Evolution and Phylogeography of Hedgehogs in North Africa

Vanessa Oliveira Lima

Mestrado em Biodiversidade, Genética e Evolução

Biologia

2013

Orientador

Dr. José Carlos Brito, Senior Scientist, Assoc. Researcher, FCUP/CIBIO

Coorientador

Dr. Zbyszek Boratyński, PhD, University of Jyväskylä

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

O Presidente do Júri,

Porto, _____/_____/_____

Agradecimentos

Ao Doutor José Carlos Brito, pela oportunidade de integrar este projeto. Pela forma como me orientou e ajudou nos momentos cruciais. E por me ter proporcionado as condições necessárias para que aprendesse tanto.

To Zbyszek Boratyński, for all the help that gave me during the development of my thesis.

À Joana Vilela pela sua amizade e preciosa ajuda laboratorial. Pela paciência que teve em me explicar muitos dos processos e por todas as dicas que se tornaram essenciais no meu trabalho.

À Vânia e à Liliana, pela enorme ajuda que me deram em perceber/resolver todos os problemas que foram aparecendo ao longo do trabalho laboratorial. E a todos aqueles que fazem parte da família CIBIO e me ajudaram a construir ideias mais sólidas.

A todos os Masterss!! Que mesmo estando fisicamente separados sempre mostraram espírito de ajuda e amizade.

A todos os grandes amigos. Pela motivação e apoio que sempre me deram.

Aos meus Pais, que nunca deixaram de acreditar em mim e aos quais devo tudo. Por sempre terem mostrado interesse e orgulho no meu trabalho e me terem confortado nos dias mais difíceis.

Ao Diogo, por ser o meu maior pilar. Por sempre ter confiado nas minhas capacidades e me ter motivado a ser melhor. Pela troca constante de ideias e por todos os conselhos que me deu. Por ter estado sempre à distância de uma palavra e por fazer, tão bem, parte da minha vida.

Sumário

Apesar do crescente interesse em diminuir a atual taxa de perda de biodiversidade, muitas áreas geográficas mantêm-se negligenciadas pelos investigadores. Os desertos e as áreas áridas são um ótimo exemplo de áreas com elevadas taxas de endemismos e micro-hotspots de biodiversidade que são geralmente mal avaliadas. Animais pequenos distribuídos em desertos, tal como os ouriços (Mammalia; Eulipotyphla) são na maioria das vezes pouco estudados. Existem quatro espécies de ouriços no Norte de África, que pertencem aos géneros *Atelerix*, *Hemiechinus* e *Paraechinus*, com distribuições e relações filogenéticas pouco esclarecidas.

O principal objetivo deste estudo foi investigar, pela primeira vez, padrões evolucionários e filogeográficos de espécies do Norte de África, recorrendo a dois marcadores moleculares. Foram analisadas 136 novas sequências mitocondriais com 1012 pares de base e 78 novas sequências nucleares com 675 pares de base, a partir de espécies Africanas e Eurasiáticas. Algumas espécies foram sequenciadas pela primeira vez com os dois marcadores moleculares. Árvores filogenéticas, mapas de distribuição, redes de haplótipos e índices de diversidade foram estimados, baseados em sequências de ADN mitocondrial e nuclear.

Os resultados suportam a relação de proximidade entre *Paraechinus, Mesechinus* e *Hemiechinus* e a elevada diferenciação das populações de *Erinaceus europaeus* das Penínsulas Europeias. Os resultados também sugerem uma distribuição mais ampla de *Atelerix algirus*, relativamente àquela que se sabia anteriormente, nas áreas mais a norte de África. Não foram encontradas evidências da presença de *Paraechinus aethiopicus* em África, tanto dentro como fora dos limites da sua distribuição. As incongruências genéticas encontradas entre o ADN mitocondrial e nuclear sugerem uma possível introgressão entre *A. algirus* e uma nova população, anteriormente não identificada (Erinaceinae cryptic). Foi encontrada uma elevada diferenciação genética entre populações de *A. albiventris*, localizadas em sítios opostos de uma suposta barreira geográfica (Rio Niger). As populações de *A. algirus* das Ilhas das Canárias provaram estar mais relacionadas com indivíduos de Marrocos e da Mauritânia, enquanto populações das Ilhas das Baleares mostraram maior semelhança com regiões geograficamente distantes.

Apesar de terem sido descobertas novas pistas para o conhecimento da evolução e da filogeografia dos ouriços, os resultados sugerem que são precisas análises mais sólidas, que usem novos marcadores moleculares e que envolvam um maior esforço de amostragem, para assim melhorar o conhecimento atual dos ouriços no Norte de África.

Palavras-chave: *Atelerix*, *Erinaceus*, evolução, ouriços, *Hemiechinus*, *Mesechinus*, mtDNA, nDNA, Norte de Africa, *Paraechinus*, filogeografia

Abstract

Despite the growing interest in slowing down the current rate of biodiversity loss many georaphic areas remain neglected by researchers. Deserts and arid regions are a great example of areas with high rate of endemism and micro-hotspots of biodiversity that are generally poorly assessed. Small mammals distributed in deserts, such as hedgehogs (Mammalia; Eulipotyphla) are mostly understudied. There are four species of hedgehogs known in North Africa, belonging to genera *Atelerix*, *Hemiechinus* and *Paraechinus,* with unclear distributions and phylogenetic relationships.

The main objective of this study was to investigate, for the first time, the evolutionary and phylogeographic patterns of the North African species, using two molecular markers. It was analysed 136 new mitochondrial sequences with 1012 bp and 78 new nuclear sequences with 675 bp from African and Eurasian species. Some species were sequenced for the first time with both molecular markers. Phylogenetic trees, distribution maps, networks and diversity indexes were estimated based on mtDNA and nDNA sequences.

Results support the close relatedness between *Paraechinus*, *Mesechinus* and *Hemiechinus* and the great differentiation between *Erinaceus europaeus* populations from the European peninsulas. Results also suggest broader distribution, than previously known, of *Atelerix algirus* in the northern areas of Africa. Evidences of the presence of *Paraechinus aethiopicus* in Africa were not found, both inside and outside their distribution range. Genetic incongruities found between mtDNA and nDNA suggest a possible introgression between *A. algirus* and a new previously unidentified population (Erinaceinae cryptic). High genetic differentiation was found between *A. albiventris* populations, located on opposite sides of a putative geographic barrier (Niger River). Populations of *A. algirus* from Canary Islands proved to be more related with Morocco and Mauritania specimens, whereas populations from the Balearic Islands showed greatest similarity with more distant geographic regions.

Although new clues to understanding the evolution and phylogeography of hedgehogs were discovered, results suggest that more robust analyses employing new molecular markers and greater sampling efforts are needed to enhance the present understanding of North African hedgehogs.

Keywords: *Atelerix, Erinaceus*, evolution, hedgehogs, *Hemiechinus*, *Mesechinus*, mtDNA, nDNA, North Africa, *Paraechinus*, phylogeography

Index

Figure Index

Table Index

1. Introduction

1.1. Decline of biodiversity

It is becoming increasingly recognizable that Earth's biota is experiencing a period of unprecedented changes which are causing disorders at the level of the abundance and species' distribution (Olff and Ritchie, 2002; Mendenhall et al., 2012). The visible changes of global biodiversity are driven by a number of factors that work in different ranges, at different time scales, in a way that not only affects the dynamics and functioning of ecosystems but also the society's dependence of natural resources (Sala et al., 2000; Cardinale et al., 2012). Natural or human-induced factors are responsible for the global loss of biodiversity. Direct drivers include changes in land use that result in habitat loss/fragmentation, invasive species (McNeely, 2001), overexploitation of particular resources (Dietz and Adger, 2003), pollution, and climate change (Pauls et al., 2013). Frequently, such drivers are influenced and induced by changes in human population and their lifestyle (Sala et al., 2000).

The habitat fragmentation/loss, and global climate change are factors of highest impact in biodiversity that are believed to negatively affect ecosystems (Fischer and Lindenmayer, 2007; Bellard et al., 2012). Changes in the availability of habitats tend to increase mortality rates, modify movement patterns, decreasing dispersal, promote the isolation of populations, which increases inbreeding levels and promotes the loss of genetic variability that leads to an increased extinction risk (Keller and Largiadèr, 2003; Dixon et al., 2007). Climate change is projected to become equally or more important than habitat change in the coming decades (Dawson et al., 2011; Bellard et al., 2012). At a basic level, climate change is able to decrease genetic diversity in space and time caused by directional selection and rapid migration (Pauls et al., 2013). It is generally recognized that species distributions mirror climatic limitations, existing a correlation between the past and actual geographical distribution of a species and some climate variable (Opdam and Wascher, 2004). In fact, this happens because while environment tolerances allow some species to persist in affected ranges, other species are only able to survive if colonize new suitable areas (D'Amen et al., 2013). As such, at a higher level of biodiversity, climate change can affect the integrity of biomes, modifying their functioning and resilience (Bellard et al., 2012).

1.1.2. Rating of biodiversity for conservation

There is urgent need in understanding how habitat and climate changes affect the distribution of life in order to develop effective conservation strategies (Cardinale et al., 2012; D'Amen et al., 2013). To quantify adequately biodiversity loss, it is necessary to fill crucial knowledge gaps, since most species living on Earth are still not described and not catalogued ('the Linnean shortfall'), and the way each aspect of biodiversity is distributed is also not yet well understood ('the Wallacean shortfall') (Richardson and Whittaker, 2010). As such, the emerging discipline of Conservation Biogeography (Whittaker et al., 2005) aims to link biogeography principles, theory and practice to improve the conservation and sustainable use of biodiversity (Richardson and Whittaker, 2010). In fact, the number of studies investigating impacts on biodiversity has increased and biogeography became helpful in providing concepts and tools about how biodiversity is evolving through time and how conservation planning should be designed (Devictor et al., 2010).

Despite the growing interest in slowing down the current rate of biodiversity loss, the improvement of analytical methods, computer technology and availability of biological and environmental data (attempting to model the distribution of species; Franklin, 2009), the overall knowledge on the world's biodiversity is far below the desired. The range of species studied represents only a small portion of known biodiversity. Generally, studies are taxonomically biased, in favor of some groups as plants, mammals and birds. The terrestrial biodiversity is favored over marine biodiversity and there are important biases in data collection across regions and ecosystems (Bellard et al., 2012). For instance, tropical forests are more studied than deserts and arid regions (Durant et al., 2012).

Deserts and arid regions are generally rated as heterogeneous areas of low diversity, poor and less valuable than biodiversity in other biogeographic regions and therefore receive less scientific attention (Ayyad, 2003; Durant et al., 2012). As such, they are understudied and much of the information available comes from historical exploratory missions and sporadic expeditions (Brito et al., 2009). But in fact, deserts and arid regions present high rate of endemism due to the influence of extreme environments (Médail and Quézel, 1999), and it is possible to find within these areas many endangered micro-hotspots of biodiversity, which should be placed as a priority for conservation given their rarity (Brito et al., 2013). North Africa gathers such characteristics and the large natural areas that still remain encourage greater conservation efforts. However, the available data and techniques to identify priorities

for safeguarding North African biodiversity are still far from the desired (Cowling and Wihelm-Rechmann, 2007).

1.2. Assessing biodiversity

To face the current crisis that defies biodiversity there are some steps to follow in order to bridge the gaps in present biodiversity knowledge. The identification of different lineages is the beginning for the recognition of populations with unique maternal histories and the starting point to define genetic groups for conservation intention (Terrasa et al., 2008). The advances in DNA technology and the incorporation of molecular tools into biodiversity and biogeographic studies have led to a vast improvement in assessing hotspots of biodiversity, species distribution patterns and areas with high biodiversity with risk of extinction (Hardy et al., 2010). But a deeper understanding will only be truly completed if we also understand events that may have contributed to one differential dispersal and consequent allelic enrichment and genetic diversity.

1.2.1. Phylogeny and phylogeographical studies

Phylogeny and Phylogeographical studies have played a significant role in the domain of biodiversity research. Phylogeny studies grew throughout the 1970s and exploded in the last decade due to the advent of polymerase chain reaction (PCR; Hewitt, 2001). Phylogenies based in molecular characters have evolved from studies based in nuclear sequences with a few hundred base-pairs and low confidence to studies developed using kilobases of data and highly resolved phylogenies (Hedges and Maxson, 1997). Generally, phylogeny tries to find the evolutionary relationships between present genetic structures of species and is currently based in molecular data, including mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) (Santucci et al., 1998), morphological data, and in methodological advances, like statistic tests or coalescent model-based methods (Kuhner, 2006). The evolutionary history of a set of taxa resulting from phylogenetic analysis is represented within a tree-like diagram where members of a group or clade are more related to each other than to members of other group (Brinkman and Leipe, 2001). On the other hand, phylogeographical studies are concerned with processes governing the geographical distributions of genealogical lineages, trying to understand microevolution and speciation in its geographic or spatiotemporal context (Kidd and Ritchie, 2006). Phylogeography allows researchers to test biogeographic hypotheses, describe the evolution of reproductive isolation of

population units and infer processes underlying the origin, distribution and maintenance of biodiversity, making inferences about temporal changes in the physical and biotic environment of a population using present-day genetic data (Beheregaray, 2008). More specifically, it allows the inference of demographic histories, divergence times, migration rates, hybridization events, hybrid zones, introgression occurrences, historical responses to landscape changes, amongst others (Bermingham and Moritz, 1998; Hickerson et al., 2010).

Many phylogenetic and phylogeographical studies were conducted worldwide, in terrestrial, freshwater and marine organisms, but while some regions have been strictly covered, others remain largely unexplored. This contrast is noted, for example, when comparing large parts of Europe and North Africa, the latter still poorly explored (Terrab et al., 2008).

1.2.2. Molecular approaches: mitochondrial and nuclear genomes

A wide range of molecular markers are nowadays available for genetic variation studies. This component coupled with powerful analyses of DNA (PCR and DNA sequencing) and the availability of easy-to-use computer packages, made it possible to answer many biological questions and contribute to much broader investigations of conservation biology, evolutionary and population themes (Sunnucks, 2000).

Since the discovery of mitochondrial DNA (mtDNA) in the late 1970s, it has been one of the most widely used molecular markers for phylogenetic studies and for phylogeographic purposes, because of its simple but efficiently genomic structure (Bajpai and Tewari, 2010). In mammals, mtDNA is a circular double stranded DNA molecule inside mitochondria that ranges in size from 16 000 to 20 000 nucleotide base-pairs (bp) and consists of about 37 functional genes and a non-coding control region (Avise, 2006). Different regions of the mitochondrial genome can provide different type of information and resolution. This genome has several important characteristics that make it a perfect option for genetic studies within and among populations (Avise, 2000). It can be easily amplified from a variety of taxa because of the abundant copies of the mitochondrial genome in a cell and its small size. Usually, mtDNA does not undergo recombination. It is highly variable in natural populations due to its elevated mutation rate and has a smaller effective population size because it is a haploid genome and is usually only maternally inherited (Ballard and Whitlock, 2004; Hurst and Jiggins, 2005; Galtier et al., 2009). But its use has some complications due to the fact that mtDNA provides only a single perspective, specifically about historical processes in females, neglecting the history of the species as a whole. The presence

of nuclear integrations of mtDNA sequences also complicates and creates conflicts in some analysis once, even with conserved primers, mtDNA incorporated in nuclear genome (NUMTs) may amplify (Hurst and Jiggins, 2005) and, mistakenly, be interpreted as a true mitochondrial gene (Ballard and Whitlock, 2004).

Despite the recognized success of mtDNA, some of its disadvantages could only be overcome by sequencing nuclear DNA (nDNA) markers. Nuclear genome encodes thousand of different proteins and exists in two copies per diploid cell. It is usually biparentally inherited with frequently recombination (Freeland, 2005) and has a rate of evolution of single copy nuclear sequence lower than the one verified in mtDNA (Zhang and Hewitt, 2003). Most of the nuclear genome accounts for non-coding sequences like introns, transposons or untranslated intergenic regions, whereas the protein-coding genes are only a small part of genome's constitution (Beebee and Rowe, 2008). In population analysis, the non-coding regions (e.g. introns) are most often used because they are more variable than the coding regions and apart from nucleotide substitutions, indels also contain a significant amount of information that contribute to the variability of the nucleotide sequence (Zhang and Hewitt, 2003).

Generally, mtDNA is preferable to use in relation to other molecular markers when addressing distant speciation events or when assessing overall genetic diversity. When inferring recent events, investigating recent loss of genetic variation, or verifying individual-level events (e.g. identity, dispersal and mating systems), it is preferable to use nuclear genes (Wan et al., 2004). Combining information deriving from nuclear and mitochondrial DNA is a powerful tool to explore complex evolutionary histories (Velo-Antón et al., 2008).

1.3. History and biogeography of the Mediterranean Basin and North Africa

The Mediterranean Basin represents one of the most remarkable places on Earth because of its level of biodiversity and also due to its incredible landscapes. Its location at the intersection of two major landmasses, Eurasia and Africa, and its huge topographical diversity and altitudinal differences promoted the amount of diversity and endemism found (Dernegi, 2010). Molecular studies demonstrate that, in some periods of the time, many Mediterranean areas acted as refugia to many species, and the isolation of populations promoted genetic differentiation, but while some refugia in southern Europe are relatively well studied, not much is known about North African refugia (Habel et al., 2011; Barata et al., 2012).

Phylogeographical surveys made in North Africa typically have detected great allelic richness (distributed heterogeneously by the highly variable topography) and evolutionary divergence of populations (Le Houérou, 1997). Distinct geological and paleoclimatic events have shaped North Africa, such as the onset of the Sahara-Sahel desert and arid region, the opening of the Strait of Gibraltar, and the setting of many rivers, and likewise contributed to the major patterns observed in present-day distribution of biodiversity (Colangelo et al., 2013).

1.3.1. The Mediterranean Basin

Considered one of the most important global biodiversity hotspots (Myers et al., 2000), the Mediterranean Basin covers more than 2 million square kilometers (second largest) over 34 countries, and is home to more than 330 mammals species and 30 000 plant species (Dernegi, 2010). Its marine portion includes 2.5 million square kilometers of sea and its circulation is dominated by the exchange of water of masses at the Strait of Gibraltar (Millot and Taupier-Letage, 2005). It ranges from westernmost islands of Madeira and Azores to northern Iraq in the east, including northern Italy and the extreme south of Morocco. The altitude ranges from 4,000 meters (Atlas Mountains) to 420 meters below sea level (Dead Sea) (Dernegi, 2010). It concentrates much of its richness in the southern European Peninsulas and the Western Maghreb (Cuttelod et al., 2008).

Superimposed forces, like the collision of the African and Eurasian plates, have profoundly shaped the topography, climate and geography of Mediterranean hotspot (Gaubert et al., 2010). However, much of the distribution patterns of biodiversity observed nowadays were influenced by paleoecological events, like the Pleistocene climate oscillations and the Messinian salinity crisis that induced the closing of the Strait of Gibraltar (Dawson, 1996; Duggen et al., 2003).

Pleistocene climatic oscillations have played a major role in the distribution of many taxa in response to climatic fluctuations, particularly those with low dispersal abilities (Salvi et al., 2013). The climate change in this period of time was characterized by alternations between glacial and interglacial periods, called by glacial-interglacial cycles. During colder periods, southern Europe acted as refugia for temperate species, when northern Europe and the Alps were covered with ice (Hughes et al., 2007), followed by subsequent northward population expansion from south during the interglacial periods (Cosson et al., 2005). Mountains and seas acted as barriers, limiting movement or even isolating populations in glacial refugia. Different species founded distinct refugia or traversed particular colonization routes, some were resilient

to environmental change while others suffered allopatric differentiation, and some may have faced local extinction (Dubey et al., 2006). The effects of Pleistocene period are noted in some species, like the grasshopper (*Chorthippus parallelus*), that has less haplotype diversity across North Europe when comparing to southern regions. The distribution and divergence of European hedgehogs (*Erinaceus europaeus* and *Erinaceus concolor*) and the brown bear (*Ursos arctos*) was instigated by western and eastern refugia during the cold period (Santucci et al., 1998; Hewitt, 1999).

The Strait of Gibraltar, currently formed by two facing peninsulas, has intermittently separated the African and Iberian plates since their collision in the late Miocene (Loget and Van Den Driessche, 2006). During the Messinian, the closing of the Mediterranean Basin and the decrease in the sea level because of a salinity crisis resulted in a land corridors that allowed intense biotic interchanges between southern Europe and northern Africa. In the early Pliocene (about 5.33 Myr), the strait reopened and the connection that existed between the two sites was closed (Krijgsman et al., 1999). This event has long been considered a cause for vicariance observed in some Mediterranean lineages and is thought to have played an important role in shaping demographic and genetic patterns in multiple amphibians, reptiles and mammals (Fonseca et al., 2008; Gaubert et al., 2010). Midwife toads (*Alytes*) have the split of some lineages related with the opening of the Strait (Martínez-Solano et al., 2004). Mediterranean snakes (*Malpolon monspessulanus* and *Hemorrhois hippocrepis*), North African wall lizards (*Podarcis*) and different genera of amphibians (*Pleurodeles*, *Discoglossus*, *Alytes*, *Pelobates*, *Rana* and *Salamandra*) reported the importance of phenomena like the opening of the Strait of Gibraltar in the diversification process of the species (Carranza et al., 2006; Pinho et al., 2006; Schmitt, 2007). Also shrews (*Crocidura russula*) and long-eared bats (*Plecotus*) demonstrated the key role of the Strait as a geographic barrier (Juste et al., 2004; Cosson et al., 2005).

1.3.2. The Sahara-Sahel desert and arid regions

Northern Africa is well known by its complex ecosystems. To the north there is the Sahara desert, the major geographical feature of the African Continent and the largest hyper-arid warm desert on the planet (Kröpelin et al., 2008). To the south exists a transition zone between the desert and the more humid savanna called Sahel, that characterized by a sharp gradient of precipitation and vegetation cover (Nicholson, 2013). Recent satellite and rainfall records show a yearly strong change in rainfall and vegetation greenness along the Sahel, but no systematic expansion of the Sahara desert is seen. But this does not mean that both regions are static, actually they are

known by their histories of dramatic regime shifts between more humid and more arid phases, occurred with no apparent warning (Foley et al., 2003).

The Sahara desert occupies an area of slightly over 8 million square kilometers, between latitudes 16 and 32°N, circumscribed within the 100 ± 50 mm isohyets of mean annual rainfall (Le Houérou et al., 1997). It is subdivided in five regions: Atlantic Sahara, Northern Sahara, Central Sahara, Southern Sahara and Eastern Sahara (Allaby and Garratt, 2008) and encompasses African countries as Algeria, Chad, Egypt, Libya, Mali, Mauritania, Morocco, Niger, Sudan, and Tunisia. The onset of desert conditions in the Sahara was estimated at approximately 7 Myr ago (Schuster et al., 2006), but accurate dating remains largely controversial (Kröpelin and Swezey, 2006; Swezey, 2009). Since the Pliocene (5.3-2.5 Myr ago), it has experienced many drastic climatic changes (Figure 1): at least eight to ten major dry-wet cycles have changed completely the landscape and enabled the occurrence of a flora and fauna speciation-extinction dynamics in response to climatic oscillations (Le Houérou, 1997). The frequency of these cycles of desert expansion and retraction are estimated to have ranged from 100.000-20.000 years, in the last 3 Myr (Le Houérou, 1992). From the late Pleistocene until the middle Holocene era (14.500 to 5.500 years ago), the Sahara was much wetter than today, covered with extensive vegetation, lakes and wetlands and inhabited by thriving animal communities and human settlements (Foley et al., 2003).

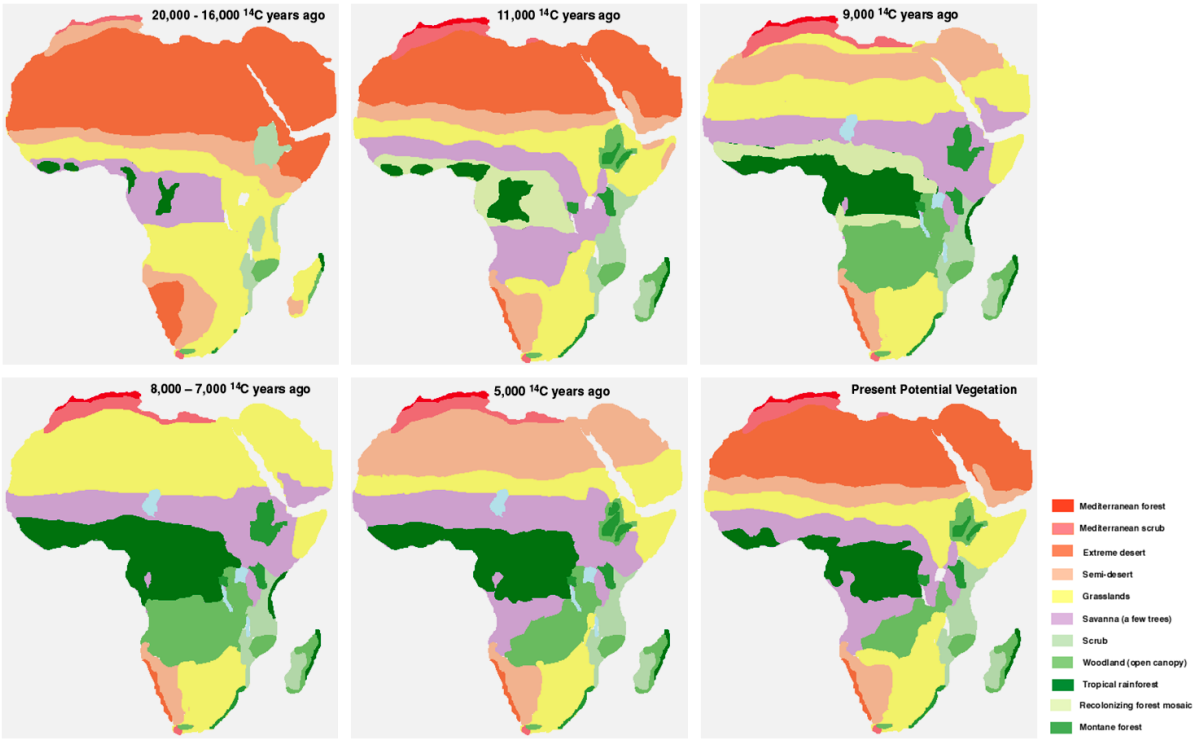


Figure 1 - Temporal changes in the distribution of African ecosystems since the Last Glacial Maximum until present-day [adapted from Adams and Faure (2004)].

This wet period came to an end between 5.000 and 6.000 years ago, when the aridity of the Sahara greatly increased and mesic vegetation communities disappeared and lake levels declined. The regime shift in the climate and ecosystems of the Sahara is thought to be due to multiple causes, such as different patterns on orbital insulation, changes in temperature of the ocean platform, and feedback mechanisms between the reduction of rainfall and vegetation cover (Holmes, 2008).

The Sahel region is a semi-arid expanse of grassland, shrubs, tropical savannas and forests with annual rainfall ranging from 100 to 200 mm in the north, where it connects with the desert, and 500 to 600 mm in southern limit (Foley et al., 2003; Nicholson, 2013). A greater knowledge about this region of Africa emerged around 1970s, when an unexpected devastating drought killed a great part of domestic livestock, caused displacement of millions of people to refugee camps and led to nearly one million people had starved to death (Nicholson, 1995). Many hypotheses were proposed to explain this event and the most satisfactory explanation was proposed by Zeng et al. (1999) which relates the drought with a complex interaction among the atmosphere, land and ocean circulation. Despite the fact that this 3-decade-long drought was considered one of the longest and most severe in recent history, when compared to the Sahara regime shift, this change was small and confined to a limited region (Foley et al., 2003).

The evolutionary history of North African species was deeply influenced by Sahara-Sahel shifts. In the Sahel, Afro-tropical fauna became rare, victims of climatic fluctuations and human impacts in the region (Nicholson, 1995). Although scarce, researches carried out in the Sahel demonstrated influence of lakes and river courses (e.g. Niger River, Lake Chad) in the distribution and diversification of species and populations (Petter, 1974; Dobigny et al., 2005; Brouat et al., 2009). The role of the Sahara desert and its mountains in promoting genetic and morphological divergences in North-African taxa is a question of great biogeographical interest. The expansion/contraction phases of the Sahara and climate changes between the Miocene and Pleistocene (Nicolas et al., 2009) have induced north-south vicariance speciation events in many species (Franck et al., 2001; Douady et al., 2003; Guillaumet et al., 2008). Nevertheless, some populations have persisted in Saharan mountains, in suitable habitat patches, where they diverged under particular selective pressures (Le Houérou, 1997; Mouline et al., 2008). The Saharan mountains hold many endemic animal species that confined to small size patches, isolation and subjected to stochastic events, are extremely vulnerable to extinction (Brito et al., 2013). The biodiversity present there are testimony to the role of mountains in the diversification of the Sahara vertebrate fauna, which makes them important habitats with high value for

conservation (Le Berre, 1989; Messerli and Winiger, 1992). A great range of studies on reptile species, the commonest inhabitants of arid areas, showed a common influence by climate fluctuations and ecological barriers where is highlighted the vicariant role of the Sahara and its mountains (Fonseca et al., 2009; Brito et al., 2011; Gonçalves et al., 2012; Metallinou et al., 2012). The same evidences were found in some studies with rodents and canids (Douady et al., 2003; Brito et al., 2009; Boratyński et al., 2012; Colangelo et al., 2013). Although these pieces of evidence reinforce the idea of diversification and isolation by geographical and ecological barriers in North Africa, we cannot extrapolate a small part of African life to a large-scale picture. In fact, much more has to be done to improve our knowledge about every levels of biodiversity, more precisely expanding the range of species studied.

1.3.3. Islands colonization: examples from the Balearics, Canary and Cyprus Islands

To a successful phylogeographic approach it is necessary to analyze a comprehensive sampling encompassing most of the current geographic distribution of taxa, not only focus on the mainland but also include islands (Juste et al., 2004). Islands' communities have two different types of species pools; one is related to endemic species with a long evolutionary history in isolation from the origin populations and the other corresponds to recently introduced species. The habitat degradation rate caused by human population and the homogenization of species threatens the worlds' biota, particularly islands (Gillespie and Roderick, 2002). Understand if species are endemic or introduced, what their origin is and how time and isolation are related in some islands are questions that need to be answered for scientific and conservation purposes, in order to mitigate some impacts (Vogel et al., 2003).

The Balearic Islands belong to the Mediterranean hotspot and are exposed peaks of a submarine promontory belonging to the Spanish Betic system. They were connected to the continent during the desiccation of the Mediterranean basin, what allowed faunal exchanges between islands and continent (Schüle, 1993). Just after the reopening of the Gibraltar Strait the islands suffered fragmentation and since then have never connected again (Colom, 1978; Schüle, 1993). The separation of Balearics from the continental area and the separation between islands into Eastern and Western groups promoted independent evolution and regional structuring in different taxa (Terrasa et al., 2008). This situation was verified in the extinct bovid (*Myotragus balearicus*) (Lalueza-Fox et al., 2005), in the midwife toad (*Alytes*) (Arntzen and García-París, 1997), in the endemic lizard (*Podarcis lilfordi*) (Brown et al., 2008;

Terrasa et al., 2008) and also in Balearic populations of holms, kermes and cork oaks (*Quercus*) (López-de-Heredia et al., 2005).

The Canary Islands are also historically and biologically of particular interest. They are considered a prime location for develop molecular phylogenetic and phylogeographic studies since they provide better laboratory conditions than continental areas to trace complex histories about human migration and animals' colonization in different animal taxa (Juan et al., 2000). The Canary Islands are composed by seven ecologically different islands distributed in the west coast of Morocco, all formed in the past 20 million years by volcanic eruptions and according to some references they have never been connected to the African continent (Carracedo, 1993). The climate is variable according to altitude, being the mean value around 21°C in coastal zones to about 9°C at higher altitudes. This variance has direct effect on the vegetation, being for example Lanzarote and Fuerteventura highly influenced by dry winds from nearby Sahara Desert. The high endemism is characterized by more than 3600 terrestrial endemic species and of a total of 13 328 terrestrial species 1434 have been introduced by man (Nogales et al., 2006). The expansion of human populations to the islands had particular impacts on the biodiversity, especially in relation to endemic species. The introduction of animals with commensal or anthropophilous behaviors, accompanying human dispersal, led to a reduction of autochthonous mammals and birds and is considered one of the most important causes of biodiversity loss (Quammen, 1996). Most of the mammals' species living nowadays in the Canaries have been introduced by man, coming originally from Europe, Africa and Asia. It has been introduced in different periods of time big mammals like the domestic cat *Felis catus*, the goat *Capra hircus*, the sheep *Ammotragus lervia* and other smaller, like rats and mouse's, rabbits, shrews, and hedgehogs, which have become established in the wild. Although is recognized the negative influence of introduced cats and herbivores on the endemic biota of the Canary Islands, further research is needed to evaluate the impact of other introduced fauna in the island ecosystems (Nogales et al., 2006).

Cyprus is the third largest island in the Mediterranean. Its ocean origin was due to a block of ocean crust uplifted over-subduction zone. The deep sea always contributed to keep the island isolated and nowadays it is situated 65 km south of Turkey and 105 km west of Syria (Hadjisterkotis et al., 2000). Given that it has never came in contact with mainland the only way of colonization was by overwater dispersals (e.g. rafting), aerial migrations or human colonization (Benda et al., 2007). The isolation let to the evolution and differentiation of many endemic, rare and important species and due to climatic variation and the variety of habitats present, Cyprus hosts considerable biodiversity (Hadjisterkotis and Masala, 1995; Hadjisterkotis et al., 2000; Myers et al., 2000). But

when compared with other regions of Mediterranean, this island is extremely poorly studied and the information available is based in old taxonomic studies, species descriptions and punctual field trips (Anlas, 2012).

1.4. Hedgehogs

1.4.1. General characteristics

Hedgehogs are spiny mammals of the order Eulipotyphla, family Erinaceidae and subfamily Erinaceinae. The old classification that titled them as insectivores was abandoned and according to the new classification, ceased to exist the order Insectivora to be replaced by the new orders Eulipotyphla and Afrosoricida (golden moles and tenrecs) (Nikaido et al., 2003). The order Eulipotyphla was the first lineage of superorder laurasiatheria to diverge and encompasses hedgehogs, moles, shrews and solenodons (Douday et al., 2002). Hedgehogs represent together with their closest living relatives, moonrats and gymnures of the subfamily Galericinae, a well-established monophyletic group (Gould, 2001; He et al., 2012). Phylogenetic analysis of both fossil and extant taxa places this lineage as far back as the late Cretaceous, making them the oldest known living placental mammals (Gould, 2001).

There are about sixteen species of hedgehogs in five genera recognized: *Atelerix* (*A. albiventris*, *A. algirus*, *A. frontalis* and *A. sclateri*), *Erinaceus* (*E. amurensis, E. concolor, E. europaeus* and *E. roumanicus*), *Hemiechinus* (*H. auritus* and *H. collaris*), *Mesechinus* (*M. dauuricus* and *M. hughi*) and *Paraechinus* (*P. aethiopicus*, *P. hypomelas*, *P. micropus* and *P. nudiventris*) (Corbet, 1988; Wilson and Reeder, 2005). Hedgehogs are generally found throughout parts of Europe, Asia, and Africa, and more recently in New Zealand due to introduction (Frost et al., 1991; Wilson and Reeder, 2005), whereas some genera are only found in some continents. Fossil records show the same distribution pattern that occurs nowadays, except for some species that used to inhabit in North America during the Oligocene-Miocene but went extinct some five million years ago (Corbet, 1988; He et al., 2012).

Hedgehogs are easily recognized by their dense coat of spines, the most obvious specialization that differs from most other mammals (Catania, 2005). The spines are modified hair, with a spongy matrix and outer keratinous shaft, which function as protection against any danger or predators. These spines are embedded in a thick muscle sheath able to form a bag into which the hedgehogs roll up their entire body, forming a dense and impenetrable ball of prickles (Riley and Chomel, 2005).

Hedgehogs are small sized (head and body: 13 to 30 cm; tail: 3 to 5 cm) with large eyes and ears, prominent whiskers on the lateral parts of their snout and also have a dense array of microvibrissae at the tip of the snout (Corbet, 1988; Catania, 2005). They are nocturnal mammals, non-territorial that solitarily inhabit extensions of land from the deserts of North Africa and Arabian Peninsula to the grasslands, woodlands and cultivated lands in Europe and Asia (Corbet, 1988). The local distribution normally depends on the density of predators and resources, such as prey and shelter (Becher and Griffiths, 1998). They are relatively sedentary and although some studies refer a juvenile dispersal phase, their dispersal ability between closely spaced populations is limited or nonexistent, but it is unclear the role of human-constructed barriers in gene flow patterns and genetic differentiation (Doncaster et al., 2001). The particular capacity to hibernate/aestivate for considerable proportions of the year enables them to endure extreme conditions, such as food shortage or low temperatures (Corbet, 1988; Symonds, 2005; He et al., 2012). Breeding in hedgehogs occurs once or twice a year and they are sexually mature by the time of the first year. Given the lack of elaborate mating behaviors and territorial behavior, mating occurs whenever they meet (Hamilton, 1999; Symonds, 2005).

Considering the historical research made on family Erinaceidae, *Erinaceus europaeus* (Linnaeus, 1758) is the most well known species of hedgehog worldwide. It is spread in western and central Europe, in UK, Ireland, southern Scandinavia, Estonia and northern Russia, and it was introduced in New Zealand around 1869 and more recently in Azores (Santucci et al., 1998; Amori et al., 2008). Multiple phylogenetic analysis have been performed in this species, mostly to determine the general hedgehog position in relation with closer relatives, such as moles and shrews (Krettek et al., 1995; Santucci et al., 1998; Nikaido et al., 2003).

1.4.2. Origin and phylogeny

The family Erinaceidae is represented in all periods since the Palaeocene, making them one of the oldest surviving lineages of placental mammals (Novacek, 1986; Gould, 2001). The split of the family in two subfamilies (Erinaceinae and Galericinae) is contemporaneous to the Cretaceous/Paleogene boundary and the division was never contested in any phylogenetic study. It appears that events that triggered the mass extinction acted also as a diversification agent (Doudady and Douzery, 2009). Fossil records of the subfamily Erinaceinae seems to be all within the actual range of the Old World with the exception of records of species in North America during the Oligocene-Miocene (Corbet, 1988). There is no record about the exact time and place where the

different hedgehogs' species arises, with the exception of the genus *Erinaceus* that have been found across Europe from the Pliocene and is considered that it entered in Europe from Asia (Butler, 1988) and genus *Atelerix*, which is thought to be a paleo-Mediterranean endemism that resulted from a pre-Messinian invasion of Africa by European populations of *Erinaceus europaeus* and represents the least modified surviving descendant (Butler and Greenwood, 1973).

Despite the fact that there is a well-studied fossil record for hedgehogs (He et al., 2012), there has never been great success in resolving the phylogenetic relationships within the Erinaceinae subfamily. Most of the studies done were only based in morphologic data and presented many discrepancies between phylogenies that, even today, remain unanswered due to a major failure in the sampling of species (Corbet, 1988; Frost et al., 1991; Gould, 1995; Grenyer and Purvis, 2003). Only recently, DNA sequences of hedgehogs started to be considered in order to solve phylogenetic evidences within the Eulipotyphla order and the Erinaceidae family. But the number of species sequenced and the molecular markers used are too small to fully understand the evolutionary history of hedgehogs (Seddon et al., 2001; Douady et al., 2002; He et al., 2012).

From the alternative hedgehogs' phylogenies proposed until know (Figure 2), most of them agree with the monophyletic relationship of *Atelerix* and *Erinaceus* clades (Gould, 1995; Grenyer and Purvis, 2003; He et al., 2012). But, over the years, they also had some points of disagreement amongst sources over the placement of *Mesechinus* genus and the constitution of a monophyletic group by *Hemiechinus* and *Paraechinus* genera (Corbet, 1988; Frost et al., 1991; He et al., 2012). Based on countless morphological characters, the *Mesechinus* was placed with some confidence as the sister taxon of the *Atelerix* + *Erinaceus* clade (Frost et al., 1991; Gould, 1995) and the monophyly of the *Hemiechinus* and *Paraechinus* (seemed strongly supported) by some cranial and external pelage characteristics (Corbet, 1988; Bannikova et al., 2002).

The most recent comprehensive phylogenetic analysis of the Erinaceinae family was achieved by He et al. (2012) (Figure 3), in which the authors combined two robust datasets, namely mitochondrial DNA of many living species and more than a hundred morphological characters. They also sequenced for the first time species of the genus *Mesechinus* for three different mitochondrial genes and completed genetic information of some genera that had available only few short genes (*Atelerix*, *Hemiechinus* and *Paraechinus*).

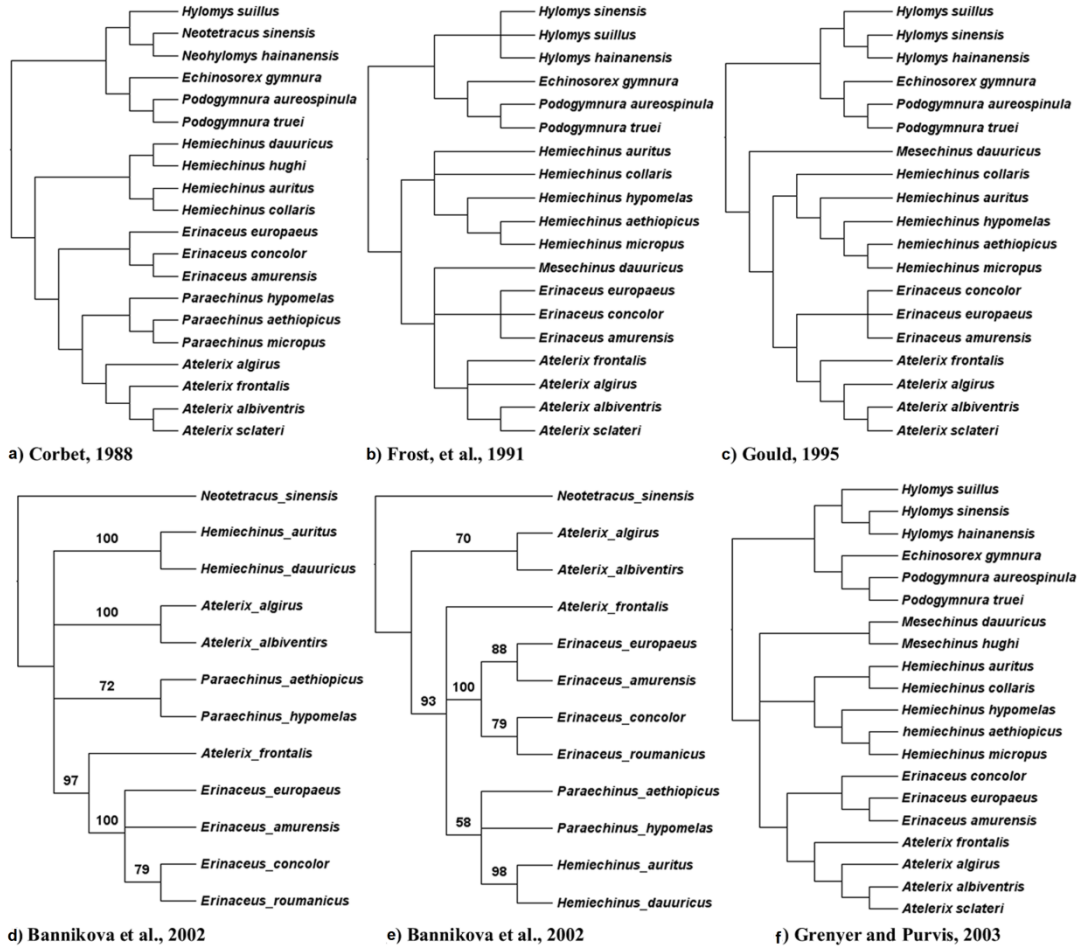


Figure 2 – Previous phylogenetic hypotheses of Erinaceidae relationships proposed by: a) Corbet. 1988; b) Frost et al., 1991; c) Gould, 1995; d) and e) Bannikova et al., 2002; f) Grenyer and Purvis, 2003. [Adapted from He et al., 2012]

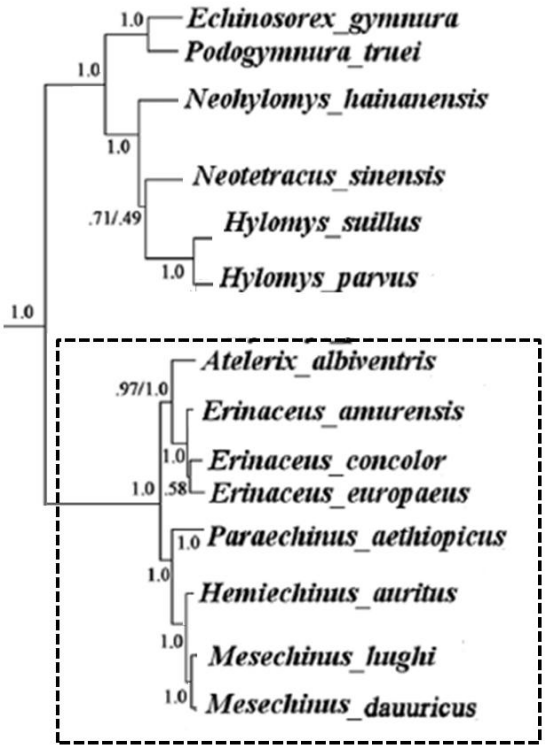


Figure 3- Phylogenetic tree of Erinaceidae family resulting from Bayesian phylogenetic analyses using combined genes and combined data set. Node numbers indicate Bayesian posterior probabilites. [Adapted from He et al., 2012]

This study and the resultant phylogenetic tree (Figure 3) clarified some unresolved positions and supported old assumptions: (1) the monophyly of *Paraechinus* and *Hemiechinus,* confirmed by morphological analyses; (2) the relationship of *Mesechinus* and *Hemiechinus* as sister taxa, unanimous in all genes and combined data; (3) and the basal position of *Paraechinus* in relation to the clade *Mesechinus* and *Hemiechinus*, observed in combined data analyses (He et al., 2012).

Some incongruence observed in the different studies along the years could be explained by adaptive evolution of phenotypic character, maternal inheritance pathway of mitochondrial genes or due to hybridization (He et al., 2012).

1.4.3. North African hedgehogs

There are four species of hedgehogs known in North Africa (Figure 4), belonging to genera *Atelerix*, *Hemiechinus* and *Paraechinus*.

a) The North African hedgehog (or Algerian hedgehog), *Atelerix algirus* (Lereboullet, 1842), is large and has its ventral pelage off-white with varying amounts of brown, but can be marginally differentiated in the insular populations (Corbet, 1988). It is naturally distributed in North Africa, from Morocco to Libya, including Djerba Island (Tunisia). It is believed that it was introduced by humans in the Balearic and Canary islands archipelagos (abundant in Lanzarote and Fuerteventura islands), Malta island, and the south-eastern littoral fringe of the Iberian Peninsula (Palomo and Gisbert, 2002; Hutterer, 2005; Morales and Rofes, 2008). This species typically occurs at lowlands and hills, up to 400 m, but in Morocco it can reach 900 m of altitude (Mitchell-Jones et al., 1999). Occupies dry and semiarid habitats but may be quite common in grassland environments, farmlands, meadows, savannas and scrub vegetations areas, avoiding deserts and dense woods (Aulagnier et al., 2009). The species is apparently related to specific habitats, as well as aridity levels that limit their distribution and reduce competition ability in face of *P. aethiopicus*, as observed in a restricted sympatry area on the high plains of Algeria (Corbet, 1988).

b) The *Atelerix albiventris* (Wagner, 1841), known as Four-toed hedgehog, is the smallest of the African hedgehogs and differs from other *Atelerix* species by the lack of a hallux (Corbet, 1988). It has black eyes, large and dark ears, the body is speckled black and white, with no dorsal stripe, the muzzle is brown, and the face, legs and ventral surface is white (Santana et al., 2010). This species is widely encountered in savannah and steppe zones of equatorial Africa, from Senegal and Guinea to Ethiopia, and south to Zambezi River (Wilson and Reeder, 2005). It is broadly widespread but sporadic in drier regions of Africa, occupying a variety of habitats, such as long-grass

savannah, scrub, suburban gardens, woodlands, agricultural lands, hills, and mountains, to a maximum of 2000 m of altitude (Corbet, 1988). There is no register of fossil records of *A. albiventris* (Santana et al., 2010).

c) The third species, *Paraechinus aethiopicus* (Ehrenberg, 1832), is known as Desert hedgehog or Ethiopian hedgehog because of its distribution and habitat preferences. It is a medium-sized hedgehog with large ears, legs are dark brown and the rest of ventral pelage is white (Corbet, 1988). In Africa, it is distributed in the Sahara, south of the Atlas Mountains, from Morocco southward to Mauritania, and then eastwards through Algeria, with populations known in Tunisia, Libya, Sudan and northern Ethiopia (Corbet, 1988; Khaldi et al., 2012). It occurs mostly in desert and in dry steppes, normally nesting in burrows under shrubs or amongst rocks (Aulagnier et al., 2009). *P. aethiopicus* is marginally sympatric with *Atelerix algirus* in Tunisia, Algeria, and Morocco, and with *A. albiventris* in Ethiopia (Corbet, 1988).

d) The Long-eared hedgehog, *Hemiechinus auritus* (Gmelin, 1770), named after the presence of long ears that are considerably longer than adjacent spines (Corbet, 1988), is also present in North Africa. Although its distribution is mostly found along the Middle East, it has been observed in coastal Libya and Egypt (Çolak et al., 1998; Wilson and Reeder, 2005). It exhibits preference for arid grassland and scrubby habitats, but it can also be found in cultivated land and around human settlements, in gardens, buildings and densely vegetated areas of coastal desert (Corbet, 1988). In arid zones, the limit of its distribution is mostly determined by physiological inability to survive in arid conditions and by competition with more arid-adapted species, like species of *Paraechinus* (Corbet, 1988).

Presently, there is a huge lack of information about systematics, evolutionary relationships, and biogeography of many species in North Africa. The case worsens in relation to North African hedgehogs that were never the main focus of phylogenetic or phylogeographical studies. The distribution and genetic variability of hedgehogs are also poorly known, the current systematics of African forms is strictly based on morphological characters, and the group needs comprehensive research. Available distribution maps of African forms are generally coarse, rendering them unsuitable for local scale conservation planning. There are uncertainties concerning the role of geological and paleoecological events in population dynamics and evolutionary trajectories. Given the wide ranges of African forms, it is possible that major haplotypes may occupy distinct ecological niches subjected to different environmental pressures, potentially leading to speciation events.

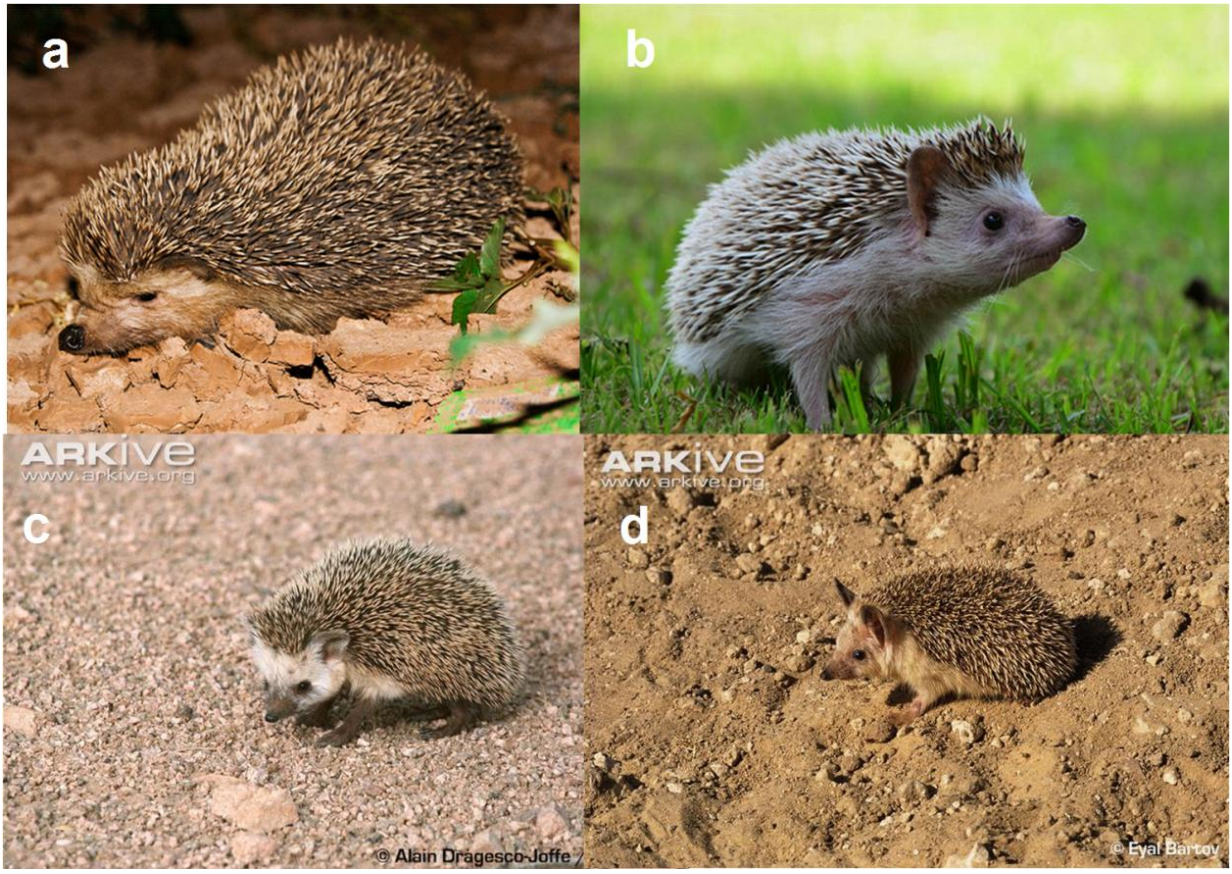


Figure 4 – Example images of North African species of hedgehog. Letters are the same as in the main text. a) *Atelerix algirus*; b) *Atelerix albiventris*; c) *Paraechinus aethiopicus*; d) *Hemiechinus auritus*. Photographs were download from flickriver and Arkive (www.flickriver.com and www.arkive.org).

1.5. Objectives

The study aims to investigate for the first time the evolutionary and phylogeographic patterns in four species of North African hedgehogs of genera *Atelerix, Hemiechinus* and *Paraechinus*. Two molecular markers, mtDNA and nDNA, will be used to answer the following questions:

1) Which are the phylogenetic relationships among African and non African hedgehogs' species? And between the African species?

2) Which are the major haplotypes?

3) Is there cryptic diversity?

4) Which are the main phylogeographic patterns?

5) Are there putative barriers to dispersal?

6) Are there potential refugia localities?

7) What is the geographic origin of introduced populations?

8) Is the systematic in accordance with the actual genetic variability?

The results of this study are expected to contribute to a better understanding of the phylogenetic relationships of hedgehog species and how genetic diversity is distributed. It is also expected to contribute to a better knowledge about African biodiversity by filling crucial knowledge gaps.

2. Methods

2.1. Study area and samples

The study area corresponds to North Africa and the Mediterranean Basin (Figure 5). A total of 289 samples of hedgehogs were collected in different countries of North Africa and the Mediterranean Basin between November 2003 and May 2013. African samples encompass northern Algeria (n=5), northeastern Egypt (n=2), northern and southern Libya (n=11), southwestern Mali (n=9), southern and coastal Mauritania (n=34), Morocco (n=95), southern Niger (n=9), northern Senegal (n=4), and northern Tunisia (n=11). Samples north of the Mediterranean Basin cover Armenia (n=1), France (n=1), Greece (n=5), Portugal (n=13), Serbia (n=1), Spain (n=39), including samples from Canary Islands (n=5), Syria (n=3), Turkey (n=25), and Ukraine (n=1). Samples from Mediterranean islands include the Baleares/Spain (n=6), Cyprus (n=3), Sardinia/Italy (n=4), and Malta (n=4).

Almost all samples (n=277) were collected during field missions to North Africa and Europe by the supervisors, CIBIO researchers, and collaborators. Samples were constituted by different parts of hedgehogs' body (e.g. spines, tissue and nails) taken from road-kill animals and placed inside labeled cap microtubes and preserved in ethanol 96%. At the moment of collection Geographical location was recorded using GPS (Global Positioning System) and represented using the Geographical Information System ArcMap 10.0 (Figure 5). In the field, samples were taxonomically named but since most of the hedgehogs collected were in a high state of degradation, putative names were given according to general color pattern, ear length, and location of the sample according to the distribution of hedgehogs, following Aulagnier et al. (2008) and IUCN (2013). As such, samples available for this study theoretically include representatives from African (*Atelerix algirus*, *Atelerix albiventris*, *Hemiechinus auritus* and *Paraechinus aethiopicus*) and non-African (*Erinaceus concolor*, *Erinaceus europaeus* and *Erinaceus roumanicus*) species. A small number of samples was also available from museum collections, including Cornell University (n=2), Centro de Recursos de Biodiversidade Animal da Universidade de Barcelona, Spain (n=2), and Natural History Museum of Crete, Greece (n=8),

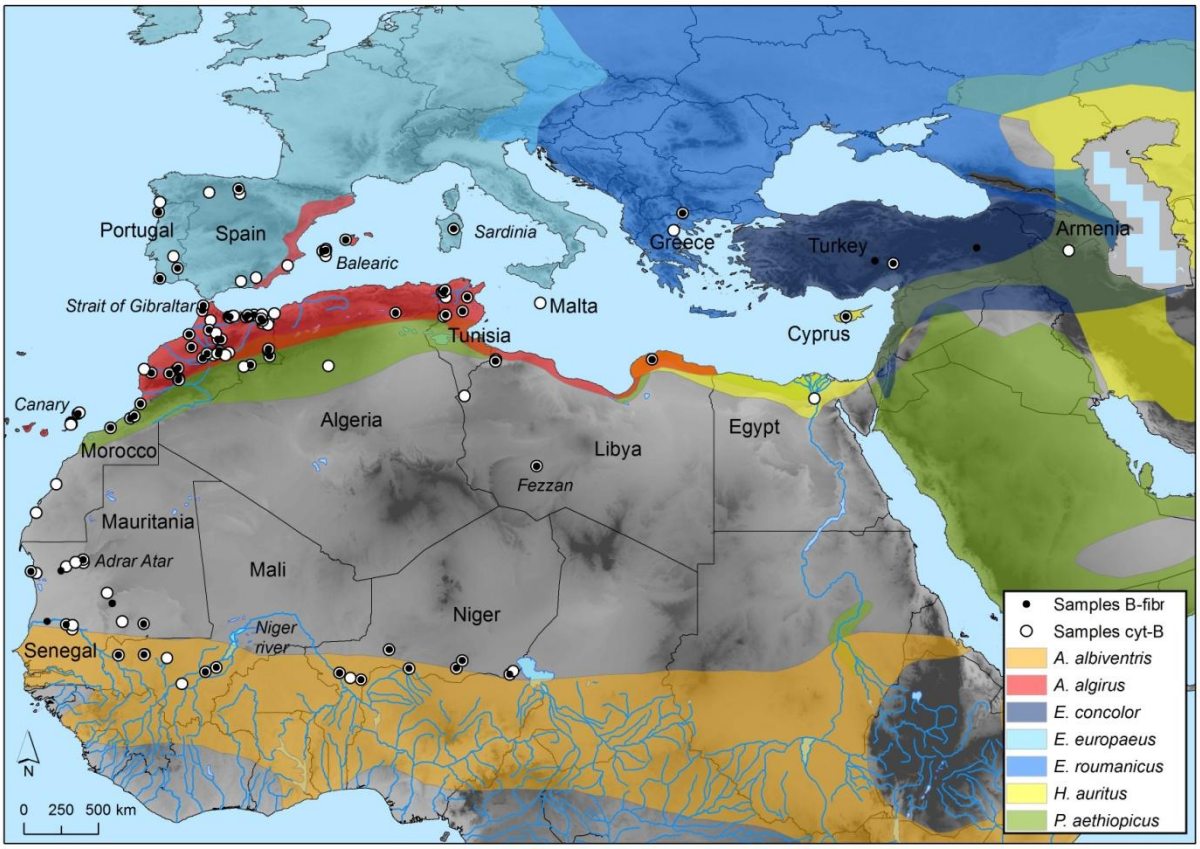


Figure 5 – Geographic distribution of hedgehog samples amplified in this study for mitochondrial (N=136; white dots) and nuclear genes (N=78; black dots). Non-amplified samples and samples taken from Genbank and outgroups are not represented. Distribution of the different taxa according to IUCN (2013) are indicated by different color patches.

2.2. Laboratory procedures

2.2.1. DNA extraction

From the available dataset, 16 samples were spines, nails or non identified tissue and were not extracted. The genomic DNA of the remaining samples was obtained using a JETQUICK Tissue DNA Spin Kit from Genomed with the buffers and solutions provided with the kit and following the manufacturer's protocol (Annex A1).

To analyze the success of the DNA extractions, samples were verified by electrophoresis on a 0.8% agarose gel (with GelRed) dipped in TBE 0.5x for 15/20 minutes at 250V. Agarose gel was visualized through UV radiation in a Bio-Rad.

All samples were frozen until needed for the Polymerase chain reactions (PCR) process. The dilutions for PCR process were only performed at the time of amplifying certain sample, because often, the same sample needed different levels of dilution.

Many samples were extracted twice due to negative results in the extraction or amplification processes.

2.2.2. mtDNA amplification and sequencing

The cytochrome b (Cyt-b) region of hedgehogs' mitochondrial DNA was almost completely amplified using different primers. In the first attempt we choose a small number of samples to amplify a fragment, combining in different ways the universal primers GluDG.L, CB3H, CB2F and CB2R. Once the results were positive for only some of the species and the size of amplified fragments were much lower than the desired we rejected this set of primers.

A pair of specific primers (forward and reverse) was designed from a complete genome sequence of *E. europaeus* available on GenBank. Primer development was performed manually using PCR primer design guidelines and subsequently confirmed in Primer3Plus, a web-interface for primer3 (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/). The primers EurCYTbF (5'- GTTGTAACTCAACTACAAG -3') and EurCYTbR (5'- TATACTACGAAGACTCTTC -3'), forward and reverse, respectively, were designed outside the ends of the Cyt-b gene fragment [14157 – 15335 base pair (bp)]. PCR reactions were performed in 10 µl reaction volumes composed of 5.6 µl of pure water, 4 µl of Master Mix (Taq PCR Master Mix, QIAGEN), and 0.2 µl of each primer (forward and reverse). For all reactions was added an amount of DNA, between 0.8µl and 5µl, depending on the quality of the sample. Every PCR was done with a negative control to check for contaminations. DNA amplifications were developed using a Biometra T-professional Thermocycler and a C1000 thermal cycler. Both thermocyclers shared the same PCR program, which began with an initial denaturation step at 94°C (15 min), followed by 45 cycles of denaturation at 94°C (30sec), annealing temperatures of 43°C, 47°C or 51°C (35 sec), extension at 72°C (1.15 min), and concluded with a final extension at 60°C (20 min).

The level of degradation of some samples was so high that those samples did not amplify and it was necessary to design new primers. These new primers were designed in order to amplify two fragments that when concatenated they formed a large fragment, with a similar length to those previously amplified sequences. Primers were designed in the most conservative regions found in the existing alignment. The first pair of primers AfricytbF (5'- CATTGATTTACCTACTCCACT -3') and HedCYTBintR (5'- GGGTGGAATGGTACTTTATC -3') was within the first half of Cyt-b gene fragment [14235 – 14846]. The second pair of primers HedCYTBintF (5'-

GCATTCTCAGTAGATAAAGC -3') and AfricytbR (5'- AATAAGGCCTGCTAATGGTAA -3') was within the last half of Cyt-b gene fragment [14689 – 15297]. PCR reactions were performed separately for each primers in 10 µl reaction volumes composed of 5.2 µl of pure water, 4 µl of Master Mix (Taq PCR Master Mix, QIAGEN), 0.4 µl of BSA (Bovine Serum Albumin 20 mg (1ml)) and 0.2 µl of each primer (forward and reverse). DNA added depended on the quality of the sample. Thermocyclers were used as before with the same PCR program for both new pair of primers, which began with an initial denaturation at 94°C (15 min) followed by a 20 cycles touchdown with a 0.5°C decrease every cycle. The 20 cycles started at 94°C (30 sec), passing 58°C (45 sec), and finishing with 72°C (45 sec). This process was followed by 20 cycles at 94° (30 sec), 48°C (45 sec), and 72°C (45 sec). A final extension step was done at 60°C (20 min).

The resulting PCR products were visualized in 2% agarose gels (with GelRed), using a M5 ladder and following the same electrophoresis' process as the one in extractions.

Every positive PCR product was cleaned up before sequencing. It was prepared a master mix with 1,5 µl of ExoSAP, being 0,5µl of Exo (Exonuclease I) to digest the excess of primers and 1 µl of SAP (FastAP Thermosensitive Alkaline Phosphatase) to degrade nucleotides. For each 4 µl of PCR sample was added 1,5 µl of ExoSAP and the resulted solutions were incubated in a thermocycler at 37°C for 15 min, followed by 15 min at 80°C. Sequencing process was done by a commercial company (Macrogen Inc., Netherlands), using the same primers used in amplification process.

The edition and alignment of sequences was performed using BioEdit 7.1.3.0. and the ClustalW Multiple alignment tool, retaining the default settings. Chromatograms, corresponding to forward and reverse directions, were verified by eye and the translation of sequences into protein sequences did not reveal any stop codons, suggesting functional sequences were obtained. Sequences available in Genbank for Cyt-b that were larger or had the same length of sequences that we obtained through our primers were added to the alignment. The final mtDNA data of the Cyt-b gene was 1012 bp.

2.2.3. nDNA amplification and sequencing

Nuclear DNA amplification was performed in order to obtain some representatives of each haplotype obtained with mitochondrial gene. We selected the intron 7 of the ß-fibrinogen (ß-fibr) and the second myoglobin intron (myo), already used in two species of hedgehogs in another study (Seddon et al., 2001). Testing of the capacity of each

gene for amplification suggested that only the ß-fibr gene could amplify almost all specimens of this study.

The ß-fibr was amplified using the primers BFIBR1 (5'-ATTCACAACGGCATGTTCTTCAG -3') and BFIBR2 (5'-AANGKCCACCCCAGTAGTATCTG -3'), forward and reverse, respectively. PCR reactions were composed of 5.6 µl of pure water, 4 µl of Master Mix (Taq PCR Master Mix, QIAGEN), and 0.2 µl of each primer (forward and reverse). The amount of DNA added depended on the quality of the samples used. A negative control was always added. PCR reactions were carried out in a Biometra T-professional Thermocycler using a touchdown program with an initial denaturation step at 94°C (15 min), followed by a 20 cycles touchdown with a 0.5°C decrease every cycle. The 20 cycles started at 94°C (30 sec), passing 58°C (45 sec), and finishing with 72°C (45 sec). This process was followed by 20 cycles at 94° (30 sec), 48°C (45 sec), and 72°C (45 sec). A final extension step was done at 60°C (20 min). Amplification results were observed through electrophoresis in 2% agarose gels (with GelRed), using a M5 ladder and following the same electrophoresis process described above.

Before DNA sequencing, every positive PCR product was purified following the same processes as previously described in mtDNA. PCRs products were sent to Macrogen for sequencing, using the same primers used in amplification process. Given that the initial sequencing results showed that the forward primer was not able to properly sequence the fragment, we used only the reverse primer in the sequencing process. Every sequence's chromatograms were visually analyzed using BioEdit 7.1.3.0. program. The sequences that showed polymorphic positions, represented by two peaks of different colors in the same nucleotide position, were sequenced twice to ensure that it was not the result of sequencing errors or bad cleaning. In those sequences, whose second sequencing result showed the same polymorphism, it was used IUB codes for the positions with double peaks and sequences were marked as heterozygous. These letter codes substitute the two possible letters of nucleotides for just one letter, and the substitution is the following: G or T = K, G or C = S, A or T = W, A or C = M, C or T = Y and A or = R.

Alignment of the ß-fibr sequences, in just reverse direction, was assisted by BioEdit 7.1.3.0. program, using the clustal method and retaining the default settings. Hedgehog sequences for the same gene were taken from Genbank and added to the alignment of 675 bp.

2.3. Data Analysis

2.3.1. mtDNA analysis

It was obtained from GenBank further sequence data from hedgehogs to complement our dataset, corresponding to species: *Erinaceus amurensis*, *Erinaceus auritus*, *Erinaceus europaeus*, *Mesechinus dauuricus*, *Mesechinus hughi* and *Paraechinus aethiopicus*. Sequences from *Neohylomys hainanensis*, *Neotetracus sinensis and Sorex unguiculatus* were also downloaded to act as outgroup, once the first two mammals are representatives of the subfamily Galericinae, belonging to the same family of hedgehogs, and *Sorex* is a distant relative from a different Order. The dataset was collapsed to unique haplotypes by use of DnaSP version 5 program (Librado and Rozas, 2009) to facilitate all subsequent phylogenetic analyses.

Preliminary analyses of the phylogeny were made by building phylogenetic trees using the program MEGA version 5.1. (Tamura et al., 2011). Definitive analyses of phylogenetic relationships of all samples sequenced and downloaded were assessed by three different methods: Maximum Likelihood (ML), Neighbor-Joining (NJ) and Bayesian inference. The best model of evolution was selected using ModelTest tool (Posada and Crandall, 1998) in the HYPHY package according to the Akaike Information Criterion (AIC). This tool give a best fit model for the General Time Reversible model with gamma distributed rate variation among sites and a proportional of invariable sites (GTR+G+I). The ML analyses were performed using PhyML version 3.0 software (Guindon et al., 2010), where was instructed to use the GTR model and to perform 1000 bootstrap replications. NJ method was done using the PHYLIP package (Felsenstein, 2005). It was generated random samples by bootstrapping using the “seqboot” command for 1000 replicates, followed by DNA distance matrix calculations using F84 distance (“Dnadist” command) and Neighbor-Joining and UPGMA tree drawing method. A consensus tree was also constructed from the 1000 random trees obtained in this program, using the “consense” command. For BI method, it was used MrBayes 3.2 software (Huelsenbeck and Ronquist, 2001) and some parameters were changed according to the suggested model. Runs were performed with unequal rates of substitution (nst=6) and gamma-shaped rate variation (rates=invgamma) for 10,000,000 generations, started with randomly generated trees. All phylogenetic programs ran two times with different seeds to compare results. The resulting trees were visualized and edited resorting to the program FigTree version 1.3.1 (Rambaut, 2009).

The haplotype network was constructed in TCS v1.21 program (Clement et al., 2000) using statistical parsimony to calculate the minimum number of mutational steps among haplotypes and connections based on a 95% confidence interval. Given that some of the haplotypes were extremely distance from each other, it was necessary to fix the connection limit at 250 steps, in order to link them and be able to see the number of mutational steps between each haplotypes. It was necessary to built three different networks for the different genera. The first network was realized taking into account the *Atelerix* sequences, the second with the sequences of *Erinaceus*, *Paraechinus*, *Mesechinus* and *Hemiechinus* and the last network included sequences from *Atelerix* and *Erinaceus* to establish the connection and obtain the number of mutations between these genera.

Finally, it was inferred genetic diversity (haplotype, Hd, and nucleotide diversity, π) of grouping populations with Arlequin version 3.5.1.2 (Excoffier et al., 2005). Two statistical tests commonly used to analyze demographic events and assess the selective neutrality of mutations (Tajima's D and Fu's Fs) were calculated using DnaSP version 5 program (Librado and Rozas, 2009). Associated statistical significances were generated using 5000 simulations. Sequence divergence between mtDNA populations based on Kimura 2-parameter model were calculated in MEGA version 5.1. (Tamura et al., 2011).

2.3.2. nDNA analysis

The hedgehog sequences available in GenBank for ß-fibr gene corresponded to *Erinaceus concolor* and *Erinaceus europaeus.* The only sequence taken from GenBank to act as outgroup was an *Anourosorex squamipes* from the family Soricidae.

Some sequences showed heterozygote positions, impossible to unravel only by the analysis of chromatograms. It was used the PHASE tool incorporated in DnaSP version 5 program (Librado and Rozas, 2009) to estimate haplotypes for each individual with uncertain phase sites. This analysis was repeated three times and checked if haplotype estimation was consistent across runs. For each run, it was used the following MCMC parameters: burn of 10,000 steps, 10 step thinning intervals, and 100,000 iterations. In the resulted phased data, sequences whose alleles estimated had a percentage value below 89% were deleted from the dataset. As it was done with mitochondrial sequences, also nuclear sequences were collapsed into haplotypes by DnaSP version 5 program (Librado and Rozas, 2009).

Phylogenetic analyses were performed using the same three methods used for mitochondrial analysis: Maximum Likelihood (ML), Neighbor-Joining (NJ) and Bayesian

inference. The ModelTest tool (Posada and Crandall, 1998) inserted in the HYPHY package was used in order to find the most adequate evolution model, according to the AIC. The selected model was the HKY85 which assumes that rates of substitution differ between each nucleotide. Bayesian analysis was run using MrBayes 3.2 software (Huelsenbeck and Ronquist, 2001). Runs were executed with unequal rates of substitution (nst=2) for 10,000,000 generations, started with randomly generated trees. The ML analysis was conducted using PhyML version 3.0 software (Guindon et al., 2010), based in the HKY85 model with 1000 bootstrap replications. NJ method was done using the PHYLIP package (Felsenstein, 2005). Random sampling was generated by bootstrapping using the "seqboot" command for 1000 replicates, followed by DNA distance matrix calculations using F84 distance ("Dnadist" command) and Neighbor-Joining. From the multiple trees obtained in this program, it was generated a consensus tree. Every program analysis was repeated to check for consistencies in the results. Trees were visualized and edited with FigTree version 1.3.1 (Rambaut, 2009).

The nuclear haplotype network was generated using parsimony calculations in TCS v1.21 program (Clement et al., 2000). Connections were based on a 95% confidence interval. The existing gaps were considered as a fifth state and since the haplotypes were relatively close, the connection limit was fixed at 30 steps.

The genetic diversity (Hd and π) of the different populations was estimated with Arlequin version 3.5.1.2 (Excoffier et al., 2005). It was also calculated Tajima's D and Fu's Fs values in DnaSP version 5 for the same defined populations and genetic distances were established using MEGA version 5.1. (Tamura et al., 2011).

3. Results

3.1. Laboratory overview

From the 289 samples initially available to perform this study, 136 were successfully amplified for the mitochondrial gene (Figure 5), with individuals representing almost every country where the sampling was taken (exceptions were France, Ukraine and Serbia). For the nuclear gene, 78 samples were amplified, which five of them were solely amplified for this molecular marker (Figure 5). Derived from their natural advanced stage of degradation, many samples presented low DNA concentration in the extractions and did not amplify with any of the PCR programs used. Detailed information of amplified sequences is available in the Annex (section A2).

3.2. Mitochondrial DNA

The final Cyt-b alignment contained 147 sequences (136 obtained by amplification and 11 downloaded from GenBank, see section A2 and A3 in the Annex for more details). Resulting sequences of 1012 bp long included 444 polymorphic sites and a total of 602 mutations (no insertions, deletions or missing data were found). A total of 89 haplotypes were identified among the analyzed samples. Samples were generally located in areas that corresponded to known ranges of distinct species, but in Africa many samples were located outside the ranges proposed by IUCN (2013): along the Atlantic Sahara and in central Libya (Figure 6A).

The different phylogenetic trees recovered unanimously eight divergent and well-supported clades, but there were some disagreements in topologies, namely in branches connecting groups of clades (Figures 6B and 7). Neighbor-joining analysis recovered the same main lineages of the other methods and the same topology of Maximum Likelihood analysis, but statistical support was generally very low and bootstrap values are not shown. The eight clades obtained in phylogenetic trees were: 1) Erinaceinae (in red); 2) A.alb1 (dark brown); 3) A.alb2 (light brown); 4) E.eur/E.amu (light blue); 5) E.con/E.rou (dark blue); 6) H.aur (yellow); 7) Mesechinus (orange); and 8) Paraechinus (green).

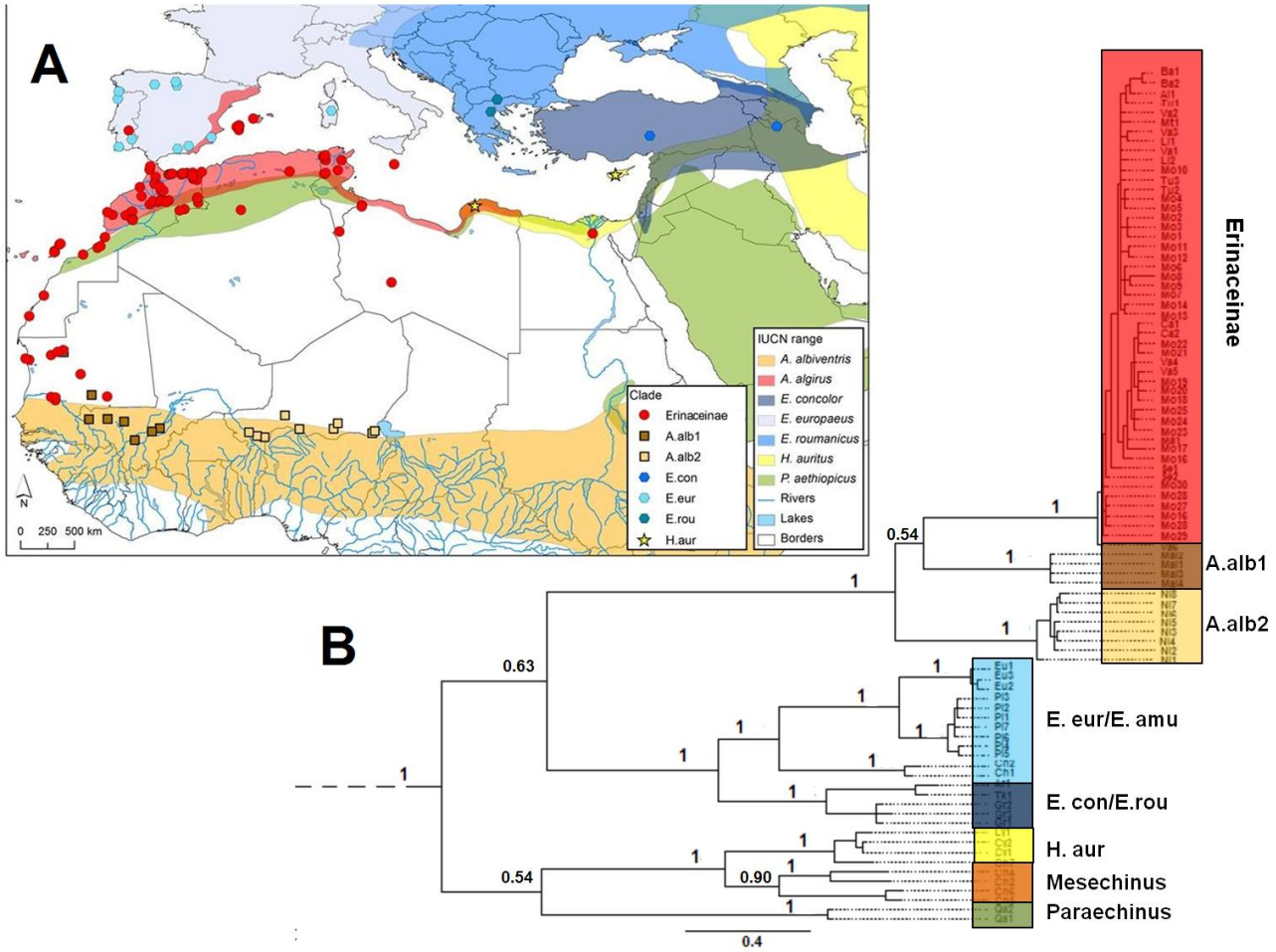


Figure 6 - (A) Geographic distribution of samples belonging to each main clade recovered by the phylogenetic tree based on Cyt-b and distribution of the different taxa according to IUCN (2013). Colors and symbols in both figures represent the main clades obtained by the phylogenetic analysis. (B) Bayesian phylogenetic tree based on mtDNA data (Cyt-b) for different hedgehog representatives and rooted with three outgroups (see Annex A4 for complete tree). Numbers above branches correspond to Bayesian posterior probabilities (BPP) obtained from Bayesian inference.

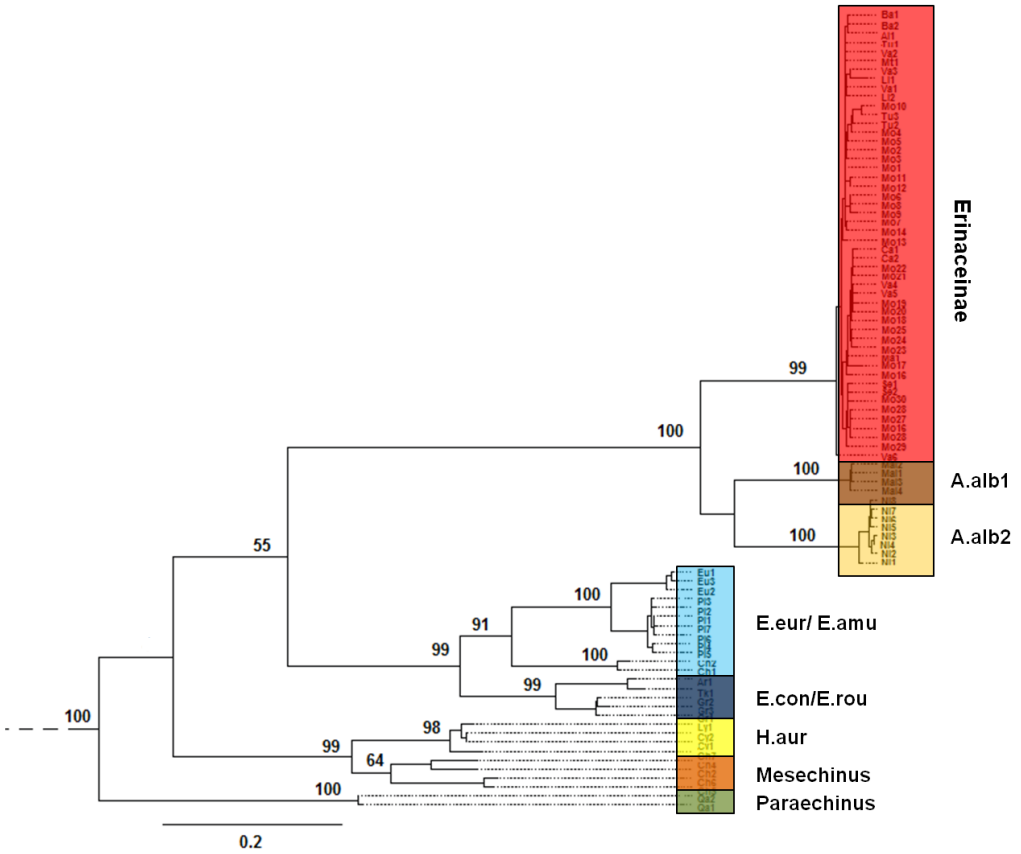


Figure 7 - Maximum Likelihood phylogenetic tree based on mtDNA data (Cyt-b) for different hedgehog representatives and rooted with three outgroups (see Annex A5 for complete tree). Numbers above branches correspond to bootstrap values greater than 50 (lower values are not shown). Colors define the main clades obtained and are the same used as in Figure 6.

The Erinaceinae clade groups' haplotypes present in Morocco, south-western Mauritania and northern Senegal, northern Algeria, Libya, northern Egypt, as well as in Canary, Balearic, and Malta islands, and in southern Portugal (Figure 6A), covering areas where the species *Atelerix algirus* is known to occur but including also many new areas where this species was previously unknown (e.g. Mauritania and southern Libya). The clades A.alb1 and A.alb2 appear in the known distribution area of *Atelerix albiventris* and both have haplotypes exclusively from African territory. The clade A.alb1 groups samples from Mali and Mauritania, west of Niger River, and one sample appears in sympatry with one sample from Erinaceinae clade, in the Adrar Atar mountains of Mauritania. The clade A.alb2 groups all samples from Niger, east of the Niger River. In both BI and NJ trees (Figure 6B and 7) E.eur/E.amu clade shows a substantial split with two subgroups, highly supported by bootstrap values. One subgroup (BPP = 1; 100% bootstrap support) contains samples distributed in the

known range of *Erinaceus europaeus* from Iberian Peninsula, Italy and Sweden, and the other one (BPP = 1; 100% bootstrap support) contains *E. amurensis* samples from China (location not represented in Figure 6A). There is a subdivision in the E.con/E.rou clade where one subgroup (BPP = 1; 100% bootstrap support) includes samples from Armenia and Turkey located in the known range of *E. concolor* and the other subgroup (BPP = 1; 94% bootstrap support) contains samples from Greece located in the known range of *E. roumanicus*. The clade H.aur comprises samples from Libya, Cyprus and China located in the known range of *Hemiechinus auritus* (location of sample from China not illustrated in Figure 6A). The Mesechinus and Paraechinus clades (not shown in the Figure 6A) contain samples from China and Qatar, respectively, located in the known range of *Paraechinus aethiopicus* and *Mesechinus dauuricus* and *Mesechinus hughi*, respectively. Common to both trees, is the basal position of Paraechinus clade (better supported in NJ tree) and the sister taxon relationship between Mesechinus and H.aur (BPP=1; 99% bootstrap support). Concerning relationships in *Erinaceus* species, *E. europaeus* and *E. amurensis* are closest to each other and the same is verified between *E. concolor* and *E. roumanicus*. Uncertainties were found in the relationships between clades Erinaceinae, A.alb1 and A.alb2: in the BI tree, the main branch split divides [Erinaceinae + A.alb1] and A.alb2 (BPP=1), while in the ML tree, the main division occurs between Erinaceinae and [A.alb1 + A.alb2].

The genetic structuring within clades is best depicted by the haplotype network (Figure 8), which exhibited similar patterns to those by the phylogenetic trees. In North Africa, the haplotypes of clade Erinaceinae (red circles) were represented by a few number of individuals and do not display clear geographical coherence and structure, demonstrating very low levels of sequence divergence. Some haplotypes are shared by representatives from Mauritania and Morocco (Va4, Va5, Va6) or even with Tunisia (Va1). Haplotype sharing was also detected between sequences from Mauritania, Libya and Egypt (Va3). The samples from Balearic (Ba1 and Ba2) and Canary Islands (Ca1 and Ca2) appear closely related to those ones existing in Algeria and Morocco, respectively. Haplotypes Mo8 and Mo17 represent more distinct lineages, separated from the main cluster of haplotypes by nine and seven mutational steps, respectively. The clades A.alb1 and A. alb2 (represented on figures by a dark and light brown, respectively) form two distinct clusters, separated by 94 mutational steps, almost with the same amount of steps that separates the clades Erinaceinae and A.alb1 (103 mutational steps). As in the Bayesian tree (Figure 6B), the clades Erinaceinae and A.alb1 appear more related to each other. The haplotype of clade H.aur from Libya (Ly1) is more closely related with haplotypes from Cyprus (Cy1 and Cy2), separated by 8 mutational steps, than to that haplotype from China (Ch7). Outside Africa, the genetic

structure is stronger and clearer. The distribution of haplotypes and the number of mutational steps between distinct groups appear to represent the distribution of the different species and genera used. The haplotypes present in Iberian Peninsula (PI1-PI7), corresponding to *Erinaceus europaeus*, appear separated by a large number of mutations (48) from the ones present in remaining Europe (Eu1-Eu3). This number of mutations is similar to the one separating clade E.con/E.rou (54 mutational steps), which correspond to *E. concolor* and *E. roumanicus*. The genera *Erinaceus*, *Hemiechinus*, *Mesechinus* and *Paraechinus* are all separated from each other by values near to one hundred mutational steps.

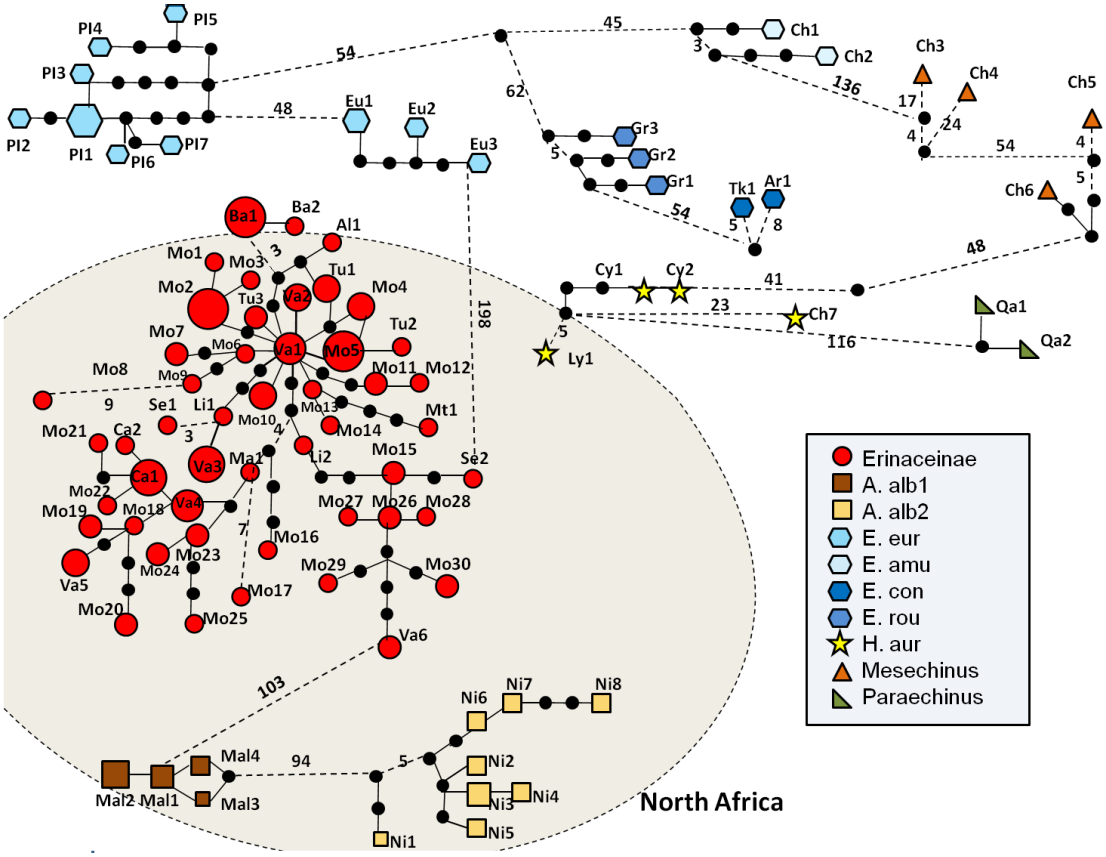


Figure 8 - Haplotype network assuming a 95% parsimony threshold, based on the same dataset used for phylogenetic analyses. The size of each haplotype symbol is proportional to its frequency and branch lengths are proportional to the number of mutational steps among haplotypes (black dot = mutation). Numbers near dashed lines correspond to mutational steps separating the main clades or in some cases to haplotypes inside groups. Names of haplotypes are also represented. Both symbols and colors are concordant with those in Figure 6A, with the exception of symbols representing E.amu and Mesechinus haplotypes from China and Paraechinus haplotypes from Qatar, that previously were not represented by symbols.

Calculations of genetic diversity of the main clades (Table 1) show consistent high levels of haplotype diversity, ranging from Hd=0.778 in A.alb1 to Hd=~1 (in most cases). Nucleotide diversity was high in the clades that share more than one species,

the case of E.eur/E.amu, E.con/E.rou and Mesechinus. Demographic analysis developed through Tajima's D and Fu's Fs tests provided different results between the groups analyzed. Erinaceinae shows significantly negative values of D and Fs (-1.727; -59.149). Significant values of tests of selective neutrality are expected for markers under selection or/and populations undergoing recent growth, with changes in population sizes, carrying an excess of rare haplotypes. Also A.alb1, A.alb2 and E.eur/E.amu obtained negative values for both tests, of which only A. alb2 obtained a significantly negative value of Fs, suggesting a strong selection or population change effect on this group. The clades E.con/E.rou and Mesechinus shows positive values of D and Fs.

Table 1 – Genetic diversity and demographic measures for the main clades found in the analyses of Cyt-b dataset. N – number of sequences; S – polymorphic sites; H – number of haplotypes; Hd – haplotype diversity; π – nucleotide diversity; D – Tajima's D; Fs – Fu's Fs. Standard deviations are shown in brackets. Just the significant values of D and Fs (P<0.05) have their correspondent p-values below. Demographic tests were not performed in *Paraechinus* clade (represented with -) due to the reduced sample size.

Clade	N	S	H	Hd	π	D	Fs
Erinaceinae	97	81	50	0.979	0.008	-1.727	-59.149
				(0.005)	(0.004)	(0.020)	(0.001)
A.alb1	10	3	4	0.778	0.001	-0.754	-2.367
				(0.091)	(0.001)		
A.alb2	9	19	8	0.972	0.005	-0.941	-3.398
				(0.064)	(0.003)		(0.020)
E.eur/E.amu	16	153	12	0.942	0.046	-0.649	-0.173
				(0.048)	(0.023)		
E.con/E.rou	5	80	5	1	0.043	1.027	1.393
				(0.127)	(0.026)		
H.aur	4	34	4	1	0.017	-0.584	0.968
				(0.177)	(0.012)		
Mesechinus	4	112	4	1	0.065	0.813	2.368
				(0.177)	(0.043)		
Paraechinus	2	3	2	1	0.003	-	-
				(0.500)	(0.003)		

Genetic distances calculated through the Kimura-2 parameter (K2p) reinforced the results obtained by the phylogenetic analyses (Table 2). The clades Erinaceinae, A.alb1 and A.alb2 exhibited the lowest levels of divergence between them, with K2p-corrected genetic distances ranging from 11.2 % (A.alb1–A.alb2) to 13.5%

(Erinaceinae–A.alb2), while the distance between H.aur and Mesechinus was relatively lower (10.9%). The highest distance values were obtained between African and Non-African groups, the latter exhibiting the most significant genetic distances: 25.7% between Erinaceinae and Paraechinus and 25.4% between H.aur and A.alb2.

Table 2 – Mean genetic distances between the main clades of hedgehogs used in this study (below diagonal). Standard error are shown above the diagonal.

	Erinaceinae	A.alb1	A.alb2	E.eur/ E.amu	E.con/ E.rou	H.aur	Mesechinus	Paraechinus
Erinaceinae	-	0.011	0.012	0.017	0.017	0.018	0.016	0.018
A.alb1	0.122	-	0.012	0.016	0.017	0.017	0.016	0.016
A.alb2	0.135	0.112	-	0.017	0.016	0.018	0.016	0.018
E.eur/E.amu	0.261	0.238	0.237	-	0.011	0.015	0.015	0.017
E.con/E.rou	0.236	0.241	0.238	0.141	-	0.016	0.014	0.016
H.aur	0.250	0.252	0.254	0.224	0.213	-	0.01	0.014
Mesechinus	0.243	0.234	0.244	0.227	0.208	0.109	-	0.014
Paraechinus	0.257	0.226	0.255	0.234	0.224	0.182	0.194	-

3.3. Nuclear DNA

Sequences for ß-fibr gene of 87 samples (78 amplified in this study and 9 acquired from GenBank, see section A2 and A3 in the Annex for more details) resulted in an alignment of 675 bp, of which 76 were polymorphic sites (71 after excluding gaps). After collapsed, nuclear dataset was reduced to 27 haplotypes for subsequent analyses.

All phylogenetic methods inferred similar trees, with high statistical support in Bayesian tree but lower values in ML and NJ trees (Figure 9B). Six main clades were retrieved with locations partially overlapping with known distributions of several species: 1) A.alg (*Atelerix algirus* – in green); 2) A.alb1 (*A. albiventris* – dark brown); 3) A.alb2 (*A. albiventris* – light brown); 4) Erinaceus (*E. europaeus*, *E. concolor* and *E. roumanicus* – light blue); 5) Erinaceinae cryptic (Uncertain species – in red); and 6) H.aur (*Hemiechinus auritus* – yellow). The phylogenies obtained with the nuclear and mitochondrial markers differed in one main feature: samples of clade Erinaceinae that clustered together in the mtDNA marker (Figures 6B and 7) now appeared separated into two different clades, named A.alg (BPP= 0.99) and Erinaceinae cryptic (BPP= 0.99). The former includes samples from northern Morocco, Algeria, and Tunisia, from southern Libya, and from the Canary and Baleares Islands (Figure 9A). The latter clade is represented by haplotypes from Mauritania, Senegal and northwestern from Libya. The clades appeared phylogenetically distant. The mtDNA clades E.eur/E.amu and E.con/E.rou here appear collapsed into clade Erinaceus (BPP= 1), because there are much less haplotypes from each species, E.amu species is not represented. The clades A.alb1 (BPP= 0.98) and A.alb2. (BPP= 0.97) are consistent in both mt- and nDNA, and are represented to the west and to the east of Niger River, respectively. Both clades branch together with A.alg (BPP= 0.99). Clade H.aur is represented by haplotypes from Libya and Cyprus, but despite its clear separation from other clades, statistical support for this division was not recovered. Sequences of *Mesechinus* and *Paraechinus* were neither amplified nor available in GenBank, and therefore these genera were not represented.

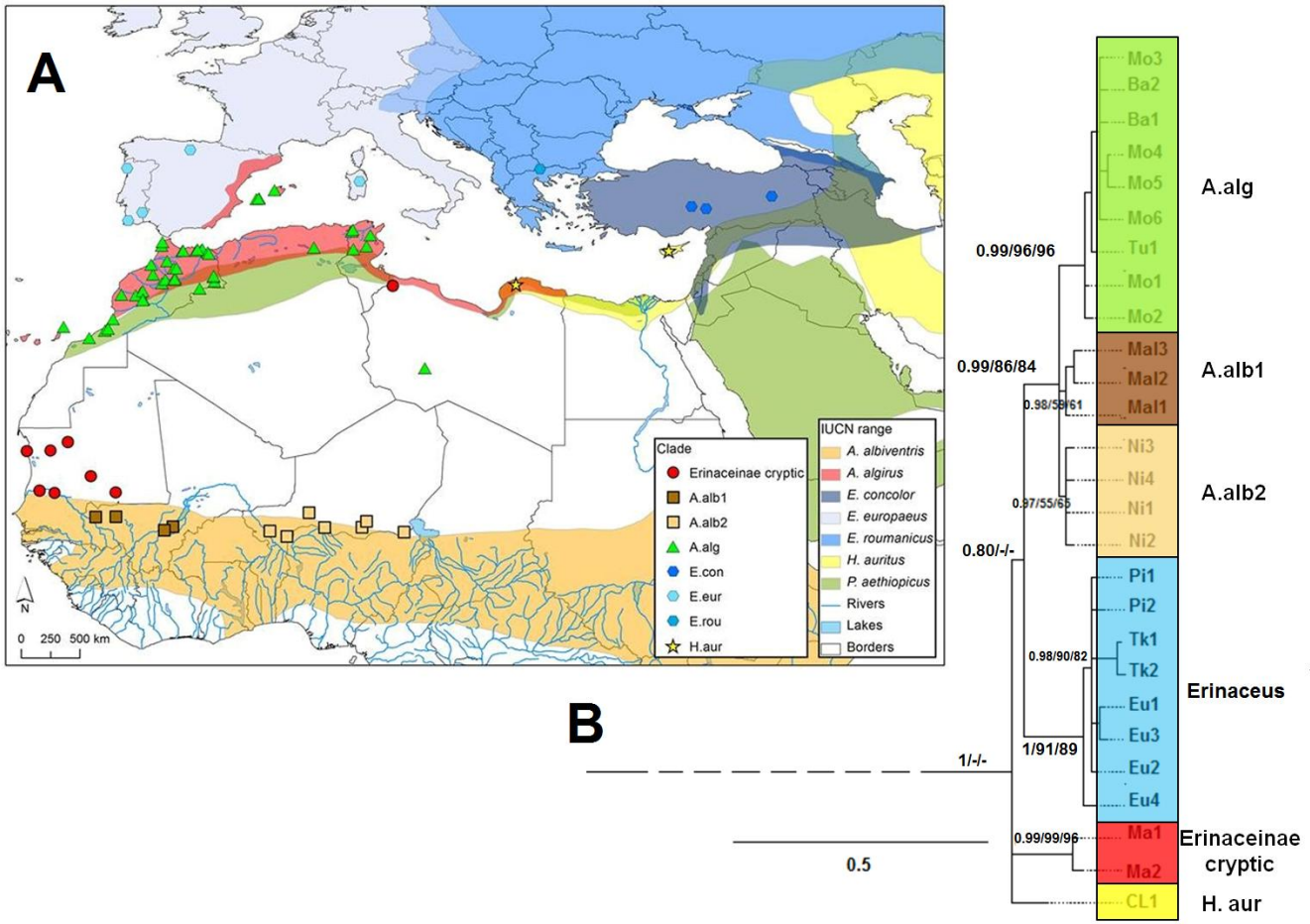


Figure 9 - (A) Geographic distribution of samples belonging to each main clade recovered by the phylogenetic tree based on ß-fibr gene and distribution of the different taxa according to IUCN (2013). Symbols and colors correspond to different group of samples and are concordant with the colors used in the phylogenetic tree, with the exception of E.con, E.eur and E.rou that in the tree are all represented by the same color (light blue). (B) Phylogenetic tree based on nDNA (ß-fibr) data for different hedgehog representatives and rooted with one outgroup (see Annex A6 for complete tree). Values of Bayesian posterior probabilities and bootstrap values of ML and NJ are above or near the branches. Colors represent the main groups obtained and are concordant with the colors seen in Fig. 9B.

Relationships between haplotypes depicted in the haplotype network (Figure 10) exhibit six haplotype networks that correspond to the clades found in the phylogenetic trees. The clade E.eur is constituted exclusively by samples from outside of Africa, while the opposite is found in clades A.alb1 and A.alb2 that just contain samples from Africa.

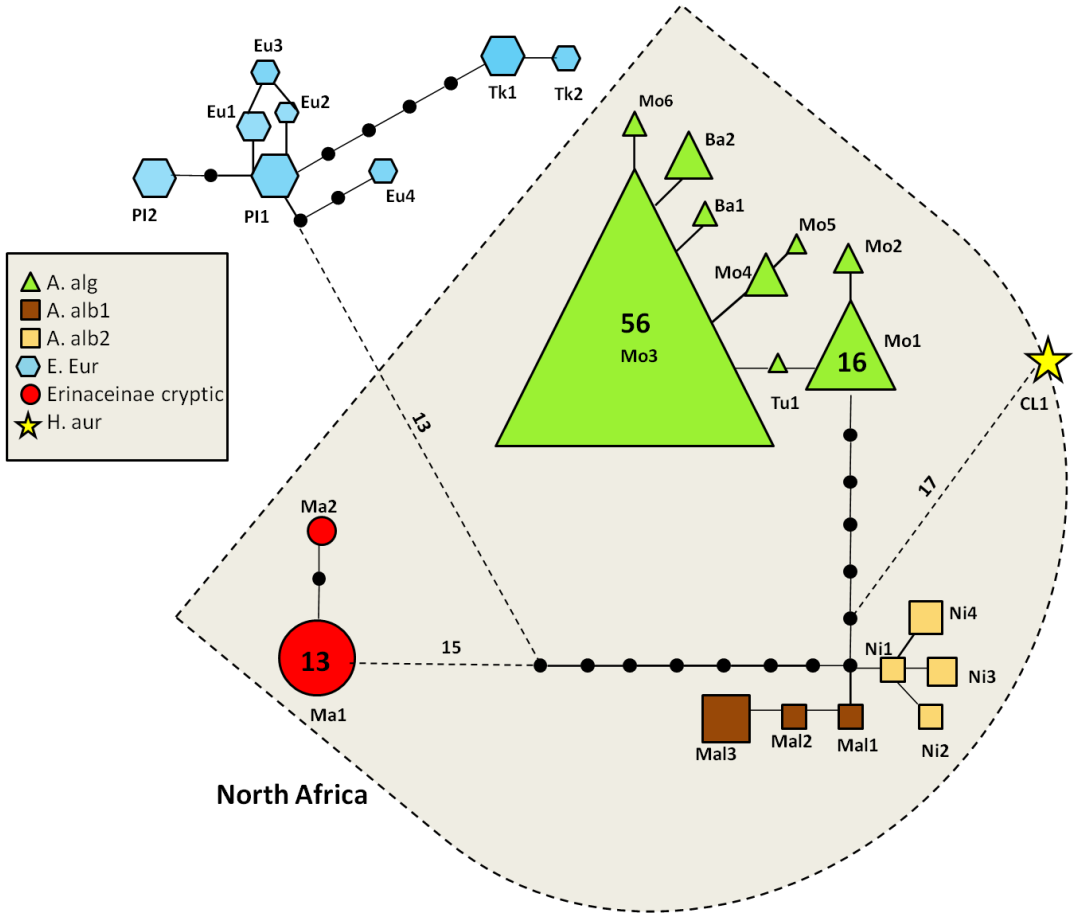


Figure 10 – Statistical parsimony haplotype network based on nDNA dataset. Figure sizes and branches are proportional to the number of shared individuals per haplotype (with the exception of bigger haplotypes that have the number inside the figure). Numbers above dashed lines refer to mutational steps separating main clades. Haplotype names are near or inside the symbols. Colors and symbols are concordant with the ones used in figure 9B.

The network exhibits high frequency of common haplotype, from clade A.alg (Mo3), shared by 56 samples (including samples from Canary Islands, Algeria, Tunisia and Morocco) with few smaller haplotypes around it, present in the Balearic Islands (Ba1 and Ba2), Tunisia (Tu1), and Morocco (Mo1, Mo2, Mo3-Mo6). This group seems relatively close to other Atelerix clades, namely A.alb1 and A.alb2 (6 mutational steps separating them), which supports previous results found in mtDNA. Clades A.alb1 and

A.alb2 are placed next to each other, demonstrating low levels of divergence. The new North African group, Erinaceinae cryptic, shows a high number of mutational steps (23) separating it from the other clades. This number is close to the one separating haplotypes from different genus: 21-26 steps between *Atelerix-Erinaceus* genera, and 19-22 mutations steps between *Atelerix–Hemiechinus* genera.

Diversity indices (Hd and π) were calculated for the six main groups (Table 3). Despite the low number of haplotypes in each clade (from H=1 in H.aur to H=9 in A.alg) some results are worth highlighting. High levels of haplotype diversity were found in clades A.alb2 and Erinaceus (Hd= 0.800 and Hd= 0.878, respectively). In the latter, it was also found one of the highest level of nucleotide diversity (π= 0.006), Tajima's D and Fu's Fs values. All other analyzed clades exhibited negative values of D and Fs. In three clades, it was not possible to perform these tests since the number of haplotypes was too small (only possible for 4 or more haplotypes). Significant negative value of Fs was found in A.alg.

Table 3 – Genetic diversity and demographic measures for the main lineages found in the previous analyses of ß-fibr dataset. N – number of sequences; S – polymorphic sites; H – number of haplotypes; Hd – haplotype diversity; π – nucleotide diversity; D – Tajima's D; Fs – Fu's Fs. Standard deviations are shown in brackets. Just the significant values of D and Fs (P<0.01) have p-values shown below. Demographic tests were not performed in some clades (represented with -) due to the reduced sample size.

Clade	N	S	H	Hd	π	D	Fs
A.alg	94	8	9	0.611	0.002	-0.855	**-7.979**
				(0.051)	(0.001)		(0.001)
A.alb1	11	2	3	0.582	0.001	-	-
				(0.142)	(0.001)		
A.alb2	11	3	4	0.800	0.002	-0.85519	-2.367
				(0.075)	(0.001)		
Erinaceus	27	13	8	0.878	0.006	-0.51889	-4.234
				(0.029)	(0.003)		
Erinaceinae cryptic	16	2	2	0.325	0.001	-	-
				(0.125)	(0.001)		
H.aur	6	0	1	0	0	-	-
				-	-		

The genetic distances for the nucleotide dataset show overall low values of interspecific divergence (Table 4) that are lower when compared with the ones obtained for mitochondrial analyses.

The lowest K2p-correct genetic distance value was obtained between A.alb1 and A.alb2 (0.6%). The same distance value was obtained between A.alg–A.alb1 and

A.alg–A.alb2 (1.4%). The highest genetic distances were obtained between Erinaceus–Erinaceinae cryptic (4.6%), A.alg-Erinaceus (4.5%), and A.alg–Erinaceinae cryptic (4.4%).

Table 4 - Mean genetic distances between the main clades of hedgehogs found in nDNA dataset (below diagonal). Standard error are shown above the diagonal.

	A.alg	A.alb1	A.alb2	Erinaceus	Erinaceinae cryptic	H.aur
A.alg	-	0.004	0.004	0.008	0.008	0.007
A.alb1	0.014	-	0.002	0.007	0.008	0.006
A.alb2	0.014	0.006	-	0.007	0.008	0.006
Erinaceus	0.045	0.037	0.037	-	0.007	0.007
Erinaceinae cryptic	0.044	0.036	0.036	0.046	-	0.007
H.aur	0.032	0.027	0.027	0.036	0.033	-

4. Discussion

In general, the results obtained in this study provide an answer to the initial proposed objectives. However, some particular results are surprising, providing information never achieved before.

4.1. Species distribution

Outside Africa, the distribution of haplotypes follows the known distribution of species with the exception of one sample from Portugal. According to IUCN (2013) distribution maps, only *Erinaceus europaeus* is recorded from Portugal, but our results suggest the presence of a different species, similar to the one existent in North Africa. As it is a single sample it is possible that it is a case of laboratory mistake, such as an exchange of identification labels. The hypothesis of human introduction can not be ruled out, and more sampling in areas near the existing sample are needed to support or exclude these hypotheses. The haplotypes of *E. europaeus* were present in Iberian Peninsula, Sardinia (Italy) occurring within the known distribution of this species (Mitchell-Jones et al., 1999; Palomo and Gisbert, 2002). Samples of *E. concolor* from Armenia and Turkey as well as samples of *E. roumanicus* from Greece were concordant with the present known range of both species (IUCN, 2013). The Mediterranean islands, Malta, Canary and Balearics are known to exhibit introduced populations of *Atelerix algirus* (Palomo and Gisbert, 2002; Hutterer, 2005; Morales and Rofes, 2008) and our study confirm this presence. The Cyprus Island and Libya have haplotypes of *Hemiechinus auritus*, which is in accordance with the known distribution of this species (Niethammer, 1969; Spitzenberger, 1978; Boye, 1991).

Samples from Africa have a more complex distribution. Our nuclear results determined the presence of *A. algirus* in Morocco, northern Algeria and northern Tunisia, also in areas where the presence of *Paraechinus aethiopicus* was expected. An isolated sample of *A. algirus* was also found in Fezzan mountains of Libya, implying that the distribution of this species should be much wider than previously thought. It is unknown if this population is linked to the populations in coastal areas along the Mediterranean or if it is isolated, as a result of climate oscillations that induced range fragmentation. Some of the haplotypes in the Atlantic Sahara corridor and Adrar Atar-Kediet ej Jill mountains were considered, by nuclear results, as belonging to an unknown species named by us as Erinaceinae cryptic (further details discussed below). An isolated sample of this unknown species was also found in northwest Libya,

suggesting a relatively wider distribution. The presence of isolated populations in Sahara-Sahel mountains and its role as refugia have been subject of studies in other mammals (Busby et al., 2009; Brito et al., 2010). It is thought that these restricted habitats have been a shelter against the impacts of extreme climate oscillations of the Plio-Pleistocene when climate became increasingly arid and endemic taxa formed in isolation (Brito et al., 2013). Although many of biodiversity corridors across the Sahara desert have been fragmented and contributed to the isolation of populations, others continue to play an important role in species dispersal and facilitate gene flow, such as the Atlantic Sahara corridor (Brito et al., 2009, 2011). Assessing the role of refuges and corridors in the evolution and phylogeography of North African hedgehogs is not possible with the present limited number of samples. In the future, it is necessary to increase sampling effort in the putative refuges (e.g. mountains) and dispersal corridors and genetic analysis targeted to these specific populations. Samples collected in the distribution area of *Atelerix albiventris* are clearly separated in two different clades in the phylogenetic results and exhibit geographic structuring as well, suggesting the existence of cryptic diversity in that range. It is possible that the Niger River, as also seen in other river systems and species (Booth, 1958; Robbins, 1978; Nicolas et al., 2006), acted as geographic barrier and isolated populations on the western (Mali and Mauritania) from eastern (Niger) banks of the river.

4.2. Phylogenetic relationships

4.2.1. General overview: differences between mtDNA and nDNA

Phylogenetic analyses of mitochondrial DNA demonstrate the existence of four main groups of species that differ on how they separate from each other according to distinct phylogenetic methods and are not well supported (Figures 6B and 7). On the contrary, all the internal divisions (species level nodes) are highly supported, ranging from 90% to 100%, except in one case of 64% in ML tree for the mtDNA marker. The first group to emerge is constituted by one clade, corresponding to the Paraechinus genera. The second group is composed by two clades, Mesechinus and H.aur (genus *Hemiechinus*). In our study, *Paraechinus* assumes a basal position in relation to sister taxa *Mesechinus* and *Hemiechinus*. This result supports the studies of Morshed and Patton (2002) and Bannikova et al. (2002) that consider *Hemiechinus* and *Paraechinus* as distinct genera. Our results also corroborate the molecular results of He et al. (2012), which used three different mitochondrial genes (12S rRNA, Cyt-b and ND2) and also found novel relationships between *Hemiechinus*, *Mesechinus* and

Paraechinus. The network results also illustrate these relationships between the three clades, suggesting that the *Paraechinus* is the most distant one. The third group is composed by two clades of Eurasian *Erinaceus* species, E.con/E.rou. and E.eur/ E.amu. Both clades have subdivisions corresponding to the different species, but inside clade E.eur/E.amu it is observed a second split in *E. europaeus* species between the haplotypes from Iberian Peninsula and those from Sweden and Italy (BPP=1), suggesting some differentiation between them. Haplotype network (Figure 8) of the same dataset also evidences separation of these two populations of *E. europaeus*. The 48 mutational steps are almost the same number of steps (54) that separate *E. concolor* and *E. roumanicus* species. A similar separation was verified by Santucci et al. (1998) between samples from Italy/Germany and samples from Spain/France/UK and dated to around 2.7Ma, when an intensification of glaciation had happened and populations were forced to find refugia in the southern European peninsulas. It is possible that the interglacial warm periods allowed the expansion of hedgehogs from these refugia areas but potential barriers, like the Pyrenees and Alps, prevented wider distribution of previously isolated clades and promoted differentiation within *E. europaeus* (Hewitt, 1993; Santucci et al., 1998). The last group, mainly composed of African haplotypes, contains species that diverged more recently than others belonging to the above Eurasian groups. Nevertheless, it is not possible to determine which of the three clades diverged first inside this group, given the low support values of both Bayesian and ML trees (Figures 6B and 7). Genetic distances corroborate the relationships between these groups since the highest distances are found between groups, ranging from 20.8% to 23.4% between the basal group (*Paraechinus*, *Mesechinus* and H.aur) and *Erinaceus* species groups and ranging from 23.7% to 26.1% between *Erinaceus* species groups and African group. As expected lower values were found inside each group, whereas the group with the smallest genetic distances between its clades is the African one (ranging from 11.2% to 13.5%).

Nuclear results suggest a slightly different topology (Figure 9), despite the number of samples analyzed being significantly lower than mtDNA samples and no *Paraechinus* and *Mesechinus* haplotypes were included in the analyses. Within the nuclear marker, one of the African mitochondrial clades is splited into two. The nDNA trees highlight the oldest and first split (BPP=1) of *Hemiechinus* haplotypes, together with one population from Africa (BPP=0.99) (Erinaceinae cryptic clade: discussed below), placing them in a basal position in relation to the rest of European and African clades. Another stongly supported group (BPP= 1) is constituted by three species of *Erinaceus* but in nuclear marker, the relationship between species is not well resolved. The last group is constituted by three mostly African clades and contrary to what is

suggested by mtDNA results, the three clades separated simultaneously (BPP= 0.99; ML and NJ support values= 86/84).

4.2.2. European and African hedgehogs

Butler (1988) suggested that hedgehogs entered Europe from Asia and that during early Pliocene cold periods, they have diverged into different species (Santucci et al., 1998). This pattern is concordant with the results described above, which include *Erinaceus* species from European countries and China, and exhibit smaller genetic distances between European and Asian clades than in between European and African clades.

The way Africa was invaded by hedgehogs is still a mystery. The studies of Butler and Greenwood (1973) indicate that the genus *Atelerix* may be a palaeo-Mediterranean endemism, resulted from pre-Messinian invasion of Africa by a northern hedgehog (*Erinaceus europaeus*). Our mitochondrial results show proximity between *E. europaeus* from Italy/Sweden and African samples from Morocco/ Senegal, while nuclear results assume that African samples are more closely related with samples from the Iberian Peninsula. Although it is not resolved in our study the relationship between Europe and Africa and which gene reflects better the history of hedgehogs, results of both molecular markers exhibit high values of genetic distances, which lead us to think in an ancient divergence in both continents. Despite the fact that *A. algirus* samples from Mediterranean islands have proven of African origin, most likely via introductions, and the high genetic values obtained between Africa and Europe, it is not possible to determine if the Gibraltar Strait is acting as barrier. As previously stated, one sample of *A. algirus,* sharing the same haplotype of African samples, was found in Portugal. As we do not know if this happened due to laboratory errors, no conclusion can be draw. To better understand the effects of the Gibraltar Strait, more samples should be collected on the eastern coast of Iberian Peninsula (where the species is also present) and in Alentejo (Portugal), in the area where the supposedly sample of *A. algirus* was found.

An isolated sample of *Hemiechinus auritus* was found in Africa, but its colonization is much older than the alleged colonization of Atelerix from Europe. Given the fact that we only found one sample of this species, it is not possible to identify colonization routes, but following its distribution range it is possible that its introduction has been given from further eastern areas.

4.2.3. African species relationships: distribution and morphologic and genetical variations

African samples showed the most surprising results and the fact that no other studies were carried out on North African species makes the results obtained in this thesis unique. As previously mentioned, the differences between mitochondrial and nuclear results contributed to a greater understanding of many processes that affected or still affect hedgehogs in North Africa.

Mitochondrial analyses define three different African clades, referred as Erinaceinae, A.alb1 and A.alb2 while nuclear analyses highlight the existence of four, namely A.alg, A.alb1, A.alb2 and Erinaceinae cryptic. The main difference was detected in the Erinaceinae clade of mtDNA analyses that, in nDNA results appears divided in two different clades (A.alg and Erinaceinae cryptic). Samples of these clades are distributed along the northwestern countries, including Libya, Tunisia, Algeria, Morocco, Mauritania and Senegal, following a more coastal distribution along the Atlantic Sahara corridor. In nuclear analyses samples belonging to northern countries were identified as *Atelerix algirus* due to their morphological appearance and geographical distribution. In terms of morphology, Aulagnier et al. (2009) described them as small, pale, with legs slightly large and ears protruding from the body. Aulagnier et al. (2009) also emphasized a dark form of *A. algirus*, which can be confused with individuals of *Paraechinus aethiopicus* species, the size of their ears and color of the spines (that are bigger and darker in the latter species) are not considered. Although we have a dataset with photographs of almost every collected sample, it is not possible to make a determinant diagnosis to the species level because of the poor condition of samples (mostly run-over sample). As an example, it was observed morphological variability between the white (codes 6866 and 7114) and dark forms (codes 339 and 905) of *A. algirus* (Figure 11). In the areas proposed for the distribution of *P. aethiopicus* species, at both morphological and genetic level we only found the presence of *A. algirus*. The southern clade was named Erinaceinae cryptic because mitochondrial results demonstrate sharing of mitochondrial genome with the *A. algirus*, but their nuclear genome showed substantial differentiation. This clade has the peculiarity of having their haplotypes scattered in areas for which there is a lack of knowledge in species occurrence (Figure 9A). The morphological analysis (codes 323 and 6066 in Figure 11) showed that those dark individuals had large ears. These characteristics may have been confusing them with the *P. aethiopicus* species, but genetically their identity can not be resolved. As in the putative area of occurrence of *P.*

aethiopicus only *A. algirus* were found, it is plausible that the distribution of *P. aethiopicus* proposed by IUCN (2013) and Aulagnier et al. (2009) is in fact occupied by the dark form of *A. algirus,* or by the Erinaceinae cryptic group. With the currently available data it is not possible to solve this problem. Further sampling from the alleged distribution of *P. aethiopicus* and extended genetic analyses with multiple independent markers are needed. Morphological analyses based on specimens in good conditions are also needed.

The nuclear genetic structuring between *A. algirus* and Erinaceinae cryptic haplotypes showed 30 mutational steps separating them, a greater number than those verified between different genera. Genetic distances found in the nuclear dataset highlighted their difference, with 4.4%, almost the same percentage obtained between *A. algirus* and *Erinaceus* (4.5%) and between Erinaceinae cryptic and *Erinaceus* (4.6%). Due to the reduced number of haplotypes found in Erinaceinae cryptic group it was not possible to measure statistical credibility of Tajima's D and Fu's Fs tests. The significantly negative values of D and Fs found in clade A.alg suggest an excess of rare alleles, due to selection on thin marker or recent population expansion and consequent changes in population size. The results obtained, besides suggesting a huge divergence between *A. algirus* and Erinaceinae cryptic groups, lead one to believe that they may belong to different genera.

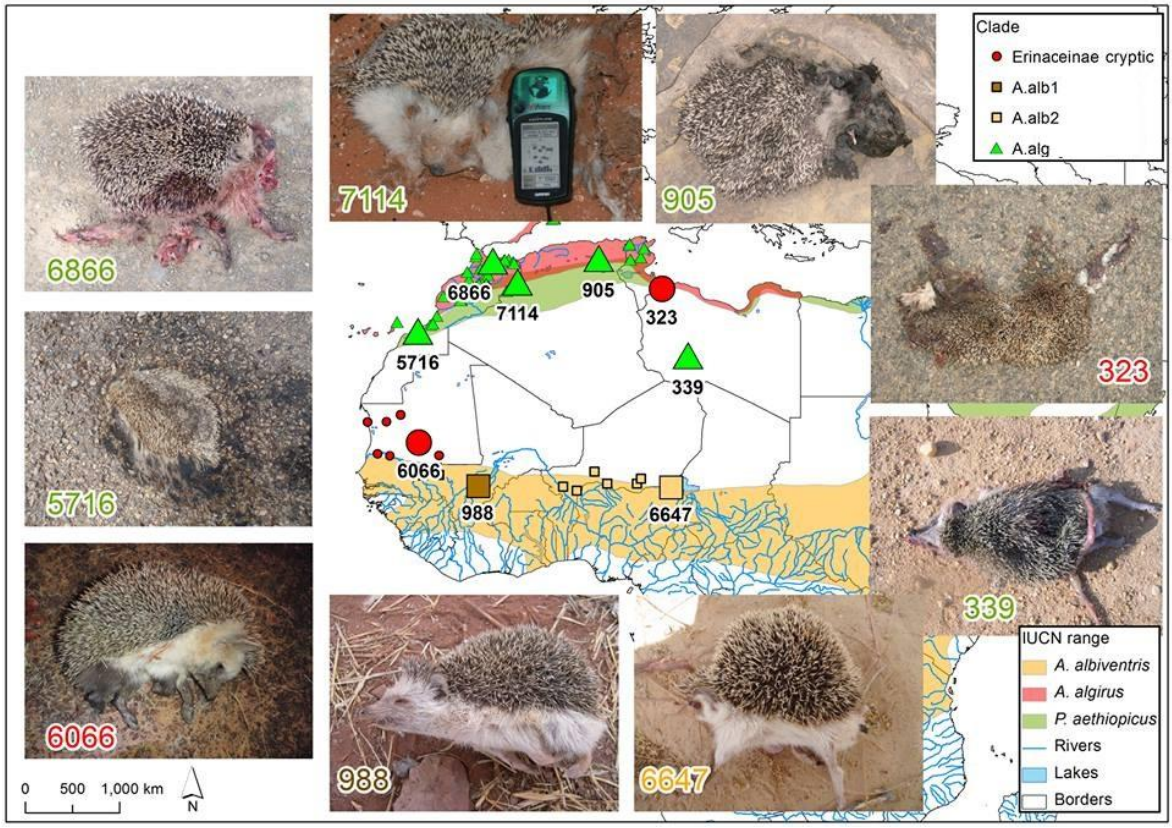


Figure 11 – Images of several sampled individuals from different species and the corresponding geographic distribution of species according to IUCN (2013). Colors of the sample number correspond to those presented in the map symbols.

Both molecular markers suggest the existence of two distinct populations in the area correspondent to the *A. albiventris* species (Le Berre, 1990; IUCN, 2013). Divergent groups were detected on both sides of the Niger River: the clade A.alb1on the western side of the river (covering Mauritania and Mali) and the clade A.alb2 on the eastern side (covering Niger). A preliminary assessment of morphologic trait of each group shows no differences between them. It is possible that both groups represent cryptic diversity, since they have diverged significantly at the genetic level but have remained morphologically similar. Although very few studies have been performed to understand the putative barrier effect of the Niger River, most of them suggest it as a dispersal barrier that appeared in a previously continuous habitat (Petter, 1974; Brouat et al., 2009). The results obtained with the hedgehogs seem to agree with the idea of the Niger River being a barrier to dispersal. The structuring and genetic results also show a consistent separation of A.alb1 and A.alb2. The mitochondrial network shows 94 mutational steps between A.alb1 and A.alb2, a value comparable to the differences between *Erinaceus* species or even to differences between genera, as in the case of *Mesechinus* - H.aur (89 mutational steps) and H.aur – *Paraechinus* (116 mutational steps) clades. Similarly, the genetic divergence of 11.2% between A.alb1 and A.alb2

was higher than the one that separates H.aur from *Mesechinus* clades (10.9%). This value is higher than the usual genetic distance (8.1%) between sister species in mammals (Bradley and Baker, 2001) for the cytochrome b marker. Although the nuclear marker did not depict such high difference, both molecular primers are concordant in reconstructed phylogenies. Also both markers depicted high levels of haplotype diversity of A.alb2, whereas the negative values of D and Fs for A.alb1 (not tested for nDNA) and A.alb2, suggests population's expansions. Those suggest these groups as being two different species, a hypothesis that needs verification through the use of additional samples and molecular markers.

The general relationship between the different African groups varies by marker analysis. In mtDNA analysis Erinaceinae large group is shown closer to A.alb1, being linked in the network by 103 mutational steps (12.2% of genetic distance). While in the nDNA analysis A.alg is equally separated from A.alb1 and A.alb2 by six mutational steps in network and 1.4% of genetic distance.

4.2.4. Signals of Introgression

Analyses of North Africa samples detected discrepancies between mitochondrial and nuclear DNA markers, disputing two different histories of the African samples relationships. The most important difference is that Erinaceinae clade in mitochondrial analyses splits into two distinct clades (A.alg and Erinaceinae cryptic) in the nuclear marker.

A possible speculation about the lack of agreement between mitochondrial and nuclear markers could be an introgression of mitochondrial genome of A.alg into the Erinaceinae cryptic samples, or incomplete lineage sorting. Curiously, the transfer of mtDNA between species has been frequently observed among animals (Sequeira et al., 2011; Bastos-Silveira et al., 2012; Melo-Ferreira et al., 2012). When the territory of a resident species is invaded by another, it is possible that rare events of hybridization can lead to a introgression from the resident species into the expanding one (Sequeira et al., 2011). It can happen through purely stochastic but can be also influenced by adaptive processes. The fact that samples corresponding to these two clades occupy the two extremes of the Atlantic Sahel corridor and one sample of Erinaceinae cryptic (known from Mauritania and Senegal) was found in southern Libya, it is possible that Erinaceinae cryptic samples have been spread beyond the areas that were identified in this study and that at some point had come into contact with mitochondrial genome of *A. algirus* populations. Further sampling from possible contact zones along southern Morocco and northern Algeria are needed to address this hypothesis.

4.3. Islands colonization

The Algerian hedgehog *Atelerix algirus* is the only known North African genus occupying Mediterranean islands, as is the case of Malta, Balearic and Canary Islands archipelagos (Corbet, 1988; Blanco, 1998; Mitchell-Jones et al., 1999; Hutterer, 2005). Our results indicate a relative closeness between African haplotypes and haplotypes from these islands. The mitochondrial network (Figure 8) shows that haplotypes from Canary (Ca1 and Ca2) are related to Mauritania and Morocco once. This suggests the first colonization by A. algirus of Canary islands by specimens from the coastal regions of Morocco and Mauritania. It was hypothesized that this species was introduced to islands just before the 20th century (Nogales et al., 2006), therefore they did not have enough time to diverge and maintain genetic characteristics similar to those found in the African coast. Balearic haplotypes showed higher separation from African haplotypes then the Canary Islands haplotypes. They are five mutational steps away from a haplotype (Va1) which share samples from Mauritania, Tunisia and Morocco and four mutational steps away from one haplotype from Algeria (Al1), suggesting that they may have had a more ancestral colonization. Morales and Rofes (2008) dated the earliest presence of *A. algirus* outside Africa in the 13th century, coinciding with the Almohads (Berber dynasty from western Atlas Mountains) settlement over Iberian lands and Balearic Islands. Its dispersion by human action may have happened from places close to Tunisia (Alcover et al., 1981) but our results also suggest colonization of Baleares from Algeria, Morocco or Mauritania. In Malta it is known the presence of *A. algirus* (Hutterer, 2005) but the time of introduction and the provenance of the colonizing samples are still unclear. Our study shows a slight differentiation between the Malta haplotype and samples from Morocco (three mutational steps) but does not show more similarities to Tunisian or Algerian haplotypes, as it would be expected in term of geographical proximity. No statement can be taken for now and efforts are needed to sample these areas as well as to produce longer sequences, allowing to estimate time of separation and test which mainland samples are closest.

The origin of the long-eared hedgehog *Hemiechinus auritus* on Cyprus is still an unresolved issue. No remains have been found between Pleistocene and prehistoric sites on this island (Davis, 1987; Held, 1989) and the hypothesis of a natural immigration is not considered. As this species does not occur in Turkey it is thought that humans took specimens from the Syrian or Palestine coast to Cyprus (Boye, 1991). Our results do not discard the hypothesis of an African origin, given that the haplotype of *H. auritus* in our study from Libya differs only with eight mutational steps from Cyprus haplotypes. To actually evaluate the proximity of Cyprus and African

hedgehogs, there should be a more detailed sampling in the northeast African coast, from Libya to Egypt and in western Asia countries, to be able to assess more genetic and geographic patterns.

4.4. Systematic inferences

The current systematic is unclear and does not reflect the genetic variability found in Europe and African samples:

1) The subdivision of *Erinaceus europaeus* found in mtDNA phylogenetic trees between populations from Iberian Peninsula and populations from Italy/Sweden indicate significant differentiation. Furthermore, also the mtDNA network emphasized this difference with 48 mutational steps between populations (almost the same (54) that separates *E. roumanicus* and *E. concolor*). This occurrence may be justified by potential refugia differentiation during glacial periods, but a review in morphological and genetic characters of both populations should be done.

2) Findings from Africa shall be considered as novel and unique as molecular studies have never been realized in the studied areas on hedgehogs.

The discovery of two highly distinct populations in the recognized area of *Atelerix albiventris* distribution suggests the existence of a possible new species. Morphological characters need to be reexamined and more sampling has to be done in order to test this hypothesis.

The presence of *Atelerix algirus* samples in areas that was supposed to be occupied by *Paraechinus aethiopicus* leads us to think that this species might not occur in North Africa or its occurrence is very limited. Its distribution may have been suggested closely to the *A. algirus* distribution due to the similarities between the dark form of *A. algirus* and *P. aethiopicus*. Other hypothesis is that the Erinaceinae cryptic group that we found in Mauritania and Senegal can be confounded with the *P. aethiopicus* (sample coded as 6066 in Figure 11 has similarities with the described *P. aethiopicus*) that should appear in North Africa, but as we do not know the distribution limits of this group no consideration can be made without further studies.

4.5. Contributions and future research

The greatest contribute of this thesis is related with the combination of two different types of molecular markers to conduct for the first time phylogeographic and population genetic analysis in North African hedgehogs. This study has produced 136 new mitochondrial and 78 new nuclear sequences, as well as two new designed pairs of primers. In addition to the sequences obtained for *E. europaeus* that already existed for Cyt-b and ß-fibr, it was sequenced for the first time samples of E. concolor for Cyt-b and sequences of E.roumanicus for Cyt-b and ß-fibr. It was also sequenced for the first time samples of *A. albiventris* for Cyt-b and ß-fibr genes and the first ever of the species *A. algirus*.

This study was conducted with samples that were occasionally collected from deceased animals along roads, during different field missions. Nevertheless many interesting and potentially important observations have been recorded that with future sampling efforts should be more focused on African species and on areas of particular interest, such as likely contact zones, mountain areas of the Sahara-Sahel, south of Niger River and along the Atlantic Sahara corridor.

Further molecular studies will help to better explain the divisions found between African and non-African specimens. In the future the size of the sequences to be amplified must be also taken into account. An interesting aspect would be to combine genetic variability with ecogeographical information, using Geographical Information System (GIS) and ecological niche-based models, following a landscape genetics approach.

5. Conclusion

Biodiversity of North Africa is still poorly known, and until now no study has focused in the phylogeny of North African hedgehogs. It is the first time that genetic relationships within Africa are assessed.

Mitochondrial data demonstrated the existence of close relationships between genera *Paraechinus*, *Mesechinus* and *Hemiechinus*. A great differentiation between species of *Erinaceus* genus was observed, and a remarkable case is seen within *E. europaeus*, between populations of the European Peninsulas. Pleistocene climate fluctuations and geographic barriers appear to be strong reasons for these differences.

African data produced divergences between mitochondrial and nuclear genes that brought new insights and raised new questions. A*telerix algirus* species is identified in new and unexpected ranges, such as distribution areas where was supposed to exist *Paraechinus aethiopicus*. No specimen of the latter species was found in North Africa. The Erinaceinae cryptic clade, for which the real identity was not established, seems to be a valid taxonomic unit according to the nuclear marker. However, mtDNA data groups this clade with *A. algirus*, suggesting mitochondrial introgression between both species. Haplotypes from west and eastern parts of the Niger River corresponding to *A. albiventris* distribution range appear genetically differentiated from each other. It is possible that the river act as barrier and populations experience cryptic divergence. There are great evidences of genetic diversity and population differentiation within North Africa.

The species *Atelerix algirus* from Canary Islands are related to haplotypes from Mauritania and Morocco, whereas samples from Balearic Islands, are closer to Tunisia, Algeria, Morocco and Mauritania haplotypes, showing a trend of geographical proximity in their colonization. Malta and Cyprus islands do not show coherent colonization origin.

Main results of this study should be considered as exploratory and preliminary, although crucial on its extent for planning future hypothesis testing approach. More thorough review of species systematic needs to be done, in order to increase the knowledge that was hitherto vague.

6. References

Alcover JA, Moyà -Solà S and Pons-Moyà J (1981). Les Quimeres del Passat. Els Vertebrats Fòssils del Plio-Quaternari de les Balears i Pitiuses. Palma de Mallorca: Moll.

Allaby M and Garratt R (2008) Deserts. Revised Edition. Facts on File, Inc.

Amori G, Hutterer R, Krystufek B, Yigit N, Mitsain G and Palomo LJ (2008) *Erinaceus europaeus*. IUCN Red List of Threatened Species. Version 2013.1.

Anlas S (2012) New and Additional Records of Paederinae (Coleoptera: Staphylinidae) from the Island of Cyprus. *Journal of the Entomological Research Society*, 14:39-43.

Arntzen JW and García-París M (1997) Phylogeny and biogeography of midwife toads (*Alytes*, Discoglossidae): a rebuttal. *Contributions to Zoology*, 66:263–268.

Aulagnier S, Haffner P, Mitchell-Jones AJ, Moutou F and Zima J (2009) Guía de los mamíferos de Europa, del norte de África y de Oriente Medio. Lynx Edicions.

Avise, JC (2000) Phylogeography: The history and formation of species. USA: Harvard University Press.

Avise, JC (2006) Evolutionary pathways in nature. A phylogenetic approach. Cambridge University Press, Cambridge, UK.

Ayyad MA (2003) Case studies in the conservation of biodiversity: degradation and threats. *Journal of Arid Environments*, 54:165-182.

Bannikova AA, Matveev VA and Kramerov DA (2002) Using inter-SINE-PCR to study mammalian phylogeny. *Russian Journal of Genetics*, 38:714-724.

Bajpai N and Tewari RR (2010) Mitochondrial DNA sequence-based phylogenetic relationship among flesh flies of the genus Sarcophaga (Sarcophagidae: Diptera). *Journal of Genetics*, 89:51-54.

Ballard JW and Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, 13:729-744.

Barata M, Perera A, Martínez-Freiría F and 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 Mda L, Heckel G (2012) Deep mitochondrial introgression and hybridization among ecologically divergent vole species. *Molecular Ecology*, 21:5309-5323.

Becher SA and Griffiths R (1998) Maximum Likelihood Analysis of the Complete Mitochondrial Genomes of Eutherians and a Reevaluation of the Phylogeny of Bats and Insectivores. *Molecular Ecology*, 7:1599-1604.

Beebee TJC and Rowe G (2008) An introduction to molecular ecology. Oxford University Press, New York.

Beheregaray LB (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, 17:3754-3774.

Bellard C, Bertelsmeier C, Leadley P, Thuiller W and Courchamp F (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters*, 15:365-377.

Benda P, Hanák V, Horácek I, Hulva P, Lucan R and Ruedi M (2007) Bats (Mammalia: Chiroptera) of the Eastern Mediterranean. Part 5. Bat fauna of Cyprus: review of records with confirmation of six species new for the island and description of a new subspecies. Acta Societatis Zoologicae Bohemicae, 71:71-130.

Bermingham E and Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology*, 7:367-369.

Blanco JC (1998). Mamíferos de Espanã, Vol. I. Barcelona: Planeta.

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

Booth AH (1958) The zoogeography of West African primates. A review. Bull. Inst. Fond. Afr. Noire, 20:587–622.

Boye P (1991) Notes on the morphology, ecology and geographic origin of the Cyprus Long-eared hedgehog. *Bonn zoological Bulletin*, 42:115-123.

Bradley RD and Baker RJ (2001) A test of the genetic species concept: Cytochrome-b sequences and mammals. *Journal of Mammalogy,* 82:960-973.

Brinkman FS and Leipe DD (2001) Phylogenetic Analysis. In: Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. John Wiley & Sons, Inc., 323-358.

Brito JC, Acosta AL, Álvares F and Cuzin F (2009) Biogeography and conservation of taxa from remote regions: An application of ecological-niche based models and GIS to North-African Canids. *Biological Conservation*, 124:3020-3029.

Brito JC, Álvares F, Martínez-Freiría F, Sierra P, Sillero N and Tarroso P (2010) Data on the distribution of mammals from Mauritania, West Africa. *Mammalia*, 74:449-455.

Brito JC, Fahd S, Geniez P, Martínez-Freiría F, Pleguezuelos JM and Trape JF (2011). Biogeography and conservation of viperids from North-West Africa: an application of ecological niche-based models and GIS. *Journal of Arid Environments*, 75:1029–1037.

Brito JC, Godinho R, Mártinez-Freiría F, Pleguezuelos JM, Rebelo H, Santos X, Vale CG, Velo-Antón G, Boratyński Z, Carvalho SB, Ferreira S, Gonçalves DV, Silva TL, Tarroso P, Campos JC, Leite JV, Nogueira J, Alvares F, Sillero N, Sow AS, Fahd S, Crochet PA and Carranza S (2013) Unravelling biodiversity, evolution and threats to conservation in the Sahara-Sahel. *Biological Reviews*, doi: 10.1111/brv.12049.

Brito JC, Fahd S, Geniez P, Martínez-Freiría F, Pleguezuelos JM and Trape JF (2011) Biogeography and conservation of viperids from North-West Africa: An application of ecological niche-based models and GIS. *Journal of Arid Environments*, 75- 1029-1037.

Brito JC, Martinez-Freiría F, Sierra P, Sillero N and Tarroso P (2011) Crocodiles in the Sahara Desert: an update of distribution, habitats and population status for conservation planning in Mauritania. *PLoS ONE*, 6:e14734.

Brouat C, Tarard C, Bâ K, Cosson JF, Dobigny G, Fichet-Calvet E, Granjon L, Lecompte E, Loiseau A, Mouline K, Piry S and Duplantier JM (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, Terrasa B, Pérez-Mellado V, Castro JA, Hoskisson PA, Picornell A and Ramon MM (2008) Bayesian estimation of post-Messinian divergence times in Balearic Island lizards. *Molecular Phylogenetics and Evolution*, 48:350-358.

Busby GBJ, Gottelli D, Wacher T, Marker L, Belbachir F, De Smet K, Belbachir-Bazi A, Fellous A, Belghoul M and Durant SM (2009) Genetic analysis of scat reveals leopard *Panthera pardus* and cheetah *Acinonyx jubatus* in southern Algeria. *Oryx*, 43:412-415.

Butler PM (1988) Phylogeny of the insectivores. In: The Phylogeny and Classification of the Tetrapods, Vol. 2, Mammals (ed. Benton MJ). Clarendon Press, London. pp. 117-141.

Butler PM and Greenwood M (1973) The Early Pleistocene hedgehog from Olduvai, Tanzania. In Fossil vertebrates from Africa. Leakey, L.S.B., Savage, R.J.G. & Coryndon, S.C. (Eds). London*: Academic Press*, 3:7–42.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, Kinzig AP, Daily GC, Loreau M, Grace JB, Larigauderie A, Srivastava DS and Naeem S (2012) Biodiversity loss and its impact on humanity. *Nature*, 586:59-67.

Carracedo JC (1993) The Canary Islands: an example of structural of structural control on the growth of large oceanic-islands volcanoes. *Journal of Volcanology and Geothermal Research*, 60:225-241.

Carranza S, Arnold EN and Pleguezuelos JM (2006) Phylogeny, biogeography, and evolution of two Mediterranean snakes, Malpolon monspessulanus and Hemorrhois hippocrepis (Squamata, Colubridae), using mtDNA sequences. *Molecular Phylogenetics and Evolution*, 40:532-546.

Catania KC (2005) Evolution of sensory specializations in insectivores. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.*, 287:1038-1050.

Clement M, Posada D and Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, 9:1657-1660.

Çolak E, Yigit N, Sozen M and Ozkurt S (1998) A Study on the Long-eared hedgehog, Hemiechinus auritus (Gmelin, 1770) (Mammalia: Insectivora) in Turkey. Turkey *Journal of Zoology*, 22:131-136.

Colangelo P, Verheyen E, Leirs H, Tatard C, Denys C, Dobigny G, Duplantier JM, Brouat C, Granjon L and Lecompte E (2013) A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. *Biological Journal of the Linnean Society*, 108:901-916.

Colom G (1978) Biogeografía de las Baleares. La formación de las islas y el origen de su flora y de su fauna (ed. by Diputación Provincial de Baleares), Instituto de Estudios Baleáricos, Consejo Superior de Investigaciones Científicas, pp. 515.

Corbet GB (1988) The family Erinaceidae: a synthesis of its taxonomy, phylogeny, ecology and zoogeography. *Mammal Review*, 18:117-172.

Cosson JF, Hutterer R, Libois R, Sarà M, Taberlet P and Vogel P (2005) Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula* (Mammalia: Soricidae). *Molecular Ecology*, 14:1151-1162.

Cowling RM and Wilhelm-Rechmann (2007) Four perspectives on conservation in Africa. *FFI-Oryx*, 41:135-139.

Cuttelod A, García N, Abdul Malak D, Temple H and Katariya V (2008) The Mediterranean: a biodiversity hotspot under threat. In: Vié J-C, Hilton-Taylor C, Stuart SN, editors. The 2008 review of the IUCN red list of threatened species: IUCN Gland, Switzerland.

D'Amen M, Zimmermann NE and Pearman PB (2013) Conservation of phylogeographic lineages under climate change. *Global Ecology and Biogeography*, 22:93-104.

Davis SJ (1987) The archaeology of animals. B.T. Batsford Ltd., London. pp. 224.

Dawson AG (1996) Ice age earth: late quaternary geology and climate. New York, NY: Routledge.

Dawson TP, Stephen TJ, House JI, Prentice IC and Mace GM (2011) Beyond Predictions: Biodiversity Conservation in a Changing Climate. *Science*, 332:53-58.

Dernegi D (2010) Ecosystem Profile – Mediterranean Basin Biodiversity Hotspot. Critical Ecosystem Partnership Fund.

Devictor V, Whittaker RJ and Beltrame C (2010) Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions*, 16:354-362.

Dietz S and Adger WN (2003) Economic growth, biodiversity loss and conservation effort. *Journal of Environmental Management*, 68:23-35.

Dixon JD, Oli MK, Wooten MC, Eason TH, McCown JW and Cunningham MW (2007) Genetic consequences of habitat fragmentation and loss: the case of the Florida black bear (*Ursus americanus floridanus*). *Conservation Genetics*, 8:455-464.

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

Doncaster CP, Rondinini C and Johnson PCD (2001) Field test for environmental correlates of dispersal in hedgehogs *Erinaceus europaeus. Journal of Animal Ecology*, 70:33-46.

Douady CJ, Catzeflis F, Raman J, Springer MS and Stanhope MJ (2003) The Sahara as a vicariant agent, and the role of Miocene climate events, in the diversification of the mammalian order Macroscelidea (elephant shrews). *PNAS*, 100:8325-8330.

Douady CJ, Chatelier PI, Madsen O, de Jong WW, Catzeflis F, Springer MS and Stanhope MJ (2002) Molecular phylogenetic evidence confirming the Eulipotyphla concept and in support of hedgehogs as the sister group to shrews. *Molecular Phylogenetics and Evolution*, 25:200-209.

Doudady CJ and Douzery EJP (2009) Hedgehogs, shrews, moles and solenodons (Eulipotyphla). In: The Timetree of Life, 157:1011-1018.

Dubey S, Zaitsev M, Cosson JF, Abdukadier A, Vogel Peter (2006) Pliocene and Pleistocene diversiWcation and multiple refugia in a Eurasian shrew (*Crocidura suaveolens* group). *Molecular Phylogenetics and Evolution*, 38:635-647.

Duggen S, Hoernle K, van den Bogaard P, Rupke L and Morgan JP (2003) Deep roots of the Messinian salinity crisis. *Nature* 422: 602–606.

Durant SM, Pettorelli N, Bashir S, Woodroffe R, Wacher T, De Ornellas P, Ransom C, Abáigar T, Abdelgadir M, El Alqamy H, Beddiaf M, Belbachir F, Belbachir-Bazi A, Berbash AA, Beudels-Jamar R, Boitani L, Breitenmoser C, Cano M, Chardonnet P, Collen B, Cornforth WA, Cuzin F, Gerngross P, Haddane B, Hadjeloum M,Jacobson A, Jebali A, Lamarque F, Mallon D, Minkowski K, Monfort S, Ndoassal B, Newby J, Ngakoutou BE, Niagate B, Purchase G, Samaïla S, Samna AK,Sillero-Zubiri C, Soultan AE, Stanley Price MR, Baillie JE. (2012) Forgotten biodiversity in desert ecosystems. *Science*, 336:1379-1380.

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

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Science, University of Washington, Seattle.

Fischer J and Lindenmayer DB (2007) Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, 16:265-280.

Fonseca MM, Brito JC, Paulo OS, Carretero MA and Harris DJ (2009) Systematic and phylogeographical assessment of the Acanthodactylus erythrurus group (Reptilia: Lacertidae) based on phylogenetic analyses of mitochondrial and nuclear DNA. *Molecular Phylogenetics and Evolution*, 51:131-142.

Fonseca MM, Brito JC, Rebelo H, Kalboussi M, Larbes S, Carretero MA and Harris DJ (2008) Genetic variation among spiny-footed lizards in the *Acanthodactylus pardalis* group from North Africa. *African Zoology*, 43:8-15.

Foley JA, Coe MT, Scheffer M and Wang G (2003) Regime shifts in the Sahara and Sahel: interactions between ecological and climatic systems in Northern Africa. *Ecosystems*, 6:524-539.

Franck P, Garnery L, Loiseau A, Oldroyd BP, Hepburn HR, Solignac M and Cornuet JM (2001) Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. *Heredity* 86:420-430.

Franklin J (2009) Mapping species distributions. Spatial inference and prediction. Cambridge University Press.

Freeland J (2005) Molecular Ecology. England: John Wiley & Sons.

Frost DR, Wozencraft WC and Hoffmann RS (1991) Phylogenetic Relationships of Hedgehogs and Gymnures (Mammalia: Insectivora: Erinaceidae). Smithsonian Contributions to Zoology, Nº 518.

Galtier N, Nabholz B, Glémin S and Hurst GD (2009) Mitochindrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, 18:4541-4550.

Gaubert P, Machordom A, Morales A, López-Bao JV, Veron G, Amin M, Barros T, Basuony M, Djagoun CAMS, San EDL, Fonseca C, Geffen E, Ozkurt SO, Couloux A and Palomares F (2010) Comparative phylogeography of two African carnivorans presumably introduced into Europe: disentangling natural versus human-mediated dispersal across the Strait of Gibraltar. *Journal of Biogeography*, 38:341-358.

Gillespie RG and Roderick GK (2002) ARTHROPODS ON ISLANDS: Colonization, Speciation, and Conservation. *Annual Review of Entomology*, 47:595-632.

Gonçalves DV, Brito JC, Crochet PA, Geniez P, Padial JM and Harris DJ (2012) Phylogeny of North African Agama lizards (Reptilia: Agamidae) and the role of the Sahara desert in vertebrate speciation. *Molecular Phylogenetics and Evolution*, 64:582-591.

Gould GC (1995) Hedgehog Phylogeny (Mammalia, Erinaceidae) – the Reciprocal Illumination of the Quick and the Dead. *American Museum Novitates*, 3131:1-45.

Gould GC (2001) The Phylogenetic Resolving Power of Discrete Dental Morphology Among Extant Hedgehogs and the Implications for Their Fossil Record. *American Museum Novitates*, 3340:1-52.

Grenyer R and Purvis A (2003) A composite species-level phylogeny of the 'Insectivora' (Mammalia: Order Lipotyphla Haeckel, 1866). *Journal of Zoology*, 260:245-257.

Guillaumet A, Crochet PA and Pons JM (2008) Climate-driven diversification in two widespread Galerida larks. *BMC Evolutionary Biology*, 8:32.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W and Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59:307-321.

Habel JC, Lens L, Rodder D and Schmitt T (2011) From Africa to Europe and back: refugia and range shifts cause high genetic differentiation in the Marbled White butterfly *Melanargia galathea*. *BMC Evolutionary Biology*, 11:215-228.

Hadjisterkotis E and Masala B (1995). Vertebrate extinction in Mediterranean islets: an example from Cyprus. *Biogeographia*, 18: 691-699.

Hadjisterkotis E, Masala B and Reese D (2000). The origin and extinction of the large endemic Pleistocene mammals of Cyprus. *Biogeographia Anatolia,* XXI: 593-606.

Hamilton WJ (1999) Potential threat of hedgehogs to invertebrates with a restricted range, Otago region. *Conservation Advisory Science Notes No. 254*, Department of Conservation, Wellington.

Hardy SM, Carr CM, Hardman M, Steinke D, Corstorphine E and Mah C (2010) Biodiversity and phylogeography of Arctic marine fauna: insights from molecular tools. *Marine Biodiversity*, 41:195-210.

He K, Chen JH, Gould GC, Yamaguchi N, Ai HS, Wang YX, Zhang YP and Jiang XL (2012) An estimation of Erinaceidae phylogeny: a combined analysis approach. *PLoS ONE*, 7:e39304.

Hedges SB and Maxson LR (1997) Complementary Uses of Molecules and Morphology: A Reply to Lee. *Molecular Phylogenetics and Evolution*, 8:445.

Held SO (1989) Colonization cycles on Cyprus 1: The Biogeographic and paleontological foundations of early prehistoric settlement. Re. Dep. Antiquities Cyprus, 1989:7-28.

Hewitt GM (1993) Postglacial distribution and species substructure: lessons from pollen, insects and hybrid zones. In: Evolutionary Patterns and Processes (eds Leeds DR, Edwards D). Linnean Society Symposium Series 14. Academic Press, London. pp.97-123.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68:87-112.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Molecular Ecology*, 10:537-549.

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF and Yoder AD (2010) Phylogeography's past, present, and future: 10 years after. *Molecular Phylogenetics and Evolution*, 54:291-301.

Holmes J (2008) How the Sahara became dry. *Science*, 320:752-753.

Huelsenbeck JP and Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, 17:754-755.

Hughes PD, Woodward JC and Gibbard PL (2007) Middle Pleistocene cold stage climates in the Mediterranean: New evidence from the glacial record. *Earth and Planetary Letters*, 253:50-56.

Hurst GD and Juggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceddings of the Royal Society B*, 272:1525-1534.

Hutterer R (2005) Order Erinaceomorpha. In: Mammal Species of the World. A Taxonomic and Geographic Reference (Ed. D.E.Wilson & D.M.Reeder). The John Hopkins University Press, Baltimore, pp. 212-219.

IUCN (2013) The IUCN Red List of Threatened Species, 2013.1. International Union for Conservation of Nature and Natural Resources. <http://www.iucnredlist.org/>

Juan C, Emerson BC, Oromí P and Hewitt GM (2000) Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *TREE*, 15:104-109.

Juste J, Ibáñez C, Muñoz J, Trujillo D, Benda P, Karatas A and Ruedi M (2004) Mitochondrial phylogeography of the long-eared bats (Plecotus) in the Mediterranean Palaearctic and Atlantic Islands. *Molecular Phylogenetics and Evolution*, 31:1114-1126.

Keller I and Largiadèr CR (2003) Recent habitat fragmentation caused by major roads leads to reduction of gene flow and loss of genetic variability in ground beetles. *Proceedings of the Royal Society B*, 270:417-423.

Khaldi M, Torres J, Samsó B, Miquel J, Biche M, Benyettou M, Barech G, Benelkadi HA and Ribas A (2012) Endoparasites (helminths and coccidians) in the hedgehogs *Atelerix algirus* and *Paraechinus aethiopicus* from Algeria. *African Zoology*, 47:48-54.

Kidd DM and Ritchie MG (2006) Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography*, 33:1851-1865.

Krettek A, Gullberg A and Arnason U (1995) Sequence Analysis of the Complete Mitochondrial DNA Molecule of the Hedgehog, Erinaceus europaeus, and the Phylogenetic Position of the Lipotyphla. *Journal of Molecular Evolution*, 41:952-957.

Krijgsman W, Hilgen FJ, Raffi I, Sierro FJ and Wilson DS (1999) Chronology, causes and progression of the Messinian salinity crisis. *Nature*, 400:652-655.

Kröpelin S and Swezey CS (2006) Revisiting the age of the Sahara desert. *Science*, 312:1138-1139.

Kröpelin S, Verschuren D, Lézine AM, Eggermont H, Cocquyt C, Francus P, Cazet JP, Fagot M, Rumes B, Russell JM, Darius F, Conley DJ, Schuster M, von Suchodoletz H and Engstrom DR (2008) Climate-Driven Ecosystem Succession in the Sahara: The Past 6000 Years. *Science,* 320:765-768.

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

Lalueza-Fox C, Castresana J, Sampietro L, Bonet TM, Alcover JA and Bertranpetit J (2005) Molecular dating of caprines using ancient DNA sequences of *Myotragus balearicus*, an extinct endemic Balearic mammal. *BMC Evolutionary Biology*, 5:1-11.

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

Le Berre M (1990) Faune du Sahara 2 – Mammiferes. Lechevalier, R. Chabaud.

Le Houérou HN (1997) Climate, flora and fauna changes in the Sahara over the past 500 million years. *Journal of Environments*, 37:619-647.

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

Loget N and Van Den Driessche (2006) On the origin of the Strait of Gibraltar. *Sedimentary Geology*, 188-189:341-356.

López-de-Heredia U, Jiménez P, Díaz-Fernández P and Gil L (2005) The Balearic Islands: a reservoir of cpDNA genetic variation for evergreen oaks. *Journal of Biogeography*, 32:939-949.

Martínez-Solano I, Gonçalves HA, Arntzen JW and García-Paris M (2004) Phylogenetic relationships and biogeography of midwife toads (Discoglossidae: *Alytes*). *Journal of Biogeography*, 31:603-618.

McNeely J (2001) Invasive species: a costly catastrophe for native biodiversity. *Land Use and Water Resources Research*, 1:1-10.

Médail F and Quézel P (1999) Biodiversity Hotspot in the Mediterranean Basin: Setting Global Conservation Priorities. *Conservation Biology*, 13:1510-1513.

Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farelo L and Alves PC (2012) Recurrent introgression of mitochondrial DNA among hares (Lepus spp.) revealed by species-tree inference and coalescent simulations. *Systematic Biology*, 61:367-381.

Mendenhall CD, Daily GC and Ehrlich PR (2012) Improving estimates of biodiversity loss. *Biological Conservation*, 151:32-34.

Messerli B and Winiger M (1992) Climate, environmental change, and resources of the African mountains from the Mediterranean to the Equator. *Mountain Research and Development*, 12:315-336.

Metallinou M, Arnold EN, Crochet PA, Geniez P, Brito JC, Lymberakis P, Baha El Din S, Sindaco R, Robinson M and Carranza S (2012) Conquering the Sahara and Arabian deserts: systematics and biogeography of Stenodactylus geckos (Reptilia: Gekkonidae). *BMC Evolutionary Biology*, 12:1-17.

Millot C and Taupier-Letage I (2005) Circulation in the Mediterranean Sea. The Handbook of Environnemental Chemestry (The Natural Environment and the Biogical Cycles), Springer-Verlag Editor.

Mitchell-Jones AJ, Amori G, Bogdanowicz W, Krystufek B, Reijnders PJH, Spitzenberger F, Stubbe C, Thissen JBM, Vohralík Vand Zima J (1999) The Atlas of European Mammals. Poyser Natural History.

Morales A and Rofes J (2008) Early evidence of the Algerian hedgehogs in Europe. *Journal of Zoology*, 274:9-12.

Morshed SH and Patton JL (2002) New records of mammals from Iran with systematic comments on hedgehog (Erinaceidae) and mouse-like hamsters (Calomyscus, Muridae). *Zoology in the Middle East*, 26:49-58.

Mouline K, Granjon L, Galan M, Tatard C, Abdoullaye D, Ag Atteyine S, Duplantier JM and Cosson JF (2008) Phylogeography of a Sahelian rodent species *Mastomys huberti*: a Plio-Pleistocene story of emergence and colonization of humid habitats. *Molecular Ecology*, 17:1036-1053.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB and Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature,* 403: 853–858.

Nicholson SE (1995) Sahel, West Africa. *Encyclopedia of Environmental Biology*, 3:261-275.

Nicholson SE (2013) The West African Sahel: A Review of Recent Studies on the Rainfall Regime and Its Interannual Variability. *ISRN Meteorology*, 2013:1-32.

Nicolas V, Granjon L, Duplantier JM, Cruaud C and Dobigny G (2009) Phylogeography of spiny mice (genus *Acomys*, Rodentia: Muridae) from the south-western margin of the Sahara with taxonomic implications. *Biological Journal of the Linnean Society*, 98:29-46.

Nicolas V, Quérouil S, Verheyen E, Verheyen W, Mboumba JF, Dillen M and Colyn M (2006) Mitochondrial phylogeny of African wood mice, genus *Hylomyscus* (Rodentia, Muridae): Implications for their taxonomy and biogeography. *Molecular Phylogenetics and Evolution*, 38:779-793.

Niethammer J (1969) Zur Taxonomie der Ohrenigel in Afghanistan (Gattung *Hemiechinus*). *Z. Sãugetierk*, 34:257-274.

Nikaido M, Cao Y, Harada M, Okada N and Hasegawa M (2003) Mitochondrial phylogeny of hedgehogs and monophyly of Eulipotyphla. *Molecular Phylogenetics and Evolution*, 28:276-284.

Nogales M, Rodríguez-Luengo JL and Marrero P (2006) Ecological effects and distribution of invasive non-native mammals on the Canary Islands. *Mammal Review*, 36:49-65.

Novacek MJ (1986) The Skull of Leptictid Insectivorans and the Higher-Level Classification of Eutherian Mammals*. Bulletin of the American Museum of Natural History*, 183: 1–111.

Olff H and Ritchie ME (2002) Fragmented nature: consequences for biodiversity. *Landscape and Urban Planning*, 58:83-92.

Opdam P and Wascher D (2004) Climate change meets habitat fragmentation: linking landscape and biogeographical scale levels in research and conservation. *Biological Conservation*, 117:285-297.

Palomo LJ and Gisbert J (2002) Atlas de los Mamíferos terrestres de España. Dirección General de Conservación de la Natureza, SECEM-SECEMU.

Pauls SU, Nowak C, Bálint M and Pfenninger M (2013) The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, 22:925-946.

Petter F (1974) Facteurs de repartition des rongeurs sahariens et périsahariens. In: Symposium Israel-France: recherches écologiques relatives au développement des zones arides (deserts méditerranéens à précipitations hivernales). Organ. Rech. Agron, 91-97.

Pinho C, Ferrand N and Harris D J (2006). Reexamination of the Iberian and North African Podarcis (Squamata: Lacertidae) phylogeny based on increased mitochondrial DNA sequencing. *Molecular Phylogenetics and Evolution*, 38:266–273.

Posada D and Crandall KA (1998) ModeltestL testing the model of DNA substitution. *Bioinformatics* 14:817-818.

Quammen D (1996) The song of the Dodo. Island biogeography in an age of extinction. Hutchinson, London.

Rambaut A (2009) FigTree, a graphical viewer of phylogenetic trees. Institute of Evolutionary Biology University of Edinburgh.

Richardson DM and Whittaker RJ (2010) Conservation biogeography – foundations, concepts and challenges. *Diversity and Distributions*, 16:313-320.

Riley PY and Chomel BB (2005) Hedgehog zoonoses. *Emerging Infectious Diseases*, 11:1-5.

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

Sala OE, Chapin III FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA,Oesterheld M, Poff NL, Sykes MT, Walker BH, Walker M, Wall DH (2000) Global Biodiversity Scenarios for the Year 2100. *Science*, 287:1770-1774.

Salvi D, Harris DJ, Kaliontzopoulou A, Carretero MA and Pinho C (2013) Persistence across Pleistocene ice ages in Mediterranean and extra-Mediterranean refugia: phylogeographic insights from the common wall lizard. *BMC Evolutionary Biology*, 13:147.

Santana EM, Jantz HE and Best TL (2010) *Atelerix albiventris* (Erinaceomorpha: Erinaceidae). *Mammalian Species*, 42:99-110.

Santucci F, Emerson BC and Hewitt GM (1998) Mitochondrial DNA phylogeography of European hedgehogs. *Molecular Ecology*, 7:1163-1172.

Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, 4:1-13.

Schüle W (1993) Mammals, vegetation and the initial human settlement of the mediterranean islands-a palaeoecological approach. *Journal of Biogeography*, 20:399–412.

Seddon JM, Santucci F, Reeve NJ and Hewitt GM (2001) DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor*: Pleistocene refugia, postglacial expansion and colonization routes. *Molecular Ecology*, 10:2187-2198.

Sequeira F, Sodré D, Ferrand N, Bernardi JAR, Sampaio I, Schneider H and Vallinoto M (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*, 11:1-15.

Spitzenberger F (1978) Die Sãugetierfauna Zyperns Teil I: Insectivora und Rodentia. Annln naturh. Mus. Wien, 81:401-441.

Sunnucks P (2000) Efficient genetic markers for population biology. TREE, 15:199-203.

Symonds MRE (2005) Phylogeny and life histories of the “Insectivora”: controversies and consequences. *Biological Reviews*, 80:93-128.

Swezey CS (2009) Cenozoic stratigraphy of the Sahara, Northern Africa. *Journal of African Earth Sciences*, 53:89-121.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony methods. *Molecular Biology and Evolution*, 28:2731-2739.

Terrab A, Hampe A, Lepais O, Talavera S, Vela E and Stuessy TF (2008) Phylogeography of North African Atlas cedar (*Cedrus atlantica*, Pinaceae): Combined molecular and fossil data reveal a complex Quaternary history. *American Journal of Botany*, 95:1262-1269.

Terrasa B, Pérez-Mellado V, Brown RP, Picornell A, Castro JA and Ramon MM (2008) Foundations for conservation of intraspecific genetic diversity revealed by analysis of phylogeographical structure in the endangered endemic lizard *Podarcis lilfordi. Diversity and Distributions*, DOI: 10.1111/j.1472-4642.2008.00520.x.

Velo-Antón G, Martínez-Solano I and García-París M (2008) ß-fibrinogen intron 7 variation in *Discoglossus* (Anura: Discoglossidae): implications for the taxonomic assessment of morphologically cryptic species. *Amphibia-Reptilia*, 29:523-533.

Vogel P, Cosson JF and Jurado LF (2003) Taxonomic status and origin of the shrews (Soricidae) from the Canary islands inferred from a mtDNA comparison with the European Crocidura species. *Molecular Phylogenetics and Evolution*, 27:271-282.

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

Wilson DE and Reeder DM (2005) Mammal Species of the World – A taxonomic and geographic reference. Johns Hopkins University Press.

Whittaker RJ, Araújo MB, Jepson P, Ladle RJ, Watson JEM and Willis KJ (2005) Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, 11:3-23.

Zeng N, Neelin JD, Lau KM and Tucker CJ (1999) Enhancement of interdecadal climate variability in the Sahel by vegetation interaction. *Science,* 286:1537-1540.

Zhang DX and Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, 12:563-584.

Annexes

A1) Adapted Protocol of JETQUICK Tissue DNA Spin Kit from Genomed

Part A: Processing of the tissue sample

1) Cut the tissue sample into small pieces with a scalpel.

2) Add 200 µl of Buffer T1 to the homogenized sample from step 1 in a 1.5 ml- or 2 ml reaction vessel (i.e. Eppendorf). Mix buffer and sample thoroughly by inverting the reaction vessel several times.

3) Add 25 µl of proteinase K (25 mg/ml) to the mixture from step 2. Mix thoroughly by inverting the tube several times and incubate overnight at 56°C. During the incubation mix several times thoroughly by inverting to get the tissue material dissolved as good as possible. If there is still particular material visible after the overnight incubation at 56°C, extend the incubation time until all material has dissolved.

Part B: JETQUICK Spin Column Procedure

1) Place the Tris-HCL in a stove at 70°C.

2) Add 200 µl Buffer T2 to the cleared lysate of the last step of part A and mix thoroughly. Incubate for 10 min at 70°C.

3) Let the mixture cool down for approximately 1 min. Then add 200 µl of absolute ethanol. Mix quickly and very thoroughly in order to prevent a precipitation of DNA due to too high local alcohol concentrations.

4) Assemble a spin unit by fitting a JETQUICK micro-spin column into the suitable receiver tube. Transfer the mixture from step 3 into the reservoir of the micro-spin column and centrifuge for 1 min at 8000 rpm.

5) Discard the flowthrough and re-combine the micro-spin column with the used receiver tube. Pipette 500 µl of reconstituted buffer TX into the reservoir of the micro-spin column and centrifuge for 1 min at 8000 rpm.

6) Discard the flowthrough, re-combine the micro-spin tube and the used receiver tube. Repeat step 5 with 500 µl of reconstituted buffer T3 as described above.

7) Discard the flowthrough and again combine the micro-spin tube and the used receiver tube. Centrifuge for 3 min at 14 000 rpm to remove residual buffer T3.

8) Discard the receiver tube and insert the JETQUICK micro-spin column into a clean, sterile 1.5 ml reaction tube. To elute the DNA, pipet 100 µl prewarmed (65-70°C) elution buffer(10 mM TrisHCl [pH 9,0]) directly onto the surface of the silica membrane, incubate for 5 min at room temperature, and centrifuge subsequently for 1 min at 8 000 rpm.
[OPTIONAL: In order to increase the elution efficiency, repeat this elution step with another 100 µl of elution buffer onto the silica membrane in the micro-spin tube. Then proceed as described before.]

A2) List of samples that amplified at list for one type of molecular marker;

Code	Country	Region	Latitude	Longitude	Marker	H Cyt-b	H B-fibr
143	Morocco	Dakhla	24.846000	-14.845100	Cyt-b	Mo19	-
306	Tunisia	Le Kef	36.167467	8.747683	Cyt-b	Va1	-
322	Libya	Az Zawiyah	32.404833	11.768533	Cyt-b	Va3	-
323	Libya	Ghadamis	32.298483	11.752083	Cyt-b /B-fibr	Va3	Ma1
327	Libya	Ghadamis	30.187450	9.872083	Cyt-b	Li2	-
339	Libya	Murzuq	25.928567	14.248417	Cyt-b /B-fibr	Li1	Mo5
398	Niger	Tahoua	14.829933	5.305150	Cyt-b /B-fibr	Ni8	Ni3
404	Niger	Dosso	13.009033	3.612667	Cyt-b /B-fibr	Ni7	Ni4
406	Niger	Tahoua	13.097033	2.970450	Cyt-b	Ni5	-
433	Mali	Koulikoro	12.747183	-7.248650	Cyt-b	Mal1	-
436	Mali	Koulikoro	14.313617	-8.140417	Cyt-b	Mal1	-
437	Mali	Kayes	14.543417	-9.509217	Cyt-b	Mal3	-
440	Mali	Kayes	14.538083	-9.524733	Cyt-b	Mal4	-
441	Mali	Kayes	14.524150	-9.544400	Cyt-b /B-fibr	Mal4	Mal3
459	Mali	Kayes	14.504000	-11.090983	Cyt-b /B-fibr	Mal1	Mal1
470	Senegal	Saint-Louis	16.076883	-13.905333	Cyt-b	Se1	-
479	Senegal	Saint-Louis	16.349150	-14.248400	Cyt-b /B-fibr	Se2	Ma1
482	Senegal	Saint-Louis	16.522917	-15.424050	B-fibr	-	Ma1
510	Mauritania	Inchiri	19.608867	-14.570167	B-fibr	-	Ma2
515	Mauritania	Inchiri	19.844800	-14.246483	Cyt-b	Va3	-
614	Tunisia	Jendouba	36.464133	8.505733	Cyt-b	Tu1	-
617	Tunisia	Kasserine	35.007633	8.613067	Cyt-b	Va2	-
717	Morocco	Essaouira	31.590500	-9.104717	Cyt-b /B-fibr	Mo20	MO3
720	Portugal	Douro Litoral	41.333333	-8.666667	Cyt-b /B-fibr	PI1	PI1
721	Portugal	Alentejo	38.641667	-7.750000	Cyt-b	Va2	-
766	Morocco	Oujda	34.693633	-2.026383	Cyt-b	Mo4	-
779	Morocco	Nador	35.102283	-2.869150	Cyt-b /B-fibr	Mo10	Mo3
781	Spain	Castilla-León	42.502350	-5.598633	Cyt-b	PI3	-
782	Portugal	Minho	41.927883	-8.587467	Cyt-b	PI1	-
810	Morocco	El Kelaa	31.782500	-7.587350	Cyt-b /B-fibr	Mo18	Mo3
848	Morocco	Ifrane - Fes	33.552300	-4.975350	Cyt-b	Mo30	-
905	Algeria	Batna	35.225883	5.708083	Cyt-b /B-fibr	Al1	Mo3
940	Morocco	Kenitra	34.773333	-5.515000	Cyt-b	Mo26	-
942	Spain	Castilla-León	42.436000	-3.720667	Cyt-b	PI4	-
986	Mali	Ségou	13.750100	-5.152900	Cyt-b /B-fibr	Mal2	Mal2
988	Mali	Ségou	13.483017	-5.812717	Cyt-b /B-fibr	Mal2	Mal3
1012	Spain	Andaluzia	37.133333	-3.666667	Cyt-b	PI1	-
1017	Spain	Andaluzia	37.368983	-2.754367	Cyt-b	PI1	-
1020	Morocco	Kenitra	34.185417	-5.582283	Cyt-b /B-fibr	Mo27	Mo3
1032	Morocco	Khénifra	32.676367	-5.330883	Cyt-b	Mo2	-
1054	Morocco	Marrakech	31.544517	-8.002867	Cyt-b /B-fibr	Mo19	Mo3
1076	Morocco	El Kelaa	31.873933	-7.456317	Cyt-b /B-fibr	Va4	Mo3
1142	Spain	Castilla-León	42.745000	-3.806500	Cyt-b /B-fibr	PI2	PI2

1311	Mauritania	Brakna	16.302783	-13.879300	Cyt-b	Va4	-
1468	Turkey	Firat	38.211680	35.865050	Cyt-b /B-fibr	Tk1	Tk1
1469	Turkey	Bingol	39.156730	40.898350	B-fibr	-	Tk2
1470	Turkey	Derinkuyu	38.385220	34.735640	B-fibr	-	Tk1
1559	Tunisia	Jendouba	36.457700	8.623017	Cyt-b /B-fibr	Va1	Tu1
2207	Mauritania	Hodh El Gharbi	16.382675	-9.564807	Cyt-b /B-fibr	Va4	Ma1
2418	Mauritania	Assaba	16.540298	-10.865143	Cyt-b	Mal2	-
2784	Italy	Sardinea	40.311440	9.206500	Cyt-b /B-fibr	Eu3	Eu1
2801	Portugal	Alentejo	37.922767	-7.507517	Cyt-b /B-fibr	PI7	PI1
2805	Portugal	Algarve	37.290300	-8.592200	Cyt-b /B-fibr	PI6	PI1
2808	Spain- Canarias	Fuerteventura	28.572814	-13.964958	Cyt-b	Ca1	-
2809	Spain- Canarias	Fuerteventura	28.486051	-13.948631	Cyt-b	Ca1	-
2811	Spain- Canarias	Fuerteventura	28.454258	-13.984501	Cyt-b	Ca2	-
2816	Spain- Canarias	Lanzarote	29.109026	-13.551832	Cyt-b /B-fibr	Ca1	Mo3
2821	Spain- Canarias	Lanzarote	29.177079	-13.462477	Cyt-b	Ca1	-
2823	Armenia	Caucaso	39.000000	46.500000	Cyt-b	Ar1	-
2824	Greece	Kedriki Makedhonía	41.237080	23.086500	Cyt-b	Gr1	-
2825	Greece	Kedriki Makedhonía	41.258710	23.095020	Cyt-b /B-fibr	Gr3	Eu4
2827	Greece	Olympus	40.220810	22.540970	Cyt-b	Gr2	-
2914	Mauritania	Adrar	20.095725	-13.203550	Cyt-b	Mal2	-
3018	Mauritania	Tagant	18.255785	-11.785698	Cyt-b	Va6	-
3525	Mauritania	Inchiri	19.462235	-16.070569	Cyt-b	Va1	-
3804	Cyprus	Agia Anna	35.000000	33.000000	Cyt-b /B-fibr	Cy1	CL1
3806	Cyprus	Fountoukodasi	35.000000	33.000000	Cyt-b /B-fibr	Cy2	CL1
3807	Libya	Cyrenaica	32.391198	21.240381	Cyt-b /B-fibr	Ly1	CL1
3817	Malta	Wardija	35.830411	14.470076	Cyt-b	Mt1	-
3827	Morocco	Tiznit	29.708333	-9.755000	Cyt-b /B-fibr	Ca1	Mo3
3828	Morocco	Essaouira	31.818333	-9.543333	Cyt-b	Mo20	
3837	Morocco	Bouarfa	32.062880	-3.082630	Cyt-b /B-fibr	Va6	Mo3
3839	Morocco	Oujda	34.597600	-2.358700	Cyt-b	Mo15	-
3841	Morocco	Ouarzazate	31.228710	-7.425470	Cyt-b /B-fibr	Mo11	Mo3
3842	Morocco	Khouribga	33.134920	-6.665773	Cyt-b /B-fibr	Mo16	Mo3
3848	Spain	Alicante	38.083700	-0.839500	Cyt-b	PI5	-
3849	Egypt	Giza	30.031944	31.075000	Cyt-b	Va3	-
3851	Unknown	Unknown			Cyt-b	Eu2	-
5643	Morocco	Ouarzazate	31.181745	-7.458934	Cyt-b /B-fibr	Mo11	Mo3
5664	Morocco	Beni Mellal	32.472984	-5.992619	Cyt-b /B-fibr	Mo25	Mo4
5665	Morocco	Khénifra	32.747173	-5.722872	Cyt-b /B-fibr	Mo2	Mo3
5666	Morocco	Khénifra	32.753637	-5.089792	Cyt-b	Mo2	-
5667	Morocco	Khénifra	32.749937	-5.031357	Cyt-b /B-fibr	Mo3	Mo3
5668	Morocco	Khénifra	32.749937	-5.031357	Cyt-b /B-fibr	Va4	Mo3
5669	Morocco	Khénifra	32.894316	-4.999255	Cyt-b /B-fibr	Mo2	Mo4
5683	Morocco	Tanger	35.368977	-6.015354	Cyt-b /B-fibr	Va1	Mo3
5684	Morocco	Tanger	35.661802	-5.951930	Cyt-b /B-fibr	Mo17	Mo1
5685	Spain-Baleares	Ibiza	39.010810	1.487720	Cyt-b /B-fibr	Ba1	Ba2
5686	Spain-Baleares	Ibiza	39.084100	1.499060	Cyt-b	Ba1	-

5687	Spain-Baleares	Ibiza	38.949800	1.283370	Cyt-b /B-fibr	Ba1	Ba1
5688	Spain-Baleares	Formentera	38.694780	1.463400	Cyt-b	Ba1	-
5689	Spain-Baleares	Ibiza	38.979530	1.441210	Cyt-b /B-fibr	Ba1	Ba1
5690	Spain-Baleares	Maiorca	39.650800	2.670030	Cyt-b /B-fibr	Ba2	Mo1
5692	Algeria	El Bayadh	31.994000	1.640167	Cyt-b	Ba1	-
5709	Morocco	Tan-Tan	28.818150	-10.373785	Cyt-b /B-fibr	Mo12	Mo3
5716	Morocco	Tan-Tan	28.271823	-11.564715	Cyt-b /B-fibr	Va5	Mo3
5732	Morocco	Rio de Oro	23.117098	-16.070763	Cyt-b	Va5	-
5914	Mauritania	Adrar	20.269873	-13.230292	Cyt-b /B-fibr	Va3	Ma1
5916	Mauritania	Adrar	20.127462	-13.691313	Cyt-b	Ma1	-
6066	Mauritania	Tagant	17.632735	-11.460103	B-fibr	-	Ma1
6311	Mauritania	Dakhlet-Nouâdhibou	19.567590	-16.383732	Cyt-b /B-fibr	Va5	Ma1
6580	Morocco	Tan-Tan	28.966112	-10.110801	Cyt-b /B-fibr	Mo22	Mo3
6610	Niger	Niamey	13.420390	2.295132	Cyt-b /B-fibr	Ni6	Ni4
6623	Niger	Zinder	13.707633	9.373178	Cyt-b /B-fibr	Ni3	Ni1
6647	Niger	Diffa	13.340055	12.614022	Cyt-b /B-fibr	Ni1	Ni4
6651	Niger	Diffa	13.511182	12.823018	Cyt-b	Ni4	-
6781	Niger	Zinder	14.161528	9.734708	Cyt-b /B-fibr	Ni3	Ni2
6791	Niger	Maradi	13.679638	6.512547	Cyt-b /B-fibr	Ni2	Ni3
6815	Tunisia	Jendouba	36.582730	8.694870	Cyt-b /B-fibr	Tu1	Mo3
6825	Tunisia	Kasserine	35.100440	8.737130	Cyt-b /B-fibr	Tu3	Mo3
6830	Tunisia	Kairouan	35.319380	9.754230	Cyt-b /B-fibr	Tu1	Mo1
6835	Tunisia	Zaghouan	36.183970	10.051690	Cyt-b /B-fibr	Tu2	Mo3
6845	Algeria	Tlemcen	35.182661	-1.647738	Cyt-b	Mo10	-
6860	Morocco	Figuig	32.595320	-1.941050	Cyt-b /B-fibr	Mo4	Mo3
6862	Morocco	Ksar es Souk	31.961490	-3.491210	Cyt-b	Mo5	-
6863	Morocco	Meknes	33.621840	-5.080800	Cyt-b /B-fibr	Mo28	Mo3
6864	Morocco	Fes	33.623820	-4.868900	Cyt-b /B-fibr	Mo30	Mo3
6866	Morocco	Al Hoceima	34.973450	-4.411630	Cyt-b /B-fibr	Mo6	Mo1
6867	Morocco	Al Hoceima	34.966400	-4.336190	Cyt-b /B-fibr	Mo2	Mo2
6868	Morocco	Al Hoceima	34.948920	-4.333440	Cyt-b	Mo24	-
6869	Morocco	Al Hoceima	35.025110	-4.161230	Cyt-b	Mo15	-
6873	Morocco	Nador	34.976480	-3.347020	Cyt-b /B-fibr	Mo29	Mo3
6874	Morocco	Nador	35.126670	-2.391170	Cyt-b	Mo8	-
6875	Morocco	Oujda	34.868010	-2.452060	Cyt-b /B-fibr	Mo5	Mo3
6878	Morocco	Oujda	34.509350	-2.050840	Cyt-b	Mo4	-
6888	Morocco	Oujda	34.738900	-2.375590	Cyt-b /B-fibr	Mo5	Mo3
6890	Morocco	Oujda	34.738900	-2.375590	Cyt-b	Mo5	-
6893	Morocco	Oujda	34.868800	-2.450320	Cyt-b	Mo5	-
6898	Morocco	Nador	35.011010	-3.211200	Cyt-b	Mo14	
6899	Morocco	Nador	35.075020	-3.249800	Cyt-b /B-fibr	Mo13	Mo4
6900	Morocco	Nador	35.075300	-3.253680	Cyt-b /B-fibr	Mo10	Mo6
6902	Morocco	Al Hoceima	35.039290	-4.116090	Cyt-b	Mo24	-
6919	Morocco	Rabat	33.902540	-6.825770	Cyt-b /B-fibr	Mo26	Mo3
6925	Morocco	Meknes	34.002700	-5.170460	Cyt-b	Mo7	-
6964	Morocco	Boulemane	32.808780	-4.415560	Cyt-b	Mo23	-

6969	Morocco	Khénifra	32.763620	-5.004140	Cyt-b /B-fibr	Mo1	Mo1
6970	Morocco	Khénifra	32.751240	-5.084740	Cyt-b	Mo2	-
7020	Morocco	Khénifra	32.676800	-4.607110	Cyt-b	Mo23	-
7037	Morocco	Ouarzazate	31.146850	-7.421310	Cyt-b	Mo21	-
7114	Morocco	Figuig	32.967360	-2.029520	Cyt-b /B-fibr	Mo9	Mo2
7119	Morocco	Figuig	33.038340	-2.003260	Cyt-b /B-fibr	Mo5	Mo3

A3) List of Cyt-b and B-fibr sequences available on GenBank and used in this study;

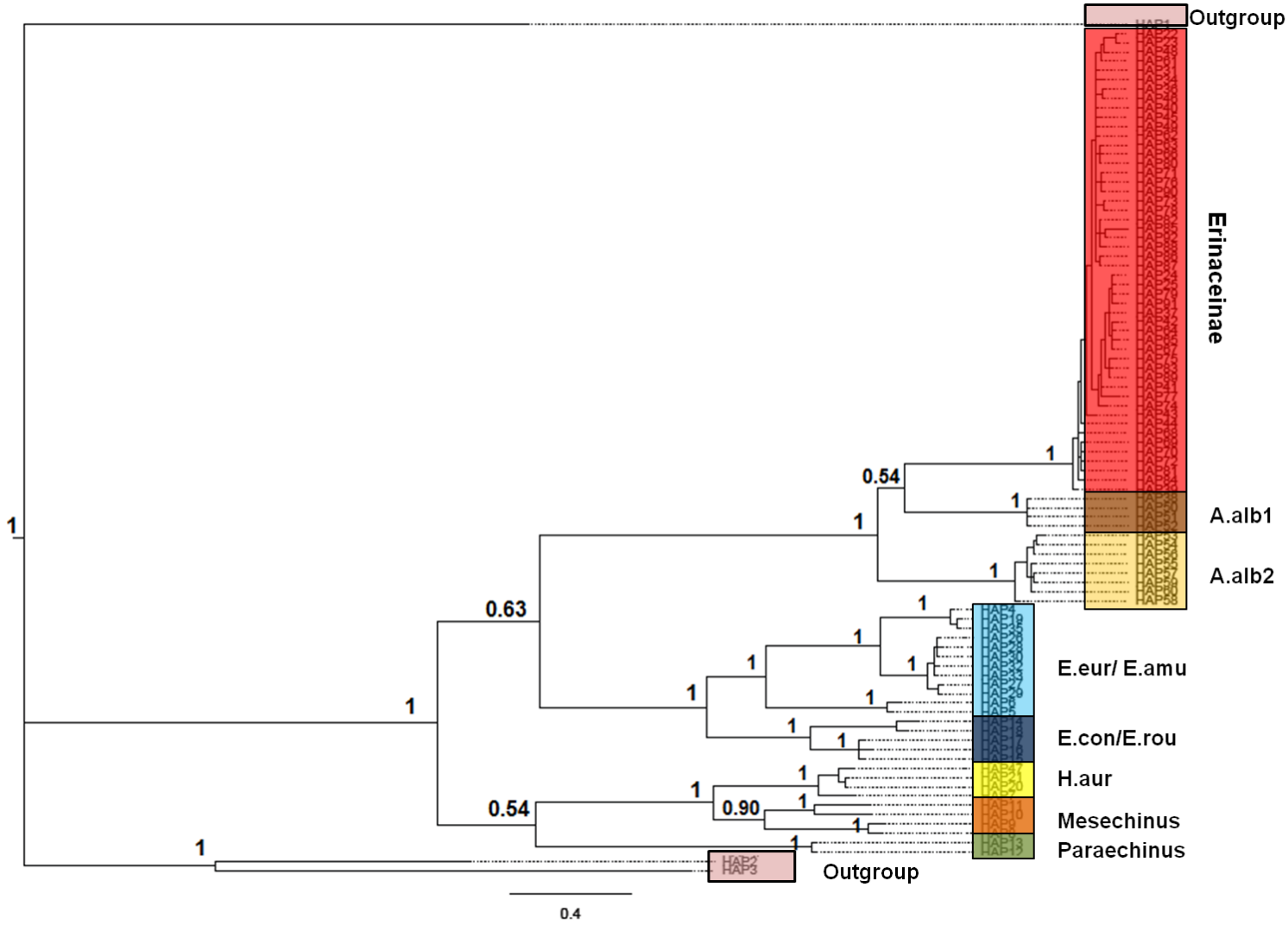
Cyt-b

CODE	Species	Country	Local	Reference
NC_002080	*Erinaceus europaeus*	Sweden	Kãvlinge	Krettek et al., 1995
X88898	*Erinaceus europaeus*	Sweden	Kãvlinge	Krettek et al., 1995
HQ857520	*Erinaceus amurensis*	China	Liaoning	He et al., 2012
HQ857521	*Erinaceus amurensis*	China	Hubei	He et al., 2012
HQ857522	*Hemiechinus auritus*	x	x	He et al., 2012
HQ857530	*Mesechinus hughi*	China	Shanxi	He et al., 2012
HQ857531	*Mesechinus hughi*	China	Shanxi	He et al., 2012
HQ857528	*Mesechinus dauuricus*	China	Ningxia	He et al., 2012
HQ857526	*Mesechinus dauuricus*	China	Liaoning	He et al., 2012
HQ857538	*Paraechinus aethiopicus*	Qatar	x	He et al., 2012
HQ857537	*Paraechinus aethiopicus*	Qatar	x	He et al., 2012

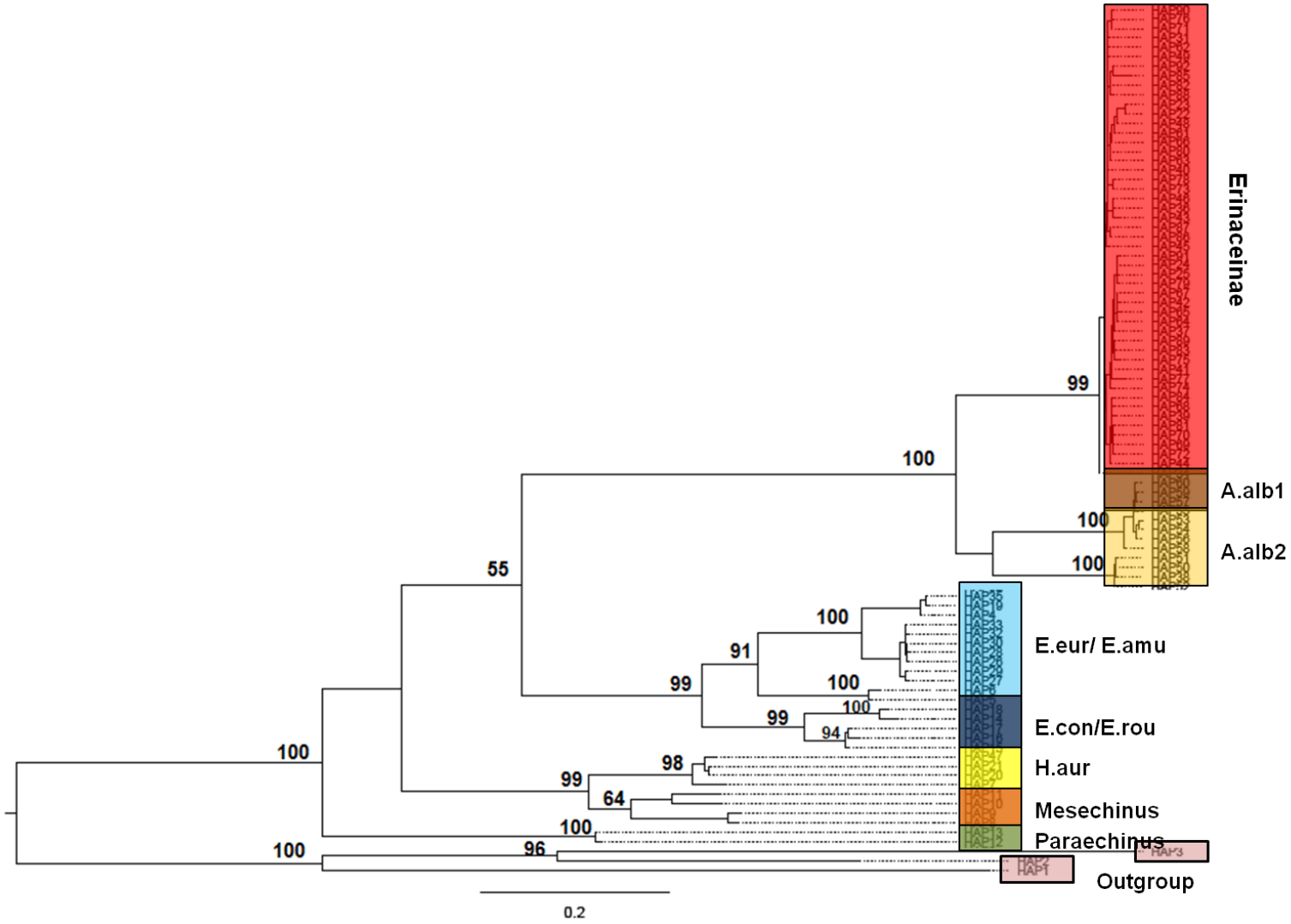
B-fibr

CODE	Species	Local	Paper
AF379832	*Erinaceus europaeus*	Europe	Seddon et al., 2001
AF379833	*Erinaceus europaeus*	Europe	Seddon et al., 2001
AF379835	*Erinaceus europaeus*	Europe	Seddon et al., 2001
AF379837	*Erinaceus europaeus*	Europe	Seddon et al., 2001
AF379840	*Erinaceus europaeus*	Europe	Seddon et al., 2001
AF379844	*Erinaceus europaeus*	Europe	Seddon et al., 2001
AF379846	*Erinaceus europaeus*	Europe	Seddon et al., 2001
AF379850	*Erinaceus concolor*	Europe	Seddon et al., 2001
AF379853	*Erinaceus concolor*	Europe	Seddon et al., 2001

A4) Bayesian phylogenetic tree of mtDNA dataset with outgroups;



A5) Phylogenetic tree inferred by Maximum Likelihood method and based on mtDNA dataset with outgroups;



A6) Phylogenetic tree based on nDNA data with Bayesian posterior probabilities and bootstrap values of ML and NJ. Outgroup sample included.