Population history and taxonomy of African hares (genus *Lepus*) inferred from genetic variation

Sara Ribeiro Barbosa Almendra Lado
Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia
CIBIO-InBIO
2015

Orientador
José Melo-Ferreira, Auxiliary Researcher, FCUP/CIBIO-InBIO

Co-orientadores
Paulo Célio Alves, Associate Professor, FCUP/CIBIO-InBIO
José Carlos Brito, Principal Researcher, FCUP/CIBIO-InBIO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Agradecimentos

"If I have seen further than others, it is by standing upon the shoulders of giants." Isaac Newton

Não poderia fechar mais um capítulo da minha vida académica sem agradecer a pessoas muito especiais pela sua ajuda incondicional.

Antes de mais agradeço ao meu orientador, José Melo-Ferreira, pelo exemplo, pela amizade e apoio que presta todos os dias aos seus alunos. Com o equilíbrio de um nível de exigência elevado, mas balançado de muita compreensão, melhor orientador seria impossível. Sem dúvida alguma, é uma pessoa que admiro bastante.

Ao meu co-orientador, Professor Paulo Célio Alves, com a boa disposição que sempre demonstra, agradeço por todas as oportunidades que me ofereceu e a confiança que apostou em mim. E claro, por todos os conhecimentos que me transmitiu e pelo apoio ao longo dos últimos anos.

Ao José Carlos Brito, meu co-orientador, pelo excelente profissional que é e pela amizade e prontidão em querer sempre ajudar os seus (muitos) alunos. E pelas amostras de campo que recolheu ao longo das diversas viagens por África, que foram essenciais para a realização deste trabalho.

Agradeço ainda aos membros do grupo das lebres (Mafalda, Fernando, Liliana e João) pelo apoio que nunca me negaram.

I thank both Dr. Frank Zachos (Natural History Museum, Berlin) and Dr. Renate Angermann (Natural History Museum, Vienna) for providing samples, and especially to Renate for all the help and kindness. I want also to thank to Dr. Zafar (National Wildlife Research Center, Saudi Arabia) for providing Saudi Arabian samples.

Como não poderia deixar de ser, agradeço aos membros do ano 2013/2014 do MBGE, pela amizade e companheirismo. E ainda a todos os meus amigos que estão sempre a meu lado.

Aos outros dois membros do trio maravilha, Filipa e Fábio, não é preciso dizer muito. São mais que amigos, são dois irmãos para mim. Um "muito obrigada" gigante por estes cinco anos maravilhosos que, sem vocês, não seria a mesma coisa (e pelos muitos anos que ainda estão para vir). Ao Miguel, agradeço por ser quem é.

Por fim, dedico este trabalho aos seis elementos da minha família, que são melhor de mim. Família essa que, apesar de pequena, é um exemplo gigante e um imenso orgulho. Sempre me ensinaram que o caminho mais fácil nunca foi o melhor, e que sem esforço nada se conquista. Tal como eu dou tudo por eles, eles dão tudo por mim. Dedico sobretudo este trabalho e este final de mais uma etapa, em memória do meu melhor amigo, o meu Avô.

Financial support

This work was financially supported by Portuguese national funds through the FCT (Fundação para a Ciência e a Tecnologia) and by FEDER funds (Fundo Europeu de Desenvolvimento Regional) through the COMPETE program (FCT-ANR/BIA-EVF/0250/2012 research projects), by Programa Operacional Potencial Humano-Quadro de Referência Estratégico Nacional (POPH-QREN) funds from the European Social Fund and Portuguese Ministério da Educação e Ciência (FCT, IF/00033/2014 FCT investigator grant to my supervisor José Melo-Ferreira). Museum samples were collected under the financial support provided by SYNTHESYS: DE-TAF-4131 and BE-TAF-5010.

Sumário

Os eventos climáticos do passado influenciaram o padrão de distribuição actual da biodiversidade Africana. Por exemplo, as glaciações durante o Pleistoceno e os ciclos de avanços e recuos das principais zonas de vegetação tiveram efeitos profundos na distribuição e evolução dos mamíferos neste continente. Estes eventos conduziram à presença de áreas biogeográficas e climáticas únicas com uma riqueza biológica assinalável, tornando África um continente de particular interesse para estudos de biodiversidade. No entanto, ainda existe um grande défice de conhecimento sobre a diversidade de espécies em África e mesmo uma robusta taxonomia dos organismos já descritos. Além disso, há regiões particulares, como as áridas, que são normalmente consideradas pobres em espécies devido às difíceis condições para a sustentação de vida. No entanto, os desertos podem ser locais muito ricos em espécies endémicas (com adaptações únicas em ambientes extremos) devido aos eventos climáticos do passado. Igualmente, a intersecção biogeográfica entre regiões a Palearcticas e Afrotropicais na bacia do Mediterrâneo é considerada um "hotspot" de biodiversidade no Norte de África. Contudo, a maioria dos estudos moleculares existentes sobre a biodiversidade em África envolvem vertebrados com baixa capacidade de dispersão, e uma boa compreensão da diversidade de organismos com maior capacidade de dispersão, como as lebres (género *Lepus*), está longe de ser alcançada.

Das 32 espécies do género *Lepus,* 6 estão descritas em África. Os estudos de taxonomia e sistemática que se têm focado neste grupo de espécies em África são baseados maioritariamente em evidências morfológicas e/ou paleontológicas. Apesar de úteis, a utilização isolada desta informação é limitada, particularmente neste género em que há uma sobreposição de características entre as espécies. A identificação e classificação das diferentes espécies está deste modo aquém de ser bem estabelecida, havendo claras divergências sobre o número de espécies ou subespécies distintas presentes no continente Áfricano, bem como as suas respectivas distribuições. Para clarificar a história populacional das lebres africanas e, por conseguinte, contribuir para uma classificação taxonómica adequada destes organismos, a identificação das entidades evolutivas distintas bem como as suas relações filogenéticas é fundamental. Neste sentido, o uso de ferramentas moleculares é fundamental.

Entre as lebres africanas, a Lebre do Cabo (*Lepus capensis)* é uma das espécies mais controversas. É uma das espécies de lebre com maior distribuição descrita, ocorrendo desde África (Norte a Sul) até à China, e compreendendo

populações com características diferentes ao longo das diferentes regiões. Esta heterogeneidade levou a que tenham sido identificadas 80 subespécies. Além disso, a sua distribuição sobrepõe-se à distribuição de outras espécies de lebre, principalmente de *L. saxatilis* e *L. microtis*, em África. Actualmente, o nível de diferenciação e divergência das populações de *L. capensis* ao longo da sua ampla distribuição são desconhecidos. Com o objectivo de compreender melhor a história evolutiva das lebres de África e adquirir conhecimentos para o esclarecimento da sua taxonomia e sistemática, neste trabalho investigou-se a estrutura populacional e os padrões filogeográficos da lebre do cabo, inferindo a história demográfica das populações. A diversidade genética foi ainda analisada no contexto das espécies com distribuições vizinhas. Para isso, foi delineado o esquema de amostragem mais amplo alguma vez realizado para *Lepus capensis*, incluindo tanto amostras recolhidas em museus de história natural como directamente do campo, compreendendo diversas localidades em África e com particular ênfase na região Noroeste. A caracterização molecular foi realizada com 18 microssatélites recém-desenvolvidos e sequências de ADN de um gene mitocondrial (mtDNA) e cinco genes nucleares.

Os resultados da análise da diversidade genética revelaram uma forte estruturação biogeográfica, inferidos usando métodos Bayesianos a partir dos dados de microssatélites, e divergências filogenéticas profundas com base tanto em sequências de mtDNA como de loci nucleares. Cinco grupos geográficos principais foram identificados: dois no Próximo Oriente, Quénia, África do Sul e Noroeste Africano. Estes níveis de divergência intra-específica inferidos sugerem que *L. capensis* é uma entidade evolutiva heterogénea que possivelmente representa várias espécies crípticas. Várias linhagens divergentes que se inferiu terem-se separado no último milhão de anos foram encontradas geograficamente bem estruturadas, o que deverá ter resultado de fragmentação das distribuições ancestrais e divergência em alopatria. Estes resultados permitiram descrever possíveis barreiras ao fluxo genético, bem como regiões geográficas onde os haplótipos mitocondriais de diferentes clados se encontram em simpatria possivelmente após contacto secundário. A análise detalhada da variação mitocondrial sugeriu fragmentação genética adicional na parte Noroeste de África, ao contrário do resultado genético homogéneo sugerido pela análise dos microssatélites. Este padrão pode resultar de filopatria das fêmeas e dispersão mediada pelos machos. Algumas linhagens de DNA foram ainda encontradas filogeneticamente mais próximas de outras espécies de lebres. A hipótese de introgressão mitocondrial, fenómeno comum em lebres, na população da parte Próximo Oriente foi testada por meio de simulações de coalescência, e revelou que de facto introgressão mitocondrial ocorreu a partir da espécie vizinha *Lepus*

europaeus. Assim, as considerações taxonómicas e a reconstrução da história evolutiva neste trabalho envolvem também outras espécies vizinhas de *L. capensis*, incluindo espécies de lebre Africanas (*L. saxatilis*, *L. microtis*) e não africanas (*L. timidus*, *L. europaeus*). Tendo em conta a multidimensionalidade do conceito de espécie, a completa compreensão do nível de isolamento entre as diversas entidades evolutivas identificadas neste estudo e alterações à taxonomia em vigor devem, no futuro, usar uma caracterização detalhada não só da variação genética, mas também da morfologia, biologia e ecologia. Esta situação é particularmente relevante para a população da Arábia, que se sabe morfologicamente distinta e se mostrou aqui fortemente divergente geneticamente das restantes populações de *Lepus capensis*.

As ferramentas moleculares aplicadas neste estudo permitiram avanços significativos na compreensão da história evolutiva, filogeografia, taxonomia e sistemática das Lebres do Cabo em África bem como das espécies de lebre vizinhas. Os resultados mostraram o papel dinâmico do deserto do Saara na diversificação das lebres da região noroeste de África, bem como a importância dos sistemas aquáticos e de montanhas que definem os maiores ramos evolutivos dentro da espécie.

Palavras-chave: África, Saara, lebres, *Lepus capensis*, estruturação genética, taxonomia, sistemática, filogenética, filogeografia, demografia, fluxo de genes, sequências de ADN, microssatélites.

Abstract

The combination of past climatic and stochastic dispersal events resulted in the contemporary pattern of species diversity and distribution in Africa. The Pleistocene glaciations and the cyclic contractions and expansions of major vegetation zones had deep effects on the distribution and evolution of African mammals. These events dictated the presence of unique biogeographic and climatic areas, making Africa a very special area for biodiversity studies. Despite Africa being generally rich in biodiversity, there is still information deficiency in the organisms that are present, and the diversity and taxonomy of the already described organisms. Additionally, arid regions are normally assumed as plain areas with difficult conditions to sustain life, and therefore with less biodiversity when compared to other climatic regions. However, deserts can be superb endemic species repositories due to the past climatic events, and the Saharan desert is one example. Moreover, the biogeographical intersection between the Palearctic and Afrotropical regions in the Mediterranean basin is considered a hotspot of biodiversity in North Africa. But most of the existing molecular studies on African biodiversity involve vertebrates with low dispersal activity while relevant information on organisms with relative dispersal ability, like hares (genus *Lepus*), is far from being achieved.

From the 32 *Lepus* species, 6 are described in Africa. However, the existing studies on this complex genus in Africa were mainly based on evidence of morphology and/or palaeontology, which are limited sources of information, due to the large overlap of characteristics between species. This species classification is far from being well established, and specialists commonly diverge on the number of distinct hares present in Africa, species or subspecies, and their distributions. To clarify the population history of these African hares and consequently contribute to a proper taxonomical classification of these organisms, the identification of the distinct evolutionary entities and their phylogenetic relationships is fundamental. The use of molecular tools therefore appears promising to this end.

Among the African hares, the cape hare (*Lepus capensis*) is one of the most controversial species. This is one of the most widespread hare species, occurring from Africa to China, with distinct features in different regions that lead to the identification of 80 subspecies. Also, its range distribution is overlapped with other hare species, mostly with *L. microtis* and *L. saxatilis* in Africa. However, the level of differentiation and divergence within the cape hare as it is currently classified along its broad distribution is unknown. In order to better understand the evolutionary history of African hares and provide valuable information to clarify its taxonomy and systematics, the population

structure and phylogeographic patterns of the species was investigated, inferring the demographic history of the populations in light of the large-scale climatic fluctuations of the Pleistocene. This genetic diversity was in addition analyzed in the context of the putative neighboring species. For this, the broadest sampling scheme ever performed in *L. capensis* was designed, including both museum and field collected samples and comprising several populations across Africa, with a particular focus on Northwest Africa, and sequences of other 4 *Lepus* species. The molecular characterization was performed with 18 newly developed microsatellites and sequences of 1 mitochondrial DNA and 5 nuclear DNA loci.

The results revealed strong biogeographic structuring at the genetic level, inferred from the microsatellite data using Bayesian clustering methods, and deep divergences based on both on mitochondrial DNA alone and on the set of sequences from 5 nuclear DNA loci, based on phylogenetic analyses. Five major geographically explicit groups were identified: Near East, Arabia, Kenya, South Africa and Norwest Africa. These inferred high levels of intraspecific divergence suggest that *L. capensis* is a heterogeneous evolutionary entity that possibly represents several cryptic species. Several divergent lineages which probably diverged in the last million years were found to be geographically structured, which likely resulted from fragmentation of ancestral ranges and divergence in allopatry. Possible barriers to gene flow were also described. Regions where mtDNA haplotypes from different clades were found in sympatry probably result from secondary contact of the divergent lineages after expansion. Whether these evolutionary entities are reproductively isolated and should be considered distinct species should be assessed in the future using a detailed characterization of their genomes, morphology, biology and ecology. This is particularly relevant for the Arabian population, which appears strongly divergent and is also morphologically distinct. Furthermore, mtDNA variation suggested additional genetic fragmentation in the species' range in Northwestern Africa, contrary to the homogenous gene pool suggested by the microsatellite analysis. This may result from female philopatry and male-mediated dispersal. Interestingly, some mtDNA lineages were found to be phylogenetically more closely related to other species. The hypothesis of mtDNA introgression, a common event in hare species, into the Near East population was tested using coalescent simulations, and revealed that mtDNA flow occurred from neighboring *Lepus europaeus*. Thus, taxonomical considerations and the reconstruction of the evolutionary history in this work also involve other neighboring hare species of *L. capensis*, both African (*L. saxatilis*, *L. microtis)* and non-African species (*L. timidus*, *L. europaeus).*

The molecular tools applied in this study shed light onto the species boundaries, evolutionary history, phylogeography, taxonomy and systematics of cape hares in Africa as well as of the neighboring hare species. The results contributed to the knowledge on the role of the landscape dynamics of the Sahara desert in the diversification of hares of the region, and how the mountain and water systems together with changes in the climate mediated the diversification. The evidences obtained open the door for more robust and thorough studies aiming at a deeper understanding of North African hare species and that seek to determine whether *L. capensis* should be decomposed into several taxonomic units.

Keywords: Africa, Sahara, hare species, *Lepus capensis*, genetic structure, taxonomy, systematics, phylogenetics, phylogeography, demography, gene flow, DNA sequences, microsatellites.

Index

List of figures

List of tables

List of abbreviations

BI - Bayesian inference

BPP - Bayesian posterior probabilities

BSP - Bayesian skyline plot

EBSP - Extended Bayesian skyline plot

ESS - Effective sample size

F_{ST} - Pairwise F-statistics

Fw – Primer forward

HWE - Hardy-Weinberg Equilibrium

IM - Isolation-with-migration model

IUPAC - The International Union of Pure and Applied Chemistry

ML - Maximum likelihood

mtDNA – Mitochondrial DNA

nucDNA – Nuclear DNA

NW Africa – Norwest Africa

Rv – Primer reverse

S. Africa – South Africa

scnDNA - Single-copy nuclear DNA

s. l. – sensu lato

TBE - Tris-Borate-EDTA Buffer

TE - Tris-EDTA

ya- Years ago

1. Introduction

Speciation is a natural process that contributes partly to a labeled view of nature, which is useful for its characterization and understanding. This concept represents an important process which leads to the divergence among entities at several levels, from genetics to morphology, providing information that can feed criteria for species classification and delimitation. The recognition of the different species and the characterization of its intraspecific diversity is not only a valuable tool to organize conservation efforts, but also an important starting point to understand relevant evolutionary processes, such as adaptation, the process of speciation itself or hybridization.

1.1 The importance of using molecular tools in taxonomy

Taxonomy describes, names and classifies living beings, so that it is possible to organize and understand biodiversity. Such task, although sometimes challenging, is important for many reasons, such as fundamental decision-making in conservation issues. The classical taxonomy makes use of morphology, geography, ecology and behavioral information to characterize the different entities (DeSalle *et al.* 2005). However, the number of described species rapidly increased for many groups of organisms, not only due to exploration of poorly known areas, but most importantly due to the complementation of “pure taxonomy” with molecular genetics tools and technological advances, creating a new concept called “integrative taxonomy” (Padial *et al.* 2010). Nevertheless, this new version of taxonomy continues to carry challenges in proving species hypotheses since it requires a careful and time-consuming labor on species delimitation and all the features and patterns that characterize these new species.

After more than 250 years of prevalence of comparative morphology in species discovery, new methods, mainly molecular, have become a successful tool for taxonomy (Padial *et al.* 2010). This happens since genetic approaches can provide a dimension of information that is not easy to obtain using only the sources of information typical of classical taxonomy. While in some cases the use of molecular tools corroborates the taxonomical classification obtained with morphological characters (e.g. Cardoso & Vogler 2005; Gompert *et al.* 2006), and can be congruent with population structure from dispersal behavior inference (Coulon *et al.* 2008), the applications of these tools extend beyond simple taxonomical classification. For instance, it is possible to describe new species (e.g. Glaw *et al.* 2010), infer the

presence of gene flow (Melo-Ferreira *et al.* 2012), estimate timing of divergence between species and populations (e.g. Jacobsen *et al.* 2012) or understand geographical patterns of diversity in taxonomically poorly studied species (e.g. Smith *et al.* 2005; Igea *et al.* 2013). Other important contribution of molecular tools is the discovery of cryptic species complexes. By definition, cryptic species are "two or more distinct species classified as a single one" which leads to taxonomic challenges in a way that it is necessary to detect different morphologically similar species (Bickford *et al.* 2006). For some of these cases, producing the correct classification has an immediate impact by aggravating the conservation status of endangered species (Bowen *et al.* 1993; Ravaoarimanana *et al.* 2004). In other cases, it leads to the split of previously considered wide-ranging species, changing the biodiversity patterns (e.g. Frynta *et al.* 2010; Wagner *et al.* 2011; Boratynski *et al.* 2012). Molecular markers have been key to identify different cases of cryptic diversity (e.g. Mulvaney *et al.* 2005; Perera & Harris 2010; Roux *et al.* 2013; Vod *et al.* 2015), allowing to identify integrative biogeographic patterns in biodiversity distribution. The number of phylogeographic studies suggesting new entities as cryptic species is increasing due to the fact that although there is occasionally underestimation of species diversity based on morphological characters, the use of genetics is exponentially growing. On the contrary, molecular approaches can also evaluate if morphologically distinguished species are only a reflection of adaptation to local conditions, belonging in fact to the same species. This highlights the need for studies that help characterizing the planet's biodiversity and promote its conservation. Nevertheless, combining with broad sampling of taxa, molecular approaches allow to unravel micro-hotspots of biodiversity, and to better support biodiversity conservation planning (see e.g. Brito *et al.* 2014).

Genetic tools contribute also to improve systematics. Systematics and taxonomy, despite being separate concepts, are often confounded and so it is important to clarify their meanings. While taxonomy is the classification and naming of individuals, systematics may be defined as the study of the diversity of organisms and the relationships between them, in an evolutionary perspective (Wheeler 2008). In other words, systematics reconstructs the relationship patterns between species at successively higher levels and through time and establishes classifications based on these patterns. Naming the groups obtained from systematic approaches is the objective of taxonomy.

1.2 The use of genetics in population ecology and conservation

Conservation genetics is the union of different disciplines, using analysis, methods and genetic techniques to direct conservation efforts, ensuring that species preserve their genetic diversity to be capable of dealing with human-induced environmental change. Knowledge on the genetic variation and structure of natural populations is essential for the conservation of biodiversity. The use of genetic markers contributes significantly to a better understanding of genetic diversity and architecture at different levels. On the one hand, at the interspecific level, genetic markers help clarifying evolutionary relationships and contribute to a better classification of natural units and, on the other hand, by looking at intraspecific patterns, they help identifying levels of diversity and population structure. Research projects on conservation genetics have been contributing to study and more properly understand the action of natural selection, the impact of hybridization in natural populations or the influence of genetic variation in the adaptive ability of natural populations. When accessing genetic variation, this contributes to evaluate the viability of both individuals and populations. Furthermore, it is possible not only to access the population dynamics and demography, but also biological factors such as ecology or behavior (Hoshino *et al.* 2012). With the loss of genetically and ecologically distinctive populations, severe problems appear as a consequence of the reduced genetic variation possibly leading to species extinctions, damaged ecosystems and destroyed ecological communities (Lacy 1997). However, increasing knowledge on patterns of genetic structuring, diversity and distribution, and identifying taxonomic and populations units, can be the basis to recommend the delineation of separate conservation units. Therefore, understanding the evolutionary history that shaped the genetic structure of populations contribute to guide conservation efforts.

Over the years, advances in molecular biology have led to the introduction of many new types of molecular markers, providing diverse kinds of information (Marsjan & Oldenbroek 2007). Understanding the nature of the information provided by each type of marker is thus important to make an informed decision about which marker suits best a particular study. Each molecular marker has its own mutation rate or expression level, and can be under different selective pressures, so the use of each marker depends on the objective to attain. Among other characteristics, they can vary in the level of polymorphism, number and nature of alleles, abundance, technical demands and costs (Maheswaran 2004). DNA sequence data and nuclear microsatellites are widely used markers. In animals, mitochondrial or nuclear genome

derived DNA sequences are frequently chosen to reconstruct phylogenies and historical patterns of population demography, biogeography and speciation.

Regardless from the marker, by choosing merely one locus for molecular analysis, the capture of the coalescent variation will be compromised and part of the evolutionary history can be lost (even if the marker being used is neutral). The use of multilocus approaches is crucial to reveal the possible complex mechanisms that may have conditioned the evolution of species. It makes possible to infer a variety of ecological processes, such as the identification of barriers to dispersal (e.g. Latch *et al.* 2008) or the inference of ancient population dynamics (e.g. Underhill & Kivisild 2007). By having information from different parts of the genome, we also have the advantage of obtaining a detailed and accurate picture of the global evolutionary history of the species. Nuclear loci are commonly used for phylogenetic tree reconstruction with multiple concatenated independent loci, being a traditional way of resolving species tree (Williams *et al.* 2013) with potential to yield supported trees if increasing the number of loci used. When research studies are working with short evolutionary timescales of very recent dynamics, the use of nuclear microsatellite data appears suitable (Nardi *et al.* 2005; Sequeira *et al.* 2008). These molecular markers are highly polymorphic with high mutation rates, abundant and fairly distributed throughout the eukaryotic nuclear genomes (Jarne & Lagoda 1996). These characteristics make microsatellites one of the most popular genetic markers for population genetics and inference of population structure (Wang *et al.* 2009). They can also provide information about genetic admixture among individuals, even when they are closely related, and it is also possible to cross with geographic information. However, the high level of polymorphism can lead to allelic size homoplasy which have been identified in different species (e.g. Viard *et al.* 1998; Culver *et al.* 2001) leading to inaccurate phylogenetics and population assessments if not taken into account.

Single-copy nuclear DNA (scnDNA) shows a slowest evolutionary rate, although mitochondrial DNA (mtDNA) still offers access to a distant range, on the order of potentially millions of years (Wan *et al.* 2004). Mitochondrial DNA is often the first molecular marker used to understand the genetic differentiation of species or populations and study phylogenies. It is technically easy to use, allowing gathering important initial information on the evolutionary history that can serve as basis to more complex and detailed molecular studies. Numerous mitochondria are found in cells, each carrying several copies of a single circular haploid molecule, which facilitates PCR amplification and analysis. Also, mitochondrial DNA has a higher mutation rate than nuclear DNA (nucDNA) and consequently provides a better resolution for recent evolutionary events. Therefore, the analysis of variation in mtDNA is useful to: 1)

reconstruct genealogies at the genus or the species level (Avise *et al.* 1987, 1998); 2) perform phylogeographic studies for species delimitation; 3) study intraspecific diversity. Among mtDNA regions, the cytochrome b is a mitochondrial protein coding gene and is also widely used for assessing intraspecific up to intergeneric level evolutionary associations (Kocher *et al.* 1989). Nonetheless, mtDNA only provides insights into the matrilineal pathways of ancestry, which represents only a small fraction of the complete historical record within the pedigree of a sexual reproducing organism (see e.g. Kryger 2002). Furthermore, mtDNA is also characterized by a substantial bottleneck due to its lower effective population size relative to the autosomes, which leads to within-individual diversity loss (Shoubridge & Wai 2007). Also, at distant evolutionary relationships (e.g. from the generic level to more ancient divergence events) mtDNA tends to accumulate redundant base substitutions (homoplasy) (Blouin *et al.,* 1998; Harris *et al.* 1998). Therefore understanding the forces that drive the molecular evolution of mtDNA is necessary for a proper use of this molecule as a marker for population genetic, phylogeographic, and phylogenetic studies (Nabholz *et al.* 2008). Moreover, introgression events often seem to affect mitochondrial DNA (Melo-Ferreira *et al.* 2012), being a problem if just one marker is applied. To obtain a complete picture of population history and evolutionary potential, data from nuclear loci is essential, as well as the inclusion of other genetic markers.

By incorporating modern phylogenetic/phylogeographic analyses, some important answers can be obtained, as for example inferences of demographic history, divergence times, migration rates, historical hybridization events or introgression (Hickerson *et al.* 2010). Combining both nuclear and mitochondrial DNA markers enriches the power of molecular data to test phylogenetic and phylogeographic hypotheses and allows to identify mito-nuclear discordances (conflicting results between the two types of markers) such as those identified in several animal systems (see Toews & Brelsford 2012). Still, the use of different types of markers for phylogeographic studies should be complemented with broad sampling across the species distribution area when possible.

1.3 Different types of biological samples for genetic analyses

Sampling is a crucial step for any population genetics and phylogeographic study and several approaches can be used to guaranty the necessary sample size. Direct handling and capture of animals to obtain samples can result in negative impacts such as altered animal behavior or inadvertent injury (e.g. blood or skin collection). As

an alternative, one recurrent way to obtain samples is from road killed animals excluding thus the need of pursuing live animals. Finding road kills is an effective way of acquiring DNA data that would be otherwise difficult to access, although the soil acidity and precipitation can be a problem in the preservation of DNA, both for tissue and bone samples. Non-invasive sampling arises as another alternative and consists on collecting animal signs, such as feces, urine or hair. Although obtaining genetic material from non-invasive sources has obvious benefits with regards to minimizing impacts on wildlife, some caveats of these samples include low DNA quantity and quality (fragmentation), contamination or presence of PCR inhibitors (Taberlet *et al.* 1996).

When fresh tissue sampling is not possible, the use of samples stored in museums is a viable and increasingly popular option that allows, for example, covering areas that are presently inaccessible. 'Museum genetics' is important for several other reasons, for instance in conservation genetics to monitor temporal (current versus historical) changes in genetic diversity in threatened species (e.g. Godoy *et al.* 2004), since most of the times it is possible to access the age of the samples, as well as relate it with climate change or with a population decline due to human impacts. However, the DNA preserved in these historic materials is generally of lower concentration, highly fragmented due to general degradation along time, and prone to contamination. The main common problems are short fragment length, an increased occurrence depurination of the DNA and an increased deamination of cytosine residues that occur primarily in the single-stranded overhangs of DNA fragments (Briggs *et al.* 2007; Brotherton *et al.* 2007; Pääbo 1989). This can lead to difficulties in PCR amplification, PCR artifacts or even genotyping complications. Even though mitochondrial DNA is more easily amplified from substandard DNA extracts than are nuclear genes (Cooper 1994), fragmented DNA template may cause incorrect bases to be incorporated in the PCR product (see Sefc *et al.* 2006 for artifact base changes in mtDNA). Also, caution and suggestions for detecting and avoiding errors in microsatellite genotyping in low DNA quantity samples have been published (e.g. Taberlet *et al.* 1996; Kalinowski *et al.* 2006). Despite all of these difficulties, there were several successful studies based on museum specimens using microsatellites, mtDNA sequencing or even with Next-Generation sequencing techniques (e.g. Bi *et al.* 2013; Themudo *et al.* 2014; Mitchell 2015). Clearly, working with this type of samples can bring some challenges, but when rigorous controls and well optimized protocols for preventing contamination by foreign DNA on the extraction process are applied it is an invaluable genetic resource.

1.4 Biodiversity in North Africa

Arid regions are normally assumed as plain areas with difficult conditions to sustain life, and therefore with less biodiversity when comparison to other climatic regions. Because of this, little attention is often paid to these regions and the biodiversity of these areas still poorly known and studied. However, deserts can be superb endemic species repositories due to past climatic events, and the Sahara desert is one example (Cloudsley-Thompson 1991).

North Africa is of considerable importance for studying local biodiversity in many ways. On the one hand, it covers the Sahara desert and the neighboring arid Sahel that constitute two major ecoregions of the African continent (Olson *et al.* 2001), characterized by patterns of low moisture, high temperatures ranges and strong winds (Villiers & Hirtle 2002; Dobigny *et al.* 2005; Brouat *et al.* 2009). The past strong climatic oscillations led to major changes in vegetation cover (Wang *et al*. 2008; Claussen 2009) and may often be reflected in geographic partitioning of genetic variation, which allows for the reconstruction of the evolutionary history of species. On the other hand, the Sahara separates the Mediterranean from the Tropical climate as well as it separates the Paleartic and Afrotropical biogeographical realms (Olson *et al.* 2001). In addition, Sahara desert not only has promoted vicariant speciation, but is also known to have acted as a barrier between mountain ranges that become isolated from each other and retain less harsh climatic conditions exchanges between North and Central Africa (Douady *et al.* 2003; Faleh *et al.* 2012). These mountains acted as refugia for many endemic species and isolated populations and display biogeographical island-like behavior (Gonçalves *et al.* 2012; Brito *et al.* 2014).

The unstable rainfall patterns and the strong aridification in the Eastern and Southern Africa are a result of global glaciations during Pleistocene (Grant & Leslie 1993). The cycles of contraction and expansion of major vegetation zones (see figure 1) led to deep effects on the distribution patterns and evolution of the African mammals (Grubb 1978; Coe & Skinner 1993), and recurrent habitat fragmentation and isolation events may have led to sub-speciation, speciation or extinction (Ewer & Cooke 1964; Grant & Leslie 1993).

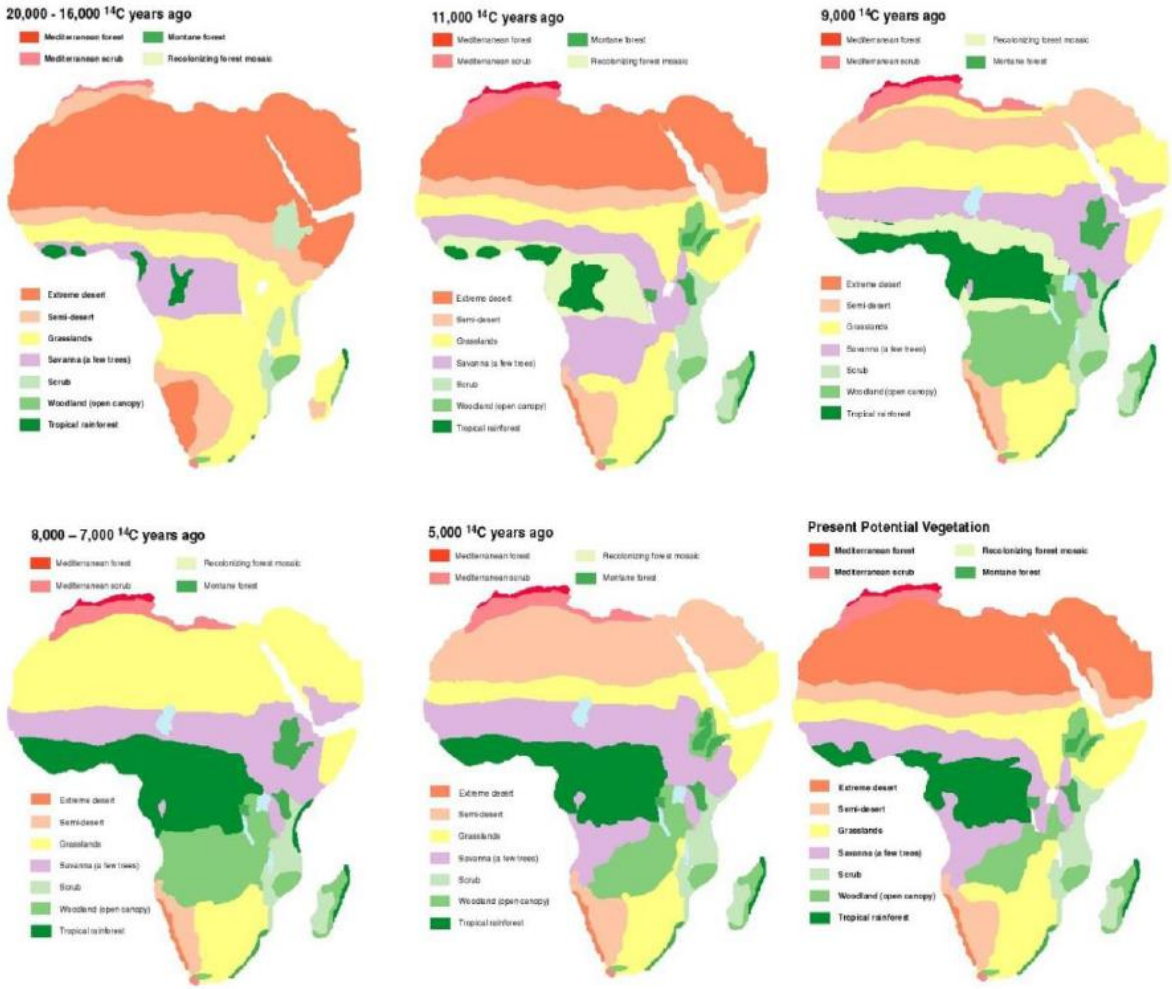


Fig. 1 - Temporal oscillations of the African ecosystems between the Last Glacial Maximum and the present; adapted from Adams and Faure (2004).

The progression of arid conditions, together with the increasing human economic activities, are negatively affecting desert biodiversity, phenology, physiology and distribution of many species, causing biodiversity loss (species extinction) (McNeely 2003; UNEP 2006; Thorton *et al.* 2008; Trape 2009; Pimm 2008). Still, the dimension, inaccessibility and political instability in the region contribute to an incomplete knowledge of the biodiversity of the Sahara-Sahel, even though the number of published scientific studies has been growing in the last years, most of them focusing on vertebrates with low dispersal activity, such as reptiles and amphibians (Brito *et al.* 2014). This means that only a few unconnected or loosely connected exploratory missions obtained data and there is no continuous biodiversity evaluation, resulting in rough species distribution maps (Le Berre 1989, 1990).

North African biodiversity is still poorly assessed and the fact that massive range oscillations occurred in the Sahara desert has probably induced events of

population isolation and opened the possibility for cryptic diversity to be formed. For instance, cryptic diversity has been reported by recent phylogeographic studies but the distribution of such diversity is mostly unknown (Brito *et al.* 2014). The North African biodiversity investigations in medium-sized mammals with relatively higher dispersal ability, such as the cape hare (*Lepus capensis*), are far from being completed due to a generalized lack of studies on this type of mammals. As such, there is a need to develop new studies, since few ones have explored the origin of Saharan biodiversity. It is of great importance to extensively sample these regions, to characterize its biodiversity, analyze taxa distributions, and relate them with environmental factors to identify biogeographic groups.

1.5 Hares as a model species

Genus *Lepus* is polytypic and belongs to the family Leporidae of the order Lagomorpha. The specimens of this genus are placental mammals and small to medium-sized herbivores that constitute the base of many predator-prey interactions (Chapman and Flux 1990; Mengoni 2011). Another important characteristic is that single individuals as well as whole populations can shift hundreds of kilometers in response to environmental changes or in search of better feeding grounds (Kryger 2002). Moreover, this is an economically and scientifically central group of mammals as they are a major human food resource, model laboratory animals, valued game and provide scientific insights into entire trophic systems (Chapman & Flux 2008).

Over the last years, hares have been used as model species in population genetics, not only to understand their phylogeography, but also to study significant evolutionary mechanisms, such as reticulate and adaptive evolution (Alves *et al.* 2008b). *Lepus* is an evolutionary recent group (Chapman & Flux 2008) in which have a considerable ecological importance (Alves *et al*. 2008a). Fossil investigations propose that this genus had a rapid radiation just within the last 2 – 2.5 million years in evolutionary terms leading to a low genetic differentiation between the recently formed taxa (Suchentrunk *et al.* 2008). Additionally, the severe climatic fluctuations during Pleistocene or Holocene may have induced major shifts in species distributions, forcing them to retract, expand, move and/or fragment their ranges (Hewitt 1996). This may have also set the conditions for competition and secondary contact among hare species, producing patterns of genetic variation that reflect introgressive hybridization (Melo-Ferreira *et al.* 2007) and incomplete lineage sorting. Eventually, this could also contribute to the prevailing taxonomic uncertainties within genus *Lepus* and to wrong

systematic conclusions (Suchentrunk *et al.* 2008). Within the genus, taxonomic difficulties may also result from low gene pool divergence among species, leading to vast sharing of characters due to recent common ancestry (Alves *et al*. 2003; Robinson & Matthee 2005).

Hares have not adapted to a single particular environment: representatives of the genus have been extraordinarily successful in occupying the most diverse habitats throughout the world, from rather arid areas to open grassland or even snow (Colbert 1980; Skinner & Smithers 1990). This genus is estimated to have arrived to Africa in the Pliocene or early Pleistocene (Carroll 1988). The oldest fossils of true representatives of the genus *Lepus* in Africa date back to the early Pleistocene (Walker 1964; Lavocat 1978). Analysing mtDNA restriction sites, Biju-Duval *et al.* (1991) estimated that the most common ancestor of the genera *Lepus*, *Oryctolagus* and *Sylvilagus* lived 6-8 million years ago. At this time the African climate was characterized by cycles of cold and dry phases alternating with warm and humid phases (Grubb 1978; DeMenocal 1995).

Over time, many different hare species have been identified in Africa, mostly relying on morphological traits (e.g. Suchentrunk *et al.* 2006; Palacios *et al.* 2008). From the approximately 32 worldwide *Lepus spp (*Hoffman and Smith 2005), six may be distributed in Africa (figure 2), in arid, semi-arid and savanna habitats - *L. capensis, L. saxatilis, L. microtis, L. habessinicus, L. fagani and L. starcki*. hese six African species are distinguished morphologically by body size, ratio of width of mesopterygoid space to minimum length of hard palate, teeth features (the shape of groove on principal incisor tooth and presence/absence of cement in that groove) and quantity of black coloration in the ears (Kingdon *et al.* 2013).

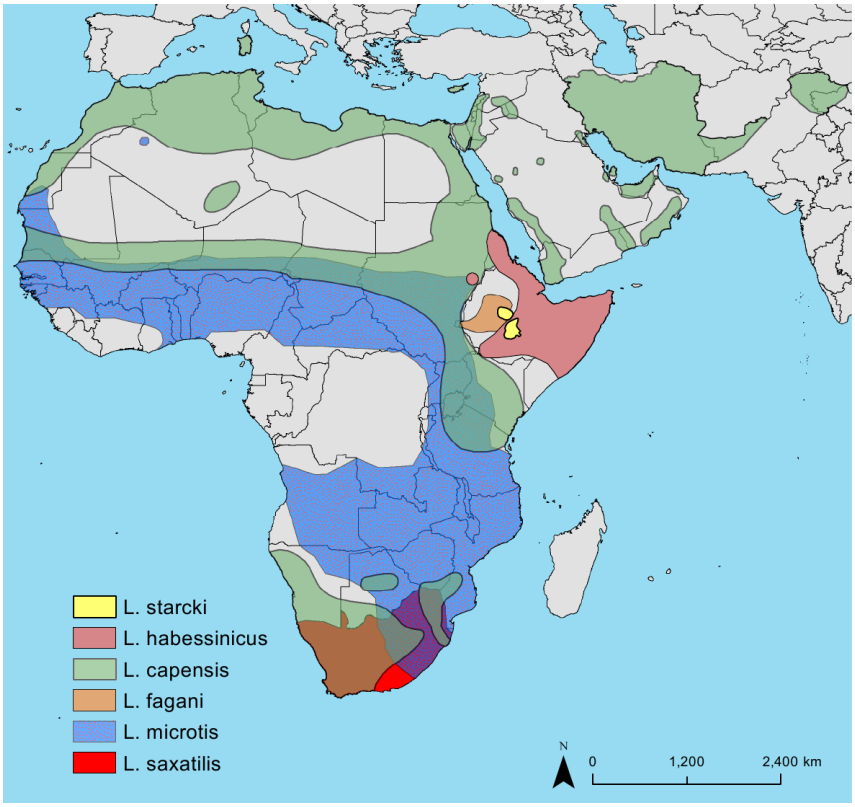


Fig. 2 - Distribution of the six African hare species.

North African hares are represented by two species – Cape Hare (*Lepus capensis* L Linnaeus, 1758*)* and African Savanna Hare (*Lepus microtis,* previously included in *L. saxatilis* and has been classified under several other names as *victoriae, crawshayi* or *whytei,* between others; although *Lepus microtis*, is considered a "*nomen dubium*" (Petter 1959; Angermann 1965), in this work we will referred to this species as it is identified at IUCN website - http://www.iucnredlist.org/). Azzaroli-Puccetti (1987) considers that *whytei* (from Malawi) is close to *L. fagani* (from Ethiopia), although they are distinguishable on cranial and dental characters, and the wide geographic separation between them merits recognition of *whytei* as a valid species. These two Northern species occupy large ranges and their distribution is overlapped as well as with the South African species *L. saxatilis.* Within the same zoogeographic region *L. microtis* has a small, isolated population in Eastern Algeria, although its large sub-Saharan range extends North through Western Mauritania to incorporate the Southern third of Western Sahara (where this species may occur sympatrically with *L. capensis*). The subspecies taxonomy for *Lepus saxatilis* has mainly been derived from characters such as body size, fur coloration and geographic locality (Roberts 1951), since few studies were done and mostly based on mtDNA sequences (e.g. Kryger *et al.* 2004). *Lepus fagani* is distributed in Northern and Western Ethiopia on the plateau and in Kenya at high altitude, and *Lepus starcki*, the Ethiopian highland hare, is the less

widespread African hare species. Also, Tolesa (2014) reinforced that *L. habessinicus* has distinct mtDNA relative to *L. capensis* and both nuclear and morphological data suggest that *L. habessinicus* is closer to other hare species in Ethiopia.

The two particularly widespread species in Africa, *L. capensis* and *L. microtis*, have many synonyms, testifying the large variation within each species, their wide distribution, complex evolutionary history and lack of comprehensive data on their phylogenetic position. The cape hare also occurs throughout most of the Middle East and Eastwards North of Himalaya Mountains to China, and its widespread distribution, inter-population variation and varying pelage colors might justify the many alternative names. Forty-four synonyms were listed by Hoffmann & Smith (2005) just in Africa and Near East, of which are considered to be subspecies (although 80 sub-species have been described by Flux & Angermann (1990), but from Africa until East China). Being a polytypic species, this shows the need to revise the taxonomy of the Cape hare. Furthermore, the confusion extends to a lack of agreement between classifications made by separate authors, with Hoffman and colleague (2005) describing five subspecies in Arabia and Near East, while Harrison & Bates (1991) referring to six sub-species of Cape Hare living on mainland Arabia and two subspecies living on offshore islands. Other important facts that also need further clarification are the taxonomic limits of this species and its uncertain relationships with *L. microtis*. Morphological comparisons between these two overlapped species demonstrated that *L. microtis,* unlike *L. capensis,* has more colorful and thicker fur, with russet areas on the nape and chest sides and has incisors with deep grooves (Moores *et al.* 2012). Despite the description of these diagnostic characteristics, distinguishing *L. capensis* and *L. microtis* in the field can be very challenging.

South Africa (Cape of Good Hope) is the type locality of the cape hare, in which a single species (*capensis sensu lato)* inhabits Africa and Near East in two separate non-forest areas. According to Hoffman and Smith (2005), there is no evidence of gene flow between the South African populations and the populations in East, West and North Africa. The same authors subdivided informally *capensis sensu lato,* in four distinct groups that might be considered as distinct species, pending on sufficient data to support a formal revision: South Africa (e.g. *centralis* Thomas, 1903; *granti* Thomas and Schwann, 1904; *aquilo* Thomas and Wroughton, 1907), East Africa (e.g. *aegyptius* Desmarest, 1822; *hawker* Thomas, 1901; *isabellinus* (cretzschmar, 1826; *sinaitocus* Ehrenberg, 1833), Arabia and Near East (e.g. *arabicus* Ehrenberg, 2833; *omanensis* Thomas, 1894), and Northwest Africa (e.g. *schlumbergeri* Remy-St. Loup, 1894; *atlanticus* De Winston, 1898; w*hitakeri* Thomas, 1902).

Despite the studies done focusing on Cape hares, no consensus was obtained concerning their phylogenetic relationships or distribution limits due to the lack of comprehensive data. But the presence of other species in the region has been suggested and contradicts Petter's (1959, 1961, 1972) concept that all these hares belong to *L. capensis*. Based on morphological data, the latter author even included brown hares (*Lepus europaeus* Pallas, 1778) from Europe and other parts of the Western Palearctis (e.g Anatolia) into *L. capensis*. Angermann (1965) likewise, based on morphological and morphometrical data, considered hares from northern Tunisia very similar to brown hares, but was later less secure of this possible classification (Angermann 1983). Conflicting results were also found when Suchentrunk and his colleagues (2009) analyzed 13 microsatellite loci and CR-1 gene in Cape hares from South Africa, which showed to be a monophyletic group. In contrast, Palacios *et al.* (2008) results suggest the split of Cape hare into two species *L. capensis* and *L. centralis* based on intensive skull morphometric data. Additionally, a population genetics study (Lado *et al.* 2013) using mitochondrial DNA sequences from sampled populations in different localities of North Africa, showed high levels of genetic diversity, and highly divergent and geographically structured lineages. This work suggested deep evolutionary fragmentation of these populations possibly with strong barriers to gene flow (figure 3) and possible cryptic species. Also, some haplotypes were found to be more closely related to other species, which can result from mtDNA introgression, a phenomenon widely described in the genus (e.g. Alves *et al.* 2003; Melo-Ferreira *et al.* 2012, 2014a; Cheng *et al.* 2014), or retention of ancestral polymorphism. These results suggest that the evolutionary history of these populations is far from being simple and may uncover relevant biogeographic and evolutionary processes.

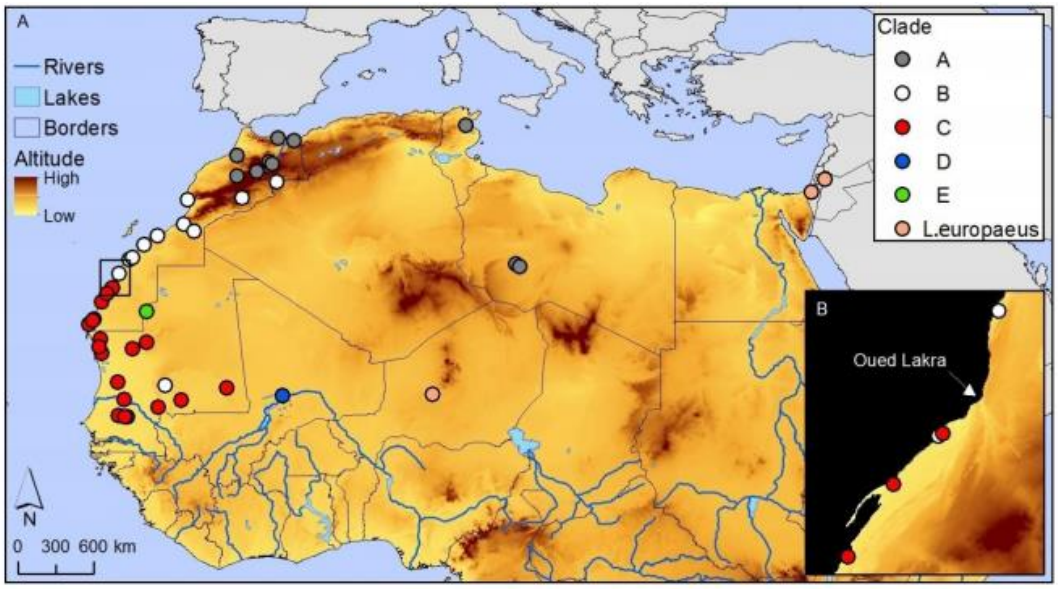


Fig. 3- A. Distribution of the phylogenetic clades in North Africa; B. Distribution of clades B and C in the contact zone of the Atlantic Costal Sahara; adapted from Lado *et al.* 2013 (internship report).

The present knowledge on some hare species is in general insufficient, and the taxonomy and the evolution of North African hares are far from being completely understood. At this stage, only a comprehensive characterization of the genetic diversity and divergence of these populations can clarify the existing taxonomic problems without *a priori* assumptions. It is necessary to clarify and precise the geographic distribution of the African species, and to collect field data on the abundance, habitat requirements and biology of the species so conservation efforts are possible.

1.6 Objectives

The delimitation of the different species and subspecies of the genus *Lepus* is still unresolved, and this situation is magnified in Africa. Previous studies have demonstrated that African species *Lepus capensis* have high genetic diversity and strong population structure. The fact that individuals are morphological similar throughout the *L. capensis* range, but genetic evidence points to highly diverse and structured populations, brings the possibility of the existence of cryptic species. Also, the evolutionary relationship between this species with African and non-African neighboring species is still unclear. In order to clarify the phylogeny and phylogeography of African hares, and identify relevant units for conservation, the global levels and patterns of genetic differentiation between African populations and

neighboring species were investigated. To do so, an increased number of molecular markers when compared with the previous studies, was used, with representations of matrilineal and biparentally inherited genetic markers (18 newly developed nuclear microsatellites, and one mitochondrial and five nuclear autosomal genes), with the broadest sampling size to date (taking advantage of museum samples) and including sequences of four more hare species. The specific objectives of this research were to:

1) determine the population structure of African hares, with a particular emphasis on *Lepus capensis*, and relate differentiation with geography;
2) assess how different evolutionary lineages are related, inferring divergence time between populations and interpreting the results in light of the climatic history of the area;
3) understand population dynamics and demographic history of the different entities;
4) infer the level of divergence among evolutionary units and identify possible cryptic diversity;
5) verify mito-nuclear concordance and possible occurrence of gene flow.

The results of the study were expected to provide valuable insights into the evolutionary history and biogeographic patterns of *L. capensis* and neighboring species, and contribute towards the clarification of Africa hares taxonomy by comparing the genetic lineages obtained in this study with the current systematics.

2. Materials and Methods

2.1 Study area and sampling

A total of 261 samples with different origins were included in this study. These samples will here be described considering the species to which they were originally attributed by their providers but, given the difficulties in properly classifying African species of hares, no *a priori* assumption of the species origin was considered for analyses.

In total, 219 samples from *Lepus capensis* were included, with different origins. Road killed samples (171) were mainly collected during fieldwork along Northwest Africa by the research group BIODESERTS (CIBIO/UP-InBIO), and the remaining ones were provided by other researchers and collaborators. All were preserved in ethanol (96%). This set of samples also included two feces and three bone samples. The remaining *Lepus capensis* samples (48) were collected in two distinct museums: the Vienna's and Berlin's Museums of Natural History. To prevent contaminations, museum samples were collected with plastic gloves, face mask, lab coat and handled with a different blade. From the South African hare species, *Lepus saxatilis,* six samples were available at the facilities and included in this study. In addition, 25 museum samples considered as *L. victoriae,* now taxonomically recognized as *Lepus microtis* and six unidentified species samples were included.

All samples used in this work (including their geographical origin and presumable taxonomical classifications) are described in appendix 1. The detailed geographical location of samples from road killed animals and a few museum samples were recorded with Global Position System (GPS). For the remaining samples, approximate coordinates were given according to the available information about the geographic origin of each sample. The plot of the geographical locations of the samples (fig. 4) was done using the Geographical Information System ArcMap 10.1.

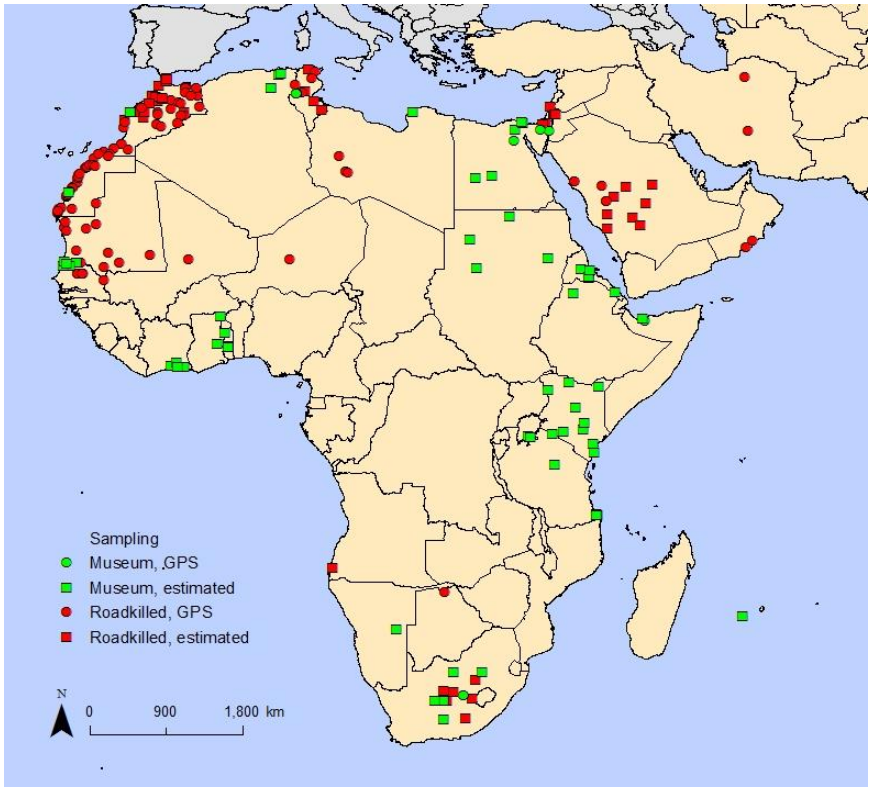


Fig. 4– Samples distribution and origin. Each square/circle corresponds to an individual sample.

2.2 DNA extraction

Genomic DNA was extracted from liver or ear tissues using the Jetquick GENOMED kit. DNA extraction of the two non-invasive samples was performed using the Tissue kit E.Z.N.A. and in an isolated and autonomous room (low-DNA-status room) with special and sterilized equipment, in order to prevent contamination.

The museum samples were extracted under the same conditions as the non-invasive ones, but before this step, the dry tissue samples were hydrated with T.E. (Tris + EDTA) solution changing the liquid every day during five days in a 37ºC incubator. After this step, the samples were conserved in 96% ethanol. Since DNA from these samples was expected to be a degraded, two different extraction kits were tested for a small museum dataset of five samples:

1. QIAGEN's QIAmp® DNA Micro Kit, following the respective protocol for tissue samples;
2. following Bi *et al.* (2013) which suggest a combination of DNeasy Blood & Tissue Kit, using its protocol and reagents, but with the QIAquick PCR Purification kit columns.

The results of both procedures were compared, from the DNA rough quantification by agarose gel electrophoresis (0.8%) to sequencing results, and the second approach was then chosen to do the DNA extraction of the other samples given the higher amounts of DNA obtained and the observation of DNA sequence chromatograms with less background noise.

For the bone samples, an ancient DNA extraction protocol (Rohland & Hofreiter 2007) was optimized for extracting DNA from bone powder. A negative control was used to check for contamination in the extraction proceeding and extractions were stored at low temperature (around -20ºC). In addition, samples that were already extracted in Lado *et al*. (2013) and extractions available at the facilities were tested to check if DNA was still in good conditions for analysis. When DNA was found in high amounts (big tight band of high molecular weight in the electrophoresis gel) or degraded (smearing band) extracts were properly diluted in ultra-pure water to avoid the amplification of PCR artifacts. Otherwise, DNA extractions were used directly for the following PCR and sequencing reactions.

2.3 Genotyping of microsatellites

2.3.1 Selection of microsatellites

In this work, species-specific microsatellite multiplexes of African *Lepus capensis* were newly developed to produce population structure analysis. To accomplish the objective, 12 samples of this species were sent to GENOSCREEN (http://www.genoscreen.fr/) to produce high throughput microsatellite libraries on a new generation sequencer 454-GsFLX ® (Roche Diagnostics). The library results information containing raw files, microsatellite motifs and number repeats, and bioinformatic validated primer pairs were sent back as a database. In order to proceed with the analysis, a preliminary selection of suited markers was done and the microsatellites tested for multiplexing (multiple microsatellites amplification in a single PCR including more than one pair of primers in the reaction).

In total, 35 markers were selected among those marked as “best” section according to the GENOSCREEN assessment. The selection of markers followed the following criteria:

1. With more than 12 repeats (high probability of polymorphism between individuals);
2. After calculating annealing temperature, primers with more than 5ºC of difference between each other per multiplex were avoided and also between

Forward (Fw) and Reverse (Rv) primers of the same marker (better for amplification by PCR);

3. After calculating the primers' sizes including tails, maximum of 250 base pairs were chosen and different repeat motifs were preferred;
4. For each multiplex, the four different types of dye (FAM, VIC, NED, PET-M13 tails) had to be balanced. The minimal difference between loci labeled with the same dye color was 70 base pairs.

AUTODIMER (Vallone & Butler 2004) was used to predict hairpin structures and primer dimers within multiplexes, in order to increase the probability of success of the multiplexed PCR reactions.

2.3.2 Optimization of PCR conditions

The test and optimization of PCR conditions of the microsatellite multiplexes was performed with a set of eight samples, selected along the samples distribution range in Africa. PCR amplifications were performed in BIOMETRA with a 10µl reaction volume of multiplex cocktail containing 1 or 2µl of DNA depending on the DNA concentration, 5µL Qiagen PCR Master Mix, 1 µl primer mix (containing forward and reverse primers and fluorescent dye of each primer – M13 tails) and water. Electrophoretic runs in agarose gel (2% p/v) with TBE (Tris-Borate-EDTA Buffer; Tris 89mM, Boric Acid 89mM and EDTA 2mM, pH 0.8) buffer with the resulting PCR products to see the amplification success. The gel was stained with 17.5µL/L GelRed (Biotarget) to allow the visualization of DNA bands under UV light. To load the gel, 3µL of each post-PCR product was used in combination with 2µL of bromophenol blue, which helps to monitor the electrophoresis progress and provides extra density to the sample. In each electrophoresis run, 2.8µL of Marker 5 (molecular weight size marker; Eurogenetec) was used in a separate well to infer the approximated size of the DNA fragments in the gel. The electrophoresis was performed at 300V and gels were visualized under UV light. Samples with no observed band in the gel were not chosen to be sequenced. For the markers with apparently low quality, individual PCR's (uniplexes) were performed for a better understanding of marker's individual amplification success and profile. A negative control was used in every reaction. With consecutive tests of each multiplex, primer concentrations (increasing primer concentration for the weakest or decrease primer concentration for the strongest ones) and temperatures during PCRs of each loci with the test samples were optimized for amplification success, until overall amplification and genotyping was possible for all

samples. Some loci were excluded due to specific problems, such as lack of polymorphism, amplification failure or unreadable profile, and the multiplexes were rearranged. From the initial 35 microsatellites, 18 were finally optimized, organized in three multiplexes and all samples were amplified (final PCR conditions in the appendix 2 and for multiplex organization, primers concentrations and amplification conditions see appendix 3). The museum samples, bones and feces were amplified four times each to ensure consistency of genotype determination. For the road killed samples, 30% was validated with repetitions. A consensus was made for all samples that were amplified more than once, where individuals with similar genotype between repetitions were accepted.

2.3.3 Genotyping

Genotyping was performed on an automatic sequencer ABI3130xl Genetic Analyzer (Applied Biosystems), where the products were separated by capillary electrophoresis using 1 or 2µl of PCR, for 10µl of LIZ NEW size standard mix (a size standard and allelic ladder – mix of 1000 µL of formamide plus 75-400 (-250) for 20 µL of liz). GENEMAPPER v4.0 (Applied Biosystems) was then used to read and score the genotyping results. In order to properly bin and score the allele calls, after each run the ladder recognition was manually edited. Automatic allele scores were then visually inspected by two independent researchers to avoid potential errors in the automatic process. Genotyping was repeated whenever uncertain allele calls were obtained. Samples with doubtful or inconsistent genotypes among replicates were considered missing data.

2.3.4 Population structure analysis

The final genotypes were filtered for missing data, with individuals with more than three loci with missing data being discarded. Using FSTAT (Goudet 1995), the number of alleles sampled were calculated. GenAlex v6.5 (Peakall & Smouse 2012) was used to produce the Genepop input file. Exact tests for pairwise linkage disequilibrium (LD) and deviation from Hardy-Weinberg expectation (HWE) for each locality using Bonferroni correction, were conducted with GENEPOP (http://genepop.curtin.edu.au/; Web interface for remote computations). Markov chain parameters for exact tests were set at 10,000 dememorizations, 100 batches, and 5000 iterations per batch (Raymond & Rousset 1995). Moreover, MICRO-CHECKER

(Van Oosterhout *et al.* 2004) was used to check for potential genotyping errors resulting from null alleles or scoring mistakes. The number of observed alleles was calculated with Genalex. STRUCTURE v.2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2003), a Bayesian tool for inferring population structure, was then used to infer the uppermost hierarchical level of genetic groups of individuals. This software assumes a model where K populations exist, and individuals are assigned to one or more populations (if they have admix genomes) where loci are in HWE and LD within populations (Pritchard *et al.* 2003). This can have several applications such as unravelling population structure and identifying migrants and admixed individuals. The parameters were defined with admixture model, providing the proportion of the genome of each individual that can be assigned to the inferred. Also, correlated allele frequencies model was assumed, for a number of clusters (K) from 1 to 25 with three replicates of each K, with a burn-in period of 1,000,000 generations and 1,000,000 MCMC repetitions after burn-in, to assure similar posterior probabilities of the data in each run. With the same parameters, the datasets were run assuming now independent allele frequencies instead of correlated. Correlated allele frequencies model assumes a level of non-independence being more conservative and providing greater power to detect distinct populations that are particularly closely related. But independent allele frequencies model requires knowledge about the correlation levels across populations. Nevertheless, the first model will tend to give the same results as the independent allele frequencies model in the absence of high levels of correlation across populations (Porras-Hurtado *et al.* 2013). To identify the most likely K, Evanno's Delta K method was used (Evanno *et al.* 2005), as implemented in STRUCTURE HARVESTER v.0.6.94 (Earl & vonHoldt 2012).

Additional substructuring of the dataset was then inferred using a hierarchical approach, where at each round of the analysis the dataset is split according to the best population structure defined in the previous generation, and STUCTURE re-run independently. This hierarchical analysis proceeded until no structure was found in each thread of the analysis. In each round only individuals with at least 80% association in a cluster were included in the succeeding analysis (Coulon *et al.* 2008; Cheng *et al.* 2014). The same model parameters were used in these analyses, just changing the K value according to the number of sampled populations in the cluster that was being examined.

Another Bayesian clustering analyses was in addition performed using BAPS v.6.0 (Corander *et al.* 2008) which takes advantage of spatial clustering of individuals. This was followed by an admixture analysis with 200 iterations, 200 reference individuals, 50 iterations of reference individuals and no need for K values, since BAPS

directly infers the number of populations from the dataset. The admixed individuals were identified with the same parameters of the STRUCTURE analysis. In addition, results of the model-based clustering were compared to a Principal Coordinate Analysis (PCA; using individual-by-individual genetic between the identified clusters on subsets of data) calculated using GenAlex.

2.4 DNA sequencing

2.4.1 DNA amplification and sequencing of mitochondrial and nuclear DNA

One mitochondrial DNA (partial Cytb) and five autosomal nuclear DNA (partial TF, SPTBN1, HPX, OXA1L and DARC) loci were amplified by polymerase chain reaction (PCR) and sequenced in this study (for primers and conditions see appendix 4). The nuclear DNA primers were already available from Melo-Ferreira *et al.* (2009), Matthee *et al.* (2004) and Wallner *et al.* (2001). Cytb primer pair, LcpCytb, was designed in Lado *et al.* (2013). A new Cytb primer pair which comprehend a smaller fragment was designed, LcpCytb2, for PCR amplification with the Primer3Plus (Untergasser *et al.* 2007), for amplification from DNA extracts from museum samples and bones, which showed high levels of fragmentation and likely caused failure of the amplification of the larger fragment. This primer pair was designed in conserved regions within the largest fragment of LcpCytb. Primers with 18-22 bp, melting temperature of 52-58ºC, and a GC constitution of 40-60% were searched for. NetPrimer computer software (Premier Biosoft International, Palo Alto, CA, USA) was used to verify if there were events of primer secondary structures. Nuclear genes were amplified for a set of 80 road kill samples covering all the study area where museum, feces and bone samples were not included, since after trying to amplify them it was not successful (it is necessary to design primers for smaller fragments). All PCR's were performed in 10µl reaction volumes in which 5µl were QIAGEN PCR MasterMix, 3µl were of pure water, 0.5µl of both reverse and forward primers, and from 1µl to 3µl of template DNA depending on the sample quality. For museum samples, all pre-amplification steps were carried out in a separate room with technical equipment that has never been used for fresh tissue samples or PCR products. Minor adjustments to conditions were required in some reactions, principally with the museum, bone and feces samples (diluting or adding 2µl of DNA instead of one). The choice of extension and annealing conditions in the programs sought a balance between high amplification

success of the target loci, and low amount of untargeted product. All the reactions took place in a BIOMETRA T-professional Thermocycler and PCRs were executed using again a negative control. The resulting PCR products were analyzed using an electrophoresis in a 2% agarose gel with TBE buffer with a concentration of 0.5X. The gel was stained with 17.5µL/L GelRed (Biotarget) to allow the visualization of DNA bands under UV light. To load the gel, 3µL of each post-PCR product was used in combination with 2µL of bromophenol blue, which helps to monitor the electrophoresis progress and provides extra density to the sample. In each electrophoresis run, 2.8µL of Marker 5 (molecular weight size marker; Eurogenetec) was used in a separate well to infer the approximated size of the DNA fragments in the gel. The electrophoresis was performed at 300V and gels were visualized under UV light. Samples with no observed band in the gel were not chosen to be sequenced. PCR products were purified using a combination of two enzymes, Exonuclease I and Shrimp Alkaline Phosphatase (USB® ExoSAP-IT® PCR Product Cleanup, Affymetrix) to remove residual primers and unincorporated nucleotides. Purified products were sequenced using in the facilities of Macrogen Inc (Netherlands) by traditional Sanger sequencing, using the same primers used in the amplification process. The museum samples were sequenced two times for Cytb due to the drawbacks caused by fragmented DNA. OXA1L fragment was sequenced with the forward and reverse primers. In cases of detection of heterozygous insertion-deletions (indels), fragments were sequenced both with forward and reverse primers. For the remaining situations, 20% of the sequences were obtained using both the forward and reverse primers for validation.

2.4.2 Sequence alignment, data treatment and genetic diversity

The software FinchTV 1.4.0 (Geospiza, Inc.; Seattle, WA, USA; http://www.geospiza.com) was used to view and edit by hand DNA sequence chromatogram data. All analyses were performed separately for the mitochondrial and nuclear genes to compare the concordance of results. The final alignments were complemented with previously published sequences downloaded from GenBank of *L. capensis* (when available), and also other species as *L. europaeus, L. saxatilis, L. granatensis and L. timidus* to take into account interspecific-divergence (note that no *Lepus microtis* sequences were available in GenBank). The European wild rabbit, *Oryctolagus cuniculus,* was used as outgroup (for GenBank accession number see appendix 5) in some analyses. The sequences were then imported into BioEdit (Hall & Hall 1999), a Sequence Aligner Editor software, for aligning and further manual editing

for each of the six genes, since not all the sequences were of the same length. Within the OXA1L sequence a region with tandem repeats was discarded.

The International Union of Pure and Applied Chemistry (*IUPAC*) nucleotide ambiguity code was used for polymorphic positions in the diploid genes. For each of the five nuclear genes, PHASE v2.1.2 (Stephens *et al.* 2001; Stephens & Scheet 2005), a Bayesian statistical method, was used to estimate the allelic states and the input files were generated on the online software SeqPHASE (Flot 2010). Information from known phased derived from heterozygote insertions-deletions (Flot *et al.* 2006) was incorporated in the analysis to improve phase determination. Three replicates of each PHASE run was performed with 1000 generations after 1000 generations of burn-in with a thinning interval of 1, to ensure the consistency of the phase determination by frequency estimates. All individuals were maintained in the final datasets, choosing the most probable reconstructed haplotypes for population genetic analyses. PHASE has been shown to have a low rate of inferring haplotype pairs incorrectly and simply excluding unresolved genotypes with low probabilities may lead reductions in overall diversity, mainly by eliminating low frequency variants, and systematic bias in estimates of effective population sizes (Garrick *et al.* 2010). The software DNAsp 5 (http://www.ub.edu/dnasp/) was used to calculate indices as nucleotide diversity, number of haplotypes, haplotype diversity and segregating sites for Cytb.

2.4.3 Gene tree phylogenies

Genealogical relationships among haplotypes were determined by constructing Median-Joining (MJ) networks (statistical parsimony) with Network version (Bandelt *et al.* 1999); http//:www.fluxus-engineering.com) for mitochondrial DNA sequence datasets. The median-joining network procedure infers haplotype genealogies frequently with median vectors (i.e. unsampled or extinct haplotypes). Two datasets corresponding to sequence alignments of different lengths, with and without museum samples, were analyzed.

To find out which was the best evolution model, the fit of the alignment to 88 models was performed using the software jModelTest V. 2.0 (https://code.google.com/p/jmodeltest2) under the Akaike's information criterion, corrected for low sample sizes (AICc) for the six genes. Molecular phylogenies were derived using Bayesian Inference (BI) and Maximum Likelihood (ML) methods for the six genes. Again, for mitochondrial DNA, both datasets with and without museum samples were used separately.

ML phylogenetic inferences were performed for each locus using Garli v1.0 (Zwickl 2008) with five independent search replicate runs by specifying the optimal mutation model (model parameters for phylogenetic trees building were not fixed) and the outgroup. No starting topology was defined and the program was set to run until no significantly better scoring topology was encountered after 5,000,000 generations for the nuclear genes and after 50,000,000 generations for the Cytb. For the cytb phylogeny, the support of the resulting nodes was estimated using 500 bootstrap replicates. Finally, the Shimodaira–Hasegawa (SH) test (Shimodaira & Hasegawa 1999) using PAUP* (Swofford 2001) was implemented to test whether the resulting best trees were statistically significantly different.

In addition, BI was performed for all genes using software BEAST v1.8.1 (Drummond & Rambaut 2007) and the previously determined mutation model. This program assumes no gene flow and no recombination within locus. The input file for BEAST was created using the application BEAUti, part of the package. The posterior probabilities were determined after runs of 50,000,000 and 100,000,000 generations, sampling trees and parameter estimates every 5,000 and 10,000 generations, for each nuclear and Cytb (with and without museum samples) haplotypes respectively, with the strict clock and the Baysian skyline coalescent (BSP) model. Three replicates of each loci were run and its stability and convergence of the Markov chain Monte Carlo (MCMC) effective sample size (ESS) were measured using the software Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). Thus the initial 10% of the run were excluded as burn-in. With LogCombiner, also part of the BEAST package, combined files .log and .tree from the multiple independent runs and summary trees were generated with TreeAnnotator v1.8.1 (part of the BEAST package), and also an output for visualization in FigTree v1.4.2 software (http://tree.bio.ed.ac.uk/software/figtree/). The same conditions were set but for the Yule tree prior applied and the uncorrelated lognormal relaxed clock (Drummond *et al.* 2006) model in which the datasets with and without museum samples were reduced to haplotypes. While the BSP is a coalescent model suitable at population level, Yule is a speciation model appropriate for analyses of inter-species relationships. Given the mixed nature of the datasets, both models were used and results compared.

Definition of the number of populations with the Cytb dataset was also assessed using BAPS v.6, the Bayesian approach of population structure but taking advantage of the spatial approach (coordinates of each sample were given) used also with the microsatellite data, excluding GenBank sequences from this analysis. The nuclear divergence between lineages was calculated by the Dxy genetic distance with MEGA 5 (Tamura *et al.*, 2011)

2.4.4 Species tree reconstruction based on nuclear loci

Due to the stochasticity of coalescent processes, it is likely that gene trees differ among loci and not matching the original species tree due to incomplete lineage sorting alone. The species-tree Bayesian inference method *BEAST (Heled & Drummond 2010), implemented in BEAST software was used to infer the phylogeny in African hares based on the five nuclear loci. *BEAST also assumes that no recombination occurs in the analyzed sequences. Here we opted to maintain the full alignments, irrespective of possible recombination events because it has been shown that the negative impact of intra-locus recombination in the estimate of species-trees in coalescent based methods is negligible (Lanier & Knowles 2012). Input files for BEAST were generated in BEAUti. Specimens were *a priori* assigned to populations, a requirement of *BEAST, based on the main clades from microsatellite results, plus *L. capensis* from China (obtained from GenBank) and the other hare species included in the previous analysis. The outgroup was not included in the species tree as this method estimates the root of each gene tree and uses the multispecies coalescent. (Heled & Drummond 2010; Melo-Ferreira *et al.* 2014b). Four different datasets were analyzed and compared:

1. All *L. capensis* populations and hare species, including five nuclear loci included to understand the relationship between populations and neighboring species;
2. Same as the first inference, but in this case excluding South African population of *L. capensis,* since this population individuals only amplified for three of the five nuclear loci;
3. The two genes which did not amplified for South African *L. capensis* population were excluded in all *L. capensis* populations and species analysed: this species tree only includes three nuclear loci, but all populations and species;
4. *L. capensis* populations (excluding Chinese population) and *L. saxatilis*, to understand the relationship between African populations, including the five loci.

Nucleotide substitution models were set for each locus as assigned by jModelTest. In cases for which the best-fit model as determined by jModelTest was not implemented on BEAST, the next most parameterized model available was used. Posterior phylogenies were determined in *BEAST using an uncorrelated lognormal relaxed clock and the Yule tree prior. Prior settings were set as default except for the

relaxed clock standard deviation prior which was set to an exponential distribution with a mean of 0.5 as recommended by the authors. Three independent runs of 250,000,000 generations sampling every 25,000 generations were performed. Convergence of the Markov chain Monte Carlo (MCMC), stationary of the runs and effective sample size (ESS) for each parameter of interest in the analysis were evaluated using the software Tracer. The initial 10% of the runs were discarded as burn-in. Summary trees were generated in TreeAnnotator. The resulting trees were then analyzed in FigTree. Note that this model assumes no gene flow among the considered species/populations. Therefore, this was verified by other means (see below).

2.4.5 Demography

Demographic inference methods were applied by Bayesian skyline plots (BSPs; Drummond et al., 2005) with mtDNA for each phylogenetic lineage and the derivative extended Bayesian skyline plot (EBSP; Heled & Drummond 2008) including all five nuclear gene data for each microsatellite cluster using software BEAST v1.8.1. Both types of markers were separated in the analysis given the strong discordances found in some instances, with possible causes that would violate the assumption of panmixia. Operators of each EBSP run were adapted as indicated in the manual on the "operators section". Three independent runs of 100,000,000 generations were performed for each Cytb and nuclear cluster using the best-fit mutation model selected with JMODELTEST or the next-most complex model implemented in the program. Tracer was used to evaluate the combined runs, and LogCombiner performed the joint of log files and tree files. EBSPs were plotted using the GraphfromCSV python script provided with BEAST package v1.6.4. The substitution rates for Cytb and nuclear genes were taken from Melo-Ferreira *et al.* (2012).

2.4.6 Inferring parameters of population divergence and testing for mtDNA introgression

The history of divergence among distinct evolutionary entities uncovered by this work was further explored using the isolation-with-migration framework (Hey 2010). This analysis was applied in three instances, with distinct objectives: i) to clarify the divergence of the Arabian population which was inferred to be a distinct evolutionary unit both for microsatellites and mtDNA analyses; ii) to understand to what extent the

geographical fragmentation between South and North African *Lepus capensis* corresponded to fragmented evolutionary history and divergence; and iii) to estimate parameters of nuclear divergence between the Near East *L. capensis* population and *L. europaeus*, which appeared as discordant in the analyses of mitochondrial and nuclear DNA. The inferred parameters of divergence, using Ima2 (Hey 2010) were effective population sizes, divergence times and migration rates (gene flow). The Isolation-with-migration model (IM) assumes the divergence of an ancestral population into two descendent populations' *t* generations ago and since then gene exchange could have occurred based on nuclear genes (Hey & Nielsen 2004). Under this model, the likelihood of gene exchange among these 3 pairs of populations and relevant demographic parameters for the nuclear data were estimated.

Given that the isolation-with-migration model assumes no recombination within each locus, each dataset was reduced to the largest non-recombining blocks, using IMgc software (Woerner *et al.* 2007), which has been shown to reduce the potential biases in the final estimates (Strasburg & Rieseberg 2010). In total, three independent runs were performed, varying the parameters' upper bound priors and the starting seeds and using the HKY mutation model (Hasegawa *et al.* 1985) (since there are multiple substitutions at single sites), to assess the consistency of the estimates. Moreover, the likelihood ratio test was applied to assess whether migration was significantly different from zero as described by Nielsen & Wakeley (2001). The locus specific mutation rates calculated by Melo-Ferreira *et al.* (2012) were used. Additionally, demographic parameters (divergence time (t), effective population size (Ne) and population migration rate (2NM)) were calculated and converted into biologically meaningful demographic parameters from the highest posterior density of each parameter and using the geometric mean of the locus-specific mutation rates times.

To investigate whether the mito-nuclear discordance inferred for the Near East *Lepus capensis* population was due to mtDNA introgression from *Lepus europaeus*, coalescent simulations were done with the software SIMCOAL V2.1.2 (Laval & Excoffier 2004). MtDNA datasets were simulated under a coalescent model with no gene flow using the parameter of divergence inferred with Ima2 from the nuclear loci. The methodology used is the same used by Melo-Ferreira *et al.* (2012, 2014). The estimates of population size and divergence time from the IM model calculations were used as input to simulate 10,000 sets of Cytb data. The assumed model applied was an ancestral haploid population of size NeA/2 which splits into two descendant populations of sizes Ne1/2 and Ne2/2, t generations ago, with no gene flow occurring after the split between the two descendant populations. The Cytb mutation rate used

was that calculated by Melo-Ferreira *et al.* (2012) for hares. Also, for each of the 10,000 replicates, the minimum pairwise uncorrected *p*-distance between the simulated descendant populations was collected and a distribution of expected these minimum distances was created. The empirical pairwise distance for Cytb (calculated with MEGA v.6 software; http://www.megasoftware.net/mega.php) was considered to reject the null hypothesis of strict lineage sorting model, i.e. to suggest introgression, if lower than the 5th percentile of the simulated minimum distances.

3. Results

3.1 Laboratory analysis of samples from different sources

The final set of samples used in this study had different origins, which corresponded to distinct rates of success in their analyses (final dataset represented in fig. 5 with samples for which at least one marker was analysed and included in this study). For mtDNA, PCR amplification and sequencing of road kill tissue samples, was generally successful, working properly for 155 out of 176 samples. From the 6 samples of *L. saxatilis*, only one amplified for Cytb. For the faeces and bones, Cytb amplification and sequencing was successful for all five samples. On the contrary, museum samples showed a low rate of amplification success, only a small part being correctly amplified (26 from the initial 79). Additionally, mtDNA sequencing showed evidences of DNA degradation in samples obtained from museum specimens: in a total of ten samples, double peaks due to deamination in the sequence electropherograms were observed. These ten sequences were excluded. In contrast, no double peaks were seen in sequences obtained from fresh tissue samples.

Microsatellite genotyping also failed for most of the museum samples, likely due to DNA degradation: from the 79 available only 29 were successfully genotyped. For the road kill samples, the successful microsatellite genotyping was 148 samples out of 176. Only one sample of *Lepus saxatilis* was correctly genotyped for the microsatellite loci used in this work. Neither bone nor faeces samples were correctly genotyped.

A subset of the samples was selected for amplification and sequencing of nuclear DNA loci and, out of the 80 samples, the maximum amplification rate includes 75 for the Transferrin gene and the minimum concerned 50 samples for OXA1L fragment. The same sample of *L. saxatilis* was successfully amplified for four of the five nuclear fragments (see appendix 1 for detailed description of samples' laboratory success).

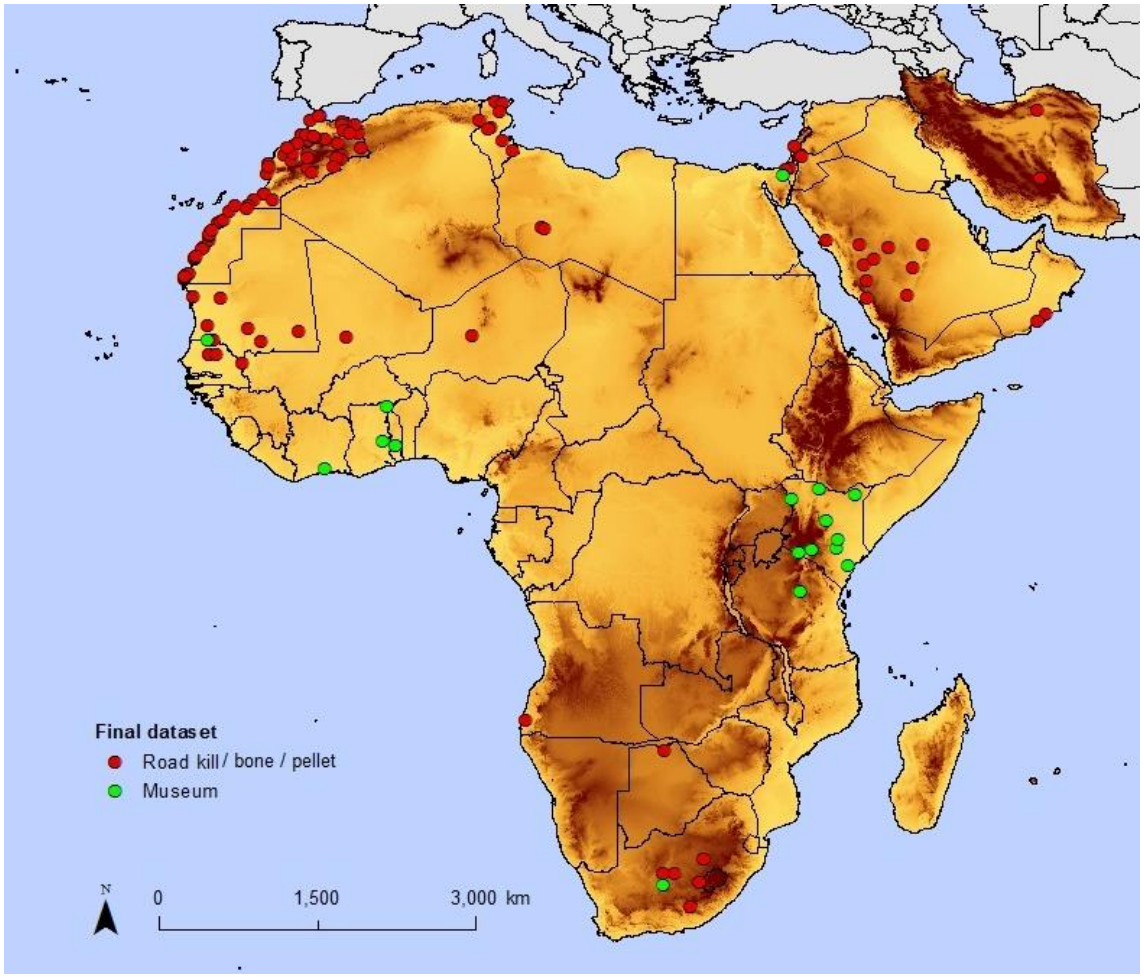


Fig. 5– Samples integrating the final dataset (faeces and bone samples included in the road kill group).

3.2 Microsatellite data

From the 35 microsatellites initially tested, 18 polymorphic markers were used for genotyping. Groups (which could correspond to populations or eventually species) were not assumed *a priori*. For mere organization of the data, individuals were divided according to their country (in total 23) where the number of samples was asymmetric between groups (from 1 to 71), though this information was not used for any inference. There was no evidence for null alleles. When analysing all 176 samples as a population, from the 18 loci, significant deviations from HWE were found in all loci except LCP_26, LCP_34 and LCP_ 38 which were found in equilibrium. Significant LD was found in 15 out of 153 pairs of loci. The high level of disequilibrium can be due to the fact of possible population structure, and a partition by geographic regions was taken to check the consistency of the deviations. The dataset was divided into 4 distinct populations: North Africa, South Africa, Southeast Africa and Near East. HWE results show that for the North Africa population, all loci are in disequilibrium except loci 10,

16, 26 and 38. For LD, results show six pairs of loci in disequilibrium for the Northern population, and only three from the six are the same loci in disequilibrium as the dataset as one large population. Significant deviations from HWE were also tested for each country and from the 18 loci, disequilibrium were found at locus 3 in Morocco, Western Sahara and Kenya; locus 37 in Kenya and Morocco; locus 6, 8 and 18 in Morocco, while the others were found to be in equilibrium. Only 5 out of the 153 pairs of loci showed a significant LD (again in Morocco and Western Sahara, including the same loci in HW disequilibrium). Still, before excluding the loci immediately and in order to test its potential influence, two independent runs were done with STRUCTURE (each with three replicates with the same parameter conditions) to detect if including or excluding the marker in apparent disequilibria induced changes in the results. Both dataset runs (with and without loci with significantly deviated from HWE) found to produce the same results (only slight differences in the proportion of admixture in a few admixed individuals was detected. The same happened with the run assuming independent allele frequencies: clusters were identical for the correlated and independent allele frequency models. Since the disequilibria observed with some of the loci were not associated more frequently with particular loci or pair of loci, this may be due to regional effects. Consequently, we retained all loci for subsequent analysis. All 18 loci were highly polymorphic where the number of alleles per locus varied between 8 (Lcp_37) and 28 (Lcp_8) (see table 1)**.**

Table 1– Number of alleles sampled per locus.

Loci	N° of alleles sampled
Lcp_2	21
Lcp_3	24
Lcp_4	11
Lcp_6	11
Lcp_8	28
Lcp_9	13
Lcp_10	17
Lcp_12	16
Lcp_16	12
Lcp_17	12
Lcp_18	15
Lcp_23	11
Lcp_26	12
Lcp_28	11
Lcp_33	18
Lcp_34	13
Lcp_37	8
Lcp_38	10

The hierarchical analysis of structure identified 22 out of the 176 specimens as potentially admixes (using a 0.80 threshold criteria). The final dataset of STRUCTURE was reduced from 176 to 154 individuals due to the maximum inferred ancestry below 0.8 being identified which did not pass through the next rounds. STRUCTURE analyses identified hierarchical population division (Fig. 6), in which on the first round, the highest likelihood model (K = 2) identified mostly a Northwestern population (with two samples of Kenya, one from Tanzania and other from South Africa, corresponding to the red samples) and the other cluster with the remnant populations (in green). The second round of STRUCTURE did not split the Northwestern cluster into sub-clusters (left plot), but the other group was divided into four sub-clusters (K=4; right plot) – Kenya, Near East, Arabia and South Africa. Further rounds of the hierarchical STRUCTURE analysis did not found additional structure in any of the groups (four plots on the right of Fig. 6). Microsatellites detected presence of gene flow between different individuals, since 22 individuals were recognized as admix where it was not possible to assign them to a single population. In the first round, regions with only one (Sudan, Angola, Egypt and Niger) or two (Algeria) samples were admixed in K=2, as well as the only *L. saxatilis* sample. Moreover, regions with considerable sampling were found to have admixed individuals too: Morocco (four samples), Mauritania (three samples), S. Africa (two samples), Western Sahara and Saudi Arabia (one sample each). In the second round, two samples from Israel showed to be admixed with Arabia, one sample from Saudi Arabia showed admixture level with Near East population and one sample from Tanzania showed to be admixed with Near East and Kenya populations. The final hierarchical STRUCTURE clusters are represented in fig. 7. Note that the colors of each group indicated in fig. 7 will be from now on used to depict the evolutionary units in the following figures.

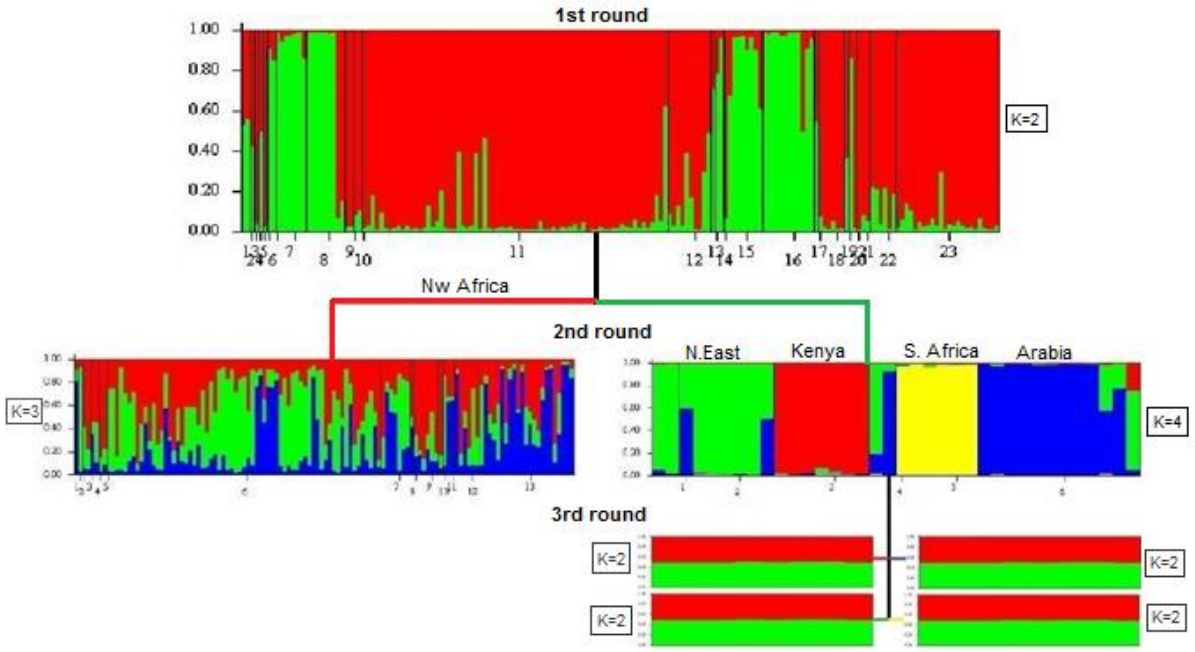


Fig. 6– STRUCTURE bar plots showing the assignment of the individuals to different hierarchical levels of population structuring according to the optimal number of clusters. The analysis was based on 18 microsatellite loci and for each plot, clusters are represented by a different color.

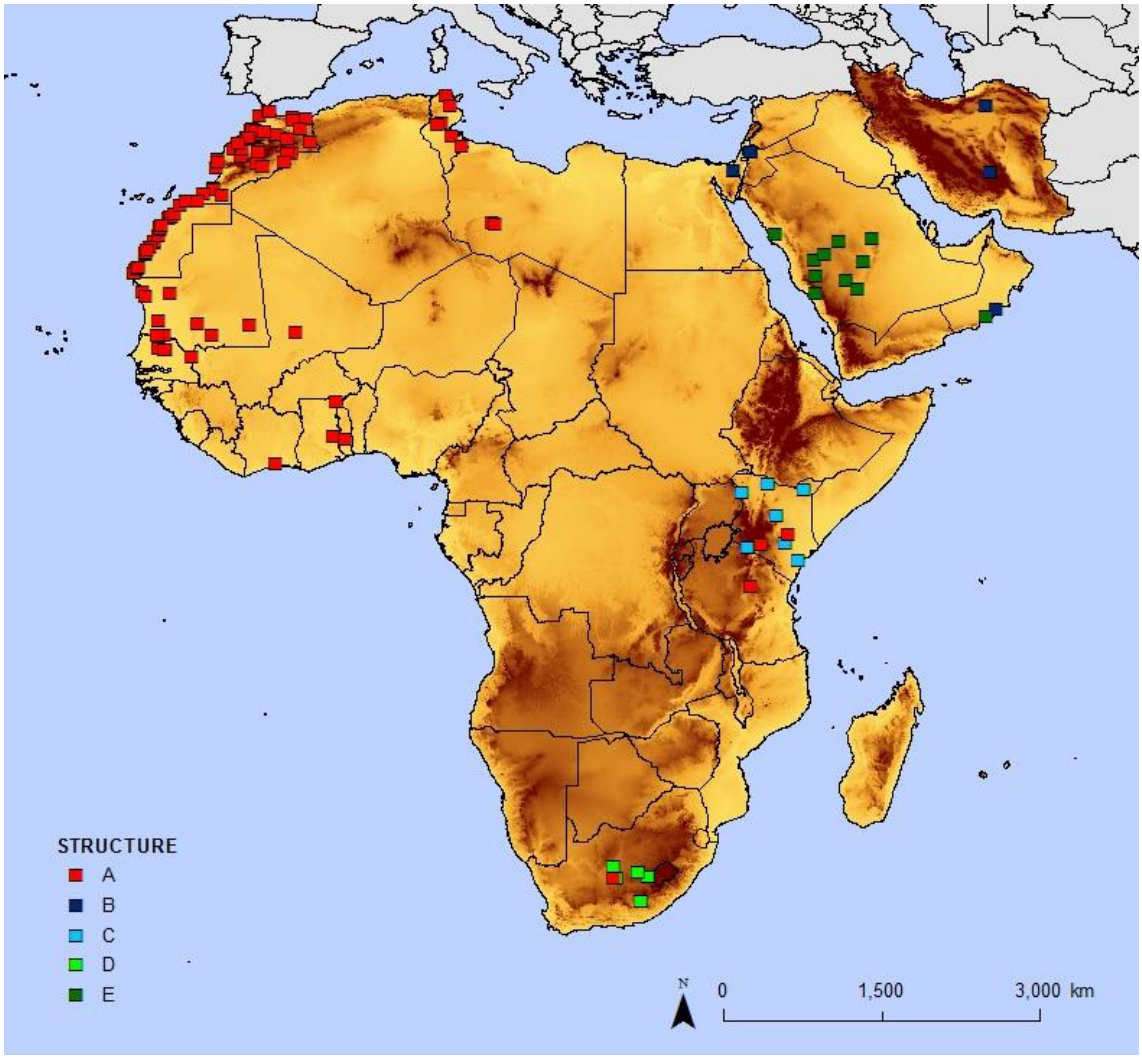


Fig. 7– STRUCTURE hierarchical results (five distinct clades) represented in a map. Each square represents an individual, and each lineage is represented by a different color (Near East – dark blue; Arabia – dark green; Kenya – light blue; South Africa- light green; the big NW African cluster – red).

PCA based on subsets of individuals corresponding to the aforementioned groups of STRUCTURE produced plots which tended to resemble the same groupings. The first three principal coordinates that resulted from PCA accounted for 13% of the total variation (third axe not shown), and allowed identifying the five clusters inferred in STRUCTURE (fig. 8).

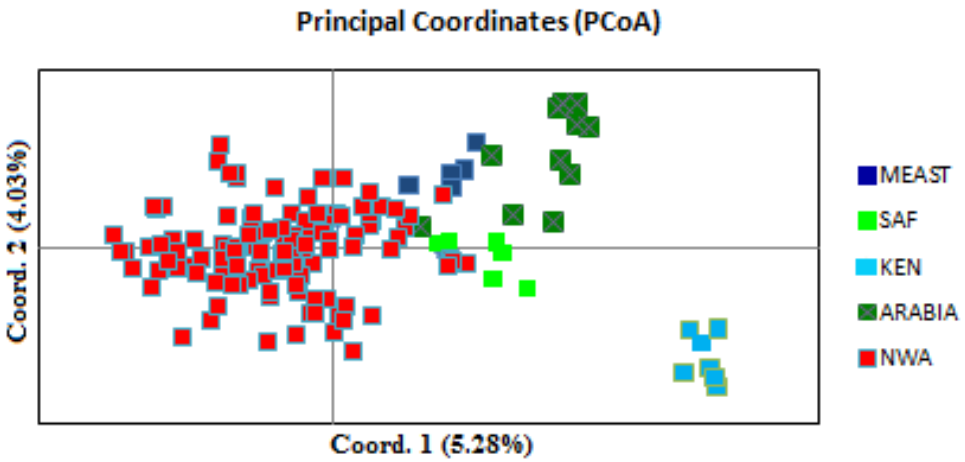


Fig. 8 – PCA output graph based on individual-by-individual genetic distances. Each square represents an individual, and each lineage is represented by a different color. The two first axes are shown. NEAST-Near East; SAF-South Africa; ARABIA-Arabia; NWA-Norwest Africa.

Measures of pairwise F_{ST} across populations showed that all the five grouping of STRUCTURE results have a significant genetic differentiation at the highest level of genetic structure (Table 2). The values ranged between 0.1 (Near East and NW Africa) and 0.29 (South Africa and Kenya).

Table 2 - Pairwise F_{ST} values of the five identified genetic clusters.

Pop 1	Pop 2	F_{ST}
S. Africa	Kenya	0.29
Kenya	Arabia	0.26
Near East	Kenya	0.26
S. Africa	Arabia	0.20
Kenya	NW Africa	0.19
Near East	S. Africa	0.16
S. Africa	NW Africa	0.13
Near East	Arabia	0.11
Arabia	NW Africa	0.10
Near East	NW Africa	0.10

Clustering results of the BAPS software (fig. 9) showed a substructure that was partly consistent with the groups found by STRUCTURE although with notorious differences. First, the South African group was included in the big NW African clade. Second, from the other four STRUCTURE groups, two were highly consistent (NW Africa and Near East) contrary to the other two clusters (Kenya and Arabia) which did not completely match the first results. In fact, BAPS structured almost at the individual level.

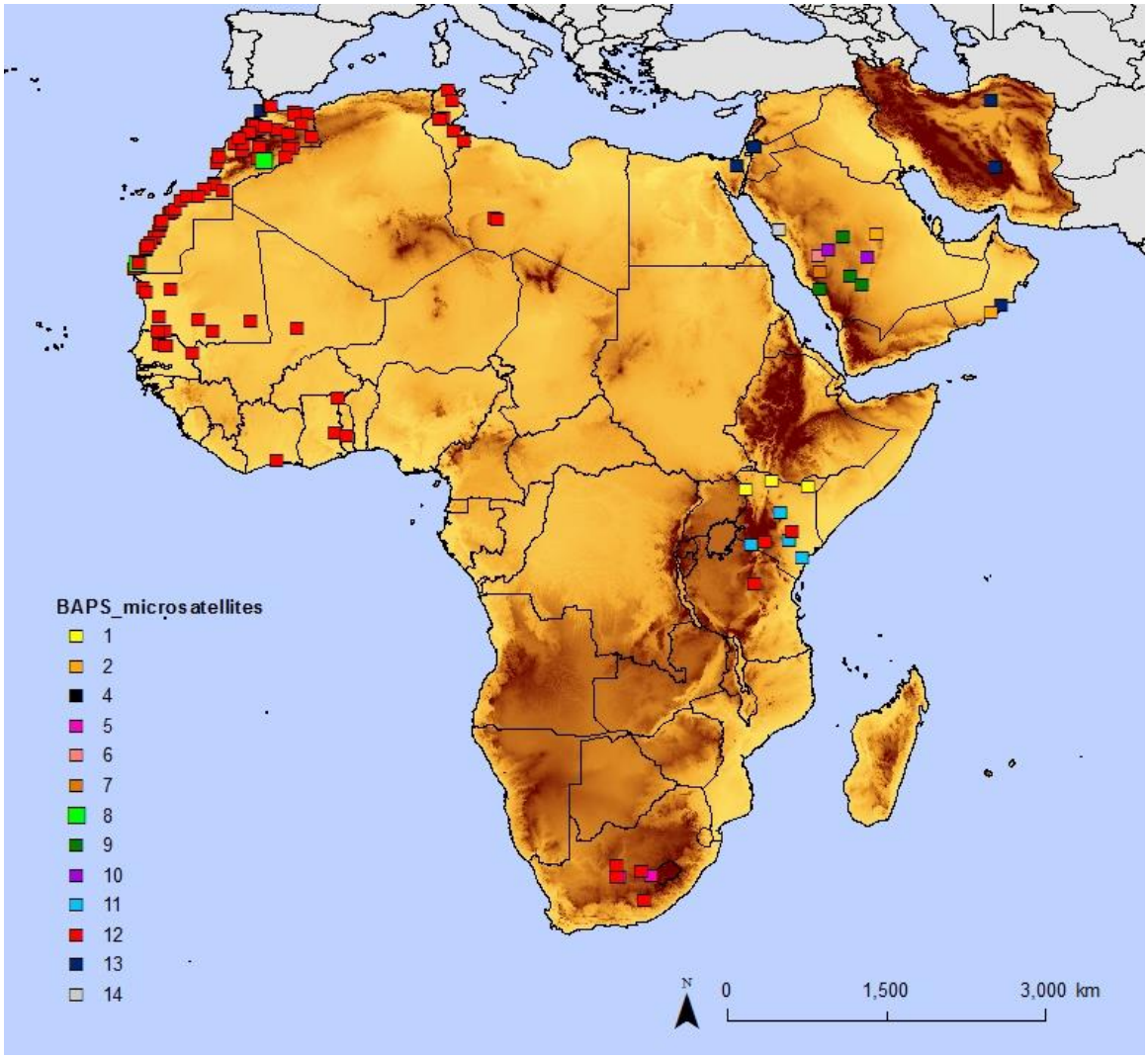


Fig. 9 – Population structure results from BAPS software. Each square represents an individual, and each lineage is represented by a different color.

Given the difficulty in defining populations a priori, we tested again HWE, but this time for the 5 distinct microsatellite groups a posteriori. The largest group (Northwest Africa) was in HW disequilibrium for several loci, as well as with significant LD, while the other four populations were in equilibrium.

3.3 DNA sequence data

The length of the final DNA sequence alignments is described in the table 3. In order to be possible to compare at the individual level, the final Cytb dataset including museum samples (the shorter fragment) used for the mtDNA analysis, was reduced to the most similar as possible to the final dataset of microsatellite genotyped samples. For the nuclear genes, note that South African population did not amplify for two of the five genes (HPX and OXA1L) and sequences of Chinese *L. capensis* from GenBank were only available for two of the five genes (TF and SPTBN1).

Table 3 - Description of the final sequence datasets.

Locus	Number of individuals	Final alignment (including GenBank sequences)	Fragment size (base pairs)	Gene position	Annealing Temp. (ºC)
Mitochondrial					
Lcp_Cytb (Cytochrome b)	156	205	417	Exon	58
Lcp_Cytb2 (Cytochrome b)	151	196	223	Exon	57
Nuclear					
DARC (Duffy blood group, chemokine receptor)	67	93	880	Exon	58
HPX (Hemopexin)	70	96	789	Intron/Exon	60
OXA1L (Oxidase Assembly 1-Like)	50	78	653	Intron/Exon	58
SPTBN1(Spectrin, beta, non-erythrocytic 1)	65	97	633	Intron	58
TF (Transferrin)	75	133	419	Intron/Exon	60

3.3.1 Phylogenetic and phylogeographic patterns

Network results show that Cytb dataset with museum samples (see fig. 10; appendix 6 for dataset without these samples) suggests that nine evolutionary groups exist with a clear geographical pattern. Representation of the Network haplogroups on the African map can be seen in fig. 11 (with museum samples; appendix 7 without museum samples). Based on the microsatellite results, from the five microsatellite groups, four are highly congruent with the Network results (Near East, Arabia, Kenya and South Africa – represented with the same colors of the microsatellites groups), although the big red fifth one in Northwest Africa is now subdivided into five distinct

clades (Northwest - North; central; South; Southeast and Kenya/Tanzania). The three most Northwestern groups (North, Central and South) have a star-like configuration with a central haplotype that was spread and the additional haplotypes were mostly connected to the central one by just one mutation. Contrary to the microsatellite results, no haplotypes are shared between very distant geographic regions within each clade.

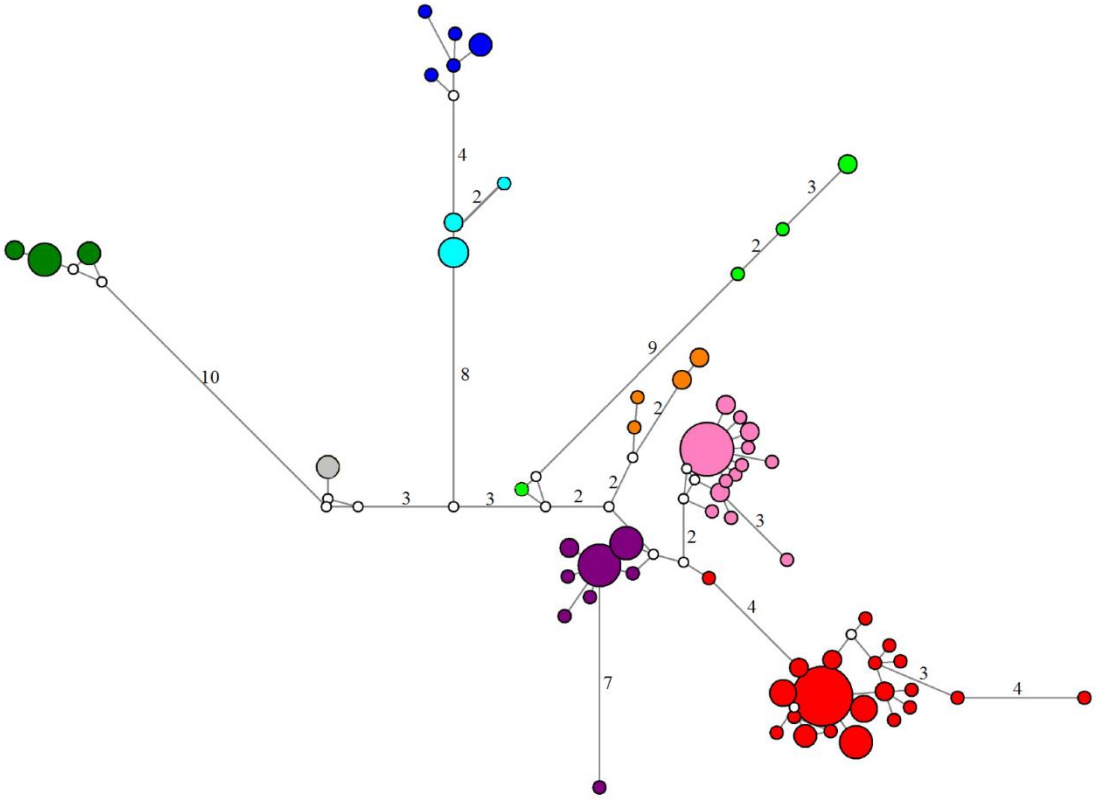


Fig. 10 - Mitochondrial haplotype network. Each circle represents a different haplotype, proportional to the number of shared individuals per haplotype. Branch lengths are proportional to genetic distance between haplogroups. The number on the branches represents the number of mutated positions. Each lineage referred in the main text is highlighted with a different color. White dots correspond to the median vectors. (Near East- Dark blue; Arabia – dark green; Kenya – light blue; South Africa – light green; Northwest Africa is divided into 5 distinct haplogroups: red- the most northern one; pink – central; purple – South; orange – Southeast and grey – Kenya/Tanzania).

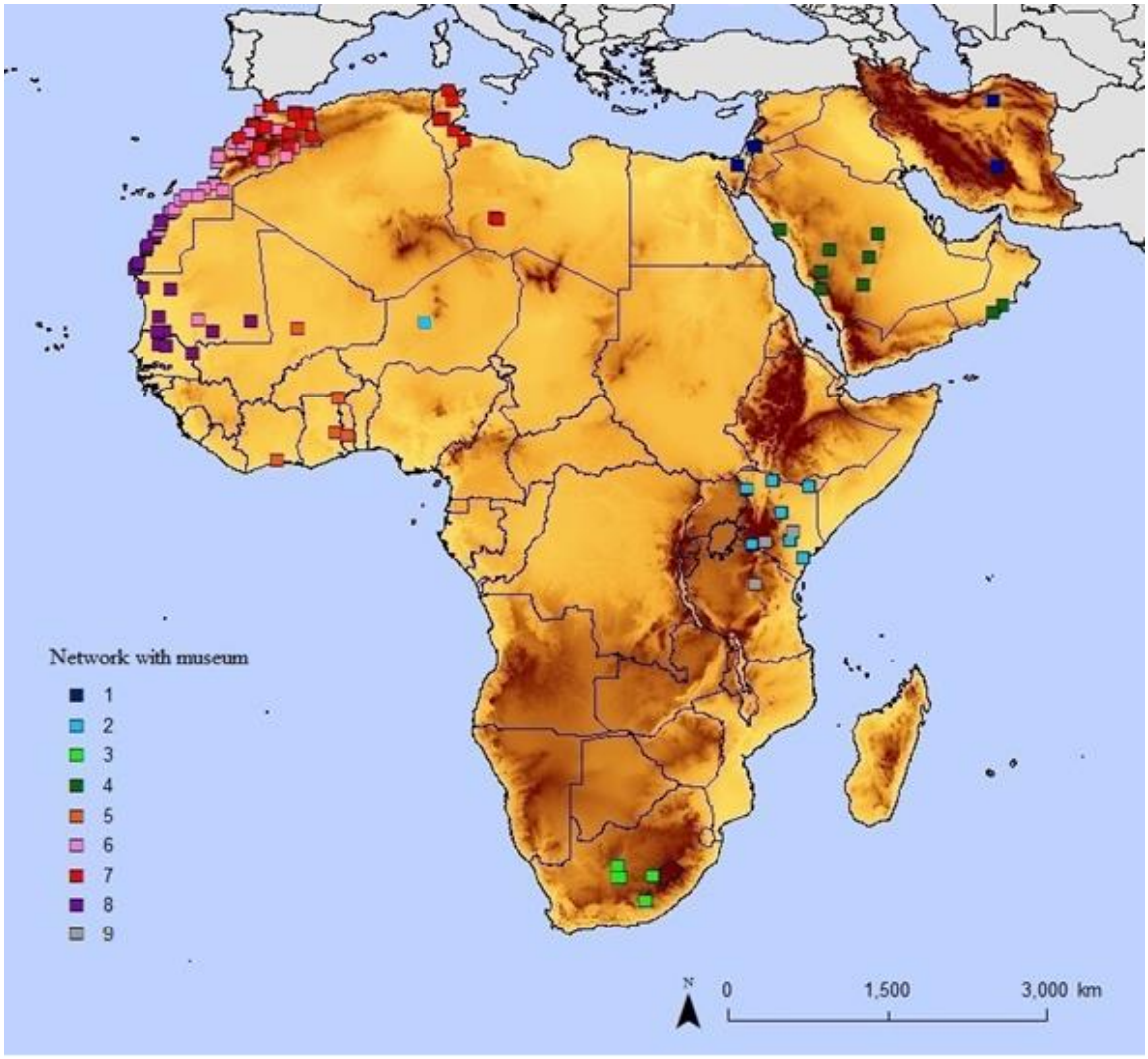


Fig. 11 - Distribution of the African mtDNA haplogroups. Each square represents a different individual.

The phylogenetic tree allows understanding better the evolutionary relationships among different clades. The combined results of the BSP Bayesian inference and Maximum Likelihood analysis of the data in an unrooted tree are represented in the figure 12, including the museum samples (see appendices 8 and 9 for full results; appendix 10 for combined tree without museum samples). Low posterior probabilities for some nodes were recovered, and only nodes with a support higher than 50% are shown in figure 12.

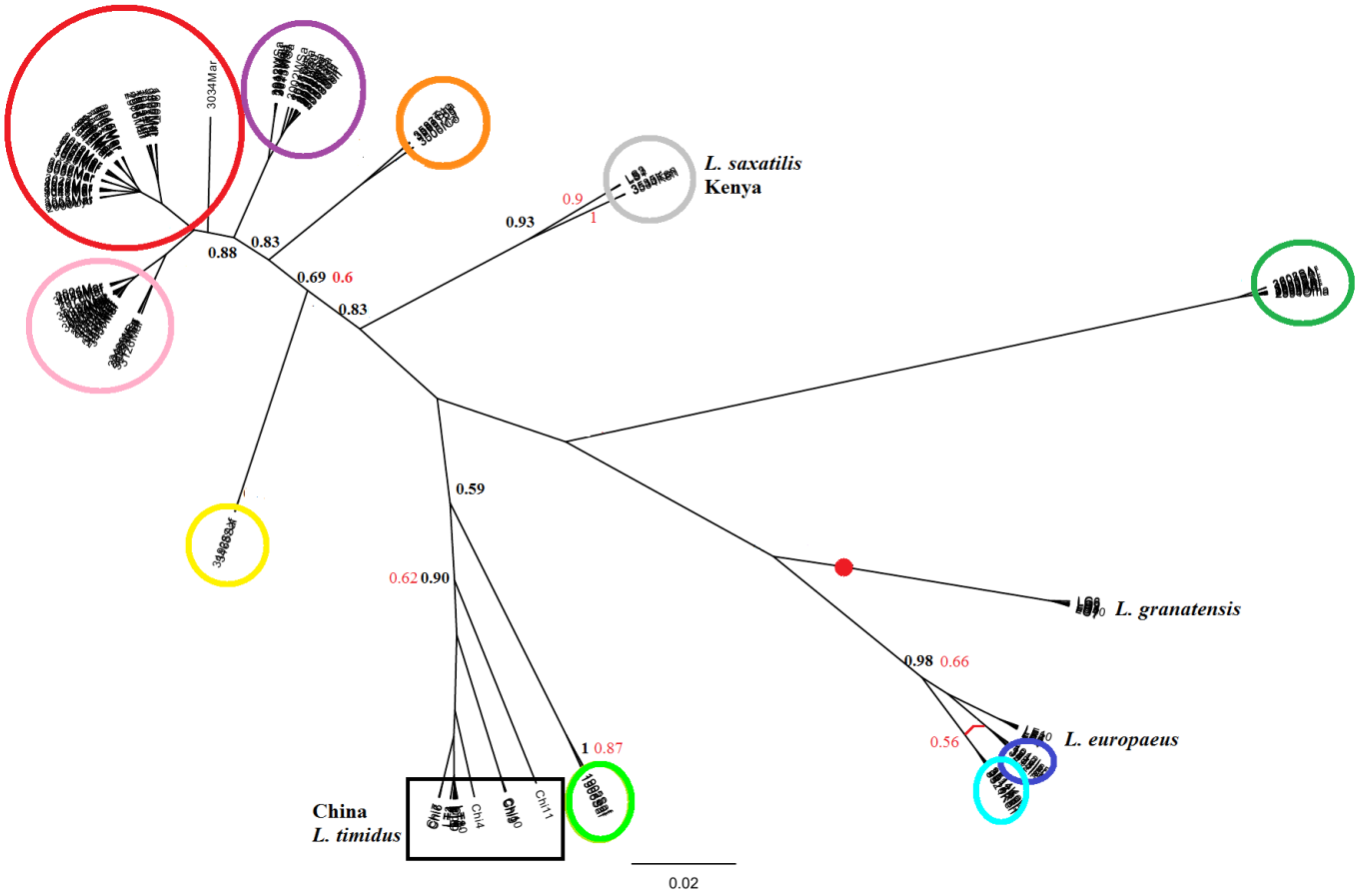


Fig. 12 - Maximum likelihood and Bayesian inference phylogenies represented in an unrooted tree, estimated for mtDNA; bootstrap and posterior probability values above 0.50 are represented in front of each node (red and black respectively). The red dot corresponds to where the outgroup, *Oryctolagus cuniculus* would appear. Lineages: 1) *L. capensis* from China and *L. timidus;* 2) *L. europaeus;* 3) South Africa (light green); 4) samples from Kenya and the sample from Niger (light blue); 5) Near East (dark blue); 6) *L. granatensis*; 7) Arabia and Oman (dark green); four groups from Northwest Africa: 8) North - (red), 9) Central (pink), 10) South (purple) and 11) Southeast (orange); 12) *L. saxatilis* from GenBank and 3samples from Kenya/Tanzania (Grey); 13) one sample of *L. saxatilis* and a sample from Angola (yellow).

The results corresponding to the smaller sequenced fragment (with museum specimens) revealed 13 main clades represented in the map in the figure 13. The clades are designated by 1) *L. capensis* from China and *L. timidus;* 2) *L. europaeus;* 3) South Africa; 4) samples from Kenya and the sample from Niger; 5) Near East (Israel and Iran); 6) *L. granatensis*; 7) Arabia and Oman; four groups from Northwest Africa: 8) North - which includes samples from Lybia to Maghreb (Morocco), 9) Central - that includes samples from Atlantic Costal Sahara (Western Sahara) and from the mountain Tagant (Mauritania), 10) South - that includes samples from Atlantic Costal Sahara, Southwestern Sahara (Senegal), and mountains Assaba and Afollé (Mauritania) and 11) Southeast - includes a sample from Central Sahel (Mali); 12) *L. saxatilis from* GenBank and samples from Kenya; 13) one sample of *Lepus saxatilis* and a sample from Angola. Without museum samples (appendix 11), there are only two small changes in the results: 1) without the museum samples from Togo, Ghana and I. Coast, the sample from Mali represents alone a distinct group; 2) without the museum

samples from Kenya, the Niger samples belongs to the Near East group. The Bayesian inference with the Yule model corroborates all the major clades of the BSP phylogenetic tree (appendix 12).

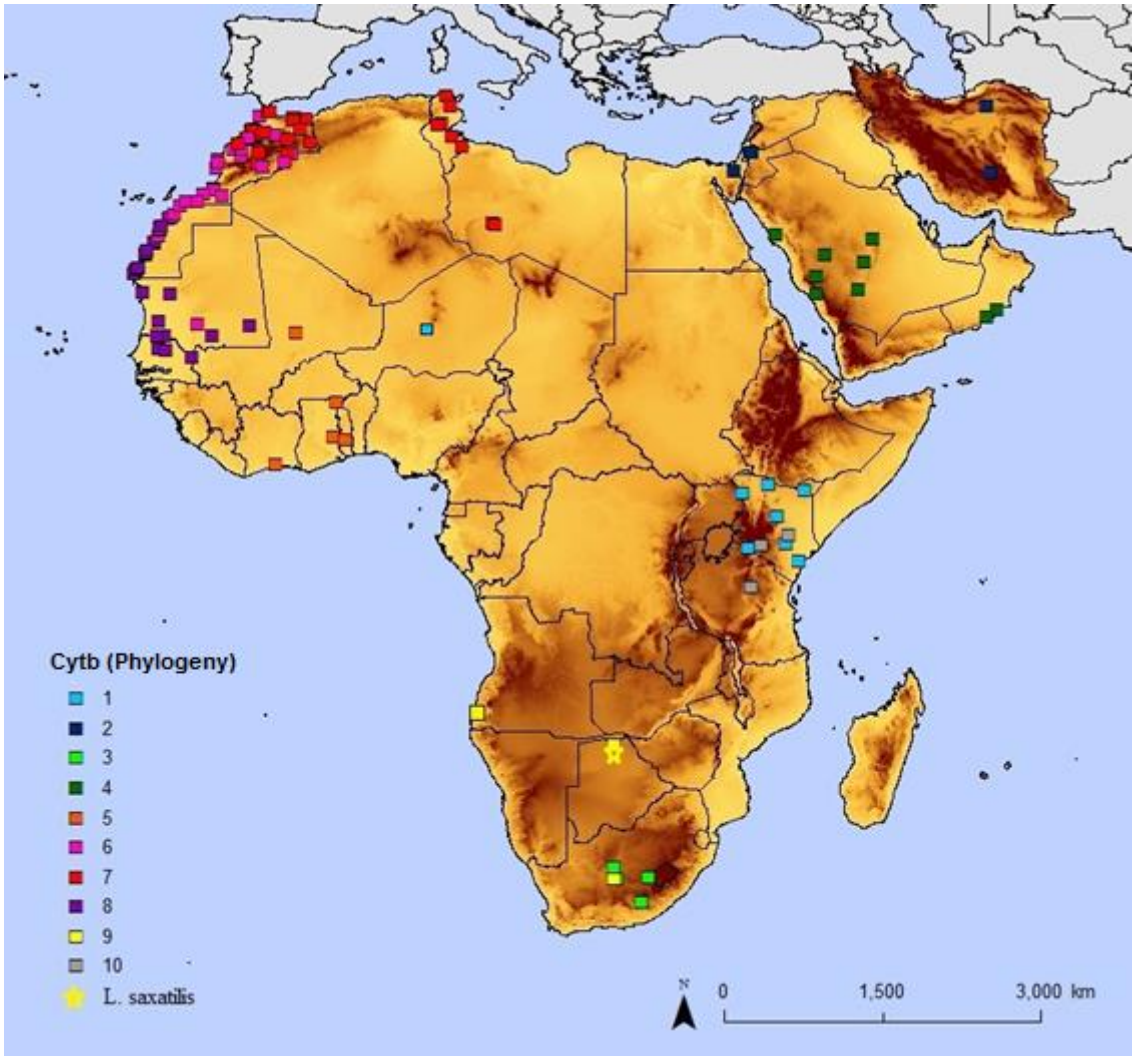


Fig. 13 - Distribution of lineages, based on the phylogenetic results. Each lineage referred in the main text is highlighted with a different color. GenBank individuals are not represented (e.g. *L. saxatilis* individuals which belong to the grey clade are not represented). The *L. saxatilis* sample from our dataset is represented with a star shape in the map.

The estimates of the phylogeny of the mtDNA showed that sequences sampled in different species were often closely related or even found in the same clade: clades of Near East and Kenya are closely related to *L. europaeus*; samples from Kenya (grey) cluster with GenBank *L. saxatilis* and Angola sample and one S. African are closely related to *L. saxatilis* sample; sequences downloaded from GenBank classified as *L. capensis* (from China) were placed within or closely related to the *L. timidus* clade. The most highly supported clade was recovered for South Africa, although bootstrap values and Bayesian posterior probabilities (BPP) were in general not high.

However, there is a good coherence in the phylogeographical groupings. The inferred mtDNA phylogeny confirms the major groups inferred with the microssatellite loci, but shows further subdivision. Estimates of divergence times were 7.4 million years for the *L. europaeus*/Near East/Kenyan *L. capensis* clade and Africa (BPP = 0.19), 5.6 million years for Arabia and Africa (BPP = 0.18), and 4.5 million years for North and South African populations (BPP = 0.25). But due to the low posterior probability these estimates can lead to erroneous conclusions. By applying BAPS to the mtDNA sequence dataset (fig.14), the principal African lineages shown in figure 12 are corroborated, with the exception of lineages 4), 11) and 13) (represented in the figure 13 by grey, orange and yellow respectively) which assembled together in just one cluster. Also, Kenya clade and Near East clustered together as well.

Genetic distances of the mtDNA BI lineages, measured by pairwise Dxy values (table 4) were calculated. All clades showed values ranging from 0.03 (between Kenya and N. East) to 0.1 (between NW Africa - North and Arabia). Additionally estimates of genetic variability of *L. capensis* samples including museum samples were high: the total haplotype diversity (*Hd*) was 0.958, the variable positions (*S*) were in total 61, the nucleotide diversity (π) was 0.065 and 60 different haplotypes were found. After calculating the haplotype diversity per lineage (for populations with a minimum of five samples) the range values were from 0.476 (Kenya clade) to 0.971 (NW Africa – Southeast lineage). Including *L. capensis* from China plus other hare species lineages, the results were for *L. europaeus* = 0.464; China/*L. timidus* = 0.914 and *L. granatensis* = 0.378. This means mtDNA populations along NW Africa have a high genetic diversity, even compared with other species. The *Hd* was not calculated for *L. saxatilis* since the dataset only included four samples.

Table 4 – Pairwise Dxy values between the principal phylogenetic lineages (NwA C/N/S/SE correspond to the four lineages of NW Africa – Center, North, South and Southeast respectively).

	Near East	**NwA_N**	**NwA_C**	**Arabia**	**NWA_S**	**NwA_SE**	**S. Africa**	***L. europaeus***	**China/ *L. timidus***	***L. granatensis***	**Kenya**	***L. saxatilis***
Near East	*											
NwA_N	0.097	*										
NwA_C	0.092	0.041	*									
Arabia	0.078	0.110	0.092	*								
NWA_S	0.088	0.043	0.028	0.095	*							
NwA_SE	0.085	0.058	0.050	0.093	0.037	*						
S. Africa	0.072	0.086	0.081	0.087	0.073	0.065	*					
L. europaeus	0.035	0.093	0.091	0.075	0.079	0.075	0.048	*				
China/ *L. timidus*	0.102	0.104	0.091	0.090	0.095	0.097	0.087	0.090	*			
L. granatensis	0.070	0.107	0.095	0.068	0.100	0.083	0.073	0.061	0.091	*		
Kenya	0.030	0.085	0.091	0.075	0.080	0.076	0.064	0.019	0.099	0.061	*	
L. saxatilis	0.088	0.087	0.072	0.059	0.065	0.068	0.081	0.081	0.094	0.073	0.071	*

The nuclear gene trees (appendices 13 and 14) show low supporting values for the nodes, where no obvious population differentiation can be identified, contrarily to what it is observed with mtDNA and microsatellites results. The same was observed in an independent investigation with the same genes, but with other hare species (Melo-Ferreira *et al.* 2012).

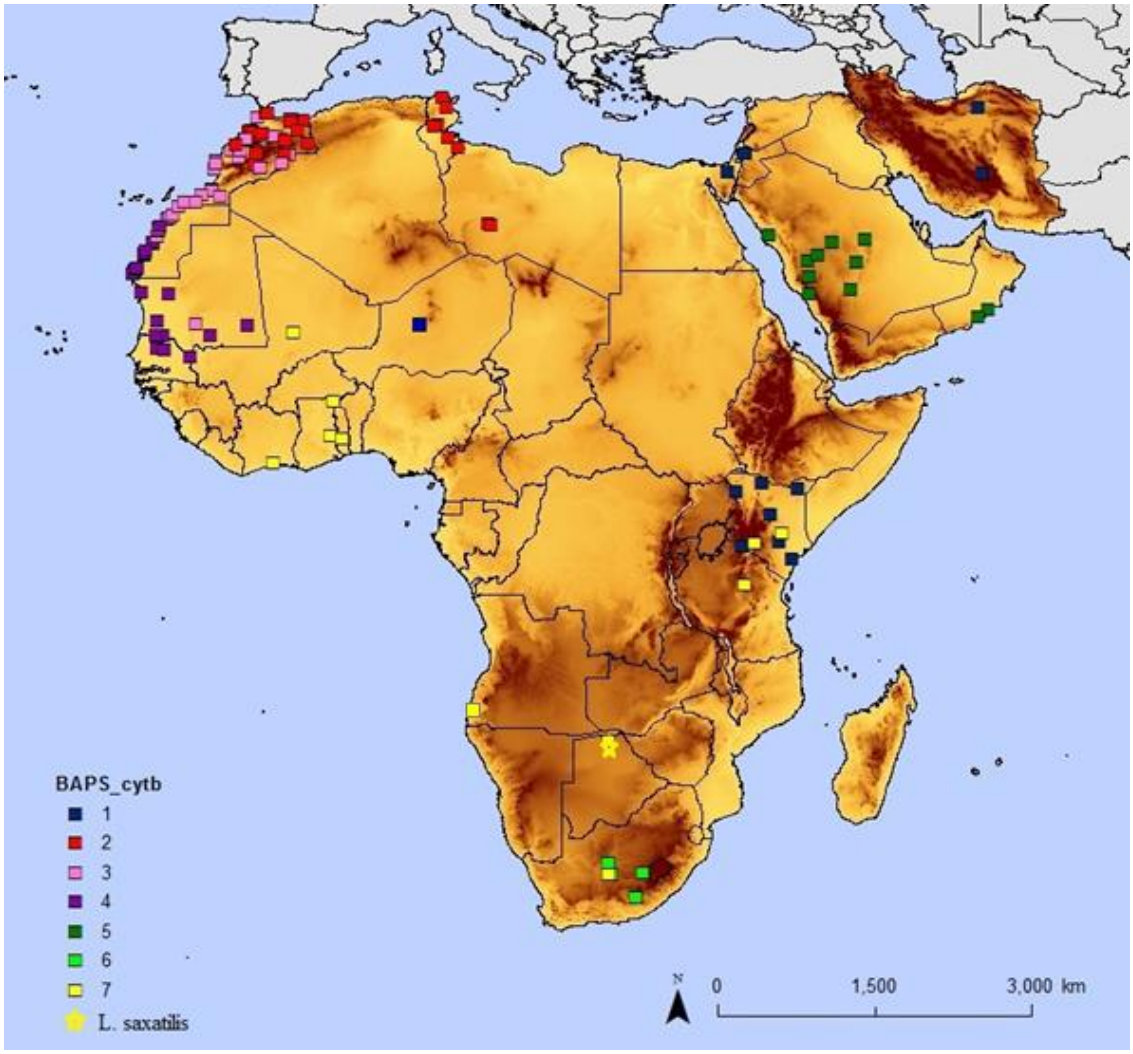


Fig. 14 - Distribution of lineages, based on the BAPS results. Each clade is highlighted with a different color. The *L. saxatilis* sample from our dataset is represented with a star shape in the map.

*Beast is a Bayesian method that takes use of multi-locus data, to infer about the relationships between different populations/species (Heled & Drummond 2010). In this study, four species tree were reconstructed based on four distinct datasets (see Material and Methods). In general, all presented congruent results (see fig. 15 and 16). When analysing the African populations' dataset (fig. 15), two major clades appear: North and South African hares. The species trees which included all populations/species and genes (fig. 16) also showed two main clades, but in this case splitting African from the non-African lineages (except for the fig.16.1, the tree with only three from the 5 nuclear loci, which splits also NW Africa from the South African populations). When including South Africa and *L. saxatilis* (fig. 15, 16.1 and 16.3), species trees showed that these two population were recovered as sister taxa (BPP = 0.63, BPP = 0.65 and BPP = 0.74 respectively). Similarly, is common in all four species

trees that Saudi Arabia and Near East are also closely related (BPP = 1). In this case, with the dataset which included other hare species, *L. europaeus* was not closely related to Near East, contrary to the mtDNA inferences. The same happened with *L. capensis* from China, which contrary to the mtDNA results, are not closely related with *L. timidus.* We attempted to date the divergence of major clades using the mutation rate calculated for the reference nuclear gene (in this case SPTBN1) by Melo-Ferreira *et al*. (2012) according to the *Lepus-Oryctolagus* divergence. Note that the low support found for most of the nodes advice extreme caution when interpreting these estimates. The split between NW Africa and S. Africa (BPP = 0.43) is estimated to be 932,098 years ago (ya) (490,000 - 1,300,000 ya HPD 95% confidence intervals), and NW Africa and Saudi Arabia (BPP = 0.56) is 883,040 ya (490,000 - 1,400,000 ya HPD 95% confidence intervals). The timing of divergence between the South African *L. capensis* population and *L. saxatilis* (BPP = 0.74*)* was estimated in 735,867 (340,000 – 1,130,000 ya HPD 95% confidence intervals) and around 686,809 ya (440,000 – 989,000 ya HPD 95% confidence intervals) *L. europaeus* and N. East have split.

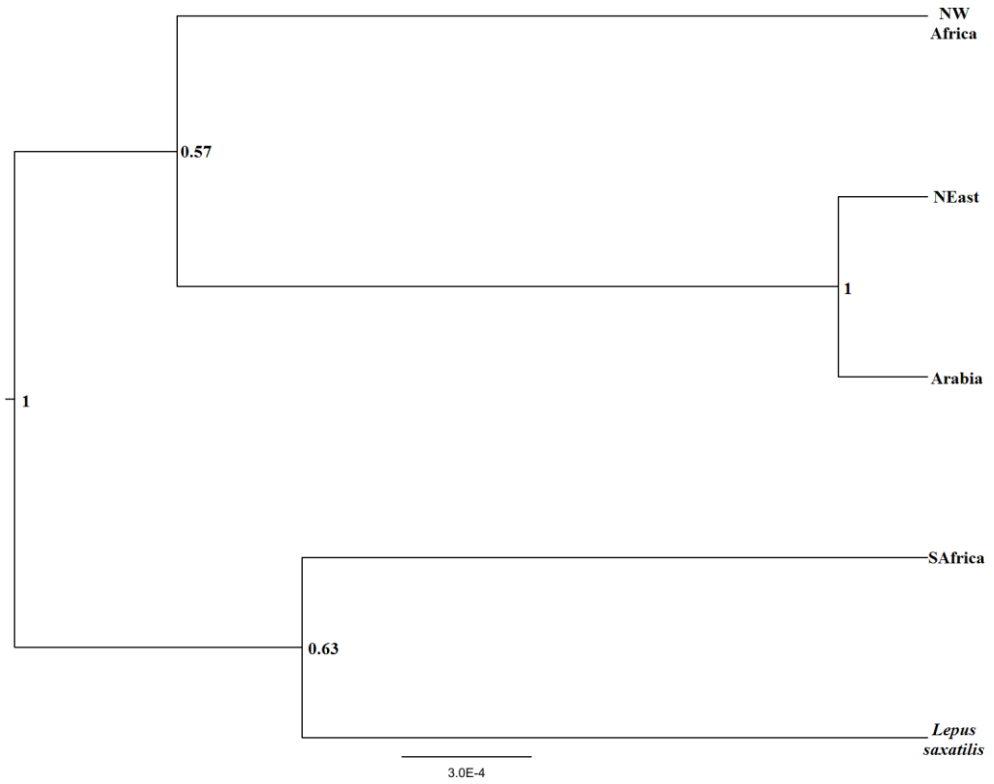
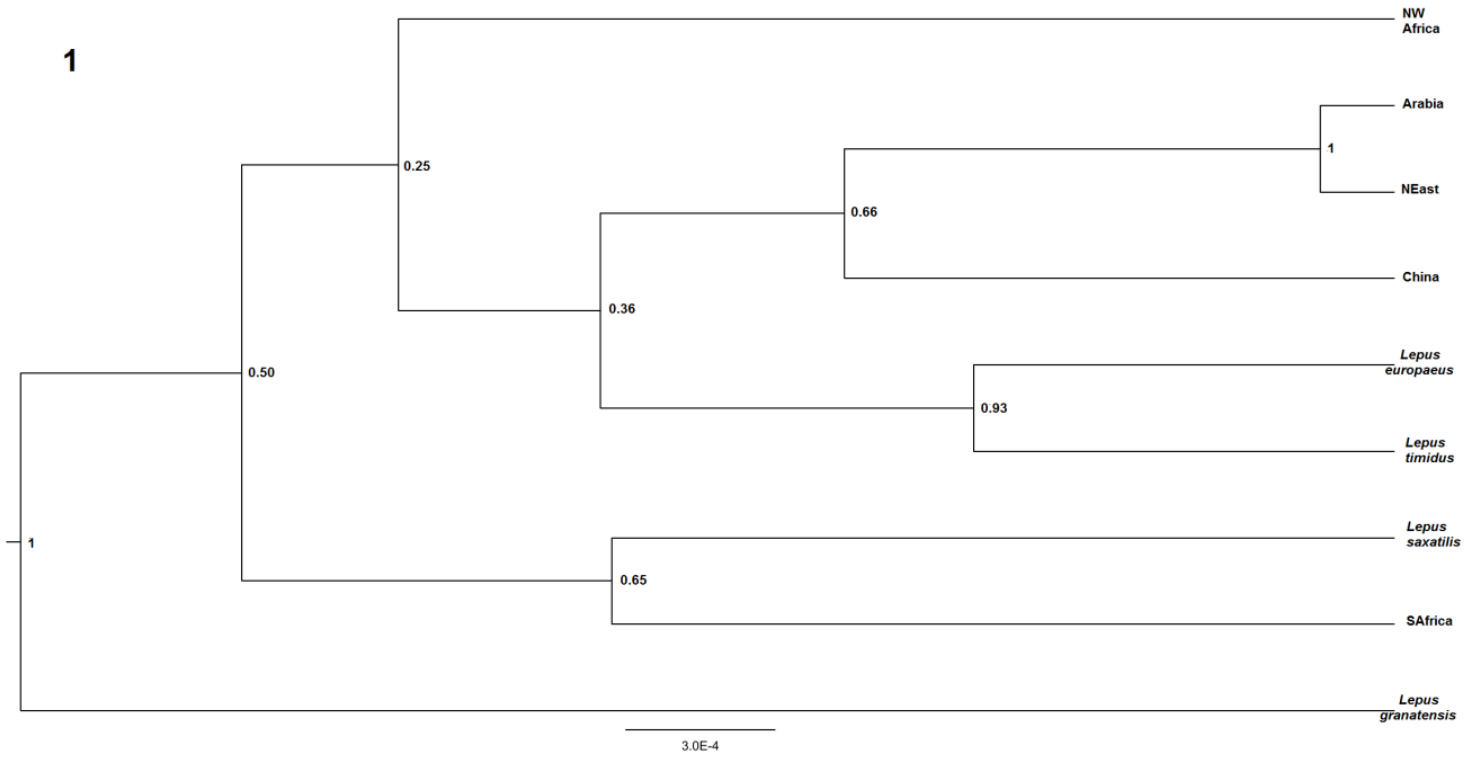
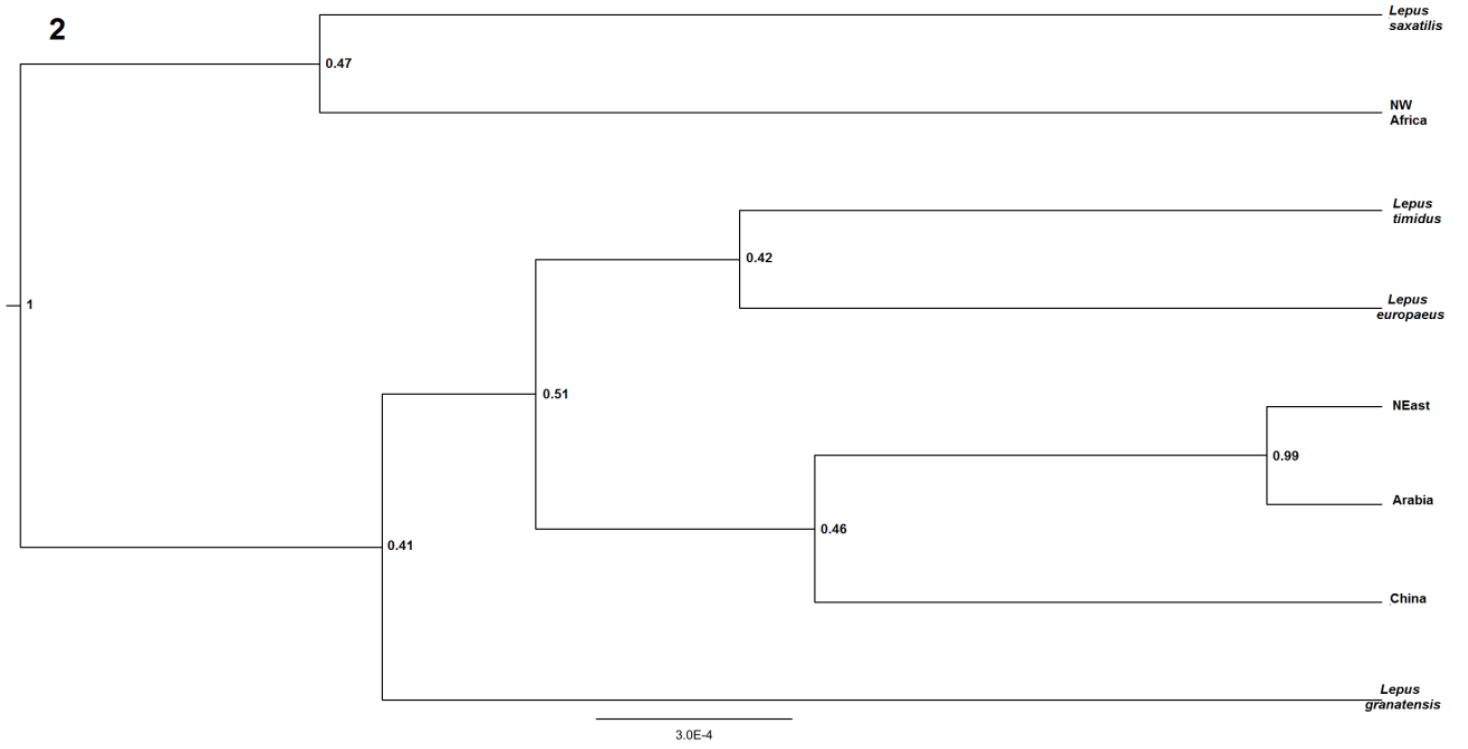


Fig. 15 - Nuclear DNA species tree for the five nuclear loci (the BPP of each clade is shown in each node) for African populations and *Lepus saxatilis*.


1
NW Africa
Arabia
1
NEast
China
Lepus europaeus
Lepus timidus
Lepus saxatilis
SAfrica
Lepus granatensis
0.25
0.66
0.36
0.50
0.93
1
0.65
3.0E-4


2
Lepus saxatilis
0.47
NW Africa
Lepus timidus
0.42
Lepus europaeus
1
0.51
NEast
0.99
Arabia
0.46
0.41
China
Lepus granatensis
3.0E-4

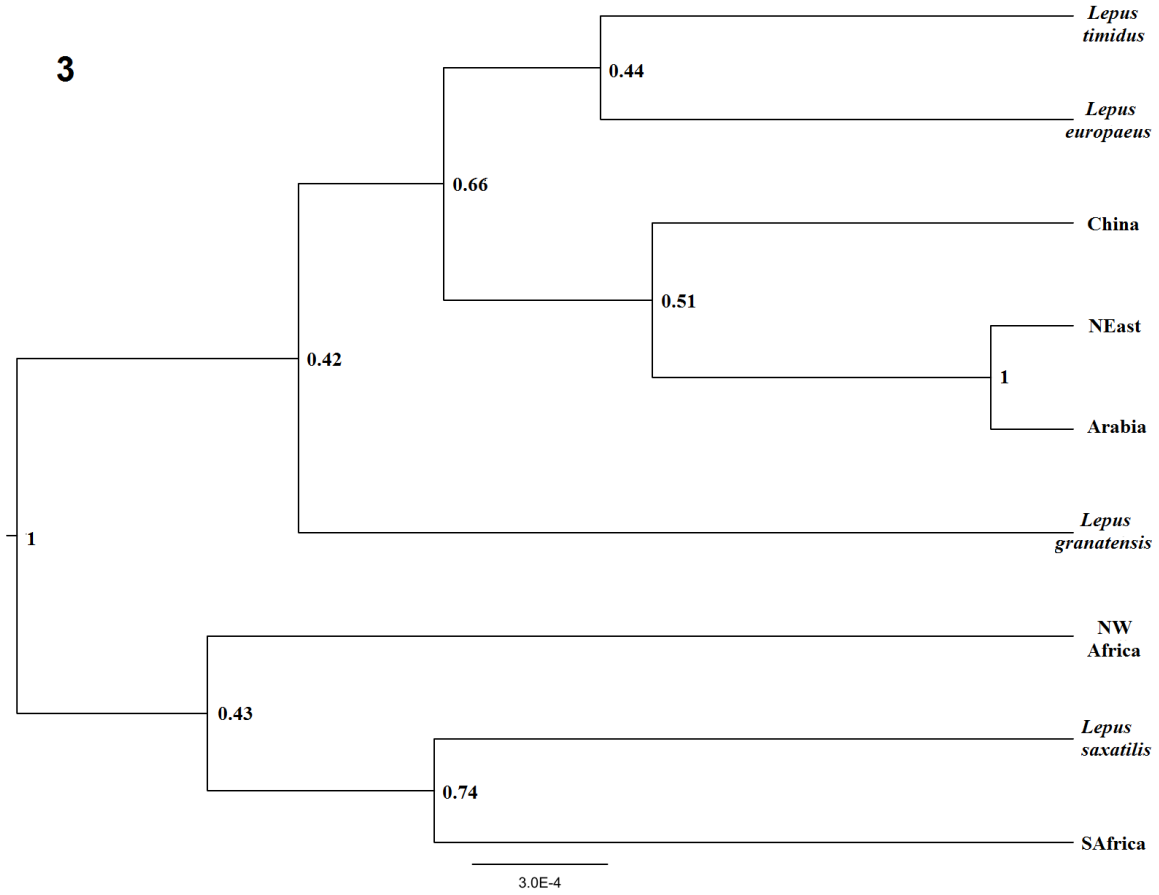


Fig. 16 - Nuclear DNA species trees (the BPP of each clade is shown in each node) for African populations and *Lepus* Species (1- without 2 of the 5 genes; 2 – without S. African population; 3 – all populations/species).

3.3.2 Demography

Historical demographic trend was inferred for mtDNA (network haplogroups) and nucDNA (microsatellite groups) independently for each African *L. capensis* cluster (given the strong discordances in population structure found in some instances), except for 4) clade (two samples from Kenya and one from Tanzania) in the mtDNA since it only included three individuals (fig. 17 and 18 respectively). In general, BSP and EBSP did not show a clear pattern of population growth through time, although with few exceptions. With mtDNA, the method inferred a more evident demographic growth only for populations NW Africa – Central and Near East (fig. 17.2 and 17.6 respectively). For the nucDNA, the NW Africa population (fig. 18.1) suggests a relatively old expansion that may have begun around 600.000 years ago and then a less abrupt continued growth. It is important to highlight that although these results are worth considering, they must be viewed with caution, since some confidence intervals are very wide.

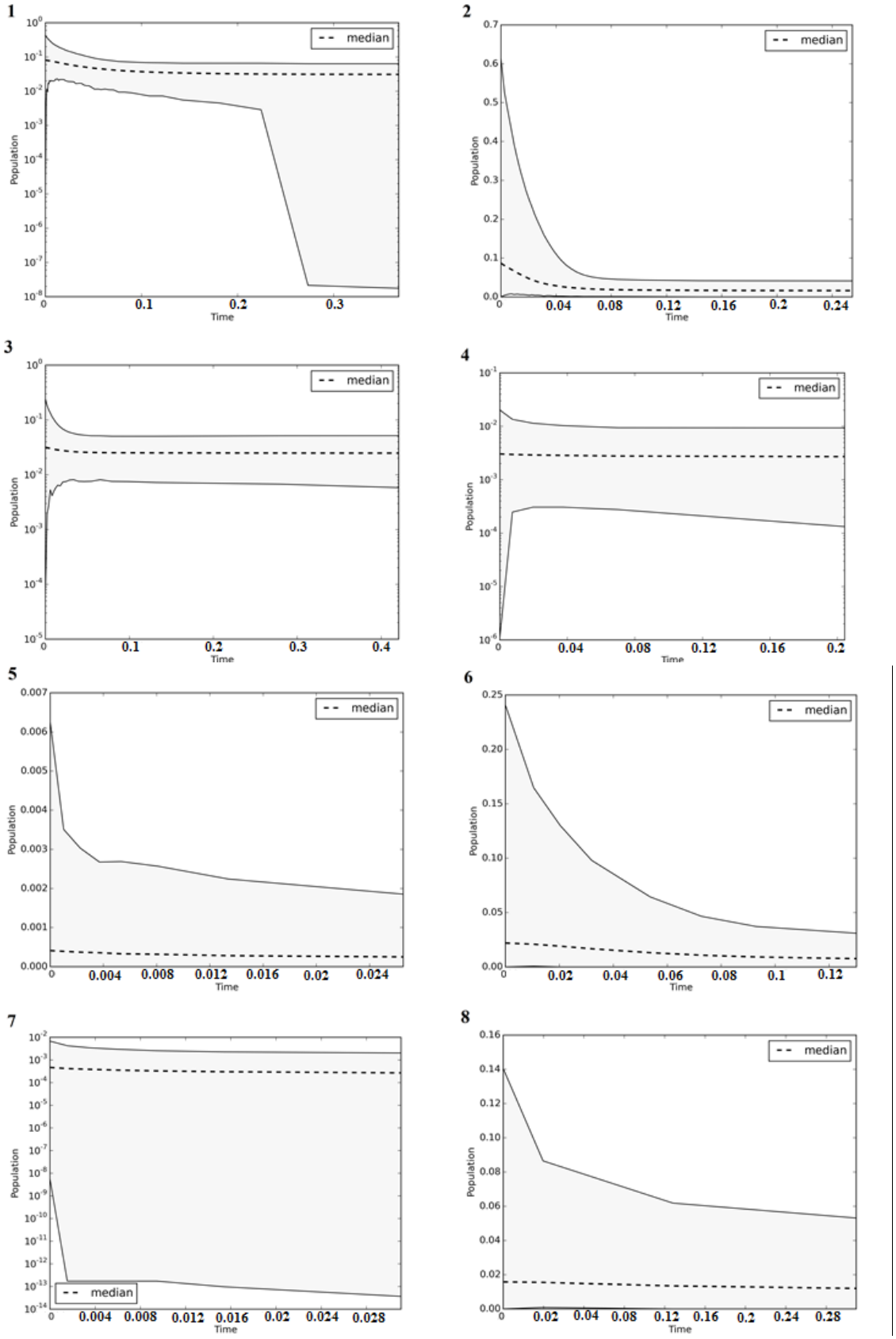


Fig. 17- Bayesian skyline plots for the African *Lepus capensis* populations with the mtDNA dataset (1-4 – NW Africa clades North, Central, South and Southeast; 5 - Saudi Arabia; 6 - N. East; 7 – Kenya; 8 - S. Africa). Relative population sizes are in units of Ne vs. mutation rate and time in million years. Grey lines represent the 95% Confidence Interval (IC).

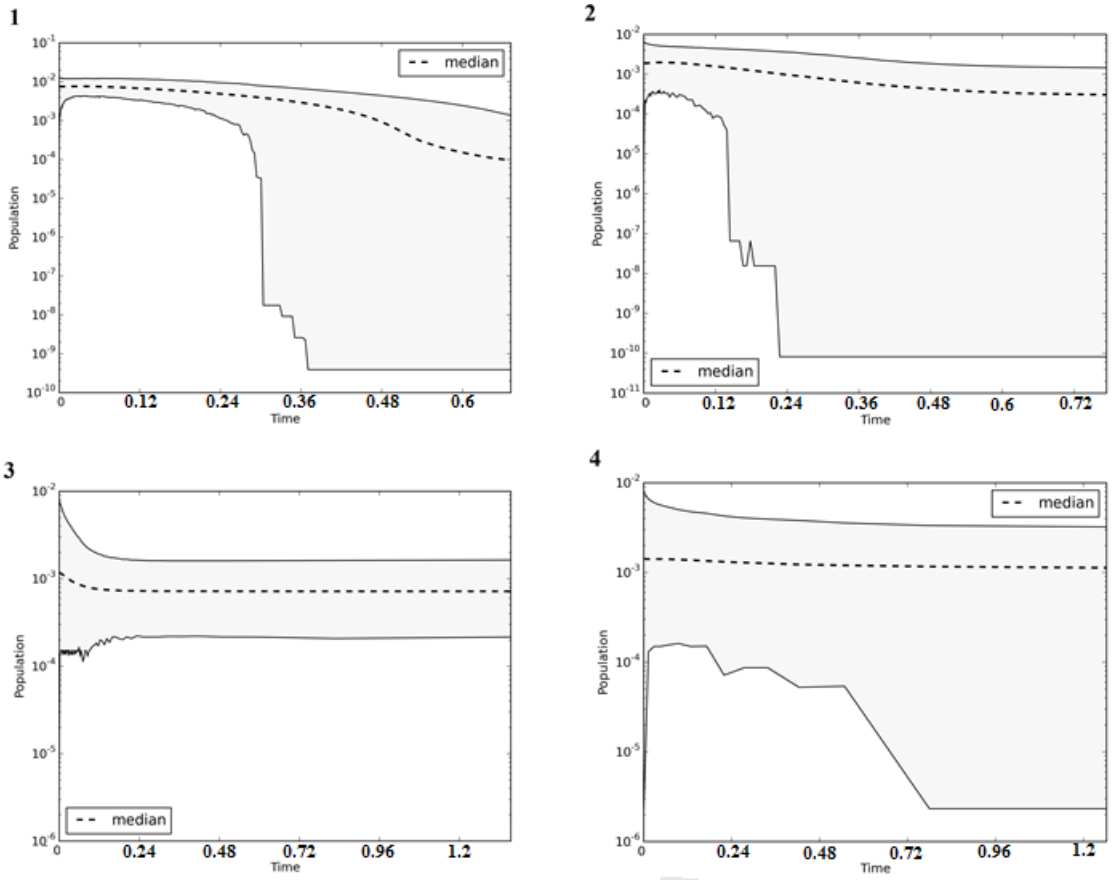


Fig. 18- Extended Bayesian skyline plots for the African *Lepus capensis* populations with the nucDNA dataset (1 - NW Africa; 2 – Saudi Arabia; 3 – N. East; 4- S. Africa). Relative population sizes are in units of Ne vs. mutation rate and time backwards in million years ago. Grey lines represent the 95% CI.

3.3.3 Population divergence – Isolation- with-migration model

Under the Isolation-with-migration model, the maximum likelihood IMa2 software generated co-estimated multi-locus population divergence parameters, inferred here for some pairs of populations (see model in fig. 19).

The consistency of the runs were accessed between each replicate by the ESS values higher than 50, the parameter estimates trends along the run and the consistency across independent runs, and the shape of the curves. In general, parameter estimates did not differ much between replicates and the density curves of the parameters' estimates for the extant populations were consistent. Among the analysed populations, gene flow was never significantly different from zero. Moreover, the right tail of the posterior density curves for the ancient population size and the divergence time between the pair of populations *L. europaeus* and Near East failed to reach zero. The geometric mean of the nuclear DNA loci was estimated from the locus-specific mutation rates calculated by Melo-Ferreira *et al.* (2012) for the estimated

parameters' values (table 5). *Lepus europaeus* - Near East and NW Africa - S. Arabia were estimated to have diverged about 811,363 and 685,569 years ago, respectively. The split time between NW African and South African population was also estimated in approximately 804,000. Among the populations, NW Africa presented the largest estimates of effective population size 1,151,002.

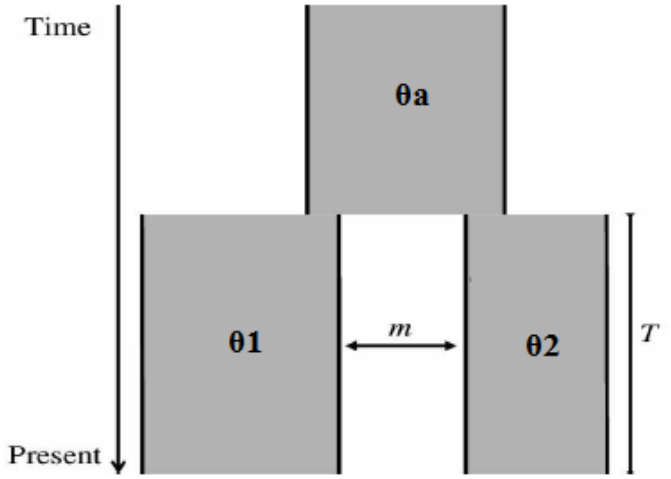


Fig. 19- The representation of the Isolation- Migration model for two populations, in which diverged *T* split between two populations at a time of divergence from a common ancestral population. The parameter *θ1/2/a* are the effective population size of population, and *m* is the migration from population 1 to population 2.

Table 5 - ML estimates (95% posterior density intervals in parentheses) of demographic parameters obtained with IMa2 between three pairs of populations.

Pop.1	Pop.2	*Ne1*[a]	*Ne2*[a]	*NeA*[a]	*t*[b]	2Nm1->2[c]	2Nm2->1[c]
N. East	*L. europaeus*	239,889	152,825	173,665	811,363	0.01	0.05
		(120,871; 445,972)	(76,413; 262,582)	-	-	-	-
NW Africa	S. Arabia	1,151,002	153,048	52,414	685,569	0.04	0.01
		(836,520; 1,589,179)	(69,186; 287,226)	-	(427,695; 997,954)	(0.00; 1.65)	-
S. Africa	NW Africa	210,602	841,319	37,549	805,403	0.01	0.03
		-	(527,321; 1,303,883)	-	(374,403; 1,436,665)	-	-

Notes: [a] Effective population size of population 1 (Ne1), 2 (Ne2) and the ancestral population (NeA); [b]Time in years since species 1 and 2 split; [c]Population migration rate into population 1 (2Nm2→1) and population 2 (2Nm1→2).

3.3.4 mtDNA introgression - Coalescent simulations

The estimates of effective population size and divergence time obtained under the IM model were used to simulate Cytb datasets under a model with no gene flow. This represents the expected distribution of minimum pairwise distances for Cytb between these population that would be expected from incomplete lineage sorting alone (fig. 20, blue bars). The observed pairwise mtDNA divergences (fig. 20 – red bars) were smaller than the 5th percentile of the simulated minimum distances, suggesting that incomplete lineage sorting does not explain the Cytb proximity, which is likely due to introgression. Distances between NW Africa and *L. europaeus* (fig. 20 – green bars) were found to lie mostly within the range of distances expected under a strict lineage sorting scenario (although some values are rather extreme emerging on the opposite 5th percentile of the Near East-*L. europaeus* results, which may indicate that our test was conservative).

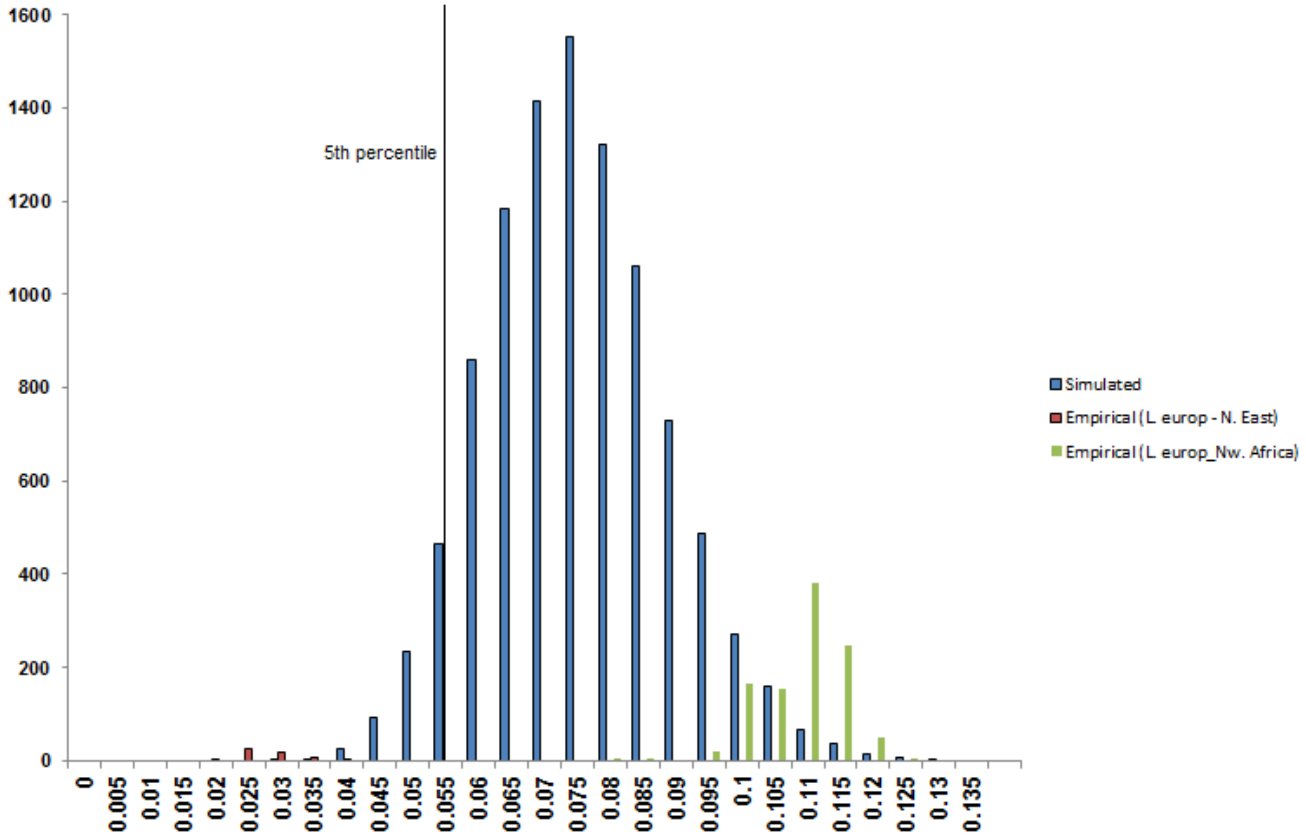


Fig. 20 - Results of the coalescent simulations of mtDNA sequences from population parameters estimated with multiple nuclear loci. The distribution of the simulated minimum pairwise uncorrected *p*-distances between *L. europaeus* and Near East population is given by the bars - lowest 5th percent (red bars), highest 95% (blue bars) and empirical data of NW Africa. Vertical line indicates the 5th percentile of the distribution of simulated distances.

4. Discussion

The magnitude of a combination of past climatic and stochastic dispersal events resulted in the contemporary pattern of species diversity and distribution in Africa (Kingdon 2013). The climatic oscillations that characterized the Pleistocene led to distinct population dynamics depending on life history and geography, and the current genetic constitution of the populations and species carry signals of these past dynamics (Hewitt 2004). Consequently, by studying these patterns caused by alternate periods of expansion and retraction, it is possible to clarify the importance of historical events in shaping Africa's biodiversity (Menegon *et al.* 2014). North Africa has been affected by climatic oscillations, mainly in the Pleistocene, that influenced the range of the Sahara and organisms inhabiting it. The biogeographical intersection between the Palearctic and Afrotropical regions in the Mediterranean basin is considered a hotspot of biodiversity (Cheylan 1991; Kowalski & Rzebik-Kowalska 1991; Myers *et al.* 2000), although the historical processes that shaped biodiversity in the North African part of this region still remains only partly understood. Moreover, due to the intersection of climatic regions, Northwest area is also compared with an "island" as it is isolated by the ocean and the desert, with natural endemism (Cooke 1963).

Despite Africa being generally rich in biodiversity, information about what organisms are present and the diversity and taxonomy of the already described organisms is still deficient. Additionally, detailed studies on the evolutionary events that shaped the existing diversity and the relationship of the evolutionary patterns related with biotic (e.g. ecology) and abiotic (e.g. climate) factors are still missing. The existing studies on the complex genus *Lepus* in Africa were mainly based on evidences of morphology and/or palaeontology, which are limited sources of information, due to the large overlap of characteristics between species. This is a consequence of the rapid radiation and potentially reticulate evolution which has been widely studied in other regions where other species of this genus occur. Few molecular investigations on African hares were done, but mainly on small and specific areas (e.g. Ben Slimen *et al.* 2005, 2006; Suchentrunk *et al.* 2009). Molecular data (e.g. Kryger 2002; Kryger *et al.* 2004) are mainly restricted to South African species (*L. saxatilis* and *L. capensis*) with a small reference made to individuals from Mozambique and Morocco (Alves *et al.* 2003a; Melo-Ferreira *et al.* 2012). Also, other molecular study for African and near African taxa include the Mediterranean hare (*L. c. mediterranus*) (Pierpaoli *et al.* 1999) and the African species *L. starcki* and *L. habessinicus* (Azzaroli Puccetti 1996; Tolesa 2014).

This study sought mainly to clarify the diversity, taxonomy and the relationship of the evolutionary patterns in Africa, using the genus *Lepus* as a model. Due to the lack of integrative investigations, the taxonomy and evolutionary history of this genus in Africa are still unclear. It is focused mainly on *Lepus capensis,* which is poorly understood and have been under less attention, being described as having one of the broadest distribution areas among hare species. We took advantage of multilocus analysis seeking to address the perspective of different types and levels of information that the different markers could provide. Molecular tools were chosen to infer possible biogeographic patterns related to intraspecific divergence, phylogenetic interrelationships and geographic distributions of the phylogenetic groupings so the taxonomy and systematics could be clarified. Although the nuclear dataset for gene and species trees datasets were not reduced to the large non-recombinant block, evidences were shown in other investigations that recombination had little influence in the phylogenies (e.g. Lanier & Knowles 2012; Melo-Ferreira *et al.* 2015).The results of this investigation not only answer the initial proposed objectives, but also open new avenues of research to be addressed in the future. Moreover, the advantage of using more than one class of markers, especially for species with large populations, high dispersal and recent colonization histories was shown.

4.1 Population structure, phylogeny and phylogeography

Analyzing genetic patterns gives insights into species colonization and sometimes uncovers unexpected genetic subdivision, distribution and mixture of species, which can greatly help our understanding of how organisms were affected by climate oscillations (Hewitt 2000). Several molecular biogeographical studies performed in North African species suggested high levels of genetic diversity, with the majority of taxa exhibited multiple endemic genetic lineages and divergence estimates for many clades date back to the Pliocene (e.g. Pook *et al.* 2009; Habel *et al.* 2012; Husemann *et al.* 2012). Also, splits between North African lineages are often deeper than among the European ones (Husemann *et al.* 2014).

One interesting comparison is the phylogeography of the African clades with the savannah ungulates just from sub-Saharan biomes, where harbors the highest diversity of these hoofed mammals. There is a striking concordance in the structure between 19 species with regional distinct lineages. This reflects the survival and divergence of isolated populations in savannah refugia during the climate oscillations of the Pleistocene in West, East, Southern and South-West Africa (Lorenzen *et al.* 2012).

These lineages are similar with the lineages found in this study in sub-Saharan areas, meaning that due to the environmental instability in some parts of Africa, it enabled several spatial refugia that is reflected in the high inter-and intraspecific diversity.

Population genetic analyses continue to be one of the most common approaches to infer population differentiation, where unraveling unknown cryptic diversity is clearer when the systematics of the species is poorly understood. The arbitrary definition of population as a basis for genetic analysis can skew the analysis, inhibiting the detection of unknown patterns, leading to misinterpretation of the inferred patterns and possible unsuitable conservation strategies. So, in this work, individuals were assigned into genetically distinct groups based on their multi-locus genotypes and this was used to detection of population boundaries and structure. The results revealed strong biogeographic structuring at the genetic level, inferred from the microsatellite data using Bayesian clustering methods, and deep divergences based on both mtDNA alone and on the set of sequences from five nuclear DNA loci, based on phylogenetic analyses. Five major geographically explicit groups were identified in *Lepus capensis* based on microsatellite analysis, and mtDNA supports and suggests further substructuring in the Northwestern part of Africa. The divisions of the most widespread microsatellite group in NW Africa which divides into distinct 5 with mtDNA (fig. 13) will be here denominated by: I – the most Northern one (red; fig.13); II – Central (pink; fig.13); III – South (purple; fig. 13); IV – Southeast (orange; fig. 13) and V – Kenya/Tanzania (grey;fig.13) to be easier to distinguish them.

The results underline the potential of these 18 loci to be useful in population and conservation genetics studies of important or endangered relevant units of African hares, where these groups might represent newly discovered evolutionary units or possible cryptic differentiation. It is also important to highlight that even with supposed less genetic resolution due to the smaller Cytb fragment, the use of museum specimens (and consequently taking advantage of more populations) allowed acquiring information that the other analysis based on modern samples and with a larger fragment did not.

4.1.1 Differentiation and diverging entities present in Africa and Near East

Microsatellite analysis showed five distinct groups with good geographical coherence: Near East; Arabia; Kenya; South Africa; NW African (plus two samples from Kenya, one from Tanzania and one from S. Africa). These groups were corroborated by STRUCTURE and PCA (fig. 7 and 8), and F_{ST} results (table 2) support this population structure as well. However, the structure inferred using BAPS was more profound, sometimes almost at the individual level. This is relatively common since it was reported by other authors (e.g. Gordon *et al.* 2008; Bohling & Waits 2011) which used both programs with microsatellite analysis. But this extra-structure can be justified by differences in the underlying algorithms (Frantz *et al.* 2009). Moreover, both software programs are assumed to be powerful Bayesian tools to detect genetic structure (Latch *et al.* 2006). STRUCTURE has been shown to better detect an admixed genotype when compared to BAPS and to correctly estimate an individual's true ancestry composition, but BAPS is more likely to classify single individuals as a group (Bohling *et al.* 2013). Alternatively BAPS may overestimate genetic structure when analysing at the individual level. Still, comparing the results of both programs, the largest populations are congruent: Northwest Africa, South Africa, Arabia, Near East and Kenya. Nevertheless, although BAPS and mtDNA datasets sub-structured more compared with STRUCTURE results, the sub-structuring is not congruent between BAPS and mtDNA. For instance, the NW African STRUCTURE group continues mainly intact with BAPS (few individuals are identified as different populations), but this same group is divided into five with the mtDNA dataset.

Although the most widespread microsatellite group is mainly in NW Africa (which divides into distinct 5 with mtDNA), 4 more museum samples were also included in the group: two samples from Kenya, one sample from Tanzania and one sample from South Africa, and this can have several possible explanations. First, it can result from the retention of ancestral polymorphism or lack of sampling of lineages between Northwest and South. However, other plausible explanation is that individuals can occupy a large area from North to South and Kenya. The possible migrating until South can be explained by the vegetation along the coast as it is observed in other species (e.g. Honey Badger (*Mellivora capensis*; Begg *et al.* 2008). Additionally, the distribution until Kenya (West to East Africa) is also common with other species (e.g. Northern Lesser Galago (*Galago selegalensis*; Bearder *et al.* 2008); Patas Monkey *(Erythrocebus patas;* Kingdon *et al.* 2008); Zorilla (*Ictonyx striatus*; Stuart *et al.* 2008)) and can be due to areas of forests, vegetation and savannah that are found right below

the desert, forming a possible ecological corridor. Alternatively, the sample from South Africa can result from miss identification of the sample locality in the museum database. This however seems rather unlikely since the mtDNA analysis groups these samples with *L. saxatilis* and Angolan samples (yellow group; fig. 13), which makes geographical sense. Regarding the phylogenetic analysis based on mtDNA obtained for NW African populations, is worth considering the clearly higher structure in geographic space for the mitochondrial DNA when compared to the microsatellite results: the red microsatellite Northwestern group divides into five mtDNA lineages (fig. 7 and 13). This can result, first, from the genomic features of the mitochondrial DNA – haploid and uniparentelly inheritated – leading to a complete process of lineage sorting faster than nuclear DNA (Funk & Omland 2003), or, second, from population evolution processes, such as sex-biased dispersal asymmetries (male mediated dispersal and female philopatry), other demographic disparities between males and females or human introduction (Toews & Brelsford 2012). Species with male-biased dispersal should have more mtDNA structure and therefore the use of nucDNA markers (in addition to mtDNA) will help prevent biases when determining population structure (Fahey *et al.* 2014). In this case, male mediated dispersal and female philopatry appears to be plausible leading to a low dispersion of the maternally inherited mtDNA. This is likely to happen since philopatry has been observed in different hare species. One example is hares in the Iberian Peninsula whose mtDNA suggests female philopatry, although in an introgression context (despite the successive introgressions, mtDNA remained in place) (Melo-Ferreira et al. 2014a). This sex-specific difference in gene flow was also evident even on small geographic scale between neighboring populations (Fickel 2003; Fickel *et al.* 2005) in European and Asian species (Hamill *et al.* 2007; Mamuris *et al.* 2010), and in the South African Cape hares and Scrub hares (*Lepus saxatilis*) by Kryger (2002) was concluded from molecular data too. One must also consider the possibility that our microsatellite dataset is not properly uncovering the detailed population structure on this region. We noted that several microsatellite loci were found to be in HW and linkage disequilibrium in the population from this region and its effective population size inferred with the IM model from nuclear sequence data is extremely large. Both patterns could be due to hidden structure. However, given that the microsatellite dataset proved efficient in detecting many other instances of population structure corresponding to variable levels of divergence, this possibility seems at this point unlikely.

Recognizable biogeographic provinces presumably exist in North Africa, because of environmental impediments (ecological and/or physical; historical as well as contemporary) to dispersal and gene flow. These impediments are conventionally

recognized as reflected in the coincidence of distributional limits for different species or populations, and lead to the divergence of evolutionary lineages. The 5 sub-divisions of mtDNA compared with one microsatellite group have shown several biogeographic aspects that are worth noticing. The distribution of the mtDNA clade I (represented in red in fig. 11 and 13) covers most of the Maghreb expanding throughout Tunisia and Libya, and the Atlas Mountains (see figure 20 for main mountain massifs and rivers) do not seem to be a barrier since some individuals occur above and below the mountain chain. However, for animals with low dispersion, for example Mediterranean Pond Turtle (*Mauremys leprosa*; Fritz *et al.* 2006) and Bibron's Agama (*Agama impalearis*; Brown *et al.* 2002), this mountain chain has been shown too to act as a barrier to gene flow, where distinct clades occur on the different sides of the mountains. On the contrary, the mtDNAclade II (represented in pink in fig 11 and 13) seems to be endemic to the Atlantic Coastal Sahara, occurring along the Atlantic coast from the southern slopes of the Atlas Mountains to the Tagant Mountain of Mauritania. The region exhibits also other endemic forms, such as the Helmethead Gecko (*Tarentola chazaliae;* Le Berre 1989) or Tarfaya Shrew (*Crocidura tarfayensis;* Le Berre 1990). Since they are limited by the desert, they benefit moisture winds from the Atlantic which leads to optimal conditions. Apparently, rivers do not constitute barriers to dispersal for the clade III (represented in purple in fig. 11 and 13), which occurs to the North and to the South of the relatively wide and deep Senegal River (fig. 21). The members of this clade are distributed also in the mountains of Adrar Atar (fig. 21) in Mauritania. Moreover, the region between clades II and III lacks evident barriers to gene flow: it is mostly flat and the single river present in the area (river Lakra; fig. 21) is dry most of the year. Thus, the contact zone apparently does not occur in an ecotone and probably constitutes a hybrid zone. The contact may result from expansion from refugia, when both clades meet in secondary contact. Demographic inferences show that, for the nucDNA dataset, NW African clade (without museum samples, which comprehends these three mtDNA lineages) suggests a relatively old expansion that may have begun around 600.000 years ago. Even though the results do not seem robust, if this inference is accurate, it is coincident with the Pleistocene glaciations.

The mtDNA clade IV (represented in orange in fig. 11 and 13) was detected from the Central Sahel (Mali) until Togo/Ghana and further sampling is needed to delimit the distribution of this clade. But, one plausible explanation is that the Niger River (e.g. Brouat *et al.* 2009; Dobigny *et al.* 2013), as also seen in other river systems (e.g Robbins 1978; Nicolas *et al.* 2006), acted as geographic barrier and isolated this population. Considering the high diversity observed in NW Africa, this *Lepus capensis* species complex has probably occurred in this region as the mountain chains may

have played a role by acting as refugia in periods of climatic instability and facilitating isolation (Barata *et al.* 2012) as mtDNA suggests. But currently, with the Atlantic-Sahara corridor (see fig. 22 for main ecological corridors), male migration along the Western coast is possible for all four lineages which is observed with microsatellites as these markers can tell a more recent history, while females tend to stay on their home places, if the female philopatry hypothesis is true. The final mtDNA clade, V (represented in grey in fig. 11 and 13), only incorporates three samples: two from Kenya and one from Tanzania. As mentioned before, it is possible that due to the vegetation that exists right below the desert, now it can act as an ecological corridor and the individuals can migrate from Northwest to East Africa, but in the past there could have been a barrier in between. But in Kenya exist 2 distinct lineages and despite this grey clade, microsatellites and mtDNA detected a distinct lineage with samples just from this country (represented in light blue in fig 7, 11 and 13). Additionally, an intriguing issue is the sample from the East Sahel (Niger; shown in fig. 11 and 13), which clusters with this second Kenya lineage, independently of the large geographic distance. Probably there are lineages missing due to the sampling gap between these two areas. Samples constituting this light blue lineage belong to the *L. capensis* distributional range according to IUCN distribution map, but this lineage shows be a distinct evolutionary entity. Further sampling is needed to determine the distribution of the haplotype as well as other possible cryptic diversity present in the region. The results also suggested that probably there is a barrier to the dispersion and female gene flow between Mali and Niger since there were two distinct lineages according to mtDNA (Niger sample was an admix individual for microsatellites).

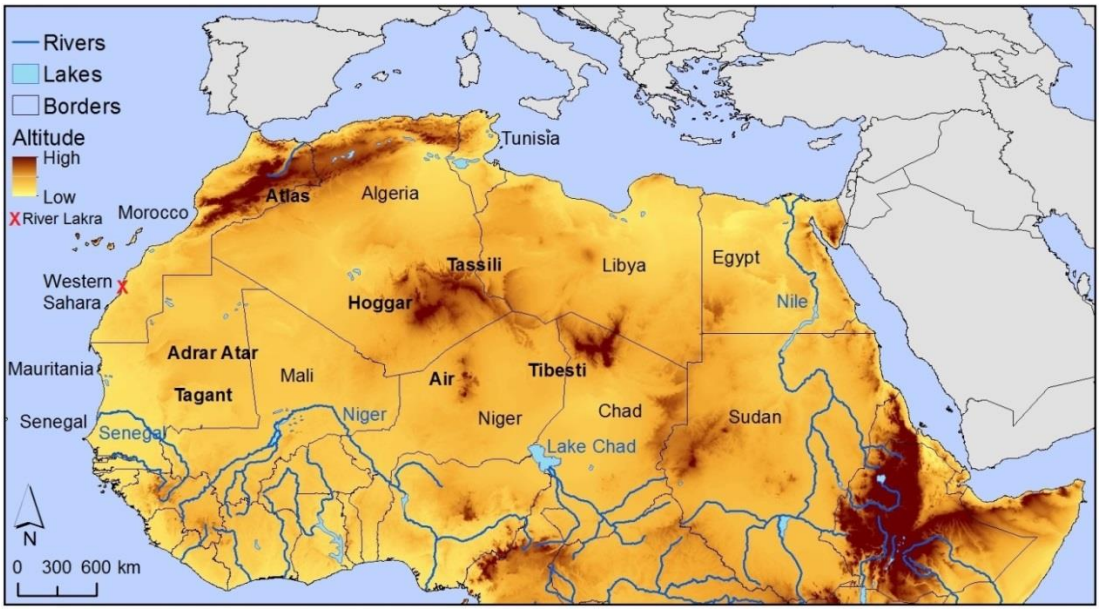


Fig. 21– Main mountain massifs and rivers of North Africa (adapted from Brito *et al.* 2014).

The mtDNA phylogeny (network and tree, fig. 10 and 12 respectively) suggested obvious lineage divergences, overall revealing nine distinct haplogroups/lineages in Africa, Arabia and Near East with geographical meaning. Some of these groups are more divergent from the Northern populations (Near East/Kenya, Saudi Arabia/Oman and South Africa as well), which may represent different diverging entities. Also, some clades like Kenya and Near East are closely related to each other. But different origins can explain these situations. First, although some biodiversity corridors across the Sahara desert have been fragmented and have contributed to the isolation of certain populations, others continue to play an important role in species dispersal and in facilitating gene flow (Brito *et al.* 2014). Besides the Atlantic Sahara corridor which was described above, Nile River and Red Sea Sahara corridors (fig. 22) perhaps explains the proximity of the Kenya and Near East clades, where individuals can freely migrate, with suitable vegetation along the East track.

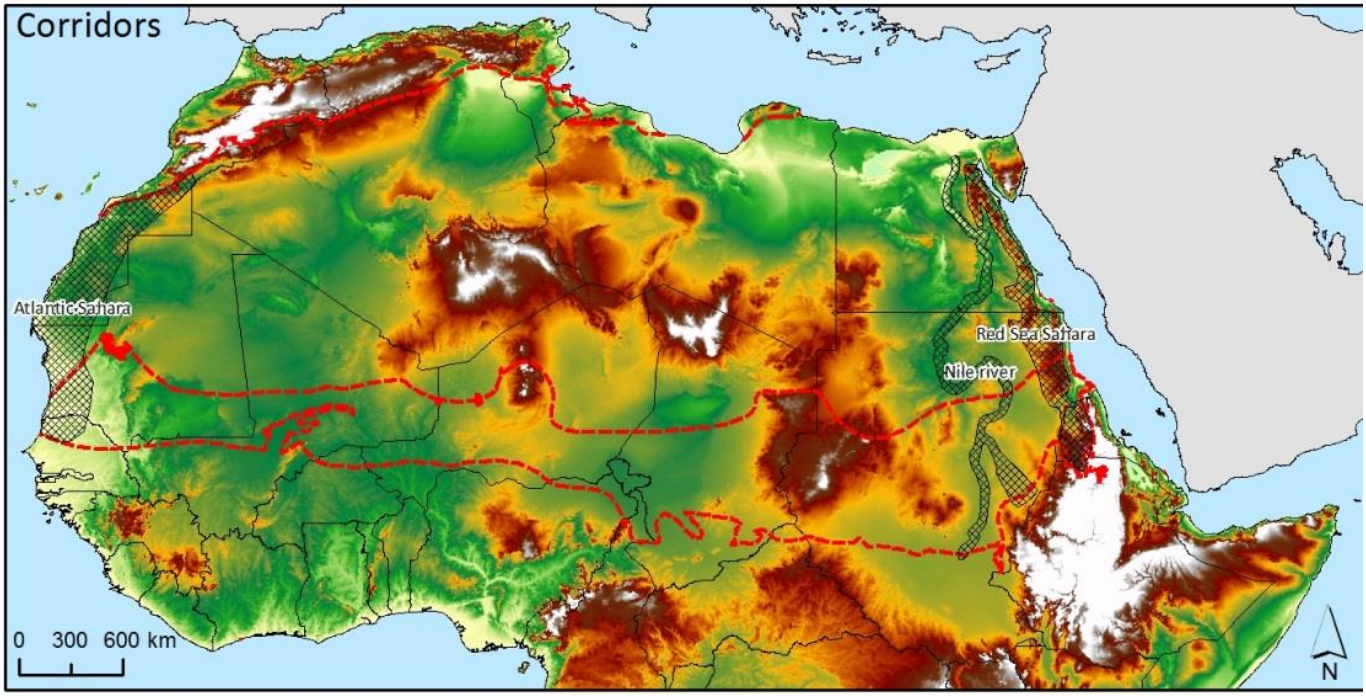


Fig. 22 – Three ecological corridors in North Africa are represented with a black grid effect. The red dashes represent ecotones (adapted from Brito *et al.* 2014).

4.1.2 Saudi Arabian hares are a divergent evolutionary unit

Both in microsatellite (except for one individual which was admixed) and mitochondrial analysis, Saudi Arabia showed to be an independent and differentiated group. After testing possible nuclear gene flow under the isolation-with-migration model between NW Africa and this population, the results did not suggest gene flow. As water systems like seas and rivers acting as a barrier to gene flow between species and populations have been reported before, it is possible that with the disjunction of the Euroasiatic plate and with the opening of the Red Sea, it acted as a geographical barrier and desert in the North part also isolated populations of the Saudi Arabia/Oman from North African populations. This split was estimated to be between 700,000 to 880,000 years ago, which dates back to the middle Pleistocene, where probably this population differentiated into a different evolutionary unit. Also, morphology supports the idea that we are in the presence of a new possible species, since the Arabian hare is much smaller than the hares in Europe and Africa, but its conservation status was not evaluated yet (Dubai desert conservation reserve; http://www.ddcr.org). Although just in this region different *L. capensis* subspecies have been already identified based on morphological characters (Hoffmann & Smith 2005), this group may be recognized for their genetic, geographical and morphological differences. The identification of a subspecies should be sufficient as first-order systematic hypothesis when the aim of conservation is to preserve biological diversity (Green 2005).

4.1.3 Phylogenetic relationships with neighboring *Lepus* species

The understanding of the deep divergences found in *L. capensis* requires an integrated analysis including other neighboring hare species. Taking into account the ML and BI trees for mtDNA, the Network results of the African hares are corroborated. But, when adding more species and populations from GenBank, namely *L. capensis* from China, *L. saxatilis, L. granatensis, L. europaeus* and *L. timidus,* the phylogenetic trees allows better understanding the evolutionary relationships among different clades and species. Although the supports are not high, there is evidence for strong biogeographical structuring of genetic diversity and population differentiation where some populations are closely related with other species (e.g. *L. saxatilis* and South African populations; *L. timidus* and Chinese *L. capensis*; also Brown hares (*L. europaeus*) show phylogenetic affinities with Near East hares and Kenya). Since the proximity of the S. African population to *L. timidus* - *L. capensis* from China (fig. 12)

was not observed in the inferences without museum samples nor in Lado *et al.* (2013), interpretation will not be done.

4.1.3.1 *L. capensis* and *L. saxatilis*

Although Cape of Good Hope, in South Africa, is the type locality of the cape hare *Lepus capensis*, a single species (*capensis sensu lato)* that inhabits Africa and Near East, Flux & Angermann (1990) and Hoffman & Smith (2005) mentioned there is no evidence of gene flow between the South Africa and the populations in East, West and North Africa. Analyzing the results of this work, the South African population shows to be an independent entity from the other *L. capensis* populations. In fact when analyzing the nuclear divergence using the species tree method (i.e. based on the distinct microsatellites groups), the results showed two main clades: Africa versus "no Africa", and North versus South African populations. The four species tree (Fig. 15 and 16) showed geographical meaning and the level of divergence between some lineages, are as deep as between other species. In the species tree, *L. saxatilis* and S. African *L. capensis* population are more closely related (sister taxa) than these *L. capensis* to the North African ones. An isolation-with-migration model analysis suggested that these populations did not exchange genes since their divergence. The absence of sharing of mtDNA variation and of admixed individuals between these populations in the STRUCTURE analysis is also compatible with the absence of gene flow between north and South African *L. capensis*. Under the IM model, the split time was estimated to be 800,000 ya in the Middle Pleistocene, and with the species tree inference (BPP = 0.74) South Africa and *L. saxatilis* were estimated to diverge 700,000 ya. This result suggests a disjunction between the taxonomic classification and the evolutionary history of these populations. Note that this pattern agrees with another independent study based on another set of nuclear genes and sampling (Tolesa 2014).

4.1.3.2 *L. saxatilis* distribution range

The geographic range of *L. saxatilis* includes South Africa, Swaziland, Lesotho and the Southern regions of Namibia (Collins *et al.* 2008). In other mtDNA studies, there were evidences that suggested *L. saxatilis* comprises different lineages (Kryger *et al.* 2004) with at least seven forms as synonyms just in South Africa.

The only *L. saxatilis* sample which could be amplified for mtDNA is from Botswana, which is not included in the original distributional range of the species, but geographically close. This sample clustered with the museum sample of cape hare from S. Africa and with the only sample from Angola (group 13 described in the fig. 13). However, the other mtDNA sequences of *L. saxatilis* specimens from GenBank clustered with the group 4 (two samples from Kenya and one from Tanzania; fig. 13). Assuming that the sample from Botswana and the sequences from GenBank truly represent *L. saxatilis*, can this translate a larger distribution of *Lepus saxatilis* corresponding to both mtDNA clades 4 and 13 represented in the figure 13. More samples are needed to address this question since due to sampling gaps the genetic diversity is underestimated, as well as the inclusion of other markers, to access the present of unknown lineages in the region.

4.1.3.3 *Lepus timidus* and Chinese *L. capensis*

Regarding mtDNA phylogeny results, Chinese cape hares clustered within the clade of *L. timidus.* This haplotype sharing may result from incomplete lineage sorting affecting mtDNA, genetic introgression events or indicate that these Chinese *L. capensis* correspond to a distinct species, closely related with *L. timidus*. When analyzing Chinese hare relationships, Wu *et al*. (2005) concluded that *L. capensis*, a species also with a broad distribution, does not occur in China since its mtDNA was similar to that of *L. timidus*. Also, Yu (2004) had previously proposed that the Chinese hares most probably derived from *L. timidus*. However, these conclusions were based on mtDNA alone which may lead to erroneous taxonomic assessments, as mtDNA introgression is frequent among hares (Alves *et al.* 2006). Considering that *L. timidus* was able to maintain a large and continuous area of distribution during the glacial periods (Melo-Ferreira *et al.* 2007), it seems possible that it could be in contact with *L. capensis*, leading to hybridization and introgression of mtDNA. Finally, the persistence of ancestral genetic variants in *L. capensis* could also eventually explain the presence of *L. capensis* haplotypes that were closely related with *L. timidus*, even if this hypothesis seems less likely. Regarding nuclear divergence results, there is mito-nuclear discordance in the phylogenetic pattern, given that for mtDNA these two population belong to the same lineage, contrary to the nucDNA results, which may indeed result from mtDNA introgression. Note however that sequences for only two loci were available for Chinese *L. capensis* and a more thorough assessment would be needed in the future.

4.1.3.4 *Lepus microtis* and *Lepus capensis*

Another interesting issue are the museum samples considered to be *L. victoriae,* now recognized as *L. microtis* that were not discriminated along the analysis. If these samples actually belong to the African Savannah hare, which in fact follow the known distribution according to IUCN distribution map (Smith & Jonhston 2008), the results did not show a distinct lineage for this species samples. According to microsatellite analysis, these samples do not show any genetically differentiation where all samples were included in the big red widespread clade. Even with BAPS, which can attribute to just one individual a single clade, there was no evidence for being a different population entity. Regarding the mtDNA phylogeny, it is clear that individuals assigned as *L. microtis* are two independent clades (IV and V), but close to the other North African lineages. Taking into account these results, several conclusions can be taken. If these samples are really representatives of *L. microtis*, genetically the divergence between the two widespread species is not sufficient to distinguish them. However, one hypothesis is the possibility of homoplasy of the microsatellites between species affecting the accuracy of the analysis, which has been reported before (see Culver *et al.* 2001). On the contrary, if these samples are in fact *L. capensis* and the samples' locality is accurate, the range distribution of these species needs to be reviewed. Another hypothesis is that the morphological differences observed and used to distinguish the two species are due to mere local adaptation, which would imply that all North African hares should be classified as *L. capensis*. But it is important to highlight that these are all suppositions based on genetic inferences, since the taxonomic difficulties are a result of inadequate diagnoses of the species. Also, due to the sampling gaps, the diversity is underestimated and since the morphology of the individuals was not accessed, this issue cannot be truly debated.

4.2 Sampling gaps

This work took advantage of a wide range of samples along Africa and Near East, both from road kill and museum specimens, incorporating several different populations. But one important point is that although there is a good coherence between geographical origin and genetic make-up of the populations, certainly there are still missing lineages to be described due to the sampling gaps. For example, the only difference between Networks considering museum samples or not is the Niger sample, which is included into distinct clades (Near East and Kenya, the two closely related clades), and this can be happening with other samples. Even this Niger

samples can constitute a third cluster. This shows that more sampling areas are necessary. Also, specifically for the mtDNA Northwestern substructure, sampling is needed for specific regions. For example, for clade I (represented in red in fig. 11 and 13), the full extension of such expansion can be better assessed with a wider sampling in the regions of Algeria and Egypt. In addition, the IVth grouping (represented in orange in fig. 11 and 13) was detected from the Central Sahel (Mali) until Togo/Ghana and further sampling is needed to delimit the distribution of this clade. Samples mainly from Mali, Guinea, Sierra Leone, Liberia and Burkina would be a great help, although some regions are military zones and therefore exist numerous mining activities which makes this process really complicated. New sampling from Nigeria, Cameroon, Democratic Republic of the Congo and Uganda would clarify the clade V, since this group only included 3 samples (represented in grey in fig. 11 and 13). Also, for mtDNA and microsatellite Kenya group (represented in light blue, fig. 7, 11 and 13), further sampling is needed across the eastern half of the Sahara-Sahel, in countries like Chad and Sudan to determine the distribution of the haplotype as well as other possible cryptic diversity present in the region. But not only for *L. capensis* more sampling is necessary. For future work, additional sampling from Angola, Namibia, Mozambique and Botswana is needed to understand in detail the population structure, distribution and the systematics of presently classified *L. saxatilis*.

For a better understanding of the distribution of the *L. capensis* lineages, the filling of such sampling gaps and molecular identification of populations may help, and consequently disentangle cryptic or unknown lineages which may be present. The molecular studies will likely continue to reveal biodiversity and genetic analysis of museum specimens will provide essential material from regions where sampling is currently nearly impossible.

4.3 Evidences of mitochondrial introgression

The lack of phylogenetic agreement between mitochondrial and nuclear markers is relatively common and can be explained by different phenomena. Although incongruence observed in gene trees among different markers are often assumed as introgression events, disagreements can be due to incomplete lineage sorting because of stochasticity of the evolutionary processes (see e.g. Melo-Ferreira *et al.* 2014b). Although distinguishing the type of event that shaped the observed patterns of multilocus genetic variation can be difficult, different methodologies have been developed to understand the impact of retention of ancestral polymorphism and gene

flow (Meng & Kubatko 2009; Hey 2010; Talavera & Vila 2011; Melo-Ferreira *et al.* 2014b).

Since the persistence of the traces of introgression are related with the migration rate, in species with male-biased dispersal and female philopatry (which is the case of several mammals), female transmitted mitochondrial DNA are more vulnerable to persistent introgression (Melo-Ferreira *et al.* 2014a). And this may be the reason why substantial mtDNA introgression has been repeatedly observed in this class of animals, and is frequently more wide-ranging than for the nuclear genome (Petit & Excoffier 2009). This process has been reported by different researchers using molecular markers for several species (Sequeira *et al.* 2011; Bastos-Silveira *et al.* 2012; Jezkova *et al.* 2013), but hares are one significant example in the study of genetic introgression (Arnold 2008). Indeed, previous works show introgression events in different hare species such as *Lepus timidus, Lepus europaeus, Lepus granatensis* or *Lepus yarkandensis* in Europe, Asia and North America (e.g. Melo-Ferreira *et al.* 2005; Thulin *et al.* 2006; Alves *et al.* 2008b; Wu *et al.* 2011; Acevedo *et al.* 2015; between others), but never in African hares.

Analyses of Northeast Africa samples detected discrepancies between mitochondrial and nuclear DNA markers. The histories of the origin of the relationships among these African samples relationships were further investigated here. This is an interesting geographic region for the group because it likely representing a zone of secondary contact, taxonomic transition and possibly admixture between cape hares and brown hares (*Lepus europaeus*). To disentangle if mtDNA gene flow occurs between the two species in the region and is the explanation for the mt-nucDNA discordance the expected mtDNA divergence under a model without gene flow was simulated using the parameters of divergence estimated from the nuclear loci. The results show that introgression is most plausible explanation for the mtDNA proximity because the empirical divergence is smaller than the expected minimum distance in a model where only incomplete lineage sorting occurs. The empirical divergence between NW African population and *L. europaeus* was also plotted and were found to lie within or even above the range of simulated distances, suggesting that our approach was realistic and even conservative.

4.4 Taxonomical remarks

Molecular data adds an important dimension for taxonomic classification: the evolutionary history. This study based on genetic inferences allowed to identify several situations where the taxonomy may need to be reviewed and clarified. Nevertheless, this review may have inherently to integrate data from other sources, including morphology, palaeontology, ecology, physiology, behaviour, among other. Taxonomy and systematics of *L. capensis* are far from being resolved and several questions are still unanswered. It is considered a polytypic species distributed across large parts of Africa (from North to South), the Near and Middle East, and large parts of Central and Far East Asia (Angermann 1965, 1983; Flux and Angermann, 1990; Hoffmann 1993), although IUCN distribution map (Drew *et al.* 2008) shows a distribution until China. Nevertheless, this work shed light onto some arguments that are believed to review at least part of the *L. capensis* taxonomy and systematics.

In the literature, only Hoffmann and Smith (2005) tended to restrict cape hares to South Africa. The authors informally suggested three more species division for *capensis*-type hares (East Africa, Northwest Africa and Arabia-Near East) which might be considered as distinct species, pending on sufficient data to support a formal revision. The basis for this division was the distribution gap between North and South populations and the presence of other species, *Lepus microtis* (synonym to *Lepus victoriae*). However, the results of this work show that not only these divisions are confirmed, but more diverging evolutionary groups were found, corroborated by mtDNA and microsatellite analysis. There is evidence for strong biogeographical structuring of genetic diversity and population differentiation North Africa and so *L. capensis s. l.* may consist of parapatric forms in various stage of divergence – “subspecies, semispecies or allospecies” (Angermann 1983). The phylogeny was not completely resolved, but the different lineages and population structure clarify some phylogenetic relationships have and taxonomic implications, where status of *L. capensis* should be thoroughly revised.

Some lineages were fairly divergent and others were closer to different *Lepus* species. Possibly, the most significant area representing a zone of taxonomic transition is in the Near East where cape hares and brown hares are in contact (Angermann 1965, 1983; Hoffmann & Smith 2005). The timing of divergence between Northwestern and Southern *L. capensis* calculated under this model was estimated to be 800,000 ya in the middle Pleistocene. South African samples are the closest samples of the dataset to the type locality of cape hare *Lepus capensis*. But this population, in all analyses was shown to be an independent lineage. Additionally, the level and time of divergence is bigger between supposed *L. capensis* populations than between a South

African *L. capensis* population and a second hare species, *L. saxatilis*. Relying just on genetic inferences, North African populations should not be considered *bona fide L. capensis*.

Other important taxonomical remark is the Arabian population, today considered to be *L. capensis*. Our results show this population to be an independent group. Since geography, morphology and molecular studies confirm and support this population to be an independent entity, hares of this region should be considered for future taxonomical revision.

More investigation in morphology, ecology, biology and species delimitation is needed concerning these African species and a focus in sampling *L. microtis* is the key. Also analysing nuclear DNA sequences would be important to reconstruct these populations' history. As long as no morphological or molecular data are available for type localities from the Cape of Good Hope region and Sudan, wherefrom Cape and African Savannah hares have originally been described, no final conclusion can be made on their identity and distribution.

Several divergent lineages which probably diverged recently were found to be geographically structured, which likely resulted from fragmentation of ancestral ranges and divergence in allopatry. Regions where haplotypes from different clades were found in sympatry probably result from secondary contact of the divergent lineages after expansion. Whether these evolutionary entities are reproductively isolated and should be considered distinct species should be assessed in the future using a detailed characterization of their genomes and ecologies. Some haplotypes were found to be more closely related to those of other species, which can result from mtDNA introgression, a phenomenon widely described in the genus, or retention of ancestral polymorphism, which may uncover different taxonomic entities.

The possible presence of gene flow, incomplete lineage sorting and the proximity of lineages leads to low posterior probabilities and bootstrap values, and due to its complex evolutionary history and rapid radiation, it is difficult to achieve the true African species tree. Otherwise, including genome-scale data may help resolving certain nodes, although there are cases where Tree of Life remains beyond resolution, even when confronted with these type of methods (Rokas & Carroll 2006).

5. Concluding remarks and future research

Even though population history and taxonomy of African hares is still poorly known, clear progress has been made in this work. This is the first genetic investigation of *Lepus capensis* with a wide sampling area, including different populations along its range, using multi-locus data, which brought new insights about the complex history of this taxon. Moreover, this investigation had a particular emphasis on Northwest Africa where sampling was more comprehensive. This is an important first step towards solving the broad uncertainties that have impeded the establishment of a clear taxonomy and systematics of these organisms, and produced valuable information for future conservation planning.

The cape hare in Africa and Near East was found to be composed by five major evolutionary units with clear geographic distributions: Near East, Arabia, Kenya, South Africa and Norwest African. Even though some differences were expected between cape hare inhabiting different regions, the level population structure identified here exceeds these expectations. In this respect, not only the use of a multilocus approach but also the different nature of the set of genetic markers used here (microsatellites, mtDNA and nucDNA sequences) contributed to i) identify the major evolutionary units without an *a priori* expectation, and ii) infer the level of divergence among the major lineages. Also relevant was the use of museum sampling to bridge some important sampling gaps, even if additional sampling efforts are need for a full understanding of the evolutionary history of the species. Within the Northwest African populations of the cape hare, mtDNA showed increased substructure, contrary to the homogeneous pattern of genetic variation found for microsatellites. This structure can be explained by multiple barriers and corridors found around the Sahara and suggest strong female philopatry.

Interestingly, the level of divergence between particular lineages of cape hare is as deep as the divergence between other well recognized hare species, which point to the possible existence of cryptic speciation within what is so far recognized as *Lepus capensis*. Note however that the phylogenetic resolution obtained from the analysis of sequences of five independent loci did not allow retrieving high support for many of the clades, even if the branching pattern perfectly fits with the geographic proximity of the distribution of lineages. Increasing the battery of nuclear loci, either using similar PCR-based approaches or next-generation sequencing will certainly contribute to solve this issue. On the reverse sense, there were no obvious genetic differentiation between *L. capensis* and the specimens identified as *L. microtis* used in this work. Also, mitochondrial DNA introgression from the brown hare was found to have affected cape

hare populations from the Near East, showing that the evolutionary history of these organisms is further complicated by reticulate events.

Despite the geographical meaning, the species tree nodes and phylogenetic analysis in general were not well supported. This can be due to the rapid evolutionary divergence that leads to short branch-lengths between speciation events, where there is little time for informative substitutions to fix in populations, yielding gene trees that are unresolved with respect to the species tree. For future work, there is the need to increase the number of genes to improve the phylogeny.

Overall, these results demonstrate that the current taxonomy of these African species is far from testifying the true relationships among the different populations. A thorough taxonomic revision is therefore needed, ideally coupling the information of different sources, such as genetic, morphological, paleontological, ecological, among many others, so that the classification of African hares reflects the biology and history of the natural populations. This is obviously a tremendous task that demands the joint effort of experts in these different areas and a strong emphasis on fieldwork, especially in remote geographical areas from which knowledge is inexistent. The revision of museum collections can also be valuable resources for this much needed taxonomic revision. In any case, this work was an important step in that direction.

6. References

Acevedo P, Melo-Ferreira J, Farelo L *et al.* (2015) Range dynamics driven by Quaternary climate oscillations explain the distribution of introgressed mtDNA of *Lepus timidus* origin in hares from the Iberian Peninsula. *Journal of Biogeography*, ***42***, 1727-1735..

Adams JM, Faure H (2004) Review and Atlas of Palaeovegetation - Preliminary land ecosystem maps of the world since the Last Glacial Maximum. Quaternary Environment Network (http://www.esd.ornl.gov/projects/qen/adams4.html).

Alves PC, Ferrand N, Hackländer K (2008) *Lagomorph biology : evolution, ecology, and conservation*. Springer.

Alves PC, Ferrand N, Suchentrunk F, Harris DJ (2003) Ancient introgression of *Lepus timidus* mtDNA into *L. granatensis* and *L. europaeus* in the Iberian Peninsula. *Molecular Phylogenetics and Evolution*, **27**, 70–80.

Alves PC, Harris DJ, Melo-Ferreira J *et al.* (2006) Hares on thin ice: introgression of mitochondrial DNA in hares and its implications for recent phylogenetic analyses. *Molecular phylogenetics and evolution*, **40**, 640-641.

Alves PC, Melo-Ferreira J, Freitas H, Boursot P (2008) The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363**, 2831–9.

Angermann R (1965) Revision der paläarkitschen und äthiopischen Arten der Gattung *Lepus* (Leporidae, Lagomorpha). Dissertation Thesis, Humboldt University of Berlin.

Angermann R (1983) The taxonomy of Old World Leptrs. *Acta Zoologica,* Fennica **174**,17-21.

Arenas M, Posada D (2010) The effect of recombination on the reconstruction of ancestral sequences. *Genetics*, **184**, 1133–1139.

Arnold ML (2008) *Reticulate evolution and humans*. Oxford University Press, New York.

Avise JC (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology*, **7**, 371 -379.

Avise JC, Arnold, J, Ball RM *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual review of ecology and systematics*, **18**, 489-522.

Azzaroli-Puccetti ML (1987) On the hares of Ethiopia and Somalia and the systematic position of *Lepus whytei* Thomas*, 1894 [ie 1984] (Mammalia, Lagomorpha)*. *Accademia nazionale dei Lince,* **19**, 1-19.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution*, **16**, 37–48.

Barata M, Perera A, Martínez-Freiría F, Harris DJ (2012) Cryptic diversity within the Moroccan endemic day geckos Quedenfeldtia (Squamata: Gekkonidae): a multidisciplinary approach using genetic, morphological and ecological data. *Biological Journal of the Linnean Society*. **106,** 828– 850.

Bastos-Silveira C, Santos SM, Monarca R, Mathias MDL, Heckel G (2012) Deep mitochondrial introgression and hybridization among ecologically divergent vole species. *Molecular Ecology*, **21**, 5309–5323.

Bearder S, Butynski TM, De Jong Y (2008) *Galago senegalensis*. The IUCN Red List of Threatened Species 2008: e.T8789A12932627.

Begg K, Begg C Abramov A (2008) *Mellivora capensis*. The IUCN Red List of Threatened Species 2008: e.T41629A10522766.

Bi KE, Linderoth T, Vanderpool DAN, Good JM, Nielsen R (2013) Unlocking the vault : next-generation museum population genomics. *Molecular ecology*, ***22*****,** 6018-6032.

Bickford D, Lohman DJ, Sodhi NS *et al.* (2006) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**, 148–155.

Biju-Duval C, Ennafaa H, Dennebouy N *et al.* (1991) Mitochondrial DNA evolution in lagomorphs: Origin of systematic heteroplasmy and organization of diversity in European rabbits. *Journal of Molecular Evolution*, **33**, 92–102.

Blanquer-Maumont A, Crouau-Roy B (1995) Polymorphism, monomorphism, and sequences in conserved microsatellites in primate species. *Journal of molecular evolution*, ***41*****,** 492-497.

Blouin MS, Yowell CA, Courtney CH, Dame JB (1998) Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. Molecular Biology and Evolution **15,** 1719– 1727.

Bohling JH, Adams JR, Waits LP (2013) Evaluating the ability of Bayesian clustering methods to detect hybridization and introgression using an empirical red wolf data set. *Molecular Ecology*, **22**, 74–86.

Bohling JH, Waits LP (2011) Assessing the prevalence of hybridization between sympatric Canis species surrounding the red wolf (*Canis rufus*) recovery area in North Carolina. *Molecular Ecology*, **20**, 2142–2156.

Boratynski Z, Brito JC, Mappes T (2012) The origin of two cryptic species of African desert jerboas (Dipodidae: *Jaculus*). *Biological Journal of the Linnean Society* **105,** 435–445.

Bowen BW, Nelson WS, Avise JC (1993). A molecular phylogeny for marine turtles: trait mapping, rate assessment, and conservation relevance. *Proceedings of the National Academy of Sciences*, **90**, 5574-5577.

Briggs AW, Stenzel U, Johnson PLF *et al.* (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences*, ***104***, 14616-14621.

Brito JC, Godinho R, Martínez-Freiría F *et al.* (2014) Unravelling biodiversity, evolution and threats to conservation in the sahara-sahel. *Biological Reviews*, **89**, 215–231.

Brotherton P, Endicott P, Sanchez JJ *et al.* (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Research*, **35**, 5717–5728.

Brouat C, Tatard C, Bâ K *et al.* (2009) Phylogeography of the Guinea multimammate mouse (*Mastomys erythroleucus*): A case study for Sahelian species in West Africa. *Journal of Biogeography*, **36**, 2237–2250.

Brown RP, Suárez NM, Pestano J (2002) The Atlas mountains as a biogeographical divide in North–West Africa: evidence from mtDNA evolution in the Agamid lizard *Agama impalearis*. *Molecular Phylogenetics and Evolution*, **24,** 324–332.

Cardoso A, Vogler AP (2005) DNA taxonomy, phylogeny and Pleistocene diversification of the *Cicindela hybrida* species group (Coleoptera: Cicindelidae). Mol Ecol **14,** 3531–3546.

Caroll RC (1988). *Vertebrate paleontology and evolution*. New York: Freeman and Company.

Chapman JA, Flux JEC (1990). Rabbits, Hares and Pikas. Status Survey and Conservation Action Plan. IUCN/SSC Lagomorph Specialist Group, Gland, Switzerland.

Chapman JA, Flux JEC (2008) Introduction to the Lagomorpha. In *Lagomorph Biology,* 1-9. Springer Berlin Heidelberg, Germany.

Cheng E, Hodges KE, Melo-Ferreira J, Alves PC, Mills LS (2014) Conservation implications of the evolutionary history and genetic diversity hotspots of the snowshoe hare. *Molecular ecology*, **23**, 2929–42.

Cheylan G (1991) Patterns of Pleistocene turnover, current distribution and speciation among Mediterranean mammals. *Biogeography of Mediterranean invasions*: 227–262. Groves, R.H. & DiCastri, F. (Eds). Cambridge: Cambridge Univ. Press

Claussen M. (2009). Late Quaternary vegetation-climate feedbacks. *Climate in the Past* **5,** 203-216.

Cloudsley-Thompson JL (1991) *Ecophysiology of desert arthropods and reptiles*. Springer Verlag GmbH & Co.

Coe MJ, Skinner JD (1993) Transactions of the Royal Society of South Africa. *Connections, disjunctions and endemism in the eastern and southern mamma*l **48**, 233-255.

Colbert EH (1980). *Evolution of the vertebrates*, 3rd ed. New York: John Wiley & Sons.

Collins K, Kryger U, Matthee C *et al.* (2008) *Lepus saxatilis*. The IUCN Red List of Threatened Species 2008: e.T41285A10433476.

Cooke HBS (1963) Pleistocene mammal faunas of Africa, with particular reference to southern Africa. *African ecology and human evolution,* **36,** 65-116.

Cooper A (1994) DNA from museum specimens. In: Herrmann B, Herrmann S (eds) *Ancient DNA: recovery and analysis of genetic material from paleontological, archeological, museum, medical and forensic specimens.* Springer, New York

Corander J, Sirén J, Arjas E (2008) Bayesian spatial modeling of genetic population structure. *Computational Statistics*, **23**, 111–129.

Coulon A, Fitzpatrick JW, Bowman R *et al.* (2008) Congruent population structure inferred from dispersal behaviour and intensive genetic surveys of the threatened Florida scrub-jay (*Aphelocoma caerulescens*). *Molecular Ecology*, **17**, 1685–1701.

Culver M, Menotti-Raymond MA, O'Brien SJ (2001) Patterns of size homoplasy at 10 microsatellite loci in pumas (*Puma concolor*). *Molecular biology and evolution*, **18**, 1151–1156.

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **360**, 1905–1916.

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

De Villiers M, Hirtle S (2003). *Sahara: The Extraordinary History of the World's Largest Desert.* Walker & Company.

Dobigny G, Aniskin V, Granjon L, Cornette R, Volobouev V (2005) Recent radiation in West African Taterillus (Rodentia, Gerbillinae): the concerted role of chromosome and climatic changes. *Heredity*, **95**, 358–368.

Dobigny G, Tatard C, Gauthier P *et al.* (2013) Mitochondrial and nuclear genes-based phylogeography of *Arvicanthis niloticus* (Murinae) and sub-saharan open habitats pleistocene history. *PLoS ONE*, **8**.

Douady CJ, Catzeflis F, Raman J, Springer MS, Stanhope MJ (2003) The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews). *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 8325–8330.

Drew C, O'Donovan D, Simkins G et al. (2008) *Lepus capensis*. The IUCN Red List of Threatened Species 2008: e.T41277A10429185.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS biology*, **4,** 699.

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

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**, 2611–2620.

Ewer RF, Cooke HBS (1964). The Pleistocene mammals of southern Africa. *Ecological studies in southern Africa. 35–48.* Davis, D. H. S., DeMeillon,B., Harington, J. S.&Kalk, M. (Eds). The Hague: Dr.W. Junk Publishers.

Fahey AL, Ricklefs RE, Dewoody JA (2014) DNA-based approaches for evaluating historical demography in terrestrial vertebrates. *Biological Journal of the Linnean Society*, ***112*****,** 367-386.

Faleh BAR, Granjon L, Tatard C *et al.* (2012) Phylogeography of the Greater Egyptian Jerboa (*Jaculus orientalis*) (Rodentia: Dipodidae) in Mediterranean North Africa. *Journal of Zoology*, **286**, 208–220.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Fickel J (2003) Molekularbiologie und Phylogenie von *Lepus europaeus* Pallas, 1778– und anderen europäischen *Lepus*-Arten. **3,** 27–34. *in* J. Niethammer and F. Krapp, editors. *Handbuch der Säugetiere Europas*, Lagomorpha, Aula-Verlag, Wiebelsheim, Germany.

Fickel J, Schmidt A, Putze M et al. (2005) Genetic structure of populations of European brown hare: implications for management. *Journal of Wildlife Management*, **69**, 760-770.

Flot JF (2010) Seqphase: A web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources*, **10**, 162–166.

Flot JF, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes*, **6**, 627–630.

Flux JEC, Angermann R (1990) The Hares and Jackrabbits. *Rabbits, Hares and Pikas: Status Survey and Conservation Action Plan*, 61–94.

Frantz AC, Cellina S, Krier A., Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493–505.

Fritz U, Barata M, Busack SD *et al.* (2005) Impact of mountain chains, sea straits and peripheral populations on genetic and taxonomic structure of a freshwater turtle, *Mauremys leprosa* (Reptilia, Testudines, Geoemydidae). *Zoologica Scripta,* **35,** 97–108.

Frynta D, Palupcikova K, Bellinvia E *et al.* (2010) Phylogenetic relationships within the cahirinus-dimidiatus group of the genus Acomys (Rodentia: *Muridae*): new mitochondrial lineages from Sahara, Iran and the Arabian Peninsula. *Zootaxa* **266,** 46–56.

Funk DJ, Omland KE (2003) Frequency, Causes, and Consequences, with insights from Animal Mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 397–423.

Garrick RC, Sunnucks P, Dyer RJ (2010) Nuclear gene phylogeography using PHASE : dealing with unresolved genotypes , lost alleles , and systematic bias in parameter estimation. *BMC Evolutionary Biology*, ***10***, 118.

Glaw F, Köhler J, De la Riva I, Vieites DR, Vences M (2010) Integrative taxonomy of Malagasy treefrogs: combination of molecular genetics, bioacoustics and comparative morphology reveals twelve additional species of Boophis. *Zootaxa*, ***2383***, 82.

Godoy JA, Negro JJ, Hiraldo F, Donázar JA. (2004) Phylogeography, genetic structure and diversity in the endangered bearded vulture (*Gypaetus barbatus,* L.) as revealed by mitochondrial DNA. *Molecular Ecology*, **13**, 371–390.

Gompert Z, Nice CC, Fordyce JA, Forister ML, Shapiro AM (2006) Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molecular Ecology* **15,** 1759–1768.

Gonçalves D V, Brito JC, Crochet P *et al.* (2012) Molecular Phylogenetics and Evolution Phylogeny of North African Agama lizards (Reptilia : Agamidae) and the role of the Sahara desert in vertebrate speciation, **64**, 582–591.

Gordon DM, Clermont O, Tolley H, Denamur E (2008) Assigning *Escherichia coli* strains to phylogenetic groups: Multi-locus sequence typing versus the PCR triplex method. *Environmental Microbiology*, **10**, 2484–2496.

Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *Journal of Heredity*, **86**, 485–486.

Grant WS, Leslie RW (1993). Biochemical divergence and biogeography of anglerfish of the genus *Lophius* (Lophiiformes). *Journal of Zoology*, **231,** 465–485.

Green DM (2005) Designatable units for status assessment of endangered species. *Conservation Biology*, **19**, 1813–1820.

Grubb P (1978). Patterns of speciation in African mammals. *Bull Carnegie Museum of Natural History, **6***, 152-167.

Habel JC, Husemann M, Schmitt T *et al.* (2012) Microallopatry caused strong diversification in Buthus scorpions (Scorpiones: Buthidae) in the Atlas Mountains (NW Africa). *PLoS ONE*, **7**, 6–13.

Hall, Hall T (1999) BioEdit : a user-friendly biological sequence alignment editor and analysis program for Windows 95 / 98 / NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hamill RM, Doyle D, Duke EJ (2007) Microsatellite analysis of mountain hares (*Lepus timidus hibernicus*): low genetic differentiation and possible sex-bias in dispersal. *Journal of mammalogy*, ***88*,** 784-792.

Harris DJ ,Arnold EN, Thomas RH (1998) Rapid speciation, morphological evolution, and adaptation to extreme environments in South African sand lizards (*Meroles*) as revealed by mitochondrial gene sequences. Molecular Phylogenetics and Evolution **10,** 37–48.

Harrison DL, Bates PJJ (1991) The mammals of Arabia. Harrison Zoological Museum, Sevenoaks, UK. 2nd Edition. 354 pp.

Hasegawa M, Kishino H, Yano T (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

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

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular biology and evolution*, **27**, 570–80.

Hewitt GM (1996.) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247–276

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

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 183–195.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hickerson MJ, Carstens BC, Cavender-Bares J *et al.* (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**, 291–301.

Hoffmann RS (1993) Order Lagomorpha. In: Wilson, D.E., Reeder, D.M. (Eds.), *Mammal Species of the World - A Taxonomic and Geographic Reference*, second ed. John Hopkins University Press, Baltimore, 807-827

Hoffmann RS, Smith AT (2005). Order Lagomorpha. In: third ed Wilson, D.E., Reeder, D.A.M. (Eds.), Mammal Species of the World, vol. 1. J. Hopkins University Press, Baltimore, 185–211.

Hoshino A, Bravo J, Nobile P, Morelli K (2012) Microsatellites as tools for genetic diversity analysis. INTECH, Open Access Publisher.

Husemann M, Schmitt T, Stathi I, Habel JC (2012) Evolution and radiation in the scorpion Buthus elmoutaouakili loureno and qi 2006 (Scorpiones: Buthidae) at the foothills of the atlas mountains (North Africa). *Journal of Heredity*, **103**, 221–229.

Husemann M, Schmitt T, Zachos FE, Ulrich W, Habel JC (2014) Palaearctic biogeography revisited: evidence for the existence of a North African refugium for Western Palaearctic biota (B Riddle, Ed,). *Journal of Biogeography*, **41**, 81–94.

Igea J, Aymerich P, Fernández-González A *et al.* (2013) Phylogeography and postglacial expansion of the endangered semi-aquatic mammal Galemys pyrenaicus. *BMC evolutionary biology*, **13**, 115.

Jacobsen MW, Hansen MM, Orlando L *et al.* (2012) Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus spp.*). *Molecular ecology*, **21**, 2727–42.

Jarne P, Lagoda PJ (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424–429.

Jezkova T, Leal M, Rodríguez-Robles J (2013) Genetic drift or natural selection? Hybridization and asymmetric mitochondrial introgression in two Caribbean lizards (*Anolis pulchellus* and *Anolis krugi*). *Journal of Evolutionary Biology*, **26**, 1458–1471.

Kalinowski ST, Taper ML, Creel S (2006) Using DNA from non-invasive samples to identify individuals and census populations: An evidential approach tolerant of genotyping errors. *Conservation Genetics*, **7**, 319–329.

Kingdon J, Butynski TM, De Jong Y (2008) *Erythrocebus patas*. The IUCN Red List of Threatened Species 2008: e.T8073A12884516.

Kingdon J, Happold D, Butynski T et al (2013). Mammals of Africa (Vol. 1). A&C Black.

Kingdon, J (2013). Mammalian evolution in Africa. In: Kingdon JS, Happold D, Hoffmann M et al (Eds.), *The Mammals of Africa, Introductory chapters and Afrotheria*, Bloomsbury, Amsterdam, **1,** 75–100

Kocher TD, Thomas WK, Meyer A *et al.* (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 6196–6200.

Kowalski K, Rzebik-Kowalska B (1991) Mammals of Algeria. Wroclaw: Ossolineum.

Krishnan NM, Seligmann H, Stewart CB, De Koning PJ, Pollock DD (2004) Ancestral sequence reconstruction in primate mitochondrial DNA: Compositional bias and effect on functional inference. *Molecular Biology and Evolution*, **21**, 1871–1883.

Kryger U (2002) Genetic variation among South African hares (*Lepus* spp.) as inferred from mitochondrial DNA and microsatellites. PhD thesis, University of Pretoria, Pretoria.

Kryger U, Robinson TJ, Bloomer P (2004) Population structure and history of southern African scrub hares, *Lepus saxatilis*. *Journal of Zoology*, **263**, 121–133.

Lado S, Melo-Ferreira J, Brito JC, Alves PC (2013). Population genetics and systematics of North African hares (genus Lepus). BSc internship at CIBIO (unpublished).

Lacy RC (1997) Importance of Genetic Variation to the Viability of Mammalian Populations. *Journal of Mammalogy*, **78**, 320–335.

Lanier HC, Knowles LL (2012) Is recombination a problem for species-tree analyses? *Systematic Biology*, **61**, 691–701.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes JOE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, ***7,*** 295-302.

Latch EK, Scognamillo DG, Fike J, Chamberlain MJ, Rhodes OE (2008) Deciphering ecological barriers to North American river otter (Lontra canadensis) gene flow in the Louisiana landscape. *Journal of Heredity*, **99**, 265–274.

Laval G, Excoffier L (2004) SIMCOAL 2.0: A program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, **20**, 2485–2487.

Lavocat R (1978). Rodentia and Lagomorpha. In *Evolution of African mammals*: 84–89. Maglio VJ, Cooke HBS (Eds). Cambridge, MA: Harvard University Press.

Le Berre M (1989) Faune du Sahara. 1. Poissons, Amphibiens et Reptiles. Lechevalier, R. Chabaud, Paris.

Le Berre M (1990). Faune du Sahara. 2. Mammifères. Lechevalier, R. Chabaud, Paris.

Lorenzen ED, Heller R, Siegismund HR (2012) Comparative phylogeography of African savannah ungulates. *Molecular ecology*, **21**, 3656–70.

Maheswaran M (2004) History , Features History , Features and Applications Markers : Molecular Markers :

Mamuris Z, Moutou KA, Stamatis C, Sarafidou T, Suchentrunk F (2010) Y DNA and mitochondrial lineages in European and Asian populations of the brown hare (*Lepus europaeus*). *Mammalian Biology-Zeitschrift für Säugetierkunde*, ***75***, 233-242.

Marsjan P, Oldenbroek J (2007) Molecular Markers, a tool for exploring gene diversity. *The state of the world's animal genetic resources for food and agriculture. FAO research report*. FAO, Rome, 359-379

Matthee C, van Vuuren BJ, Bell D, Robinson TJ (2004) A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Systematic biology*, **53**, 433–447.

McNeely JA (2003) Biodiversity in arid regions: values and perceptions. Journal of Arid Environments **54**, 61–70.

Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, 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. *Molecular ecology*, **18**, 2643–58.

Melo-Ferreira J, Boursot P, Carneiro M *et al.* (2012) Recurrent introgression of mitochondrial DNA among hares (*Lepus spp.*) revealed by species-tree inference and coalescent simulations. *Systematic biology*, **61**, 367–81.

Melo-Ferreira J, Boursot P, Randi E *et al.* (2007) The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. *Molecular ecology*, **16**, 605–18.

Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC (2005) Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular ecology*, **14**, 2459–64.

Melo-Ferreira J, Farelo L, Freitas H *et al.* (2014a) Home-loving boreal hare mitochondria survived several invasions in Iberia: the relative roles of recurrent hybridisation and allele surfing. *Heredity*, **112**, 265–73.

Melo-Ferreira J, Lemos de Matos A, Areal H *et al.* (2015) The phylogeny of pikas (Ochotona) inferred from a multilocus coalescent approach. *Molecular Phylogenetics and Evolution*, **84**, 240–244.

Melo-Ferreira J, Seixas F a, Cheng E, Mills LS, Alves PC (2014b) The hidden history of the snowshoe hare, *Lepus americanus*: extensive mitochondrial DNA introgression inferred from multilocus genetic variation. *Molecular ecology*, **23**, 4617–30.

Menegon M, Loader SP, Marsden SJ *et al.* (2014) The genus Atheris (Serpentes : Viperidae) in East Africa : Phylogeny and the role of rifting and climate in shaping the current pattern of species diversity. *Molecular phylogenetics and evolution*, ***79*,** 12-22.

Meng C, Kubatko LS (2009) Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theoretical Population Biology*, **75**, 35–45.

Mengoni C (2011) Phylogeny and genetic diversity of Italian species of hares (genus *Lepus*) PhD thesis. Università di Bologna, Bologna.

Mitchell A (2015) Collecting in collections: a PCR strategy and primer set for DNA barcoding of decades-old dried museum specimens. *Molecular ecology resources*, ***15***, 1102-1111.

Moores R, Brown D, Martin R, Lees AC (2012) Status and identification of hares L*epus sp*. in Western Sahara. *Go-South Bul,* **9**, 126-130.

Mulvaney A, Castoe T, Ashton KG, Krysko KL, Parkinson CL (2005) Evidence of Population Genetic Structure within the Florida Worm Lizard, *Rhineura floridana* (Amphisbaenia: Rhineuridae). *Journal of Herpetology*, **39**, 118–124.

Myers N, Mittermeier R, Mittermeier CG, da Fonseca GB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–8.

Nabholz B, Glémin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals - The longevity hypothesis. *Molecular Biology and Evolution*, **25**, 120–130.

Nardi F, Carapelli A, Dallai R, Roderick GK, Frati F (2005) Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, Tephritidae). *Molecular Ecology*, **14**, 2729–2738.

Nicolas V, Quérouil S, Verheyen E *et al.* (2006) Mitochondrial phylogeny of African wood mice, genus Hylomyscus (Rodentia, Muridae): Implications for their taxonomy and biogeography. *Molecular Phylogenetics and Evolution*, **38**, 779–793.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Olson DM, Dinerstein E, Wikramanayake ED *et al.* (2001) Terrestrial Ecoregions of the World : A New Map of Life on Earth. *BioScience*, **51**, 933–938.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Pääbo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 1939–1943.

Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. *Frontiers in zoology*, **7**, 16.

Palacios F, Angelone C, Alonso G, Reig S (2008) Morphological evidence of species differentiation within Lepus capensis Linnaeus, 1758 (Leporidae, Lagomorpha) in Cape Province, South Africa. *Mammalian Biology - Zeitschrift für Säugetierkunde*, **73**, 358–370.

Peakall R, Smouse PE (2012) GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, **28**, 2537–2539.

Perera A, Harris DJ (2010) Genetic variability in the ocellated lizard *Timon tangitanus* in Morocco. *African Zoology*, **45,** 321-329.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology & evolution,* **24,** 386-393.

Petter F (1959) Eléments d'une révision des lièvres africains du sous-genre *Lepus. Mammalia* **23**, 41–67.

Petter F (1972) Les lagomorphes du Maroc. *Société Des Sciences Naturelles et Physiques du Maroc* **52**, 122–129

Petter F 1961: Eléments d'une révision des lièvres europé ens et asiatiques du sous-genre *Lepus*. *Z. f. Säugetierkunde* **26**, 30–40.

Pierpaoli M, Riga F, Trocchi V, Randi E (1999) Species distinction and evolutionary relationships of the Italian hare (*Lepus corsicanus*) as described by mitochondrial DNA sequencing. *Molecular Ecolology* **8,** 1805–1817.

Pimm SL (2008). Biodiversity: climate change or habitat loss - which will kill more species? *Current Biology* **717**, 117-119.

Pook CE, Joger U, Stümpel N, Wüster W (2009) When continents collide: Phylogeny, historical biogeography and systematics of the medically important viper genus

Echis (Squamata: Serpentes: Viperidae). *Molecular Phylogenetics and Evolution*, **53**, 792–807.

Porras-Hurtado L, Ruiz Y, Santos C *et al.* (2013) An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Frontiers in Genetics*, **4**, 1–13.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pritchard JK, Wen W, Falush D (2003) Documentation for structure software: version 2.

Ravaoarimanana LB, Tiedemann R, Montagnon D, Rumpler Y (2004) Molecular and cytogenetic evidence for cryptic speciation within a rare endemic Malagasy lemur, the Northern Sportive Lemur (*Lepilemur septentrionalis*). *Molecular Phylogenetics and Evolution*, **31,** 440-448.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Robbins LW (1978) Karyotypic data for African mammals, with a description of an *in vivo* bone marrow technique. *Bulletin of Carnegie Museum of Natural History*, ***6***, 188-210.

Roberts A (1951). The Mammals of South Africa. T*rustees of the Mammals of South Africa Book Fund*, Johannesburg.

Robinson TJ, Matthee CA (2005). Phylogeny and evolutionary origins of the Leporidae: a review of cytogenetics, molecular analyses and a supermatrix analysis. *Mammal Review* **35,** 231–247

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

Rokas A, Carroll SB (2006) Bushes in the tree of life. *PLoS Biology*, **4**, 1899–1904.

Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: genomic hotspots of introgression between two highly divergent Ciona intestinalis species. *Molecular biology and evolution*, **30**, 1574–87.

Sefc KM, Payne RB, Sorenson MD (2006) Single base errors in PCR products from avian museum specimens and their effect on estimates of historical genetic diversity. *Conservation Genetics*, **8**, 879–884.

Sequeira F, Alexandrino J, Weiss S, Ferrand N (2008) Documenting the advantages and limitations of different classes of molecular markers in a well-established phylogeographic context: Lessons from the Iberian endemic Golden-striped salamander, *Chioglossa lusitanica* (Caudata: Salamandridae). *Biological Journal of the Linnean Society*, **95**, 371–387.

Sequeira F, Sodré D, Ferrand N *et al.* (2011) Hybridization and massive mtDNA unidirectional introgression between the closely related Neotropical toads *Rhinella marina* and *R. schneideri* inferred from mtDNA and nuclear markers. *BMC Evolutionary Biology*, **11**, 264.

Shimodaira H, Hasegawa M (1999) Letter to the Editor Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular biology and evolution*, ***16***, 1114-1116.

Shoubridge EA, Wai T (2007) Mitochondrial DNA and the mammalian oocyte. *Current topics in developmental biology*, ***77,*** 87-111.

Skinner JD, Smithers RHN (1990) The mammals of the South African subregion. 2nd edition. Pretoria: University of Pretoria.

Slimen H, Suchentrunk F, Memmi A *et al.* (2006) Evolutionary relationships among hares from North Africa (*Lepus sp. or Lepus spp*.), cape hares (*L. capensis)* from South Africa, and brown hares (*L. europaeus*), as inferred from mtDNA PCR-RFLP and allozyme data. *Journal of Zoological Systematics and Evolutionary Research*, **44**, 88–99.

Slimen H, Suchentrunk F, Memmi A, Ben Ammar Elgaaied A (2005) Biochemical genetic relationships among Tunisian hares (*Lepus sp*.), South African cape hares (*L. capensis*), and European brown hares (*L. europaeus*). *Biochemical genetics*, **43**, 577–96.

Smith AT, Johnston CH (2008) *Lepus microtis*. The IUCN Red List of Threatened Species 2008: e.T41879A10565789.

Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Success Stories in Implementation of the Programmes of Work on Dry and Sub-Humid Lands and the Global Taxonomy Initiative (2005):* ***164.***

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American journal of human genetics*, **76**, 449–462.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American journal of human genetics*, **68**, 978–989.

Strasburg JL, Rieseberg LH (2010) How robust are "isolation with migration" analyses to violations of the im model? A simulation study. *Molecular Biology and Evolution*, **27**, 297–310.

Stuart C, Stuart T, Hoffmann (2008) *Ictonyx striatus*. The IUCN Red List of Threatened Species 2008: e.T41646A10527451.

Suchentrunk F, Ben Slimen H, Kryger U (2009) Molecular evidence of conspecificity of South African hares conventionally considered Lepus capensis L., 1758. *Mammalian Biology - Zeitschrift für Säugetierkunde*, **74**, 325–343.

Suchentrunk F, Slimen HB, Sert H (2008). Phylogenetic aspects of nuclear and mitochondrial gene-pool characteristics of South and North African cape hares (*Lepus capensis*) and European hares (*Lepus europaeus).* In Lagomorph Biology (pp. 65-85). *Springer* Berlin Heidelberg.

Suchentrunk F, Slimen BH, Stamatis C *et al.* (2006) Molecular Approaches Revealing Prehistoric, Historic, or Recent Translocations and Introductions of Hares (Genus *Lepus*) by Humans. *Human Evolution*, **21**, 151–165.

Swofford LD (2001) PAUP* phylogenetic analysis using parsimony (* and other methods). Version 4.}.

Taberlet P, Griffin S, Goossens B *et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic acids research*, **24**, 3189–3194.

Talavera G, Vila R (2011) What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. *BMC evolutionary biology*, **11**, 315.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28,** 2731–2739.

Themudo GE, Rufino C, Campos PF (2014) Complete mitochondrial DNA sequence of the endangered giant sable antelope (Hippotragus niger variani): Insights into conservation and taxonomy. *Molecular phylogenetics and evolution*, **83**, 242–249.

Thorton PK, Jones PG, Owiyo T et al. (2008). Climate change and poverty in Africa: Mapping hotspots of vulnerability. *African Journal of Agricultural and Resource Economics* **2,** 24-44.

Thulin CG, Fang MY, Averianov O (2006) Introgression from *Lepus europaeus* to *L. timidus* in Russia revealed by mitochondrial single nucleotide polymorphisms and nuclear microsatellites. *Hereditas*, **143**, 68–76.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular ecology*, **21**, 3907–30.

Tolesa, ZG (2014) Evolutionary Relationships among Hares (*Lepus spp*.) from Ethiopia: Multivariate Morphometry, Molecular Phylogenetics and Population Genetics. *PhD dissertation. Addis Ababa University. Ethiopia.*

Trape S (2009). Impact of climate change on the relict tropical fish fauna of Central Sahara: threat for the survival of Adrar mountains fishes, Mauritania. *PLoS ONE **4**,* e4400.

Underhill P, Kivisild T (2007) Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual review of genetics*, **41**, 539–564.

UNEP (2006). Ecosystems and biodiversity in deep waters and high seas. UNEP Regional Seas Reports and Studies. Switzerland: UNEP/IUCN: 58.

Untergasser A, Nijveen H, Rao X *et al.* (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, **35**, 71–74.

Vallone PM, Butler JM (2004) AutoDimer: A screening tool for primer-dimer and hairpin structures. *BioTechniques*, **37**, 226–231.

Viard F, Franck P, Dubois MP, Estoup A, Jarne P (1998) Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *Journal of Molecular Evolution*, ***47,*** 42-51.

Vod R, Dapporto L, Dinc V, Vila R (2015) Why Do Cryptic Species Tend Not to Co-Occur ? A Case Study on Two Cryptic Pairs of Butterflies. *PloS one*, ***10***, 1–18.

Walker EP, Wamick F, Lange KI *et al.* (1964). Mammals of the World. John Hopkins Press, Baltimore, MD. **2**, 647-1500

Wallner B, Huber S, Achmann R (2001) Short communications-Non-invasive PCR sexing of rabbits (*Oryctolagus cuniculus*) and hares (*Lepus europaeus*). *Mammalian Biology-Zeitschrift fur Saugetierkunde*, **66**, 190.

Wan QH, Wu H, Fujihara T, Fang SG (2004) Which genetic marker for which conservation genetics issue? *Electrophoresis*, **25**, 2165–2176.

Wang ML, Barkley N, Jenkins TM (2009) Microsatellite markers in plants and insects. Part I: applications of biotechnology. *Genes, genomes and, genomics*, **3**, 54–67.

Wang Y, Notaro M, Liu Z et al. (2008). Detecting vegetation precipitation feedbacks in mid-Holocene North Africa from two climate models. *Climate in the Past* **4**, 59-67.

Wagner P, Melville J, Wilms TM, Schmitz A (2011) Opening a box of cryptic taxa—the first review of the North African desert lizards in the *Trapelus mutabilis* Merrem, 1820 complex (Squamata: Agamidae) with descriptions of new taxa. *Zoological Journal of the Linnean Society* **163**, 884–912.

Wheeler QD (2008). *The new taxonomy*. *CRC Press.*

Williams JS, Niedzwiecki JH, Weisrock DW (2013) Species tree reconstruction of a poorly resolved clade of salamanders (Ambystomatidae) using multiple nuclear loci. *Molecular Phylogenetics and Evolution*, **68**, 671–682.

Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851–1853.

Wu Y, Xia L, Zhang Q, Yang Q, Meng X (2011) Bidirectional introgressive hybridization between Lepus capensis and Lepus yarkandensis. *Molecular phylogenetics and evolution*, **59**, 545–55.

Wu C, Wu J, Bunch TD *et al* (2005) Molecular phylogenetics and biogeography of *Lepus* in Eastern Asia based on mitochondrial DNA sequences. *Molecular Phylogenetic Evolution* **37**, 45-61.

Yu X (2004) Molecular systematics of the genus *Lepus* in China. MSc thesis, Institute of Zoology, Chinese Academy of Sciences, Beijing.

Zwickl DJ (2008) GARLI manual (program version 0.96 , manual revision 1). *Time*.

Appendices

Appendix 1 – Dataset information (bone samples are highlighted in green and feces samples are highlighted in blue). LCP – *Lepus capensis;* LSX – *Lepus saxatilis;* LVT – *Lepus victoriae (=L. microtis); XXX* – no information. **---** Represents no amplification/no genotyping result. Poli – double peaks in the electropherogram. The species name was attributed by who sampled and not according to the results of this study.

	Code	Country	Extraction			Sequencing / Genotyping						
			Intership	This work	Available	Cytb	Microsatellite	HPX	SPTB	OXA1L	TF	DARC
1	LCP.SAF.1902	South Africa			X	X	X	---	X	---	X	---
2	LCP.SAF.1903	South Africa			X	X	X	---	X	---	X	---
3	LCP.SAF.1904	South Africa			X	---	---					
4	LCP.SAF.1905	South Africa			X	X	X	---	X	---	X	X
5	LCP.SAF.1906	South Africa			X	X	X	---	X	---	X	---
6	LCP.SAF.1907	South Africa			X	X	X	---	X	---	X	---
7	LCP.ISR.1910	Israel			X	X	X	X	X	X	X	X
8	LCP.ISR.1911	Israel			X	---	---	---	---	---	X	---
9	LCP.ISR.1912	Israel			X	X	X	X	X	X	X	X
10	LCP.ISR.1913	Israel			X	X	X	X	X	X	X	---
11	LCP.ISR.1914	Israel			X	X	X	X	X	X	X	X
12	LCP.ISR.1916	Israel			X	---	---	---	X	---	X	---
13	LCP.ISR.1918	Israel			X	X	X	X	X	X	X	X
14	LCP.KHE.1921	Morocco			X	X	X					
15	LCP.KHE.1922	Morocco			X	X	X					
16	LCP.KHE.1923	Morocco			X	X	X					
17	LCP.RAB.1925	Morocco			X	X	X					
18	LCP.RAB.1928	Morocco			X	X	X					
19	LCP.RAB.1929	Morocco			X	---	---					
20	LCP.RHE.1932	Morocco			X	X	X					
21	LCP.RHE.1934	Morocco			X	X	X					
22	LCP.RHE.1935	Morocco			X	X	X					
23	LCP.RHE.1936	Morocco			X	X	X					
24	LCP.RHE.1937	Morocco			X	X	X					
25	LCP.TAJ.1938	Morocco			X	X	X					
26	LCP.TET.1939	Morocco			X	X	X					
27	LCP.TET.1940	Morocco			X	X	X					
28	LCP.TET.1941	Morocco			X	X	X					
29	LCP.TET.1942	Morocco			X	X	X					
30	LCP.FOG.1947	Western Sahara			X	X	X					
31	LCP.HEC.1948	Mauritania			X	X	X	X	X	X	X	---
32	LCP.HEG.1949	Mauritania			X	X	X	X	X	---	X	X
33	LCP.TUN.1950	Tunisia		X		---	X					
34	LCP.TUN.1951	Tunisia			X	X	X					
35	LCP.TUN.1952	Tunisia			X	X	X					
36	LCP.TUN.1953	Tunisia			X	X	X					
37	LCP.OMA.2394	Oman		X		X	X	X	X	X	X	X
38	LCP.OMA.2395	Oman		X		X	X	X	X	X	X	X
39	LCP.CEU.2430	Spain	X			---	---	---	---	---	---	---
40	LCP.CEU.2431	Spain	X			---	---	---	---	---	---	---
41	LCP.CEU.2432	Spain	X			---	---	---	---	---	---	---
42	LCP.CEU.2433	Spain	X			---	---	---	---	---	---	---
43	LCP.CEU.2434	Spain	X			---	---	---	---	---	---	---
44	LCP.CEU.2435	Spain	X			---	---	---	---	---	---	---
45	LCP.CEU.2436	Spain	X			---	---	---	---	---	---	---

46	LCP.CEU.2437	Spain	X	---	---	---	---	---	---	---
47	LCP.CEU.2438	Spain	X	---	---	---	---	---	---	---
48	LCP.CEU.2439	Spain	X	---	---	---	---	---	---	---
49	LCP.CEU.2440	Spain	X	---	---	---	---	---	---	---
50	LCP.DAK.2778	Mauritania	X	X	X	X	X	X	X	X
51	LCP.JER.2975	Morocco	X	X	---					
52	LCP.OUA.2976	Morocco	X	X	X					
53	LCP.MEK.2977	Morocco	X	X	X	X	X	X	X	X
54	LCP.FES.2978	Morocco	X	X	X					
55	LCP.BOK.2982	Tunisia	X	X	---					
56	LCP.CHB.2983	Tunisia	X	X	---					
57	LCP.GAF.2986	Tunisia	X	---	X					
58	LCP.MZQ.2990	Libya	X	X	X	X	X	X	X	X
59	LCP.MZQ.2991	Libya	X	X	X	X	X	---	X	X
60	LCP.AGA.2992	Niger	X	X	X	X	X	---	X	---
61	LCP.SLO.2993	Senegal	X	X	X	X	X	---	X	X
62	LCP.ADR.2994	Mauritania	X	X	X	---	X	---	---	---
63	LCP.FOG.2995	WestSahara	X	X	X	X	X	X	X	---
64	LCP.DAK.2996	WestSahara	X	X	X	X	X	---	X	X
65	LCP.DAK.2997	WestSahara	X	X	X	X	X	---	X	X
66	LCP.BJD.2998	WestSahara	X	---	X					
67	LCP.BJD.2999	WestSahara	X	X	X					
68	LCP.LAA.3000	WestSahara	X	X	X	X	X	X	X	X
69	LCP.LAA.3001	WestSahara	X	X	X					
70	LCP.FOG.3002	WestSahara	X	---	X					
71	LCP.MAR.3003	Morocco	X	X	X	X	X	X	X	X
72	LCP.TAN.3004	Morocco	X	X	X	X	X	X	X	X
73	LCP.TAN.3005	Morocco	X	X	X					
74	LCP.TBT.3006	Mali	X	X	X	X	X	---	X	X
75	LCP.BJD.3007	WestSahara	X	X	X	X	X	---	X	X
76	LCP.DAK.3008	WestSahara	X	X	X					
77	LCP.GUI.3009	Mauritania	X	X	X	X	X	X	X	X
78	LCP.LAA.3010	WestSahara	X	X	X	X	X	---	X	X
79	LCP.DAK.3011	WestSahara	X	X	X	X	X	X	---	X
80	LCP.DAK.3012	WestSahara	X	X	X					
81	LCP.INC.3013	Mauritania	X	X	X	X	X	---	X	X
82	LCP.ASB.3015	Mauritania	X	---	X	X	X	X	X	X
83	LCP.SEH.3016	WestSahara	X	X	X					
84	LCP.ORO.3017	WestSahara	X	---	X					
85	LCP.ORO.3018	WestSahara	X	X	X					
86	LCP.TRZ.3019	Mauritania	X	X	X	X	X	X	X	X
87	LCP.DAK.3020	Mauritania	X	X	X	X	X	---	X	X
88	LCP.ORO.3021	WestSahara	X	X	X	X	X	X	X	X
89	LCP.KHE.3022	Morocco	X	X	X	X	X	---	X	X
90	LCP.TAN.3023	Morocco	X	X	X	X	X	---	X	X
91	LCP.ORO.3024	WestSahara	X	X	X	X	X	X	X	X
92	LCP.TIR.3025	Mauritania	X	X	---	---	---	---	---	---
93	LCP.ADR.3026	Mauritania	X	X	X	X	X	---	X	X
94	LCP.SEH.3027	WestSahara	X	X	X	X	X	X	---	X
95	LCP.SEH.3028	WestSahara	X	X	X	X	X	X	X	X
96	LCP.LOU.3030	Senegal	X	X	X	X	X	X	X	X
97	LCP.LOU.3031	Senegal	X	X	X					
98	LCP.KAI.3032	Tunisia	X	X	X	X	X	---	X	X
99	LCP.OUJ.3033	Morocco	X	X	X	X	X	---	---	X
100	LCP.NAD.3034	Morocco	X	X	---					
101	LCP.NAD.3035	Morocco	X	X	X	X	X	---	---	X
102	LCP.BOU.3036	Morocco	X	X	X	X	X	---	X	X
103	LCP.BOU.3037	Morocco	X	X	X	X	X	X	X	X

104	LCP.KEN.3038	Morocco	x				x	x	x	---	---	x	x
105	LCP.SLO.3039	Senegal	x				x	x	x	x	x	x	x
106	LCP.SLO.3040	Senegal	x				x	x	x	x	x	x	x
107	LCP.MIM.3041	Morocco		x			x	x					
108	LCP.MIM.3042	Morocco		x			x	x					
109	LCP.MIM.3043	Morocco		x			x	x					
110	LCP.MIM.3044	Morocco		x			x	x					
111	LCP.MIM.3045	Morocco		x			x	x					
112	LCP.MIM.3046	Morocco		x			x	x					
113	LCP.JMA.3047	Morocco		x			x	x					
114	LCP.JMA.3048	Morocco		x			x	x					
115	LCP.ROM.3049	Morocco		x			x	x					
116	LCP.BOU.3050	Morocco		x			x	---					
117	LCP.BOU.3051	Morocco		x			x	x					
118	LCP.EGA.3052	Morocco		x			x	x					
119	LCP.EGA.3053	Morocco		x			x	x					
120	LCP.EGA.3054	Morocco		x			x	x					
121	LCP.EGA.3055	Morocco		x			x	x					
122	LCP.EGA.3056	Morocco		x			x	x					
123	LCP.EGA.3057	Morocco		x			x	x					
124	LCP.SEH.3058	Morocco		x			x	x					
125	LCP.SEH.3059	Morocco		x			x	x					
126	LCP.TID.3061	Morocco			x		x	x					
127	LCP.BEH.3063	Morocco			x		x	x					
128	LCP.MIM.3066	Morocco			x		x	x					
129	LCP.MIM.3067	Morocco			x		x	x					
130	LCP.MIM.3068	Morocco			x		x	x					
131	LCP.BAH.3069	Morocco			x		x	x					
132	LCP.EGA.3070	Morocco			x		x	x					
133	LCP.EGA.3071	Morocco			x		x	x					
134	LCP.SFA.3072	Morocco			x		x	x					
135	LCP.ESS.3075	Morocco			x		x	x					
136	LCP.TAN.3124	Morocco	x				x	x	x	x	---	x	x
137	LCP.ORO.3125	WestSahara	x				x	x	x	x	x	x	x
138	LCP.OUA.3126	Morocco	x				x	x	x	---	x	x	x
139	LCP.IRA.3292	Iran		x			x	x	x	x	x	x	x
140	LCP.IRA.3293	Iran		x			x	x	x	x	x	x	x
141	LCP.XXX.3318	Angola		x			x	x	x	x	x	x	---
142	LCP.XXX.3456	Marocco		x			x	x					
143	LCP.TAN.3457	Marocco		x			x	x					
144	LCP.ORO.3458	WestSahara		x			x	x					
145	LCP.ORO.3459	WestSahara		x			x	x					
146	LCP.ORO.3460	WestSahara					x	---					
147	LCP.XXX.3461	Mali		x			x	x	x	x	x	x	x
148	LCP.ESS.3462	Marocco		x			x	x					
149	LCP.BOU.3463	Marocco		x			x	x					
150	LCP.NAD.3464	Marocco		x			x	x					
151	LCP.TAZ.3465	Marocco		x			x	---					
152	LCP.OUJ.3466	Marocco		x			x	x					
153	LCP.AZI.3467	Marocco		x			x	x					
154	LCP.XXX.3487	Marocco		x			x	x					
155	LCP.XXX.3488	Marocco		x			x	x					
156	LCP.TAN.3489	Marocco		x			x	x					
157	LCP.SDA.3598	S. Arabia		x			x	x	x	x	x	x	x
158	LCP.SDA.3599	S. Arabia		x			x	x	x		x	x	x
159	LCP.SDA.3600	S. Arabia		x			x	x	x	x	x	x	x
160	LCP.SDA.3601	S. Arabia		x			x	x	x	x	x	x	x
161	LCP.SDA.3602	S. Arabia		x			x	x	x	x	x	x	x

162	LCP.SDA.3603	S. Arabia	x		x	x	x		---	x	x
163	LCP.SDA.3604	S. Arabia	x		x	x	x	x	x	x	x
164	LCP.SDA.3605	S. Arabia	x		x	x	x		x	x	x
165	LCP.SDA.3606	S. Arabia	x		x	x	x	x	x	x	x
166	LCP.SDA.3607	S. Arabia	x		x	x	x		x	x	x
167	LCP.SDA.3608	S. Arabia	x		x	x	x	x	x	x	x
168	LCP.SDA.3609	S. Arabia	x		x	x	x	x	x	x	x
169	LCP.SDA.3610	S. Arabia	x		x	---					
170	LCP.ORO.3613	WestSahara	x		x	x					
171	LCP.ORO.3614	WestSahara	x		x	---					
172	LCP.TAN.3804	Marocco	x		x	x	x	x	---	x	x
173	LCP.TAN.3805	Marocco	x		x	x	x	x	x	x	x
174	LCP.TAN.3806	Marocco	x		x	---	x	---	---	x	x
175	LCP.SEH.3807	Western Sahara	x		x	---		---	x	x	x
176	LCP.SEH.3808	Western Sahara	x		x	---					
177	LSX.BTS.3492	Botswana		x	x	x	x	x	---	x	x
178	LSX.sax1				---	---	---	---	---	---	---
179	LSX.sax7				---	---	---	---	---	---	---
180	LSX.mz2				---	---	---	---	x	---	x
181	LSX.sax4				---	---	---	---	---	---	---
182	LSX.sax5				---	---	---	---	---	---	---

Museum samples

	Code	Museum	Country	Extraction			Cytb Seqeuncing	Microsat Genotyping
				Intership	This work	Available		
183	LCP.SAF.3468	MNHVienna	South Africa		x		x	x
184	LCP.SAF.3469	MNHVienna	South Africa		x		x	x
185	LCP.ISR.3470	MNHVienna	Israel		x		x	x
186	LCP.ISR.3471	MNHVienna	Israel		x		x	x
187	LCP.SAF.3472	MNHVienna	South Africa		x		---	---
188	LCP.SAF.3473	MNHVienna	South Africa		x		---	---
189	LCP.SOM.3474	MNHVienna	Somalia		x		---	---
190	LCP.TUN.3475	MNHVienna	Tunisia		x		poli	---
191	LCP.SAF.3476	MNHVienna	South Africa		x		---	---
192	LCP.SAF.3477	MNHVienna	South Africa		x		poli	x
193	LCP.SAF.3478	MNHVienna	South Africa		x		poli	x
194	LCP.SUD.3479	MNHVienna	Sudan		x		---	x
195	LCP.TUN.3480	MNHVienna	Tunisia		x		poli	---
196	LCP.TUN.3481	MNHVienna	Tunisia		x		---	---
197	LCP.TUN.3482	MNHVienna	Tunisia		x		poli	---
198	LCP.TUN.3483	MNHVienna	Tunisia		x		---	---
199	LCP.TUN.3484	MNHVienna	Tunisia		x		---	---
200	LCP.TUN.3485	MNHVienna	Tunisia		x		---	---
201	LCP.EGI.3486	MNHVienna	Egypt		x		---	---
202	LCP.EGI.3493	MNHBerlin	Egypt		x		---	---
203	LCP.ALG.3494	MNHBerlin	Algeria		x		x	x
204	LCP.SUD.3496	MNHBerlin	Sudan		x		---	---
205	LCP.SUD.3497	MNHBerlin	Sudan		x		x	x
206	LCP.LIB.3498	MNHBerlin	Lybia		x		---	---
207	LCP.SUD.3499	MNHBerlin	Sudan		x		---	---
208	LCP.ALG.3502	MNHBerlin	Algeria		x		---	---
209	LCP.ALG.3503	MNHBerlin	Algeria		x		x	x
210	LCP.MAR.3504	MNHBerlin	Morocco		x		---	---
211	LCP.EGI.3510	MNHBerlin	Egypt		x		x	x
212	LCP.KEN.3511	MNHBerlin	Kenya		x		x	x
213	LCP.KEN.3512	MNHBerlin	Kenya		x		x	x
214	LCP.ETI.3513	MNHBerlin	Eritrea		x		---	---

215	LCP.KEN.3514	MNHBerlin	Kenya	X	X	X
216	LCP.KEN.3515	MNHBerlin	Kenya	X	X	X
217	LCP.KEN.3516	MNHBerlin	Kenya	X	X	X
218	LCP.KEN.3517	MNHBerlin	Kenya	X	X	X
219	LCP.SOM.3518	MNHBerlin	Somalia	X	---	---
220	LCP.ERI.3519	MNHBerlin	Eritrea	X	poli	---
221	LCP.ETI.3520	MNHBerlin	Ethiopia	X	poli	---
222	LCP.KEN.3521	MNHBerlin	Kenya	X	X	X
223	LCP.ETI.3523	MNHBerlin	Eritrea	X	---	---
224	LCP.ETI.3524	MNHBerlin	Ethiopia	X	---	---
225	LCP.ORO.3525	MNHBerlin	West Sahara	X	poli	---
226	LCP.ERI.3533	MNHBerlin	Eritrea	X	---	---
227	LCP.ERI.3548	MNHBerlin	Eritrea	X	---	---
228	LCP.EGI.3578	MNHBerlin	Egypt	X	poli	---
229	LCP.EGI.3579	MNHBerlin	Egypt	X	---	---
230	LCP.XXX.3592	MNHBerlin	Egypt	X	poli	---
231	XXX.ALG.3495	MNHBerlin	Algeria	X	---	---
232	XXX.MAR.3500	MNHBerlin	Marocco	X	---	---
233	XXX.ALG.3501	MNHBerlin	Algeria	X	X	X
234	LVT.CMA.3505	MNHBerlin	Ivory Coast	X	---	---
235	LVT.CMA.3506	MNHBerlin	Ivory Coast	X	X	X
236	LVT.CMA.3507	MNHBerlin	Ivory Coast	X	---	---
237	LVT.CMA.3508	MNHBerlin	Ivory Coast	X	---	---
238	LVT.CMA.3509	MNHBerlin	Ivory Coast	X	---	---
239	LVT.TZN.3528	MNHBerlin	Tanzania	X	---	---
240	LVT.TZN.3529	MNHBerlin	Tanzania	X	---	---
241	LVT.TZN.3530	MNHBerlin	Tanzania	X	X	X
242	LVT.SNG.3531	MNHBerlin	Senegal	X	---	---
243	LVT.SNG.3532	MNHBerlin	Senegal	X	X	X
244	LVT.KEN.3534	MNHBerlin	Kenya	X	X	X
245	LVT.KEN.3535	MNHBerlin	Kenya	X	X	X
246	LVT.SNG.3536	MNHBerlin	Senegal	X	---	---
247	LVT.SNG.3537	MNHBerlin	Senegal	X	---	---
248	LVT.SNG.3538	MNHBerlin	Senegal	X	---	---
249	LVT.SNG.3539	MNHBerlin	Senegal	X	---	---
250	LVT.TZN.3540	MNHBerlin	Tanzania	X	X	X
251	XXX.REU.3563	MNHBerlin	Reunion island	X	---	---
252	XXX.KEN.3581	MNHBerlin	Kenya	X	---	---
253	LVT.TOG.3582	MNHBerlin	Togo	X	---	---
254	LVT.TOG.3583	MNHBerlin	Togo	X	X	X
255	XXX.NMB.3584	MNHBerlin	Namibia	X	---	---
256	LVT.TZN.3585	MNHBerlin	Tanzania	X	---	---
257	LVT.TZN.3586	MNHBerlin	Tanzania	X	---	---
258	LVT.GHA.3587	MNHBerlin	Ghana	X	X	X
259	LVT.TOG.3588	MNHBerlin	Togo	X	X	X
260	LVT.TOG.3589	MNHBerlin	Togo	X	X	X
261	LVT.TZN.3591	MNHBerlin	Tanzania	X	---	---

Appendix 2 – PCR-touchdown program conditions for the microsatellite multiplexes

Amplification step	Temp (ºC)	Time	Nº cycles
Initial denaturation	95	15'	1
Denaturation	95	30''	x40
Annealing	56-52	60''	
Extension	72	30''	
Denaturation	95	30''	x31
Annealing	52	60''	
Extension	72	30''	
Final extension	60	30'	1

Appendix 3 – Microsatellite multiplexes information

Multiplex	Number	Code	PCR product size	Primers Fw/Rv (5'-3')	Repeat number	Motif	Qt. of primers (50µl mix)	Tail
1	1	LCP_2	149	TGTAAAACGACGGCCAGTCCGTGAAGACTATCCTGATTCC GTTTGAGTGCATGTCTTGTTTAGCTCAT	15	ga	4	FAM
	2	LCP_3	244	TGTAAAACGACGGCCAGTAGGTAGGGGCAAAAGGAGAA GTTTCATCTGGAATGAAGATTGAACA	12	ag	1.5	
	3	LCP_4	92	TAATACGACTCACTATAGGGTTTACCTCTATTGCTGATTGCC GTTTCCTTGTCAGATGTAACTTTAACTAGA	12	ca	2.5	VIC
	4	LCP_6	240	TAATACGACTCACTATAGGGGGGTATGCCCCTCTAGTTCC GTTTGAAACCTAGAGAAAGATGGCA	12	ac	1.55	
	5	LCP_8	210	GATAACAATTTCACACAGGCTGAAGGCTGGGAGATCAAG GTTTCACCCACATGAGTTGGTCTG	19	ga	1.2	PET
	6	LCP_17	123	GATAACAATTTCACACAGGACAGACAGGTTGCCCAATGT GTTTACTGATGTGAGCACCACTGC	15	gag	0.6	
	7	LCP_9	91	TTTCCCAGTCACGACGTTGTCCTATTATCTTCTCCCTCCCTCT GTTTCAATGTGAGAGAGCAGCAGG	12	ca	0.4	NED
	8	LCP_10	237	TTTCCCAGTCACGACGTTGCCGTGTCCGTCCTTGATTAT GTTTCGATTTTCATTTGAACTGCC	12	ca	1.7	
2	1	LCP_23	105	TCACTAGTCTCCAACAGCAGTC CATTTCATTATCATATTGGATTCTCA	12	ca	1	FAM
	2	LCP_12	169	TAGCCCAGTCCCAGCTATTG GCTAGCTTTGGGGAAGAGAAA	12	tc	0.7	
	3	LCP_16	233	TCCACCGTCCTTTCTGTTTC GCCAGGACTCCAACTCATGT	13	tg	0.3	VIC
	4	LCP_26	95	TTGTGTCTAGCTGCTCCCCT ATGGCCTGGTTCCTGACTCT	14	ag	1	
	5	LCP_28	98	TTCTTTATGCTGCCAAATCG AGAAAGTCCCTCCCTCTCCA	14	tg	0.7	PET
3	1	LCP_33	195	CCAAGGATTTGAAATGTCTATGA CCCTTGAATCTTTGTTGGGA	13	ca	0.4	VIC
	2	LCP_34	123	GTTCTTTTCCCAGCCTCCTC AACCCACGCTGGTAAGTCTG	15	tg	0.4	
	3	LCP_18	195	GGAAATCAAGACGCAACTGG CCCGCATAATTGTAGGCACT	12	ca	0.7	PET
	4	LCP_37	184	AGCCAAATGCTCTACATGCC ATCATCCACTCCAACTCCCA	12	ac	0.8	NED
	5	LCP_38	112	GCAGTTACACATTCCTGATTTGA GCAGTTACACATTCCTGATTTGA	13	ac	0.3	

Appendix 4 – Primers and PCR conditions for the nuclear and mitochondrial sequenced genes

Loci		PCR				PCR primers	
Number	Symbol	$MgCl_2$[a]	AT[b]	E[c]	NC[d]	Fw/Rv (5'-3')	Reference
1	LcpCytb	1.8	58	40''	35x	ACATATTTGTCGAGACGTGAAC	**Lado *et al* (2013); unpublished**
						GCGTAGGCAAATAGGAAGTATC	**Lado *et al* (2013); unpublished**
2	LcpCytb2	1.8	57	60''	40x	CGGCTGACTTATTCGTTAC	**This work**
						GTTGAATGAATCTGAGGAGG	**This work**
3	DARC	1.8	58	40''	35x	CTCTCAGTTGACCCAAATTC	Melo-Ferreira *et al.* 2009
						GCCTTTAATTCAGGTTGACG	Melo-Ferreira *et al.* 2009
4	HPX	1.8	60	40''	35x	GTGGAGAATGCCAAGATGAAGG	Melo-Ferreira *et al.* 2009
						ACAGGATTGAAGCGGAGGAACT	Melo-Ferreira *et al.* 2009
5	SPTBN1	1.8	58	40''	35x	CTCTGCCCAGAAGTTTGCAAC	Matthee *et al.* 2004
						TGATAGCAGAACTCCATGTGG	Matthee *et al.* 2004
6	TF	1.8	60	40''	35x	GCCTTTGTCAAGCAAGAGACC	Wallner *et al.* 2001
						CACAGCAGCTCATACTGATCC	Wallner *et al.* 2001
7	OXA1L	1.8	58	45''	35x	TCGAATCAGAGAGGCCAAGTT	Melo-Ferreira *et al.* 2009
						ACCTGTCTCAGCACCTAGCTG	Melo-Ferreira *et al.* 2009

[a]Concentration (mM); [b]Annealing temperature; [c]Extension step length; [d]Number of cycles.

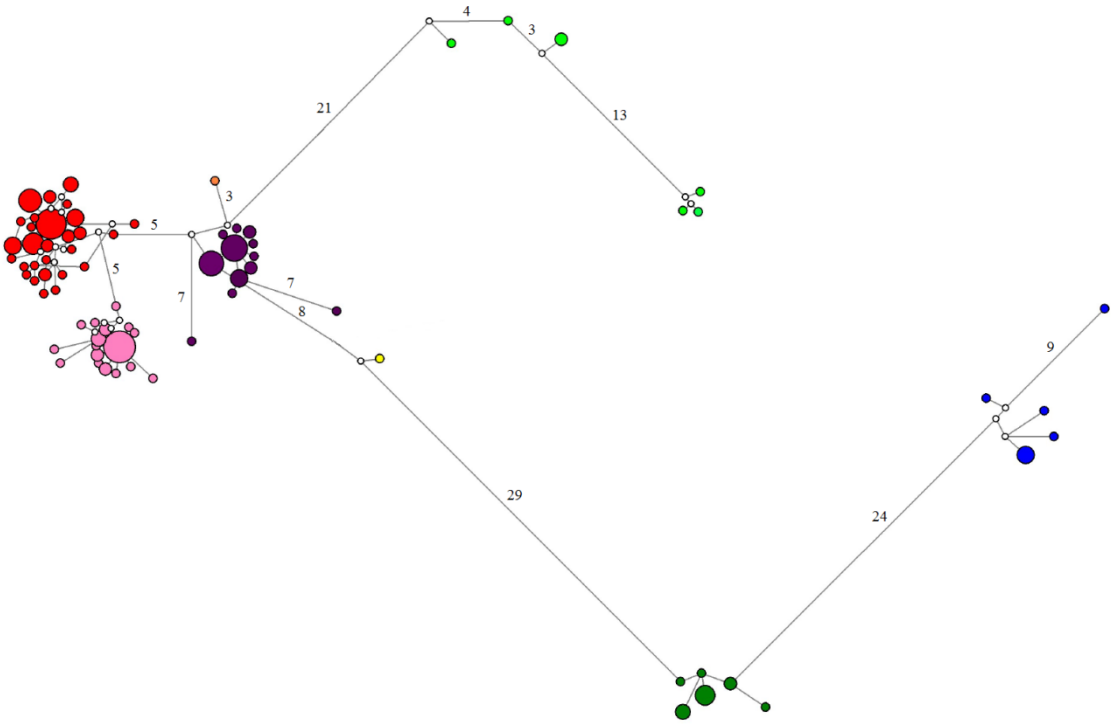
Appendix 5 – Genbank accession number for sequence datasets

Cytb		**TF**		**DARC**		**OXA1L**		**SPTBN1**		**HPX**	
Code	**Access. Nº**	**Code**	**Access. Nº**	**Code**	**Access. Nº**	**Code**	**Access. Nº**	**Code**	**Access. Nº**	**Code**	**Access. Nº**
Chi1	HM233085.1	Mar1	JN037061.1	Mar1	JN036923.1	Mar1	JN037141.1	Mar1	JN037035.1	Mar1	JN036951.1
Chi2	HM233083.1	Mar2	JN037062.1	Mar2	JN036924.1	Mar2	JN037142.1	Mar2	JN037036.1	Mar2	JN036952.1
Chi3	HM233013.1	Tun	JN037063.1	Tun	JN036925.1	Tun	JN037143.1	Tun	JN037037.1	Tun	JN036953.1
Chi4	HM233016.1	China1	HM233541.1	Leup1	JN036917.1	Leur1	JN037135.1	China1	HM233497.1	Leup1	JN036945.1
Chi5	HM233008.1	China2	HM233540.1	Leup2	JN036916.1	Leur2	JN037134.1	China2	HM233496.1	Leup2	JN036944.1
Chi6	HM233022.1	China3	HM233539.1	Leup3	FJ811758.1	Leur3	FJ811601.1	China3	HM233495.1	Leup3	FJ811733.1
Chi7	HM233039.1	Raba	AY176278.1	Leup4	FJ811757.1	Leur4	FJ811600.1	China4	HM233494.1	Leup4	FJ811732.1
Chi8	HM233019.1	Tet	AY176277.1	Leup5	FJ811756.1	Leur5	FJ811599.1	China5	HM233493.1	Leup5	FJ811731.1
Chi9	HM233082.1	Leur1	JN037056.1	Leup6	FJ811755.1	Leur6	FJ811598.1	Leur1	JN037029.1	Leup6	FJ811730.1
Chi10	HM233078.1	Leur2	JN037057.1	Leup7	FJ811754.1	Leur7	FJ811597.1	Leur1	JN037028.1	Leup7	FJ811725.1
Chi11	HM233042.1	Lsax	AY176279.1	Leup8	FJ811753.1	Leur8	FJ811596.1	Leur3	FJ811651.1	Lgra1	JN036943.1
Lsax1	AF009731.1	Leur3	FJ811626.1	Lgra1	JN036915.1	Lgra1	JN037133.1	Leur4	FJ811650.1	Lgra2	JN036942.1
Lsax2	AY292730.1	Leur4	FJ811625.1	Lgra2	JN036914.1	Lgra2	JN037132.1	Leur5	FJ811649.1	Lgra3	FJ811724.1
Lsax3	HQ596480.1	Leur5	FJ811624.1	Lgra3	FJ811752.1	Lgra3	FJ811595.1	Leur6	FJ811648.1	Lgra4	FJ811723.1
LE1	JN037344.1	Leur6	FJ811623.1	Lgra4	FJ811751.1	Lgra4	FJ811594.1	Leur7	FJ811647.1	Lgra5	FJ811722.1
LE2	JN037345.1	Leur7	FJ811622.1	Lgra5	FJ811750.1	Lgra5	FJ811593.1	Leur8	FJ811646.1	Lgra6	FJ811721.1
LE3	JN037346.1	Leur8	FJ811621.1	Lgra6	FJ811749.1	Lgra6	FJ811592.1	Leur9	FJ811645.1	Lgra7	FJ811720.1
LE4	JN037347.1	Leur9	FJ811620.1	Lgra7	FJ811748.1	Lgra7	FJ811591.1	Lgra1	JN037027.1	Lgra8	FJ811719.1
LE5	JN037348.1	Leur10	AY176267.1	Lgra8	FJ811747.1	Lgra8	FJ811590.1	Lgra2	JN037026.1	Ltim1	FJ811746.1
LE6	JN037349.1	Leur11	AY176266.1	Ltim1	FJ811770.1	Lgra9	FJ811589.1	Lgra3	FJ811644.1	Ltim2	FJ811745.1
LE7	AF010161.1	Leur12	AY176265.1	Ltim2	FJ811769.1	Lgra10	FJ811588.1	Lgra4	FJ811643.1	Ltim3	FJ811744.1

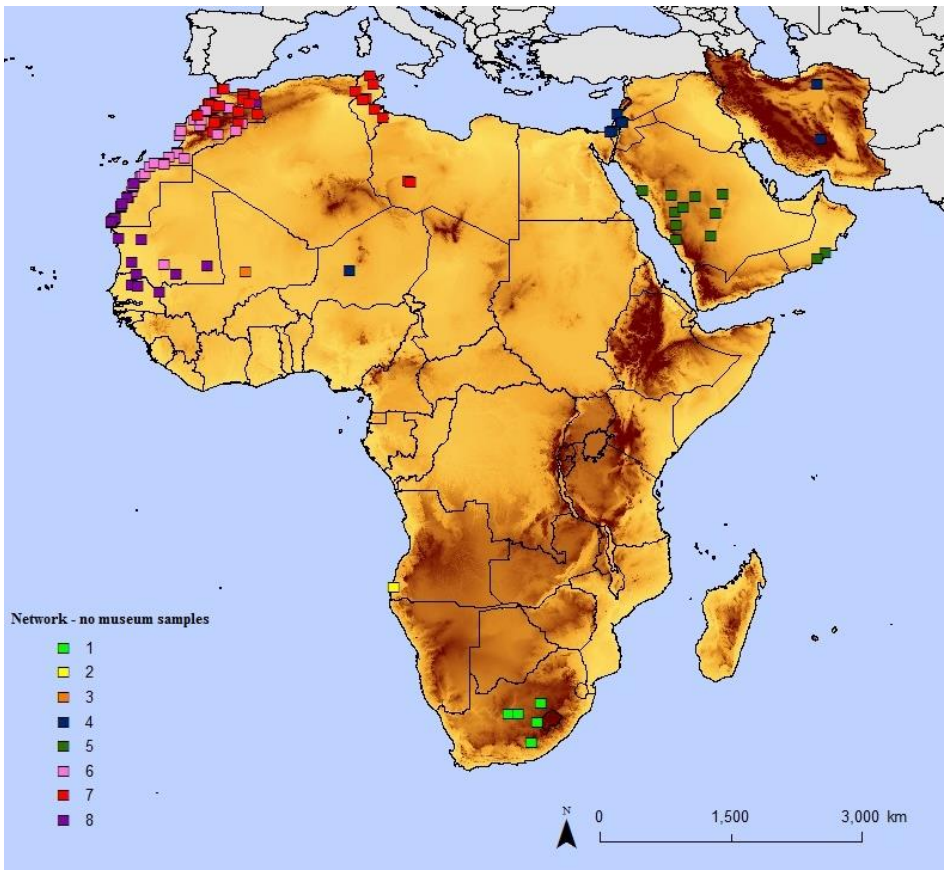
LE8	AF010162.1	Leur13	AY176264.1	Ltim3	FJ811768.1	Ltim1	FJ811613.1	Lgra5	FJ811642.1	Ltim4	FJ811743.1
LG1	JF298950.1	Leur14	AY176263.1	Ltim4	FJ811767.1	Ltim2	FJ811612.1	Lgra6	FJ811641.1	Ltim5	FJ811742.1
LG2	JF299027.1	Leur15	AY176262.1	Ltim5	FJ811766.1	Ltim3	FJ811611.1	Lgra7	FJ811640.1	Ltim6	FJ811741.1
LG3	JF298953.1	Lgra1	JN037055.1	Ltim6	FJ811765.1	Ltim4	FJ811610.1	Lgra8	FJ811639.1	Ltim7	FJ811740.1
LG4	JF298944.1	Lgra2	JN037054.1	Ocun	JN036940.1	Ltim5	FJ811609.1	Ltim1	FJ811663.1	Ocun	JN036968.1
LG5	JF298955.1	Lgra3	FJ811619.1			Ltim6	FJ811608.1	Ltim2	FJ811662.1		
LG6	JF298963.1	Lgra4	FJ811618.1			Ocun	JN037158.1	Ltim3	FJ811661.1		
LG7	JF298968.1	Lgra5	FJ811617.1					Ltim4	FJ811660.1		
LG8	JF298971.1	Lgra6	FJ811616.1					Ltim5	FJ811659.1		
LG9	JF299012.1	Lgra7	FJ811615.1					Ltim6	FJ811658.1		
LG10	JF299034.1	Lgra8	FJ811614.1					Ocun	JN037052.1		
LT1	DQ882893.1	Lgra9	EU196169.1								
LT2	DQ882890.1	Lgra10	AY176258.1								
LT3	DQ882906.1	Lgra11	AY176257.1								
LT4	DQ882899.1	Lgra12	AY176256.1								
LT5	DQ882911.1	Lgra13	AY176255.1								
LT6	DQ882915.1	Lgra14	AY176254.1								
LT7	DQ882936.1	Lgra15	AY176253.1								
LT8	DQ882937.1	Lgra16	AY176252.1								
LT9	DQ882944.1	Lgra17	AY176251.1								
LT10	DQ882957.1	Lgra18	AY176250.1								
Ocun	AJ001588.1	Lgra19	AY176249.1								
		Ltim1	FJ811638.1								
		Ltim2	FJ811637.1								
		Ltim3	FJ811636.1								
		Ltim4	FJ811635.1								
		Ltim5	FJ811634.1								
		Ltim6	FJ811633.1								
		Ltim7	AY176276.1								
		Ltim8	AY176275.1								
		Ltim9	AY176274.1								
		Ltim10	AY176273.1								
		Ltim11	AY176272.1								
		Ltim12	AY176271.1								
		Ltim13	HM233544.1								
		Ltim14	HM233535.1								
		Ltim15	HM233534.1								
		Ocun	AH005776.1								

Leur – *Lepus europaeus*; Lgra – *Lepus granatensis*; Ltim – *Lepus timidus*, Ocun – *Oryctolagus cuniculus*; the others are from *Lepus capensis*. Acess. Nº - accession number from GenBank.

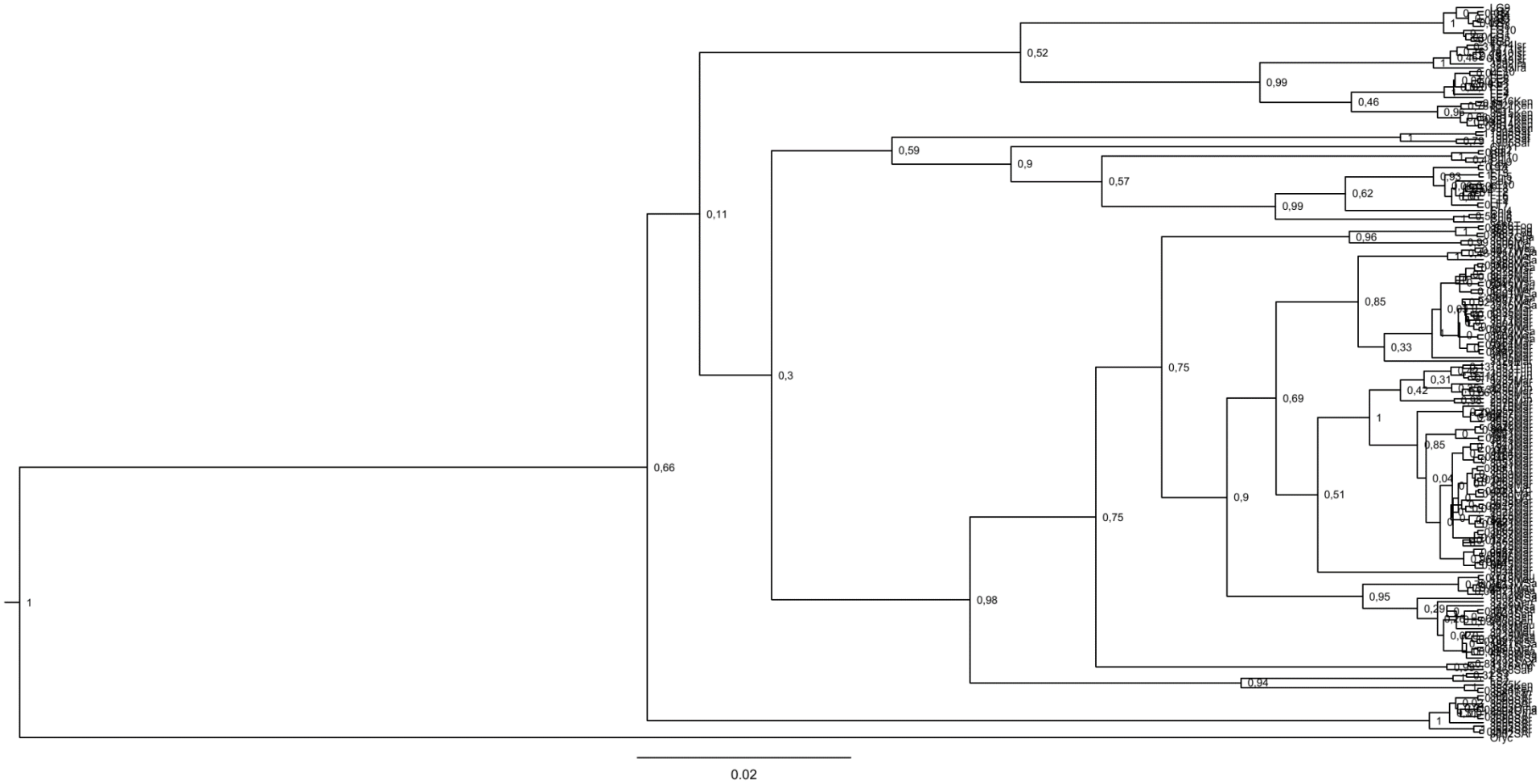
Appendix 6 – mtDNA network without museum samples.



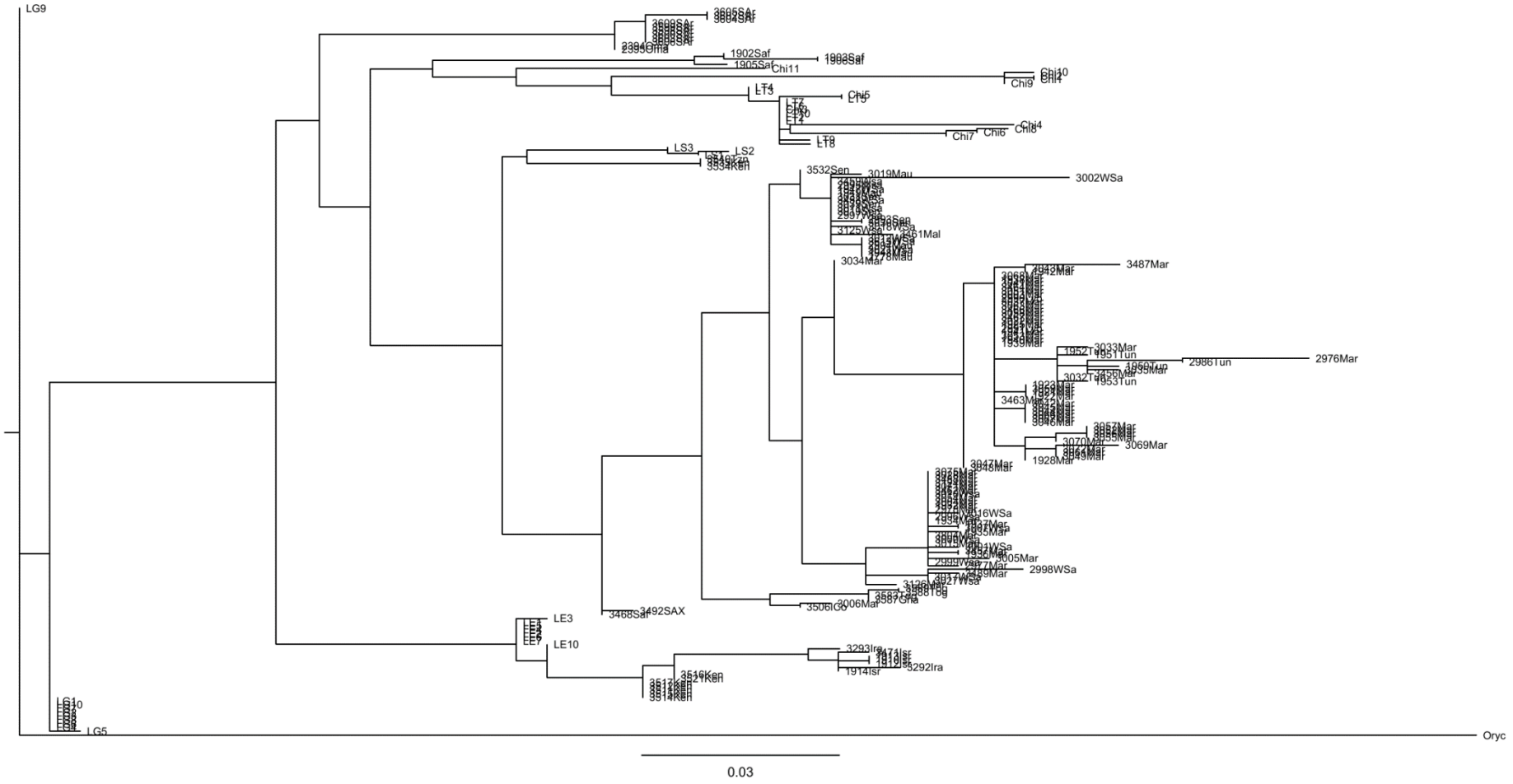
Appendix 7 – Representation of the Network clades in the map, from the dataset without museum samples.



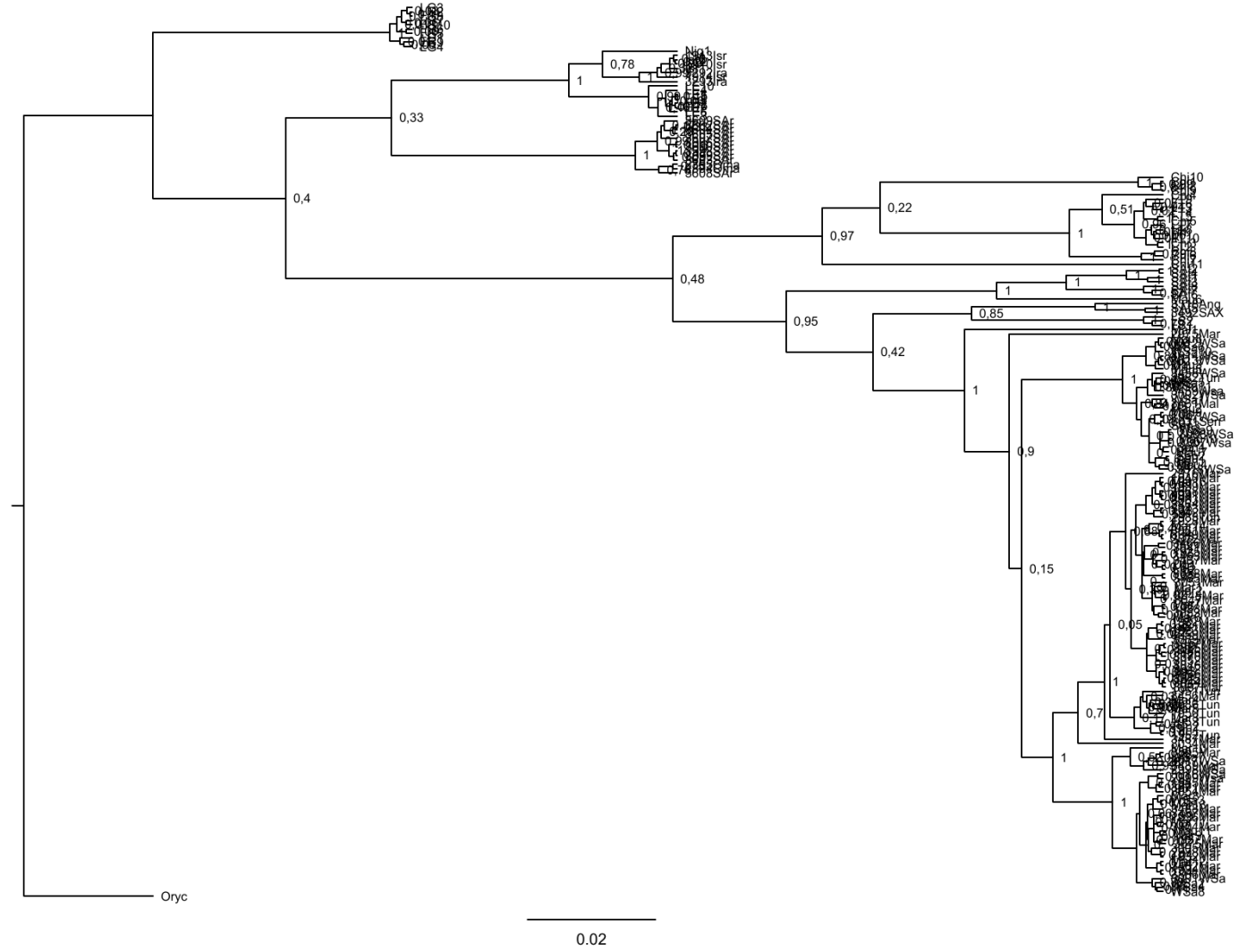
Appendix 8 – BSP Inference of the dataset with museum samples.



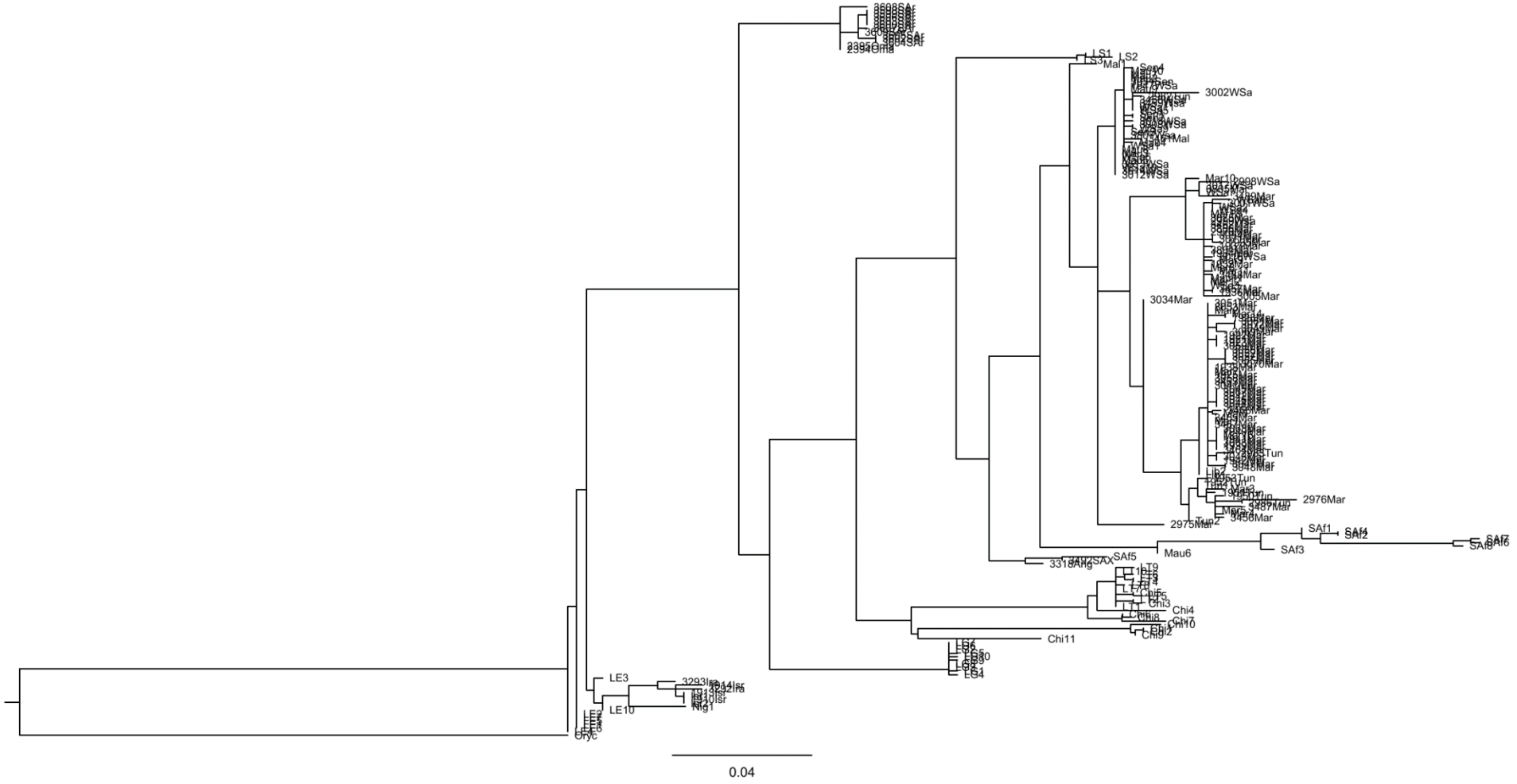
Maximum likelihood inference of the dataset with museum samples.



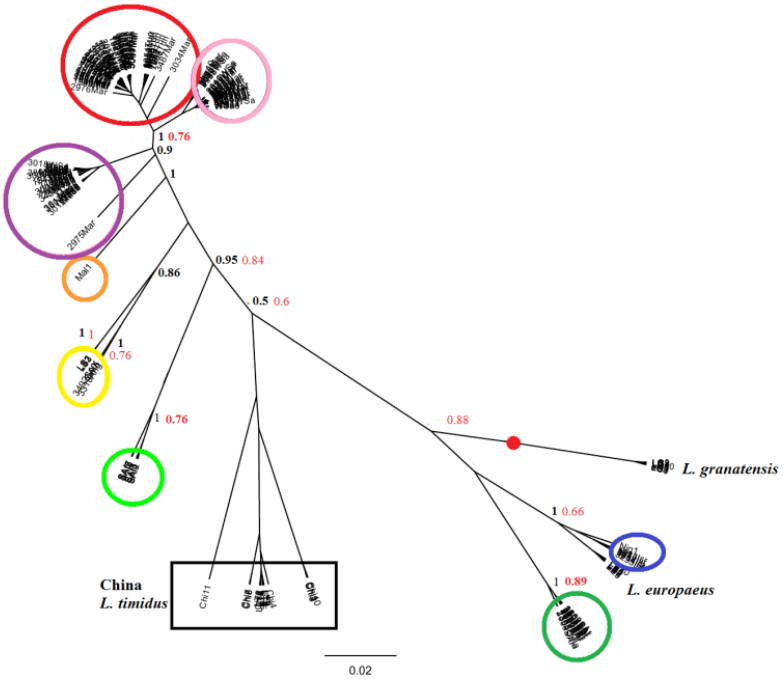
Appendix 9 - BSP Inference of the dataset without museum samples



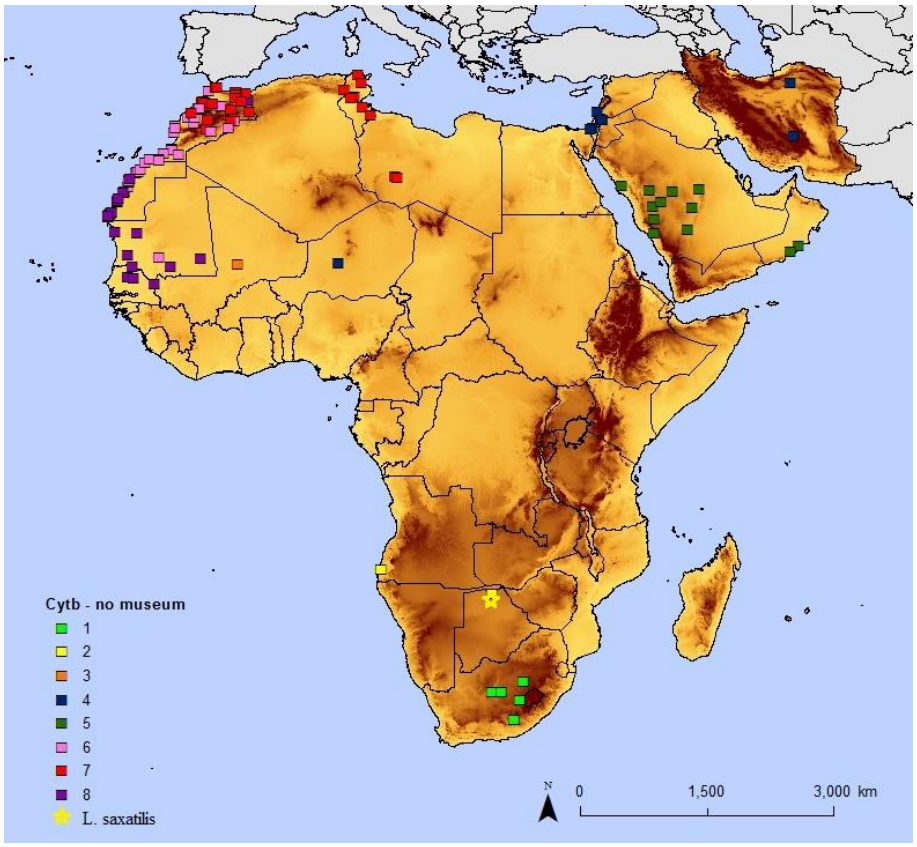
Maximum likelihood inference of the dataset without museum samples.



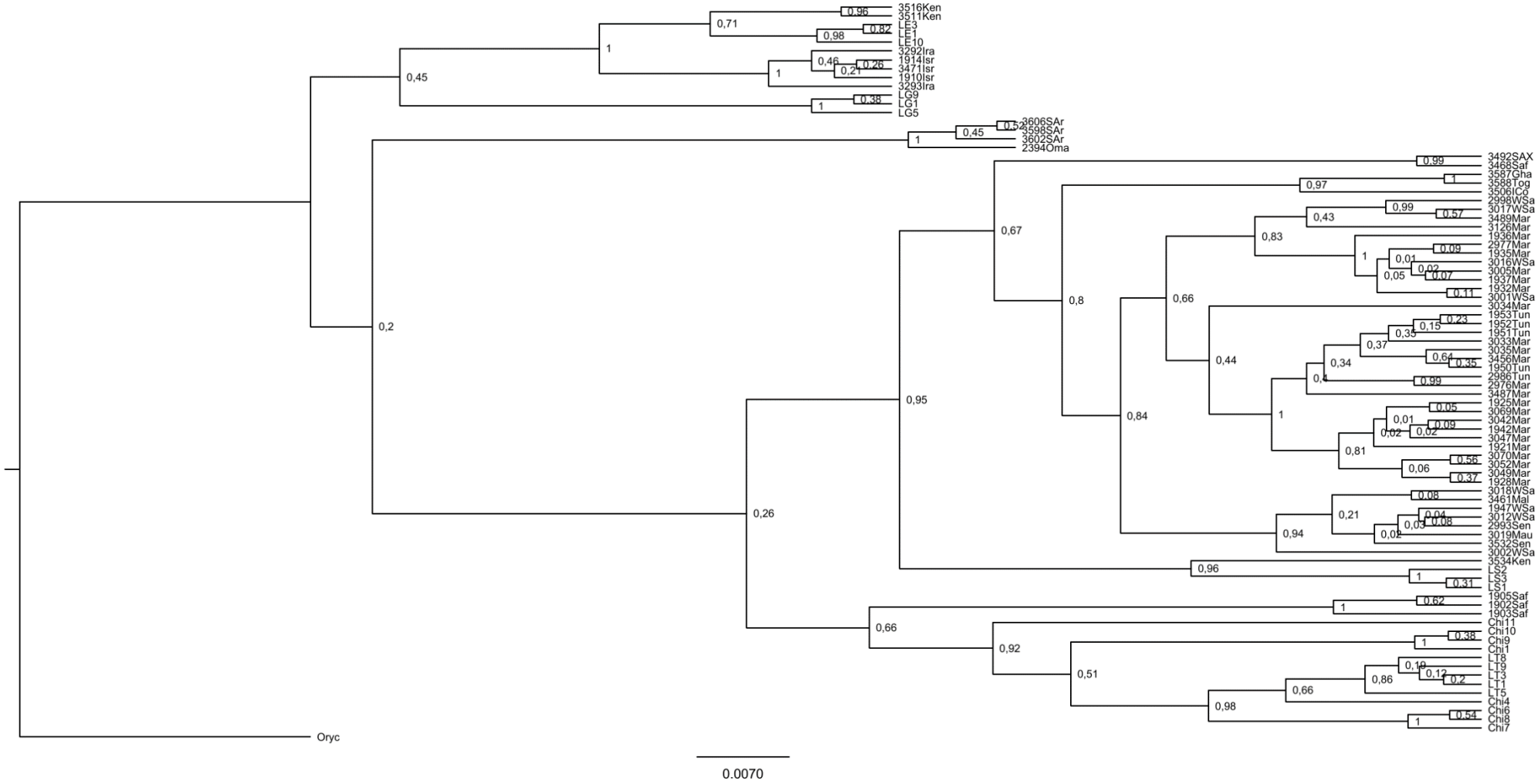
Appendix 10 – Unrooted tree with BI (black) and ML (red) phylogenetic combined results without museum samples. Just the supports of the nodes above 0.5 are represented. Each lineage is represented by a different color. Red dot corresponds to where the outgroup would appear.



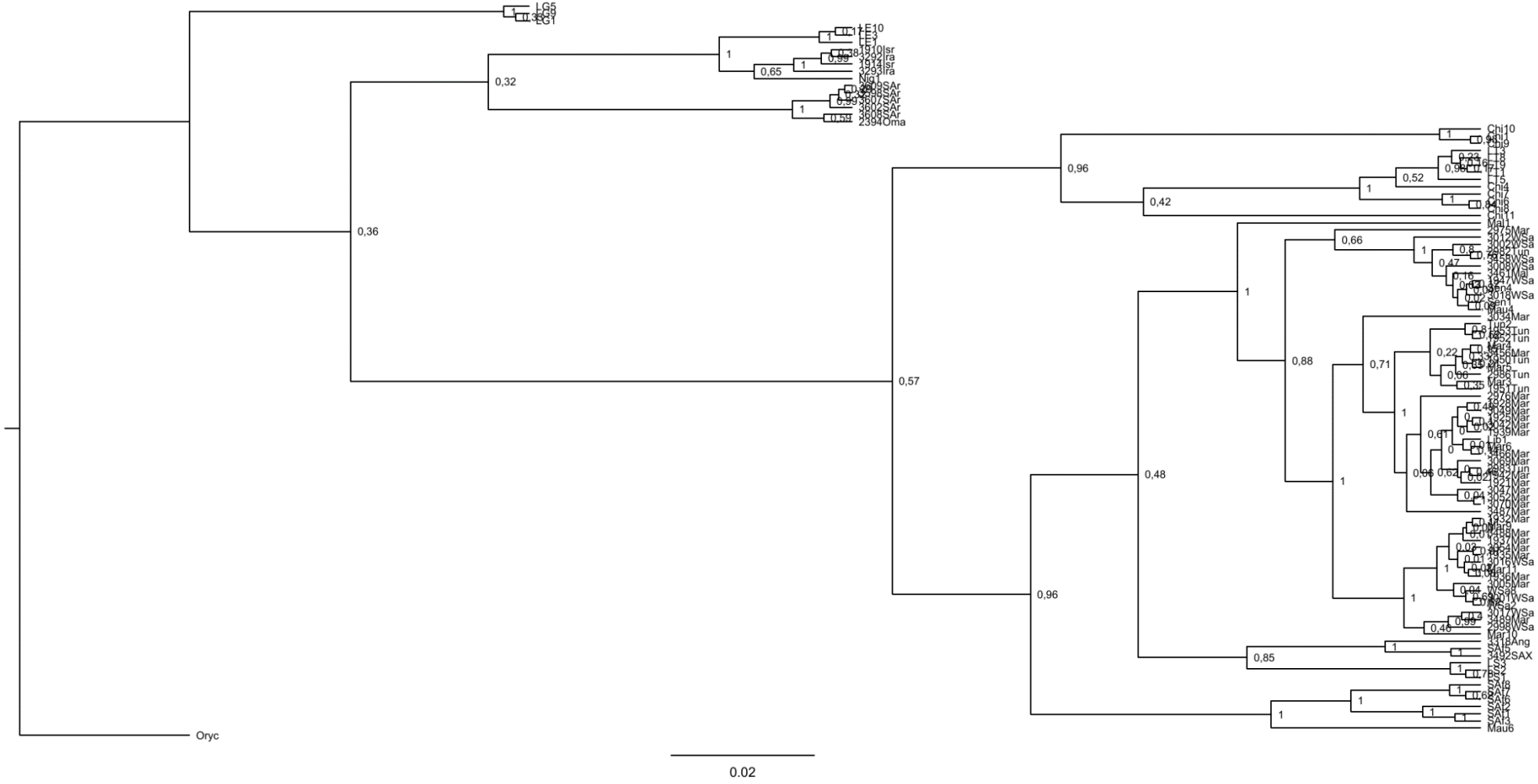
Appendix 11 – Map representation of the phylogeny result without museum samples.



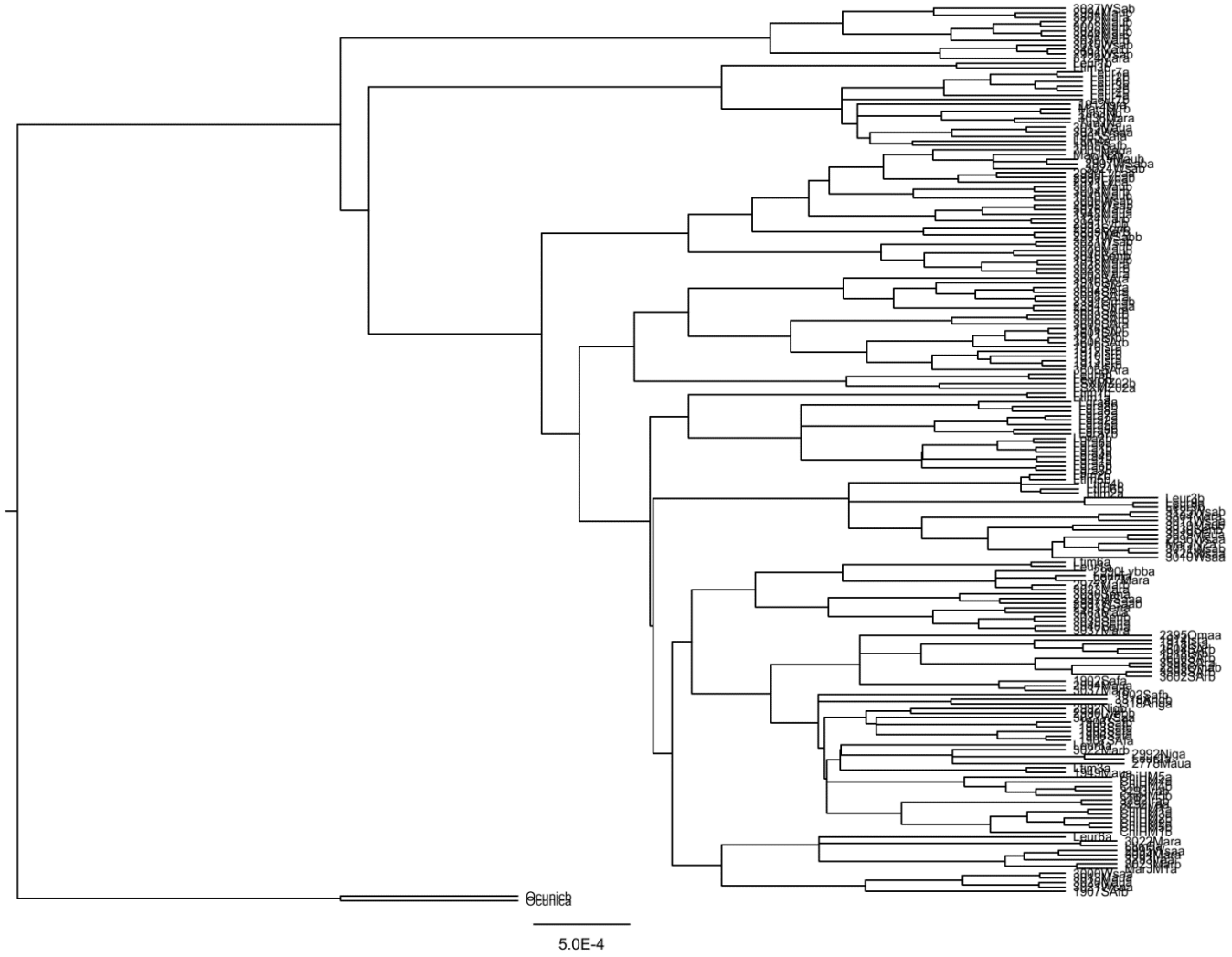
Appendix 12 – Bayesian inference with the YULE model, with the haplotypes corresponding to the dataset without museum samples, which confirms the principal phylogenetic clades.

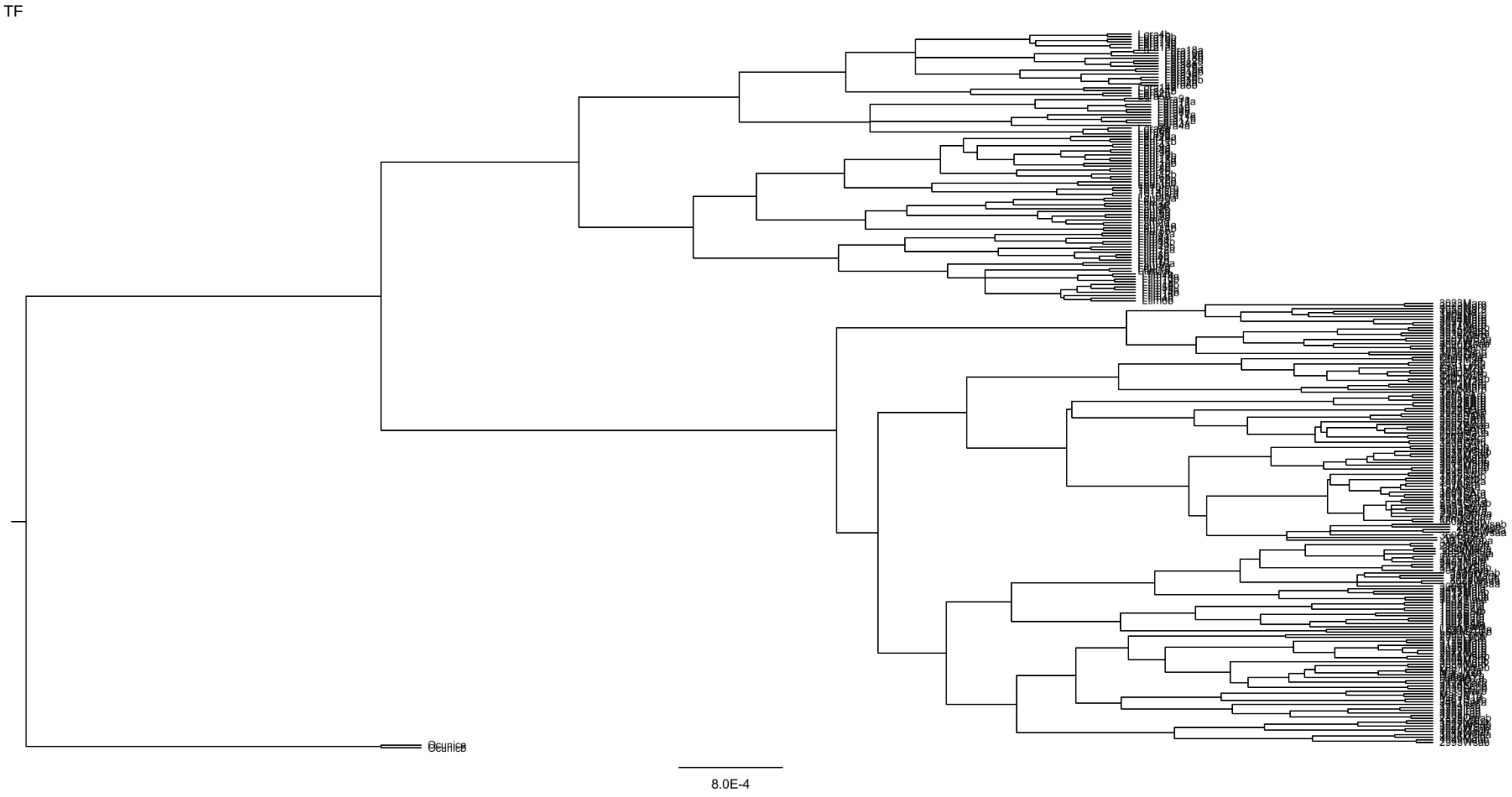


Bayesian inference with the YULE model, with the haplotypes corresponding to the dataset without museum samples, which confirms the principal phylogenetic clades.

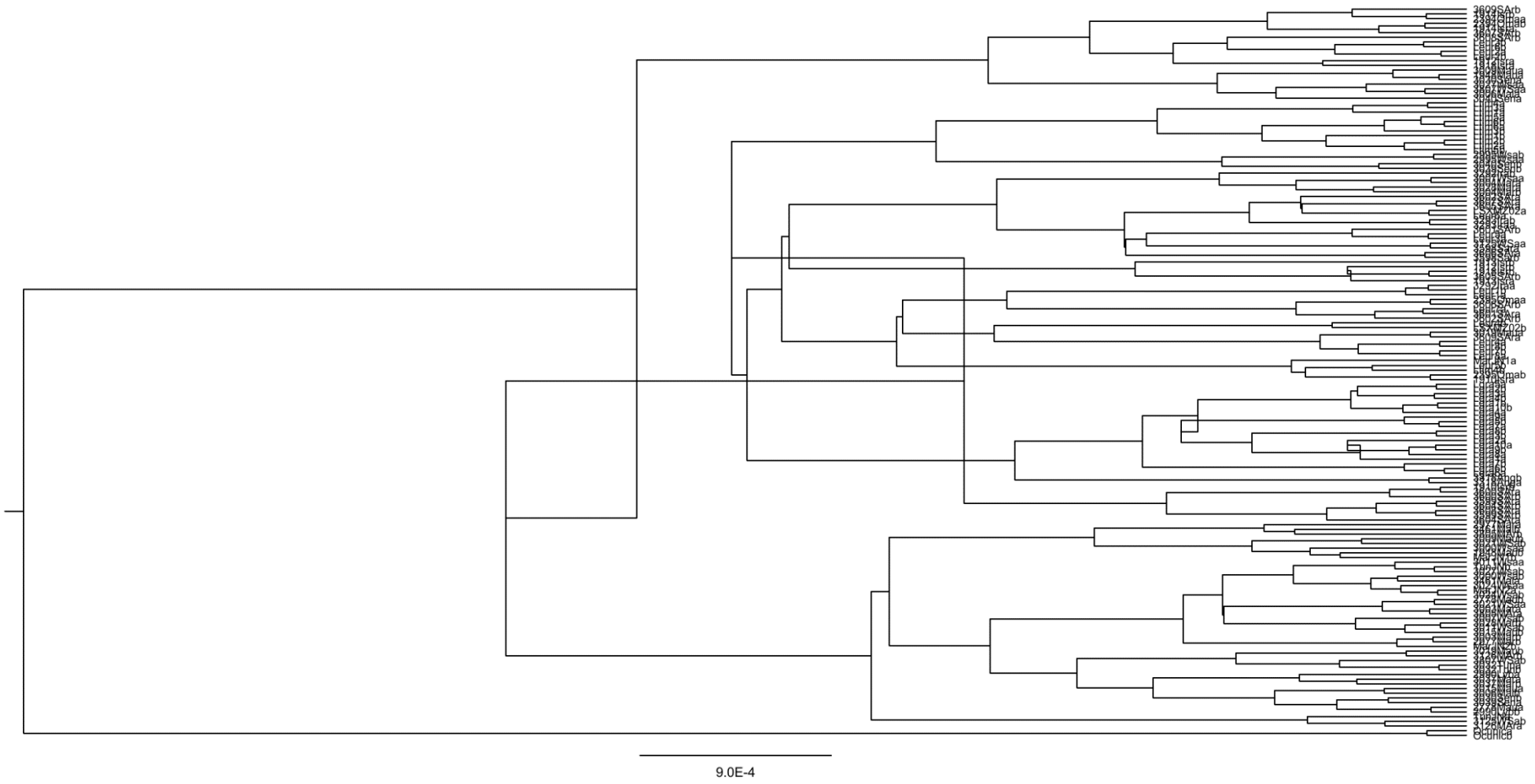


Appendix 13 – Nuclear gene tree with Bayesian inference. Posterior probabilities were low, not represented.
SPTBN1

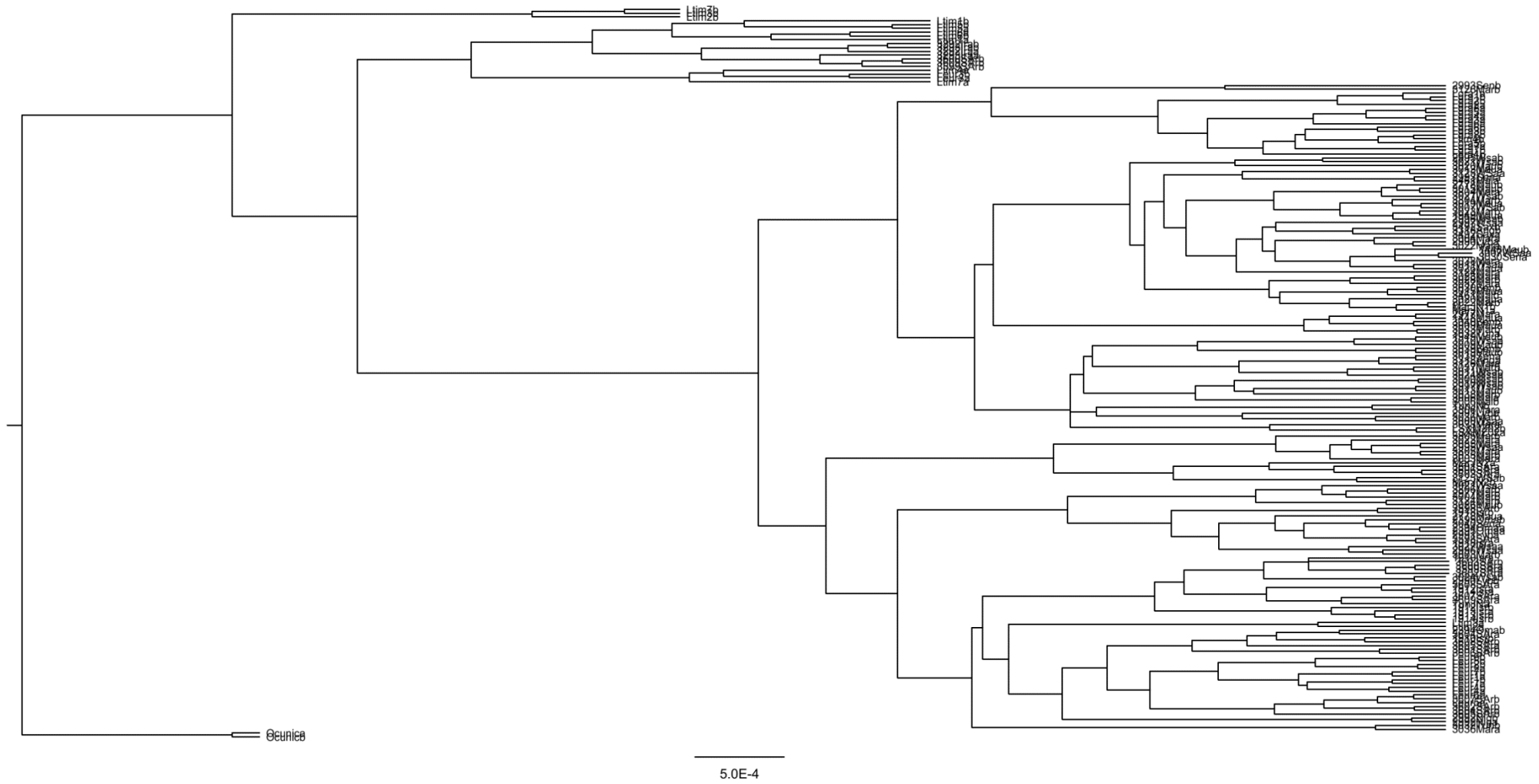



TF
Ocunicu
8.0E-4

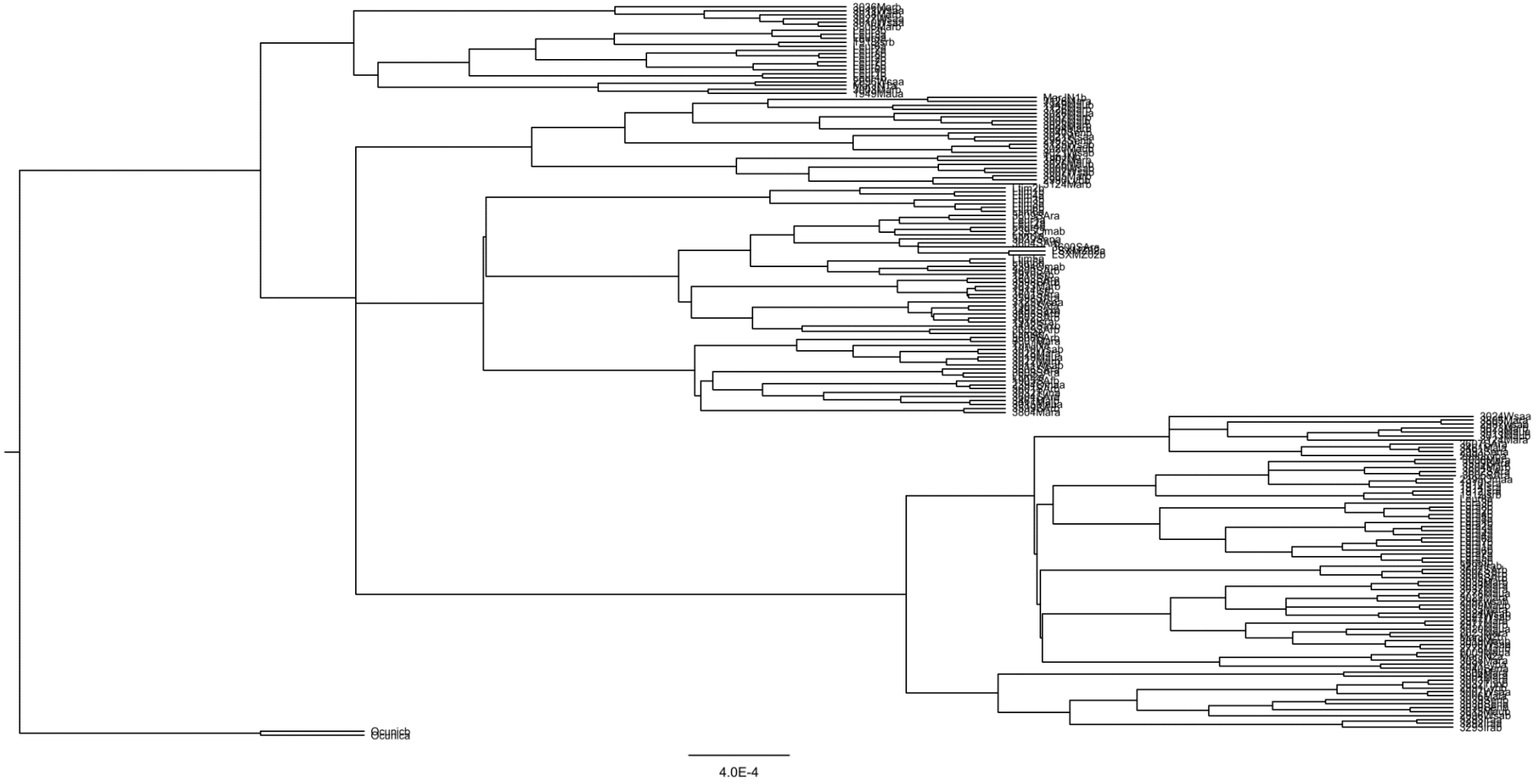
OXA1L



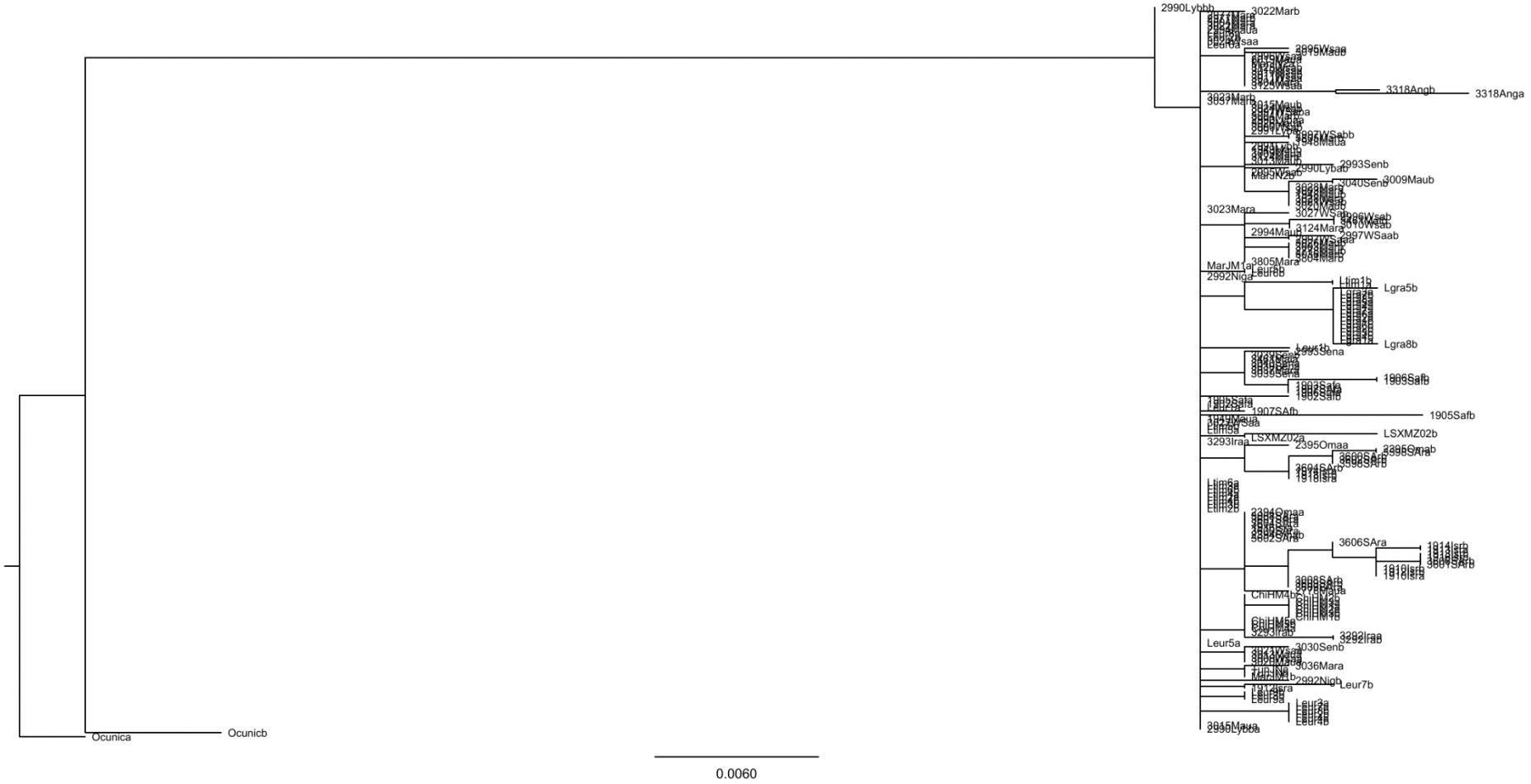
HPX



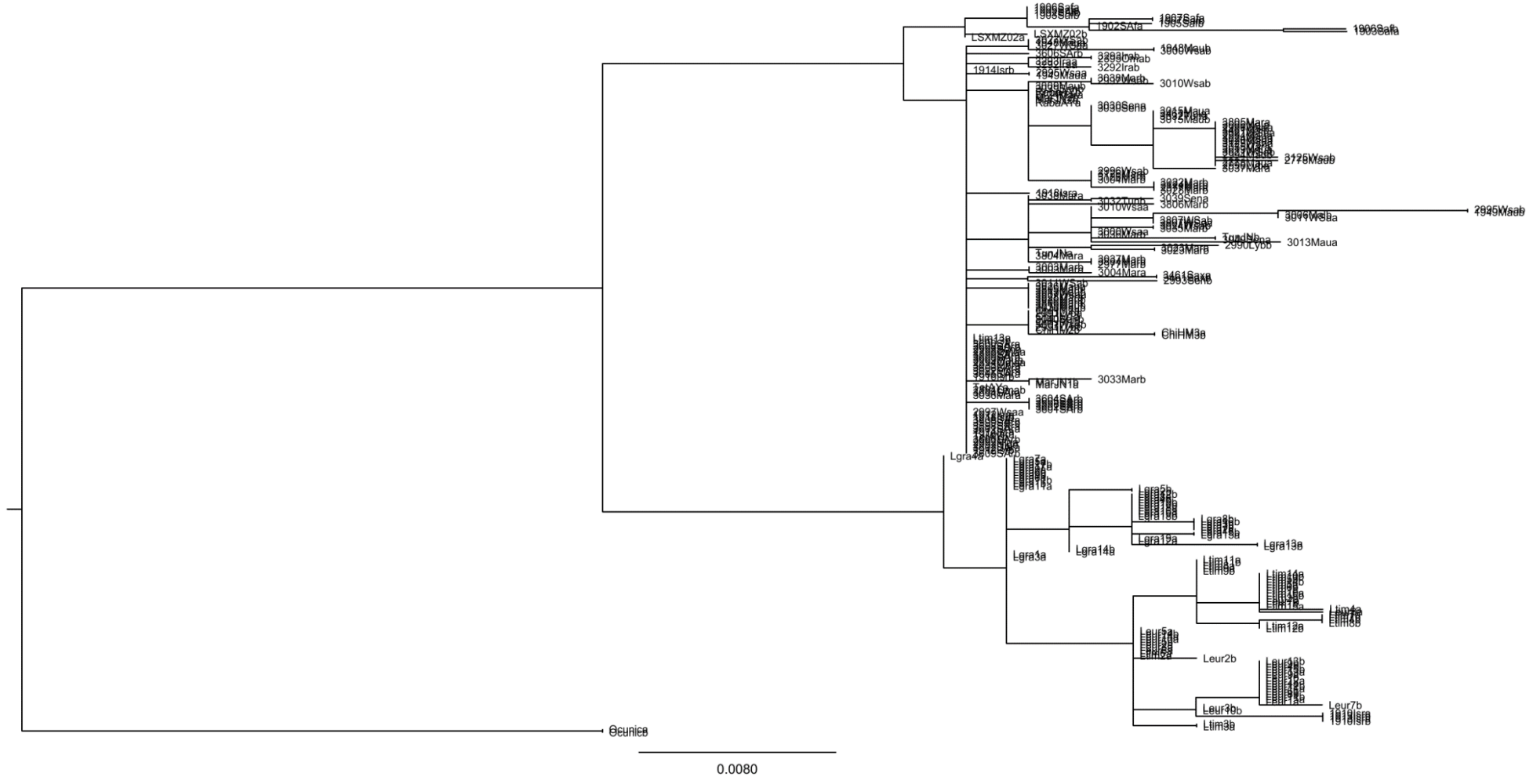
DARC



Appendix 14 – Nuclear gene tree with ML inference.
SPTBN1

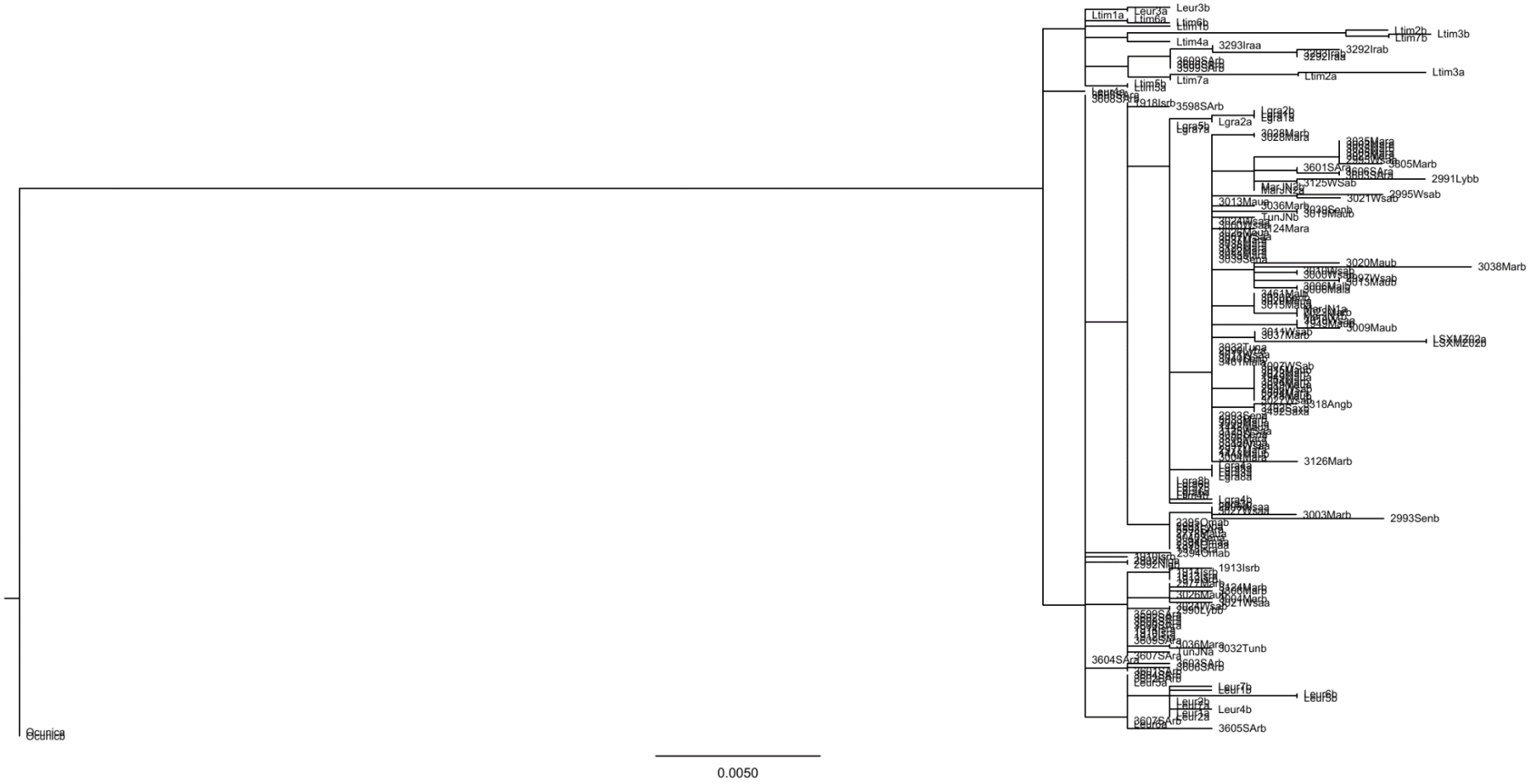


TF

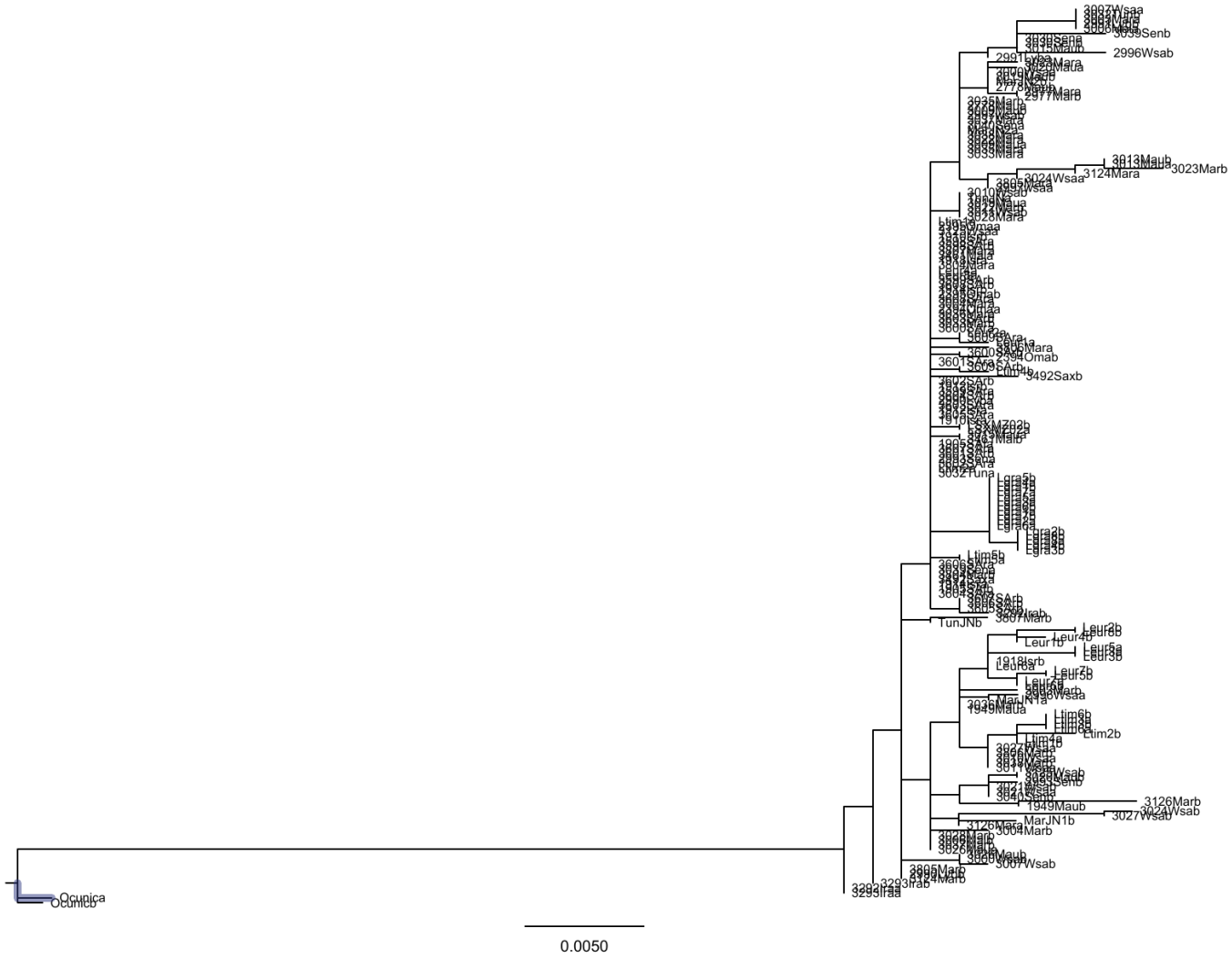

3010Wsab
3292Irab
3033Marb
3013Maua
Leur2b
Leur7b
Ocunica
0.0080

OXA1L

HPX



DARC


3007Wsaa
3039Senb
2996Wsab
3013Maub
3023Marb
3024Wsaa
3124Mara
3492Saxb
3126Marb
3024Wsab
3027Wsab
1949Maub
MarJN1b
3004Marb
3007Wsab
Ocupica
0.0050