q water (final volume of 10 uL; 1XExoSAP mix)] (Affymetrix). Then, the PCR samples were incubated in a PCR machine at 37˚C for 15 minutes, and then at 85˚C for 15 minutes. Single PCR bands were sequenced for both strands using the Cycle sequencing protocol BigDye Terminator v3.1 (Applied Biosystems). The Cycle sequencing thermo cycling program consisted in 24 cycles of an initial denaturation step of 3min at 94°C, followed by annealing at 96°C for 10s, and extension at 55°C for 5s. The final extension was at 60°C for 4min. After the cycle sequencing reaction, the sequenced products were cleaned with 400 uL of Sephadex (66.7g/L). After this treatment, the sequences were dried, and then treated with 15 uL of Formamid. Then, the sequenced products were separated and analyzed by capilar electrophoresis in a 3130 XL Genetic Analyzer. Both forward and reverse sequences were aligned, corrected and compared in the software Seq-Scape 2.0 (Applied Biosystems).

Text 8_SI: In order to correct the control region (CR) for all modern samples we first aligned the consensus mtDNA sequence obtained from NGS data and its respective Sanger sequenced CR sequence with ClustalW v.2.0.10 (Larkin et al., 2007).For regions of the complete mitochondrial genome that did not overlapped with the Sanger CR, bases with an error probability greater than 125 on a PHRED scale were masked. For the parts of the mitochondrial genome Sanger sequences and NGS sequencing reads were available, the following was used to call a consensus: 1) First, for the positions in which there was a gap in the NGS CR but none in the Sanger, the base given by the Sanger sequence was retained. This is because Sanger sequences are less likely to have spurious insertions or deletions than NGS reads. 2) Secondly, when there was match between the Sanger sequence and the NGS, that base was retained regardless of the NGS error probability. 3) When both bases did not match, the NGS base was retained if its quality score exceeded 125 (PHRED scale) otherwise, the Sanger base was used. 4) When there were gaps in the Sanger sequence but the NGS sequence had a base, the NGS base was called if its quality score exceeded 125; if not, no bases were produced. 5) Finally, for the situations in which the NGS sequences did not had enough coverage to call a base, the Sanger bases were retained. This approach was done for all individuals within the modern dataset, to avoid biasing our analyses towards certain individuals, and make sure that all modern sequences received the same treatment.

Text 9_SI:PCR amplifications of the X-linked intron PGK1 (phosphoglycerate kinase 1), selected as a control to the mtDNA pattern, were performed on modern samples using the

Bioline MyTaq HS DNA Polymerase protocol on a T100- Cycler (Bio-Rad), based on the manufacturer's instructions. This allowed the amplification of a fragment with 809 bp, correspondent to the region between Exon 5 and 6 of the gene (Exon5F-TA GAAGCCTTCCGAGCTTCA; Exon6R-GGCCTTGGCAAAGTAGTTCA). Amplifications were performed in a final volume of 10 uL using approximately 10-20 ng of DNA (1uL of DNA extraction), 5 uL of Bioline PCR MyTaq/Qiagen PCR MasterMix, and 0.4 uL for each primer (0.4uM). PCR conditions consisted of an initial denaturation step of 10min at 95°C, followed by a touchdown program with 9 cycles of a denaturation step at 95°C for 30s, annealing at 56°C for 45s, decreasing 0.5°C in each cycle, and extension at 72°C for 1min, followed by 31cycles of 95°C for 30s, 52°C for 45s and 72°C for 1min. The final extension was at 72°C for 10min. The PCR products were cleaned-up by adding 1 uL of ExoSAP mix to each PCR sample [0.025 uL Exonuclease I; 0.250 uL SAP; 9.725 uL Milli Q water (final volume of 10 uL; 1XExoSAP mix)] (Affymetrix). Then, the PCR samples were incubated in a PCR machine at 37˚C for 15 minutes, and then at 85˚C for 15 minutes. Single PCR bands were sequenced for both strands using the Cycle sequencing protocol BigDye Terminator v3.1 (Applied Biosystems). The Cycle sequencing thermo cycling program consisted in 24 cycles of an initial denaturation step of 3min at 94°C, followed by annealing at 96°C for 10s, and extension at 55°C for 5s. The final extension was at 60°C for 4min. After the cycle sequencing reaction, the sequenced products were cleaned with 400 uL of Sephadex (66.7g/L). After this treatment, the sequences were dried, and then treated with 15 uL of Formamid. Then, the sequenced products were separated and analyzed by capilar electrophoresis in a 3130 XL Genetic Analyzer. Both forward and reverse sequences were aligned, corrected and compared in the software Seq-Scape 2.0 (Applied Biosystems). The resultant alignment was realigned with ClustalW v.2.0.10 (Larkin et al., 2007), and trimmed so that all individuals in the alignment had the exact same sequence length.

Text 10_SI: For phase I analysis we created a dataset of almost complete mitochondrial genomes (excluding the control region), from a representative set of sequences belonging to the different *H. niger* haplogroups (37 sequences, including the *Hippotragus niger* reference genome JN632648.1) and *H. equinus* generated in this study, and 9 sequences imported from the Bovidae dataset published by Bibi (2013), belonging the following species: *Damaliscus pygarus*, *Alcephalus busephalus lichtensteinii*, *Connochaetes gnou*, *Addax nasomasculatus*, *Oryx beisa*, *Oryx gazella*, *Oryx leucocoryx*, *Oryx dammah* and *Hippotragus equinus* (sequences available on the TreeBase repository: http://treebase.org/treebase-web/search/study/summary.html?id=14132).

B-Tables

Table 1_SI:

Table 1 - Sample information: list of all modern samples with respective country and population of origin, subspecies according to the classification proposed by Ansell (1971), sex, material and downer.

Sample ID	Country	Local Population	Subspecies	Sex	Material	Sent by:
PA10	Angola	Cangandala	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
PA11	Angola	Cangandala	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
PA12	Angola	Cangandala	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
PA14	Angola	Cangandala	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
PA16	Angola	Cangandala	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
PA17	Angola	Cangandala	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
PA18	Angola	Cangandala	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
PA19	Angola	Cangandala	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
PA20	Angola	Cangandala	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN145	Angola	Cangandala	*Hippotragus niger variani*	-	Hair	Pedro Vaz Pinto
PA02	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
PA03	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
PA04	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
PA05	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
PA06	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
PA07	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
PA08	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
PA09	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
PA13	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
HN143	Angola	Luando	*Hippotragus niger variani*	-	Hair	Pedro Vaz Pinto
HN144	Angola	Luando	*Hippotragus niger variani*	-	Hair	Pedro Vaz Pinto
HN146	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN147	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN148	Angola	Luando	*Hippotragus equinus*	F	Tissue	Pedro Vaz Pinto

HN151	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN152	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN153	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
HN154	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
HN158	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
HN159	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN160	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN163	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN164	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN166	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN167	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
HN168	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN169	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN170	Angola	Luando	*Hippotragus niger variani*	F	Tissue	Pedro Vaz Pinto
HN171	Angola	Luando	*Hippotragus niger variani*	M	Tissue	Pedro Vaz Pinto
HN323	Kenya	Shimba Hills	*Hippotragus niger roosevelti*	F	Tissue	Hans Siegismund
HN325	Kenya	Shimba Hills	*Hippotragus niger roosevelti*	M	Tissue	Hans Siegismund
HN320	Kenya	Shimba Hills	*Hippotragus niger roosevelti*	M	Tissue	Hans Siegismund
HN268	Kenya	Shimba Hills	*Hippotragus niger roosevelti*	F	Tissue	Hans Siegismund
HN269	Kenya	Shimba Hills	*Hippotragus niger roosevelti*	F	Tissue	Hans Siegismund
HN270	Kenya	Shimba Hills	*Hippotragus niger roosevelti*	F	Tissue	Hans Siegismund
HN276	Kenya	Shimba Hills	*Hippotragus niger roosevelti*	F	Tissue	Hans Siegismund
HN347	Mozambique	Niassa	*Hippotragus niger roosevelti*	-	Skin	Marco Silva
HN348	Mozambique	Niassa	*Hippotragus niger roosevelti*	-	Skin	Marco Silva
HN349	Mozambique	Niassa	*Hippotragus niger roosevelti*	-	Skin	Marco Silva
HN83	Malawi	Liwonde National Park	*Hippotragus niger kirkii*	F	Blood	Andre Uys
HN84	Malawi	Liwonde National Park	*Hippotragus niger kirkii*	F	Blood	Andre Uys
HN85	Malawi	Liwonde National Park	*Hippotragus niger kirkii*	F	Blood	Andre Uys
HN86	Malawi	Liwonde National Park	*Hippotragus niger kirkii*	F	Blood	Andre Uys
G018	Zambia	Masebe, Mkushi	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren

G075	Zambia	Masebe, Mkushi	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
G076	Zambia	Masebe, Mkushi	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
G077	Zambia	Masebe-Mkushi	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
G079	Zambia	Masebe-Mkushi	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
G124	Zambia	Masebe-Mkushi	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G138	Zambia	Masebe-Mkushi	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
G180	Zambia	Masebe-Mkushi	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G182	Zambia	Masebe-Mkushi	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G230	Zambia	Masebe-Mkushi	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
G241	Zambia	Masebe-Mkushi	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G244	Zambia	Masebe-Mkushi	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G030	Zambia	Mufumbwe	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G064	Zambia	Mufumbwe	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
G065	Zambia	Mufumbwe	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
G067	Zambia	Mufumbwe	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
G273	Zambia	Mufumbwe	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G196	Zambia	Mulobezi	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G203	Zambia	Mulobezi	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G208	Zambia	Mulobezi	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G211	Zambia	Mulobezi	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
G236	Zambia	Mulobezi	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
HN208	Zambia	Lusaka-Kafue	*Hippotragus niger kirkii*	F	Tissue	Hans Siegismund
HN213	Zambia	Lusaka-Kafue	*Hippotragus niger kirkii*	F	Tissue	Hans Siegismund
HN215	Zambia	Lusaka-Kafue	*Hippotragus niger kirkii*	F	Tissue	Hans Siegismund
HN216	Zambia	Lusaka-Kafue	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN217	Zambia	Lusaka-Kafue	*Hippotragus niger kirkii*	F	Tissue	Hans Siegismund
HN312	Zambia	Mulobezi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN313	Zambia	Mulobezi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN273	Zambia	Mulobezi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN282	Zambia	Lusaka-Kafue	*Hippotragus niger kirkii*	F	Tissue	Hans Siegismund

HN283	Zambia	Lusaka-Kafue	*Hippotragus niger kirkii*	F	Tissue	Hans Siegismund
HN286	Zambia	Lusaka-Kafue	*Hippotragus niger kirkii*	F	Tissue	Hans Siegismund
HN290	Zambia	Lusaka-Kafue	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN291	Zambia	Lusaka-Kafue	*Hippotragus niger kirkii*	F	Tissue	Hans Siegismund
V04	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
V05	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
V06	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
V07	Zambia	Kafue National Park	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
V08	Zambia	Kafue National Park	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
V09	Zambia	Kafue National Park	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
V10	Zambia	Kafue National Park	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
V12	Zambia	Kafue National Park	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
V13	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
V14	Zambia	Kafue National Park	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
V15	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
V16	Zambia	Kafue National Park	*Hippotragus niger kirkii*	M	Tissue	Bettine van Vuuren
V17	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
V18	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
V19	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
V20	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
V30	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
V31	Zambia	Kafue National Park	*Hippotragus niger kirkii*	F	Tissue	Bettine van Vuuren
HN44	Tanzania	Wembere	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN45	Tanzania	Wembere	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN77	Tanzania	Wembere	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN263	Tanzania	Wembere	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN264	Tanzania	Wembere	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN304	Tanzania	Mlele	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN37	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN51	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund

HN52	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN53	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN54	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN55	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN56	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN57	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN74	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN75	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN81	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN329	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN331	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN334	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN335	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN327	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN328	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN199	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN200	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN306	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN302	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN303	Tanzania	Kigosi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN38	Tanzania	Kizigo	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN39	Tanzania	Kizigo	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN40	Tanzania	Kizigo	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN48	Tanzania	Kizigo	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN49	Tanzania	Kizigo	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN72	Tanzania	Kizigo	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN73	Tanzania	Kizigo	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN78	Tanzania	Kizigo	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN79	Tanzania	Kizigo	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN190	Tanzania	Rungwa	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund

HN319	Tanzania	Kizigo	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN265	Tanzania	Rungwa	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN277	Tanzania	Rungwa	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN279	Tanzania	Rungwa	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN295	Tanzania	Rungwa	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN296	Tanzania	Rungwa	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN297	Tanzania	Rungwa	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN298	Tanzania	Rungwa	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN299	Tanzania	Rungwa	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN32	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN33	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN58	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN62	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN64	Tanzania	Ugalla	*Hippotragus niger kirkii*	F	Tissue	Hans Siegismund
HN65	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN66	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN67	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN70	Tanzania	Ugalla	*Hippotragus niger kirkii*	-	Tissue	Hans Siegismund
HN80	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN82	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN330	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN332	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN184	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN185	Tanzania	Niensi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN186	Tanzania	Niensi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN187	Tanzania	Niensi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN188	Tanzania	Niensi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN189	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN201	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN318	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund

HN275	Tanzania	Niensi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN278	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN292	Tanzania	Niensi	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN305	Tanzania	Ugalla	*Hippotragus niger kirkii*	M	Tissue	Hans Siegismund
HN202	Botswana	Chobe	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN257	Botswana	Chobe	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN258	Botswana	Chobe	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN259	Botswana	Chobe	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN260	Botswana	Chobe	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN284	Botswana	Chobe	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN17	Namibia	Mahango	*Hippotragus niger niger*	-	Tissue	Mark Jago
HN18	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Mark Jago
HN222	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN223	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN224	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN225	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN226	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN227	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN228	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN229	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN230	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN231	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN248	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN249	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN250	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN251	Namibia	Mahango	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN252	Namibia	Mahango	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN253	Namibia	Mahango	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN254	Namibia	Mahango	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
SUN30	Namibia	Mahango	*Hippotragus equinus*	F	Tissue	Bettine van Vuuren

SUN652	Namibia	Mahango	*Hippotragus niger niger*	M	Tissue	Bettine van Vuuren
HN25	Zimbabwe	Matetsi	*Hippotragus niger niger*	F	Hair	Bruce Fivaz
HN26	Zimbabwe	Matetsi	*Hippotragus niger niger*	F	Hair	Bruce Fivaz
HN27	Zimbabwe	Matetsi	*Hippotragus niger niger*	F	Hair	Bruce Fivaz
HN28	Zimbabwe	Matetsi	*Hippotragus niger niger*	F	Hair	Bruce Fivaz
HN29	Zimbabwe	Matetsi	*Hippotragus niger niger*	F	Hair	Bruce Fivaz
HN30	Zimbabwe	Matetsi	*Hippotragus niger niger*	F	Hair	Bruce Fivaz
HN31	Zimbabwe	Matetsi	*Hippotragus niger niger*	F	Hair	Bruce Fivaz
HN31-2	Zimbabwe	Matetsi	*Hippotragus niger niger*	F	Hair	Bruce Fivaz
HN218	Zimbabwe	Triangle	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN219	Zimbabwe	Triangle	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN220	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN221	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN232	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN233	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN234	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN235	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN236	Zimbabwe	Triangle	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN237	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN238	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN239	Zimbabwe	Triangle	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN240	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN241	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN242	Zimbabwe	Triangle	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN243	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN244	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN245	Zimbabwe	Triangle	*Hippotragus niger niger*	F	Tissue	Hans Siegismund
HN246	Zimbabwe	Triangle	*Hippotragus niger niger*	M	Tissue	Hans Siegismund
HN247	Zimbabwe	Triangle	*Hippotragus niger niger*	M	Tissue	Hans Siegismund

Table 2_SI:

Table 2-Sample information: list of all historic samples, with respective age, country and population of origin, museum where the sample was collected, or the downer of the samples, material and subspecies according to the classification proposed by Ansell (1971).

Sample ID	Year	Country	Population	Museum information	Material	Subspecies
HnNI33	1992	Congo	Katanga	provided by Mike Buser	Tooth	*Hippotragus niger kirkii*
HnNI131	1975-1985	Congo	Katanga	provided by Michael Hasson	Tooth	*Hippotragus niger kirkii*
HnNI132	1975-1985	Congo	Katanga	provided by Michael Hasson	Tooth	*Hippotragus niger kirkii*
HnNI133	1975-1985	Congo	Katanga	provided by Michael Hasson	Skin	*Hippotragus niger kirkii*
HnNI145	1911	Zambia	undetermined	Peabody Museum of Natural History, Yale University , USA	Bone	*Hippotragus niger kirkii*
HnNI146	1911	Zambia	undetermined	Peabody Museum of Natural History, Yale University , USA	Bone	*Hippotragus niger kirkii*
HnNI149	undetermined	Zambia	undetermined	Peabody Museum of Natural History, Yale University , USA	Bone	*Hippotragus niger kirkii*
HnNI162	1900-1975	Mozambique	Gorongosa NP	Museum of Vila Viçosa, Portugal	Skin	*Hippotragus niger niger*
HnNI163	1900-1975	Mozambique	Gorongosa NP	Museum of Vila Viçosa, Portugal	Skin	*Hippotragus niger niger*
HnNI164	1900-1975	Mozambique	Gorongosa NP	Museum of Vila Viçosa, Portugal	Skin	*Hippotragus equinus*
HnNI165	1900-1975	Mozambique	Gorongosa NP	Museum of Vila Viçosa, Portugal	Skin	*Hippotragus equinus*
HnNI169	1955	Mozambique	Gorongosa NP	Museum of Vila Viçosa, Portugal	Skull	*Hippotragus niger niger*
HnNI170	1955	Mozambique	Gorongosa NP	Museum of Vila Viçosa, Portugal	Skull	*Hippotragus niger niger*
HnNI172	1913	Zambia	undetermined	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger kirkii*
HnNI173	1913	Kenya	Shimba Hills	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger roosevelti*
HnNI174	1913	Zambia	undetermined	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger kirkii*
HnNI175	undetermined	Mozambique	Sofala-Manica	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger niger*
HnNI177	1913	Kenya	Shimba Hills	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger roosevelti*
HnNI178	undetermined	Mozambique	Sofala-Manica	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger niger*
HnNI179	undetermined	Mozambique	Sofala-Manica	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger niger*
HnNI180	undetermined	Mozambique	Sofala-Manica	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger niger*
HnNI181	undetermined	Kenya	Shimba Hills	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger roosevelti*
HnNI182	undetermined	Mozambique	Sofala-Manica	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger niger*
HnNI183	1913	Zambia	undetermined	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger kirkii*
HnNI184	1914	Zambia	undetermined	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger kirkii*
HnNI185	undetermined	Mozambique	Sofala-Manica	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger niger*

HnNI186	1913	Zambia	undetermined	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger kirkii*
HnNI187	undetermined	Kenya	Shimba Hills	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger roosevelti*
HnNI188	1913	Zambia	undetermined	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger kirkii*
HnNI189	1913	Zambia	undetermined	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger kirkii*
HnNI190	undetermined	Zambia	undetermined	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger kirkii*
HnNI191	1913	Zambia	undetermined	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger kirkii*
HnNI192	1913	Zambia	undetermined	Smithsonian Institution; Washington DC, USA	Bone	*Hippotragus niger kirkii*

Abbreviations: NP – National Park;

Table 3_SI: Reaction conditions for long-range PCR amplification of *H. niger* mitochondrial genomes

Master Mix Components	**Volume (uL) per sample**
Water	14.4
PCR buffer with MgCl2	5
dNTP mix	1.25
Foward primer (10uM)	1.5
Reverse primer (10 uM)	1.5
DNA polymerase	0.35
Sample volume	1
Total reaction	25uL

Table 4_SI: Thermal-cycler conditions used for long-range PCR amplification of *H. niger* mitochondrial genomes in historic samples

Step	**Temperature**	**Time**	
1	92ºC	2 min	
2	92 ºC	10 sec	
3	60-56 ºC (for P1and P2); 66-62 ºC (P3) and 63-59 ºC (P4)	20 sec	decreasing 0.5ºC/cycle
4	68 ºC	6 min	
5	Goto step 2 **8X**		
6	92ºC	10 sec	
7	56ºC (for P1 and P2); 62ºC (P3); 59ºC (P4)	20 sec	
8	68ºC	6 min	
9	Goto step 6 **20x** (for P1,3,4) **27 x** (P2)		
10	68ºC	6 min	
11	7 ºC	forever	

P1= Primer pair Hn1377F-Hn6895R; P2= Primer pair Hn6538F-Hn12024R; P3= Primer pair Hn10583F-Hn14756R; P4= Primer pair Hn14637F-Hn3305R

II-Supplementary Information for the Results

I-Results in Methods

A-Texts

Text 11_SI: For both sequential analyses different clock models and tree priors were specified. In the first analysis we used an uncorrelated relaxed log-normal molecular clock with a Yule speciation process, as several distinct species were being considered. The ucld.stdev parameter checked with Tracer v.1.6.0 provided an indication of how clock-like our dataset was. According to the software's manual (Drummond and Rambaut, 2007; Drummond et al., 2007), if this parameter is closer to 0.0 then the data is clock-like. If it has a value greater than 1.0, then the data clearly exhibits rate heterogeneity among lineages. The same can be said for the coefficient of rate variation, which is defined as the standard deviation of the rate divided by the mean, with values close to zero implying a good fit to the strict molecular (i.e. low rate of variation across all lineages) and higher values implying among-lineage rate variation, or deviation from the strict molecular clock. As our phase I dataset showed ucld.stdev and coefficient of rate variation parameters closer to zero (0.1631 with 95% confidence interval between 0.0732 and 0.2932, and 0.1608 with 95% confidence interval between 0.0737 and 0.2866, respectively), we assumed our dataset to be quite clock-like, and performed a less complex second run using a strict clock model with constant coalescent prior. In both cases, no substantial differences were observed between the posterior estimates and fossil calibration priors. We also found little difference between the two models by comparing the marginal likelihoods of both runs (as suggested by Drummond and Rambaut, 2007), indicating that our dataset could be treated as clock-like. The posteriors distributions were also compared in the two runs. There were no discrepant posterior distributions, with all median TMRCA (time to the most recent common ancestor) estimations being approximately the same, and the respective 95% highest posterior density intervals (HPDI) almost entirely overlapped in both runs. As expected, the same results were obtained for the phase II dataset. Given that second runs did not showed significant differences from the first runs, were less complex and yielded narrower 95% HPDI and higher effective sample sizes (ESS), second run parameters were used for both phases of analyses.

B-Tables and Figures

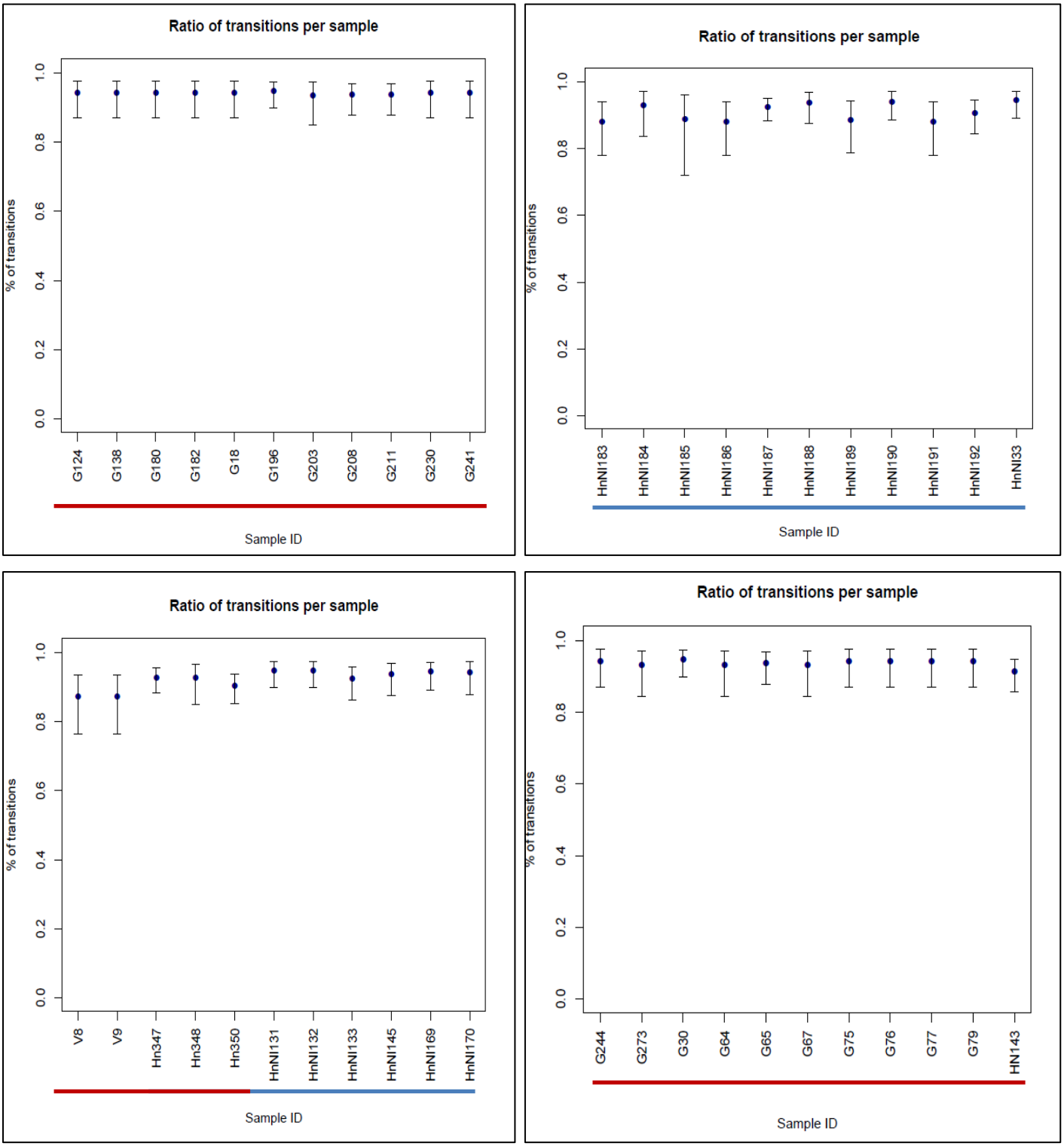

Fig 1_SI- Ratio of transitions per samples for a subset of both modern and historic samples. Historic samples are underlined in blue, while modern samples are underlined in red. The plots show that there are no significant differences between the transitions/transversions ratio of historic and modern samples. From these observations it is possible to infer that there is not a bias towards transitions in historic samples as a result of deamination.

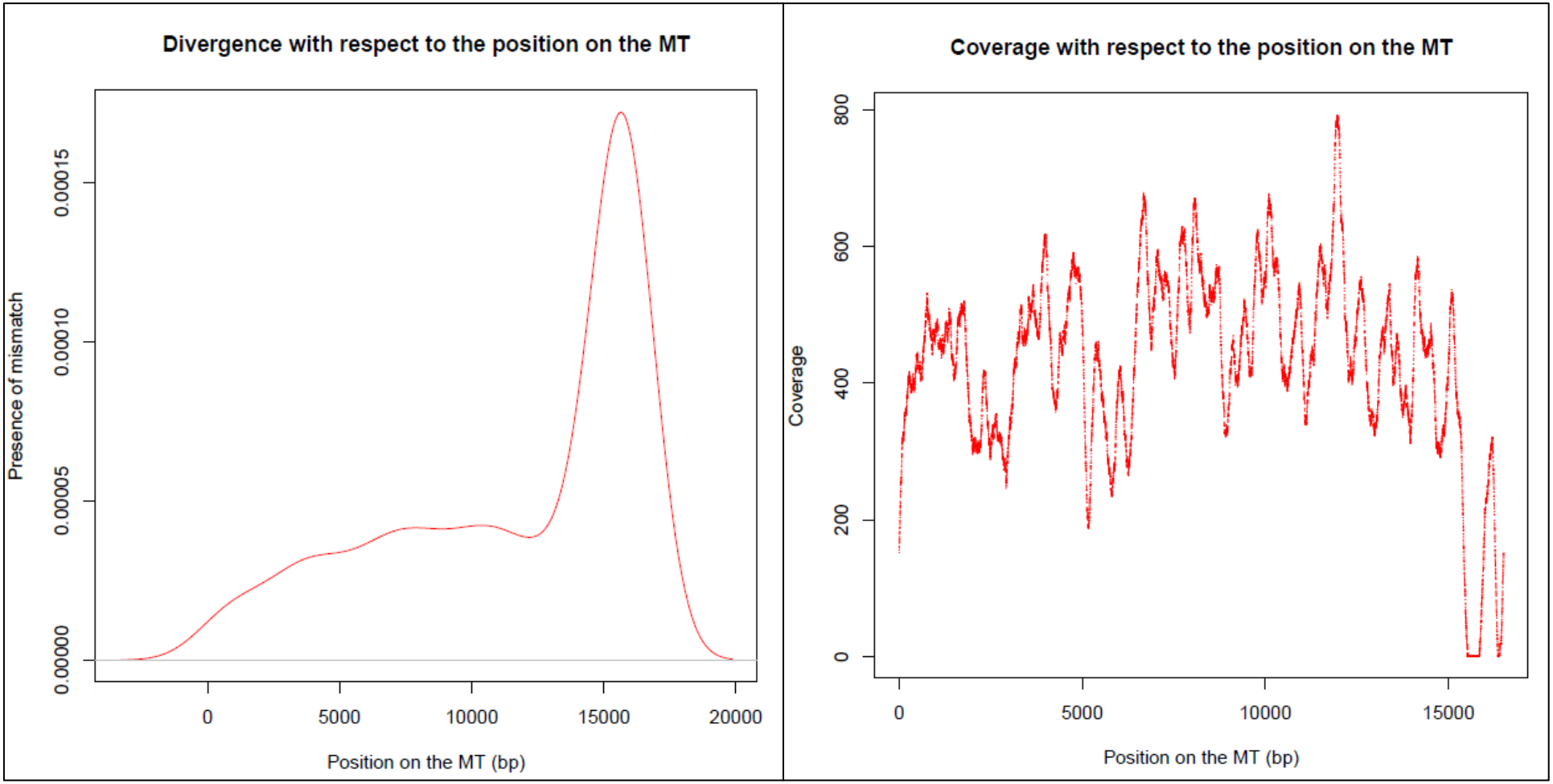

Fig 2_SI- Divergence (left) and coverage (right) with respect to the position in the mitochondrial genome, using as example a sample with an average coverage of 432.242 (HN332 from Ugalla, Tanzania). These plots illustrate a typical case in which it is evident the extreme divergence at the control region (15421-16507 bp), which is coincident with a drop in coverage. Note that, apart from the control region, all positions in the remainder sections of the mitochondrial genome have a very high coverage (from 200 to 800).

Table 5_SI: Prior distributions used for Bayesian analyses of divergence times. Posterior distributions from Bibi (2013) were used as priors in the first phase of analyses (Phase I). The priors used for the second phase of analyses (Phase II) came from Phase I posteriors on the genus *Hippotragus* and on the highly divergent lineage from Tanzania/remainder *H. niger*. Mya=million years ago.

		Priors		Posteriors			
Phase	Taxa	median (Mya)	95% HPDI	Mean	s.d.	median (Mya)	95% HPDI
I	Alcelaphinae + Hippotraginae	11.5	10.6-12.5	11.2428	0.7074	11.2	9.8-12.6
I	Alcelaphini	6.3	5.4-7.2	6.2902	0.4174	6.3	5.5- 7.1
I	*Addax* + *Oryx*	3.2	2.6-3.9	2.8352	0.1996	2.8	2.4- 3.2
I	*Alcephalus* +*Damaliscus*	4.2	3.2-5.1	4.5198	0.3138	4.5	3.9-5.1
I	Hippotraginae	6.6	5.6-7.6	6.9199	0.4523	6.9	6.0-7.8
I→II	*Hippotragus*	-	-	5.9968	0.4098	6.0	5.2-6.8
I→II	*All H. niger* + Tanzanian lineage	-	-	1.666	0.1287	1.7	1.4-1.9

II-Results for data analysis

A-Texts

Text 12_SI: The alignment with most parsimonious and higher log likelihood scored trees was the one generated by MUSCLE v3.8.31 (Edgar, 2004). Such alignment was used in all subsequent analysis. From a total of 266 full-length mitochondrial genomes, including 262 *H. niger* sequences (231 modern and 31 historic) and 4 sequences of *H. equinus* (2 modern and 2 historic), 50 were eliminated according to the previously described criteria for individuals with masked bases (N). These 50 sequences included 47 *H. niger* sequences (40 modern and 7 historic) and 3 *H. equinus* sequences (1 modern and 2 historic). Since the outgroup species was reduced to only one individual, the *H. equinus* complete mitochondrial genome, downloaded from Genbank (NC_020712.1), was added to our dataset. In the end, our final dataset included 215 *H. niger* sequences (corresponding to 191 modern and 24 historic samples) and 2 *H. equinus* sequences (one from our original dataset and the reference genome downloaded from Genbank).

Text 13_SI: A multidimensional scaling (MDS) analysis was performed to assess population relationships averaged across all sequences. The results evidenced a complete lack of geographic correlation at a population scale, which is not unexpected given that most populations share divergent haplogroups and are highly differentiated (Figure 7_SI).

B-Tables and Figures

I. Phylogenetic relationships of *H. niger*

Table 6_SI – DNA polymorphism measures calculated for the complete mtDNA dataset, excluding a total of 429 bp of the hypervariable regions, and for the nuclear intron PGK1 dataset (*H. equinus* sequences included in both datasets)

	MtDNA	PGK1
Number of sequences used	217	273
Total number of sites	16114	660
Total number of sites (excluding sites with gaps / missing data)	16084	659
Sites with alignment gaps or missing data	30 (29 gaps+ 1 N at position 4751)	1
Invariable (monomorphic) sites *	14147	652
Variable (polymorphic) sites*	1937	7
Singleton variable sites*	49	0
Parsimony informative sites*	1888	7

Note: Although the X-linked intron PGK1 was not sequenced in historic samples, the number of sequences for PGK1 is higher than for the mtDNA because each female specimen gives the information for two sequences. The * stands for parameters that do not account for insertion/deletions (gaps) and missing data (N).

Table 7_SI- Molecular diversity indexes for the complete mtDNA dataset, excluding a total of 429 bp of the hypervariable regions

Statistics	*H. niger*	*Hippotragus sp.*	*H. equinus*
Number of transitions	408	93	49
Number of transversions	30	3	1
Number of substitutions	438	96	50
Number of indels	4	2	1
Number of polymorphic sites	439	98	51
Nº of sites with transitions	407	93	49
Nº of sites with transversions	407	93	49
Number of substitution sites	435	96	50
Number of private substitution sites	190	72	35
Number of indel sites	4	2	1

Note: all parameters were estimated accounting for insertion/deletions and allowing 5% of missing data.

Table 8_SI- Genetic divergence between different *H. niger* haplogroups, in comparison to the outgroup (*H. equinus*), and the highly divergent haplogroup from Tanzania (*Hippotragus sp*.)

Comparisons	Dxy (JC)	Da (JC)	Fd.
A vs E1	0.00958 ± 0.00333	0.00956 ± 0.00333	153
A vs E2	0.00930 ± 0.00304	0.00898 ± 0.00304	131
A vs CE	0.00330 ± 0.00093	0.00233 ± 0.00094	27
A vs S1	0.00509 ± 0.00071	0.00415 ± 0.00071	51
A vs S2	0.00492 ± 0.00050	0.00417 ± 0.00050	56
E1 vs E2	0.00450 ± 0.00195	0.00420 ± 0.00196	57
E1 vs CE	0.01032 ± 0.00387	0.00937 ± 0.00387	139
E1 vs S1	0.00949 ± 0.00193	0.00856 ± 0.00193	123
E1 vs S2	0.00927 ± 0.00153	0.00854 ± 0.00153	123
E2 vs CE	0.00976 ± 0.00262	0.00851 ± 0.00264	114
E2 vs S1	0.00921 ± 0.00151	0.00798 ± 0.00151	104
E2 vs S2	0.00911 ± 0.00131	0.00808 ± 0.00132	105
CE vs S1	0.00575 ± 0.00081	0.00387 ± 0.00082	39
CE vs S2	0.00568 ± 0.00070	0.00400 ± 0.00071	45
S1 vs S2	0.00404 ± 0.00027	0.00238 ± 0.00027	20
Mean all *H. niger*	0.00729 ± 0.00253	0.00631 ± 0.00272	86
H. niger vs *Hippotragus sp.*	0.03436 ± 0.00140	0.03157 ± 0.00141	406
H. niger vs *H. equinus*	0.09134 ± 0.01550	0.08764± 0.01552	1211
Hippotragus sp. vs *H. equinus*	0.09342 ± 0.01987	0.09121 ± 0.01988	1350

Abbreviations as follows: Average number of nucleotide substitutions (Dxy) and Number of net nucleotide substitutions (Da) per site between populations, with Jukes and Cantor correction (JC); Fd. Number of fixed differences.

Table 9_SI- Molecular diversity indexes for the nuclear intron PGK1

Statistics	*H. niger*	*H. equinus*	All dataset
Number of transitions	3	0	5
Number of transversions	0	0	2
Number of substitutions	3	0	7
Number. of indels	0	0	1
Number of polymorphic sites	3	0	8
Nº of sites with transitions	3	0	5
Nº of sites with transversions	3	0	5
Number of substitution sites	3	0	7
Number of private substitution sites	0	0	2
Number of indel sites	0	0	1

Note: all parameters were estimated accounting for insertion/deletions and allowing 5% of missing data.

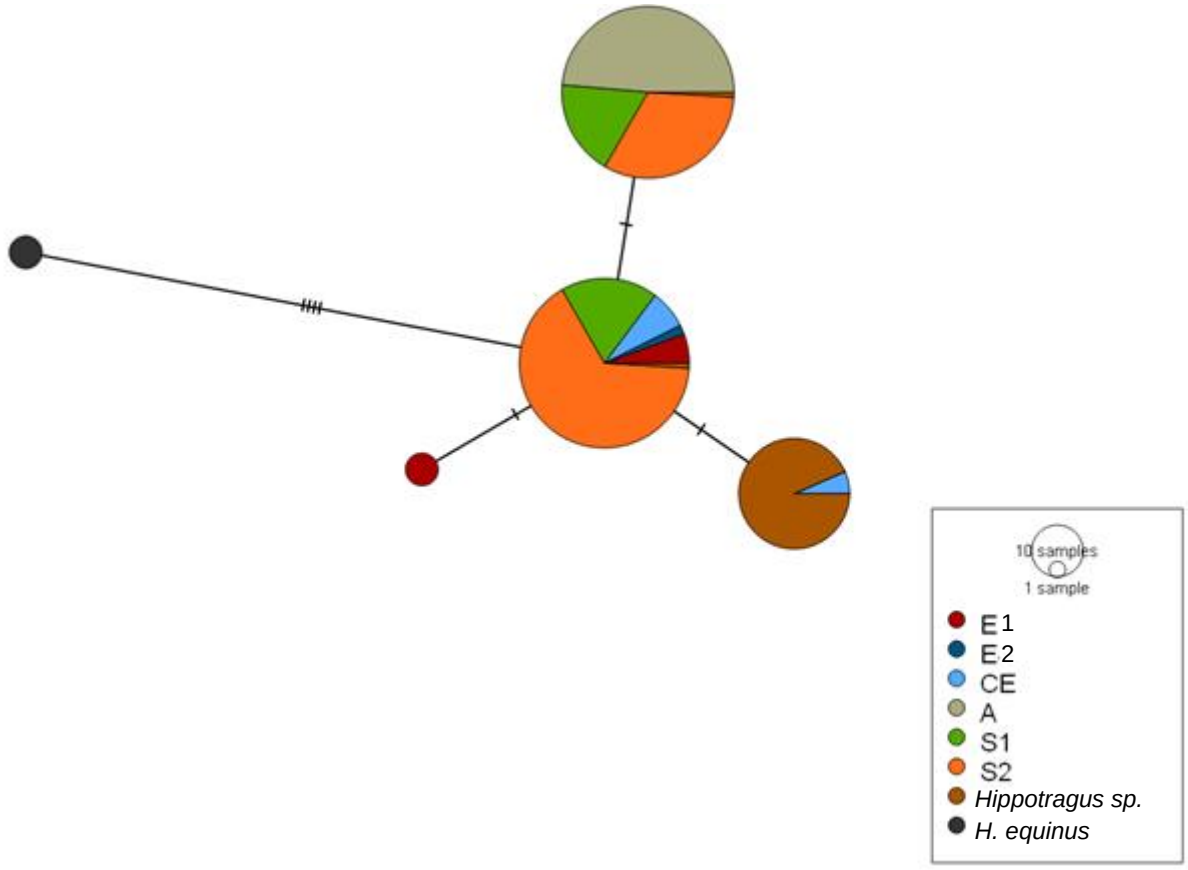

Fig 3_SI- Median-joining network of the PGK1 intron sequences including all haplotypes of *H. niger* and one haplotype belonging to the outgroup, *H. equinus*. Given that the same dataset than mtDNA was used, individuals were labeled accordingly to the respective mtDNA haplogroup. The individuals carrying the highly divergent haplogroup from Tanzania in mtDNA (*Hippotragus sp.*) share the same haplotypes with individuals belonging to the reminder *niger*-like mtDNA haplogroups, with no extreme intraspecific divergence being observed in the nuclear data.

II. Divergence Times

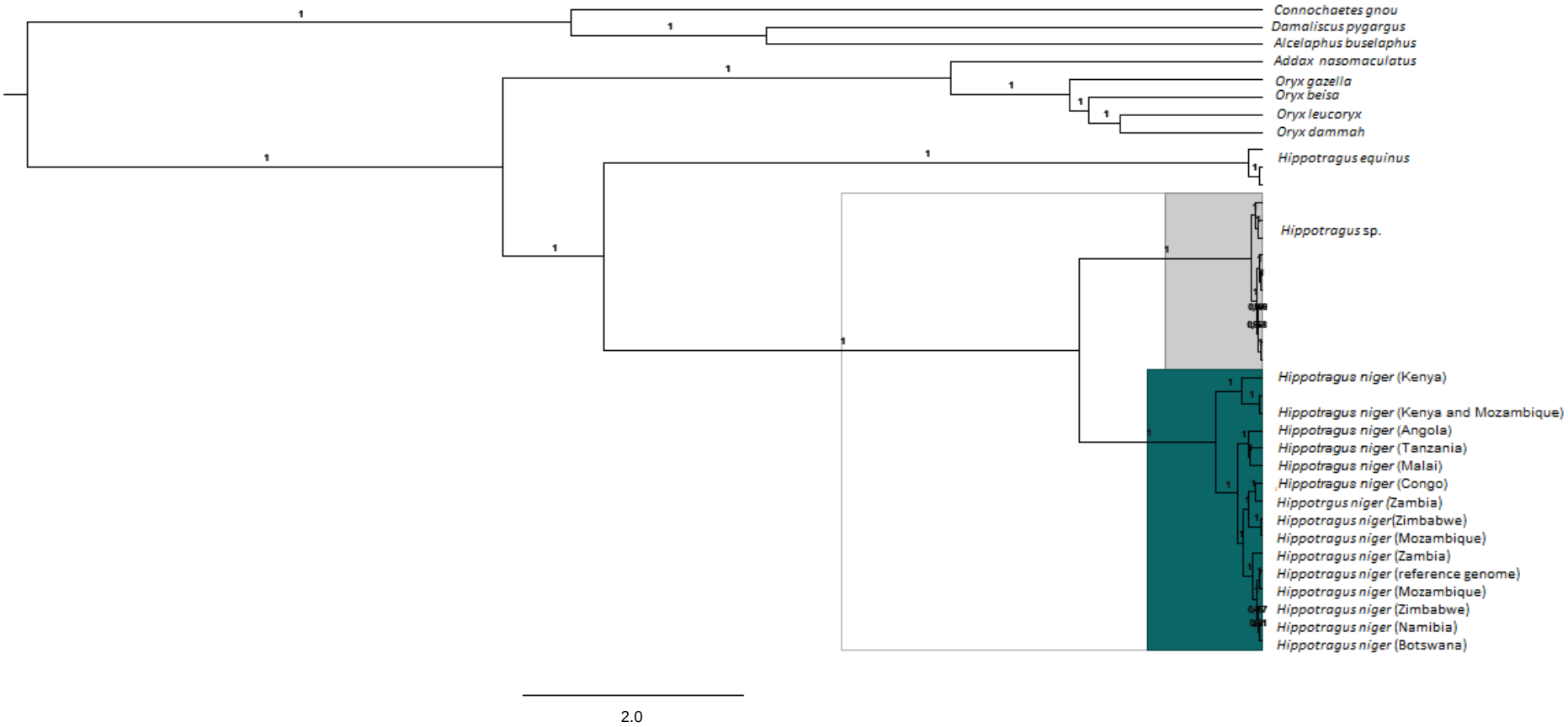

Fig. 4_SI- Bayesian phylogenetic tree of Hippotraginae phylogeny of whole mitochondrial genomes (excepting the control region) from public databases (TreeBase repository of the Bovidae dataset published by Bibi (2013)) and new sequences from this study. Posterior probabilities for nodes of interest are indicated on top of the tree branches (note these are not the same as branch nodes). Clades corresponding to *niger*-like haplogroups are highlighted in blue, while the highly divergent clade of sables from Tanzania (described by Pitra et al. (2002)) named as *Hippotragus sp.* is highlighted in grey.

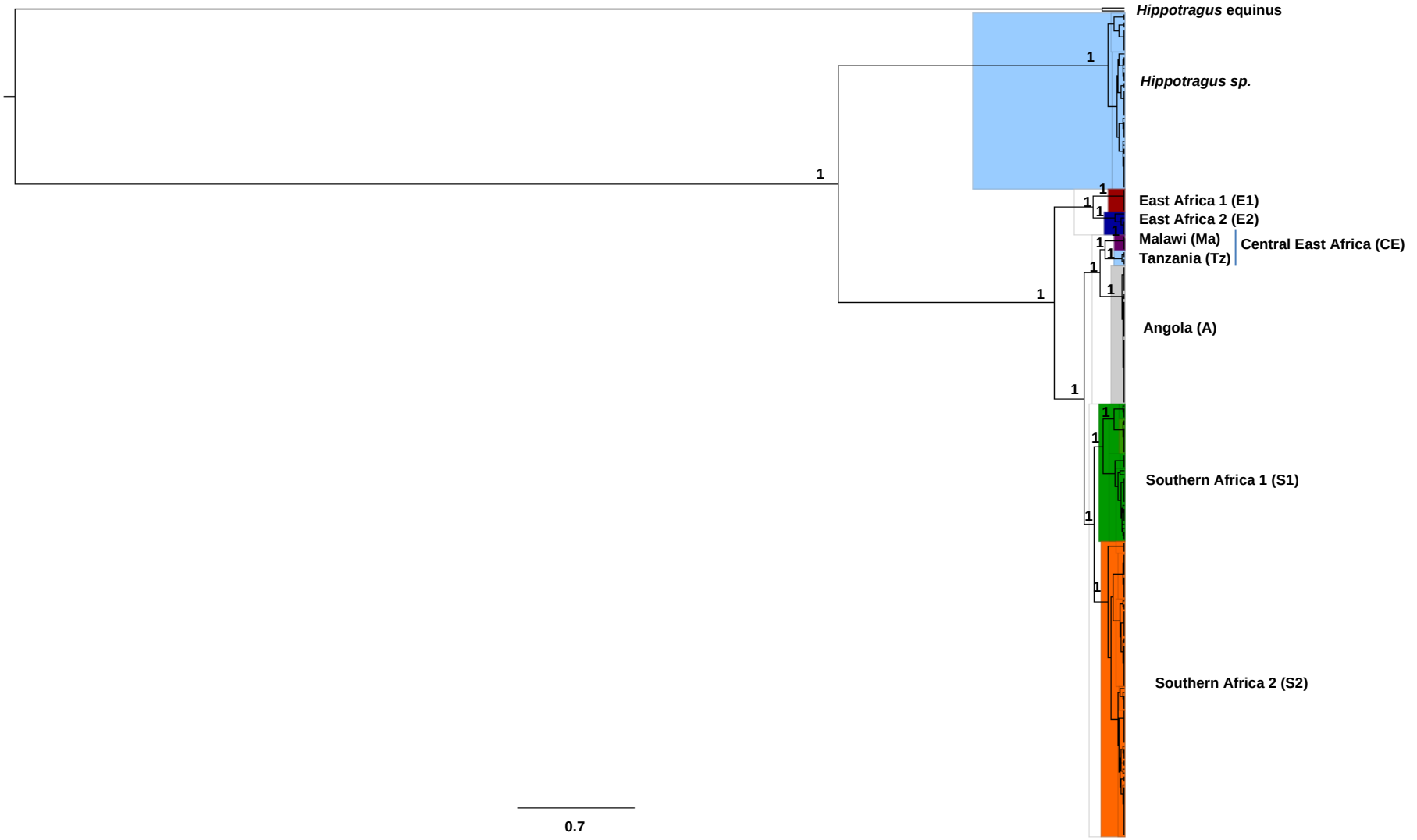

Fig. 5_SI- Bayesian phylogenetic tree of 215 *H. niger* and 2 *H. equinus* whole mitochondrial genomes (excluding 429 bp of the HVR). Clades corresponding to *niger*-like haplogroups are highlighted in different colors, followed by the respective label. Posterior probabilities are indicated for nodes of interest.

III. Phylogeographic and mtDNA population genetics analyses

Table 10_SI – DNA polymorphism measures calculated for the dataset including complete mtDNA sequences from the highly divergent lineage from Tanzania (*Hippotragus sp*).

Number of sequences	46
Total number of sites	16505
Total number of sites (excluding sites with gaps / missing data)	16502
Sites with alignment gaps or missing data	3
Invariable (monomorphic) sites*	16369
Variable (polymorphic) sites*	133
Singleton variable sites*	23
Parsimony informative sites*	110

The * stands for parameters that do not account for insertion/deletions (gaps) and missing data (N).

Table 11_SI- Molecular diversity indexes for the dataset including complete mtDNA sequences from the highly divergent lineage from Tanzania (*Hippotragus sp*.).

Statistics	
Number of transitions	128
Number of transversions	5
Number of substitutions	133
Number. of indels	3
Number of polymorphic sites	136
Nº of sites with transitions	128
Nº of sites with transversions	5
Number of substitution sites	133
Number of private substitution sites	2
Number of indel sites	3

Note: all parameters were estimated accounting for insertion/deletions and allowing 5% of missing data.

Table 12_SI- Genetic Diversity summary statistics, and respective standard deviations, for the highly divergent complete mtDNA sequences from Tanzania

Population	N	S	h	Hd	π	MPD
***Hippotragus sp.* from Tanzania**	**46**	**136**	**29**	**0.966 ± 0.013**	**0.00166 ± 0.00082**	**27.370 ± 12.207**
Kigosi	15	69	6	0.800 ± 0.071	0.00161 ± 0.00084	26.571 ± 12.348
Kizigo	6	67	5	0.933 ± 0.122	0.00167 ± 0.00099	27.600 ± 14.146
Mlele	1	0	1	1.000 ± 0.000	0.00000 ± 0.00000	0.000 ± 0.000
Niensi	3	1	2	0.667 ± 0.314	0.00004 ± 0.00005	0.667 ± 0.667
Rungwa	6	33	6	1.000 ± 0.096	0.00085 ± 0.00051	14.000 ± 7.342
Ugalla	12	86	10	0.970 ± 0.044	0.00139 ± 0.00074	22.924 ± 10.867
Wembere	3	43	3	1.000 ± 0.272	0.00174 ± 0.00132	28.667 ± 17.486

Abbreviations as follows: N, number of samples; S, number of polymorphic (segregating) sites; h, number of haplotypes; Hd, haplotype diversity; π, nucleotide diversity; MPD, mean pairwise distance. Note: all parameters were estimated accounting for insertion/deletions and allowing 5% of missing data.

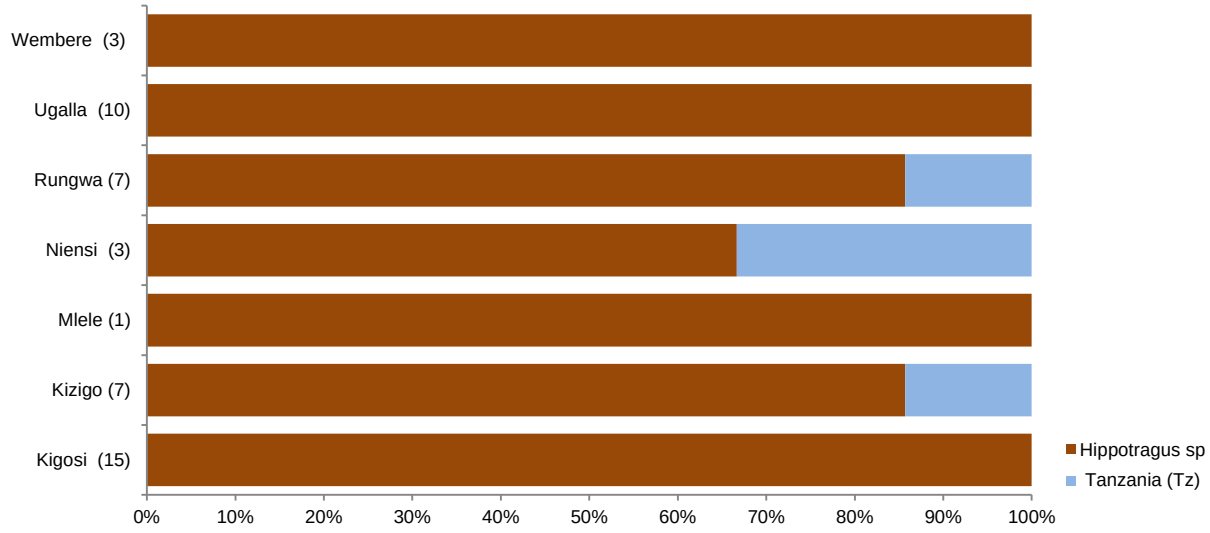

Fig. 6_SI- Haplotype Frequency for *Hippotragus sp*. and Haplogroup Tz in the seven sampled populations from Tanzania.

Table 13_SI- DNA polymorphism measures for *H. niger* complete mitochondrial sequences	
Number of sequences used	169
Number of Sites	16514
Total number of sites (excluding sites with gaps / missing data)	16501
Sites with alignment gaps or missing data	13 (12 gaps+1 N at position 4745)
Invariable (monomorphic) sites*	15978
Variable (polymorphic) sites*	523
Singleton variable sites*	17
Parsimony informative sites*	506

The * stands for parameters that do not account for insertion/deletions (gaps) and missing data (N).

Table 14_SI- Molecular diversity indexes for complete mtDNA sequences of *H. niger*

Statistics	A	E1	E2	CE	S1	S2	all dataset
Nº of transitions	1	0	23	74	120	135	497
Nº of transversions	1	0	0	6	2	8	31
Nº of substitutions	2	0	23	80	122	143	528
Nº of indels	0	1	2	3	2	4	11
Nº of polymorphic sites	2	1	25	83	124	147	535
Nº of sites with transitions	1	0	23	74	120	135	496
Nº of sites with transversions	1	0	23	74	120	135	496
Nº of substitution sites	2	0	23	80	122	143	524
Nº of private substitution sites	0	0	0	46	0	0	496
Nº of indel sites	0	1	2	3	2	4	11

Abbreviations are as follows: ts =transitions; tv =transversions; subst. =substitutions;

Table 15_SI- Genetic diversity summary statistics, and respective standard deviations, for the different populations sampled, based on complete mtDNA genome sequences of *H. niger*

Populations	N	S	h	Hd	π	MPD
All	159	532	49	0.945 ± 0.009	0.00551 ± 0.00264	91.166 ± 39.396
Luando	27	0	1	0.000 ± 0.000	0.00000 ± 0.00000	0.000 ± 0.000
Cangandala	9	1	2	0.389 ± 0.164	0.00002 ± 0.00003	0.389 ± 0.399
Shimba Hills	10	100	5	0.756 ± 0.129	0.00323 ± 0.00173	53.267 ± 25.217
Tanzanian populations	4	11	3	0.833 ± 0.222	0.00043 ± 0.00031	7.167 ± 4.258
Liwonde NP	4	1	2	0.500 ± 0.265	0.00003 ± 0.00004	0.500 ± 0.519
Niassa	2	14	2	1.000 ± 0.500	0.00085 ± 0.00088	14.000 ± 10.247
Sofala-Manica	5	16	2	0.400 ± 0.237	0.00039 ± 0.00026	6.400 ± 3.651
Gorongosa NP	2	1	2	1.000 ± 0.500	0.00006 ± 0.00009	1.000 ± 1.000
Katanga	4	42	4	1.000 ± 0.177	0.00154 ± 0.00103	25.500 ± 14.289
Mufumbwe	4	88	3	0.833 ± 0.222	0.00267 ± 0.00177	44.167 ± 24.492
Lusaka-Kafue	9	21	3	0.639 ± 0.125	0.000592 ± 0.00034	9.778 ± 4.948
Mulobezi	6	118	4	0.867 ± 0.129	0.00367 ± 0.00214	60.600 ± 30.648
Kafue	16	19	4	0.350 ± 0.148	0.00021 ± 0.00012	3.425 ± 1.847
Masebe-Mkushi	12	4	2	0.545 ± 0.061	0.00013 ± 0.00009	2.1818 ± 1.297
Matesi	6	81	3	0.600 ± 0.215	0.00260 ± 0.00153	43.000 ± 21.847
Triangle	18	118	8	0.830 ± 0.064	0.00296 ± 0.00151	48.922 ± 22.228
Mahango	17	24	3	0.3235 ± 0.136	0.00027 ± 0.00015	4.412 ± 2.290
Chobe	4	20	3	0.8333 ± 0.222	0.00065 ± 0.00045	10.667 ± 6.176

Abbreviations as follows: N, number of samples; S, number of polymorphic (segregating) sites; h, number of haplotypes; Hd, Haplotype diversity; π, nucleotide diversity; MPD, mean pairwise distance. All parameters were estimated accounting for insertion/deletions and allowing 5% of missing data. Individuals from undetermined populations (10 museum specimens from Zambia) are not represented. Given the lack of representativeness in each population within Tanzania, and that the haplotypes within these populations are all Tanzanian specific, these were clustered together as one population.

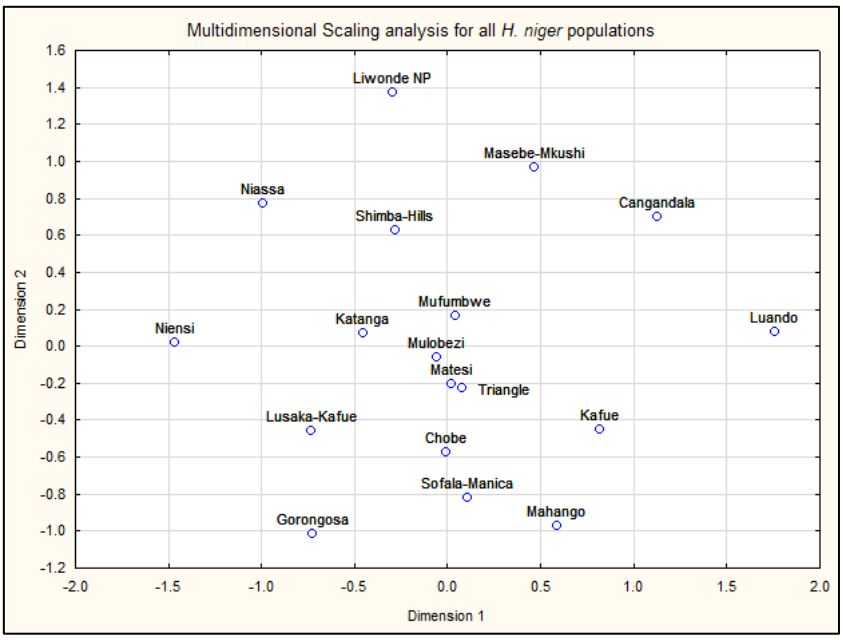

Fig.7_SI- Multidimensional Scaling analysis for all *H. niger* populations. Only populations with more than 2 individuals were considered. Stress value was 0.2473, which is considered fit according to the multidimensional scaling stress evaluation table proposed by Sturrock and Rocha (2000).

Table 16_SI- Neutrality tests results for each haplogroup of *H. niger*. All tests were non-significant in all haplogroups (p-value > 0.05).

Statistics	**A**	**E1**	**E2**	**CE**	**S1**	**S2**
Tajima's D						
Tajima's D	0.04572	0	-0.25337	2.24927	1.11867	0.48285
p-value	0.644	1	0.456	0.9996	0.885	0.758
Fu's FS test						
FS	0.15259	-0.00275	0.9038	6.98029	5.69037	12.64805
p-value	0.444	0.264	0.617	0.9949	0.959	0.995
Ewens-Watterson test						
Observed F value	0.6034	0.72222	0.22222	0.25	0.1034	0.09766
Expected F value	0.60459	0.63667	0.22222	0.25678	0.08395	0.09327
Watterson F p-value	0.555	1	1	0.7838	0.908	0.68
Slatkin's exact p-value	0.409	1	1	0.7838	0.95	0.786

Table 17_SI- AMOVA results for a hierarchical arrangement in two groups: populations located easterly to the Rift Valley and populations located westerly to the eastern branch of the Rift Valley

Type of variation	Among groups	Among populations within groups	Within populations
d.f.	1	18	139
% of variation	58.67*	38.71*	2.62*
ϕ-Statistics	ϕ_{CT}	ϕ_{SC}	ϕ_{ST}
Fixation indices	0.56202**	0.79390*	0.90973*

All values represented are significant: *p-value <0.0001; **p-value<0.01

III-Supplementary Information for the Discussion

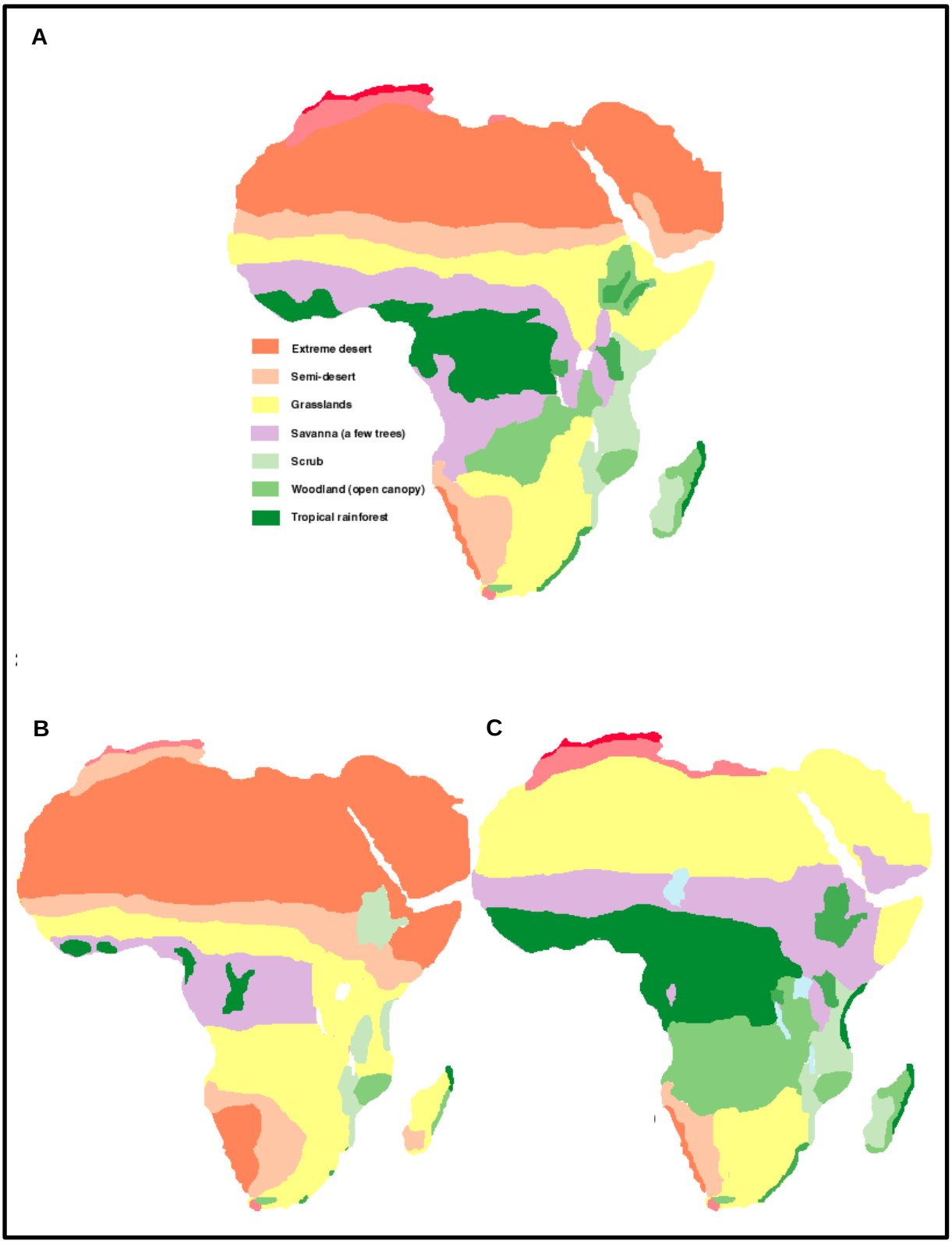

Fig.8_SI- Potential distribution of the major vegetation zones of Africa in the present day (A), during the maximum of the penultimate glaciation (thought to be characterized by the same climatic conditions as the Last Glacial Maximum) (B) and during the moistest phase of the Eemian Interglacial (thought to be characterized by the same climatic conditions of the early Holocene, around 8,000 14C years ago) (C). Maps from Adams (1997) accessible at http://www.esd.ornl.gov/ern/qen/nerc.html.

References

Bibi, F. (2013). A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. BMC Evol. Biol. *13*, 166.

Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. *7*, 214.

Drummond, A., Ho, S., Rawlence, N., and Rambaut, A. (2007). A rough guide to BEAST 1.4. 1–41.

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

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. a, Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. U. S. A. *110*, 2223–2227.

Hoelzel, a R., Hancock, J.M., and Dover, G. a (1991). Evolution of the cetacean mitochondrial D-loop region. Mol. Biol. Evol. *8*, 475–493.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448.

Pitra, C., Hansen, A.J., Lieckfeldt, D., and Arctander, P. (2002). An exceptional case of historical outbreeding in African sable antelope populations. Mol. Ecol. *11*, 1197–1208.

Shields, G.F., and Kocher, T.D. (1991). Phylogenetic relationships of north American ursids based on analysis of mitochondrial DNA. Evolution (N. Y). *45*, 218–221.

Sturrock, K., and Rocha, J. (2000). A Multidimensional Scaling Stress Evaluation Table. Field Methods *12*, 49–60.