Phylogeography and current patterns of genetic diversity and structure of the Mediterranean pond turtle

Joana Filipa da Silva Veríssimo
Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia
2014

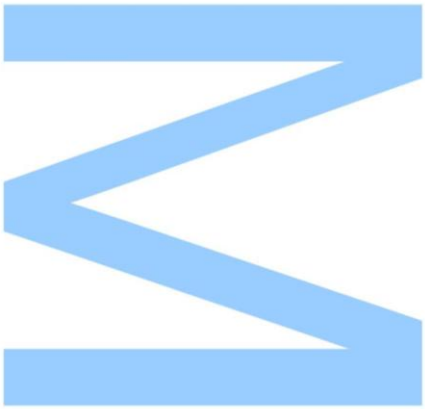
Orientador
José Fernando Alves Teixeira, Investigador CIIMAR

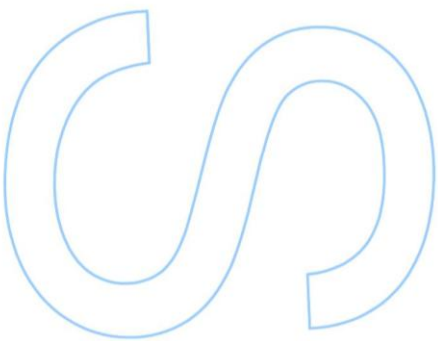
Coorientador
Guillermo Velo-Antón, Post-Doc, CIBIO

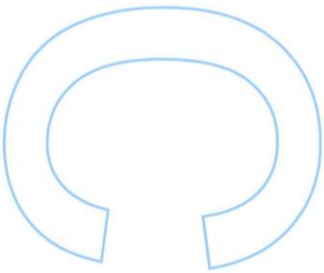
Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____







Agradecimentos

Em primeiro lugar, quero agradecer ao Doutor José Teixeira e ao Doutor Guillermo Velo-Antón, pela ajuda, empenho e disponibilidade durante a preparação desta tese.

Ao Doutor José Carlos Brito e restantes membros do grupo Biodeserts, por me terem acolhido tão bem.

Um obrigado especial à Teresa Silva e à Sónia Ferreira pelos últimos dias desta tese.

A todas as pessoas que comigo partilharam os seus conhecimentos no laboratório, principalmente à Susana, Sofia, Patrícia e Filipa.

Quero agradecer também à equipa técnica do Parque Biológico de Gaia, principalmente à Ana Mafalda Cruz, por toda a ajuda disponibilizada.

A todas as meninas do 304, as que ainda lá estão e às que por lá já passaram, um obrigado pela convivência e amizade.

A todos os elementos do Masterss, incluindo os adotados, obrigado por tudo. Um obrigado ao Hugo por conseguir encontrar sempre os papers que lhe pedi e à Diana pela ajuda nos últimos dias.

À Joanita, um pedido de desculpas pelas ausências em Agosto.

À Ana e à Leonor, muito obrigado por estarem sempre nos momentos melhores e piores.

Ao Tiago e à Luísa pela presença constante.

A todos os meus familiares e amigos que me apoiaram, principalmente neste último ano um muito obrigado pelo carinho.

Um enorme obrigado aos meus pais por tudo o que têm feito por mim. Um obrigado à minha mãe, por ter sido tão forte, durante todo este doloso processo, pelo qual passámos. Um muito obrigado ao meu pai, por tudo o que foi e continuará a ser. Por nunca ter desistido de sorrir e lutar até ao fim, mesmo após saber que nunca o poderia vencer.

Ao Paulo, pelo apoio nos momentos difíceis e companheirismo nos melhores.

Este estudo foi parcialmente financiado por LIFE Trachemys Project (LIFE09 NAT/ES/000529).

Abstract

The complex palaeogeographic history of the Mediterranean Basin lead to the high levels of diversity and endemisms in the area, as of that, this region is now included in the 25 global hotspots of biodiversity. For instance, the Milankovitch climatic oscillations induced range retractions to the Southern European Peninsulas and the Maghreb, where temperate species found refuge, allowing them to survive the colder periods. This process also induced allopatric diversification in some species as different populations took refuge in different regions and so became isolated. Ectothermic species depend on climate induced temperatures to survive, making them more susceptible to suffer retraction/expansion events during those times of climatic instability.

The Mediterranean pond turtle, *Mauremys leprosa,* occurs widely throughout the Iberian Peninsula and most of the Maghreb region. Currently, two subspecies are recognized: *M. l. saharica* (ranging from southern of the Atlas Mountains to Tunisia) and *M. l. leprosa* (northern of the Atlas Mountains and in the Iberian Peninsula). For this work, we aim to explore the effect that past climatic oscillations and landscape barriers produced in the current patterns of genetic diversity and structure of *M. leprosa*, to do so, we used two fragments of mitochondrial DNA. Also, we intend to assess population genetic patterns and structure within this species at a more recent-scale. As of that, microsatellite loci were here optimized for the first time for *Mauremys leprosa* by cross-amplification of two closely related species.

Mitochondrial DNA (cyt-*b* and D-loop), retrieved from 163 specimens, showed deep genetic structure and higher levels of genetic diversity in North Africa, reinforcing the hypothesis of an African origin of the Iberian populations. Moreover, a secondary contact zone within the species was found in the Rif and Middle Atlas region. Microsatellite loci (genotyped in 556 individuals) revealed lower genetic structure in Morocco than in the Iberian Peninsula. However, for the latter, no geographical patterns were found. Furthermore, the high levels of genetic diversity found in southern populations of Iberian Peninsula might indicate a late Pleistocene *refugia* in the area, however, further studies are needed to clarify the role of this area during climatic oscillations. Regarding the secondary contact zone, this fast evolving marker revealed gene flow between the subspecies.

Overall, this study sheds new light into the role of both geographical and climatic features on the genetic diversity and structure patterns of *Mauremys leprosa*, complementing the current knowledge on the importance of North Africa as a *refugia.*

Keywords: Biogeography, Climatic Oscillations, Glacial *Refugia*, Iberian Peninsula, Maghreb, *Mauremys leprosa,* Microsatellites, mtDNA, Phylogeography, Secondary contact zones.

Resumo

A bacia do Mediterrâneo alberga um elevado número de endemismos e diversidade, em resultado da sua complexa história paleogeográfica, valendo-lhe um lugar entre os 25 *hotspots* de biodiversidade. Por exemplo, as oscilações climáticas induziram contrações na distribuição de espécies temperadas, levando a que estas se refugiassem nas penínsulas do sul da Europa, permitiu-lhes persistir durante os períodos glaciares. Algumas populações refugiaram-se em diferentes locais geográficos, levando assim ao seu isolamento, tendo como consequência a ocorrência de divergência alopátrica. As espécies ectotérmicas são altamente dependentes das temperaturas induzidas pelo clima para a sua sobrevivência, tornando-as extremamente suscetíveis a eventos de retração/expansão durante o período de instabilidade climática.

O cágado-mediterrânico, *Mauremys leprosa,* encontra-se distribuído pela Península Ibérica e na maioria da região Magrebina. Atualmente, duas subespécies são reconhecidas: *M. l. saharica* (distribuindo-se do sul das montanhas do Atlas até à Tunísia) e *M. l. leprosa* (distribuindo-se do norte das montanhas do Atlas até à Península Ibérica). O nosso objetivo neste trabalho foi, não só avaliar o efeito que as oscilações climáticas e barreiras geográficas à dispersão produziram nos atuais padrões de diversidade genética e de estrutura em *M. leprosa,* através de ADN mitocondrial, mas também, avaliar os mesmos padrões a uma escala mais recente. Para tal, otimizámos, pela primeira vez, microssatélites para *M. leprosa* através de amplificação-cruzada, utilizando marcadores previamente desenvolvidos para duas espécies próximas.

ADN mitocondrial (cyt-*b* e D-loop) de 163 indivíduos revelou níveis profundos de estrutura genética e elevada diversidade no Norte de África, reforçando assim a hipótese de que os indivíduos da Península Ibérica têm origem africana. Além disso, foi possível identificar uma possível zona de contacto entre as duas subespécies no Rif e Médio Atlas. Os dados obtidos através da genotipagem de 556 indivíduos revelou uma menor estrutura genética em Marrocos, em comparação com a Península Ibérica. No entanto, nesta última, não foi encontrado nenhum padrão geográfico. Ainda assim, a ocorrência de populações com elevados valores de diversidade no sudoeste da Península Ibérica, leva-nos a ponderar sobre a existência de um refúgio glaciar na área durante o Pleistoceno Superior, no entanto, uma análise mais detalhada será necessária para determinar qual o papel desta região durante as oscilações climáticas.

Tendo em conta a zona de contato, quando esta é analisada com marcadores de elevada *taxa* mutacional, aparenta fluxo génico entre as subespécies.

Em suma, este estudo enriquece o conhecimento sobre o papel que as características geográficas e climáticas tiveram nos padrões de estrutura e diversidade genética observados em *Mauremys leprosa.* Para além disso, foi possível complementar o conhecimento atual em torno da importância do Norte de África como refúgio.

Palavras-chave: Biogeografia, Oscilações climáticas, Refúgios glaciares, Península Ibérica, Magrebe, *Mauremys leprosa,* Microssatélites, mtDNA, Filogeografia, Zona de Contato.

Index

Figure Index

Table Index

List of Abbreviations

µL – microlitre

AFLP - amplified fragment length polymorphisms

AICc – Akaike's information criterion corrected for finite sample size

AMOVA - Analysis of molecular variance

AR – allelic richness

bp - base pairs

ca. – circa

cyt-*b* - Cytochrome-*b*

D – Tajima's D

DNA - Deoxyribonucleic acid

F_{IS} – Inbreeding coefficient

F_{ST} – Fixation index

H_E – expected heterozygosity

H_O – observed heterozygosity

H-W E – Hardy-Weinberg equilibrium

IUCN – International Union for Conservation of Nature

K – Number of population assumed by STRUCTURE software

LD – Linkage disequilibrium

mtDNA – mitochondrial DNA

nuDNA – nuclear DNA

PCA – Principal Component Analysis

PCR – polymerase chain reaction

SSR - short sequence repeats

uHe – unbiased expected heterozygosity

π – Nucleotide diversity

Chapter 1: General Introduction

1.1 Biodiversity loss and Climate Change

Biodiversity can be defined as all the variety present in genes, species and ecosystems along with all the services that are provided to society (Wilson 1988; Rands *et al.* 2010). These services are extremely valuable to all species, however, we are losing biodiversity at an alarming pace (Global Biodiversity Outlook 2). For instance, in the Millennium Ecosystem Assessment (2005), it was concluded that roughly 60% of the provided ecosystems services were being unsustainably used and ultimately degraded. Thus, several nations signed agreements to preserve biodiversity in an attempt to halt biodiversity loss (Rands *et al.* 2010), such as the 2010 Biodiversity Target, signed in 2002. However, the several objectives set were not achieved (Global Biodiversity Outlook 3). This mass extinction, called by some as the sixth extinction (Wake & Vredenburg 2008), is being mainly provoked by five factors – habitat change, alien species, overexploitation, pollution, and climate change – which can all be linked to human activities (Global Biodiversity Outlook 3). However, a recent study considers that calling it the sixth mass extinction may be currently overestimating it, but they also point that we are moving faster towards it (Barnosky *et al.* 2011).

From all the factors harming biodiversity, climate change and habitat destruction and/or degradation are considered to be the most nefarious of all, with some estimates predicting up to 37% of species going extinct by 2050 (based on climate change) (Thomas *et al.* 2004). The magnitude of the climate change we are facing is intensified by the industrial era we have been going through for the past decades, and it is very important to establish both long (e.g. reducing greenhouse gases) and short-term actions (e.g. designing reserves) to prevent global warming (Botkin *et al.* 2007). Despite all attempts, we are still losing biodiversity (Rands *et al.* 2010). So, current laws should be improved by trying to include sustainable use and the economic value of biodiversity in ecosystems, as well as monitoring their response once they are implemented, so their efficiency can be assessed (Butchart *et al.* 2010).

Studies show that these major threats are also greatly impacting reptile populations which, are declining at an alarming rate on a global scale (Gibbons *et al.* 2000). In particular aquatic or semi-aquatic freshwater turtle species are particular vulnerable to wetland destruction or pollution, terrestrial habitat degradation and to changes of hydrological patterns due to climate change.

1.2 Testudines: The genus *Mauremys*

Turtles are considered one of the most peculiar groups based on both anatomic and physiological characters. They have one of the highest lifespan among tetrapods, are able to survive a variety of environmental conditions (e.g. severe cold and hypoxia), and some have temperature-dependent sex determination (Gilbert & Corfe 2013). The order Testudines comprises all species of terrestrial, marine, and freshwater turtles, which are *ca.* 331 recognized species (Dijk *et al.* 2012). Testudines are divided into two major monophyletic extant clades – Cryptodira and Pleurodira – which initially were assumed to have diverged *ca.* 210 million years ago (Near *et al.* 2005, 2008; Hugall *et al.* 2007). However, recent studies infer this divergence to have occurred later, *ca.* 170 million years (Chiari *et al.* 2012; Lourenço *et al.* 2012). One of the most important characteristic that differentiates the specimens contained in these clades is how they retract the head. If the individuals retract their neck accordingly to a horizontal plane into the shell they belong to Pleurodira, and by contrast, individuals belonging to Cryptodira retract the neck accordingly to a vertical plane lodging the head between the shoulders girdles (Shaffer 2009).

The Geoemydidae family, formerly known as Bataguridae (Bour & Dubois 1986), is composed by 69 species of freshwater and semi-aquatic turtles and is part of the Testudinoidea superfamily (within Cryptodira) (Dijk *et al.* 2012). The majority of the species in this family can only be found in the Indo-Malayan region (Bour 2008). Phylogenetic relationships between all Geoemydidae genera were studied by Spinks et al. (2004), where the genus *Cuora* and *Mauremys* were revealed as sister groups, and *Mauremys* was paraphyletic with *Chynemis* and *Ocadia*.

The genus *Mauremys* contains nine species (Dijk *et al.* 2012), however, two of them have hybrid origin – *Mauremys pritchardii* and *Mauremys iversoni* (Parham *et al.* 2001). *Mauremys reevesii*, only recent was included in the Mauremys genus and its distributed throughout East Asia (including China, Korean Penisula, Taiwan, and Japan) (Fritz & Havaš 2007; Dijk *et al.* 2012). Regarding the remaining *Mauremys* species, they currently present a patchy distribution. *Mauremys mutica, M. japonica,* and *M. annamensis* are distributed through eastern Palaearctic and *M. leprosa, M. caspica,* and *M. rivulata* are distributed to western Palaearctic (Fritz & Havaš 2007) (Fig 1.1). This type of distribution pattern is usually associated with different *refugia* during glaciations periods which eventually lead to allopatric speciation (Gómez & Lunt 2007; Stewart *et al.* 2010; Hewitt 2011a). However, a study based on cytochrome-*b* (cyt-*b*) revealed enough genetic differentiation within *Mauremys* that lead the authors to

assume the current distribution of the species was the result of numerous radiation events prior to the allegedly Pleistocene extinctions (Barth *et al.* 2004). In 2012, Guillon et al. (2012) attempted to clarify turtles phylogeny through nuclear and mitochondrial DNA of 230 individuals. Regarding Geomydidae, they mostly achieved the same conclusions as Spinks et al. (2004), specifically the inclusion of *Chinemys* and *Ocadia* as *Mauremys* species. Moreover, given the higher number of genes (mtDNA and nuDNA) used they were able to clarify the *Mauremys* phylogeny, recovering *M. mutica* and *M. annamensis* as the basal clades for the remaining *Mauremys* species (Guillon *et al.* 2012), contradicting the previous results that portrayed *M. leprosa* as the most basal taxon of Mauremys + Chinemys + Ocadia (Barth *et al.* 2004). The different results achieved in these two studies may be connected to the different genes used to re-construct the phylogeny. When solely using mtDNA the conflict between gene tree vs. species tree arises (Pamilo & Nei 1988; Avise 1989; Maddison 1997).

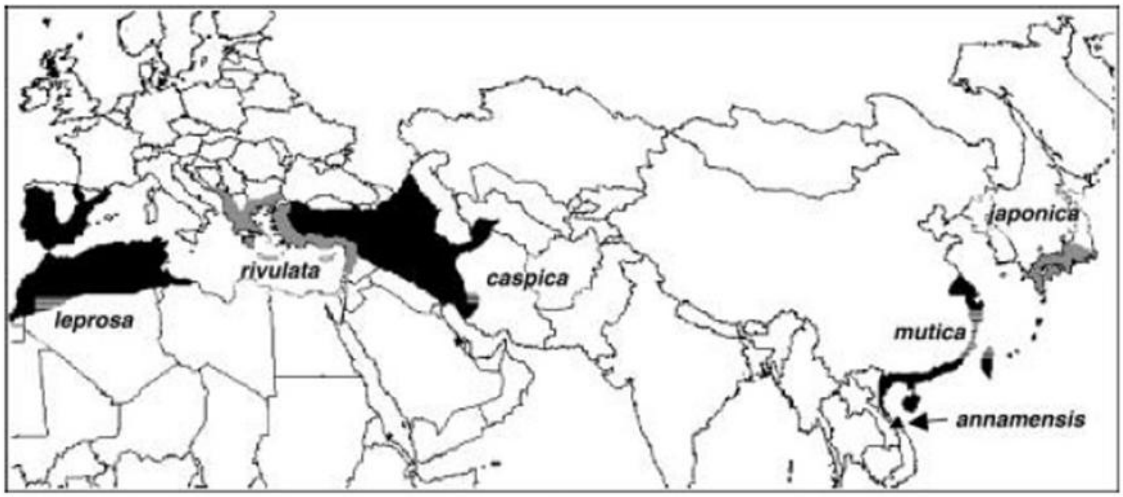


Fig. 1.1: Distribution map of six *Mauremys* species. Withdrawn from (Barth *et al.* 2004).

1.3 *Mauremys leprosa*

The Mediterranean pond turtle, *Mauremys leprosa* (Schweigger, 1812), also known as the stripe-necked terrapin, is a small size freshwater turtle (maximum carapace length can be up to 23 cm (Andreu *et al.* 1998), with a life expectancy of 30 years (Da Silva 2002). Males achieve sexual maturity ca. 7 years old, while in females only happens between 8 and 9 years old (Keller 1997). This species has high plasticity regarding carapace coloration, it can range from olive tones to brownish coloration. Neck and forelimbs in juveniles have orange or yellow strips, losing them with age, and presenting only a green tone when older. It is an omnivorous species, however, their primarily choice are aquatic invertebrates (Ernst & Barbour 1989). It shows sexual dimorphism, with females being bigger than males (when adults), with a flat or convex

plastron (concave in males), and a short tail with the cloaca next to the plastron (long tail in males) (Muñoz & Nicolau 2006). Accordingly to Muñoz and Nicolau (2006) the bigger female size would help maximizing the number of carried eggs. Even under favourable conditions (thermal regulated environment and enhanced diet), and contrary to expected, females do not mature earlier. Instead, the maturation process is delayed so the individual can grow bigger to be able to carry an higher number of eggs (Lovich *et al.* 2010).

1.3.1 *Mauremys leprosa* ssp.

When described for the first time, in 1812 by August F. Schweigger, it was denominated *Emys leprosa* (Bour & Maran 1999), and since then has been evaluated in several taxonomic revisions. *Mauremys leprosa* was split into eight subspecies – *M. l. leprosa; M. l. atlantica; M. l. erhardi; M. l. marokkensis; M. l. saharica; M. l. vanmeerhaeghei; M. l. wernerkaestlei;* and *M. l. zizi* (Schleich 1996; Bour & Maran 1999). This division was based on several morphological characters, such as the type of dorsal ornamentation, the plastron pattern, and the tympanic spot (Schleich 1996; Bour & Maran 1999). Seven of the described subspecies occurred in Morocco: *M. l. atlantica, M. l. erhardii, M. l. marokkensis, M. l. wernerkaestlei* were distributed through patchy areas north of the Atlas Mountains, while *M. l. saharica, M. l. vanmeerhaghei,* and *M. l. zizi* occurred south of the Atlas Mountains. Regarding *M. l. leprosa*, it was thought to only occur in Europe (Schleich 1996; Bour & Maran 1999). However, in 2005, a study based on mitochondrial DNA (cytochrome *b*) revealed that this diversity was overestimated (Fritz *et al.* 2005). The following year, this study was enhanced by adding more samples and since then, only *M. l. leprosa* and *M. l. saharica* are considered as subspecies, with the Atlas Mountains identified as a barrier, not the Strait of Gibraltar (Fritz *et al.* 2006). Therefore, *M. l. leprosa* is described to occur northern of the Atlas Mountains to France, while *M. l. saharica* is described to occur southern of the Atlas Mountains to Tunisia. The Atlas Mountains has been described as a barrier to, at least, two other species - *Agama impalearis* (Brown *et al.* 2002) and *Tarentola* sp. (Rato *et al.* 2012). This comes from suitable environmental conditions during the colder periods near the mountains, which allows species to shelter, although in some cases being the motor force for allopatric divergence (Brown *et al.* 2002; Fritz *et al.* 2006; Rato *et al.* 2012).

1.3.1 Habitat and Distribution

The species is distributed throughout most of the Iberian Peninsula, with some populations in southern France, and it is also present in the Mediterranean Maghreb region (Morocco, Algeria, Tunisia and Libya) (see Fig. 1.2). There are also some old reports of populations in the Aïr Mountains (Niger) and Libya (pink mark in Fig.1.2) (Papenfuss 1969; Busack & Ernst 1980; Schleich 1996); however, those reports are old and there is the possibility of a misidentification with *Pelomedusa subrufa olivacea* (see Fig. 1.3 for comparison). The isolated population in Fderîck (Mauritania), which could be the result of human introductions, (red mark in Fig.1.2), is considered extinct since the 1996 (Schleich 1996). *Mauremys leprosa* can inhabit different freshwater habitats, such as dams, rivers and ponds, (Ernst & Barbour 1989), and it is known to be able to inhabit locations with a high degree of pollution (Keller 1997).

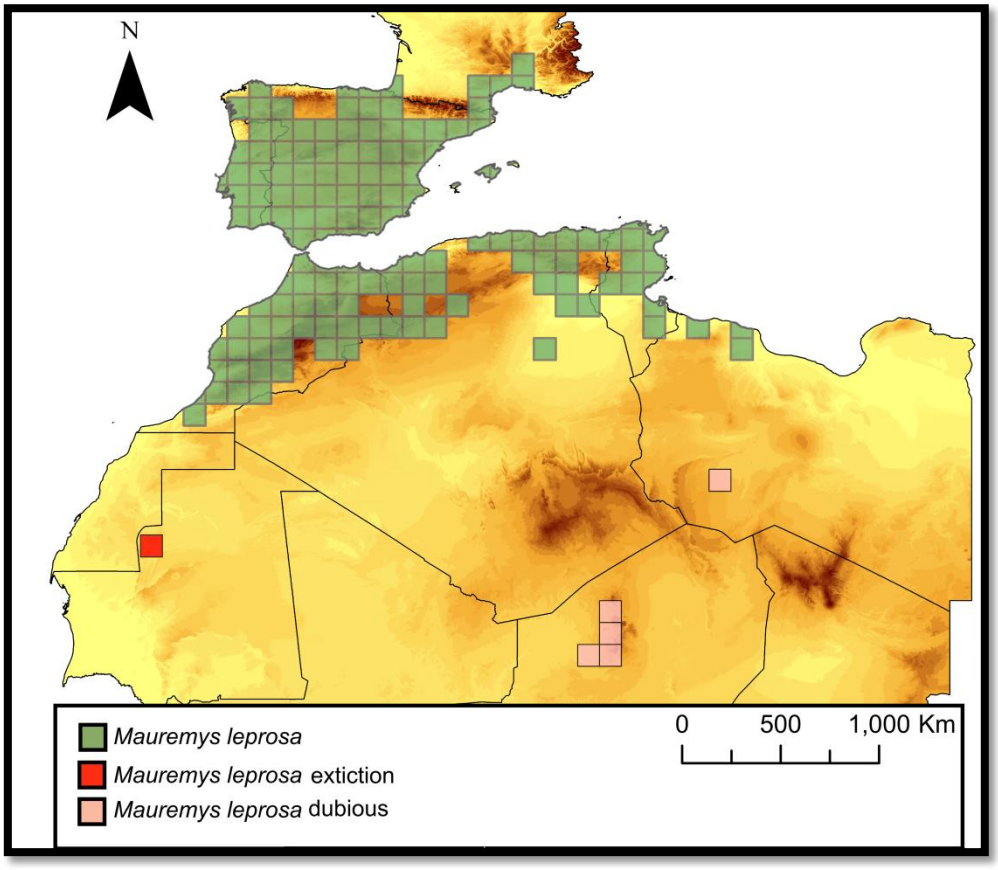


Fig. 1.2: Distribution map of *Mauremys leprosa*. Currently distribution of the species (green), the extinct population in Fderîck (red) and the dubious populations (pink).

Fig. 1.3: Morphological comparison between *Pelomedusa subrufa olivacea* (left picture) and *Mauremys leprosa* (right picture). Pictures of carapaces, plastrons and heads (in *P. s. olivacea* can be seen the parallel retraction of the head characteristic of pleurodirians)

1.3.1 Conservation

In 2012, 58.8% of the listed turtle species were considered as globally threatened according to IUCN (Dijk *et al.* 2012). Most turtles' populations are decreasing in numbers, especially due to anthropogenic effects, such as habitat fragmentation, pet trade, the use of some species in traditional medicine, and as food (Parham & Shi 2001; Spinks & Shaffer 2006; TURTLE CONSERVATION COALITION, 2011; Natusch & Lyons 2012). The conservation of freshwater habitats and water dependent species is particularly threatened by overexploitation, water pollution, destruction and/or degradation of habitat and invasive species (Sala *et al.* 2000). *Mauremys leprosa* is known to be locally abundant at certain parts of its range. Yet, is considered Vulnerable by the IUCN and is listed in the Appendix II of the Berne Convention and in Appendix II and IV of Habitat Directive due to decreasing numbers in several populations (92/43/CEE) (Cox & Temple 2009).

1.4 The Mediterranean basin: a biodiversity hotspot

The Mediterranean Basin has been identified as one of the 25 world biodiversity hotspots, which are defined mainly on the number of endemisms and threats (Myers *et al.* 2000). The amount of diversity that can be found in the Mediterranean basin has its origin in the heterogeneous palaeogeographic history. Tectonic movements in the basin during the Miocene completely rearranged the topology of the area, for instance with the uplift of the Iberian and Moroccan plates connecting the Rif-Betic mountain range, the Mediterranean and the Atlantic got separated inducing the dissecation of the basin (Krijgsman *et al.* 1999; Duggen *et al.* 2003). This event, known as the Messinian

Salinity Crisis, allowed for several *taxa* to cross between the two continents due to the emergence of land bridges (Krijgsman *et al.* 1999; Duggen *et al.* 2003; Hewitt 2011a). The Zanclean flood is thought to be the event responsible for the refilling of the Mediterranean Basin (Garcia-Castellanos *et al.* 2009), and therefore inducing divergence through vicariance between European and North African species (Veith *et al.* 2004; Sousa *et al.* 2012; Velo-Antón *et al.* 2012).

During the Pliocene and the Quaternary, divergence within and/or between species was induced by Milankovitch climatic oscillations, even though they were more frequent and intense during the Pleistocene, inducing several ice ages (Hewitt 2000; Dynesius & Jansson 2000). During these periods, when the temperatures reached the lowest values, the ice sheets spread and covered large geographic areas, forcing some species to retreat ranges into small areas currently known as *refugia*, usually located in the southern European peninsulas (Taberlet *et al.* 1998; Hewitt 1999, 2000). A recent review brought attention to the similar species composition between the Maghreb region and the European peninsulas, which is an evidence of very similar ecological and climatic conditions during the Pleistocene and Pliocene (Husemann *et al.* 2014). When temperatures started to rise, some species expanded to nearby territories in search for suitable habitats, some of the times originating complex phylogeographic patterns (Taberlet *et al.* 1998; Hewitt 1999, 2004). Even though, the Strait of Gibraltar appears to be one of the major barriers to dispersion in the Mediterranean, several *taxa* has crossed it after its re-opening and in different directions (Griswold & Baker 2002; Paulo *et al.* 2002; Carranza *et al.* 2004; Cosson *et al.* 2005; Recuero *et al.* 2007; Kaliontzopoulou *et al.* 2011; Habel *et al.* 2012; Rato *et al.* 2012; Santos *et al.* 2012; Stuckas *et al.* 2014; see Husemann *et al.* 2014 for more references therein).

1.5 Objectives

A northern African origin for *Mauremys leprosa* is currently accepted (Fritz *et al.* 2006), however, fossil records in the Iberian Peninsula can be dated back to the Pliocene and Holocene (Fèlix *et al.* 2006; Soler *et al.* 2012). This discordance implies that climatic oscillations induced a major retraction of this species to the surrounding areas of the Atlas Mountains in Morocco, which in itself acts as a barrier to gene flow. Moreover, one specimen belonging to *M. l. saharica* was found in sympatry with *M. l. leprosa* in northern Morocco, which at the time was attributed to human-mediated translocation (Fritz *et al.* 2006).

For this study we significantly increased the number of samples, in comparison to the previous phylogeographic study by Fritz et al. (2006), and also the number of molecular markers. In here, we will be using two mtDNA fragments and more recent demographic processes will be inferred through microsatellite loci. Thus, the main objectives for this thesis are as follows:

→ Assess how past climatic oscillations and landscape barriers (Strait of Gibraltar and Atlas Mountains in Morocco) have shaped current genetic diversity and structure in *M. leprosa* **Manuscript I (Chapter 3)**;
→ Test and optimize a set of 16 microsatellite markers for *M. leprosa* that were initially developed for two closely related species (*M. caspica* and *M. rivulata*) **Manuscript II (Chapter 4)**;
→ Obtain a more recent and fine-scale genetic pattern across Morocco and the Iberian Peninsula to infer spatial population structure and diversity **Manuscript III (Chapter 5)**.

At the end of this thesis, we expect to fully understand if the Strait of Gibraltar is indeed a permeable barrier, as it is for many other reptiles, and if the gene flow barrier induced by the Atlas Mountains has became more permeable through time. Also, taking into account the recent expansion process in the Iberian Peninsula, we expect a decline of diversity towards north. Regarding the *M. l. saharica* individual found in northern Morocco we expect to be able to discern if it is a human-mediated translocation or a natural colonization of the area indicating the presence of a contact zone within this species.

Chapter 2: Common Methodologies

This section describes common methods used for the remaining chapters of this thesis.

2.1 Study Area and Sampling

Given the species distribution range and the goals proposed, the study area covers the majority of *Mauremys leprosa* distribution. Thus, our sample collection comprises populations from the Iberian Peninsula and part of the Maghreb region. The majority of samples used in this study were provided by Dr. Guillermo Velo-Antón, who collected them throughout the past years, while others were collected by me and other LIFE:*Trachemy*s team members. The final dataset includes a total of 653 samples, including 37 individuals belonging to Recovery Centres. All of the sampled animals were individually marked by making small notches on the carapace following a code based on Ernst et al. (1974) (see Fig. 2.1), in doing so, it made possible to be aware of re-captures during field-work and prevent drawing blood from the same individual twice. Blood was drawn from the jugular vein, occipital venous sinus or subcarapacial vein (see Fig 2.2).

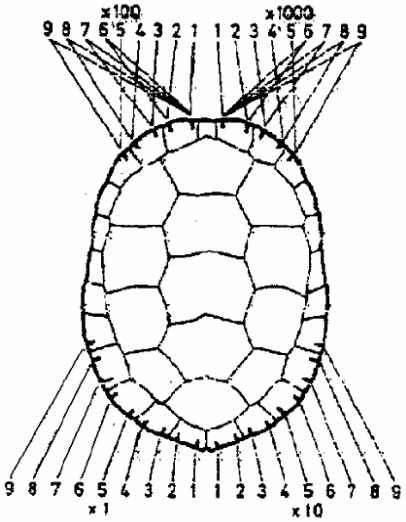


Fig. 2.1: Illustration explaining carapace notching scheme used when sampling for this thesis.

Fig. 2.2: Examples of blood withdrawn from a) jugular vein, b) subcarapacial vein, and c) from occipital venous sinus in freshwater turtles (the three photographs portraiting *Emys orbicularis* specimens).

2.2 DNA extraction

DNA extraction were performed for the majority of the samples (blood and tissue) using EasySpin® (Citomed) following manufactures' protocol, except for a minor modification in which the lyses step was extended from 2 hours to overnight. In the case of *M. rivulata* and *M. caspica* samples, the Quiagen DNeasy Blood & Tissue Kit was used since it can achieve better results for small amounts of sample. For this extraction kit manufactures' protocol was followed. *Mauremys rivulata* and *Mauremys caspica* were used as positive controls, which were kindly provided by Doctor Uwe Fritz.

Electrophoresis of 0.8% agarose gels stained with GelRed™ (Biotium) was used to assess the quality and quantity of DNA extracted, to visualise the gels through UV radiation it was used the BioRad Universal Hood II Quantity One 4.4.0. A roughly estimation was made and, if necessary, DNA was diluted with ultra-pure water. DNA was storage at -20ºC till further use.

All procedures performed after DNA extraction will be detailed described in the corresponding chapter.

2.3 Molecular Markers

In order to obtain a more clear picture on the processes that shaped the evolutionary history of a given *taxa,* the combined use of slower evolving markers (e.g. mtDNA genes) with hyper-variable markers (e.g. microsatellites) allows to address both contemporary and historical events (Zhang & Hewitt 1996, 2003; Selkoe & Toonen 2006).

Mitochondrial DNA has been widely used in taxonomic, phylogenetic, biogeographic and population studies. It has a limited repair capability and lack of histones, which enables the higher mutation rate found when compared with nuclear DNA (Jansen 2000). Moreover, different mitochondrial genes have distinct mutation rates so, we can question different taxonomic levels and assess different time scales (Wan *et al.* 2004). However, all conclusions drawn based on this marker alone may be biased since this marker is maternally inherited (Jansen 2000) and thus, we are only assessing the maternal history (Wan *et al.* 2004).

Microsatellites are tandemly repeated sequences, of up to six bases, which occur throughout the euchromatic part of the genome, usually in non-coding regions. They

are co-dominantly inherited and have very high mutation rates as a result of its mutation mechanism (DNA-replication slippage), turning them into a highly polymorphic marker (Schlötterer 2000; Wan *et al.* 2004). Also, the number of bases composing the repeat motif influences the mutation rate. Microsatellite loci can be under different selection pressures associated with its location, thus the need to perform equilibrium tests analyses to discriminate which are under neutral selection in order to perform further analysis (Wan *et al.* 2004). Allelic dropout and null alleles are two problems that microsatellite loci should be tested for (Wan *et al.* 2004; Selkoe & Toonen 2006). Nonetheless, this type of fast evolving marker have become a helpful tool to infer demographic patterns, due to several characteristic, like the high mutation rate, co-dominant inheritance and the possibility existence of several alleles in a population, which provides them high degree of polymorphism that cannot be found in more stable parts of the genome (Angers & Bernatchez 1998; Schlötterer 2004; Ellegren 2004; Wan *et al.* 2004). Microsatellite markers constitute a valuable genetic tool to infer recent evolutionary histories, demographic processes, current patterns of gene flow and kinship.

Chapter 3: Manuscript I

Pliocene-Pleistocene divergence in North-Western Maghreb and recent demographic expansion across the Iberian Peninsula in the Mediterranean pond turtle (*Mauremys leprosa*)

Joana Veríssimo[1,2],Mohamed Znari[3], Heiko Stuckas[4], Uwe Fritz[4], Paulo Pereira[1,2], José Teixeira[1,5], Marco Arculeo[6], Federico Marrone[6], Francesco Sacco[6]& Guillermo Velo-Antón[1*]

[1]CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto. Instituto de Ciências Agrárias de Vairão. R. Padre Armando Quintas. 4485-661 Vairão Portugal

[2]Departamento de Biologia da Faculdade de Ciências da Universidade do Porto, Rua Campo Alegre, 4169-007 Porto, Portugal

[3]Laboratory ''Biodiversity and Ecosystem Dynamics'', Department of Biology, Faculty of Science–Semlalia, Cadi Ayyad University,Avenue Prince Moulay Abdellah, 40000 Marrakech, Morocco

[4]Museum of Zoology, Senckenberg Dresden, A. B. Meyer Building, 01109 Dresden, Germany

[5]CIMAR-LA/CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Porto, Portugal

[6]Dipartimento di Scienze e Tecnologie Biologiche, Chimiche eFarmaceutiche (STEBICEF), Università di Palermo, Via Archirafi18, 90123 Palermo, Italy

Abstract

The Mediterranean basin harbours a high degree of endemisms and species richness as a result of its complex palaeogeographic history. Events such as the Messinian Salinity Crisis, the formation and/or transformation of orographic features, and the Quaternary climatic oscillations have strongly influenced the distribution and genetic diversification of species occurring in the area.

We aim to evaluate the effect of climatic oscillations and main geographic barriers in the current diversity patterns of the Mediterranean pond turtle, *Mauremys leprosa*. This species occurs widely throughout the Iberian Peninsula and most of the Maghreb region. Two *M. leprosa* subspecies are recognized: *M. l. saharica* (ranging from southern of the Atlas Mountains to Tunisia) and *M. l. leprosa* (northern of the mountains and in Iberian Peninsula). We used 164 individuals from the entire range to amplify two mitochondrial fragments: cyt-*b* and D-loop. Phylogenetic relationships and the most common recent ancestor were assessed under Bayesian inferences. Furthermore, we tested for demographic expansions through three tests of selective

neutrality. Also, genetic distances and nucleotide diversity were interpolated under the kriging method to assess spatially the genetic structure and variability patterns. We successfully identify three sublineages for *M. l. leprosa* and four to *M. l. saharica*. The lack of genetic diversity and structure of the species in the Iberian Peninsula points for an African origin, despite several fossil records dated from the Pliocene in Europe. As in for several other *taxa*, the Strait of Gibraltar acted as a crossing point between the two continents, even after its re-opening. In Morocco, several individuals of *M. l. saharica* were found in the Riff and Middle Atlas, which in conjunction with higher nucleotide diversity point for a secondary contact zone.

Keywords: phylogeography, mitochondrial DNA, *Mauremys leprosa*, secondary contact zone.

3.1 Introduction

The Mediterranean basin harbors a high degree of species richness and endemism (Myers *et al.* 2000), which has been mostly associated to a combination of geologic and climatic events. The palaeogeographic history of the western Mediterranean occurred during the Miocene, such as the tectonic movements that originated the Mediterranean islands (Balearic Islands, Sardinia and Corsica) and the split of the Rif-Betic mountain range, can be matched with diversification events in different taxonomic groups (e.g. Martínez-Solano *et al.* 2004; Magri *et al.* 2007; Bidegaray-Batista & Arnedo 2011; Miraldo *et al.* 2011). Furthermore, at the end of this epoch, the Strait of Gibraltar closed leading to the desiccation of the Mediterranean basin (Krijgsman *et al.* 1999; Duggen *et al.* 2003), which produced land bridges connecting the European and African continents facilitating migration through these corridors for terrestrial organisms that used them to expand their ranges. The refilling of the Mediterranean sea *ca*. 5.3 Ma, induced again divergence between European and North African *taxa* (Veith *et al.* 2004; Sousa *et al.* 2012; Velo-Antón *et al.* 2012). Then, diversification events continued during the Pliocene-Pleistocene when the effects of the Milankovitch climatic oscillations became more frequent and intense, especially during the Pleistocene (Hewitt 2000; Dynesius & Jansson 2000). Glacial periods forced Mediterranean species to move southwards in search for suitable habitats where to shelter until climate amelioration, usually at the southern European peninsulas (Taberlet *et al.* 1998; Hewitt 1999, 2000) and North Africa (Husemann *et al.* 2014). Then, during subsequent range expansions, species could encounter barriers to dispersal hampering species to colonize further areas (Taberlet *et al.* 1998; Hewitt 1999, 2000). The combination of palaeogeographic and climatic events has rendered distinct

phylogeographic patterns for Mediterranean species whose response to the above events could differ due to their ecological constrains, such as their dispersal abilities and ectothermal physiology.

Ectothermic species, which are extremely dependent on climate induced temperatures to regulate body temperature and perform normal physiological functions (Huey & Kingsolver 1993), are highly prompt to suffer retraction/expansion events during climatic oscillations. Low temperatures like those experienced during the glacial periods have a high impact on a species' thermoregulation processes, easily inducing migration events perhaps contracting the species range to microrefugia, where it was possible to experience slightly warmer temperatures. The Mediterranean pond turtle, *Mauremys leprosa* (Schweigger, 1812), is one of most wide and abundant reptiles in the Iberian Peninsula (with a few and scattered populations in the south-western France) and Northern Maghreb (from Western Morocco to Tunisia). The species occurrence in the Iberian Peninsula dates back from Pliocene and Holocene based on fossil records (Fèlix *et al.* 2006; Soler *et al.* 2012). Altogether, make it a good model to understand the effects of glaciations and consequent range expansions and/or contractions in the Western Mediterranean basin. Two main lineages have been identified: *Mauremys leprosa leprosa* inhabits the Iberian Peninsula and northern Morocco, and *M. l. saharica* occurs in southern Morocco, easternmost Algeria and Tunisia (Fritz *et al.* 2006).

In this work we aim to investigate how past climatic oscillations and landscape barriers (Strait of Gibraltar and Atlas Mountains in Morocco) have shaped current genetic diversity and structure in *M. leprosa*. By increasing the available genetic information (cytochrome *b* and control region mitochondrial fragments) and expanding the sample size and studied localities, we infer phylogenetic relationships and spatially interpolate the genetic diversity and divergence within *M. leprosa* to: 1) identify the geographic origin of diversification within the species; 2) determine the genetic structure within each subspecies and estimate the origin of major lineages; 3) evaluate the effect of climatic oscillations and main geographic barriers, the Atlas Mountains in Morocco and the Strait of Gibraltar, in shaping current diversity patterns of *M. leprosa*; and 4) identify potential contact zones of both subspecies as a result of recent population expansions.

3.2 Material and Methods

3.2.1 Sampling and sequencing

Blood or tissue (tail tips) samples of 163 *Mauremys leprosa* were collected across the Iberian Peninsula and the Maghreb region and preserved in absolute ethanol (Figure 1, and Supplementary Material Table S. M. 1). Genomic DNA was extracted with a commercial kit (Easyspin), following manufacturer's protocol, and extending the lyses period to enhance the extraction.

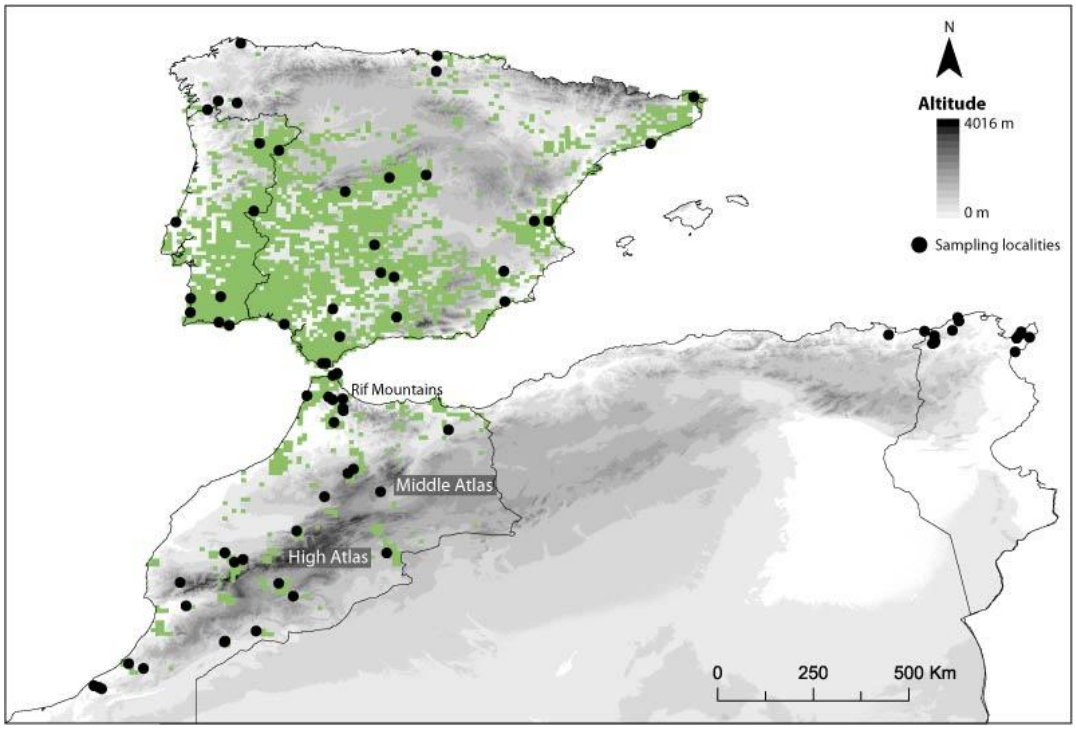


Fig. 3.1: Present distribution of *Mauremys leprosa* in the Iberian Peninsula and Morocco. Sampling locations are marked as black circles. Main mountain chains in Morocco are identified on the map.

Two mitochondrial fragments were targeted in this study: the cytochrome b (cytb) and the control region (D-loop). The former gene was selected in order to increase the available sequences produced in previous works (see Fritz et al. 2006), and to better unveil the spatial distribution of genetic diversity within *M. leprosa*, which yielded in a total of 163 sequences (Table S. M. 1). These samples were also amplified with D-loop to increase the genetic information that is needed to fully resolve the phylogenetic relationships within *M. leprosa*. Cyt-b was amplified using the primers mt-a-neu (Lenk & Wink 1997) and H-15909 (Lenk *et al.* 1999). D-loop was amplified with specific primers designed with OLIGOEXPLORER 1.2 (http://www.genelink.com/tools/gl-oe.asp): MauMut_tThr.for (forward, 5'- ACT CTA GTA GCT TAA CCC AT-3') and MauMut_Dloop_2.rev (reverse, 5'- TCA GTT TAG TTG CTC TCG GA-3'). PCR reactions were conducted in a final volume of 10 µL from which 5 µL corresponded to

MyTaq™ Mix (Bioline), 0.4μM of each primer, 3.2μL of ultra-pure water, and 1μL of DNA. PCRs were carried out on a BioRad T100 Thermal Cycler with the following procedure: initial denaturation at 95 ºC for 10 min (minutes); 10 cycles at 95 ºC for 30 s (seconds), 55 ºC (57 ºC for D-loop) decreasing 0.5 ºC per cycle for 20 s; 72 ºC for 1 min; 30 cycles at 95 ºC for 30 s, 50 ºC for 20 s; 72 ºC for 1 min; and a final elongation step at 72 ºC for 10 min. Both reactions were cleaned for removal of non-used primers and nucleotides with ExoSap (USB® ExoSAP-IT® PCR Product Cleanup, Affymetrix) following manufactures instructions. Four independent sequencing reactions (one for each primer) were performed on a BioRad T100 Thermal Cycler with BigDye® Terminator v3.1 Cycle Sequencing Kits (AB Applied Biosystems) following manufactures protocol. Finally, the four strands were sequenced on an ABI 3130xl genetic analyzer (Applied Biosystems, Foster City, Ca, USA). All the obtained chromatograms were verified, aligned, and corrected by eye using Geneious Pro v4.8.5 (http://www.geneious.com/). MUSCLE algorithm was used for the alignments that were later manually checked.

3.2.2 Phylogenetic analyses

Sequences were collapsed into haplotypes using DnaSP v5.10 (Librado & Rozas 2009) and phylogenetic relationships were assessed using a Bayesian inference (BI) approach. jMODELTEST v.2.1.4 (Darriba *et al.* 2012) was used to test for the best fitting model of nucleotide substitution for our dataset, under Akaike information criteria correction (AICc; TIM+I+G for both markers). The dataset was partitioned by gene to be run under the corresponding evolutionary model. BIs were conducted using BEAST v1.7.5 (Drummond *et al.* 2012). Markov Chain Monte Carlo (MCMC) analyses were run in three independent runs of 10 million generations with four chains, with a sampling frequency of 100 generations, and discarding 25% trees as burn-in. Parameter convergence was verified by examining the effective sample sizes (ESSs) using TRACER v1.6 (all parameter values of ESS were above 300), and used the remaining trees to obtain the subsequent maximum clade credibility summary tree with posterior probabilities for each node using TREEANNOTATOR. A substitution rate of 0.00626 substitutions/site/million years suggested for the mitochondrial DNA in turtles (Lourenço *et al.* 2013) was used to estimate time to most recent common ancestor (TMRCA) of supported mtDNA lineages, with a standard deviation of 0.0002. A lognormal relaxed clock and a coalescence constant size model were used as tree priors.

3.2.3 Genetic diversity and demographic analyses

DnaSP v5.10 was used to assess the number of segregating sites (S), and the nucleotide (π) and haplotype diversity (Hd). Three tests of selective neutrality (Tajima's D, R2, and Fu's FS) were performed in DnaSP v5.10 to infer signatures of demographic expansion in each sublineage, using 10,000 bootstrap replicates.

Two haplotype networks were constructed to visualize haplotypes relationships within *M. leprosa*: a haplotype network using statistical parsimony implemented in TCS v1.21 (Clement *et al.* 2000); and a neighbour-net network based on uncorrected patristic distances and bootstrap analysis with 1,000 replicates, using SPLITSTREE v4.6 (Huson & Bryant 2006). Uncorrected p-distances were calculated for the concatenated dataset in MEGA v5.1 (Tamura *et al.* 2011) to estimate genetic divergence between main lineages and sublineages.

3.2.4 Spatial analyses

To visualize geographical patterns of genetic structure and variability, both genetic distances (uncorrected p-distances) and nucleotide diversity (π) were spatially interpolated following a kriging interpolation method. Samples from Tunisia and Algeria were excluded in order to avoid the sampling gap across this region that would induce artefacts on the analysis. To identify potential barriers to gene flow, pairwise uncorrected p-distances (Nei & Kumar 2000) were calculated between each pair of sequences (see Supplementary Material R code 3 for methodology). From the resulting pairwise uncorrected p-distances matrix, we treated each column as a different variable, as each column corresponds to the distances from one point to all the others, which we interpolated using the kriging interpolation method (Oliver & Webster 1990). A Principal Components Analysis was then used to summarize the results, using the Principal Components tool found in the "Spatial Analyst Tools" extension of GIS ArcMap 9.3 (ESRI 2008). In order to identify areas that present sharp changes in the values of uncorrected p-distances, which correspond to potential barriers to gene flow, we have used the function slope, located in the "Spatial Analyst" extension of GIS ArcMap 9.3 (ESRI 2008). To identify the spatial distribution of genetic diversity, nucleotide diversity values were calculated by pooling samples contained in a buffer with a radius of 0.449 decimal degrees (approximately 50km), which represent the potential genetic diversity of the original point (see Supplementary Material R code 1 and 2 for methodology). Nucleotide diversity values were then interpolated by generating a continuous surface with a kriging interpolation method (Oliver & Webster

1990), implemented in the “Geostatistical Analyst” extension of GIS ArcMap 9.3. The resulting raster was then reclassified into five classes, using Natural Jenks as the division criteria.

3.3 Results

3.3.1 Phylogenetic analyses

We obtained 75 unique haplotypes from 1769 bp concatenated mtDNA dataset (933bp of cytb and 862 of D-loop) in 163 samples. Bayesian inferences show a resolved phylogeny with two major clades (BPP > 0.95), which corresponds to the known subspecies, *Mauremys leprosa leprosa* and *Mauremys leprosa saharica* (Fig. 3.2). *M. l. leprosa* haplotypes are distributed in North Africa (north of the Atlas Mountains) and the Iberian Peninsula, while *M. l. saharica* haplotypes occur in Morocco (north and south of the Atlas Mountains), easternmost Algeria and Tunisia.

We also identified three and four well supported sublineages in *M. l. leprosa* and *M. l. saharica* respectively (BPP > 0.95; Fig. 3.2). Two sublineages (A1 and A2) of *M. l. leprosa* occur in southwestern and central Morocco, respectively, and north of the Atlas Mountains, while the third sublineage (A3) is distributed throughout the Iberian Peninsula and the Rif, with two haplotypes (A3-9 and A3-15) occurring at both sides of the strait of Gibraltar (see Table SM1). For *M. l. saharica*, two sublineages (B1 and B2) are admixed across the south of the Atlas Mountains, a third sublineage (B3) ranges from the Rif and Middle Atlas to Tunisia and a fourth sublineage (B4) occurs in the Rif. Average sequence divergence (uncorrected p-distance) between *M. l. leprosa* and *M. l. saharica* is 1.5% (range: 0.1-2.2%), and high genetic divergence values are also found between sublineages (Table 3.1).

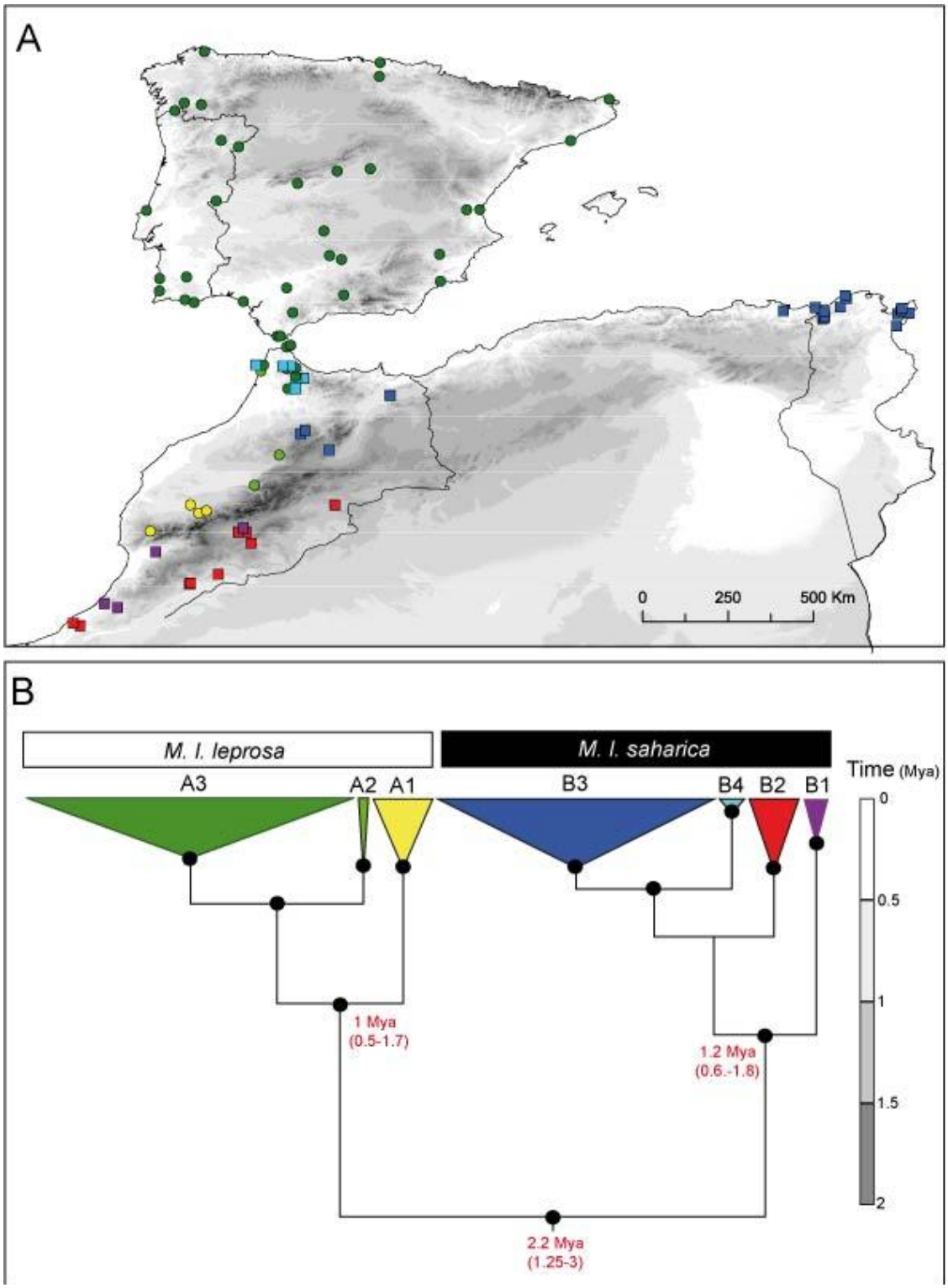


Fig. 3.2: A: Spatial distribution of the sublineages based on the phylogenetic tree. Circles correspond to *M. l. leprosa*, while squares correspond to *M. l. saharica*. **B:** Bayesian consensus phylogram based on mtDNA data (D-loop and Cyt-*b*) for *M. leprosa.* Times to most recent common ancestor for the split between and within lineages are written in red (TMRCA; 95% highest posterior density interval).Both symbol shapes and colours are concordant with Figure 3.3.

Table 3.1: Genetic distances between sublineages. Bellow the diagonal mean uncorrected p-distances and standard deviations in percentages. In bold on the diagonal the mean uncorrected p-distance within each sublineage.

Group	A1	A2	A3	B1	B2	B3	B4
A1	**0.002**						
A2	0.9 (±0.2)	**0.003**					
A3	1 (±0.2)	0.5 (±0.2)	**0.001**				
B1	2 (±0.3)	2.3 (±0.3)	2.1 (±0.3)	**0.002**			
B2	1.9 (±0.3)	2 (±0.3)	1.9 (±0.3)	1.4 (±0.3)	**0.001**		
B3	1.7 (±0.3)	1.9 (±0.3)	1.8 (±0.3)	1.1 (±0.3)	0.9 (±0.2)	**0.001**	
B4	1.5 (±0.3)	1.8 (±0.3)	1.7 (±0.3)	1 (±0.2)	0.7 (±0.2)	0.3 (±0.1)	**0**

Assessment of divergence times using BEAST estimates the time to the MRCA for *M. l. leprosa* and *M. l. saharica* at the upper Pliocene (mean = 2.2; 95% HPD = 1.25-3 Myr), and the time to the MRCA for each subspecies at the Middle Pleistocene (*M. l. leprosa*, mean = 1; 95% HPD = 0.5-1.7 Myr; *M. l. saharica*, mean = 1.13; 95% HPD = 0.6-1.8 Myr) (Fig. 3.2). Sublineages within each subspecies diverged during the Upper and Middle Pleistocene (Fig. 3.2). We should bear in mind that these dates represent the coalescence time of the different mtDNA haplotypes, and thus the above lineages could have diverged at a much more recent time.

Table 3.2: Summary table of all genetic diversity and demographic parameters measured for the different haplogroups of *Mauremys leprosa*. N, sample size; S, polymorphic sites; π, nucleotide diversity; Hn, number of haplotypes; Hd, haplotype diversity; R_2, Ramos-Osins and Rosas; D, Tajima´s D; Fs, Fu´ s Fs. Significant results for D and Fs shown in bold (P < 0.01). When unable to calculate demographic measures due to low sampling size represented as NA (Not Available).

Group	N	π	S	Hn	Hd	R2	D	FS
A1	13	0.0023	13	11	0.92	0.13	-0.13	**-3.81**
A2	3	0.0030	8	3	1	NA	NA	NA
A3	67	0.001	29	21	0.75	**0.03**	**-21.44**	**-15.48**
A	83	0.0036	52	35	0.83	0.09	-11.36	**-9.77**
B1	6	0.0022	9	6	1	0.14	-0.11	**-2.69**
B2	11	0.0016	9	6	0.69	0.15	-0.28	1.34
B3	57	0.0007	24	19	0.46	**0.03**	**-23.14**	**-12.28**
B4	6	0.0003	1	2	0.53	0.26	0.85	0.62
B	80	0.0043	58	33	0.71	0.06	-12.27	-4.80
ALL	163	0.0115	106	68	0.87			

3.3.2 Genetic diversity and demographic analyses

Both subspecies showed similar values of genetic diversity, although nucleotide diversity (π) was higher in *Mauremys leprosa saharica* (Table 3.2). Within *M. l. saharica*, the two sublineages distributed south of the Atlas Mountains (B1 and B2) show higher genetic diversity (π and Hd) than the ones distributed across the Middle Atlas, Rif and northeastern Maghreb (B3 and B4). Within *M. l. leprosa*, the highest genetic diversity is found in the two sublineages on the north slope of the Atlas Mountains (A1 and A2), with a much less variation in the sublineage distributed in both continents (A3). When this sublineage A3 was divided into two groups (North Africa and Iberian Peninsula) we found higher genetic diversity in North Africa (π = 0.00099) than in the Iberian Peninsula (π = 0.00083), even though the higher number of samples analyzed from the Iberian Peninsula. The most widely distributed sublineages, A3 and B3, showed negative and significant values of Tajima's; R2 and Fu's statistics (Table 3.2).

Parsimony analyses in TCS yielded independent haplotype networks for *M. l. leprosa* and *M. l. saharica*, which can be manually connected with 26 mutational positions (Fig. 3.3). Within each subspecies, all sublineages are also well separated from each other by 7-15 mutational positions, except for the sublineage B4 that differs from B3 in only 3 positions. Sublineages A3 and B3 clearly show a star-like network which are characteristic from demographic expansion scenarios, with A3-1 widely spread across the Iberian Peninsula and B3-1 widely distributed in Tunisia. Splitstree network shows identical relations to the above described for all lineages and sublineages.

3.3.3 Spatial analyses

The kriging interpolation produced a continuous surface of nucleotide diversity that clearly shows the highest genetic diversity in North Africa, particularly in the Rif Mountains (π ranges from ~0.0056 to ~0.0086; Fig. 4). We observe a latitudinal pattern of genetic diversity loss across the Iberian Peninsula, where the southern region presents moderate levels of genetic diversity that is reduced in the Northwest and Central Iberian Peninsula (π ranges from ~0.0001 to ~0.0004).

The interpolated genetic distances surface detected an abrupt change in North Africa, ranging from the High to the Middle Atlas and the Rif mountains (Fig. 3.5). As for the genetic diversity, the Iberian Peninsula shows a homogeneous surface reflecting the lack of genetic divergence across these populations (Fig. 3.5).

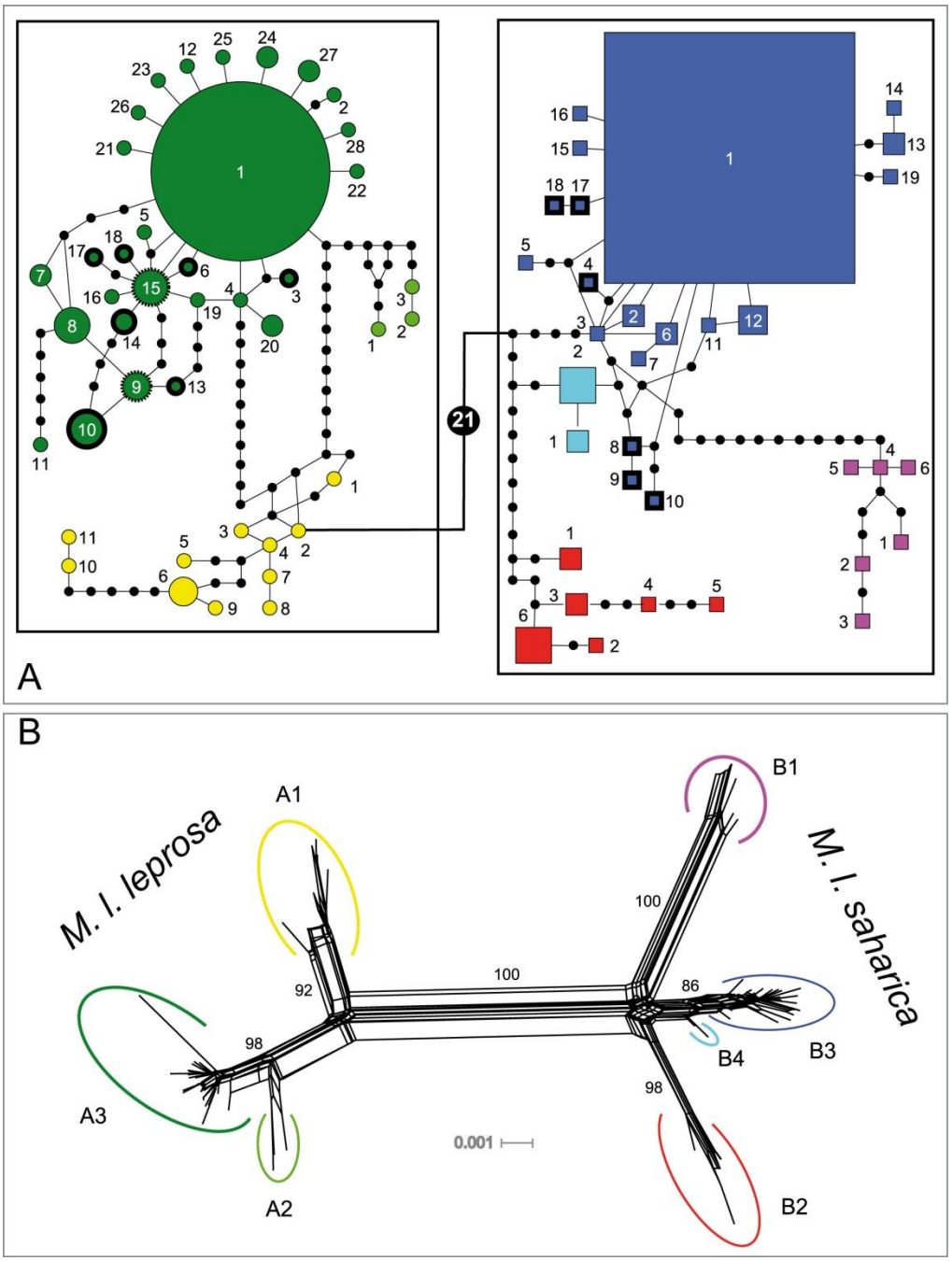


Fig. 3.3: A: Haplotype networks inferred by TCS under the 95% criterion. Circles correspond to *M. l. leprosa* and squares to *M. l. saharica*. The size of each haplotype symbol is proportional to its frequency and lines represent mutational steps separating observed haplotypes. Bold haplotype outline corresponds to haplotypes only found in Morocco, while dashed outline correspond to haplotypes found in Morocco and Iberian Peninsula (for sublineages A3, in green, and B3, in blue). **B:** Mitochondrial neighbour-net networking inferred by SplitsTree. Scale bar represents 1% sequence divergence while numbers correspond to bootstrap values. Both symbol shapes and colours are concordant with Fig. 3.2.

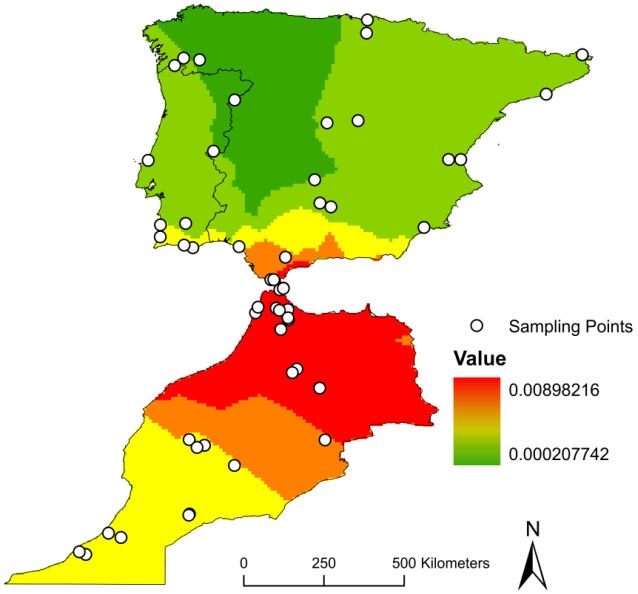


Fig. 3.4: Geographic genetic differentiation in *M. leprosa*. First axis of the spatial principal component analysis applied to the interpolations of the uncorrected p-distances matrix. The white circles correspond to samples used for the interpolation.

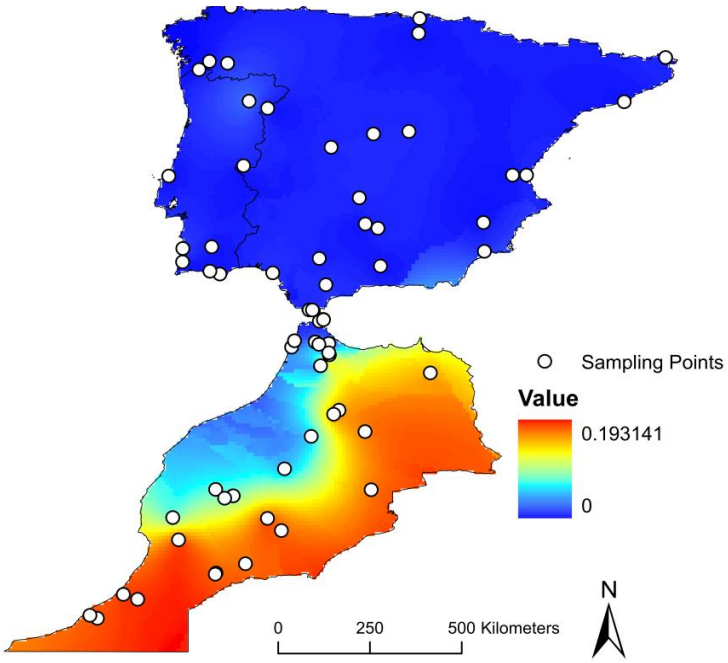


Fig. 3.5: Geographic genetic variation in *M. leprosa*. Surface of interpolated genetic diversity based on nucleotide diversity, resulting raster file was reclassified into 5 different classes using natural breaks. The white circles correspond to samples used for the interpolation, darker green colors correspond to areas with the lowest nucleotide diversity while darker red areas are assigned to areas with high nucleotide diversity.

3.4 Discussion

3.4.1 North African diversification

A wide distribution gap in the central Mediterranean region separates *M. leprosa* from its closest relatives: two western Paleartic species, *M. caspica* and *M. rivulata*, which are distributed in the south-eastern Balkans, the Near and the Middle East (Fritz *et al.* 2008). This distribution suggests and old and allopatric divergence between *M. leprosa* and their relatives. A previous phylogeographic study by Fritz et al. (2006), identified two distinct phylogenetic clades of *M. leprosa* in Morocco (*M. l. leprosa* and *M. l. saharica*), suggesting a North African origin for the species. However, the scarce spatial distribution of the samples and the use of one single mtDNA marker precluded a thorough evaluation of the lineage diversification in *M. leprosa*. Our study approximately doubled both the number of samples and the number of nucleotide positions that were used in a previous study, which allowed a better support on the monophyly of both subspecies. Our study also shows a deep genetic structuration and high levels of genetic diversity within each subspecies in Morocco. The fact that all sublineages found in *M. l. leprosa* and *M. l. saharica* occur in Morocco, as well as the highest genetic diversity present in this country, reinforce the hypothesis of a north-western Maghreb origin (Fritz *et al.* 2006).

When comparing the genetic divergence of *M. leprosa* with other chelonians populations inhabiting North Maghreb, which shares low evolutionary rate and long generation times (Avise *et al.* 1992), is higher than in *Emys orbicularis occidentalis* (0.66 %, Stuckas et al. 2014) and identical to *Testudo graeca* (up to 2.3 % Fritz et al. 2009). Deep levels of genetic structuration and high genetic diversity have also been observed in other *taxa* (e.g. Cosson et al. 2005; Habel et al. 2011; Barata et al. 2012; Sousa et al. 2012; Rato et al. 2012; Velo-Antón et al. 2012; Coelho et al. 2014), for which the combination of topographic heterogeneity and climatic factors are thought to be the main drivers promoting speciation events and genetic differentiation during long periods of allopatric populations (Husemann *et al.* 2014).

3.4.2 The role of climate and geographic barriers

The high elevation (up to 4167 m) and large extent (2500 km) of the Atlas Mountains in Morocco has served as a major barrier to dispersal for North African species. This was previously evaluated in *M. leprosa* by Fritz et al. (2006) and concluded that these mountains are a strong barrier to dispersal in *M. leprosa*, separating *M. l. leprosa* and

M. l. saharica to northern and southern regions of the mountain chain, respectively. Likewise, the Atlas Mountains was described as a barrier to other reptiles, for instance it separates the two lineages of *Agama impalearis* (Brown *et al.* 2002), and its formation is thought to have been the vicariant event responsible for the *Tarentola* diversification (Rato *et al.* 2012). Our study allowed us to observe that these mountains clearly acted as a barrier since the old split between the two major lineages (late Pliocene), but could not prevent individuals' dispersal, and subsequent contact, along the eastern part of these mountains. The most widespread sublineage of *M. l. saharica* (B3) is distributed from Moroccan Rif and Middle Atlas to Tunisia, showing a pattern of recent population expansion (in the last two hundred thousand years). Moroccan populations of this lineage show higher diversity than Tunisian populations, suggesting a west-east expansion from a glacial refugium located between Middle and High Atlas, although further work is needed in Algeria to fill in the sampling gap on this region and to accurately investigate the population expansion and directionality within this lineage. Further genetic sub-structuration is present in the north (sublineages A1, A2, A3, B3, B4) and south of the High-Middle Atlas (sublineages B1, B2, B3). The spatial distribution of some of these sublineages are partially similar to the ones described in other *taxa* (e.g. *Buthus*, Sousa et al. 2012; *Tarentola*, Rato et al. 2012), which could emerged during the Pleistocene through population isolation when river valleys temporally transgressed by the sea, or due to other orographic structures at both sides of the Altas Mountains.

The Strait of Gibraltar was previously considered one of the major barriers to disperse for the Mediterranean *taxa*, however, for the past decades, several studies revealed that many species managed to cross it, regardless of the expansion direction (see Husemann et al. 2014 and references therein). Our study also supports previous findings in *M. leprosa* (Fritz *et al.* 2006), suggesting the strait of Gibraltar as a permeable barrier to dispersal and gene flow between African and Iberian populations. The lack of genetic differentiation between Iberian and north African populations and the recent origin of linage A3 suggest a very recent re-colonization in Europe (upper Pleistocene or post-glacial period). Moreover, a series of evidences suggest a recent expansion throughout Iberia: the genetic homogeneity and low genetic diversity in Iberian populations, as well as the network position of highly frequent haplotypes such as A3-1, supports the hypothesis of a rapid expansion throughout the Iberian Peninsula. This recent colonization in Europe might be associated to the drops of the sea water level along the Strait of Gibraltar (the separation between continents could reach only 5km; Brandt et al. 1996; Zazo 1999), following a stepping-stone model of

colonization. *Mauremys leprosa* is known to be able to survive in brackish waters (Keller & Busack 2001), facilitating the crossing of the Strait, and a recent study on *Mauremys caspica* has shown the ability of this species for transoceanic dispersal on much higher distances (Vamberger *et al.* 2014). Other *taxa* have crossed the Strait of Gibraltar since the re-opening till more recent times, evidencing a migration pathway in this region (e.g. scorpions, Habel et al. 2012; birds, Griswold & Baker 2002; amphibians, Recuero et al. 2007; reptiles, Paulo et al. 2002; Carranza et al. 2004, 2006; Kaliontzopoulou et al. 2011; Rato et al. 2012; Santos et al. 2012; Stuckas et al. 2014; mammals, Cosson et al. 2005; see Husemann et al. 2014 for more references therein). Yet, since the Iberian haplotype (A3-15) occurs in Ceuta (northernmost African population), and that a second haplotype (A3-9), shared also between continents, is only present in Tetouan (North Morocco) and Cádiz (South Spain), we could not rule out the hypothesis of human mediated transportation of the species between both sides of the strait, as it could happened for chamaleons (Paulo *et al.* 2002) and hylids (Recuero *et al.* 2007), although dating very recent divergence events is a difficult challenge that complicate to distinguish late Pleistocene colonizations from anthropogenic introductions (e.g. Graciá et al. 2013).

Several fossil records of *M. leprosa* have been found in the Iberia Peninsula, dated from Pliocene (Soler *et al.* 2012) and Holocene (Fèlix *et al.* 2006). Therefore, the most parsimonious scenario to explain the mismatch of old fossil records and extremely low genetic diversity and shallow population differentiation would be an ancient (Pliocene or earlier) invasion of Europe, followed by a massive extinction of the species in this region due to Pleistocene climatic oscillations, and a later re-colonization from North Africa and rapid population expansion throughout Iberia. Interestingly, the co-distributed terrapin, *Emys orbicularis*, in the Iberian Peninsula and northern Maghreb (*E. o. occidentalis*) shows a similar pattern of re-colonization from North Africa (Stuckas *et al.* 2014), and a rapid population expansion throughout the Iberian Peninsula (Velo-Antón *et al.* 2008) that likely caused present carapace scute anomalies through a series of bottlenecks effects (Velo-Antón *et al.* 2011a). On the contrary, the eastern Algerian and Tunisian *Emys* populations (identified as a new subspecies, see Stuckas et al. 2014) are genetically different from the Moroccan *Emys* populations, in contrast to the pattern here observed for *M. leprosa*. Thus, it appears that climatic conditions occurred in the Iberian Peninsula during glacial phases were too harsh for both terrapins, which likely caused large extinctions of *Emys* and *Mauremys* populations. However, the paleaocological similarities of southern Europe and North Africa would point to other unidentified factors that would better explain the vanishing of these

thermophilic species in the Iberian Peninsula, where many other reptiles remained in suitable *refugia* during Pleistocene climatic fluctuations.

3.4.3 Contact zones within *Mauremys leprosa*

The presence of a single specimen of *M. l. saharica* in north Morocco (Fritz *et al.* 2006) led to these authors to suggest a likely anthropogenic introduction as reptile trade and capture collection has commonly occurred in Morocco. However, in this study we show that the area with highest genetic diversity for *M. leprosa* is located in the Rif and Middle Atlas as a result of the presence of several sublineages of both subspecies (A2 and A3, *M. l. leprosa*; B3 and B4, *M. l. saharica*). The high number of *M. l. saharica* samples widely distributed in this region point to a natural colonization from a potential refugium in the Rif and the Middle Atlas, leading to a secondary contact zone with *M. l. leprosa* populations that also remained in north Morocco during past climatic oscillations. We found both species occurring in syntopy in several Moroccan localities of the Rif (Tazia, Tetouan, Fifi, Zoumi) and Middle Atlas (Sidi Mimoun), allowing us to identify populations that should be further evaluated with nuclear markers to study whether potential natural hybridization occurs between sublineages or if they evolved isolation mechanisms (prezygotic or postzygotic barriers) that still maintain the distinctiveness generated through past allopatric isolation during Pleistocene climatic oscillations.

Acknowledgments

We thank M. Casal, C. Ayres, A. Cordero, A- Acuña, M, Franch, G. Alarcos, A. Álvarez, V. Sancho, X. Rubio, J. Pleguezuelos, L. García-Cardenete, I. Martínez-Solano, E. Recuero, A. Egea, B. El Marnisi and BIOSESERTS members for helping during field-work or providing samples, and S. Lopes for technical support in the lab. This work was funded by Instituto de Estudios Ceutíes in 2012 and by LIFE-Trachemys (LIFE09/NAT/ES/000529). GVA and JT were supported by post-doctoral fellowships from Fundação para a Ciência e Tecnologia (FCT, Portugal: SFRH/BPD/74834/2010 and SFRH/BPD/27173/2006).

Chapter 4: Manuscript II

This chapter corresponds to manuscript published in Amphibia-Reptilia, in June, 2013.

Cross-amplification of microsatellite loci for the Mediterranean stripe-necked terrapin (*Mauremys leprosa*)

Joana Veríssimo[1,2], Paulo Pereira[1,2], Susana Lopes[1], José Teixeira[1], Uwe Fritz[3], Guillermo Velo-Antón[1]

1 - CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, R. Padre Armando Quintas, 4485-661 Vairão, Portugal

2 - Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal

3 - Museum of Zoology (Museum fürTierkunde), Senckenberg Dresden, A.B. Meyer Building, 01109 Dresden, Germany

Abstract

To accurately infer population structuring and manage species, it is advised to combine data obtained from mitochondrial DNA (mtDNA) with data from fast evolving markers such as microsatellites. To date, the evolutionary history of a threatened Mediterranean species, *Mauremys leprosa*, was inferred based solely on mtDNA data, which may lead to an incomplete, or partially explained, population structuring. We tested the cross-amplification of 16 microsatellite loci in 190individuals of *M. leprosa* from six Iberian and two African populations. We obtained a successful set of 11 polymorphic loci with 2-18 alleles and observed heterozygosity ranging from 0.007-0.783. This panel of loci can be used for future research in *M. leprosa*, such as population structuring, analysis of gene flow in secondary contact zones, paternity analyses, changes in phenotypic traits and to assemble a comprehensive genetic dataset (mtDNA and nuDNA) that will allow the geographic assignment of individuals of unknown origin. These tools will help managing *M. leprosa* populations throughout the species' range.

***Keywords*:** conservation genetics, cross-amplification, *Mauremys leprosa*, microsatellites, phylogeography, population genetics.

4.1 Introduction

Inferring genetic structuring patterns based on molecular analysis of mitochondrial DNA usually produces an incomplete picture of the species' biogeographic processes

due to its maternal heritability. These patterns are even more incomplete when studying species, such as turtles, with lower mtDNA mutation rate than other vertebrates, which is partially explained by their long life-span and average long generation time (Avise *et al.* 1992; Bromham 2002; Lourenço *et al.* 2013). Therefore, the development and optimization of fast evolving nuclear markers such as microsatellite loci is crucial to unveil genetic diversity and structure patterns in this taxonomic group.

The range of the Mediterranean stripe-necked terrapin *Mauremys leprosa* (Schweigger, 1812) embraces the Northwestern Africa and the Iberian Peninsula, with a few populations located in southwestern France (Keller & Busack 2001). This species is currently threatened by habitat fragmentation and/or destruction, pet trade, alien species (Polo-Cavia *et al.* 2011) and pathogens (Hidalgo-Vila *et al.* 2008; Verneau *et al.* 2011). *Mauremys leprosa* is considered vulnerable by the IUCN and is listed in Appendix II of the Berne Convention and in Appendix II and IV of Habitat Directive (92/43/CEE) (Cox & Temple 2009). A phylogeographic study of *M. leprosa* identified two major mitochondrial lineages (Fritz *et al.* 2006) classified in two subspecies: *M. l. saharica* (southern Morocco, eastern Algeria and Tunisia) and *M. l. leprosa* (Iberian Peninsula and northern Morocco). However, genetic differentiation was inferred solely from mtDNA data and population structuring within the two subspecies was not well resolved by the use of this marker. For developing conservation strategies and identification of management units, genetic population structure and diversity needs to be analysed in more detail. In this study, we tested and optimized a set of 16 microsatellite markers for *M. leprosa* that were developed for two closely related species (*M. caspica* and *M. rivulata*).

4.2 Material and Methods

We tested 13 microsatellite loci developed for *M. caspica* (Vamberger *et al.* 2011) and three microsatellite loci developed for *M. rivulata* (Mantziou *et al.* 2005) for cross-amplification in 190 *Mauremys leprosa* belonging to six Iberian (Algarve, Castro Verde, Castelo Branco, Caldas da Rainha, Madrid, Murcia) and two African populations (Ceuta and Tazia). Genomic DNA was extracted from blood using EasySpin (for *M. leprosa*) or Quiagen (for *M. caspica* and *M. rivulata*) extraction kits, following the manufacturer's protocol. In order to ensure that primers were amplifying correctly in *M. leprosa*, we used samples from *M. caspica* and *M. rivulata* as a positive control. We divided the PCR reaction in two multiplexes (see table 4.1); forward primers were labelled with fluorescent dye markers (FAM, NED, VIC and PET; Oetting *et al.* 1995).

PCR amplifications were performed on 10 μl final volume containing 2 μl of DNA, 5 μl of Quiagen Multiplex PCR Kit and 0.14 μl-0.32 μl at 10 μM of each primer (table 4.1). Touchdown PCR conditions started with an initial denaturation for 95°C for 15 min (minutes) followed by 7 cycles of 95°C for 30 s (seconds), 58°C for 1 min (decreasing 0.5°C per cycle to 55°C), 72°C for 30 s; 24 cycles of 95°C for 30 s, 55°C for 1 min, 72°C for 30 s; 8 cycles of 95°C for 30 s, 53°C for 1 min, 72°C for 30 s, and a final elongation step at 60°C for 30 min. PCRs were performed on BioRad C1000 Thermocycler and genotyped on an ABI 3130*xl* genetic analyzer (Applied Biosystems, FosterCity, CA, USA). GeneScan™-500 Liz was used as fragment size standard to score amplicons sizes on GeneMapperv4.0 (Applied Biosystems). GENEPOP v4.2 (Rousset 2008) was used to assess deviations from Hardy-Weinberg equilibrium (H-WE) for each locus and population and linkage disequilibrium (LD) between loci at each population using the Markov chain method with 1000 batches and 10000 iterations per batch. Bonferroni correction for multiple comparisons was applied for both cases. Observed and expected heterozygosities were calculated using GenAlExv6.4 (Peakall & Smouse 2006) and the possible existence of null alleles at each locus for all populations was assessed with MICROCHECKER v2.2.3 (Van Oosterhout *et al.* 2004).

4.3 Results and Discussion

All loci were in H-WE in all populations except for three cases that showed a deviation caused by heterozygote deficit (MC18 and MR-1 in Murcia, and MC20 in Tazia). Nevertheless, this pattern was not observed across all populations and therefore the loci were not discarded. The fact that we observed this pattern in only two populations might suggest that they could be inbred or that they are non-panmictic populations. Another explanation might be the presence of null alleles in MR1. This marker, together with MC21 showed signs of null alleles, since they have a low number of alleles sampled (MR1 – 4 alleles and MC21 – 6 alleles) and appeared as homozygotes in the majority of populations. No cases of linkage disequilibrium were found. None of the tested primers failed to amplify. However, three loci turned out to be monomorphic (MR-3, MC8 and MC25). Allele variation ranged from 2 (MC1, MC17) to 18 (MC3) (table 4.1).

Table 4.1: Characteristics of 16 microsatellite loci tested in M. leprosa. Loci whose codes begin with MR were designed for Mauremys rivulata and the ones that begin with MC were designed for Mauremyscaspica. GenBank accession numbers arein brackets below each locus name abbreviation. Microsatellite repeat motif; Ta (C°) = PCR annealing temperature; Primer (µl) = quantity of primer (µl) of a 10 µM Primer solution; N = number of sampled individuals; Na = number of alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; P-value (H-WE) = Hardy-Weinberg probability test (Fisher's exact test).

Multiplex	Locus	Repeat Motif	T_a	Primer (µL)	N	Na	Allele size range (bp)**	Ho	He	P (HWE)***
1	Mr-1 [AY934859]	$(AC)_{11}$	56	0.2	117	4	205-213	0.019	0.124	Highsig.
1	* MC5 [HQ010418]	$(ATCT)_{12}$	56	0.32	156	15	184-244	0.773	0.805	0.6213
1	* MC6 [HQ010407]	$(ATCT)_{21}$	56	0.2	186	12	114-186	0.770	0.764	0.0063
1	MC8 [HQ010411]	$(AC)_{15}$	56	0.12	186	1	192	0	0	-
1	* MC12 [HQ010410]	$(TG)_{14}$	56	0.14	187	3	84-92	0.166	0.154	1
1	* MC17 [HQ010417]	$(TAGA)_8$	56	0.12	189	2	106-114	0.007	0.006	-
1	* MC22 [HQ010413]	$(CT)_6…(ATCT)_8$	56	0.14	189	6	100-112	0.562	0.542	0.2687
1	* MC24 [HQ010412]	$(AGAT)_8$	56	0.24	176	13	104-152	0.767	0.794	0.4777
2	MR-3 [AY934861]	$(GT)_8$	56	0.2	47	1	182	0	0	-
2	* MR-9 [AY934864]	$(CT)^{16}$	55.5	0.14	186	13	93-123	0.611	0.615	0.0185
2	* MC1 [HQ010420]	$(AGAT)_{12}$	56	0.14	189	2	89-105	0.02	0.018	-
2	* MC3 [HQ010419]	$(TAGA)_{14}$	56	0.24	175	18	195-263	0.783	0.8	0.2282
2	* MC18 [HQ010416]	$(ATCT)_{10}$	56	0.16	157	17	219-287	0.627	0.769	0.0158
2	* MC20 [HQ010415]	$(TATC)_{14}$	56	0.2	184	13	160-212	0.483	0.547	0.0099
2	MC21 [HQ010414]	$(TC)_{12}$	56	0.2	189	6	111-125	0.467	0.486	0.1821
2	MC25 [HQ010409]	$(AG)_{18}$	56	0.16	123	1	168	0	0	-

* Informative polymorphic microsatellite markers in *M. leprosa;*
**We've corrected the allele sizes by discounting the fluorescent tail size to the allele size.
***'-'correspond to cases where the marker showed to be monomorphic or with only two alleles, but one only had one copy.

Overall, we successfully optimized for cross-amplification 11 polymorphic microsatellites in *M. leprosa*, which allow estimating fine scale genetic diversity and structuring across the species' distribution. In addition, past demographic events can also be assessed with these markers, allowing to draw a more detailed picture on the biogeographic history of the species. *Mauremys leprosa* shares its range with another freshwater terrapin, the European pond turtle (*Emys orbicularis*), across the Iberian Peninsula and northwestern Africa. The two species inhabit similar water bodies, although the latter is thought to be more sensitive against pollution. While the evolutionary history and contemporary genetic structure is well-studied in Iberian populations of *E. orbicularis* using evidence from mtDNA and microsatellites (Velo-Antón *et al.* 2008), fast-evolving biparentally inherited microsatellites have never been used before for *M. leprosa*, which would be promising for a comparative phylogeographic study of both species. Moreover, the use of microsatellites and a comprehensive sampling would be particularly important for conservation studies because turtles are amongst the most common vertebrates associated to pet trade and

illegal translocations (Dijk *et al.* 2000; Moll & Moll 2004), which has an important effect on specie's genetic structure and genetic variability within populations (Gong *et al.* 2009; Velo-Antón *et al.* 2011b). For instance, genetic characterization of Iberian populations of *E. orbicularis* allowed the allocation of unknown samples from Recovery Centres to the most likely region of origin (Velo-Antón *et al.* 2007) and we expect that this will be possible for *M. leprosa* as well using the cross amplified microsatellite markers. For *M. leprosa*, many individuals are thought to be translocated to areas distant from their home populations across its entire distribution range, including terrapins from Morocco that were introduced to Iberian populations. Therefore, an accurate genetic characterization of *M. leprosa* populations will allow building a feasible tool to assign individuals of unknown origin to their natural populations, and help to better manage this species. Furthermore, we expect that these microsatellite markers will contribute to a better understanding of potential gene flow in the contact zone of the two subspecies in North Africa (e.g. Pedall et al. 2011 for *E. orbicularis*), paternity analyses (e.g. Roques et al. 2006) and changes in phenotypic traits (e.g. Velo-Antón et al. 2011a).

Acknowledgements

We thank A. Cordero Rivera, M. Casal, A. Álvarez, A. Egea, B. El Marnisi, S. Fahd, J. Cruz, Defense of Paul da Tornada Association – PATO, AltriFlorestal and all LIFE *Trachemys* team, in particular B. Martins and V. Soeiro, for their help during sample collection and fieldwork. This study was funded by LIFE-*Trachemys* (LIFE09/NAT/ES/000529) and Instituto de Estudios Ceutíes. GVA and JT are supported by post-doctoral fellowships from Fundação para a Ciência e Tecnologia (FCT, Portugal: SFRH/BPD/74834/2010 and SFRH/BPD/27173/2006).

Chapter 5: Manuscript III

Genetic footprint of a secondary contact zone and recent demographic expansion across the Iberian Peninsula of Mediterranean pond turtle (*Mauremys leprosa*)

Joana Veríssimo[1,2], Paulo Pereira[1,2], José Teixeira[1,3], & Guillermo Velo-Antón[1*]

1CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto. Instituto de Ciências Agrárias de Vairão. R. Padre Armando Quintas. 4485-661 Vairão Portugal

2Departamento de Biologia da Faculdade de Ciências da Universidade do Porto, Rua Campo Alegre, 4169-007 Porto, Portugal

3CIMAR-LA/CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Porto, Portugal

Abstract

During the climatic oscillations in the Quaternary, several *taxa* suffered range contractions, using the southern European peninsulas as *refugia*. Only recently, the Maghreb region was included as another refugium in the Mediterranean basin. When multiple *refugia* within species occurred, the resulting isolation of populations could lead to allopatric divergence. During post-glacial times, ranges expansions could promote secondary contact between divergent units. Given the recent findings of a secondary contact zone within *Mauremys leprosa* in northern Morocco, we hope to assess the possibility of gene flow between subspecies. Moreover, the effects of the recent European colonization on genetic diversity and structure are here analysed. A group of 11 microsatellite loci was used in 556 individuals, distributed through the majority of the species range (Morocco and Iberian Peninsula). We did not find any considerable differences of genetic diversity between the two continents. In North Africa the genetic structure has geographic concordance and the contact zone suggest gene flow between the subspecies. For the Iberian Peninsula, genetic structure does not show a geographic pattern, however, the high values of diversity indices suggests the possibility of a glacial *refugia* in SW Iberian Peninsula.

Keywords: Glacial *refugia*, Secondary contact zones, *Mauremys leprosa,* microsatellites.

5.1 Introduction

The Mediterranean Basin is considered one of the world biodiversity hotspots (Myers *et al.* 2000). Its three peninsulas – Iberian, Italian, and the Balkan – harbour a diverse group of species, mostly resulting from complex evolutionary histories as a response to climatic oscillations during the Quaternary (Hewitt 2011a). More recently, the Maghreb region was pointed as another refugium area in the Mediterranean basin, since it could be the source of genetic diversity for many species present in Europe (Husemann *et al.* 2014). Moreover, these areas have complex topographic features allowing for divergence and/or speciation processes due to population isolation in suitable climate *refugia* during the Quaternary climatic oscillations (see Hewitt 2004, 2011a; Husemann *et al.* 2014 and references therein).

After climate amelioration, larger areas of suitable habitat became available allowing species to expand (Hewitt 2011b). During these range expansions, populations could suffer repeated bottlenecks and founder effects leading to a decrease of genetic diversity in the northern populations (Hewitt 2004). However, multiple *refugia* within a species could also occur enabling allopatric divergence (Gómez & Lunt 2007; Stewart *et al.* 2010; Hewitt 2011a), with subsequent population expansions that may lead to the occurrence of secondary contact between or within species (Taberlet *et al.* 1998; Sequeira *et al.* 2005; Babik *et al.* 2005; Martínez-Solano *et al.* 2006; Gonçalves *et al.* 2009; Miraldo *et al.* 2013).

The Mediterranean pond turtle, *Mauremys leprosa* (Schwieegger, 1812), is a small terrapin endemic to the Iberian Peninsula (with a few and scattered population in south-western France), is also present in northern Africa, from Morocco to Tunisia (Keller & Busack 2001). Two subspecies are known: *Mauremys leprosa leprosa* (north of the Atlas Mountains in Morocco till the northern species range) and *Mauremys leprosa saharica* (south of the Atlas Mountains in Morocco, Algeria and Tunisia). Moreover, it was recently found the two subspecies became in contact in the Middle Atlas and the Rif in Morocco (see Chapter 3), however only mtDNA was used in previous phylogeographic studies and information from a bi-parentally inherited marker should be added in order to assess gene flow dynamics between both lineages.

Given the recent findings (see Chapter 3) we aim to obtain a more recent and fine-scale genetic pattern across Moroccan and the Iberian Peninsula populations. By using a set of microsatellites polymorphic loci we infer spatial genetic structure and diversity to 1) verify the current impact of the Atlas Mountains as a landscape barrier; 2)

determine gene flow possibility between subspecies on the contact zone, and 3) assess the resulting genetic pattern of the recent population expansion across the Iberian Peninsula.

5.2 Material and Methods

5.2.1 Sampling and Microsatellite Genotyping

Genomic DNA was extracted from 556 individuals sampled along 51 different locations across the Iberian Peninsula and Morocco (see Supplementary Material Table S. M. 2) covering the species distribution in this area (Fig. 5.1). A set of 11 polymorphic microsatellites previously tested were used (See chapter 4, Veríssimo et al., 2013).

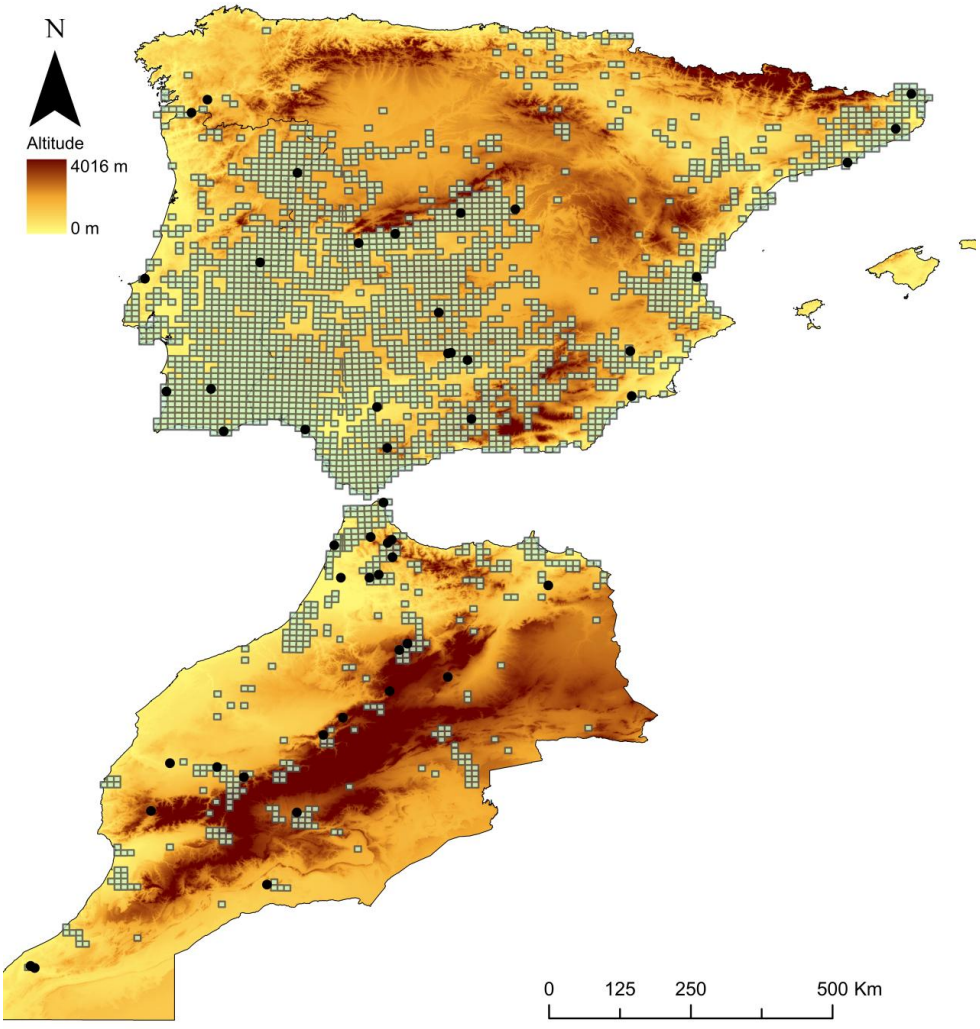


Fig. 5.1: Current species distribution of *Mauremys leprosa* across the study area (species presence in pale green) and sampling locations (marked as black dots).

5.2.2 Genetic diversity analysis

These analyses were performed in populations with 10 or more individuals. Deviations from Hardy-Weinberger equilibrium and linkage disequilibrium were assessed through GENEPOP v4.2 (Rousset 2008), using 1000 batches and 10000 iterations per batch. In both cases, Bonferroni correction for multiple comparisons was applied (Rice 1989). Observed (Ho) and expected (He) heterozygosities and mean number of alleles per locus (Na) were calculated using GENALEX v6.5.1 (Peakall & Smouse 2006). This software was also used to conduct a Principal Coordinate Analysis (PCA). HP-Rare (Kalinowski 2005) was used to calculate allelic richness (Ar) through a rarefaction method in order to eliminate the effect of different sample sizes. ARLEQUIN v3.5 (Excoffier & Lischer 2010) was used to calculate pairwise F_{ST} distances using 1000 permutations between all the selected populations.

5.2.3 Genetic structure analysis

From all successfully genotyped samples, we eliminated from further analysis each specimen that failed to amplify less than 75% of markers, to avoid erroneous results (Amos 2006). We assess population structure within the microsatellite dataset using STRUCTURE v2.3.4 (Pritchard *et al.* 2000) without providing a prior information about sampling location. The model implemented in this software assumes the existence of K populations, where K may be unknown, and each K is composed by a set of alleles' frequencies at each locus. A total of 500000 MCMC iterations preceded by 50000 of burn-in were performed in four independent runs for each K ($1 \leq K \geq 20$). Then, we used STRUCTURE HARVESTER Web v0.6.93 (Earl & vonHoldt 2011) to identify the number of K that would better explain our data. The K was chosen based on the posterior probabilities (highest lnP(D)) and ΔK method (Evanno *et al.* 2005). To better infer the genetic structure within Iberian and Moroccan populations we split the data by continent and run STRUCTURE v2.3.4 using the same parameters.

Table 5.1: Summary table of the dataset divided into sampling locality. ID: identification number of each sampling locality; Area: assigned area for graphical STRUCTURE outputs; Lat: latitude; Long: longitude; Locality: sampling locality or its description; *n*: number of samples per locality.

ID	Area	Lat	Long	Locality	*n*
1	S of Atlas Mountains	28.4969667	-10.8856	Tan-Tan, Guelta Ez Zerga	11
2	S of Atlas Mountains	28.53154952	-10.9504	Tan Tan, Draa river	5
3	S of Atlas Mountains	29.8233417	-7.1991	Tata, Oued Tissint	1
4	S of Atlas Mountains	30.968611	-6.72389	Embalse Ouarzzazate	6
5	N of Atlas Mountains	30.99038333	-9.03982	12km N of Timezgadiouine	1
6	N of Atlas Mountains	31.529717	-7.56338	Douer Targa	26
7	N of Atlas Mountains	31.6892	-7.98978	Marrakech (Palmeral) River Ouad Tansift	2
8	N of Atlas Mountains	31.74957833	-8.73844	Near Sidi-Chikér	1
9	N of Atlas Mountains	32.199494	-6.30223	Sidi Mimoun	2
10	N of Atlas Mountains	32.474118	-5.99285	Sidi Mimoun	4
11	Contact Zone	32.893142	-5.25043	Sidi Mimoun	1
12	Contact Zone	33.1187583	-4.32936	Boulemane, Oued Sebb Ousfa	2
13	Contact Zone	33.548236	-5.09752	Sidi Mimoun	1
14	Contact Zone	33.649233	-4.96812	Sidi Mimoun	3
15	Contact Zone	33.650625	-4.96812	Sidi Mimoun	1
16	Contact Zone	34.57302	-2.73322	near Douira	1
17	Contact Zone	34.69665	-6.02587	road to Moulay Bousselhaim	1
18	Contact Zone	34.696982	-5.57259	Sidi Mimoun	1
19	Contact Zone	34.745983	-5.42282	5km before Zoumi	1
20	Contact Zone	35.022558	-5.20518	Fifi	18
21	Contact Zone	35.21	-6.13	Agadir, Loukkos	1
22	Contact Zone	35.247357	-5.282	Sidi Mimoun	1
23	Contact Zone	35.299448	-5.2187	Sidi Mimoun	1
24	Contact Zone	35.341814	-5.55192	Tazia	17
25	Contact Zone	35.89036	-5.34878	Ceuta_Embalse Renegado	29
26	Iberian Peninsula	36.75649	-5.29005	Málaga	3
27	Iberian Peninsula	37.020458	-7.88625	Algarve	21
28	Iberian Peninsula	37.049271	-6.59136	Doñana	14
29	Iberian Peninsula	37.217609	-3.95261	Granada, Brácana	1
30	Iberian Peninsula	37.406293	-5.4519	Sevilla, Fuentes de Andalucia	1
31	Iberian Peninsula	37.581687	-1.40875	Murcia: Las Moreras Mazarron	2
32	Iberian Peninsula	37.652608	-8.79358	Almograve	16
33	Iberian Peninsula	37.693791	-8.0863	Castro Verde	16
34	Iberian Peninsula	38.152175	-4.01453	Andújar, arroyo de la Cabrera	13
35	Iberian Peninsula	38.258223	-4.32406	Cardeña, centro de información	15
36	Iberian Peninsula	38.266715	-4.27808	Cardeña, embalse Tejoneras	6
37	Iberian Peninsula	38.293806	-1.43219	Murcia: Cieza (Embalse del Judío)	28
38	Iberian Peninsula	38.907735	-4.4721	Ciudade Real	11
39	Iberian Peninsula	39.4447	-9.13751	Caldas da Rainha	31
40	Iberian Peninsula	39.470222	-0.37714	Valencia - Peñíscola	16
41	Iberian Peninsula	39.70255	-7.30815	Castelo Branco	30

ID	Area	Lat	Long	Locality	*n*
42	Iberian Peninsula	40.010028	-5.7425	Cáceres: Jaraiz de la Vera	29
43	Iberian Peninsula	40.158179	-5.16166	Ávila: Poyales del Hoyo	25
44	Iberian Peninsula	40.48825	-4.12439	Madrid: Fresnedillas de la Oliva	28
45	Iberian Peninsula	40.548417	-3.25661	Guadalajara	16
46	Iberian Peninsula	41.125028	-6.71611	Salamanca: Vilvestre	32
47	Iberian Peninsula	41.287028	2.016194	Barcelona: Delta del Llobregat	30
48	Iberian Peninsula	41.824222	2.781694	Girona: Caldes de Malabella	3
49	Iberian Peninsula	42.081165	-8.39662	As neves	3
50	Iberian Peninsula	42.287553	-8.1435	Ribadavia	9
51	Iberian Peninsula	42.377583	3.030556	Girona: Albera	19

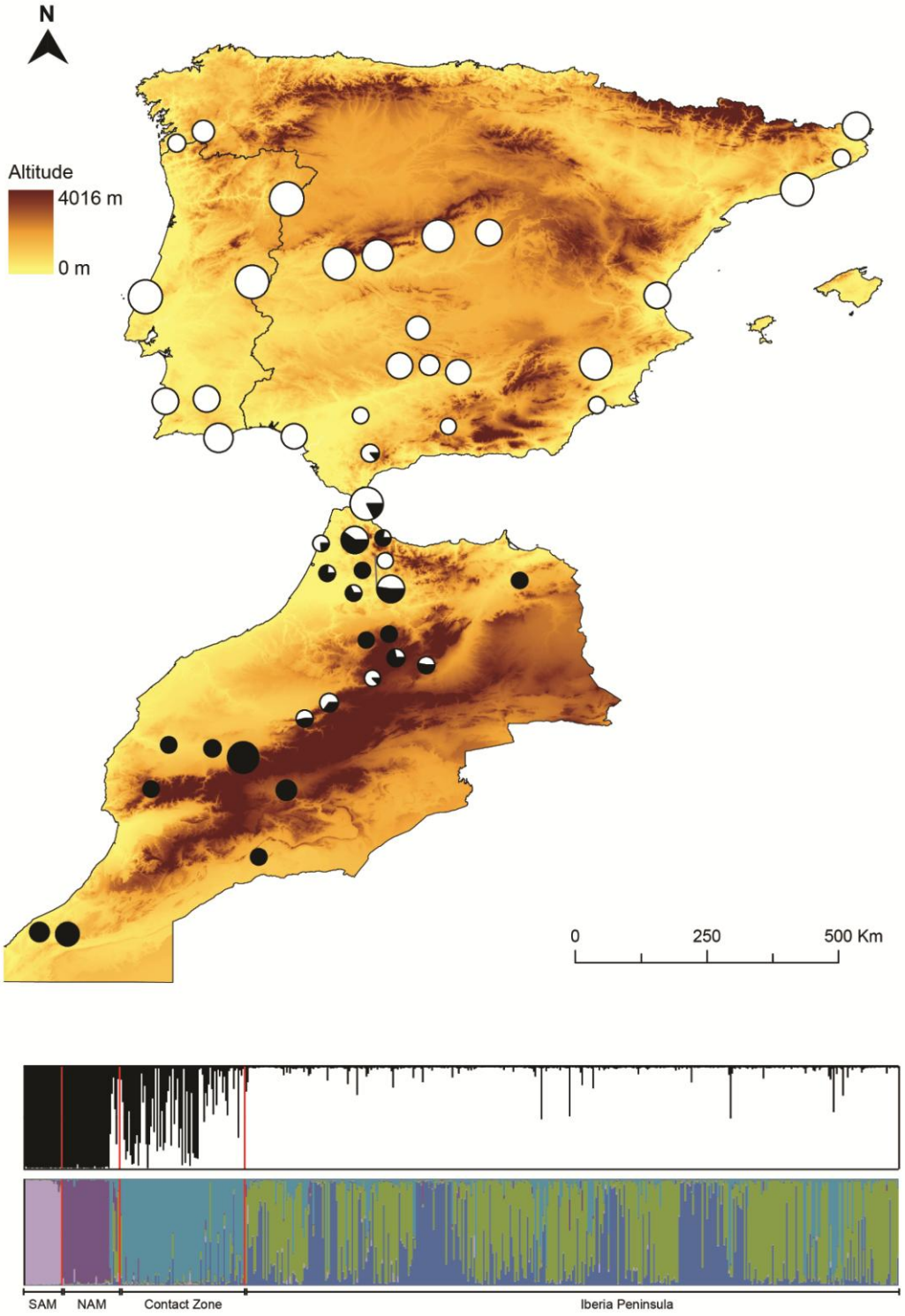


Fig. 5.2: On top: Population structure pattern of *Mauremys leprosa* across the study area for K=2 (most probable K; see Supplementary Material S.M. Fig.1 for Structure Harvester output). Each pie chart corresponds to a sample location and the size is proportional to the number of samples. Below: STRUCTURE output for K=2 (on top; most probable K) and K=5 (below). Samples are distributed from right to left following an orientation from south to north. The red lines represent spatial areas divisions, with SAM corresponding to South of the Atlas Mountains, NAM corresponding to North of the Atlas Mountains, the Contact Zone and the Iberian Peninsula. Each line corresponds to a single individual. For more information regarding samples, localities and areas see Table 5.1.

Table 5.2: Summary table of the African dataset divided into sampling locality. ID: identification number of each sampling locality; Area: assigned area for graphical STRUCTURE outputs; Lat: latitude; Long: longitude; Locality: sampling locality or its description; *n*: number of samples per locality.

ID	Area	Lat	Long	Locality	*n*
1	S of Atlas Montains	28.49697	-10.8856	Tan-Tan, Guelta Ez Zerga	11
2	S of Atlas Montains	28.53155	-10.9504	Tan Tan, Draa river	5
3	S of Atlas Montains	29.82334	-7.1991	Tata, Oued Tissint	1
4	S of Atlas Montains	30.96861	-6.72389	Embalse Ouarzzazate	6
5	N of Atlas Mountains	30.99038	-9.03982	12km N of Timezgadiouine	1
6	N of Atlas Mountains	31.52972	-7.56338	Douer Targa	26
7	N of Atlas Mountains	31.6892	-7.98978	Marrakech (Palmeral) River Ouad Tansift	2
8	N of Atlas Mountains	31.74958	-8.73844	Near Sidi-Chikér	1
9	N of Atlas Mountains	32.19949	-6.30223	Sidi Mimoun	2
10	N of Atlas Mountains	32.47412	-5.99285	Sidi Mimoun	4
11	Contact Zone	32.89314	-5.25043	Sidi Mimoun	1
12	Contact Zone	33.11876	-4.32936	Boulemane, Oued Sebb Ousfa	2
13	Contact Zone	33.54824	-5.09752	Sidi Mimoun	1
14	Contact Zone	33.64923	-4.96812	Sidi Mimoun	3
15	Contact Zone	33.65063	-4.96812	Sidi Mimoun	1
16	Contact Zone	34.57302	-2.73322	near Douira	1
17	Contact Zone	34.69665	-6.02587	road to Moulay Bousselhaim	1
18	Contact Zone	34.69698	-5.57259	Sidi Mimoun	1
19	Contact Zone	34.74598	-5.42282	5km before Zoumi	1
20	Contact Zone	35.02256	-5.20518	Fifi	18
21	Contact Zone	35.21	-6.13	Agadir, Loukkos	1
22	Contact Zone	35.24736	-5.282	Sidi Mimoun	1
23	Contact Zone	35.29945	-5.2187	Sidi Mimoun	1
24	Contact Zone	35.34181	-5.55192	Tazia	17
25	Contact Zone	35.89036	-5.34878	Ceuta_Embalse Renegado	29

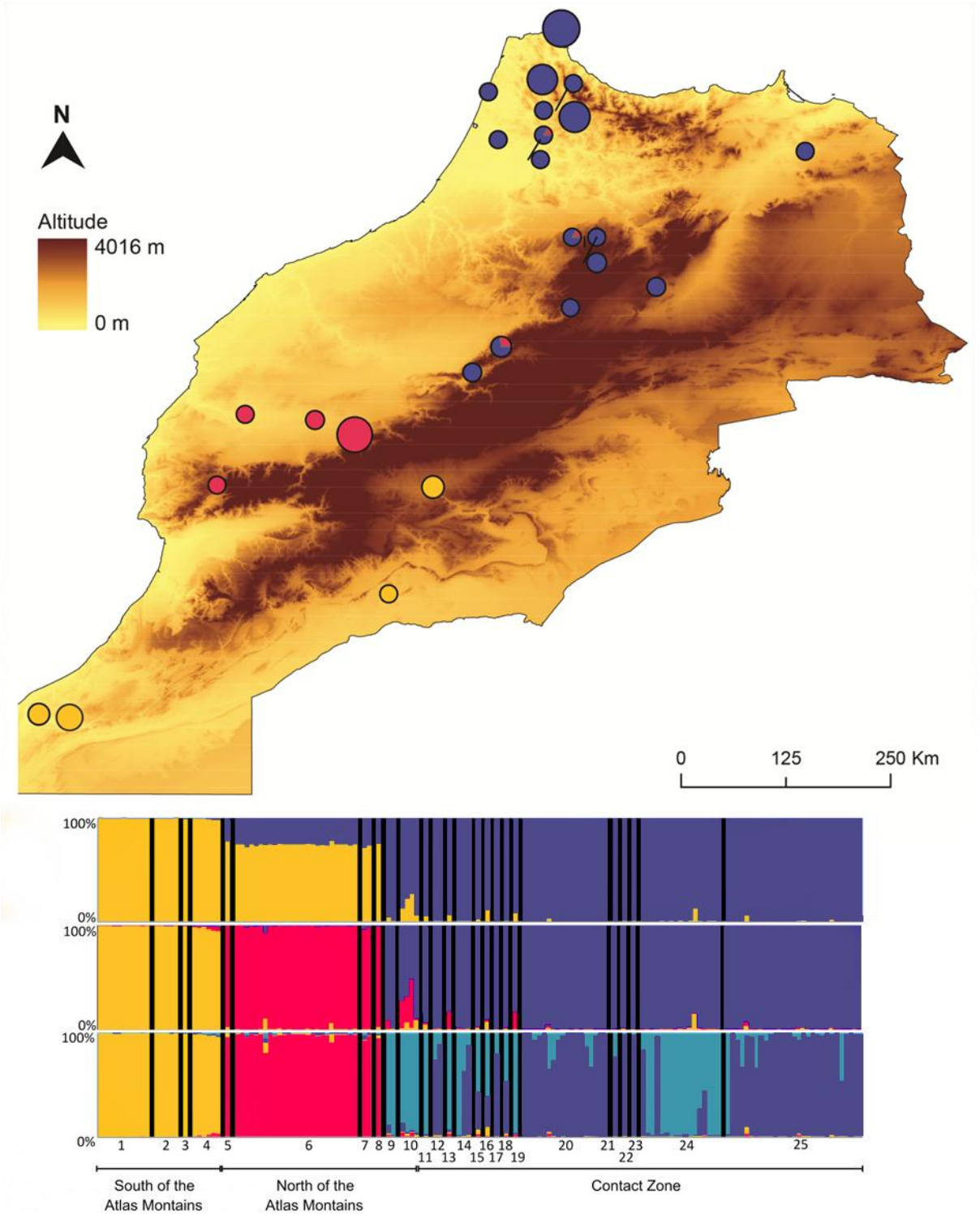


Fig. 5.3: On top: Population structure pattern of *Mauremys leprosa* in Africa for K=3 (most probable K; see Supplementary Material S.M. Fig.1 for Structure Harvester output). Each pie chart corresponds to a sample location and the size is proportional to the number of samples. **Below:** STRUCTURE output for K=2 (on top), K=3 (middle) and K=4 (below). Samples are distributed from right to left following an orientation from south to north. The black lines represent spatial areas divisions. Each line corresponds to a single individual. For more information regarding samples, localities and areas see Table 5.2.

Table 5.3: Summary table of the Iberian Peninsula dataset divided into sampling locality. ID: identification number of each sampling locality; Lat: latitude; Long: longitude; Locality: sampling locality or its description; *n*: number of samples per locality.

ID	Lat	Long	Locality	*n*
1	36.75649	-5.29005	Málaga	3
2	37.02046	-7.88625	Algarve	21
3	37.04927	-6.59136	Doñana	14
4	37.21761	-3.95261	Granada, Brácana	1
5	37.40629	-5.4519	Sevilla, Fuentes de Andalucia	1
6	37.58169	-1.40875	Murcia: Las Moreras Mazarron	2
7	37.65261	-8.79358	Almograve	16
8	37.69379	-8.0863	Castro Verde	16
9	38.15218	-4.01453	Andújar, arroyo de la Cabrera	13
10	38.25822	-4.32406	Cardeña, centro de información	15
11	38.26672	-4.27808	Cardeña, embalse Tejoneras	6
12	38.29381	-1.43219	Murcia: Cieza (Embalse del Judío)	28
13	38.90774	-4.4721	Ciudade Real	11
14	39.4447	-9.13751	Caldas da Rainha	31
15	39.47022	-0.37714	Valencia - Peñíscola	16
16	39.70255	-7.30815	Castelo Branco	30
17	40.01003	-5.7425	Cáceres: Jaraiz de la Vera	29
18	40.15818	-5.16166	Ávila: Poyales del Hoyo	25
19	40.48825	-4.12439	Madrid: Fresnedillas de la Oliva	28
20	40.54842	-3.25661	Guadalajara	16
21	41.12503	-6.71611	Salamanca: Vilvestre	32
22	41.28703	2.016194	Barcelona: Delta del Llobregat	30
23	41.82422	2.781694	Girona: Caldes de Malabella	3
24	42.08117	-8.39662	As neves	3
25	42.28755	-8.1435	Ribadavia	9
26	42.37758	3.030556	Girona: Albera	19

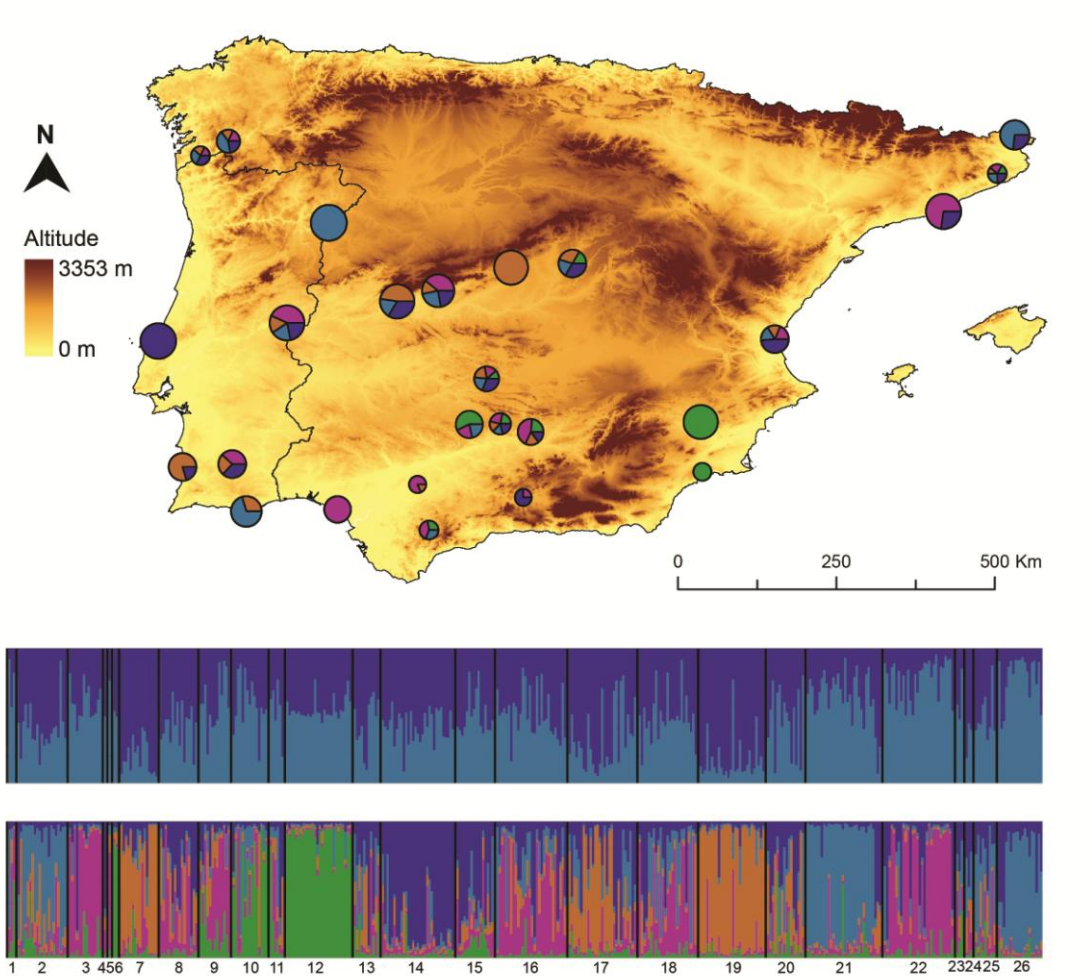


Fig. 5.4: On top: Population structure pattern of *Mauremys leprosa* in the Iberian Peninsula for K=5 (most probable K; see Supplementary Material S.M. Fig.1 for Structure Harvester output). Each pie chart corresponds to a sample location and the size is proportional to the number of samples. **Below:** STRUCTURE output for K=2 (on top) and K=5 (below). Samples are distributed from right to left following an orientation from south to north. The black lines represent spatial areas divisions. Each line corresponds to a single individual. For more information regarding samples, localities and areas see Table 5.3.

5.3 Results

Number of alleles range from 4 (MC1) to 21 (MC5), with an average of 13.9. Regarding further representation of the sampling sites, for easier graphical representation, Moroccan localities were pooled into three groups: South of the Atlas Mountains, North of the Atlas Mountains, and the Contact zone region (See Table 5.1 for details).

5.3.1 Genetic diversity analysis

After analysing populations sample size, we discarded 65 individuals (corresponding to populations with less than 10 individuals) obtaining a total of 491 samples divided into 23 populations (Table 5.4). From those 65 individuals, 37 have African origin, diminishing the African dataset when comparing it with the European (Morocco dataset: N= 101; Iberian Peninsula: N =390). Tan-Tan (ID=1, Morocco) and Ciudad Real (ID=38, Spain) are our smallest sampled populations (N=11), and Salamanca (ID=46, Spain) is the largest sampled population with 32 individuals. All populations were under Hardy-Weinberg equilibrium (HW-E), except for Ceuta (Maucas 6), Tazia (Maucas 20), and Douer Targa (Maucas 12), which showed signs of heterozygosity deficiency. Signs of linkage disequilibrium (LD) were found in Barcelona (Maucas 18 and Maucas 3) and Almograve (Maucas 18 and Maucas 3; Maucas 18 and Maucas 5). Given that no geographical pattern of LD was detected across all studied populations we may assume that there is no physical linkage since it only occurs in two populations and thus none of the loci were removed from further analysis.

Table 5.4: Summary table of the dataset used for population analysis. ID: original identification number of each sampling locality; New_ID: defined identification number to be used for population analysis; Area: assigned area for graphical STRUCTURE outputs; Population: sampling locality or its description; *n*: number of samples per population.

ID	New_ID	Area	Population	*n*
1	1	S of Atlas Mountains	Tan-Tan, Guelta Ez Zerga	11
6	2	N of Atlas Mountains	Douer Targa	26
20	3	Contact Zone	Fifi	18
24	4	Contact Zone	Tazia	17
25	5	Contact Zone	Ceuta_Embalse Renegado	29
27	6	Iberian Peninsula	Algarve	21
28	7	Iberian Peninsula	Doñana	14
32	8	Iberian Peninsula	Almograve	16
33	9	Iberian Peninsula	Castro Verde	16
34	10	Iberian Peninsula	Andújar, arroyo de la Cabrera	13
35	11	Iberian Peninsula	Cardeña, centro de información	15
37	12	Iberian Peninsula	Murcia: Cieza (Embalse del Judío)	28
38	13	Iberian Peninsula	Ciudade Real	11
39	14	Iberian Peninsula	Caldas da Rainha	31
40	15	Iberian Peninsula	Valencia - Peñíscola	16
41	16	Iberian Peninsula	Castelo Branco	30
42	17	Iberian Peninsula	Cáceres: Jaraiz de la Vera	29
43	18	Iberian Peninsula	Ávila: Poyales del Hoyo	25
44	19	Iberian Peninsula	Madrid: Fresnedillas de la Oliva	28
45	20	Iberian Peninsula	Guadalajara	16
46	21	Iberian Peninsula	Salamanca: Vilvestre	32
47	22	Iberian Peninsula	Barcelona: Delta del Llobregat	30
51	23	Iberian Peninsula	Girona: Albera	19

Table 5.5: Summary table of diversity indices for *Mauremys leprosa* populations. ID: identification number to be used for population diversity analysis; *n*: number of samples per location; Na: number of alleles; Ne: number of effective alleles; He: expected heterozygosity; uHe: unbiased expected heterozygosity; PA: number of private alleles; Ar: allelic richness. Columns with * the value represent the population average.

ID	Population	*n*	Na*	Ne*	Ho*	He*	uHe*	PA	Ar
1	Tan-Tan, Guelta Ez Zerga	11	4.636	3.214	0.534	0.533	0.559	7	4.14
2	Douer Targa	26	7.818	5.121	0.525	0.568	0.579	14	5.03
3	Fifi	18	5.545	3.659	0.528	0.553	0.569	3	4.31
4	Tazia	17	4.727	3.184	0.475	0.554	0.575	-	3.92
5	Ceuta	29	6.909	4.074	0.535	0.610	0.623	4	4.77
6	Algarve	21	4.818	3.143	0.483	0.488	0.500	-	3.74
7	Doñana	14	5.727	3.678	0.550	0.543	0.564	1	4.48
8	Almograve	16	5.182	3.463	0.506	0.545	0.563	-	4.15
9	Castro Verde	16	6.455	3.904	0.575	0.592	0.614	4	5
10	Andújar	13	5.545	4.109	0.601	0.576	0.599	1	4.64
11	Cardeña	15	4.909	3.548	0.559	0.557	0.577	-	4.18
12	Murcia	28	4.909	2.778	0.482	0.507	0.516	-	3.6
13	Ciudade Real	11	5.182	3.180	0.466	0.509	0.538	-	4.54
14	Caldas da Rainha	31	5.364	3.125	0.462	0.450	0.457	-	3.67
15	Valencia	16	6.000	3.972	0.498	0.562	0.581	-	4.66
16	Castelo Branco	30	6.636	3.911	0.539	0.553	0.563	1	4.48
17	Cáceres	29	6.636	3.900	0.559	0.550	0.560	-	4.47
18	Ávila	25	6.273	4.082	0.584	0.559	0.570	1	4.44
19	Madrid	28	4.818	3.027	0.552	0.524	0.535	-	3.57
20	Guadalajara	16	5.273	3.326	0.544	0.531	0.551	2	4.28
21	Salamanca	32	5.727	3.153	0.482	0.518	0.527	-	3.84
22	Barcelona	30	6.182	3.508	0.568	0.564	0.574	-	4.14
23	Gerona	19	3.455	1.890	0.294	0.315	0.325	-	2.63

In general, observed (H_O) and expected (H_E) heterozygosity showed similar values throughout the study area, ranging between *ca.* 0.3 and 0.6. The north easternmost population, Girona, presented the lowest values across all measured diversity indices (H_O =0.294 and H_E =0.315; Ar=2.63; Table 5.5). While Douer Targa and Ceuta (both in Morocco), presented the highest diversity indices (Douer Targa: H_O =0.525 and H_E =0.568; Ar=5.03; Ceuta: H_O =0.535 and H_E =0.610; Ar=4.77; Table 5.5). In central and southern Iberian Peninsula, three populations showed diversity values as high as Moroccan populations (Andújar: H_O = 0.601, Ar = 4.64; Ávila: H_O =0.584, Ar = 4.44, and Castro Verde: H_O = 0.575, Ar = 5). All Moroccan populations presented private alleles, with the populations near the Atlas Mountains scoring the higher values (Douer Targa, PA = 14, and Tan-Tan (Guelta Ez Zerga, PA = 7). The exception was found in Tazia

with no private alleles. As expected, the majority of Iberian Peninsula populations did not present private alleles.

5.3.2 Genetic structure analysis

When analysing genetic structure for the complete dataset the most probable number of genetic clusters was K=2. All populations in the Iberian Peninsula were assigned to the same cluster, with the exception of individuals from Malaga which show admixture. In northern Morocco, the majority of populations also showed admixture, while the more southern populations were assigned to a single cluster (Fig 5.2). It is interesting to note that in the Atlas Mountains, where the two subspecies exist (geographically separated by the Anti and High Atlas), both subspecies were assigned to the same cluster. When solely analysing the African dataset, the best number of genetic clusters was K=3 (see Supplementary Material S.M. Fig.1 for Structure Harvester output) and thus, the nuclear pattern became more similar to the one found for mtDNA, with the clear separation of the subspecies (Fig. 5.3). The Iberian Peninsula (when analysed solely) shows higher structure than Morocco, with five clusters being found (see Supplementary Material S.M. Fig.1 for Structure Harvester output). However, there is a high degree of admixture between populations across the area, with more than half of populations having three or more assigned clusters. Generally, microsatellites recovered little geographic structure since no locality or group of localities had a unique cluster assigned. Nonetheless, the majority of the individuals from Doñana, Murcia, Caldas da Rainha, Madrid, and Salamanca were each assigned to single clusters, denoting a unique ancestry and less admixture with other populations (Fig. 5.4).

F_{ST} pairwise comparisons revealed the highest value between the most distant populations, Tan-Tan and Gerona (F_{ST} = 0.576). The lowest value was found between Guadalajara and Ávila (F_{ST} = 0.015; geographic distance between these populations is *ca.*170 km), however, five pairwise F_{ST} comparisons revealed to be non-significant (Table 5.6). Furthermore, when only comparing populations from Morocco the higher F_{ST} value is between Tan-Tan and Douer Targa (F_{ST} = 0.528) and between European population the value drops to F_{ST} = 0.268 between Doñana and Gerona. The PCA results were congruent with the previously obtained STRUCTURE and mitochondrial results, where the Iberian Peninsula populations were pooled together almost overlapping with each other. The first axis was able to explain 37.77% of the genetic diversity found and the second 21.11% (Fig. 5.5).

Table 5.6: Pairwise FST values calculated through Arlequin based on allele frequencies. Light red highlights the highest values, while light blue highlights the lowest value. Values in bold are non-significant after 10000 permutations.

	Tan-Tan	Douer Targa	Fifi	Tazia	Ceuta	Algarve	Doñana	Almograve	Castro Verde	Andújar	Cardeña	Murcia	Ciudade Real	Caldas da Rainha	Valencia	Castelo Branco	Cáceres	Ávila	Madrid	Guadalajara	Salamanca	Barcelona
Douer Targa	0.528																					
Fifi	0.495	0.406																				
Tazia	0.460	0.347	0.167																			
Ceuta	0.437	0.304	0.114	0.079																		
Algarve	0.494	0.424	0.322	0.288	0.160																	
Doñana	0.469	0.419	0.263	0.255	0.144	0.098																
Almograve	0.462	0.404	0.238	0.222	0.078	0.106	0.125															
Castro Verde	0.435	0.369	0.178	0.155	0.060	0.073	0.044	0.038														
Andújar	0.448	0.375	0.248	0.179	0.074	0.036	0.096	0.021*	0.023													
Cardeña	0.471	0.396	0.298	0.221	0.109	0.101	0.148	0.078	0.075	0.026												
Murcia	0.489	0.408	0.336	0.289	0.154	0.130	0.173	0.098	0.126	0.072	0.104											
Ciudade Real	0.482	0.407	0.288	0.180	0.090	0.123	0.123	0.045	0.028	0.034	0.062	0.137										
Caldas da Rainha	0.539	0.465	0.235	0.279	0.162	0.178	0.126	0.106	0.059	0.139	0.170	0.210	0.122									
Valencia	0.466	0.360	0.235	0.141	0.055	0.110	0.132	0.059	0.042	**0.022**	0.061	0.080	**0.029**	0.144								
Castelo Branco	0.490	0.399	0.239	0.196	0.084	0.088	0.119	0.036	0.027	0.049	0.103	0.166	**0.021**	0.095	0.049							
Cáceres	0.447	0.387	0.260	0.199	0.095	0.130	0.145	0.024	0.046	0.043	0.069	0.125	**0.003**	0.133	0.044	0.044						
Ávila	0.474	0.392	0.257	0.218	0.100	0.096	0.115	0.031	0.028	0.044	0.092	0.114	0.026	0.100	0.029	0.024	0.017					
Madrid	0.474	0.415	0.260	0.273	0.123	0.067	0.117	0.051	0.057	0.051	0.115	0.105	0.107	0.115	0.088	0.072	0.077	0.059				
Guadalajara	0.494	0.389	0.238	0.173	0.066	0.135	0.158	0.034	0.038	0.038	0.079	0.100	**0.023**	0.130	**0**	0.024	0.022	0.015	0.057			
Salamanca	0.517	0.431	0.292	0.254	0.133	0.056	0.111	0.071	0.049	0.068	0.128	0.139	0.075	0.101	0.076	0.020	0.091	0.041	0.060	0.061		
Barcelona	0.445	0.371	0.226	0.184	0.089	0.082	0.044	0.067	0.038	0.038	0.086	0.093	0.063	0.078	0.051	0.075	0.091	0.063	0.085	0.078	0.059	
Girona	0.576	0.485	0.406	0.287	0.199	0.194	0.268	0.195	0.143	0.167	0.202	0.215	0.086	0.250	0.128	0.107	0.149	0.131	0.216	0.140	0.117	0.191

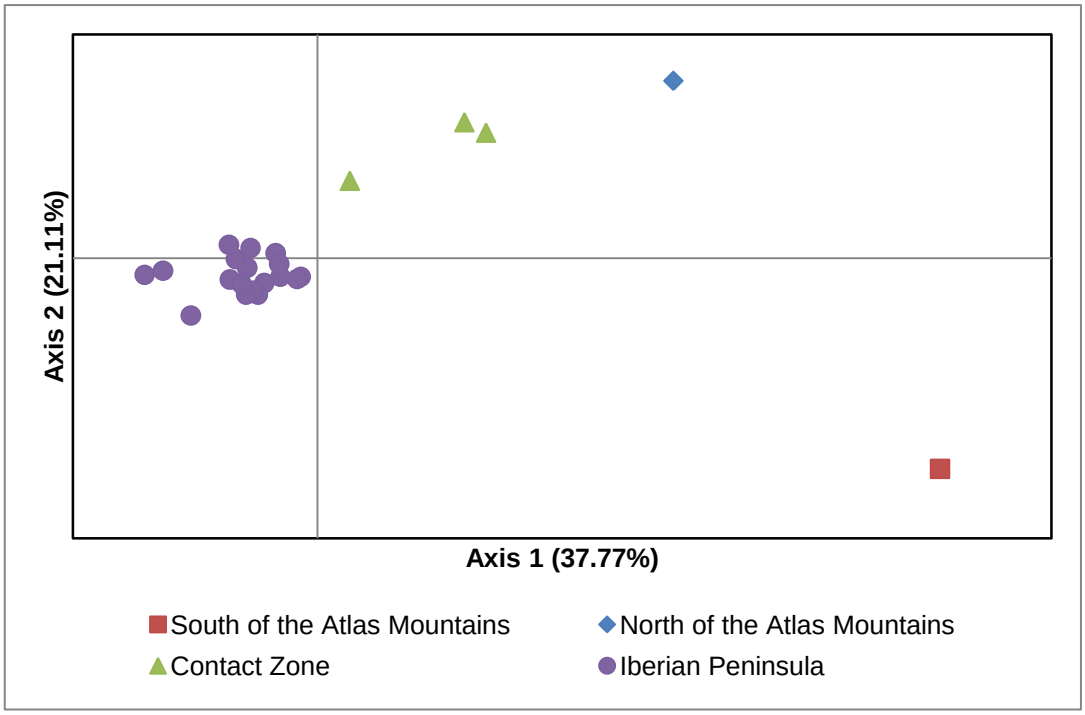


Fig. 5.5: PCA based on populations' genetic distances of 11 microsatellite loci. All populations are represented, but grouped by previously defined areas. For details see table 5.4.

5.4 Discussion

5.4.1 Genetic structure and gene flow across the contact zone

In 2006, Fritz *et al.* pointed the north-western Maghreb as the origin of the species. This hypothesis was later reinforced, and high levels of genetic structuring within each subspecies – *M. l. leprosa* and *M. l. saharica* - were assessed through two mitochondrial markers (see Chapter 3), showing that the current genetic diversity and structure of the species has been shaped by the climate-induced expansions. Therefore, and as expected, we found higher genetic diversity in Morocco than in the Iberian Peninsula, although the latter presents higher genetic structure.

Regarding F_{ST} pairwise comparisons within each continent, Iberian Peninsula populations are more similar between them presenting a maximum F_{ST} value two times lower than the maximum found in Morocco (Morocco: F_{ST}=0.528, Iberian Peninsula: F_{ST} = 0.268). Namely, this high F_{ST} value found between African populations (Tan-Tan and Douer Targa) corresponds to the divergence resulting of these populations representing the two *M. leprosa* subspecies. Nonetheless, the barrier effect of the Atlas Mountains seems to have diminished, allowing for gene flow across the southwestern mountain chain. Individuals from the Anti-Atlas area and north of the Middle Atlas were clustered together indicating recent and mild gene flow in the area between both

subspecies. However, for higher clustering divisions (in K=5 for the complete dataset, see Fig.5.2; and in K=3 for the African dataset, see Fig.5.3) this group appears divided accordingly to subspecies, which also occurs in the PCA analysis.

A secondary contact between *M. leprosa* subspecies was recently reported, when several *M. l. saharica* individuals were found in syntopy with *M. l. leprosa* in the Rif and Middle Atlas (see Chapter 3). Regarding mtDNA results, this area presented the highest diversity across the species range and the same occurs for nuDNA. In fact, Ceuta is the population carrying more genetic diversity (H_E), which can be the effect of the sublineages contact (since this population has individuals assigned to both lineages). It is important to note that the genetic structure here inferred shows a larger contact zone than with mtDNA. In the latter, this contact zone was restricted to the northernmost part of Morocco, while here we found that admixture between the two subspecies extends further and reaches the Middle Atlas (Fig. 3.2 in Chapter 3 and Fig.5.2). The presence of the Oued Ouergha between the Rif and the Middle Atlas might have been the corridor used by *M. l. saharica* individuals towards the Tetouan region.

While in the Iberian Peninsula the vicariant effect produced by the Inner Plateau and post-glacial expansions followed by possible secondary contact has already been reviewed (Hewitt 1999, 2001; Gómez & Lunt 2007), the information regarding secondary contact in northern Morocco are scarce (e.g. *Buteo buteo buteo* x *Buteo rufinus cirtensis* in northern Morocco (Elorriaga & Muñoz 2013)). Given the observable pattern of genetic structure in the contact zone, one can assume that the two subspecies are admixed in terms on nDNA, allowing us to conclude that no reproductive barrier exists between them.

5.4.2 Population expansion in the Iberian Peninsula

During the upper-Pleistocene the Strait of Gibraltar might act as a permeable barrier, allowing for a single *M. leprosa* mitochondrial lineage to colonize the Iberian Peninsula. This event resulted in a lack of diversity (in comparison with Morocco) and an overall genetic homogeneity (see Chapter 3). Other studies in the area have shown a decrease of genetic diversity northwards (Carranza & Arnold 2004; Carranza *et al.* 2004, 2006a; Pinho *et al.* 2007; Recuero *et al.* 2007; Velo-Antón *et al.* 2008) as the result of the different migration waves reducing population variability along a colonization route (Hewitt 1999, 2000). However, this is not the case of *M. leprosa,* which presents similar values of diversity across the Peninsula (H_O; H_E; see table 5.5).

Moreover, some populations in south and central Iberia presented diversity values as high as in some Morocco populations (e.g.: Douer Targa Ar = 5.03 and Castro Verde Ar=5; Andújar H_O = 0.601 and Tan-Tan H_O =0.534). This high diversity areas in the Iberian Peninsula were unexpected based on previously inferences where the species showed very low nucleotide diversity along the area pointing for a current African origin of the individuals (see Chapter 3; Fritz et al. 2006) and the weak genetic differentiation between populations based on nuDNA. These highly diverse localities may be explained by a glacial refuge during the late Pleistocene in south-western Iberia, where temperatures could be more suitable for an ecthotermic species. Also, the most distant mtDNA haplotype found within the A3 lineage also occurs in this area (see chapter 3). The occurrence of a glacial refuge, or genetic isolation, in this area has already been reported to other species (e.g. cyprinid (Mesquita *et al.* 2005), Iberian newt (Martínez-Solano *et al.* 2006), *Blanus* sp. (Albert *et al.* 2007), Iberian emerald lizard (Godinho *et al.* 2008), midwife toads (Gonçalves *et al.* 2009), and southern smooth snake (Santos *et al.* 2012)). Nonetheless, the role of the Iberian Peninsula as a *refugia* for *Mauremys leprosa* needs more attention and further analysis should be performed to test this hypothesis.

It is interesting to note that for *Emys orbicularis* (a co-distributed terrapin in the Iberian Peninsula) genetic pattern shows higher geographical concordance with each population being assigned to a single cluster, denoting very low values of admixture (Velo-Antón *et al.* 2008), which indicates very different life histories and historical biogeographies for these two species. Even though, both *E. orbicularis* and *M. leprosa* had their origin in Africa with posterior re-colonization of the Iberian Peninsula, no evidence of possible *refugia* for *E. orbicularis* was found (Stuckas *et al.* 2014). In fact, the little genetic divergence observed between *E. orbicularis* lineages is very weak, indicating that this species suffered a very recent range expansion from Africa to the Iberian Peninsula, with clear signs of heterozygosity loss in the south-north axis (Velo-Antón *et al.* 2008; Stuckas *et al.* 2014).

Another interesting case appeared in south-eastern Spain. There is one cluster (presented in Fig. 5.4 in green) that only appears in a few populations, with a moderate probability of assignment, while appearing as the only cluster for Murcia. This pattern could be produced by the Betic system, which has been described to induce lineage differentiation in some reptile *taxa* (Fromhage *et al.* 2004; Velo-Antón *et al.* 2008; Fonseca *et al.* 2009; Miraldo *et al.* 2011, 2013; Kaliontzopoulou *et al.* 2011; Santos *et al.* 2012). Nonetheless, the effect induced by those mountains in *Mauremys leprosa* occurred much later than for the above cited examples, which might be explained by a

more contemporary barrier effect of this mountain, instead of vicariance induced by the Rif-Betic system separation.

Overall, this study contributed to fill the knowledge gap regarding the population structure of *Mauremys leprosa* across the Iberian Peninsula and Morocco. Also, these results give us a new perspective regarding gene flow across the subspecies, shedding new light into the contact zone dynamics. By using a set of bi-parentally inherited and highly polymorphic markers, we have access to a more detailed picture given the higher mutation rate of microsatellite loci allowing us to detect more recent demographic events, such as the possibility of a refugium later in the Pleistocene in the Iberian Peninsula.

Chapter 6: Final Remarks

The complex palaeogeographic history of the Mediterranean basin allied with the Milankovitch climatic oscillations have shaped the phylogeographic patterns of Western Palearctic species (Taberlet *et al.* 1998; Hewitt 1999, 2000; Martínez-Solano 2004; Veith *et al.* 2004; Magri *et al.* 2007; Miraldo *et al.* 2011; Sousa *et al.* 2012; Velo-Antón *et al.* 2012; Husemann *et al.* 2014). However, together with historical events, individual species' biology and ecological constraints plays key roles in the determination of actual biogeographical ranges and populations' genetic structure. So, each species depicts a unique interplay of these factors and consequent evolutionary history. With this study we were able to increase the knowledge regarding this matter by clarifying the phylogeography of *Mauremys leprosa*.

Regarding the phylogeography of this species, we were able to uncover deep divergent and highly structured sublineages in the Maghreb, when compared with the Iberian Peninsula, where only one sublineage is found. The genetic homogeneity and low genetic diversity in Iberian populations, as well as the network position of highly frequent haplotypes such as A3-1, supports the hypothesis of a recent re-colonization and rapid expansion of this species throughout the Iberian Peninsula, suggesting that the Strait of Gibraltar act as a permeable barrier to dispersal and gene flow between the Iberian and Moroccan populations. Moreover, we found that the Atlas Mountains acted as an important barrier for the *M. leprosa* since it is associated with the old split (late Pliocene) between the two major mitochondrial lineages (chapter 3). We were also able to detect a secondary contact zone between subspecies in northern Morocco, unknown until now.

Within this work we also successfully optimized by cross-amplification 11 polymorphic microsatellites in *M. leprosa* (chapter 4), which constitutes an important new genetic tool to allow the estimation of fine scale genetic diversity and structuring across the species' distribution.

The use of these highly polymorphic and bi-parentally inherited genetic markers allowed us to assess current gene flow between populations and to detect more recent demographic events, such as the possibility of a refugium in the later Pleistocene in the Iberian Peninsula (chapter 5). For the Iberian Penisula, microsatellites showed little geographic structure since no locality or group of localities had a unique cluster assigned. Given these observable patterns, we were unable to effectively establish a genetic tool capable to detect the origin of individuals present in Recovery Centers (see S. M. Fig. 8.2 in Supplementary Material), with the purpose of reallocating them to natural populations. In contrast, this type of genetic assignment tool was successfully

established for *Emys orbicularis,* with assignment probabilities achieving values over 90%, and never below 60% (Velo-Antón *et al.* 2007).

It was interesting to note vastly different genetic structure patterns and evolutionary histories between *Mauremys leprosa* and a co-distributed terrapin – *Emys orbicularis* – in the Iberian Peninsula, despite the colonization event of the area being considered recent and rapid in both cases (Velo-Antón *et al.* 2008; Stuckas *et al.* 2014).

After this thesis, the nuDNA and mtDNA available information for this species was largely increased, with 136 individuals sequenced for cyt-*b* and D-loop and 556 individuals screened for the microsatellite dataset optimized by us.

The currently available data for the Maghreb suggest that the sublineage B3 (found from Morocco to Tunisia) underwent a recent demographic or range expansion. However, the direction of expansion of this sublineage is still unclear, due to the sampling gap in Algeria. And so, regarding future prospects, Algerian samples should be added to recover in more detail the demographic history of *M. l. saharica* subspecies. Furthermore, both Tunisian and Algerian specimens should be screened for the microsatellite loci in order to assess the genetic diversity and structure patterns in the eastern part of the species range. It would also be interesting to ally this genetic data with ecological niche-based models, following a landscape approach.

Chapter 7: References

Albert EM, Zardoya R, García-París M (2007) Phylogeographical and speciation patterns in subterranean worm lizards of the genus Blanus (Amphisbaenia: Blanidae). *Molecular Ecology*, **16**, 1519–1531.

Amos W (2006) The hidden value of missing genotypes. *Molecular Biology and Evolution*, **23**, 1995–6.

Andreu A, Bea A, Braña F *et al.* (1998) *Fauna Ibérica. Volume 10 - Reptiles*. Museo Nacional de Ciencias Naturales-CSIC, Madrid.

Angers B, Bernatchez L (1998) Combined use of SMM and Non-SMM methods to infer fine structure and evolutionary history of closely related brook charr (*Salvelinus fontinalis*, Salmonidae) populations from microsatellites. *Molecular Biology and Evolution*, **15**, 143–159.

Avise JC (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution*, **43**, 1192–1208.

Avise JC, Bowen BW, Lamb T, Meylan AB, Bermingham E (1992) Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. *Molecular Biology and Evolution*, **9**, 457–473.

Babik W, Branicki W, Crnobrnja-Isailović J *et al.* (2005) Phylogeography of two European newt species-discordance between mtDNA and morphology. *Molecular Ecology*, **14**, 2475–2491.

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

Barnosky AD, Matzke N, Tomiya S *et al.* (2011) Has the Earth's sixth mass extinction already arrived? *Nature*, **471**, 51–57.

Barth D, Bernhard D, Fritzsch G, Fritz U (2004) The freshwater turtle genus *Mauremys* (Testudines, Geoemydidae) - a textbook example of an east-west disjunction or a taxonomic misconcept? *Zoologica Scripta*, **33**, 213–221.

Bidegaray-Batista L, Arnedo MA (2011) Gone with the plate: the opening of the Western Mediterranean basin drove the diversification of ground-dweller spiders. *BMC Evolutionary Biology*, **11**, 317.

Botkin DB, Saxe H, Araújo MB *et al.* (2007) Forecasting the Effects of Global Warming on Biodiversity. *BioScience*, **57**, 227–236.

Bour R (2008) Global diversity of turtles (Chelonii; Reptilia) in freshwater. *Hydrobiologia*, **595**, 593–598.

Bour R, Dubois A (1986) Nomenclature ordinale et familiale des Tortues (Reptilia). Note complémentaire. *Bulletin Mensuel de la Société Linnéenne de Lyon*, **55**, 87–90.

Bour R, Maran J (1999) Taxinomie de *Mauremys leprosa* (Schweigger, 1812) dans le sud du Maroc: la “tortue aux yeux bleus” (Reptilia, Chelonii, Geoemydidae). *Manouria*, **1**, 22–52.

Brandt P, Alpers W, Backhaus JO (1996) Study of the generation and propagation of internal waves in the Strait of Gibraltar using a numerical model and synthetic aperture radar images of the European ERS 1 satellite. *Journal of Geophysical Research*, **101**, 237–252.

Bromham L (2002) Molecular clocks in reptiles: life history influences rate of molecular evolution. *Molecular Biology and Evolution*, **19**, 302–309.

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

Busack S, Ernst C (1980) Variation in Mediterranean populations of *Mauremys* Gray 1869 (Reptilia, Testudines, Emydidae). *Annals of Carnegie Museum*, **49**, 251–264.

Butchart SHM, Walpole M, Collen B *et al.* (2010) Global biodiversity: indicators of recent declines. *Science*, **328**, 1164–1168.

Carranza S, Arnold EN (2004) History of West Mediterranean newts, *Pleurodeles* (Amphibia: Salamandridae), inferred from old and recent DNA sequences. *Systematics and Biodiversity*, **1**, 327–337.

Carranza S, Arnold EN, Pleguezuelos JM (2006a) 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.

Carranza S, Arnold EN, Wade E, Fahd S (2004) Phylogeography of the false smooth snakes, *Macroprotodon* (Serpentes, Colubridae): mitochondrial DNA sequences show European populations arrived recently from Northwest Africa. *Molecular Phylogenetics and Evolution*, **33**, 523–532.

Carranza S, Harris DJ, Arnold EN, Batista V, Gonzalez de la Vega JP (2006b) Phylogeography of the lacertid lizard, *Psammodromus algirus*, in Iberia and across the Strait of Gibraltar. *Journal of Biogeography*, **33**, 1279–1288.

Chiari Y, Cahais V, Galtier N, Delsuc F (2012) Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biology*, **10**, 65.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Coelho P, Sousa P, Harris DJ, van der Meijden A (2014) Deep intraspecific divergences in the medically relevant fat-tailed scorpions (*Androctonus*, Scorpiones). *Acta Tropica*, **134**, 43–51.

Cosson J-F, Hutterer R, Libois R *et al.* (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.

Cox N, Temple H (2009) *European red list of reptiles*. Office for Official Publications of the European Communities, Luxembourg.

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

Dijk PP van, Iverson JB, Bradley Shaffer H, Bour R, Rhodin AGJ (2012) Turtles of the World, 2012 Update: Annotated Checklist of Taxonomy, Synonymy, Distribution, and Conservation Status. In: *Conservation Biology of Freshwater Turtles and Tortoises: A Compilation Project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group. Chelonian Research Monographs*

Dijk PP van, Stuart B, Rhodin AGJ (2000) Asian Turtle Trade. In: *Proceedings of a Workshop on Conservation and Trade of Freshwater Turtles and Tortoises in Asia*, pp. 1–164. Chelonian Research Foundation, Lunenburg.

Drummond AJ, Suchard M a, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Duggen S, Hoernle K, Bogaard P van den, Rüpke L, Morgan JP (2003) Deep roots of the Messinian salinity crisis. *Nature*, **422**, 602–605.

Dynesius M, Jansson R (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 9115–9120.

Earl D a., vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Ecosystems and Human Well-being: Synthesis (2005) Island Press, Washington, DC.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**, 435–445.

Elorriaga J, Muñoz A-R (2013) Hybridisation between the Common Buzzard *Buteo buteo buteo* and the North African race of Long-legged Buzzard *Buteo rufinus cirtensis* in the Strait of Gibraltar: prelude or preclude to colonisation? *Ostrich*, **84**, 41–45.

Ernst JB, Barbour RW (1989) *Turtles of the World*. Smithsonian Institution Press, Washington.

Ernst CH, Hershey MF, Barbour RW (1974) A new coding system for hardshelled turtles. *Transactions of the Kentucky Academy of Science*, **35**, 27–28.

ESRI (2008) *ArcGIS 9: ArcMap Tutorial*. New York.

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

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Fèlix J, Budó J, Capalleras X, Mascort R (2006) The Fossil Register of the Genera *Testudo*, *Emys* and *Mauremys* of the Quaternary in Catalonia. In: *Chelonii 4*

Fonseca MM, Brito JC, Paulo OS, Carretero MA, 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.

Fritz U, Ayaz D, Buschbom J *et al.* (2008) Go east: phylogeographies of *Mauremys caspica* and *M. rivulata*- discordance of morphology, mitochondrial and nuclear genomic markers and rare hybridization. *Journal of Evolutionary Biology*, **21**, 527–540.

Fritz U, Barata M, Busack SD, Fritzsch G, Castilho R (2006) Impact of mountain chains, sea straits and peripheral populations on genetic and taxonomic structure of a freshwater turtle, *Mauremys leprosa* (Reptilia, Testudines, Geoemydidae). *Zoologica Scripta*, **35**, 97–108.

Fritz U, Fritzsch G, Lehr E, Ducotterd J (2005) The Atlas Mountains, not the Strait of Gibraltar, as a biogeographic barrier for *Mauremys leprosa* (Reptilia: Testudines). *Salamandra*, **41**, 97–106.

Fritz U, Harris DJ, Fahd S *et al.* (2009) Mitochondrial phylogeography of *Testudo graeca* in the Western Mediterranean: Old complex divergence in North Africa and recent arrival in Europe. *Amphibia-Reptilia*, **30**, 63–80.

Fritz U, Havaš P (2007) Checklist of chelonians of the world. *Vertebrate Zoology*, **57**, 149–368.

Fromhage L, Vences M, Veith M (2004) Testing alternative vicariance scenarios in Western Mediterranean discoglossid frogs. *Molecular Phylogenetics and Evolution*, **31**, 308–322.

Garcia-Castellanos D, Estrada F, Jiménez-Munt I *et al.* (2009) Catastrophic flood of the Mediterranean after the Messinian salinity crisis. *Nature*, **462**, 778–781.

Gibbons JW, Scott DE, Ryan TJ *et al.* (2000) The global decline of reptiles, déjà vu amphibians. *BioScience*, **50**, 653–666.

Gilbert SF, Corfe I (2013) Turtle origins: picking up speed. *Developmental Cell*, **25**, 326–328.

Godinho R, Crespo EG, Ferrand N (2008) The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Molecular Ecology*, **17**, 4670–4683.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography of Southern European Refugia* , pp. 155–188.

Gonçalves H, Martínez-Solano I, Pereira RJ *et al.* (2009) High levels of population subdivision in a morphologically conserved Mediterranean toad (*Alytes cisternasii*) result from recent, multiple refugia: evidence from mtDNA, microsatellites and nuclear genealogies. *Molecular Ecology*, **18**, 5143–5160.

Gong S-P, Chow AT, Fong JJ, Shi H-T (2009) The chelonian trade in the largest pet market in China: scale, scope and impact on turtle conservation. *Oryx*, **43**, 213–216.

Graciá E, Giménez A, Anadón JD *et al.* (2013) The uncertainty of Late Pleistocene range expansions in the western Mediterranean: a case study of the colonization of south-eastern Spain by the spur-thighed tortoise, *Testudo graeca*. *Journal of Biogeography*, **40**, 323–334.

Griswold CK, Baker AJ (2002) Time to the Most Recent Common Ancestor and Divergence Times of Populations of Common Chaffinches (*Fringilla Coelebs*) in Europe and North Africa: Insights Into Pleistocene Refugia and Current Levels of Migration. *Evolution*, **56**, 143–153.

Guillon J, Guéry L, Hulin V, Girondot M (2012) A large phylogeny of turtles (Testudines) using molecular data. *Contributions to Zoology*, **81**, 147–158.

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

Habel JC, Lens L, Rödder D, 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.

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

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

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

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

Hewitt GM (2011a) Mediterranean Peninsulas: The Evolution of Hotspots. In: *Biodiversity Hotspots* (eds Zachos FE, Habel JC), pp. 123 – 147. Springer-Verlag, Berlin Heidelberg.

Hewitt GM (2011b) Quaternary phylogeography: the roots of hybrid zones. *Genetica*, **139**, 617–38.

Hidalgo-Vila J, Díaz-Paniagua C, Pérez-Santigosa N, de Frutos-Escobar C, Herrero-Herrero A (2008) Salmonella in free-living exotic and native turtles and in pet exotic turtles from SW Spain. *Research in Veterinary Science*, **85**, 449–452.

Huey RB, Kingsolver JG (1993) Evolution of resistence to high temperature in ectotherms. *The American Naturalist*, **142**, S21–S46.

Hugall AF, Foster R, Lee MSY (2007) Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Systematic Biology*, **56**, 543–563.

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

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Jansen RP (2000) Origin and persistence of the mitochondrial genome. *Human Reproduction*, **15**, 1–10.

Kalinowski ST (2005) HP-RARE 1.0: A computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kaliontzopoulou A, Pinho C, Harris DJ, Carretero MA (2011) When cryptic diversity blurs the picture: a cautionary tale from Iberian and North African *Podarcis* wall lizards. *Biological Journal of the Linnean Society*, **103**, 779–800.

Keller C (1997) Ecologia de las poblaciones de *Mauremys leprosa* y *Emys orbicularis* del Parque Nacional de Doñana, Unpubl. Dissertation, Universität Sevilla, Spanien.

Keller C, Busack SD (2001) *Mauremys leprosa* (Schweigger, 1812)—Maurische Bachschildkröte. *Handbuch der Reptilien und Amphibien Europas*, **3**, 57–88.

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

Lenk P, Fritz U, Joger U, Wink M (1999) Mitochondrial phylogeography of the European pond turtle, *Emys orbicularis* (Linnaeus 1758). *Molecular Ecology*, **8**, 1911–1922.

Lenk P, Wink M (1997) A RNA/RNA heteroduplex cleavage analysis to detect rare mutations in populations. *Molecular Ecology*, **6**, 687–690.

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

Lourenço JM, Claude J, Galtier N, Chiari Y (2012) Dating cryptodiran nodes: origin and diversification of the turtle superfamily Testudinoidea. *Molecular Phylogenetics and Evolution*, **62**, 496–507.

Lourenço JM, Glémin S, Chiari Y, Galtier N (2013) The determinants of the molecular substitution process in turtles. *Journal of Evolutionary Biology*, **26**, 38–50.

Lovich JE, Znari M, Baamrane MA ait, Naimi M, Mostalih A (2010) Biphasic Geographic Variation in Sexual Size Dimorphism of Turtle (*Mauremys leprosa*) Populations Along an Environmental Gradient in Morocco. *Chelonian Conservation and Biology*, **9**, 45–53.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Magri D, Fineschi S, Bellarosa R *et al.* (2007) The distribution of *Quercus suber* chloroplast haplotypes matches the palaeogeographical history of the western Mediterranean. *Molecular Ecology*, **16**, 5259–5266.

Mantziou G, Antoniou A, Poulakakis N *et al.* (2005) Isolation and characterization of six polymorphic microsatellite markers in the freshwater turtle *Mauremys rivulata* (Testudines: Geoemydidae). *Molecular Ecology Notes*, **5**, 727–729.

Martínez-Solano Í (2004) Phylogeography of Iberian *Discoglossus* (Anura: Discoglossidae). *Journal of Zoological Systematics and Evolution Research*, **42**, 298–305.

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

Martínez-Solano I, Teixeira J, Buckley D, García-París M (2006) Mitochondrial DNA phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple refugia in an Iberian endemic. *Molecular Ecology*, **15**, 3375–88.

Mesquita N, Hänfling B, Carvalho GR, Coelho MM (2005) Phylogeography of the cyprinid *Squalius aradensis* and implications for conservation of the endemic freshwater fauna of southern Portugal. *Molecular Ecology*, **14**, 1939–1954.

Miraldo A, Faria C, Hewitt GM, Paulo OS, Emerson BC (2013) Genetic analysis of a contact zone between two lineages of the ocellated lizard (*Lacerta lepida* Daudin 1802) in south-eastern Iberia reveal a steep and narrow hybrid zone. *Journal of Zoological Systematics and Evolutionary Research*, **51**, 45–54.

Miraldo A, Hewitt GM, Paulo OS, Emerson BC (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, **11**, 170.

Moll D, Moll E (2004) *The Ecology, Exploitation, and Conservation of River Turtles*. Oxford University Press, New York.

Muñoz A, Nicolau B (2006) Sexual Dimorphism and Allometry in the Stripe-Necked Terrapin, *Mauremys leprosa*, in Spain. *Chelonian Conservation and Biology*, **5**, 87–92.

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

Natusch DJD, Lyons J a. (2012) Exploited for pets: the harvest and trade of amphibians and reptiles from Indonesian New Guinea. *Biodiversity and Conservation*, **21**, 2899–2911.

Near TJ, Meylan P a, Shaffer HB (2005) Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *The American Naturalist*, **165**, 137–46.

Near TJ, Meylan PA, Shaffer HB (2008) Caveats on the Use of Fossil Calibrations for Molecular Dating: A Reply to Parham and Irmis. *The American Naturalist*, **171**, 137–140.

Nei M, Kumar S (2000) *Molecular Evolutionand Phylogenetics*.

Oetting W, Lee H, Flanders D *et al.* (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics*, **30**, 450–458.

Oliver MA, Webster R (1990) Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information Systems*, **4**, 313–332.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. *Molecular Ecology Notes*, **4**, 535–538.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568–583.

Papenfuss TJ (1969) Preliminary analysis of the reptiles of arid Central West Africa. *Wasmann j. biol. Wasmann journal of biology.*, **27**, 249–325.

Parham JF, Shi H (2001) The Discovery of *Mauremys iversoni*-like Turtles at a Turtle Farm in Hainan Province, China: The Counterfeit Golden Coin. *Asiatic Herpetological Research*, **9**, 71–76.

Parham JF, Simison WB, Kozak KH, Feldman CR, Shi H (2001) New Chinese turtles: endangered or invalid? A reassessment of two species using mitochondrial DNA, allozyme electrophoresis and known-locality specimens. *Animal Conservation*, **4**, 357–367.

Paulo OS, Pinto I, Bruford MW, Jordan WC, Nichols RA (2002) The double origin of Iberian peninsular chameleons. *Biological Journal of the Linnean Society*, 1–7.

Peakall R, Smouse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pedall I, Fritz U, Stuckas H, Valdeón A, Wink M (2011) Gene flow across secondary contact zones of the *Emys orbicularis* complex in the Western Mediterranean and evidence for extinction and re-introduction of pond turtles on Corsica and Sardinia (Testudines: Emydidae). *Journal of Zoological Systematics and Evolutionary Research*, **49**, 44–57.

Pinho C, Harris DJ, Ferrand N (2007) Contrasting patterns of population subdivision and historical demography in three western Mediterranean lizard species inferred from mitochondrial DNA variation. *Molecular Ecology*, **16**, 1191–205.

Polo-Cavia N, López P, Martín J (2011) Aggressive interactions during feeding between native and invasive freshwater turtles. *Biological Invasions*, **13**, 1387–1396.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–59.

Rands MRW, Adams WM, Bennun L *et al.* (2010) Biodiversity conservation: challenges beyond 2010. *Science*, **329**, 1298–303.

Rato C, Carranza S, Harris DJ (2012) Evolutionary history of the genus *Tarentola* (Gekkota: Phyllodactylidae) from the Mediterranean Basin, estimated using multilocus sequence data. *BMC Evolutionary Biology*, **12**, 14.

Recuero E, Iraola A, Rubio X, Machordom A, García-París M (2007) Mitochondrial differentiation and biogeography of *Hyla meridionalis* (Anura: Hylidae): an unusual phylogeographical pattern. *Journal of Biogeography*, **34**, 1207–1219.

Rice WR (1989) Analyzing Tables of Statistical Tests. *Evolution*, **43**, 223–225.

Roques S, Díaz-Paniagua C, Portheault A, Pérez-Santigosa N, Hidalgo-Vila J (2006) Sperm storage and low incidence of multiple paternity in the European pond turtle, *Emys orbicularis*: A secure but costly strategy? *Biological Conservation*, **129**, 236–243.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–6.

Sala OE, Chapin FS, Armesto JJ *et al.* (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–4.

Santos X, Rato C, Carranza S, Carretero M a., Pleguezuelos JM (2012) Complex phylogeography in the Southern Smooth Snake (*Coronella girondica*) supported by mtDNA sequences. *Journal of Zoological Systematics and Evolutionary Research*, **50**, 210–219.

Schleich HH (1996) Beitrag zur Systematik des Formenkreises von *Mauremys leprosa* (Schweigger) in Marokko. Teil 1. (Reptilia, Chelonii, Emydidae). *Spixiana Supplement*, **22**, 29–59.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma*, **109**, 365–371.

Schlötterer C (2004) The evolution of molecular markers-just a matter of fashion? *Nature reviews. Genetics*, **5**, 63–9.

Secreteriat of the Convention on Biological Diversity (2006) *Global Biodiversity Outlook 2*. Montréal.

Secreteriat of the Convention on Biological Diversity (2010) *Global Biodiversity Outlook 3*. Montréal.

Selkoe K a, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–629.

Sequeira F, Alexandrino J, Rocha S, Arntzen JW, Ferrand N (2005) Genetic exchange across a hybrid zone within the Iberian endemic golden-striped salamander, *Chioglossa lusitanica*. *Molecular Ecology*, **14**, 245–54.

Shaffer HB (2009) Turtles (Testudines). In: *The Timetree of Life* (eds Hedges SB, Kumar S), pp. 398–401. Oxford University Press.

Da Silva E (2002) Mauremys leprosa (Schweiger, 1812). *Atlas y Libro Rojo de Los Anfibios y Reptiles de España*, 143–146.

Soler BG de, Vall-llosera GC, Made J van der *et al.* (2012) A new key locality for the Pliocene vertebrate record of Europe: the Camp dels Ninots maar (NE Spain). *Geologica Acta*, **10**, 1–17.

Sousa P, Harris DJ, Froufe E, Meijden A Van Der (2012) Phylogeographic patterns of *Buthus* scorpions (Scorpiones: Buthidae) in the Maghreb and South-Western Europe based on CO1 mtDNA sequences. , **288**, 66–75.

Spinks PQ, Bradley Shaffer H, Iverson JB, McCord WP (2004) Phylogenetic hypotheses for the turtle family Geoemydidae. *Molecular Phylogenetics and Evolution*, **32**, 164–182.

Spinks PQ, Shaffer HB (2006) Conservation phylogenetics of the Asian box turtles (Geoemydidae, *Cuora*): mitochondrial introgression, numts, and inferences from multiple nuclear loci. *Conservation Genetics*, **8**, 641–657.

Stewart JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings. Biological sciences / The Royal Society*, **277**, 661–71.

Stuckas H, Velo-Antón G, Fahd S *et al.* (2014) Where are you from, stranger? The enigmatic biogeography of North African pond turtles (*Emys orbicularis*). *Organisms Diversity & Evolution*.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular ecology*, **7**, 453–464.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–9.

Thomas C, Cameron A, Green R (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.

TURTLE CONSERVATION COALITION [Rhodin A, Walde AD, Horne BD *et al.* (2011) Turtles in Trouble: The World's 25+ Most Endangered Tortoises and Freshwater Turtles, 2011. In:. Lunenburg, MA.

Vamberger M, Stuckas H, Ayaz D *et al.* (2014) Massive transoceanic gene flow in a freshwater turtle (Testudines: Geoemydidae: *Mauremys rivulata*). *Zoologica Scripta*, **000**, 000–000.

Vamberger M, Stuckas H, Fritz U (2011) Fifteen microsatellite markers for the stripe-necked terrapin *Mauremys caspica* (Testudines: Geoemydidae) and cross-amplification tests in M. rivulata. *Conservation Genetics Resources*, **3**, 87–89.

Veith M, Mayer C, Samraoui B, Barroso DD, Bogaerts S (2004) From Europe to Africa and vice versa: evidence for multiple intercontinental dispersal in ribbed salamanders (Genus *Pleurodeles*). *Journal of Biogeography*, **31**, 159–171.

Velo-Antón G, Becker CG, Cordero-Rivera A (2011a) Turtle carapace anomalies: the roles of genetic diversity and environment. *PloS one*, **6**, e18714.

Velo-Antón G, García-París M, Cordero Rivera A (2008) Patterns of nuclear and mitochondrial DNA variation in Iberian populations of *Emys orbicularis* (Emydidae): conservation implications. *Conservation Genetics*, **9**, 1263–1274.

Velo-Antón G, Godinho R, Ayres C, Ferrand N, Rivera AC (2007) Assignment tests applied to relocate individuals of unknown origin in a threatened species, the European pond turtle (*Emys orbicularis*). *Amphibia-Reptilia*, **28**, 475–484.

Velo-Antón G, Godinho R, Harris DJ *et al.* (2012) Deep evolutionary lineages in a Western Mediterranean snake (*Vipera latastei/monticola* group) and high genetic structuring in Southern Iberian populations. *Molecular Phylogenetics and Evolution*, **65**, 965–73.

Velo-Antón G, Wink M, Schneeweiß N, Fritz U (2011b) Native or not? Tracing the origin of wild-caught and captive freshwater turtles in a threatened and widely distributed species (*Emys orbicularis*). *Conservation Genetics*, **12**, 583–588.

Veríssimo J, Velo-Antón G, Lopes S *et al.* (2013) Cross-amplification of microsatellite loci for the Mediterranean stripe-necked terrapin (*Mauremys leprosa*). *Amphibia-Reptilia*, **34**, 259–262.

Verneau O, Palacios C, Platt T *et al.* (2011) Invasive species threat: parasite phylogenetics reveals patterns and processes of host-switching between non-native and native captive freshwater turtles. *Parasitology*, **138**, 1778–92.

Wake DB, Vredenburg VT (2008) Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences of the United States of America*, **105 Suppl** , 11466–73.

Wan Q-H, Wu H, Fujihara T, Fang S-G (2004) Which genetic marker for which conservation genetics issue? *Electrophoresis*, **25**, 2165–76.

Wilson E (1988) *Biodiversity*. National Academy of Sciences/ Smithsonian Institution, Washington.

Zazo C (1999) Interglacial sea levels. *Quaternary International*, **55**, 101–113.

Zhang D, Hewitt G (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, **11**, 247–251.

Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–84.

Chapter 8: Supplementary Material

Table S. M. 1: Summary table of the *Mauremys leprosa'* samples used for Chapter 3 (Manuscript I). Lineage and haplotype outcome resulted from the analysis of the concatenated dataset.

Country	Locality	Samples	Latitude	Longitude	Ref	Lineage	Haplotype
Algeria	Skikda Wilayat: Hadjar Ediss, 20 km SSW Annaba	MTD T 1222	36.797	7.608	MVZ Herp 235704	*M. l. saharica*	B3-2
Algeria	El Kala	MTD T 3249	36.886	8.455	This study	*M. l. saharica*	B3-1
Morocco	Tan Tan (Guelta Ez Zerga)	GVA3168	28.497	-10.886	This study	*M. l. saharica*	B2-3
Morocco	Tan Tan (Guelta Ez Zerga)	GVA3169	28.497	-10.886	This study	*M. l. saharica*	B2-4
Morocco	Agadir: Loukkos	GVA3525	35.210	-6.130	This study	*M. l. leprosa*	A2-1
Morocco	15km N fo Tan Tan (Draa)	MTD T 6869	28.567	-11.067	This study	*M. l. saharica*	B2-2
Morocco	Agadir: 17km E Guelmine	MTD T 784	28.974	-9.904	(Fritz et al., 2006)	*M. l. saharica*	B1-4
Morocco	Agadir: 17km E Guelmine	MTD T 780	28.974	-9.904	(Fritz et al., 2006)	*M. l. saharica*	B1-6
Morocco	Agadir: N Tiliouine (Oued Noun Canyon)	MTD T 785	29.085	-10.252	(Fritz et al., 2006)	*M. l. saharica*	B1-5
Morocco	Agadir: 18km S Tata, Oued Tata near El-Khemis	MTD T 774	29.592	-8.000	(Fritz et al., 2006)	*M. l. saharica*	B2-6
Morocco	Sidi El Mehadou,i 13km S of Tata	MTD T 6876	29.617	-7.983	This study	*M. l. saharica*	B2-6
Morocco	Agadir: SE Tissint (Oued Tissint)	MTD T 773	29.852	-7.255	(Fritz et al., 2006)	*M. l. saharica*	B2-6
Morocco	Agadir: 3km SW Taroundannt (Oued Souss Valley)	MTD T 1204	30.440	-8.899	(Fritz et al., 2006)	*M. l. saharica*	B1-1
Morocco	Ouarzazate: Tamnougalt (Oued Drâa)	MTD T 770	30.670	-6.381	(Fritz et al., 2006)	*M. l. saharica*	B2-6
Morocco	Ouarzzazate	GVA457	30.969	-6.724	This study	*M. l. saharica*	B1-2
Morocco	Ouarzzazate	GVA455	30.969	-6.724	This study	*M. l. saharica*	B1-3
Morocco	Ouarzzazate	GVA456	30.969	-6.724	This study	*M. l. saharica*	B2-3
Morocco	Ouarzzazate	GVA453	30.969	-6.724	This study	*M. l. saharica*	B2-5
Morocco	12km N of Timezgadiouine	GVA2707	30.990	-9.040	This study	*M. l. leprosa*	A1-7
Morocco	Marrakech: near Aït-Ourir	MTD T 765	31.468	-7.766	(Fritz et al., 2006)	*M. l. leprosa*	A1-10
Morocco	Marrakech: near Aït-Ourir	MTD T 764	31.468	-7.766	(Fritz et al., 2006)	*M. l. leprosa*	A1-11
Morocco	Marrakech Province (Oued Zat)	GVA462	31.530	-7.563	This study	*M. l. leprosa*	A1-1
Morocco	Marrakech Province (Oued Zat)	GVA460	31.530	-7.563	This study	*M. l. leprosa*	A1-2
Morocco	Douer Targa	GVA1327	31.530	-7.563	This study	*M. l. leprosa*	A1-3
Morocco	Douer Targa	GVA1326	31.530	-7.563	This study	*M. l. leprosa*	A1-4
Morocco	Douer Targa	GVA1325	31.530	-7.563	This study	*M. l. leprosa*	A1-6

Morocco	Marrakech Province (Oued Zat)	GVA459	31.530	-7.563	This study	*M. l. leprosa*	A1-6
Morocco	Marrakech Province (Oued Zat)	GVA461	31.530	-7.563	This study	*M. l. leprosa*	A1-6
Morocco	Douer Targa	GVA1323	31.530	-7.563	This study	*M. l. leprosa*	A1-9
Morocco	Ksar-es-Souk: Aoufous (Oued Ziz)	MTD T 1908	31.682	-4.183	(Fritz et al., 2006)	*M. l. saharica*	B2-1
Morocco	Ksar-es-Souk: Aoufous (Oued Ziz)	MTD T 1909	31.682	-4.183	(Fritz et al., 2006)	*M. l. saharica*	B2-1
Morocco	Marrakech (Ouad Tansift)	GVA451	31.689	-7.990	This study	*M. l. leprosa*	A1-5
Morocco	Marrakech (Ouad Tansift)	GVA452	31.689	-7.990	This study	*M. l. leprosa*	A1-8
Morocco	Sidi Mimoun	GVA1351	32.199	-6.302	This study	*M. l. leprosa*	A2-3
Morocco	Sidi Mimoun	GVA1349	33.001	-5.648	This study	*M. l. leprosa*	A2-2
Morocco	Boulemane (Oued Sebb Ousfa)	GVA3179	33.119	-4.329	This study	*M. l. saharica*	B3-17
Morocco	Boulemane (Oued Sebb Ousfa)	GVA3180	33.119	-4.329	This study	*M. l. saharica*	B3-18
Morocco	Sidi Mimoun	GVA1350	33.548	-5.098	This study	*M. l. saharica*	B3-10
Morocco	Sidi Mimoun	GVA1343	33.649	-4.968	This study	*M. l. saharica*	B3-8
Morocco	Sidi Mimoun	GVA1344	33.649	-4.968	This study	*M. l. saharica*	B3-9
Morocco	Douira	GVA2713	34.573	-2.733	This study	*M. l. saharica*	B3-4
Morocco	Zoumi	GVA2710	34.746	-5.423	This study	*M. l. leprosa*	A3-6
Morocco	Zoumi	GVA2711	34.746	-5.423	This study	*M. l. saharica*	B4-2
Morocco	Fifi	GVA1307	35.023	-5.205	This study	*M. l. leprosa*	A3-14
Morocco	Fifi	GVA1306	35.023	-5.205	This study	*M. l. saharica*	B4-1
Morocco	Fifi	GVA1308	35.023	-5.205	This study	*M. l. saharica*	B4-1
Morocco	Tétouan: 11km E Chefchaouene	MTD T 1211	35.074	-5.215	(Fritz et al., 2006)	*M. l. leprosa*	A3-10
Morocco	Tétouan: NW Chefchaouene	MTD T 1210	35.278	-5.459	(Fritz et al., 2006)	*M. l. leprosa*	A3-10
Morocco	Sidi Mimoun	GVA1348	35.299	-5.219	This study	*M. l. leprosa*	A3-14
Morocco	Tazia	GVA2520	35.342	-5.552	This study	*M. l. leprosa*	A3-18
Morocco	Tazia	GVA2523	35.342	-5.552	This study	*M. l. saharica*	B4-2
Morocco	Tazia	GVA2544	35.342	-5.552	This study	*M. l. saharica*	B4-2
Morocco	Tétouan: 14.7km S Asilah	MTD T 1205	35.370	-6.060	(Fritz et al., 2006)	*M. l. leprosa*	A3-10
Morocco	Tétouan: 14.7km S Asilah	MTD T 1206	35.370	-6.060	(Fritz et al., 2006)	*M. l. leprosa*	A3-10

Morocco	Tétouan: 14.7km S Asilah	MTD T 1209	35.370	-6.060	(Fritz et al., 2006)	*M. l. saharica*	B4-2
Morocco	Tétouan: 9.9km E Ksar-es-Srhir	MTD T 1213	35.848	-5.453	(Fritz et al., 2006)	*M. l. leprosa*	A3-9
Portugal	Algarve	MI0104	37.020	-7.886	This study	*M. l. leprosa*	A3-1
Portugal	Algarve	MI0114	37.020	-7.886	This study	*M. l. leprosa*	A3-15
Portugal	Vilamoura	MTD T 1286	37.093	-8.129	(Fritz et al., 2006)	*M. l. leprosa*	A3-8
Portugal	Aljezur	Aljezur 4	37.324	-8.801	(Fritz et al., 2006)	*M. l. leprosa*	A3-8
Portugal	Almograve	GVA2106	37.653	-8.794	This study	*M. l. leprosa*	A3-1
Portugal	Almograve	GVA2105	37.653	-8.794	This study	*M. l. leprosa*	A3-5
Portugal	Castroverde	GVA2179	37.694	-8.086	This study	*M. l. leprosa*	A3-1
Portugal	Castroverde	GVA2187	37.694	-8.086	This study	*M. l. leprosa*	A3-1
Portugal	Paul da Tornada	MI0312	39.445	-9.138	This study	*M. l. leprosa*	A3-1
Portugal	Paul da Tornada	MI0313	39.445	-9.138	This study	*M. l. leprosa*	A3-1
Portugal	Castelo Branco	MI0230	39.703	-7.308	This study	*M. l. leprosa*	A3-1
Portugal	Castelo Branco: M. Galisteu	MI0219	39.703	-7.308	This study	*M. l. leprosa*	A3-23
Portugal	Castelo Branco: Monte Barata	MI0224	39.703	-7.308	This study	*M. l. leprosa*	A3-8
Portugal	Vila Flôr	MTD T 1597	41.295	-7.173	(Fritz et al., 2006)	*M. l. leprosa*	A3-11
Spain	Ceuta	MTD T 1299	35.890	-5.349	(Fritz et al., 2006)	*M. l. leprosa*	A3-13
Spain	Ceuta	GVA2481	35.897	-5.349	This study	*M. l. leprosa*	A3-15
Spain	Ceuta	GVA2480	35.897	-5.349	This study	*M. l. leprosa*	A3-17
Spain	Ceuta	GVA2498	35.897	-5.349	This study	*M. l. leprosa*	A3-3
Spain	Cádiz: Facinas	MTD T 1217	36.133	-5.700	(Fritz et al., 2006)	*M. l. leprosa*	A3-9
Spain	Cádiz: Facinas	MTD T 1218	36.133	-5.617	(Fritz et al., 2006)	*M. l. leprosa*	A3-9
Spain	Málaga	GVA3521	36.756	-5.290	This study	*M. l. leprosa*	A3-15
Spain	Málaga	GVA3520	36.756	-5.290	This study	*M. l. leprosa*	A3-16
Spain	Huelva: Doñana National Park	GVA2120	37.049	-6.591	This study	*M. l. leprosa*	A3-20
Spain	Huelva: Doñana National Park	GVA2129	37.049	-6.591	This study	*M. l. leprosa*	A3-20
Spain	Granada: Brácana	GVA3527	37.218	-3.953	This study	*M. l. leprosa*	A3-1
Spain	Sevilla: Fuentes de Andalucia	GVA3526	37.406	-5.452	This study	*M. l. leprosa*	A3-15

Spain	Murcia: Mazarrón	GVA2714	37.582	-1.409	This study	*M. l. leprosa*	A3-1
Spain	Murcia: Mazarrón	GVA2715	37.582	-1.409	This study	*M. l. leprosa*	A3-28
Spain	Jaén: Andújar	GVA3144	38.152	-4.015	This study	*M. l. leprosa*	A3-27
Spain	Córdoba: Cardeña	GVA3111	38.258	-4.324	This study	*M. l. leprosa*	A3-1
Spain	Córdoba: Cardeña	GVA3112	38.258	-4.324	This study	*M. l. leprosa*	A3-1
Spain	Murcia	GVA2026	38.294	-1.432	This study	*M. l. leprosa*	A3-1
Spain	Ciudad Real	GVA2142	38.908	-4.472	This study	*M. l. leprosa*	A3-1
Spain	Ciudad Real	GVA2138	38.908	-4.472	This study	*M. l. leprosa*	A3-21
Spain	Valencia: Chiva	MTD T 1435	39.467	-0.717	(Fritz et al., 2006)	*M. l. leprosa*	A3-8
Spain	Valencia: Serpis	Ml0714	39.470	-0.377	This study	*M. l. leprosa*	A3-19
Spain	Valencia: Peñíscola	Ml0701	39.470	-0.377	This study	*M. l. leprosa*	A3-25
Spain	Ávila: Poyales del Hoyo	GVA2265	40.158	-5.162	This study	*M. l. leprosa*	A3-27
Spain	Madrid	GVA2033	40.488	-4.124	This study	*M. l. leprosa*	A3-24
Spain	Madrid	GVA2043	40.488	-4.124	This study	*M. l. leprosa*	A3-24
Spain	Guadalajara	GVA2063	40.548	-3.257	This study	*M. l. leprosa*	A3-22
Spain	Guadalajara	GVA2062	40.548	-3.257	This study	*M. l. leprosa*	A3-26
Spain	Salamanca: Vilvestre	GVA2226	41.125	-6.716	This study	*M. l. leprosa*	A3-1
Spain	Salamanca: Vilvestre	GVA2227	41.125	-6.716	This study	*M. l. leprosa*	A3-1
Spain	Barcelona: Delta del Llobregat	GVA2296	41.287	2.016	This study	*M. l. leprosa*	A3-1
Spain	Barcelona: Delta del Llobregat	GVA2297	41.287	2.016	This study	*M. l. leprosa*	A3-12
Spain	Ourense: As Neves	GVA2084	42.081	-8.397	This study	*M. l. leprosa*	A3-1
Spain	Ourense: Arnoia river	GVA2092	42.240	-7.699	This study	*M. l. leprosa*	A3-2
Spain	Ourense: Ribadavia	GVA2082	42.288	-8.143	This study	*M. l. leprosa*	A3-1
Spain	Ourense: Ribadavia	GVA2096	42.288	-8.143	This study	*M. l. leprosa*	A3-1
Spain	Ourense: Ribadavia	GVA2097	42.288	-8.143	This study	*M. l. leprosa*	A3-1
Spain	Girona: Albera	GVA2320	42.378	3.031	This study	*M. l. leprosa*	A3-1
Spain	Girona: Albera	GVA2321	42.378	3.031	This study	*M. l. leprosa*	A3-1
Spain	Gerona: Orlina river, Rabos d'Empordà	MTD T 1571	42.379	3.028	(Fritz et al., 2006)	*M. l. leprosa*	A3-7

Spain	Gerona: Orlina river, Rabos d'Empordà	MTD T 1577	42.379	3.028	(Fritz et al., 2006)	*M. l. leprosa*	A3-7
Spain	Araba: Poza Tertanga	GVA3554	42.982	-3.018	This study	*M. l. leprosa*	A3-1
Spain	Bizkaia: Humedal Bolue	GVA3553	43.346	-2.995	This study	*M. l. leprosa*	A3-4
Spain	Lugo: Ribadeo	GVA2091	43.638	-7.615	This study	*M. l. leprosa*	A3-1
Tunisia	Al Watan al Quibli: Hammamet	MTD T 1360	36.401	10.583	(Fritz et al., 2006)	*M. l. saharica*	B3-19
Tunisia	Firnanah, Djebel Rmila	MTD T 8442	36.593	8.640	This study	*M. l. saharica*	B3-1
Tunisia	Firnanah, Djebel Rmila	MTD T 8443	36.593	8.640	This study	*M. l. saharica*	B3-1
Tunisia	Firnanah, Djebel Rmila	MTD T 8444	36.593	8.640	This study	*M. l. saharica*	B3-1
Tunisia	Firnanah, Djebel Rmila	MTD T 8445	36.593	8.640	This study	*M. l. saharica*	B3-1
Tunisia	Firnanah, Djebel Rmila	MTD T 8446	36.593	8.640	This study	*M. l. saharica*	B3-1
Tunisia	Firnanah, Oued Ghrib	MTD T 8450	36.616	8.686	This study	*M. l. saharica*	B3-1
Tunisia	Firnanah, Oued Ghrib	MTD T 8453	36.616	8.686	This study	*M. l. saharica*	B3-1
Tunisia	Firnanah, Oued Ghrib	MTD T 8455	36.616	8.686	This study	*M. l. saharica*	B3-1
Tunisia	Firnanah, Oued Ghrib	MTD T 8452	36.616	8.686	This study	*M. l. saharica*	B3-2
Tunisia	Firnanah, Oued Ghrib	MTD T 8451	36.616	8.686	This study	*M. l. saharica*	B3-6
Tunisia	Firnanah, Oued Ghrib	MTD T 8454	36.616	8.686	This study	*M. l. saharica*	B3-6
Tunisia	Firnanah, Oued Ghezala	MTD T 8447	36.643	8.700	This study	*M. l. saharica*	B3-1
Tunisia	Firnanah, Oued Ghezala	MTD T 8448	36.643	8.700	This study	*M. l. saharica*	B3-1
Tunisia	Firnanah, Oued Ghezala	MTD T 8449	36.643	8.700	This study	*M. l. saharica*	B3-1
Tunisia	Douaar Zaaba, Oued Bzhig	MTD T 8435	36.720	10.621	This study	*M. l. saharica*	B3-1
Tunisia	Douaar Zaaba, Oued Bzhig	MTD T 8436	36.720	10.621	This study	*M. l. saharica*	B3-1
Tunisia	Douaar Zaaba, Oued Bzhig	MTD T 8437	36.720	10.621	This study	*M. l. saharica*	B3-1
Tunisia	Douaar Zaaba, Oued Bzhig	MTD T 8438	36.720	10.621	This study	*M. l. saharica*	B3-1
Tunisia	Douaar Zaaba, Oued Bzhig	MTD T 8439	36.720	10.621	This study	*M. l. saharica*	B3-1
Tunisia	Douaar Zaaba, Oued Bzhig	MTD T 8440	36.720	10.621	This study	*M. l. saharica*	B3-1
Tunisia	Douaar Zaaba, Oued Bzhig	MTD T 8441	36.720	10.621	This study	*M. l. saharica*	B3-1
Tunisia	Manzil Hurr, Barrage Lebna	MTD T 8420	36.740	10.922	This study	*M. l. saharica*	B3-1
Tunisia	Manzil Hurr, Barrage Lebna	MTD T 8421	36.740	10.922	This study	*M. l. saharica*	B3-1

Tunisia	Manzil Hurr, Barrage Lebna	MTD T 8422	36.740	10.922	This study	*M. l. saharica*	B3-1
Tunisia	Manzil Hurr, Barrage Lebna	MTD T 8423	36.740	10.922	This study	*M. l. saharica*	B3-1
Tunisia	Manzil Hurr, Barrage Lebna	MTD T 8426	36.740	10.922	This study	*M. l. saharica*	B3-1
Tunisia	Manzil Hurr, Barrage Lebna	MTD T 8427	36.740	10.922	This study	*M. l. saharica*	B3-1
Tunisia	Manzil Hurr, Barrage Lebna	MTD T 8419	36.740	10.922	This study	*M. l. saharica*	B3-5
Tunisia	Manzil Hurr, Barrage Lebna	MTD T 8424	36.740	10.922	This study	*M. l. saharica*	B3-7
Tunisia	Aïn Draham (Oued Sidi Youssef)	MTD T 4177	36.777	8.687	This study	*M. l. saharica*	B3-1
Tunisia	Aïn Draham (Oued Sidi Youssef)	MTD T 4179	36.777	8.687	This study	*M. l. saharica*	B3-1
Tunisia	Aïn Draham (Oued Sidi Youssef)	MTD T 4180	36.777	8.687	This study	*M. l. saharica*	B3-1
Tunisia	Douaar Zaaba, Barrage El Abid	MTD T 8434	36.811	10.699	This study	*M. l. saharica*	B3-1
Tunisia	Al Makhzan (Oued El Abid)	MTD T 8428	36.867	10.725	This study	*M. l. saharica*	B3-1
Tunisia	Al Makhzan (Oued El Abid)	MTD T 8429	36.867	10.725	This study	*M. l. saharica*	B3-1
Tunisia	Al Makhzan (Oued El Abid)	MTD T 8430	36.867	10.725	This study	*M. l. saharica*	B3-1
Tunisia	Al Makhzan (Oued El Abid)	MTD T 8431	36.867	10.725	This study	*M. l. saharica*	B3-1
Tunisia	Al Makhzan (Oued El Abid)	MTD T 8432	36.867	10.725	This study	*M. l. saharica*	B3-11
Tunisia	Al Makhzan (Oued El Abid)	MTD T 8433	36.867	10.725	This study	*M. l. saharica*	B3-15
Tunisia	Tunisia (Oued Ordha)	MTD T 4178	36.903	9.110	This study	*M. l. saharica*	B3-3
Tunisia	Dawwar Mraf (Oued Sejenane)	MTD T 8459	37.121	9.263	This study	*M. l. saharica*	B3-1
Tunisia	Dawwar Mraf (Oued Sejenane)	MTD T 8456	37.121	9.263	This study	*M. l. saharica*	B3-12
Tunisia	Dawwar Mraf (Oued Sejenane)	MTD T 8457	37.121	9.263	This study	*M. l. saharica*	B3-12
Tunisia	Dawwar Mraf (Oued Sejenane)	MTD T 8458	37.121	9.263	This study	*M. l. saharica*	B3-12
Tunisia	Sidi Ferdjani (Oued Serrat)	MTD T 8460	37.206	9.232	This study	*M. l. saharica*	B3-13
Tunisia	Sidi Ferdjani (Oued Serrat)	MTD T 8461	37.206	9.232	This study	*M. l. saharica*	B3-13
Tunisia	Sidi Ferdjani (Oued Serrat)	MTD T 8462	37.206	9.232	This study	*M. l. saharica*	B3-14
Tunisia	Sidi Ferdjani (Oued Serrat)	MTD T 8463	37.206	9.232	This study	*M. l. saharica*	B3-16

R code 1: Pairwise Euclidean Distances for Chapter 3 interpolations

```
distances <- function (dataTemplate) {
  data <- dataTemplate
  #Transform both Lat and Long variables into numeric
  data$Lat <- as.numeric(data$Lat)
  data$Long <- as.numeric(data$Long)

  #Creates a vector with the row names to be used
  rNames <- c('distance', data$Sample)

  #Creates a matrix with the required size to receive all pairwise
euclidean distances
  distance <- matrix(  nrow= length(data$Sample)+1,
ncol=length(data$Sample)+1)

  #"Simultaneously" name both rows and columns
  distance[,1] <-rNames
  distance[1,] <-rNames

  #Loop all pairs of points and calculates the Euclidean distance
between them in the Euclidean plane, the distance is given by the same
formula as the Pythagorean theorem as in fact, here, we are using
triangulation to determine the direct distance
  for (i in 1 : length(data$Sample) ) {

    for ( j in 1 : length(data$Sample ) ) {
      if (i == j) {
        distance[i+1,j+1] = 0

      }

      y = sqrt(  (abs(data$Lat[i] - data$Lat[j])^2) + (abs(data$Long[i]
- data$Long[j])^2)    )


      distance[i+1,j+1] = y
      distance[j+1,i+1] = y


    }

  }
  dist <- return (distance)
}
```

R code 2: π diversity calculation and group compilation for Chapter 3 interpolations

```
#This function requires several parameters, a datafile, a pairwise distance matrix, the distance threshold to group individuals, and the output directory and file name to save the results
nDiversity <- function (dataTemp, distanceTemp, distThereshold, outDir)
{
  #Prepare the starting variables
  data <- dataTemp
  distance <- distanceTemp
  dist <- distThereshold
  #A first loop through all lines of the pairwise euclidean distance with a inner loop covering all columns
  for (iter in 2:(dim(distance)[1])) {
    where = iter - 1
    i = 1
    group = list()
    Seq <- data.frame()

    #Here we pool the sample corresponding to the line in the pairwise distance matrix and all samples that have a euclidean distance to it below the threshold
    while (i <= dim(distance)[2]) {
      #Here the first individual is included in the group
      if (i < 2){

        group <- append(group, as.character(distance[iter,i]))


      }
      #All individuals whose distance to the individual of the correspondent line is below the threshold are here added to the group
      else if (as.numeric(distance[iter,i])<= dist &
!(as.character(distance[1,i]) %in% group)) {

        group <- append(group, as.character(distance[1,i]))

      }



      i=i+1



    }
    #If the group has only one sequence in it, we do not consider it for further analysis
    if (length(group)< 2){
      next
    }
    ################################################
```

```
    ##Create the Sample Sequence Dataframe
    #Here we prepare the dataset for each group by looping through all
available samples and adding the sequence data of only the samples in
the group
    for (i in 1: length(data$Sample)){
      if (as.character(data$Sample[i]) %in% group){


        rbind(Seq,data.frame(Sample = (as.character(data$Sample[i])),
Sequence = (as.character(data$Sequence[i])))) -> Seq


      }

    }

    ##################################################
    ##Calculates each haplotype frequency
    Seq["Freq"]<-as.numeric(0)
    #To calculate the haplotype frequency, for each sequence we match it
against all sequences (including itself) in order to calculate the
number of times that an haplotype occurs in one group
    for (m in 1:length(Seq$Sequence)){
      ntimes = 0

      for (n in 1:length(Seq$Sequence)){
        if (Seq$Sequence[m] == Seq$Sequence[n]) {

          ntimes = ntimes + 1

        }

      }
      #For each sequence the haplotype frequency is then added (it will
include duplicates in this phase)
      Seq$Freq[m] <- ntimes/(length(Seq$Sequence))
    }
    ##################################################

    ##################################################
    ##Creates a new dataframe with unique haplotypes
    Seqs = data.frame()
    #Unique haplotypes are extracted from the total list by creating a
new empty dataset which is only fed with a new sequence if this sequence
is different from previously added ones
    for (s in 1:length(Seq$Sequence)) {
      if (Seq$Sequence[s] %in% Seqs$Sequence) {

        next

      }
      else  {

        rbind(Seqs,data.frame(Sample = (Seq$Sample[s]), Sequence =
```

```
(Seq$Sequence[s]),Freq= (Seq$Freq[s]))) -> Seqs

      }

    }

    Seqs$Freq = as.numeric(Seqs$Freq)
    Seqs$Sequence = as.character(Seqs$Sequence)
    Seqs$Sample = as.character(Seqs$Sample)
    ################################################
    ## If less than two unique haplotypes were sampled in a group, its
nucleotide diversity was 0
    if (length (Seqs$Sequence)< 2){
      data$Ndiversity[where]<-0
    }
    ## If more than two unique haplotypes were found we calculated the
nucleotide diversity by following Nei & Kumar (2000) formula

    else if (length(Seqs$Sequence) >= 2) {


      ##Calculates the ndiversity

      n = as.numeric(length(Seq$Sequence))

      nDiversity = 0


      for (i in 1:(length(Seqs$Sequence)-1)){

        for (j in i+1:(length(Seqs$Sequence)-i)){

          s1=strsplit(Seqs$Sequence[i], "")[[1]]
          s2=strsplit(Seqs$Sequence[j], "")[[1]]

          diff = 0
          total = 0

          for (k in 1:length(s1)) {

            if (s1[k] == "-" | s2[k]== "-" | s1[k] == "N" | s2[k] ==
"N") {

              next
            }
            else if(s1[k]!=s2[k]){

              diff = diff + 1
              total = total +1

            }
            else if(s1[k]==s2[k]){

              total = total +1
```

```
            }

          }

          nDiversity = nDiversity + ((2*(n/(n-
1)))*((Seqs$Freq[i])*(Seqs$Freq[j])*(diff/total)))

        }
      }


      #Data is then added to a data.frame and saved in the output
directory
      data$Ndiversity[where]<-nDiversity
    }

  }
  write.csv (data, file=outDir)


}
```

R code 3: Pairwise uncorrected p-distances for Chapter 3 interpolations

```
uncorpDist <- function(file,printevery){
  #Choose to print every n loops
  printeveryN <- printevery
  #Reads the data into a data.frame
  data <- read.csv(file, sep=";", header = TRUE, colClasses=
"character")

  #Creates an empty matrix with the necessary dimensions
  n <- matrix(nrow=length(data$Sample), ncol = length(data$Sample))

  #A simple loop Tracker allowing us to know the stage of the loop
  where = 0

  #Calculates the number of iterations necessary to complete the matrix
  itotal = length(n)

  #A double loop is here created to compare differences between
sequences
  for (i in 1:length(data$Sequence)) {
    for (j in 1:length(data$Sequence))  {

      #Splits the string of the sequence into single nucleotide strings
allowing for individual nucleotide comparisons
      s1=strsplit(data$Sequence[i], "")[[1]]
      s2=strsplit(data$Sequence[j], "")[[1]]

      #The diff variable keeps track of the counted differences between
two sequences
      diff = 0
      #while the total will be equal to the length of the analysed
sequences minus any ambiguous area
      total = 0

      #This inner loop will compare each nucleotide of the two sequences
      for (l in 1:length(s1)) {
        #Our method discards regions between two sequences where missing
data is present
        if (s1[l] == "-" | s2[l]== "-" | s1[l] == "N" | s2[l] == "N"
|s1[l] == "?" | s2[l]== "?") {
          next
        }


        #When two different nucleotides are encountered, the diff
variable is raised by one
```

```
        else if(s1[l]!=s2[l]){

          diff = diff + 1
          total = total +1

        }

        #If the two nucleotides are the same, only the total amount of
sampled nucleotides is increased
        else if(s1[l]==s2[l]){

          total = total +1

        }

      }
      #The uncorrected p-distance  (Nei & Kumar 2000) are given by
p=nd/n where nd corresponds to the number of sites with different
nucleotides (variable diff) and n corresponds to the number of sampled
sites (in this case n may vary between pairs of sequences as missing
date is discarded
      p = diff/total



      #This statement evaluate if the current loop is located on the
matrix diagonal. If it is, the cell will be filled with a 0 and the next
iteration will start
      if (i == j) {
        n[i,j]= 0
        where = where + 1
        next

      }
      #This condition checks for non-diagonal cells that have already be
filled as this function fills the matrix by mirroring the cells (e.g. if
we are calculating the differences between the first and second samples,
both first row second column and second row first column cells would be
filled
      else if (i!= j && !is.na(n[i,j])) {
        where = where + 1
        next
      }

      n[i,j] = p
      n[j,i] = p
```

```
      where = where + 1
      #The cat serves only to inform the used on how many interactions
are left!
      #Note that repetitive printing of data to the console considerably
slows the process, for this reason in very large datasets the printevery
should have an high value


      if (where%%printeveryN == 0) {
        cat ("iteration ", where, " of ", itotal," ",
((where*100)/itotal)," % completed",sep="" , fill=TRUE       )
      }

    }
  }
  return(n)
}
```

Table S. M. 2: Summary table of the complete dataset used in Chapter 5 (Manuscript III). ID: identification number of each sampling locality; Locality: sampling locality or its description; Lat: latitude; Long: longitude; Sample: sample identification.

ID	Locality	Lat	Long	Sample
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3167
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3168
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3169
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3170
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3171
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3172
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3173
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3174
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3175
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3176
1	Tan-Tan, Guelta Ez Zerga	28.496967	-10.885570	GVA3177
2	Tan Tan, Draa river	28.531550	-10.950410	GVA3185
2	Tan Tan, Draa river	28.531550	-10.950410	GVA3186
2	Tan Tan, Draa river	28.531550	-10.950410	GVA3187
2	Tan Tan, Draa river	28.531550	-10.950410	GVA3188
2	Tan Tan, Draa river	28.531550	-10.950410	GVA3189
3	Tata, Oued Tissint	29.823342	-7.199100	GVA3178
4	Embalse Ouarzzazate	30.968611	-6.723889	GVA453
4	Embalse Ouarzzazate	30.968611	-6.723889	GVA454
4	Embalse Ouarzzazate	30.968611	-6.723889	GVA455
4	Embalse Ouarzzazate	30.968611	-6.723889	GVA456
4	Embalse Ouarzzazate	30.968611	-6.723889	GVA457
4	Embalse Ouarzzazate	30.968611	-6.723889	GVA458
5	12km N of Timezgadiouine	30.990383	-9.039817	GVA2707
6	Douer Targa	31.529717	-7.563383	GVA1320
6	Douer Targa	31.529717	-7.563383	GVA1321
6	Douer Targa	31.529717	-7.563383	GVA1322
6	Douer Targa	31.529717	-7.563383	GVA1323
6	Douer Targa	31.529717	-7.563383	GVA1324
6	Douer Targa	31.529717	-7.563383	GVA1325
6	Douer Targa	31.529717	-7.563383	GVA1326
6	Douer Targa	31.529717	-7.563383	GVA1327
6	Douer Targa	31.529717	-7.563383	GVA1328
6	Douer Targa	31.529717	-7.563383	GVA1329
6	Douer Targa	31.529717	-7.563383	GVA1330
6	Douer Targa	31.529717	-7.563383	GVA1332
6	Douer Targa	31.529717	-7.563383	GVA1333

ID	Locality	Lat	Long	Sample
6	Douer Targa	31.529717	-7.563383	GVA1334
6	Douer Targa	31.529717	-7.563383	GVA1335
6	Douer Targa	31.529717	-7.563383	GVA1336
6	Douer Targa	31.529717	-7.563383	GVA1337
6	Douer Targa	31.529717	-7.563383	GVA1338
6	Douer Targa	31.529717	-7.563383	GVA1339
6	Douer Targa	31.529717	-7.563383	GVA1340
6	Douer Targa	31.529717	-7.563383	GVA1341
6	Douer Targa	31.529717	-7.563383	GVA1342
6	Río Zat	31.529717	-7.563383	GVA459
6	Río Zat	31.529717	-7.563383	GVA460
6	Río Zat	31.529717	-7.563383	GVA461
6	Río Zat	31.529717	-7.563383	GVA462
7	Marrakech (Palmeral) River Ouad Tansift	31.689200	-7.989780	GVA451
7	Marrakech (Palmeral) River Ouad Tansift	31.689200	-7.989780	GVA452
8	Near Sidi-Chikér	31.749578	-8.738442	GVA2709
9	Sidi Mimoun	32.19949	-6.302234	GVA1351
9	Sidi Mimoun	32.19949	-6.302234	GVA1352
10	Sidi Mimoun	32.47412	-5.992854	GVA1353
10	Sidi Mimoun	32.47412	-5.992854	GVA1354
10	Sidi Mimoun	32.47412	-5.992854	GVA1355
10	Sidi Mimoun	32.47412	-5.992854	GVA1356
11	Sidi Mimoun	32.89314	-5.250433	GVA1357
12	Boulemane, Oued Sebb Ousfa	33.118758	-4.329360	GVA3179
12	Boulemane, Oued Sebb Ousfa	33.118758	-4.329360	GVA3180
13	Sidi Mimoun	33.54824	-5.097517	GVA1350
14	Sidi Mimoun	33.649233	-4.968117	GVA1343
14	Sidi Mimoun	33.649233	-4.968117	GVA1345
14	Sidi Mimoun	33.649233	-4.968117	GVA1346
15	Sidi Mimoun	33.65063	-4.968117	GVA1359
16	near Douira	34.573020	-2.733220	GVA2713
17	road to Moulay Bousselhaim	34.696650	-6.025867	GVA2712
18	Sidi Mimoun	34.69698	-5.572592	GVA1358
19	5km before Zoumi	34.745983	-5.422817	GVA2711
20	Fifi	35.022558	-5.205183	GVA1301
20	Fifi	35.022558	-5.205183	GVA1302
20	Fifi	35.022558	-5.205183	GVA1303
20	Fifi	35.022558	-5.205183	GVA1304
20	Fifi	35.022558	-5.205183	GVA1305
20	Fifi	35.022558	-5.205183	GVA1306

ID	Locality	Lat	Long	Sample
20	Fifi	35.022558	-5.205183	GVA1307
20	Fifi	35.022558	-5.205183	GVA1308
20	Fifi	35.022558	-5.205183	GVA1309
20	Fifi	35.022558	-5.205183	GVA1310
20	Fifi	35.022558	-5.205183	GVA1311
20	Fifi	35.022558	-5.205183	GVA1312
20	Fifi	35.022558	-5.205183	GVA1313
20	Fifi	35.022558	-5.205183	GVA1314
20	Fifi	35.022558	-5.205183	GVA1315
20	Fifi	35.022558	-5.205183	GVA1317
20	Fifi	35.022558	-5.205183	GVA1318
20	Fifi	35.022558	-5.205183	GVA1319
21	Agadir, Loukkos	35.210	-6.130	GVA3525
22	Sidi Mimoun	35.247357	-5.282003	GVA1347
23	Sidi Mimoun	35.299448	-5.218703	GVA1348
24	Tazia	35.341814	-5.551915	GVA2519
24	Tazia	35.341814	-5.551915	GVA2522
24	Tazia	35.341814	-5.551915	GVA2523
24	Tazia	35.341814	-5.551915	GVA2535
24	Tazia	35.341814	-5.551915	GVA2538
24	Tazia	35.341814	-5.551915	GVA2539
24	Tazia	35.341814	-5.551915	GVA2540
24	Tazia	35.341814	-5.551915	GVA2541
24	Tazia	35.341814	-5.551915	GVA2542
24	Tazia	35.341814	-5.551915	GVA2543
24	Tazia	35.341814	-5.551915	GVA2544
24	Tazia	35.341814	-5.551915	GVA2545
24	Tazia	35.341814	-5.551915	GVA2547
24	Tazia	35.341814	-5.551915	GVA2549
24	Tazia	35.341814	-5.551915	GVA2550
24	Tazia	35.341814	-5.551915	GVA2551
24	Tazia	35.341814	-5.551915	GVA2552
25	Ceuta_Embalse Renegado	35.890360	-5.348775	GVA2478
25	Ceuta_Embalse Renegado	35.890360	-5.348775	GVA2480
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2482
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2483
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2486
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2488
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2489
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2490

ID	Locality	Lat	Long	Sample
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2491
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2492
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2493
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2494
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2495
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2496
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2497
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2502
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2503
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2505
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2506
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2507
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2508
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2509
25	Ceuta_Embalse Infierno	35.890360	-5.348775	GVA2511
25	Ceuta_Embalse Renegado	35.890360	-5.348775	GVA2515
25	Ceuta_Embalse Renegado	35.890360	-5.348775	GVA2516
25	Ceuta_Embalse Renegado	35.896885	-5.349489	GVA2481
25	Ceuta_Embalse Renegado	35.896885	-5.349489	GVA2500
25	Ceuta_Embalse Renegado	35.896885	-5.349489	GVA2513
25	Ceuta_Embalse Renegado	35.896885	-5.349489	GVA2517
26	Málaga	36.756490	-5.290048	GVA3520
26	Málaga	36.756490	-5.290048	GVA3521
26	Málaga	36.756490	-5.290048	GVA3522
27	Algarve	37.020	-7.886	MI0109
27	Algarve	37.020	-7.886	MI0111
27	Algarve	37.020	-7.886	MI0110
27	Algarve	37.020	-7.886	MI0118
27	Algarve	37.020	-7.886	MI0117
27	Algarve	37.020	-7.886	MI0119
27	Algarve	37.020	-7.886	MI0121
27	Algarve	37.020	-7.886	MI0120
27	Algarve	37.020	-7.886	MI0101
27	Algarve	37.020	-7.886	MI0102
27	Algarve	37.020	-7.886	MI0103
27	Algarve	37.020	-7.886	MI0104
27	Algarve	37.020	-7.886	MI0105
27	Algarve	37.020	-7.886	MI0106
27	Algarve	37.020	-7.886	MI0107
27	Algarve	37.020	-7.886	MI0108

ID	Locality	Lat	Long	Sample
27	Algarve	37.020	-7.886	MI0114
27	Algarve	37.020	-7.886	MI0115
27	Algarve	37.020	-7.886	MI0116
27	Algarve	37.020	-7.886	MI0112
27	Algarve	37.020	-7.886	MI0113
28	Doñana	37.049271	-6.591358	gone
28	Doñana	37.049271	-6.591358	GVA2120
28	Doñana	37.049271	-6.591358	GVA2121
28	Doñana	37.049271	-6.591358	GVA2123
28	Doñana	37.049271	-6.591358	GVA2126
28	Doñana	37.049271	-6.591358	GVA2127
28	Doñana	37.049271	-6.591358	GVA2128
28	Doñana	37.049271	-6.591358	GVA2129
28	Doñana	37.049271	-6.591358	GVA2130
28	Doñana	37.049271	-6.591358	GVA2131
28	Doñana	37.049271	-6.591358	GVA2133
28	Doñana	37.049271	-6.591358	GVA2134
28	Doñana	37.049271	-6.591358	GVA2135
28	Doñana	37.049271	-6.591358	GVA2136
29	Granada, Brácana	37.217609	-3.952606	GVA3527
30	Sevilla, Fuentes de Andalucia	37.406293	-5.451904	GVA3526
31	Murcia: Las Moreras Mazarron	37.581687	-1.408753	GVA2714
31	Murcia: Las Moreras Mazarron	37.581687	-1.408753	GVA2715
32	Almograve	37.652608	-8.793578	GVA2100
32	Almograve	37.652608	-8.793578	GVA2101
32	Almograve	37.652608	-8.793578	GVA2102
32	Almograve	37.652608	-8.793578	GVA2103
32	Almograve	37.652608	-8.793578	GVA2104
32	Almograve	37.652608	-8.793578	GVA2105
32	Almograve	37.652608	-8.793578	GVA2106
32	Almograve	37.652608	-8.793578	GVA2107
32	Almograve	37.652608	-8.793578	GVA2108
32	Almograve	37.652608	-8.793578	GVA2109
32	Almograve	37.652608	-8.793578	GVA2110
32	Almograve	37.652608	-8.793578	GVA2111
32	Almograve	37.652608	-8.793578	GVA2112
32	Almograve	37.652608	-8.793578	GVA2113
32	Almograve	37.652608	-8.793578	GVA2114
32	Almograve	37.652608	-8.793578	GVA2115
33	Castro Verde	37.693791	-8.086303	GVA2170

ID	Locality	Lat	Long	Sample
33	Castro Verde	37.693791	-8.086303	GVA2171
33	Castro Verde	37.693791	-8.086303	GVA2172
33	Castro Verde	37.693791	-8.086303	GVA2173
33	Castro Verde	37.693791	-8.086303	GVA2174
33	Castro Verde	37.693791	-8.086303	GVA2175
33	Castro Verde	37.693791	-8.086303	GVA2178
33	Castro Verde	37.693791	-8.086303	GVA2179
33	Castro Verde	37.693791	-8.086303	GVA2180
33	Castro Verde	37.693791	-8.086303	GVA2182
33	Castro Verde	37.693791	-8.086303	GVA2183
33	Castro Verde	37.693791	-8.086303	GVA2184
33	Castro Verde	37.693791	-8.086303	GVA2185
33	Castro Verde	37.693791	-8.086303	GVA2186
33	Castro Verde	37.693791	-8.086303	GVA2187
33	Castro Verde	37.693791	-8.086303	GVA2188
34	Andújar, arroyo de la Cabrera	38.152175	-4.014532	GVA3142
34	Andújar, arroyo de la Cabrera	38.152175	-4.014532	GVA3143
34	Andújar, río Jándula	38.152175	-4.014532	GVA3144
34	Andújar, río Jándula	38.152175	-4.014532	GVA3145
34	Andújar, río Jándula	38.152175	-4.014532	GVA3146
34	Andújar, río Jándula	38.152175	-4.014532	GVA3147
34	Andújar, río Jándula	38.152175	-4.014532	GVA3148
34	Andújar, río Jándula	38.152175	-4.014532	GVA3149
34	Andújar, río Jándula	38.152175	-4.014532	GVA3150
34	Andújar, río Jándula	38.152175	-4.014532	GVA3151
34	Andújar, río Jándula	38.152175	-4.014532	GVA3152
34	Andújar, río Jándula	38.152175	-4.014532	GVA3153
34	Andújar, río Jándula	38.152175	-4.014532	GVA3154
35	Cardeña, centro de información	38.258223	-4.324061	GVA3107
35	Cardeña, centro de información	38.258223	-4.324061	GVA3108
35	Cardeña, centro de información	38.258223	-4.324061	GVA3109
35	Cardeña, centro de información	38.258223	-4.324061	GVA3110
35	Cardeña, centro de información	38.258223	-4.324061	GVA3111
35	Cardeña, centro de información	38.258223	-4.324061	GVA3112
35	Cardeña, centro de información	38.258223	-4.324061	GVA3113
35	Cardeña, centro de información	38.258223	-4.324061	GVA3114
35	Cardeña, centro de información	38.258223	-4.324061	GVA3115
35	Cardeña, centro de información	38.258223	-4.324061	GVA3116
35	Cardeña, centro de información	38.258223	-4.324061	GVA3131
35	Cardeña, centro de información	38.258223	-4.324061	GVA3132

ID	Locality	Lat	Long	Sample
35	Cardeña, centro de información	38.258223	-4.324061	GVA3133
35	Cardeña, centro de información	38.258223	-4.324061	GVA3134
35	Cardeña, centro de información	38.258223	-4.324061	GVA3135
36	Cardeña, embalse Tejoneras	38.266715	-4.278076	GVA3120
36	Cardeña, embalse Tejoneras	38.266715	-4.278076	GVA3121
36	Cardeña, embalse Tejoneras	38.266715	-4.278076	GVA3122
36	Cardeña, embalse Tejoneras	38.266715	-4.278076	GVA3123
36	Cardeña, embalse Tejoneras	38.266715	-4.278076	GVA3125
36	Cardeña, embalse Tejoneras	38.266715	-4.278076	GVA3126
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2001
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2002
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2003
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2004
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2005
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2006
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2008
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2009
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2010
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2011
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2012
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2013
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2014
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2015
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2016
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2017
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2018
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2019
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2020
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2021
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2022
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2023
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2024
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2025
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2026
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2027
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2028
37	Murcia: Cieza (Embalse del Judío)	38.293806	-1.432194	GVA2029
38	Ciudade Real	38.907735	-4.472099	GVA2138
38	Ciudade Real	38.907735	-4.472099	GVA2139
38	Ciudade Real	38.907735	-4.472099	GVA2141

ID	Locality	Lat	Long	Sample
38	Ciudade Real	38.907735	-4.472099	GVA2142
38	Ciudade Real	38.907735	-4.472099	GVA2143
38	Ciudade Real	38.907735	-4.472099	GVA2144
38	Ciudade Real	38.907735	-4.472099	GVA2145
38	Ciudade Real	38.907735	-4.472099	GVA2151
38	Ciudade Real	38.907735	-4.472099	GVA2160
38	Ciudade Real	38.907735	-4.472099	GVA2161
38	Ciudade Real	38.907735	-4.472099	GVA2163
39	Caldas da Rainha	39.444700	-9.137514	MI0301
39	Caldas da Rainha	39.444700	-9.137514	MI0303
39	Caldas da Rainha	39.444700	-9.137514	MI0304
39	Caldas da Rainha	39.444700	-9.137514	MI0305
39	Caldas da Rainha	39.444700	-9.137514	MI0306
39	Caldas da Rainha	39.444700	-9.137514	MI0307
39	Caldas da Rainha	39.444700	-9.137514	MI0308
39	Caldas da Rainha	39.444700	-9.137514	MI0309
39	Caldas da Rainha	39.444700	-9.137514	MI0310
39	Caldas da Rainha	39.444700	-9.137514	MI0311
39	Caldas da Rainha	39.444700	-9.137514	MI0312
39	Caldas da Rainha	39.444700	-9.137514	MI0313
39	Caldas da Rainha	39.444700	-9.137514	MI0314
39	Caldas da Rainha	39.444700	-9.137514	MI0315
39	Caldas da Rainha	39.444700	-9.137514	MI0318
39	Caldas da Rainha	39.444700	-9.137514	MI0319
39	Caldas da Rainha	39.444700	-9.137514	MI0320
39	Caldas da Rainha	39.444700	-9.137514	MI0321
39	Caldas da Rainha	39.444700	-9.137514	MI0322
39	Caldas da Rainha	39.444700	-9.137514	MI0323
39	Caldas da Rainha	39.444700	-9.137514	MI0324
39	Caldas da Rainha	39.444700	-9.137514	MI0325
39	Caldas da Rainha	39.444700	-9.137514	MI0326
39	Caldas da Rainha	39.444700	-9.137514	MI0327
39	Caldas da Rainha	39.444700	-9.137514	MI0328
39	Caldas da Rainha	39.444700	-9.137514	MI0329
39	Caldas da Rainha	39.444700	-9.137514	MI0330
39	Caldas da Rainha	39.444700	-9.137514	MI0331
39	Caldas da Rainha	39.444700	-9.137514	MI0332
39	Caldas da Rainha	39.444700	-9.137514	MI0333
39	Caldas da Rainha	39.444700	-9.137514	MI0335
40	Valencia - Peñíscola	39.470	-0.377	MI0701

ID	Locality	Lat	Long	Sample
40	Valencia - Peñíscola	39.470	-0.377	MI0710
40	Valencia - Peñíscola	39.470	-0.377	MI0702
40	Valencia - Peñíscola	39.470	-0.377	MI0704
40	Valencia - Peñíscola	39.470	-0.377	MI0705
40	Valencia - Peñíscola	39.470	-0.377	MI0706
40	Valencia - Peñíscola	39.470	-0.377	MI0707
40	Valencia - Peñíscola	39.470	-0.377	MI0708
40	Valencia - Serpis	39.470	-0.377	MI0711
40	Valencia - Serpis	39.470	-0.377	MI0720
40	Valencia - Serpis	39.470	-0.377	MI0713
40	Valencia - Serpis	39.470	-0.377	MI0714
40	Valencia - Serpis	39.470	-0.377	MI0716
40	Valencia - Serpis	39.470	-0.377	MI0717
40	Valencia - Serpis	39.470	-0.377	MI0718
40	Valencia - Serpis	39.470	-0.377	MI0719
41	Castelo Branco	39.702550	-7.308150	MI0230
41	Castelo Branco	39.702550	-7.308150	MI0231
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0201
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0207
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0219
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0220
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0221
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0223
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0208
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0209
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0210
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0211
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0212
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0213
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0202
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0214
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0215
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0216
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0217
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0218
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0203
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0204
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0205
41	Castelo Branco - Monte Galisteu	39.702550	-7.308150	MI0206
41	Castelo Branco - Monte Barata	39.702550	-7.308150	MI0224

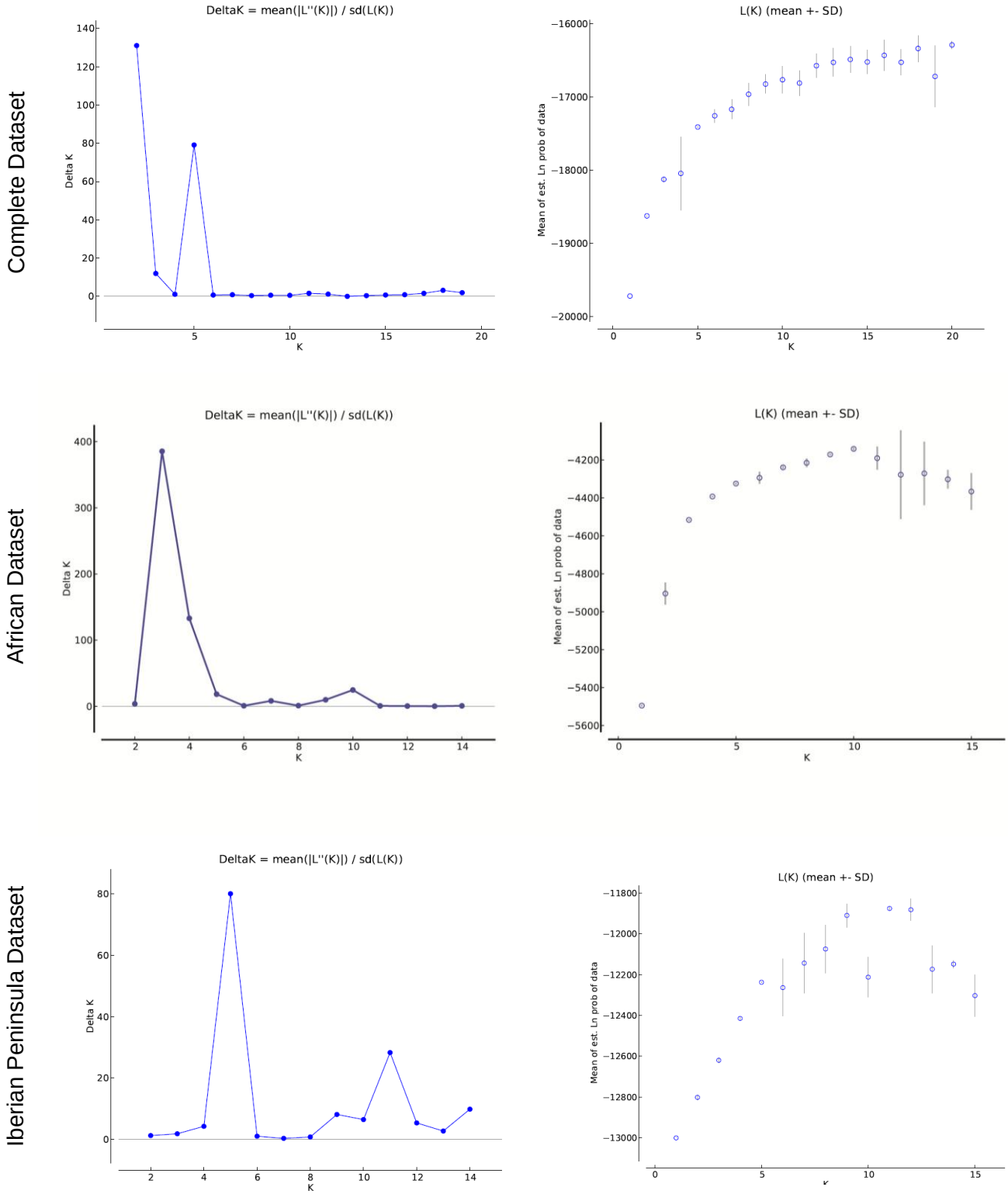
ID	Locality	Lat	Long	Sample
41	Castelo Branco - Monte Barata	39.702550	-7.308150	MI0225
41	Castelo Branco - Monte Barata	39.702550	-7.308150	MI0226
41	Castelo Branco - Monte Barata	39.702550	-7.308150	MI0229
41	Castelo Branco - Monte Barata	39.702550	-7.308150	MI0227
41	Castelo Branco - Monte Barata	39.702550	-7.308150	MI0228
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2189
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2190
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2191
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2192
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2193
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2194
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2195
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2196
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2197
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2198
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2199
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2200
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2201
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2202
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2203
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2205
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2206
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2207
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2208
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2209
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2211
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2212
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2213
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2214
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2215
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2216
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2217
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2218
42	Cáceres: Jaraiz de la Vera	40.010028	-5.742500	GVA2219
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2256
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2257
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2258
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2259
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2261
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2262

ID	Locality	Lat	Long	Sample
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2263
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2264
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2265
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2266
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2267
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2268
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2269
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2270
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2271
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2272
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2273
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2274
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2275
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2276
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2277
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2278
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2279
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2280
43	Ávila: Poyales del Hoyo	40.158179	-5.161660	GVA2281
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2030
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2031
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2032
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2033
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2034
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2035
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2036
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2037
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2038
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2039
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2040
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2041
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2042
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2043
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2044
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2045
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2046
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2047
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2048
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2049
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2050

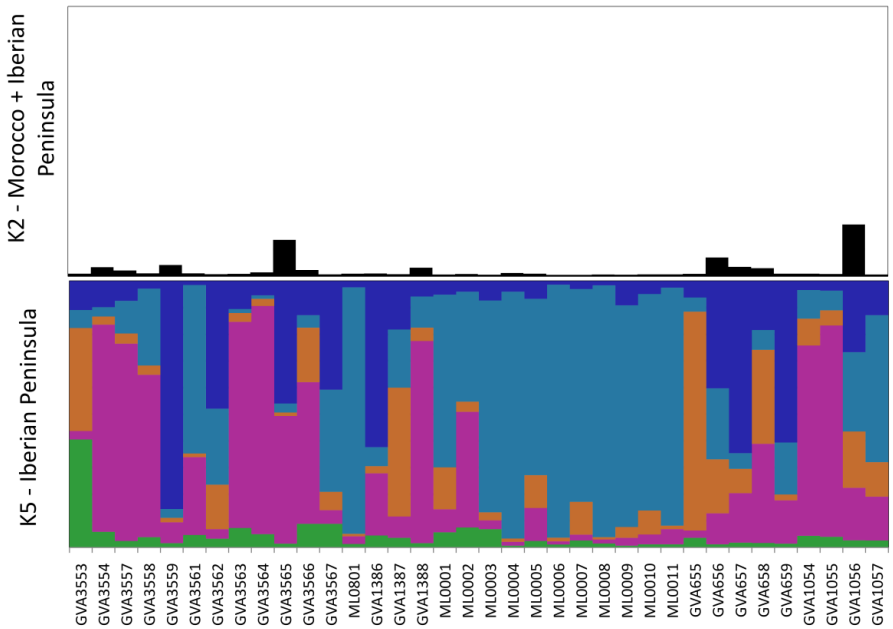
ID	Locality	Lat	Long	Sample
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2051
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2052
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2053
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2054
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2055
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2056
44	Madrid: Fresnedillas de la Oliva	40.488250	-4.124389	GVA2057
45	Guadalajara	40.548417	-3.256611	GVA2058
45	Guadalajara	40.548417	-3.256611	GVA2059
45	Guadalajara	40.548417	-3.256611	GVA2061
45	Guadalajara	40.548417	-3.256611	GVA2062
45	Guadalajara	40.548417	-3.256611	GVA2063
45	Guadalajara	40.548417	-3.256611	GVA2064
45	Guadalajara	40.548417	-3.256611	GVA2065
45	Guadalajara	40.548417	-3.256611	GVA2066
45	Guadalajara	40.548417	-3.256611	GVA2067
45	Guadalajara	40.548417	-3.256611	GVA2068
45	Guadalajara	40.548417	-3.256611	GVA2069
45	Guadalajara	40.548417	-3.256611	GVA2070
45	Guadalajara	40.548417	-3.256611	GVA2071
45	Guadalajara	40.548417	-3.256611	GVA2072
45	Guadalajara	40.548417	-3.256611	GVA2073
45	Guadalajara	40.548417	-3.256611	GVA2074
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2226
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2227
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2228
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2229
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2230
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2231
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2232
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2233
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2234
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2235
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2236
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2237
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2238
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2239
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2240
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2241
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2242

ID	Locality	Lat	Long	Sample
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2243
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2244
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2245
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2246
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2247
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2248
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2249
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2250
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2251
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2252
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA2254
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA650
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA651
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA652
46	Salamanca: Vilvestre	41.125028	-6.716111	GVA653
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2286
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2287
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2288
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2289
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2290
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2291
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2292
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2293
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2294
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2295
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2296
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2297
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2298
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2299
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2300
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2301
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2302
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2303
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2304
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2305
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2306
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2307
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2308
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2309
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2310

ID	Locality	Lat	Long	Sample
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2311
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2312
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2313
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2314
47	Barcelona: Delta del Llobregat	41.287028	2.016194	GVA2315
48	Girona: Caldes de Malabella	41.824222	2.781694	GVA2316
48	Girona: Caldes de Malabella	41.824222	2.781694	GVA2317
48	Girona: Caldes de Malabella	41.824222	2.781694	GVA2318
49	As neves	42.081165	-8.396621	GVA2083
49	As neves	42.081165	-8.396621	GVA2084
49	As neves	42.081165	-8.396621	GVA2085
50	Ribadavia	42.287553	-8.143496	GVA2093
50	Ribadavia	42.287553	-8.143496	GVA2094
50	Ribadavia	42.287553	-8.143496	GVA2095
50	Ribadavia	42.287553	-8.143496	GVA2096
50	Ribadavia	42.287553	-8.143496	GVA2097
50	Ribadavia	42.287553	-8.143496	GVA2098
50	Ribadavia	42.287553	-8.143496	GVA2099
50	Ribadavia (L.C.)	42.287553	-8.143496	GVA2081
50	Ribadavia (L.C.)	42.287553	-8.143496	GVA2082
51	Girona: Albera	42.377583	3.030556	GVA2320
51	Girona: Albera	42.377583	3.030556	GVA2321
51	Girona: Albera	42.377583	3.030556	GVA2322
51	Girona: Albera	42.377583	3.030556	GVA2324
51	Girona: Albera	42.377583	3.030556	GVA2325
51	Girona: Albera	42.377583	3.030556	GVA2326
51	Girona: Albera	42.377583	3.030556	GVA2327
51	Girona: Albera	42.377583	3.030556	GVA2328
51	Girona: Albera	42.377583	3.030556	GVA2330
51	Girona: Albera	42.377583	3.030556	GVA2331
51	Girona: Albera	42.377583	3.030556	GVA2332
51	Girona: Albera	42.377583	3.030556	GVA2333
51	Girona: Albera	42.377583	3.030556	GVA2334
51	Girona: Albera	42.377583	3.030556	GVA2335
51	Girona: Albera	42.377583	3.030556	GVA2336
51	Girona: Albera	42.377583	3.030556	GVA2337
51	Girona: Albera	42.377583	3.030556	GVA2339
51	Girona: Albera	42.377583	3.030556	GVA2341
51	Girona: Albera	42.377583	3.030556	GVA2342



S.M. Fig. 1: Structure Harvester graphic output of Delta K and Mean L(K). Top: Outputs for the complete dataset; Middle: Outputs for the African dataset; Bottom: Outputs for the Iberian Peninsula dataset;



S.M. Fig. 2: Structure output of the assignment tool results for samples collected from Iberian Recovery Centers and Basque Country (introduced population). For more details on the parameters used in STRUCTURE see Chapter (5) Methods. **Above:** Individuals with unknown origin output for K=2 when analysed in conjunction with the complete dataset. **Down:** Individuals with unknown origin output for K=5 when analysed in conjunction with solely the Iberian dataset.