Biogeographical History and Range Expansion Dynamics of the European Pond Turtle

Paulo Jorge de Andrade Dias Pereira
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, _____/_____/_____

Acknowledgments

To my supervisors, José Teixeira and Guillermo Velo-Antón, for the support during this work.

To Biodeserts team members, for all the help and advice given during all this time, a especial thank you to José Carlos Brito for all the help during my learning.

To the Masterss for all the help and funny moments that we shared during this journey.

To my family for being there all the time.

To Joana, for being Joana.

To my friends that helped alot during this process

This work was partially financed by the LIFE project LIFE+ Trachemys (LIFE09 NAT/ES/000529).

Abstract

The complex paleogeographic history of the Mediterranean Basin allowed for high levels of biodiversity in the region. In fact, several endemism occur in the area, granting the Mediterranean Basin the status of hotspot for biodiversity. For example, events such as the Messinian Salinity Crisis, that lead to the partial dissecation of the Mediterranean Sea, which in turn allowed for species to cross between Europe and Africa; and the Glacial-Interglacial cycles that promoted range shifts in several temperate species while tracking for suitable habitat, promoted in one hand vicariance events between several species, while in the other hand allowed for secondary contact zones to occur after species expanded from their glacial *refugias* during climate amelioration. Furthermore, the role of the European Peninsulas and the Maghreb as *refugia* during glacial periods promoted allopatric diversification as different populations of a same species would sometimes take refuge in different glacial *refugia* therefore allowing for the diversification between them.

The European pond turtle, *Emys orbicularis*, alongside its sister taxa, *Emys trinacris*, are the only representatives of the *Emys* genus in the old world. *Emys orbicularis* occurs widely throughout Europe, with populations reaching Asia, and North-western part of Africa. Currently 9 distinct mitochondrial lineages have been found using the cytochrome *b* gene. Nonetheless, across its entire range, demographical contractions have been noticed and prompted several conservation measures.

One of the described lineages is native to the Ibero-Maghbreian region, and has been subject of a lot of interest due to its very complex biogeographical history. Two different patterns arise in light of different data. Fossil records seem to indicate a longer presence of *Emys orbicularis* in the Iberian Peninsula when compared to the Maghreb. However, in light of genetic data, the hypothesis of a colonization of the Iberian Peninsula from North Africa seems more likely.

With this work we aim to reinforce the current knowledge on the biogeographical history of the species and to assess the role of the past climatic oscillations, and more specifically, the role that range expansions and contractions that occurred during the Quaternary to the Holocene, had on the current pattern of genetic structure. To do so both slow and fast evolving markers were here used (cytochrome *b* and microsatellites respectively) in an attempt to understand the directionality of the expansion and its consequences at the level of the genetic structure.

The addition of 80 new sequences to the already large data collected for this species allowed for the first time the detection of African haplotypes in the Iberian Peninsula, also the fact that North African populations show higher levels of genetic structure and diversity lead to a further support of North Africa as the origin for the Ibero-Maghrebian lineage. As for the role that range expansions had in shaping the current patterns of genetic diversity and structure, we found strong signs of allelic frequency clines alongside the axis of expansion, and a strong decrease in genetic diversity. Furthermore, the strong genetic structure present in the Northwest of the Iberian Peninsula seems to be concordant with known consequences of range expansions.

Overall, this study allowed for the first time a complete survey of the effects that the range expansions from southern *refugia* had in the Iberian Peninsula. Furthermore, the collected genetic data permitted us to improve a tool for the genetic allocation of individuals of unknown origin to a probable putative population of origin.

Keywords: *Emys orbicularis*, range expansions, phylogeography, historical biogeography, conservation

Resumo

Resumo

A complexa história paleogeográfica da Bacia do Mediterrâneo possibilitou a existência de níveis elevados de biodiversidade na região. De facto, a Bacia do Mediterrâneo é hoje considerada como um hotspot de biodiversidade devido em parte ao vasto número de endemismos que ocorrem na área. Por exemplo, eventos como a Crise Salina do Messiniano, que levou à dissecação parcial do Mar Mediterrânico, mas que em contrapartida permitiu a passagem de espécies entre a Europa e África; e os ciclos Glaciares-Interglaciares que promoveram alterações na distribuição de várias espécies de zonas temperadas que procuravam habitat adequado, promovam por um lado eventos de vicariância entre várias espécies, e, por outro lado, possibilitaram a ocorrência de zonas de contacto secundário após a expansão das espécies dos seus refúgios aquando o clima melhorou. Mais ainda, o papel desempenhado pelas Penínsulas Europeias e o Magreb como *refugia* glaciar, promoveu divergência alopátrica entre espécies, uma vez que populações distintas de uma mesma espécie poderiam refugiar-se em diferentes refúgios promovendo divergência entre as mesmas.

O cágado-de-carapaça-estriada, *Emys orbicularis,* em conjunto com a sua espécie irmã, *Emys trinacris,* são os únicos representantes do género *Emys* no mundo velho. *Emys orbicularis,* ocorre amplamente na Europa, com algumas populações a chegarem à Asia, e em parte do noroeste africano. Contudo, em toda a sua distribuição, contrações demográficas foram detetadas, levando à criação de várias medidas de conservação.

Uma das linhagens descritas é nativa da região Ibero-Magrebiana, e tem sido alvo de elevada atenção devido a sua complexa história biogeográfica. Consoante os dados observados, dois diferentes padrões surgem nesta região, pois, se considerarmos o registo fóssil, *E. orbicularis* apresenta estar presente na Península Ibérica à mais tempo que no Norte de África. No entanto, geneticamente esta linhagem apresenta ter a sua origem no Norte de África.

O nosso objetivo com este trabalho é o de reforçar o conhecimento atual sobre a biogeografia da espécie e discernir qual o papel das flutuações climáticas, e mais especificamente, o papel que os fenómenos de expansão e retração da distribuição das espécies que ocorreram entre o Quaternário e o Holoceno, tiveram nos padrões de estrutura genética que observamos hoje. Para tal, usamos tanto marcadores com elevadas como com baixas taxas de mutação (microssatélites e citocromo *b*,

respetivamente) numa tentativa de perceber a direccionalidade da expansão e as suas consequências ao nível da estrutura genética.

A adição de 80 novas sequências ao já vasto *dataset* colhido para esta espécie, permitiu pela primeira vez detetar haplótipos Africanos na Península Ibérica, mais ainda, o facto de as populações do Norte de África apresentarem elevados níveis de diversidade e estrutura genética, permitem-nos inferir com mais certeza sobre a possibilidade de a linhagem Ibero-Magrebiana se ter originado no Norte de África. Qual o papel de expansões de distribuição, no moldar os padrões de diversidade e estrutura genética. Encontra-mos sinais de variação clinal na frequência alélicas ao longo do eixo de expansão e um forte decréscimo na diversidade genética, padrões congruentes com expansões de distribuição.

De forma geral, este estudo possibilitou pela primeira vez uma completa analise dos efeitos que as expansões tiveram na Península Ibérica. Mais ainda, a quantidade de dados genéticos colhidos neste trabalho, permitiu o melhoramento de uma ferramenta genética para a alocação de indivíduos de origem desconhecida a sua população.

Palavras chave: *Emys orbicularis*, expansões, filogeografia, biogeografia, conservação.

Table of contents

ACKNOWLEDGMENTS I

ABSTRACT II

RESUMO IV

FIGURE INDEX IX

TABLE INDEX XIII

LIST OF ABBREVIATIONS XIV

CHAPTER 1: GENERAL INTRODUCTION 15

1.1 CONSERVATION OF BIODIVERSITY 16

1.1.1 Why conserve biodiversity? 16

1.1.2 Freshwater biodiversity 17

1.1.3 Major threats to Freshwater Biodiversity 18

1.2 TESTUDINES, THE EMYDIDAE FAMILY AND THE *EMYS* GENUS 18

1.2.1 *Emys* genus 19

1.2.2 *Emys orbicularis* distribution and threats 20

1.2.3 The complex biogeographical history of the European Pond Turtle 21

1.3 PALEOGEOGRAPHY OF THE MEDITERRANEAN BASIN 22

1.3.1 The role of the Messinian Salinity Crisis 22

1.3.2 Climatic oscillations of the Pleistocene 23

1.4 IMPACTS OF RANGE EXPANSION ON GENETIC PATTERNS 24

1.5 OBJECTIVES 26

CHAPTER 2: MANUSCRIPT 1 27

2.1 ABSTRACT: 28

2.2 INTRODUCTION 29

2.3 MATERIAL AND METHODS 31

2.3.1 Sampling 31

2.3.2 Cyt-B amplification through PCR, sequencing and haplotype analysis 31

2.3.3 Microsatellites amplification 32

2.3.4 Genetic structure and diversity indexes 33

2.4 RESULTS 35

2.4.1 Mitochondrial DNA analyses 35
2.4.2 Microsatellites analysis 35
2.5 DISCUSSION 36
2.5.1 Re-evaluating the role of North Africa as the source of the Iberian populations 36
2.5.2 Genetic relationships at both sides of the Gibraltar Strait 38
2.5.3 Genetic structure in Morocco 39
2.5.4 Conservation Implications 40

CHAPTER 3: MANUSCRIPT 2 43

3.1 INTRODUCTION 45
3.2 MATERIAL AND METHODS 47
3.2.1 Sampling, DNA extraction and dataset preparation 47
3.2.2 Microsatellites amplification 48
3.2.3 Microsatellites data analysis 48
3.2.4 Spatial patterns of genetic diversity 49
3.2.5 Allele frequency clines and trends of genetic diversity along the axis of colonization 50
3.3 RESULTS 51
3.3.1 Microsatellites analysis 51
3.3.2 Allelic Frequency Clines 51
3.4 DISCUSSION 55
3.4.1 Impacts of the range expansion in the genetic structure 55
3.4.2 Impacts of geographical barriers on the range expansion dynamics 56
3.4.3 The West Coast as a colonization route for the Northwestern populations 62

CHAPTER 4: MANUSCRIPT III 63

4.1 ABSTRACT 64
4.2 INTRODUCTION 65
4.3 MATERIAL AND METHODS 66
4.3.1 Sampling and lab procedures 66
4.3.2 Assignment tests 66
4.4 RESULTS 67
4.4.1 Proportion of assignment of each putative population 67
4.4.2 Accuracy of the test 69

4.4.3 Estimations of the population of origin of animals found in Recovery Centres 69

4.5 DISCUSSION 69

CHAPTER 5: FINAL REMARKS 74

CHAPTER 6: REFERENCES 77

CHAPTER 7: SUPPLEMENTARY MATERIAL 91

Figure Index

Figure 1.1: Diagram representing the five major threats to freshwater biodiversity and their possible interactions (in Dudgeon *et al.* 2006).....18

Figure 1.2: Global distribution of the genus *Emys*, in Spinks & Shaffer (2009). 19

Figure 1.3: Distribution range of *Emys orbicularis*. The Ibero-Maghrebian lineage is here highlighted in brown. Adapted from Spinks & Shaffer (2009). 20

Figure 2.1: Top: Geographical distribution of the three main haplogroups found in the study area, Red circles represent the Iberian haplogroup, which includes the haplotypes VIa, VIe, VIg, VIb, VIk, VIj and VId. Light green circles represent the haplogroup tightly linked with the Rif Mountain and comprises the haplogroups VIf, VIm, VIl, VIh, VIn and VIo. It is important to note that the last three haplotypes are either shared with the Iberian Peninsula or only have been found in there (case of VIo). Dark green circles correspond to the Atlas haplogroup comprising the haplotype Vic and VIi. Blue Circles indicate other lineages found in the study area. The current distribution of the subspecies is denoted in 10km by10 km squares marked in beige. **Bottom:** Haplotype network, in the left the complete haplotype network for the entire range of the species is shown the position of the lineage VI. On the right, the haplotype network for the lineage VI is shown. In this case red circles correspond to haplotypes found in the Iberian Peninsula, Light Green Circles, haplotypes found in the Rif Mountains and Dark Green haplotypes found in the Atlas Mountains. 34

Figure 2.2: Top: A spatial representation of the genetic structure found in Morocco for k=2. Each pie chart corresponds to a sampling location, identified by a number that corresponds to the population in the graphic below. Rivers are here represented as blue lines. **Bottom:** Structure output for k=2, each vertical bar corresponds to a individual, where the proportion of colours indicates its probability of assignment to the corresponding cluster. Black lines separate sampling locations. 37

Figure 3.1: Current distribution of the subspecies *Emys orbicularis occidentalis* (as denoted by the 10x10km squares in beige). The sampled locations are marked with red circles. 47

Figure 3.2: Decline of the genetic diversity alongside the south-north axis of expansion. Red circles correspond to the mean heterozygosity retrieved at each window of the sliding window analysis. Red line represents the linear regression used to summarize the trend, and the two dotted lines correspond to the 95% confidence interval. 52

Figure 3.3: This figure summarizes the overall trends found in the allelic frequencies. **Top:** A clear sign of allelic surfing, where an allele with very low frequencies at the core of the expansion, presents very high frequencies at the edge of the expansion. **Middle:** A common pattern in several of our alleles, this allele hints for a possible role of a barrier to constrict the surfing of alleles in the wave front, it is worth mentioning that due to the rapid shifts in the allele frequencies, the resulting interpolations are very weak and present several artefacts. **Bottom:** With the increase of frequencies of surfing alleles. Several alleles have their allelic frequencies decaying throughout the expansion axis. The graphics on the right side were produced through the analysis of the allelic frequency with a sliding window. The interpolations on the left were based on the allele frequencies observed in each population. 54

Figure 3.4: Top: Spatial representation of the genetic structure found for the entire distribution of *Emys orbicularis occidentalis* for k=2. Each pie chart corresponds to a sampling location, identified by a number that corresponds to the population in the structure graphic below. **Bottom:** Structure output for k=2, each horizontal line corresponds to a individual, where the proportion of the colours indicates its probability of assignment to the corresponding cluster. Black lines separate sampling locations. 57

Figure 3.5: Top: Spatial representation of the genetic structure found for the entire distribution of *Emys orbicularis occidentalis* for k=6. Each pie chart corresponds to a sampling location, identified by a number that corresponds to the population in the structure graphic below. **Bottom:** Structure output for k=6, each horizontal line corresponds to a individual, where the proportion of the colours indicates its probability of assignment to the corresponding cluster. Black lines separate sampling locations. 58

Figure 3.6: Summary display of the sPCA; **Top left:** The resulting connection networker after applying the Delaunay triangulation. This type of network was the only that could detect some type of connection between the Easter Iberian Coast and the South of the Iberian Peninsula; **Top right:** Spatial Interpolation of the scores of the first eigenvalue of the sPCA. The red lines denote regions of abrupt change in allelic frequencies; **Middle Left:** Assignment of the populations to one of the two groups recovered by the method, the method show a segregation between populations at North of the Central System to the others; **Middle Right:** A somewhat similar approach to the one in the left but in this case it uses the lagged scores to interpolate the assignment; **Bottom Left:** This graphic shows how much each eigenvector explains of the data. The eigenvector being observed in all previous graphics is underlined in black, while other selected eigenvectors for the sPCA are shown in light gray. **Bottom Right:** Decomposition of the eigenvalues in accordance in their spatial autocorrelation and Moran's I. We can observe here that, by the fact that

component one and two are isolated from other eigenvalues, they are good predictors to infer global structure 59

Figure 3.7: Summary display of the sPCA; **Top left:** The resulting connection networker after applying the Delaunay triangulation. This type of network was the only that could detect some type of connection between the Easter Iberian Coast and the South of the Iberian Peninsula; **Top right:** Spatial Interpolation of the scores of the second eigenvalue of the sPCA. The red lines denote regions of abrupt change in allelic frequencies; **Middle Left:** Assignment of the populations to one of the two groups recovered by the method, the division between a group in the Western Coast from other populations might be indicative of a possible colonization route; **Middle Right:** A somewhat similar approach to the one in the left but in this case it uses the lagged scores to interpolate the assignment; **Bottom Left:** This graphic shows how much each eigenvector explains of the data. The eigenvector being observed in all previous graphics is underlined in black, while other selected eigenvectors for the sPCA are shown in light gray. **Bottom Right:** Decomposition of the eigenvalues in accordance in their spatial autocorrelation and Moran's I. We can observe here that, by the fact that components one and two are isolated from other eigenvalues, they are good predictors to infer global structure.....60

Figure 3.8: First component of the PCA of genetic distances. A strong relationship is found between the western Coast and the Southwestern Peninsula. 61

Figure 4.1: **Top:** Spatial representation of the genetic structure found for the entire distribution of *Emys orbicularis occidentalis* for k=16. Each pie chart corresponds to a sampling location, identified by a number that corresponds to the population in the structure graphic below. **Bottom:** Structure output for k=16, each horizontal line corresponds to a individual, where the proportion of the colours indicates its probability of assignment to the corresponding cluster. Black lines separate sampling locations. 68

Figure 7.1: Graphical output of the Structure Harvester results for Delta K and L(k); **Top:** Results for the dataset used in Chapter 2; **Bottom:** Results for the dataset used in Chapters 3 and 4. 112

Figure 7.2: Allelic patterns of all alleles showing signs of allelic frequency clines. 119

Figure 7.3: Allelic patterns of all alleles showing signs of allelic frequency clines. 120

Figure 7.4: Allelic patterns of all alleles showing signs of allelic frequency clines. 121

Figure 7.5: Allelic patterns of all alleles showing signs of allelic frequency clines. 122

Figure 7.6: Allelic patterns of all alleles showing signs of allelic frequency clines. 123

Figure 7.7: Most important alleles contributing to the patterns observed in the first eigenvalue of the sPCA. 124

Figure 7.8: Most important alleles contributing to the patterns observed in the second eigenvalue of the sPCA 125

Table Index

Table 2.1: Summary table of diversity indices for both complete and partial dataset. *n*: number of samples per population; H_O: observed heterozygosity; H_E: expected heterozygosity; PA: private alleles; MNA: Mean Number of Alleles. 36

Table 3.1: A summary of the genetic diversity indices calculated for *Emys orbicularis occidentalis* populations. Lat: Latitude; Long: Longitude; n: number of samples for that location; Na: Number of alleles; Ne: Number of effective alleles; H_O: Observed heterozygosity; H_E: Expected heterozygosity; uH_E: unbiased expected heterozygosity; Ar: Allelic Richness. Populations with less than 10 samples, marked with *, were not used when calculating the Allelic Richness, and were not considerated for any populational inference. 52

Table 4.1: Proportion of membership of each putative population in each of the 16 clusters. Populations that failed to reach at least 0.5 of proportion of membership are marked with a *. 70

Table 4.2: Summary table for the genetic assignment results of individuals with unknown origin to a putative population. Note that to facilitate reading, Genetic Clusters have inherited the name of the putative population with a higher proportion of membership. 71

Table 7.1: Characteristics of the microsatellite loci used in *Emys orbicularis*. Primer (µL) = quantity of primer (µL) of a 10 µM primer solution; NA = Number of alleles sampled; primers marked with * are common for all three manuscripts 92

Table 7.2: Summary table of all sequenced samples in the study area. Locality: Sampling location; Lat: Latitude; Long: Longitude. ND: information not available. 93

Table 7.3: Summary table of all samples used for the microsatellite data used in this work. Sample: sample code; Country: country from where the sample was collected; Locality: sampling locality; Lat: Latitude; Long: Longitude; Reference: Source of the genotype. 100

Table 7.4: List of non private alleles that were tested for signs of allelic frequency clines. Highlighted in red are those alleles that had a p-value > 0.001 and a Amplitude of frequency of at least 0.2. Allele: List of non-private allele here tested; p-val: significance value of the linear regression; r^2: measures if a good fit to the linear model was found or not; minFreq: minimum observed frequency of a allele in all populations; maxFreq: maximum observed frequency of a allele in all populations; Amplitude: Absolute difference between minFreq and maxFreq. 116

List of Abbreviations

Mya – Million years ago

A_R – allelic richness

bp - base pairs

DNA - Deoxyribonucleic acid

nuDNA – nuclear DNA

mtDNA – mitochondrial DNA

PCR – polymerase chain reaction

cyt-*b* – Cytochrome-*b*

F_{ST} – Fixation index

H_E – expected heterozygosity

H_O – observed heterozygosity

uH_E – unbiased expected heterozygosity

H-W E – Hardy-Weinberg equilibrium

IUCN – International Union for Conservation of Nature

K – Number of genetic clusters assumed by the software STRUCTURE

LD – Linkage disequilibrium

µL – microlitre

µM - micromolar

PCA – Principal Component Analysis

AFC – Allelic Frequency Clines

SDD – Short Distance Dispersal

SPCA – Spacial Principal Component Analysis

Na – Number of alleles

Ne – Number of effective alleles

PA – private alleles

MNA – mean number of alleles

Chapter 1: General Introduction

1.1 Conservation of Biodiversity

Biodiversity – the variety of ecosystems, species, populations within species and genes within populations that constitutes life on earth (Allem 2000; Rands *et al.* 2010) of the planet is being exhausted at a rapid pace due to both indirect and direct anthropogenic actions (Frankham 2003). Despite the fact that public awareness on the matter has increased substantially in the past years, with several commitments from world leaders to halt biodiversity loss by 2010 (Butchart *et al.* 2010), we are still observing strong rates of biodiversity loss. In fact, these effects on biodiversity are so severe that some authors argue that we are losing biodiversity at closely the same rate as the past five mass extinctions, considering our time as the "sixth extinction" (Leakey & Lewin 1996). More recently, Barnosky *et al.* (2011) demonstrated that in fact, although the current rates of extinction are not as severe as those expected to have occurred in the past five mass extinctions, we are losing biodiversity at an alarming rate and, if we don't contradict the current trend, in a few centuries the extinction rate will match that of the five mass extinctions.

1.1.1 Why conserve biodiversity?

As Humans, we have a biological need for food, water, clean air, shelter and a certain set of climatic conditions (Millennium Ecosystem Assessment 2005). As of that, we are dependent of the world's biodiversity to exist as a species. In fact, not only we depend on it as a source for Bioresources – food, pharmaceutical components, natural fibres, timber, etc. – but we also depend on the services provided by the ecosystems – climate regulation, soil formation, oxygen production, carbon sequestration, etc. Furthermore, we benefit from its aesthetic and cultural value and a sustainable exploitation of biodiversity induce wealth and will improve human well-being (Nunes & van den Bergh 2001; Díaz *et al.* 2006).

1.1.2 Freshwater biodiversity

Of the world's Hydrosphere, only 2.5% is composed of freshwater. As it is well known, a huge amount of the world's freshwater is stored in the form of ice and permanent snow (68.7%). The second most abundant source of freshwater are in groundwater (30.1%) and only 0.29% of the total freshwater is concentrated in lakes, ponds, wetlands, river systems and biota (Gleick 1993).

Freshwater ecosystem services have been estimated to value approximately 6.5 trillion US dollars per year, corresponding to almost 20% of the estimated value of all Ecosystems on Earth (Costanza *et al.* 1997). Nonetheless, more than a billion of people lack safe-to-drink water, almost three billions lack access to proper sanitation (Gleick 1998; Millennium Ecosystem Assessment 2005). Each day, 14 to 30 thousand people have been estimated to die of water related diseases (Gleick 1998).

Although freshwater is, in part, considered a renewable resource, and even though we currently only use about 10% of the maximum renewable freshwater in the world, due to the variation in the availability of water through time and space, we can advocate that water scarcity is a problem that calls for appropriated water management solutions (Oki & Kanae 2006).

The freshwater ecosystems are possibly the most endangered in the world (Dudgeon *et al.* 2006). According to Ricciardi & Rasmussen (1999), the projected mean future extinctions for freshwater ecosystems are as five time greater than those projected for terrestrial ecosystems. A similar, although weaker, trend was later found by Collen *et al.* (2013).

1.1.3 Major threats to Freshwater Biodiversity

According to Dudgeon *et al*. (2006) threats to freshwater biodiversity can be summarized in 5 categories: Over-exploitation, pollution, flow modification, invasive species and habitat degradation (Figure 1.1).

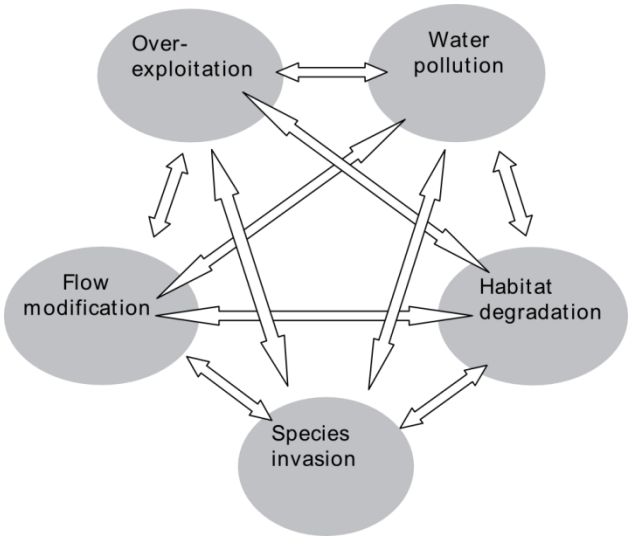

Figure 1.1: Diagram representing the five major threats to freshwater biodiversity and their possible interactions (in Dudgeon *et al.* 2006)

Since the Neolithic revolution, humans have preferred to settle in areas that are close to any easy source of freshwater (e.g. rivers or lakes) (Johnson 1996). Since then, we have overexploited the resources provided for our own advantage, with severe consequences for the ecosystems.

1.2 Testudines, the Emydidae family and the *Emys* genus

Testudines are characterized by the presence of a shell composed of a dorsal carapace and ventral plastron (Meylan 2002). Hedges & Kumar (2009) placed the emergence of the group in the Triassic (about 200 Mya) from a terrestrial ancestor (Joyce & Gauthier 2004). At the present, Testudines occupy marine, freshwater and land ecosystem, with 331 recognized species (Van Dijk *et al.* 2012), divided into two distinct sub orders; Pleurodira and Cryptodira (Guillon *et al.* 2012).

The Emydidae family is part of the Cryptodira sub-order, with two described subfamilies: the Deirochelyinae with all 6 described genus native to the American Continent; and the Emydinae with 4 described genus, where 3 are native to the American continent, while the genus *Emys* presents two species in the old world.

1.2.1 *Emys* genus

The *Emys* genus is characterized by medium-sized pond turtles ranging from highly aquatic (*E. marmorata, E. orbicularis, E. trinacris*) to semi terrestrial (*E. blandingii)*. The genus is present in both the new world (*E. marmorata* and *E. blandingii*) and the old world (*E. orbicularis* and *E. trinacris*) (see Figure 1.2; Fritz & Havaš 2007; Van Dijk *et al.* 2012). The ancestor of the *E. orbicularis*/*E. trinacris* species complex is thought to have colonized Eurasia from North America around 16 Mya (Miocene) via the Bering bridge (Spinks & Shaffer 2009).

Figure 1.2: Global distribution of the genus *Emys*, in Spinks & Shaffer (2009).

1.2.2 *Emys orbicularis* distribution and threats

While the Sicilian pond turtle (*Emys trinacris*) distribution is confined to the island of Sicily (Fritz *et al.* 2005), its sister taxa *Emys orbicularis* (Linneus, 1758) has a much wider and patchy distribution ranging from the Western Asia and both Eastern and Central Europe, to the European Peninsulas and Maghreb (see Figure 1.3; Stuckas *et al.* 2014).

The habitat requirements of *E. orbicularis* are very strict, the species is dependent on good freshwater quality in order to forage for food, as well as good terrestrial habitats in the surrounding areas, with open areas for basking and nesting (Ficetola *et al.* 2004). Those requirements sometimes conflict with human interests as drainage of wetlands is a common practice given the increasing necessity for agriculture and urban areas (Wood *et al.* 2003). Also, the increased water pollution, crescent amount of invasive species and overall habitat degradation are impacting the populations of *E. orbicularis* throughout its entire range, with clear signs of population regression (Cadi & Joly 2003, 2004; Andueza & Alcayde 2004; Matson *et al.* 2005; Velo-Antón *et al.* 2007, 2011; Trakimas & Sidaravi 2008; Fritz & Chiari 2013; Velo-Antón *et al.* unpublished).

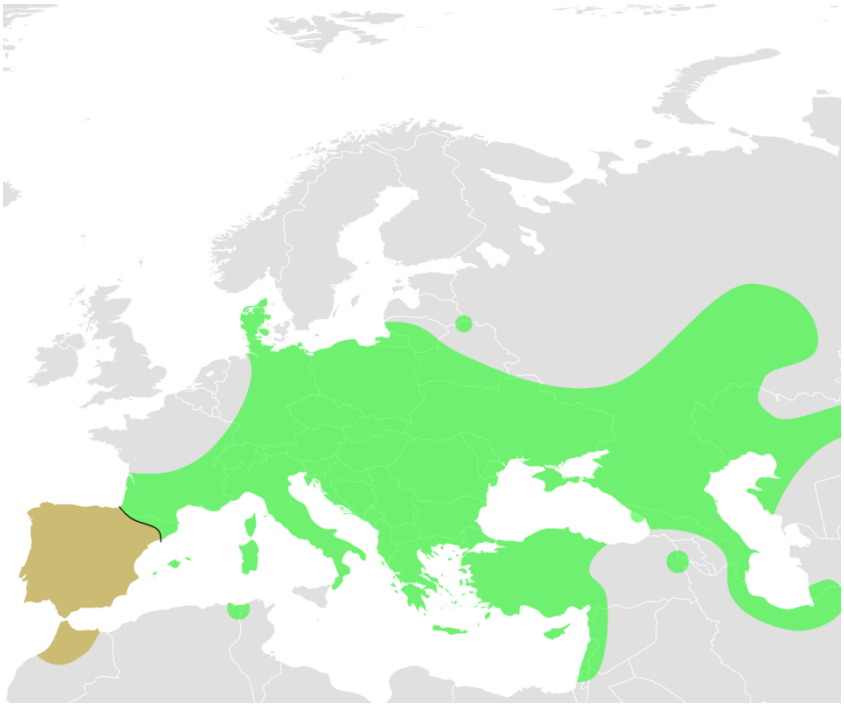

Figure 1.3: Distribution range of *Emys orbicularis*. The Ibero-Maghrebian lineage is here highlighted in brown. Adapted from Spinks & Shaffer (2009).

Moreover, freshwater turtles are among the most popular pets in the world (Moll & Moll 2004). During the 1980's the commercialization of *E. orbicularis* was common (Cordero & Ayres 2004). Due to the longevity and growth of terrapins, there is a tendency for pet owners to release animals in the wild, where admixture between native and non-native individuals may occur. Furthermore, some individuals are left on Recovery Centers, where it is common to use those animals to reinforce vulnerable populations (Velo-Antón *et al.* 2007). However, determining the population of origin from individuals delivered in Recovery Centres is hard but important, as the introduction of individuals from another differentiated population may lead to the genetic homogenization and consequent loss of genetic diversity.

1.2.3 The complex biogeographical history of the European Pond Turtle

Several studies show a deep phylogeographic structure throughout the species range where nine distinct mitochondrial lineages have been identified (Lenk *et al.* 1999; Fritz *et al.* 2007, 2009; Velo-Antón *et al.* 2011b; Stuckas *et al.* 2014). The distribution of these lineages results from the cyclical glacial-interglacial induced range expansions and contractions. During the Pleistocene, populations of *E. orbicularis* found refuge in the European Peninsulas, Anatolia (Lenk *et al.* 1999; Fritz *et al.* 2007, 2009; Sommer *et al.* 2007, 2009) and in the Maghreb (Stuckas *et al.* 2014), resulting in several divergent lineages.

The post-glacial re-colonization routes are well defined for the North-western, Central and Eastern Europe where the main sources for colonization were the Lineage I from the Anatolia region that colonized all Eastern Europe, and the Lineage II that originated in the Balkans that colonized Central and Western Europe, with a small number of populations present in Northeast of the Iberian Peninsula (Sommer *et al.* 2007, 2009). The remaining lineages played little to no role in the colonization of the Northern European latitudes, expanding little from their glacial *refugia*. However, the biogeographical history of the Ibero-Maghrebian lineage (lineage VI as described in Lenk *et al.* 1999; Fritz *et al.* 2007) is not as clear. This lineage inhabits both Morocco and the Iberian Peninsula and until very recently, the populations of the Maghreb were understudied and little was known about the role of North-Africa in the biogeographical history of this lineage. In fact, the work of Stuckas *et al.* (2014) was the first to target this region in order to assess their phylogenetic patterns. Their findings show a higher

genetic diversity and deeper phylogenetic structure in the Maghreb, and that Moroccan sequences are basal to the Iberian sequences. These results indicate a recent origin of the Iberian populations from Morocco and this hypothesis is concordant with the findings of Velo-Antón *et al.* (2008), that described a decline of genetic diversity in the Iberian Peninsula in a south-north axis.

1.3 Paleogeography of the Mediterranean Basin

The Mediterranean basin is considered one of the world's hotspots for biodiversity (Myers *et al.* 2000). Overall, this hotspot covers approximately 2 million square kilometers, from the Portuguese Islands of Madeira and Azores in the west to northern Iraq, crossing 34 countries, and harboring several reptile species (357 with 48% endemism), amphibians (115 with 62% endemism) and a great variety of plants (30000 with 43% endemism). Much of this richness is concentrated in the southern European Peninsulas and the Western Maghreb (Cuttelod *et al.* 2008).

The Western Mediterranean went through several geological, historical and environmental events making it a very interesting target for phylogenetic and biogeographical studies (de Jong 1998).

1.3.1 The role of the Messinian Salinity Crisis

In the Messinian stage (around 5.6 Mya) of the Miocene (23.04 to 5.33 Mya), the Gibraltar strait was closed due to tectonic movements that resulted in the uplift of the Southern Iberian and Morocco margins and subsequent closure of the gateways between the Mediterranean Sea and the Atlantic Ocean (Krijgsman *et al.* 1999; Duggen *et al.* 2003), which lead to the desiccation of the Mediterranean Sea exposing land bridges between North Africa and the Iberian Peninsula. This event had two major consequences in the biodiversity of the region: first the connection between both continents facilitated the movement of terrestrial species between North Africa and the Iberian Peninsula (Hsü *et al.* 1977); second, the changes in the level of the Mediterranean Sea brought significant impacts on the mediterranean climate where an

increase in temperatures and precipitation in the surrounding areas was observed (Murphy *et al.* 2009; Jiménez-Moreno *et al.* 2010).

At the Miocene-Pliocene boundary (around 5.33 Mya), the Atlantic Ocean reconnected with the Mediterranean Basin in an event known as the Zanclean or post-Messinian flood (Garcia-Castellanos *et al.* 2009). This event closed the land bridges between both continents allowing for several *taxa* to diverge in result of vicariance processes (e.g. *Acanthodactylus erythrurus*, Harris et al. 2004; *Pleurodeles*, Carranza & Arnold 2004; *Alytes*, Fromhage *et al.* 2004; *Chalcides bedriagae*, Carranza *et al.* 2008).

1.3.2 Climatic oscillations of the Pleistocene

The Pleistocene climatic oscillations also played a major role in the distribution of several *taxa*. This period was characterized by several glacial-interglacial cycles (Hewitt 2004). With the decrease in temperature and subsequently advance of the Ice sheets, several western Palaearctic species retracted to several *refugia* mainly in the southern European peninsulas (Hewitt 2011) and in the Maghreb (Husemann *et al.* 2014), where more mild temperatures were found, followed by northwards expansion during the Interglacial periods, tracking the availability of suitable habitat. The isolation of several populations of one species in separated Mediterranean *refugia* allowed for allopatric differentiation and, in some cases, speciation. During the Holocene, at the beginning of the post-glacial period, environmental conditions at the northern latitudes improved, allowing species to disperse once more to these areas. In fact, this pattern is observed in several European species (Michaux *et al.* 2005; see Hewitt 2004; Weiss & Ferrand 2007 for a review).

As for the role of the Maghreb as *refugia* during the climatic oscillations of the Pleistocene, Husemann *et al.* (2014) reviewed several phylogeographical studies where European lineages are nested within African clades (e.g. *Testudo graeca*, Graciá *et al*. 2013; *Malpolon monspessulanus* and *Hemorrhois hippocrepis,* Carranza *et al*. 2006; *Mauremys leprosa*, Fritz *et al*. 2006; *Emys orbicularis*, Stuckas *et al*. 2014; *Crocidura russula,* Cosson *et al.* 2005), indicating a northward colonization from North Africa to Europe, suggesting that North Africa played an important role in Pleistocene as *refugia* and source for the post-glacial colonization of Europe.

1.4 Impacts of range expansion on genetic patterns

Most, if not all, species have experienced a range expansion at some time in their history (Excoffier *et al.* 2009; Petit 2011). Understanding the impacts of range expansion on a species' genetic patterns is of great importance, as range expansions are linked with several important events, such as: the expansion of species from *refugia* during interglacial periods (Hewitt 2000, 2004), the spread of pathogens during epidemics (Biek *et al.* 2007; Velo-Antón *et al.* 2012b), and species range shift due to current climate changes (Parmesan & Yohe 2003).

In-silico studies have demonstrated that range expansions are fundamentally different from demographic expansions. In fact, range expansions will result in a reduction of genetic diversity and stochastic loss of alleles in the axis of expansion as populations suffer consecutive founder effect and consequent genetic drift (Austerlitz *et al.* 1997), while promoting genetic structure (Nei *et al.* 1975; Hallatschek *et al.* 2007; Hallatschek & Nelson 2008; Excoffier & Ray 2008; Excoffier *et al.* 2009). Also, during range expansion, new and/or extant alleles present in the edge of the wave of expansion, might "surf" the wave of expansion, reaching very large frequencies and might even fixate in the front of expansion (Edmonds *et al.* 2004; Klopfstein *et al.* 2006; Excoffier & Ray 2008), forming allelic frequency clines (Klopfstein *et al.* 2006).

The surfing of alleles has several potentiating factors; for once, the population size and dispersal dynamics of the expanding species might affect the probability of an allele to surf. In fact, Klopfstein *et al.* (2006) showned that alleles tend to surf more often in small populations than in large populations. Also, if alleles surf in large populations, they usually don't have the chance to fixate in the new colonized areas due to the overall low frequencies of the allele. In terms of dispersal strategy, long distance dispersal might mitigate the effects of sequential founder effects, preventing the loss of genetic diversity, while balancing the allelic frequencies in the peripheral populations (Berthouly-Salazar *et al.* 2013).

The surfing phenomena is not restricted to neutral alleles, as in fact any allele present at the wave front, being it neutral, advantageous or deleterious, may surf at the edge of the expansion (Travis *et al.* 2007; Excoffier *et al.* 2009; Lehe *et al.* 2012). The potential evolutionary consequences of such patterns are immense; Klopfstein *et al.* (2006) suggest that the surfing phenomenon could increase the rates of evolution at

range margins. Also, range expansions could be propitious to the spread of selected traits. For example, Phillips *et al.* (2006) observed that the annual rate of invasion for the Cane toad (*Rhinella marina)* in Australia has increased about fivefold since the first introduction. The authors also found that individuals at the expansion front have longer legs and expand faster than those in the core, possibly due to selection on the wave front. However, the surfing of deleterious mutations in the expansion wave, associated with the sequential reduction of genetic diversity might slow or halt the rate of expansion.

1.5 Objectives

The biogeographical history of the Ibero-Maghrebian lineage is very complex. The genetic relationships between populations at both sides of the Gibraltar Strait are still unclear due to the lack of samples in the southernmost regions of the Iberian Peninsula. Also, the pattern observed by Velo-Antón *et al.* (2008) in the genetic diversity of the Iberian Peninsula calls for more attention as it might be an opportunity to understand the impacts of range expansions in this system.

With this work, we aimed to increase the available information on the Ibero-Maghrebian lineage of *E. orbicularis* by focusing on several aspects:

Manuscript 1 [Chapter 2]: Here, we re-evaluate the genetic relationships between the Iberian Peninsula and Morocco by increasing the amount of genetic sequences of both regions. Also, through the use of microsatellite markers, we propose to shed new light into the genetic structure between three major Moroccan populations of *E. orbicularis occidentalis*.

Manuscript 2 [Chapter 3]: Here, we aim to determine the role of the recent range expansion from Morocco to the Iberian Peninsula in shaping the genetic structure and distribution of genetic diversity. We hypothesize that allele surfing might have had a major role into shaping the current observable genetic structure. Also, we evaluate the potential effects of major geographical barriers on the expansion dynamics.

Manuscript 3 [Chapter 4]: Taking into consideration the numerous conservation actions to protect *Emys orbicularis,* we followed the methodology described by Velo-Antón *et al.* (2007) to relocate individuals of unknown origin. In this work, we increase the baseline information with the aim of a more precise origin assignment.

With this work, we aim to contribute to the available knowledge about the relationship between the Iberian Peninsula and Morocco, with new insights on the impacts of the recent range expansion that this lineage underwent.

Chapter 2: Manuscript 1

Revisiting the biogeographical history of the Ibero-Maghrebian lineage of *Emys orbicularis* with insights on the Moroccan genetic structure.

Authors:

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

Affiliations:

[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]CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Rua dos Bragas, 289. 4050-123 Porto, Portugal

2.1 Abstract:

The Mediterranean basin harbours a high number of endemisms and species richness as a result of the palaeogeographic history of the region. Events such as the Messinian Salinity Crisis and the Quaternary climatic oscillations had a profound impact on the current species distribution and genetic patterns. The European pond turtle, *Emys orbicularis*, is distributed across Eastern and Central Europe, as well, as in Mediterranean countries and the Maghreb. However, the species faces several threats and has now a patchy distribution.

For this work we aimed to re-evaluate genetic relationships between the Iberian Peninsula and Morocco in order to discern the colonization direction. To do so, we used 80 cytochrome-*b* sequences and 126 individuals were genotyped for a set of 11 microsatellite loci. Our goals are to: 1) understand the relevance of North African populations as source for the colonization of the Iberian Peninsula; 2) discern any relevant genetic relationships between both sides of the Gibraltar Strait; 3) assess the genetic structure and diversity patterns; and 4) understand the impacts of anthropogenic actions on the genetic diversity of Moroccan populations.

We found three haplotypes shared between the two continents, which were highly differentiated from the remaining found in the Iberian Peninsula. Moreover, the star-like haplotype network found for the Iberian Peninsula leads to believe that a recent and rapid colonization occurred in the area. The nuclear data is concordant given the overall low allelic richness found in the region, when in comparison to Morocco. In Morocco, we found a possible gene flow barrier in the Rif Mountains, which could be explained by the topography in the area.

Overall, we were able to re-enforce the idea of a North African origin for the Iberian Peninsula *Emys orbicularis* and the Moroccan populations should be further evaluated to create protection policies in the area.

Keywords: Phylogeography, *Emys orbicularis,* population structure, conservation

2.2 Introduction

The palaeogeographic history of the western Mediterranean resulted in high species richness and endemism in the Iberian Peninsula and the Maghreb (region that spreads from Morocco, northern Algeria to Tunisia) giving the region the status of hotspot of biodiversity (Myers *et al.* 2000). In the late Miocene (around 5.6 Mya), the closure of the Gibraltar Strait, possibly due to tectonic movements (Krijgsman *et al.* 1999; Duggen *et al.* 2003), lead to the desiccation of the Mediterranean Sea, forming land bridges that connected the two continents allowing for several terrestrial organisms to easily disperse in both directions (Hsü *et al.* 1977). Around 5.5-5.3 Mya, the land bridge connecting both continents collapsed and the Mediterranean Sea refilled from the Atlantic, isolating populations that were in contact allowing for vicariant processes to act on genetic diversification, predicting considerable genetic differentiation between the two continents. Even though this pattern is observed in several organisms (e.g. *Acanthodactylus erythrurus*, Harris *et al.* 2004; *Pleurodeles*, Carranza & Arnold 2004; *Alytes*, Fromhage *et al.* 2004; *Chalcides bedriagae*, Carranza *et al.* 2008), several *taxa* present relatively less marked patterns of genetic differentiation, suggesting that several species dispersed after the re-opening of the Gibraltar Strait through water instead of land (e.g. *Testudo graeca*, Graciá *et al.* 2013; *Malpolon monspessulanus* and *Hemorrhois hippocrepis* Carranza *et al.* 2006; *Mauremys leprosa,* Fritz *et al.* 2006 and *Emys orbicularis*, Stuckas *et al.* 2014).

Here, we focus on the European pond turtle, *Emys orbicularis* (Linnaeus, 1758), a widely spread species but with a patchy distribution, ranging from the Eastern and Central Europe, to the Mediterranean countries and the Maghreb (Fritz 2001). European pond turtles tolerate a strict range of habitat conditions, as they require good freshwater quality in order to forage for food, as well as terrestrial habitats with some open areas for nesting and basking (Ficetola *et al.* 2004). In fact, throughout its entire distribution and due to the degradation of habitat, the populations of *E. orbicularis* are in clear regression (Cadi & Joly 2003, 2004; Andueza & Alcayde 2004; Matson *et al.* 2005; Velo-Antón *et al.* 2007, 2011a; Trakimas & Sidaravi 2008; see Fritz & Chiari 2013). Throughout its distribution, nine geographically coherent mitochondrial lineages (based on Cytochrome *b*) have been found (Fritz *et al.* 2007, 2009; Sommer *et al.* 2007; Velo-Antón *et al.* 2011b; Stuckas *et al.* 2014). The Ibero-Maghrebian lineage (hereby lineage VI as first described in Lenk *et al.* 1999; Fritz *et al.* 2007) inhabits

Morocco and the Iberian Peninsula, where it overlaps with both lineages II and V along northeastern Iberian Peninsula. Studies characterizing the distribution of genetic diversity and its structure throughout the distribution of this lineage have described higher levels of genetic divergence in Moroccan populations when compared with the Iberia Peninsula (Stuckas *et al.* 2014). Two main haplogroups are described throughout the lineage range (Stuckas *et al.* 2014), the first representing Moroccan populations, sub-structured into two different groups, one encompassing the Middle Atlas Mountains and Moroccan Atlantic coast, and the other corresponding to the Rif Mountains. The second haplogroup corresponds to the Iberian populations, where little differentiation is found between haplotypes pointing to a recent expansion (Stuckas *et al.* 2014). Also, Velo-Antón *et al.* (2008) shown that genetic diversity in the Iberian Peninsula decreases northwards, further reinforcing the hypothesis of a single and recent colonization of the Iberian Peninsula from Morocco.

Due to the biogeographic history of this lineage and the historical importance of the Moroccan populations, further sampling efforts are necessary both in Morocco and Southern Iberian Peninsula, as several isolated populations at both sides of the Gibraltar Strait are still unstudied.

In this study, we aim to: 1) confirm the relevance of North African populations as the source for the colonization of the Iberian Peninsula, where, we hypothesize that indeed, North Africa acted as *refugia* during past climatic oscillations; 2) discern any relevant genetic relationships between both sides of the Gibraltar Strait; 3) assess the genetic structure and diversity patterns, both at mitochondrial and nuclear level; and 4) understand the impacts of anthropogenic actions on the genetic diversity of Moroccan populations of *Emys orbicularis*. To achieve our goals, we take advantage of the available genetic database (cytochrome *b* sequences from previous studies (Fritz *et al.* 2009; Velo-Antón *et al.* 2011b; Stuckas *et al.* 2014)), and complement it with new sequences from isolated populations collected in the study area. Also, microsatellite markers were used to assess the contemporary genetic structure and diversity in Moroccan populations, allowing us to uncover more detailed relationships between these populations, and to obtain as well a better picture of how anthropogenic impacts might have affected the genetic diversity in the region.

2.3 Material and Methods

2.3.1 Sampling

A total of 126 blood (either conserved in ethanol or in dried blood spots) or tissue (tail tips or buccal swabs) samples from the Iberian Peninsula (56 samples from 16 sites) and Morocco (70 samples from 6 sites), where collected.

Genomic DNA was extracted from both blood and tissue samples using the EasySpin commercial kit, following the manufacturer's protocols, with an extended lyses time to maximize the yield.

2.3.2 Cyt-B amplification through PCR, sequencing and haplotype analysis

A total of 83 samples were selected for cytochrome *b* sequencing (56 from the Iberian Peninsula and 27 from Morocco covering all sampled sites). A total of 80 new sequences were generated (55 from the Iberian Peninsula and 25 from Morocco), increasing the number of sampled populations and expanding the covered range (see Figure 2.1).

Primers mt-A-neu and H-15909 (Lenk *et al.* 1999) were used to amplify the selected fragment, resulting in an amplicon of approximately 1031 bp. PCR reactions were carried out in a final volume of 10μL, with 5μL of MyTaq™ Mix (Bioline), 0.4μL of each primer (primer concentration of 10μM), 3.2μL of ultra-pure water and 1μL of DNA. A BioRad T100 Thermal Cycler was used to carry out the PCR under the following program: initial denaturation at 95ºC for 15 minutes; 40 cycles at 95ºC for 45 seconds, 52ºC for 45 seconds, and 72ºC for 1 minute and 10 seconds. A final elongation step at 60ºC was performed throughout 30 minutes. Afterwards, the resulting product was depleted of non-used primers and nucleotides through an ExoSap (USB® ExoSAP-IT® PCR Product Cleanup, Affymetrix) cleaning step following manufacture's protocol. Sequencing reactions were then carried out on a BioRad T100 Thermal Cycler with BigDye® Terminator v3.1 Cycle Sequencing Kits (AB Applied Biosystems) following manufactures protocol. Finally, sequences were produced on an ABI 3130xl genetic analyzer (Applied Biosystems, Foster City, Ca, USA). The resulting chromatograms

were verified, aligned and corrected by eye using Geneious Pro v4.8.5 (http://www.geneious.com/).

Sequences were then added to the genetic dataset available from previous studies (Lenk *et al.* 1999; Fritz *et al.* 2007; Velo-Antón *et al.* 2008; Pedall *et al.* 2011; Stuckas *et al.* 2014), resulting in a total dataset of 274 sequences (1031 bp long) from both lineage VI (n= 180) and the other two lineages that occur in the Northeastern Iberian Peninsula (II and V). To avoid confusion when naming new haplotypes, we followed the nomenclature adopted in the mentioned papers.

A parsimony haplotype network was constructed using TCS v1.21 (Clement *et al.* 2000), using as threshold the default 95% probability.

2.3.3 Microsatellites amplification

A total of 66 samples spread across three populations (Fifi, Sidi Mimoun and Moulay Abdesalam; see supplementary material table 7.3) were amplified for eleven microsatellite loci. These primers were developed for *Glyptemys muhlenbergii* and tested for cross-amplification in *Emys orbicularis* (King & Julian 2004), and 7 of them (D88, D114, D16, D93, D87, D51 and B08) were previously used in Iberian populations of *Emys orbicularis* (Roques *et al.* 2006; Velo-Antón *et al.* 2007, 2008). For each locus a third primer was used, following the M13 tailed primer method (Oetting *et al.* 1995). This primer was labeled with FAM, NED, VIC or PET depending on the selected dye for each locus, and was used at the same concentration as the reverse primer. To adjust for this, forward primers were used at a ten-fold dilution.

PCR reactions were performed in two multiplex reactions (Supplementary Material Table 7.1) at a final volume of 10μL, with 5μL of Quiagen Multiplex PCR Kit, 1μL of DNA and 1μL of the primer mix, with tails and primers (at 10μM). For each multiplex a touchdown PCR reaction was conducted under the following conditions: initial denaturation at 95ºC for 15 minutes; 9 cycles with a denaturation step at 95ºC for 38 seconds, annealing at 60ºC for 1 minute (decreasing 0.5ºC each cycle) and extension at 72ºC for 30 seconds; 23 cycles of 95ºC for 30 seconds, 56ºC for 1 minute, 72ºC for 30 seconds; 8 cycles of 95ºC for 30 seconds, 53ºC for 1 minute, 72ºC for 30 seconds, followed by a final elongation step at 60ºC for 30 minutes. PCR reactions were performed on a BioRad T100 Thermocycler and genotypes were obtained from an ABI 3130xl genetic analyzer (Applied Biosystems, Foster City, CA, USA). GeneScan^{TM}-500

Liz was used as size standard. Allele scoring was performed in GeneMapper v4.0 (Applied Biosystems). For comparison purposes, a small set of ten samples used in Velo-Antón *et al.* (2008) were re-genotyped in this study in order to correct the allele scoring for the seven common microsatellites (D88, D114, D16, D93, D87, D51 and B08).

2.3.4 Genetic structure and diversity indexes

Two datasets of microsatellites were created, a first dataset containing the eleven microsatellites (hereby complete dataset) and a second (hereby partial dataset) only containing the seven common microsatellites (Velo-Antón *et al.* 2007, 2008). This second dataset was constructed in order to allow us to compare our results with those previously found for the Iberian Peninsula.

Deviations from Hardy-Weinberg equilibrium and signs of linkage disequilibrium across all populations and loci were inferred in GENEPOP v4.2 on the web (Raymond & Rousset 1995; Rousset 2008). Genalex 6.5 (Peakall & Smouse 2006, 2012) was used to calculate genetic diversity indexes such as Observed Heterozygosity (H_O), Expected Heterozygosity (H_E) and Number of effective alleles (Na). Mean number of alleles per locus across loci was also calculated using Genetix V4.05 (Belkhir *et al.* 2004) for the partial dataset.

Genetic structure in North Africa was evaluated using the Bayesian approach implemented in STRUCTURE (Pritchard *et al.* 2000). STRUCTURE was run from k=1 to k=10. This upper bound was set to 10 as it is high enough to allow for possible substructure in each population, ten replicates were run for each k for 1,000,000 generations per run, including 250,000 generations of burn-in. Our runs accounted for admixture and correlated allele frequencies. In order to determine the most likely number of clusters we used the Evanno's method (Evanno *et al.* 2005) implemented in the Structure Harvester (Earl & vonHoldt 2012).

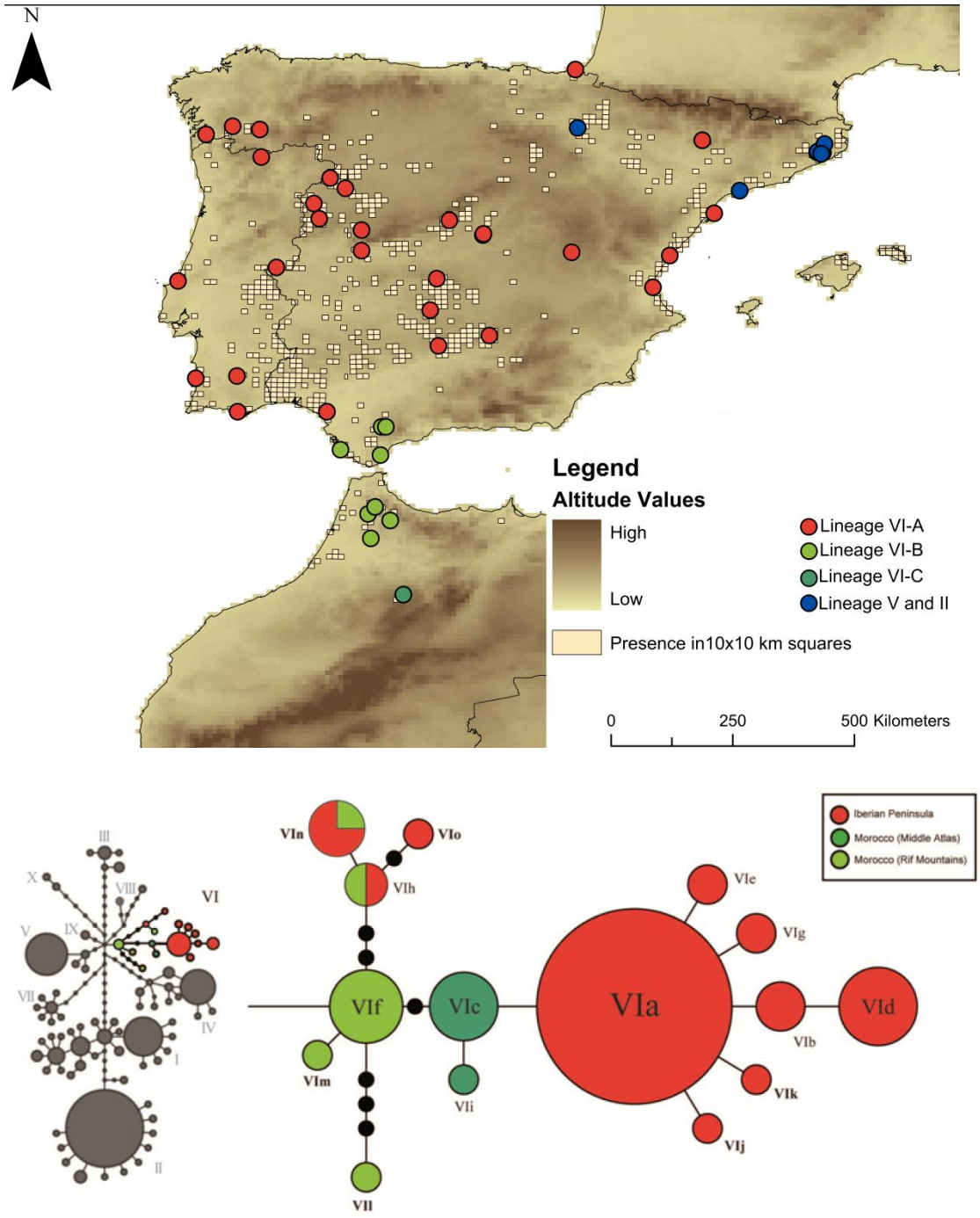

Figure 2.1: Top: Geographical distribution of the three main haplogroups found in the study area, Red circles represent the Iberian haplogroup, which includes the haplotypes VIa, VIe, VIg, VIb, VIk, VIj and VId. Light green circles represent the haplogroup tightly linked with the Rif Mountain and comprises the haplogroups VIf, VIm, VII, VIh, VIn and VIo. It is important to note that the last three haplotypes are either shared with the Iberian Peninsula or only have been found in there (case of VIo). Dark green circles correspond to the Atlas haplogroup comprising the haplotype Vic and VIi. Blue Circles indicate other lineages found in the study area. The current distribution of the subspecies is denoted in 10km by10 km squares marked in beige. **Bottom:** Haplotype network, in the left the complete haplotype network for the entire range of the species is shown the position of the lineage VI. On the right, the haplotype network for the lineage VI is shown. In this case red circles correspond to haplotypes found in the Iberian Peninsula, Light Green Circles, haplotypes found in the Rif Mountains and Dark Green haplotypes found in the Atlas Mountains.

2.4 Results

2.4.1 Mitochondrial DNA analyses

In addition to the nine haplotypes previously found in the Ibero-Maghrebian lineage (Stuckas *et al.* 2014), six new haplotypes were found (supplementary material table 7.2). Three of the newly encountered haplotypes were found in Morocco (VIl, VIm, VIn) and three in the Iberian Peninsula (VIj, VIk, VIo) (Figure 2.1). Here we find for the first time shared haplotypes between both sides of the Gibraltar Strait, particularly between the Rif Mountains (Jeramena) and the two southernmost populations of the Iberian Peninsula, Málaga and Cádiz. Also, the haplotypes found in these populations (VIh, VIn and VIo) are highly divergent from those found across the Iberian Peninsula.

In the Iberian Peninsula the most commonly represented haplotype is VIa (n=101), while in Morocco the most frequent haplotype is the VIf (n=25), even though most of the sampling effort is concentrated in the Iberian Peninsula, the same number of haplotypes was found for both regions. The haplotype network divides both regions into two groups; a first one comprising all Moroccan haplotypes and the haplotype VIo, even though the latter was only found in the Iberian Peninsula, presenting some geographical structure, with the Middle Atlas separated from the Rift Mountains. The second group collapses all Iberian haplotypes into a start-shaped network, with no evident geographical structure.

2.4.2 Microsatellites analysis

No evidence of linkage disequilibrium or deviation to Hardy-Weinberg equilibrium was found in the eleven microsatellites used, so all of them were used in subsequent analysis. Overall, the number of alleles per locus ranged from 2 (Locus A32 and D121) to 16 (Locus D16), averaging 9 alleles per locus, amounting to a total of 99 alleles across all loci. For the Moroccan populations, the highest expected heterozygosity values were observed in Fifi (0.695 for the complete dataset, 0.757 for the partial dataset), while in Moulay Abdesalam the lowest values of expected heterozygosity were found (0.578 for the complete dataset, 0.663 for the partial dataset). These values were still lower than those presented in Doñana (0.80 partial dataset only; southernmost population of the Iberian Peninsula sampled in Velo-Antón *et al.* 2008). As for the Mean number of alleles per locus, Fifi presents again the highest values

Table 2.1: Summary table of diversity indices for both complete and partial dataset. *n*: number of samples per population; H_O: observed heterozygosity; H_E: expected heterozygosity; PA: private alleles; MNA: Mean Number of Alleles.

		Complete Dataset		Partial Dataset		
Population	***n***	**H_O**	**H_E**	**H_O**	**H_E**	**MNA**
Fifi	26	0.736	0.695	0.777	0.757	8.4286
Sidi Mimoun	16	0.614	0.607	0.765	0.751	7.2857
Moulay Abdesalam	24	0.655	0.578	0.763	0.663	4.8571
Doñana	36	---	---	0.820	0.800	6.180

(8.43) and Moulay Abdesalam presents the lowest values (4.86). Nonetheless, Fifi and Sidi Mimoun (7.29) present higher values than those found in Doñana (6.18) (Table 2.1).

As for the genetic structure, two clusters were presented as the most probable umber of genetic clusters by the Evanno method (Supplementary material Figure 7.1 Top). In this solution, Moulay Abdesalam (Rif Mountains) is assigned to its own genetic group, while Fifi (Rif Mountains) and Sidi Mimoune (Middle Atlas) are grouped together (Figure 2.2). We have found very little genetic admixture, especially in the Fifi-Sidi Mimoune genetic cluster, in the Moulay Abdesalam cluster, two individuals present some genetic admixture (Figure 2.2).

2.5 Discussion

2.5.1 Re-evaluating the role of North Africa as the source of the Iberian populations

North Africa played an important role as *refugia* in the glacial-interglacial cycles. When temperatures started to decrease, various European species started to contract their ranges to the south, where mild temperatures could be found, settling in several areas that acted as *refugia*, both in the European Peninsulas and in North Africa (Hewitt 2000; Husemann *et al.* 2014). Nowadays, the origin of several taxa that are currently present in both margins of the Mediterranean Sea can be traced to a North African origin; such is the case in arthoprods (*Buthus,* Sousa *et al.* 2012), reptiles (*Mauremys leprosa,* Fritz *et al.* 2006; various snakes, Carranza *et al.* 2004, 2006; *Testudo graeca*, Álvarez *et al.* 2000; *Chamaeleo chamaeleon*, Paulo *et al.* 2002).

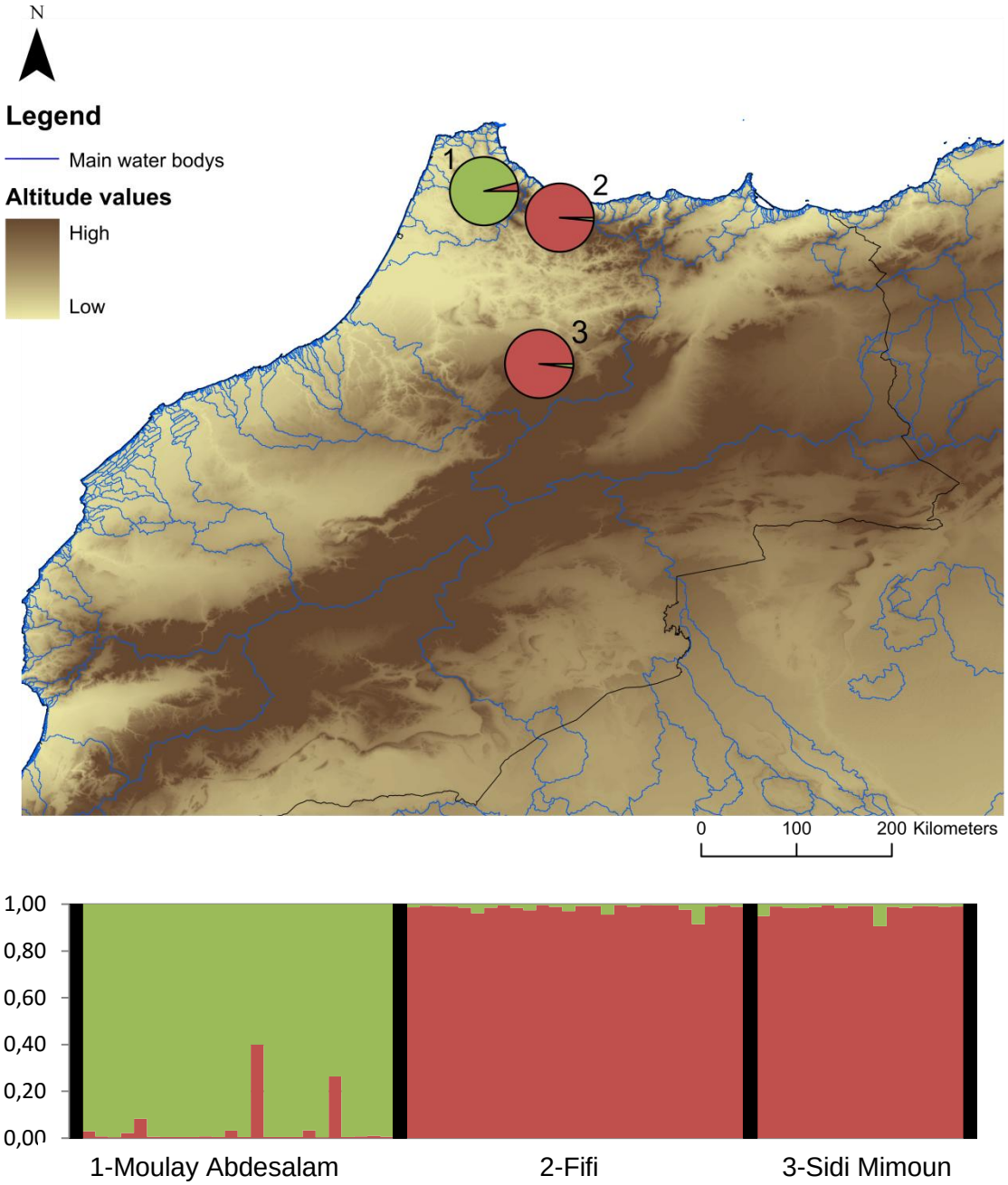

Figure 2.2: Top: A spatial representation of the genetic structure found in Morocco for k=2. Each pie chart corresponds to a sampling location, identified by a number that corresponds to the population in the graphic below. Rivers are here represented as blue lines. **Bottom:** Structure output for k=2, each vertical bar corresponds to a individual, where the proportion of colours indicates its probability of assignment to the corresponding cluster. Black lines separate sampling locations.

In a recent study, Stuckas *et al.* (2014) uncovered two major groups for the lineage VI of *E. orbicularis*, one in North Africa, with a higher number of haplotypes and a more structured network, and a second group, derived from Morocco, confined to the Iberian Peninsula with a widespread and common haplotype (VIa) and several other satellite haplotypes differing in only one or two nucleotides, a pattern that is very typical of lineages that are undergoing a demographic expansion. The low haplotype diversity found is expected as turtles tend to have slower evolutionary rates when compared to other vertebrates (Avise *et al.* 1992). Also, Stuckas *et al.* (2014) uncovered deeply divergent lineages in North Africa, between Morocco and the Eastern Maghreb. These results seem to indicate that *E. orbicularis* have been present in North Africa longer than on the Iberian Peninsula.

In light of our results, the hypothesis of a North African origin for *E. orbicularis* seems very plausible. This hypothesis is also supported by the microsatellite data where a higher allelic richness was found in North Africa even though genetic diversity is lower. Widmer & Lexer (2001) suggested that potential glacial *refugia* will present higher Allelic Richness than the recently colonized regions, as several alleles might disappear through genetic drift (Austerlitz *et al.* 1997). Nonetheless, this might not be true for heterozygosity as the genetic drift acts faster on allelic richness than on the expected heterozygosity (Widmer & Lexer 2001). We should not rule out the possibility that the low levels of genetic diversity observed in Morocco, especially in Moulay Abdesalam, might be associated with possible anthropogenic pressures (Fahd *et al.* 2009; Velo-Antón *et al.* unpublished data).

2.5.2 Genetic relationships at both sides of the Gibraltar Strait

Here we discuss two distinct hypotheses for the presence of shared haplotypes between Morocco and the southernmost Iberian Populations (Cádiz and Málaga): 1) either by natural transmarine dispersal of Moroccan individuals to the Iberian Peninsula or by 2) Human mediated introductions.

Transmarine dispersal as been observed in other Chelonians (Caccone *et al.* 1999; Gerlach *et al.* 2006; Vamberger *et al.* 2014), and several other cases document possible transmarine migration through the Strait of Gibraltar (Carranza *et al.* 2006b; Kaliontzopoulou *et al.* 2011). The distance between both sides of the Strait is relatively small (14 km at the present to 4-5 km in some glacial cycles; Brandt *et al.* 1996; Zazo

1999), facilitating the dispersal of individuals between continents. Similar scenarios can be observed in chameleons (Paulo *et al.* 2002), as haplotypes found in Málaga populations are shared with Eastern Moroccan haplotypes, and in *Testudo graeca* (Graciá *et al.* 2013b), where shared haplotypes are found between North-African and Spanish populations. Nonetheless, in *E. orbicularis* as in the two previous examples, it is hard to disentangle possible transmarine dispersal movements from human mediated introductions. Terrapins were part of the primitive Mediterranean cultures alimentary habits (Blasco *et al.* 2011), and are amongst the most requested animals in pet trade (Moll & Moll 2004). Due to the long longevity of these animals, it is not uncommon for pet owners to release them in the wild, resulting in the admixture of native and non native lineages in one population as, for example, in the Iberian Peninsula, Velo-Antón *et al.* (2011a) found evidences of translocated animals from the Lineage I (distributed all around the Black Sea and from Poland eastwards) in wild populations.

2.5.3 Genetic structure in Morocco

Stuckas *et al.* (2014) assessed for the first time the genetic structure of Moroccan populations, where they found two genetic clusters. A first group corresponding to the two locations in the Rif (n=9 for the Rif), while the second group aggregated the Middle Atlas (n=14) with the individuals from the Kenitra Province (Ghrab plains, n=2), connecting two geographically distant populations. Our results show a somewhat contradictory pattern, while the same number of genetic clusters for Morocco was observed. The two sampled populations in the Rif Mountains seem to be genetically distinct, with the eastern population of Fifi being aggregated with the Middle Atlas. The observed differences between both studies might be explained by the differences in samples size and chosen markers (Pritchard *et al.* 2000).

Our results show the presence of a possible barrier to gene flow in the Rif Mountains. This pattern might result from a natural barrier, as, for example, in the North African fire salamander (*Salamandra algira*), two different subspecies (*S. algira algira* and *S. algira tingitana*) co-exist in the Rif Mountains, but no geneflow was detected between both margins of the Oued Laou (Beukema *et al.* 2010). Nonetheless the widespread agriculture practiced in the region, especially for the production of cannabis (Labrousse & Romero 2001) and consequent dissection of water bodies (Velo-Antón *et*

al. unpublished), might constrain potential gene flow between the Rif populations, leading to the genetic isolation of Moulay Abdesalam.

A link between the Middle Atlas and populations from Northern Morocco was also detected in Stuckas *et al.* (2014), where samples from the Ghrab plains are assigned to the same genetic cluster as the Middle Atlas. In this study, we show that some populations in the Rif might be linked to the Middle Atlas as well, hinting for a reminiscent connectivity between the Rif (east of the Oued Laou) and the Middle Atlas. These two mountains systems are separated by the Oued Sebou, a river that has its source in the Middle Atlas and runs to the Atlantic Sea, that presents a very extensive hydrological basin (Shahin 2002). Several tributary rivers, namely the Oued Ouerrha, link the Rif to the Oued Sebou and consecutively to the Middle Atlas and the Ghrab plains. It is possible that *E. orbicularis* is taking advantage of this hydrological basin as a corridor to disperse and maintain gene flow between the two mountains and the Ghrab plains. In fact, water bodies (permanent or temporary) are thought to be the main pathway for dispersal in semi-aquatic organisms (Bilton *et al.* 2001; Campbell Grant *et al.* 2010; Velo-Antón *et al.* 2014). Nonetheless, the ability for *E. orbicularis* to disperse such long distances needs to be further evaluated, possibly through the combined used of molecular tools and telemetry.

2.5.4 Conservation Implications

E. orbicularis is the rarest testudine species in Morocco (Pleguezuelos *et al.* 2010), with a very fragmented distribution, where few populations are known in the Rif Mountains and Middle Atlas, and few isolated individuals observed in the Ghrab plains (Fahd *et al.* 2009; Velo-Antón *et al.* unpublished). In Morocco, the habitat transformation due to the expansion of agriculture areas, especially in the Rif with the extensive cannabis plantation, water pollution and desiccation, and the increased desertification are some of the most concerning threats for the populations of *E. orbicularis*. In light of our results, the most vulnerable region in Morocco is in fact the Northwest region of the Rif Mountains, where anthropogenic impacts also threaten viper populations (Brito *et al.* 2011). Further sampling in the region is required to determine the extent of the genetic cluster here found, in order to establish possible conservation measures in the region. The populations in Fifi and Sidi Mimoune present

high values of genetic diversity and allelic richness, providing no evidence for the decline of genetic diversity in both regions.

Chapter 3: Manuscript 2

Range expansions shapes the genetic structure of the European pond turtle, *Emys orbicularis occidentalis*, via the re-colonization of the Iberian Peninsula

Authors:

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

Affiliations:

[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]CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Rua dos Bragas, 289. 4050-123 Porto, Portugal

Abstract

Past climatic fluctuations have molded species´ range distributions. Harsh climatic periods force species to retract their range to suitable areas that serve as *refugia*, and from which subsequent population expansions occurred in response to climate amelioration. These expansions are usually characterized by sequential founder events, which may lead to the surfing of rare alleles at the fringe of the expansion, allowing their fixation in newly occupied territories. Nonetheless, more empirical studies are needed to support the theoretical background on how range expansions impact the patterns of genetic structure.

The European pond turtle, *Emys orbicularis*, invaded North Africa from the Iberian Peninsula, where it is thought to have gone extinct due to past climatic oscillations. The re-colonization of the Iberian Peninsula and subsequent population expansions northwards make this species a good model to study the role of demographic expansions and landscape barriers in shaping the genetic structure of the Ibero-Maghrebian lineage of *Emys orbicularis*. In this work we aim to identify: 1) signals of allele surfing shaping the genetic structure of Iberian populations, and 2) barriers to gene flow that may enhance the effect of founder effects typical of range expansions.

We use a dataset of 453 genotyped individuals (7 microsatellites) from 21 populations distributed throughout the Ibero-Maghrebian lineage. We inferred the genetic structure of this lineage and obtained two major groups: one group spanning from North Africa to the south of the Iberian Central Mountains, and the second group occupying the areas north of these mountains. 26 out of the 108 alleles identified in this dataset showed signals of allelic frequency clines, with a pronounced shift of allele frequencies north of the Iberian Central Mountains, indicating that north Iberian populations might have had more pronounced founder effects due to the Iberian Central Mountains acting as a barrier to dispersal. Furthermore, we found evidences for a coastal corridor for dispersal in the western coast of the Iberian Peninsula.

Keywords: Range expansions, genetic structure, *Emys orbicularis*

3.1 Introduction

Throughout time, species distribution range have suffered consecutive expansions and retractions shaped by the sequential glacial-interglacial cycles (Hewitt 2004). During glacial periods, species tend to contract their range and settle in suitable areas that acted as refugia. After climate amelioration, species respond by expanding back their ranges, following the habitat availability.

Range expansions, especially when characterized by Short Distance Dispersal (SDD), will cause founder effects due to genetic drift in the expansion front (Austerlitz *et al.* 1997). Such events will reduce the genetic diversity in the new populations in the edge of expansion while promoting genetic structuring (Nei *et al.* 1975; Hallatschek *et al.* 2007; Hallatschek & Nelson 2008; Excoffier & Ray 2008; Excoffier *et al.* 2009). During the expansion, new mutation and/or extant alleles that are present at the wave front might "surf" the wave of expansion (Edmonds *et al.* 2004; Klopfstein *et al.* 2006). Surfing alleles might tend to either reach very high frequencies or even be fixed in newly colonized regions. This phenomena result in allelic frequency clines (AFC).

The decline in genetic diversity throughout the expansion axis might constrain the adaptability of edge populations to cope with new ecological conditions (Bridle & Vines 2007). Also, allele surfing might play a role in the evolutionary history, as advantageous and deleterious alleles may also surf the wave of expansion (Travis *et al.* 2007; Excoffier *et al.* 2009), allowing for either an accelerated adaptation to new environment conditions at the edge of the expansion or a slower rate of the expansion due to higher extinction rate brought by the surfing of deleterious alleles.

Up to the moment, most of the knowledge on the impacts of range expansions is derived from computer models and simulations (Hallatschek & Nelson 2008) and micro-cosmos experiments (Hallatschek *et al.* 2007), leaving a gap in the empirical knowledge available to support the theoretical background, where most of the work is focused in human mediated introductions (White *et al.* 2013; Berthouly-Salazar *et al.* 2013), with few examples on natural range expansions (Swaegers *et al.* 2013).

The European pond turtle, *Emys orbicularis* (Linnaeus, 1758), is distributed throughout a wide range, covering northwestern Africa and a great area in Europe, where nine distinct lineages based on the mitochondrial DNA cytochrome *b* gene are

present (Stuckas *et al.* 2014). The Ibero-Maghrebian lineage, *Emys orbicularis occidentalis* has a patchy distribution across most of the Iberian Peninsula and North Morocco (Velo-Antón *et al.* 2008). Phylogenetic studies have suggested that the ancestor of this lineage colonized North Africa, from an unknown location in Europe, where there was a phase of diversification with a posterior re-colonization of the Iberian Peninsula, where *Emys orbicularis* had probably gone extinct due to past climate oscillations (Stuckas *et al.* 2014). The hypothesis that the current populations of *Emys orbicularis occiddentalis* are derived from North Morocco is supported by the fact that haplotypes found in the Iberian Peninsula form a start shape network, indicative of a expanding population, and Moroccan populations show a higher haplotype diversity in the latter (Stuckas *et al.* 2014). Also there is a decreasing pattern of genetic diversity in the south- north axis in the Iberian Peninsula (Velo-Antón *et al.* 2008), with a strong genetic structure observed in the northern fringe of the range.

In this work we aim to identify:

1) Signals of allelic frequency clines shaping the genetic structure of Iberian populations. We hypothesize that if some alleles have an accentuated shift in frequency following the axis of expansion, with either an increase through allele surfing or a decrease in frequency, a high differentiation between northern and southern populations is expected. This will in turn accentuate the genetic structure and promote the “southern richness to northern purity” pattern reviewed in Hewitt (2000).

2) Barriers to gene flow that may enhance the founder effect typical of range expansions. We hypothesize that major geographic barriers (e.g. Strait of Gibraltar, Iberian Central System), might funnel the direction of expansion, since, as seen in *Mauremys leprosa*, mountain chains might pose a strong barrier to gene flow in pond turtles (Fritz *et al.* 2006). If any geographical barrier effect is evident in *E. orbicularis occidentalis* we expect to observe stronger allelic frequency clines north of the barrier, since the allele availability in the northern side of the barrier would be constrained by a reduced migrant load, promoting even stronger allele surfing events.

3.2 Material and Methods

3.2.1 Sampling, DNA extraction and dataset preparation

A total of 164 blood or tissue samples from both Morocco (66 samples from 3 localities) and the Iberia Peninsula (98 samples from 8 localities), covering all *E. orbicularis occidentalis* range, were here genotyped for the first time. The genotypes produced were then merged with a previous published microsatellite dataset for the Iberian Peninsula (Velo-Antón *et al.* 2008), resulting in a complete dataset of 453 samples, with 387 genotypes from 18 Iberian populations and 66 from 3 Moroccan populations (Figure 3.1; see supplementary material table 7.3).

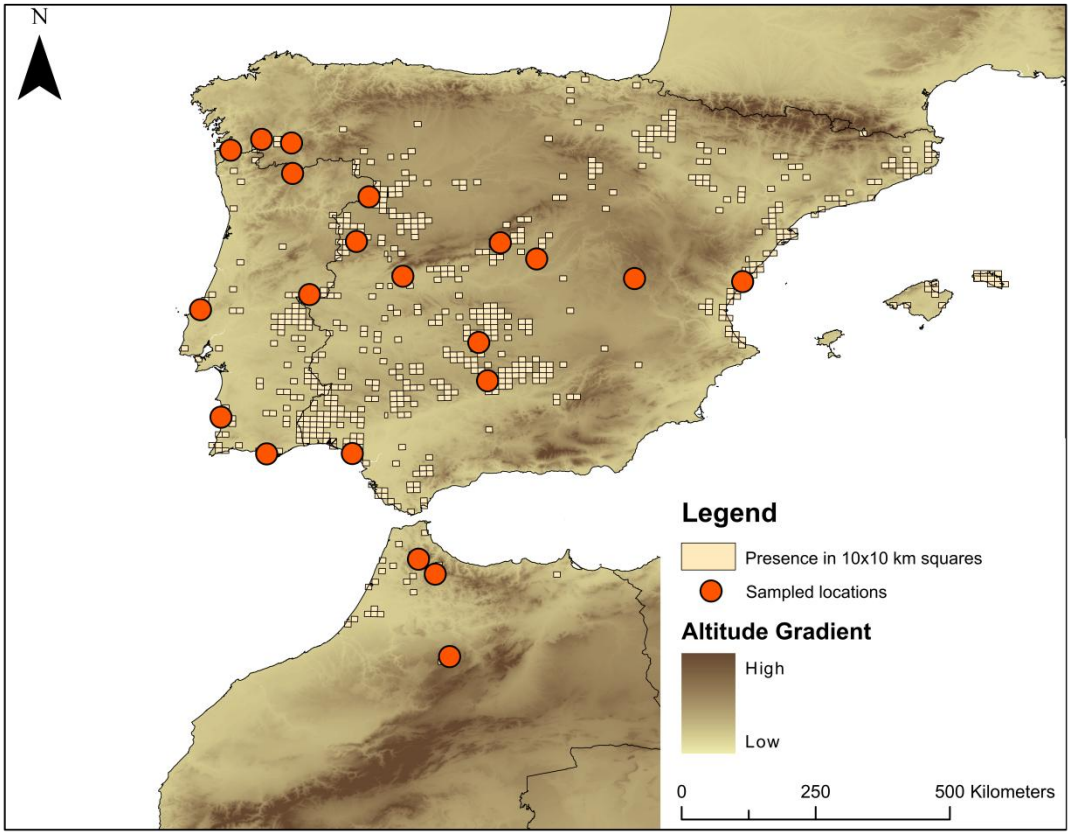

Figure 3.1: Current distribution of the subspecies *Emys orbicularis occidentalis* (as denoted by the 10x10km squares in beige). The sampled locations are marked with red circles.

Genomic DNA was extracted from both blood and tissue samples using the EasySpin commercial kit, following the manufacturer's protocols, while extending lyses time to maximize the yield.

3.2.2 Microsatellites amplification

Seven hyper-variable loci were amplified (Gmu: B08, D16, D51, D87, D88, D93, D114). The corresponding primers were first developed for *Glyptemys muhlenbergii* and tested for cross-amplification in *Emys orbicularis* (King & Julian 2004). For each locus a third primer was used, following the M13 tailed primer method (Oetting *et al.* 1995). This primer was labeled with FAM, NED, VIC or PET depending on the selected dye for each locus, and was used at the same concentration as the reverse primer. To adjust for this, forward primers were used at a ten-fold dilution.

PCR reactions were divided into two multiplex assays (Supplementary material Table 7.1), and performed at a final volume of 10 µL with 5µL of Quiagen Multiplex PCR Kit, 1µL of DNA, 1µL of the primer mix, with tails and primers (at 10µM) and 3 µL of ultra-pure water. For each multiplex a touchdown PCR reaction was conducted under the following conditions: initial denaturation at 95ºC for 15 minutes; 9 cycles with a denaturation step at 95ºC for 38 seconds, annealing at 60 ºC for 1 minute (decreasing 0.5ºC each cycle) and extension at 72ºC for 30 seconds; 23 cycles of 95ºC for 30 seconds, 56ºC for 1 minute, 72ºC for 30 seconds; 8 cycles of 95ºC for 30 seconds, 53ºC for 1 minute, 72ºC for 30 seconds, followed by a final elongation step at 60ºC for 30 minutes. PCR reactions were performed on a BioRad T100 Thermocycler and genotypes were obtained from an ABI 3130xl genetic analyzer (Applied Biosystems, Foster City, CA, USA). GeneScan^{TM}-500 Liz was used as size standard. Allele scoring was performed in GeneMapper v4.0 (Applied Biosystems).

In order to standardize allele scores between the two datasets, a small set of 10 samples from Velo-Antón *et al.* (2008) were re-genotyped.

3.2.3 Microsatellites data analysis

Deviations from Hardy-Weinberg equilibrium and tests for linkage disequilibrium across all populations and loci were performed in GENEPOP v4.2 on the web (Raymond & Rousset 1995; Rousset 2008). Genalex 6.5 (Peakall & Smouse 2006, 2012) was used to calculate genetic diversity indexes such as observed heterozygosity (H_O), expected heterozygosity (H_E), Number of effective alleles (Na) and Allele

frequencies per population. Allelic richness was calculated in HP-Rare (Kalinowski 2005), using rarefaction, for all populations with more than 10 individuals. Pairwise F_{ST} values were calculated in Arlequin 3.5 (Excoffier & Lischer 2010).

3.2.4 Spatial patterns of genetic diversity

STRUCTURE (Pritchard *et al.* 2000) was used to identify the potential number of genetic groups found across the study area. To do so, we run STRUCTURE from k=1 to k=20, where k=20 was set as an arbitrary highest value that allowed to account for substructure inside populations, with a burn-in length of 1000000 steps with 4000000 iterations after the initial burn-in, for each k four independent iterations were run. We also accounted for potential admixture between genetic groups and correlated allelic frequencies. The Evanno's method (Evanno *et al.* 2005), implemented in Structure Harvester (Earl & von Holdt 2012), was used to determine the most likely number of genetic clusters present in our dataset.

A Spatial Principal Component Analysis (sPCA; Jombart *et al.* 2008) was performed to evaluate the existence of spatial patterns of genetic variability using the *adegenet* package (Jombart 2008) in R (R Core Team 2012). To do so, the allele frequencies of each population were used. A Delaunay triangulation connection network was used. This network links populations by creating triangles where: 1) Nodes (populations) are exclusively located in the triangle vertices; 2) the interiors of the triangle are pairwise disjointed; 3) The union of triangles represents the smallest convex that contains all populations and 4) the interior of the circumcircle of each triangle contains no node. The choice of this connection network was based on the lack of samples on the east coast of the Iberian Peninsula, which potentially constrained the connectivity alongside the region. To identify if the sPCA scores were able to assess global (found in eigenvalues with high variance and positive autocorrelation) and local (found in eigenvalues with high variance and negative autocorrelation) structure, two separate Monte Carlo tests were performed with 100000 randomizations each. Selected eigenvalues were then interpolated for the study area to identify abrupt shifts in allelic frequencies that could result in separated genetic groups. Furthermore, the weight that each allele had in the observed patterns was calculated.

To identify spatial patterns of the inferred genetic distances, the resulting pairwise F_{ST} matrix was spatially interpolated. To do so, each column of the matrix was treated as a different variable, as each column represents the genetic distance of one

population to all other populations, and underwent interpolation using the kriging interpolation method (Oliver & Webster 1990) implemented in the "Geostatistical Analyst" extension in GIS ArcMap 9.3. To summarize the results, a Principal Component Analysis was performed using the Raster package (Hijmans 2014) for R (R Core Team 2012; see supplementary material Script 1 for the script used).

3.2.5 Allele frequency clines and trends of genetic diversity along the axis of colonization

A sliding window approach was used to evaluate shifts in allele frequencies and expected heterozygosity (H_E) throughout the south-north expansion axis. To do so, a window with a height of two decimal geographical degrees and spanning the entire longitudinal range slide across the study area, starting in the southern population as its center, and then moving 0.2 degrees north each step until the center of the window reaches the northernmost population. At each window the mean of the frequency for each allele and heterozygosity index of all populations present in the window were calculated. Even though this method disregards the potential radial effect of the expansion (Hallatschek *et al.* 2007), it compensates for stochastic differences between populations, allowing the identification of global patterns of variation throughout the colonization axis.

Linear regressions were used to summarize the trend in both expected heterozygosity and allelic frequencies. Even though, in the latter, clines might not possess a truly linear distribution, especially if any major barrier to dispersal constricted gene flow and promoted new founder effects. Nonetheless, linear regressions allow to summarize the overall trend. Also, a smooth curve was use for each allele to facilitate the trend assessment by eye.

Following the work of Currat & Excoffier (2005) and Klopfstein *et al.* (2006), we considered an allele as surfing if the p-value for the linear regression was statistically significant with a very strict alpha value ($\alpha = 0.001$). Also, we only considered as surfing alleles when the difference between the maximum frequency and the minimum frequency per population was at least 0.2 (in order to exclude alleles that were stable across the study area). This method was implemented in R (R Core Team 2012; see supplementary material Script 2 for the script used). Furthermore, all alleles had their frequencies interpolated through the kriging interpolation method (Oliver & Webster 1990) implemented in the "Geostatistical Analyst" extension in GIS ArcMap 9.3.

3.3 Results

3.3.1 Microsatellites analysis

From the seven microsatellites used in 453 individuals, 108 different alleles were sampled. The mean number of alleles per locus was 15, with a minimum of 5 (GmuB08) and a maximum of 21 (GmuD16). No evidences of linkage disequilibrium or deviations from Hardy-Weinberg equilibrium were found between all combination of loci and populations.

The highest value of genetic diversity across the study area was found in Doñana (H_E: 0.80), while the lowest was found in Ribadavia (H_E: 0.55). As for allelic richness the highest value here observed was found in Castelo Branco (Ar: 7.33) and the lowest value was again found in Ribadavia (Ar: 3.23). Overall, the mean genetic diversity of African populations is higher (mean H_E: 0.723) than that of the Iberian populations (mean H_E: 0.693; see Table 3.1).

The sliding window across the axis of expansion revealed a linear decay of genetic diversity (p-value=$6.888xe^{-12}$; r^2=0.796) from south to the north (Figure 3.2).

3.3.2 Allelic Frequency Clines

From the 108 sampled alleles, 15 alleles were private to one population and therefore discarded from the allelic surfing analysis. In the remaining 93 alleles, 26 showed signs of a latitudinal cline on the allele frequency (15 with an increase on the frequency along the axis of expansion, and thus considered as surfing, and 11 with a decrease in frequency (see figure 3.3 for examples; the latitudinal pattern of the 26 alleles can be found in supplementary material Figure 7.2 to Figure 7.6 and the significance of the linear regressions in supplementary material Table 7.4).

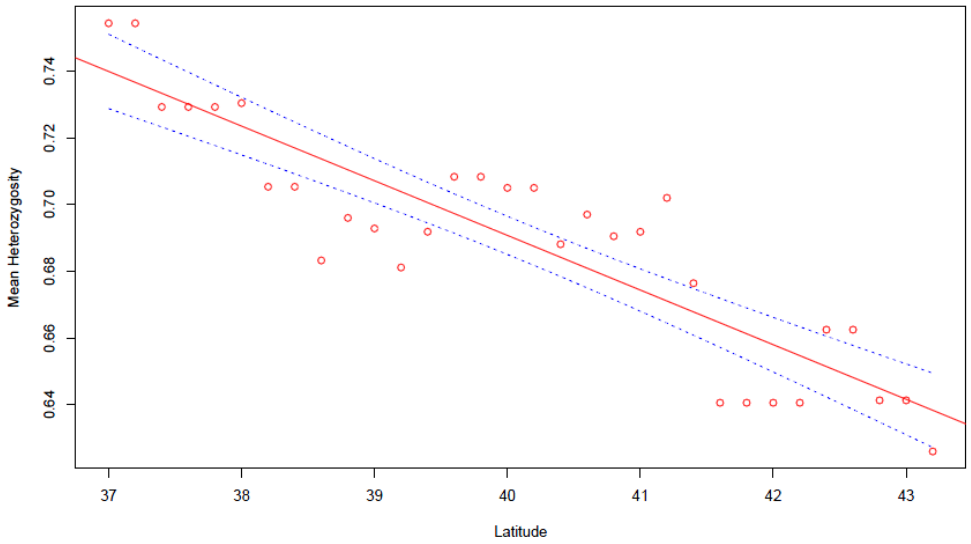

Figure 3.2: Decline of the genetic diversity alongside the south-north axis of expansion. Red circles correspond to the mean heterozygosity retrieved at each window of the sliding window analysis. Red line represents the linear regression used to summarize the trend, and the two dotted lines correspond to the 95% confidence interval.

Table 3.1: A summary of the genetic diversity indices calculated for *Emys orbicularis occidentalis* populations. Lat: Latitude; Long: Longitude; n: number of samples for that location; Na: Number of alleles; Ne: Number of effective alleles; H_O: Observed heterozygosity; H_E: Expected heterozygosity; uH_E: unbiased expected heterozygosity; Ar: Allelic Richness. Populations with less than 10 samples, marked with *, were not used when calculating the Allelic Richness, and were not considerated for any populational inference.

Population	Lat	Long	n	Na	Ne	H_O	H_E	uH_E	Ar
Ribadavia	42.3	-8.11	35	4.143	2.321	0.644	0.547	0.555	3.23
Orense	42.24	-7.61	32	6.571	4.011	0.673	0.705	0.717	5.46
Porrino	42.12	-8.63	30	5.857	3.420	0.727	0.672	0.684	4.5
Boticas	41.73	-7.59	28	6.286	4.269	0.701	0.726	0.740	5.6
Zamora	41.34	-6.31	17	3.857	2.471	0.612	0.553	0.572	3.98
Salamanca	40.59	-6.52	36	8.714	5.062	0.746	0.757	0.768	6.3
Madrid	40.57	-4.11	31	7.286	5.208	0.759	0.775	0.788	6.14
Argana*	40.3	-3.5	5	5.286	4.199	0.714	0.729	0.814	----
Robledillo*	40.01	-5.74	3	3.286	2.678	0.619	0.611	0.733	----
Cuenca	39.97	-1.86	18	4.286	2.982	0.665	0.652	0.671	4.04
Valencia	39.92	-0.05	24	7.286	4.551	0.756	0.721	0.739	5.7
Castelo Branco	39.70	-7.31	12	7.571	5.503	0.691	0.778	0.816	7.33
Paul da Tornada*	39.45	-9.13	5	4.143	2.965	0.543	0.617	0.686	----
Ciudad Real	38.9	-4.47	22	6.857	4.586	0.782	0.735	0.754	6.78
Cardena*	38.26	-4.32	6	4.429	3.287	0.548	0.654	0.719	----
Almograve	37.65	-8.79	27	7.286	4.515	0.763	0.727	0.743	6.75
Doñana	37.04	-6.59	36	8.429	5.532	0.821	0.800	0.811	6.42
Algarve	37.03	-8.03	20	7.857	4.804	0.759	0.736	0.756	6.26
Moulay Abdesalam	35.27	-5.48	24	4.714	3.131	0.763	0.663	0.678	4.26
Fifi	35.02	-5.2	26	8.429	6.005	0.777	0.757	0.772	6.59
Sidi Mimoun	33.64	-4.96	16	7.286	4.444	0.765	0.751	0.775	6.02

3.3.1 Genetic Structure

According to the Evanno Method, the most likely number of genetic clusters found in our dataset is two (k=2; Supplementary material Figure 7.1). Under k=2, Moroccan and Southern Iberian populations are clustered together while the Northern Iberian populations are grouped into a distinct cluster (see Figure 3.4). The two following most likely possibilities that arised to explain the genetic structure were k=6 and k=16. Under k=6, the three Moroccan populations are grouped together, while in the Iberian Peninsula two clusters seem to appear in the northern populations, a third group covers the western coast of the Peninsula with the exception of the Almograve population and the two remaining groups are spread across the region (see Figure 3.5). As of k=16, in Morocco, the three populations are split into individual clusters. Regarding the Iberian Peninsula, with a few exceptions in the central Peninsula and in the western Coast, all populations are assigned to their own cluster (see Chapter 4 Figure 4.1).

When performing the sPCA, the presence of global structure in the data was significant (p= 0.002), while no statistical significance was found for local structure (p= 0.714). The two first positive eigenvalues addressed around 12% of the variance. The first axis splits the data into two groups. A first northern group composed of all the populations north of the Iberian Central System, with a second group composed of all southern Iberian populations and Morocco (Figure 3.6). A closer look to the weight of each allele to explain the observable pattern revealed that 20 out of the 25 most important alleles to explain this pattern were significant for the test of allelic frequency cline (Supplementary Material figure 7.3; Supplementary material Figure: 7.7). The second most important eigenvalue seems to indicate a relationship between the populations of the Western Coast of the Peninsula, aggregating them in a group (Figure 3.7, supplementary material figure 7.8). This pattern is also shown by the interpolated F_{ST} distances, with a high correlation of the genetic distances between these populations (Figure 3.8).

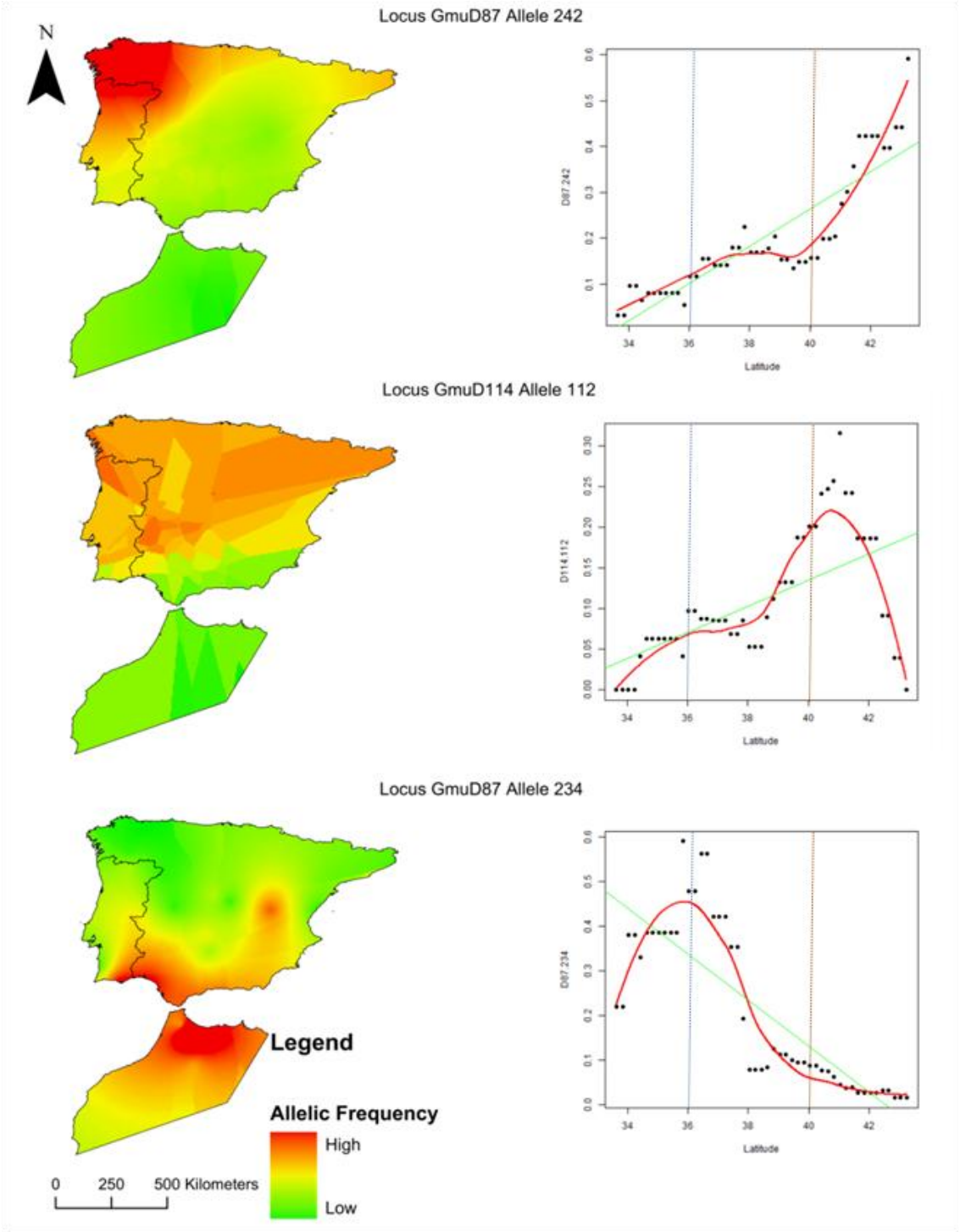

Figure 3.3: This figure summarizes the overall trends found in the allelic frequencies. **Top:** A clear sign of allelic surfing, where an allele with very low frequencies at the core of the expansion, presents very high frequencies at the edge of the expansion. **Middle:** A common pattern in several of our alleles, this allele hints for a possible role of a barrier to constrict the surfing of alleles in the wave front, it is worth mentioning that due to the rapid shifts in the allele frequencies, the resulting interpolations are very weak and present several artefacts. **Bottom:** With the increase of frequencies of surfing alleles. Several alleles have their allelic frequencies decaying throughout the expansion axis. The graphics on the right side were produced through the analysis of the allelic frequency with a sliding window. The interpolations on the left were based on the allele frequencies observed in each population. Blue dashed lines on the figures on the right indicate roughly the position of the Gibraltar Strait, Brown dashed lines indicate roughly the position of the Iberian Central System.

3.4 Discussion

3.4.1 Impacts of the range expansion in the genetic structure

In previous studies on *Emys orbicularis*, using mitochondrial DNA (Chapter 2; Stuckas *et al.* 2014), the strait of Gibraltar seemed to act as a barrier to gene flow between Morocco and Iberian Peninsula, and only recently (see Chapter 2) shared haplotypes between the two continents were found. However, when looking to microsatellite data, we do not observe the same pattern. In fact, Structure, under k=2, segregates both Southern Iberian populations and Morocco from the Northern Iberian populations (Figure 3.4), and the same pattern is retrieved by the first component of the sPCA (Figure 3.6). Contrarily, this pattern is not observed in other Testudines species inhabiting the same region (*Mauremys leprosa*, Veríssimo 2014; *Testudo graeca*, Graciá *et al.* 2013). Two different hypotheses might explain this pattern. First, a natural or human mediated dispersal of individuals from Morocco to the Iberian Peninsula might lead to gene flow between the two groups, leading to the reduction of the genetic distance between southern Iberian and Moroccan populations. In fact, the three shared haplotypes found between Málaga and Morocco seem to support this hypothesis. Nonetheless, such hypothesis would require a massive translocation of individuals between both continents in order to homogenize the gene pool of both regions, which is unlikely. As for the second hypothesis, the observed structure might be the result of the recent expansion that this species went over. As shown in the work of Hallatschek *et al.* (2007) and later discussed by Excoffier & Ray (2008), genetic drift at the edge of the expansion wave might promote clines in allelic frequencies across the expansion axis, leading to the substructure between the edge populations and those at the core or origin of the expansion. Our findings of a decline in genetic diversity (H_E) along the expansion axis (Figure 3.2) supports the hypothesis that the populations of *Emys orbicularis* in the Iberian Peninsula underwent a recent range expansion (Austerlitz *et al.* 1997; Hewitt 1999, 2000; Excoffier *et al.* 2009). Furthermore, the identification of 26 alleles with signs of allelic frequency clines (AFC) throughout the expansion axis, alongside the fact that 20 of those are amongst the 25 alleles which contribute to the pattern observed in the first component of the sPCA, support the hypothesis that the

present genetic structure is a product of the genetic drift that occurred during the range expansion.

Overall, the fact that 1) we observe a significant decrease in genetic diversity along the axis of expansion; 2) several allelic frequency clines are observed throughout unlinked and neutral loci; and 3) stronger genetic structure is observed in edges of the expansion, are strong signals that support the hypothesis that *E. orbicularis* genetic structure is promoted by range expansion.

3.4.2 Impacts of geographical barriers on the range expansion dynamics

Geographical features, such as mountain chains, may act as strong barriers to gene flow for several *taxa*. For example, the Atlas Mountains in Morocco, a mountain system extending across 2500 km, from Morocco to Tunisia, and with its maximum height at 4167 meters, as been proposed as a barrier for several species (*Mauremys leprosa,* Fritz *et al.* 2006; *Agama impalearis,* Brown *et al.* 2002).

A closer look into the unveiled genetic structure of *Emys orbicularis* show that the two different groups seem to be separated by the Iberian Central System. Some alleles show a curious pattern that might be indicative of such. For example, the allele 112 of the Locus GmuD114 show a steep increases in its frequency from the south until around the 40º latitude mark, where the Iberian Central System is located. Afterwards, it rapidly decays in the northern populations (Figure 3.3 middle). Such pattern is present in several other alleles, even though some of them were not statistically significant for an AFC, and might be indicative of a second founder effect, leading to the lost of these alleles by a second genetic drift event. The importance of the Iberian Central System as a barrier is well known for several other reptiles and amphibians species (*Hyla meridionalis*, Recuero *et al.* 2007; *Triturus pygmaeus* and *T. marmoratus* García-París *et al.* 2001; *Vipera latastei* and *Vipera Monticola,* Brito *et al.* 2008). Nonetheless, we refrain from taking any definitive conclusions about the role of the Iberian Central System as a barrier to gene flow as our current sampling strategy can be insufficient.

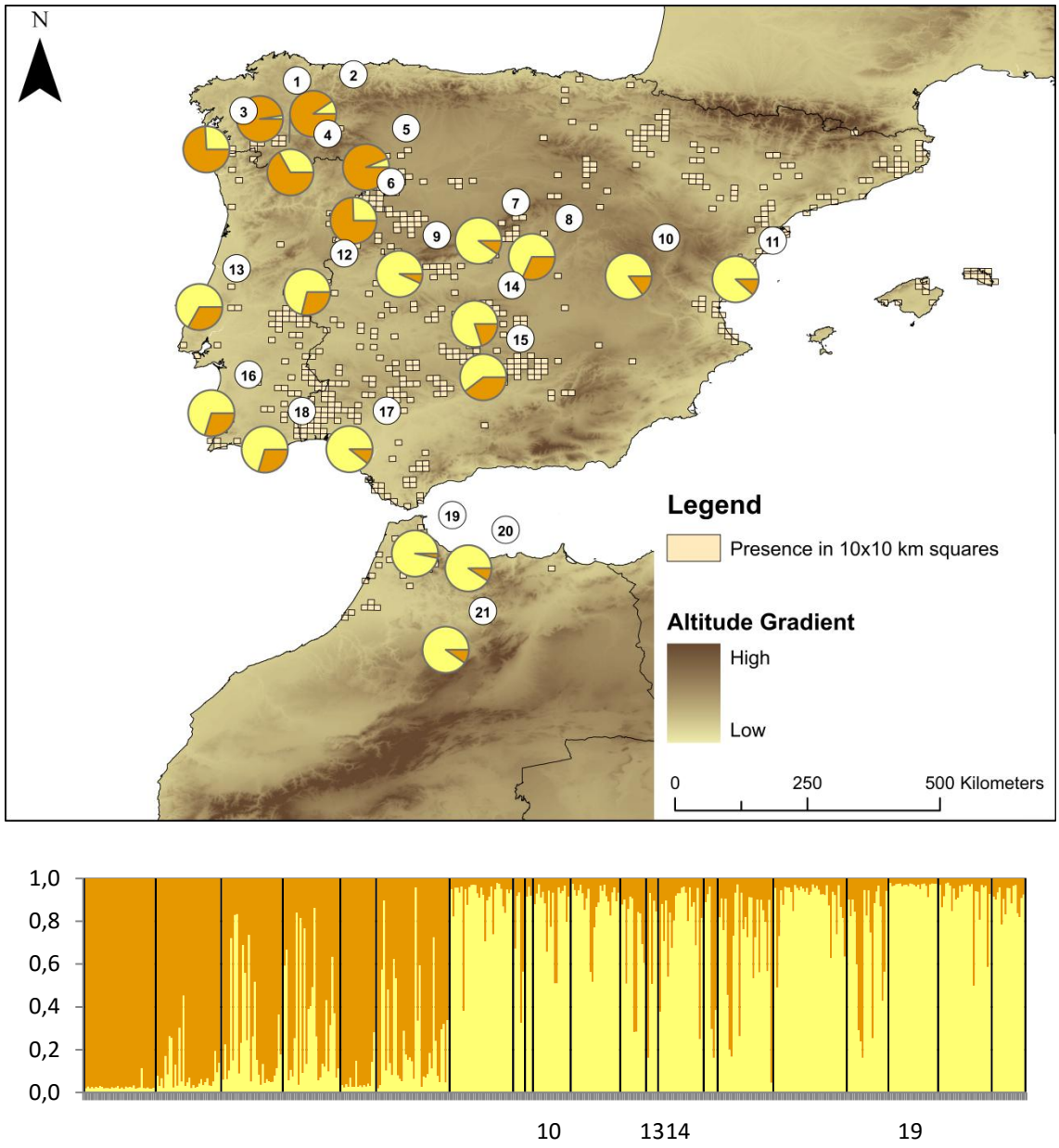

Figure 3.4: Top: Spatial representation of the genetic structure found for the entire distribution of *Emys orbicularis occidentalis* for k=2. Each pie chart corresponds to a sampling location, identified by a number that corresponds to the population in the structure graphic below. **Bottom:** Structure output for k=2, each horizontal line corresponds to a individual, where the proportion of the colours indicates its probability of assignment to the corresponding cluster. Black lines separate sampling locations.

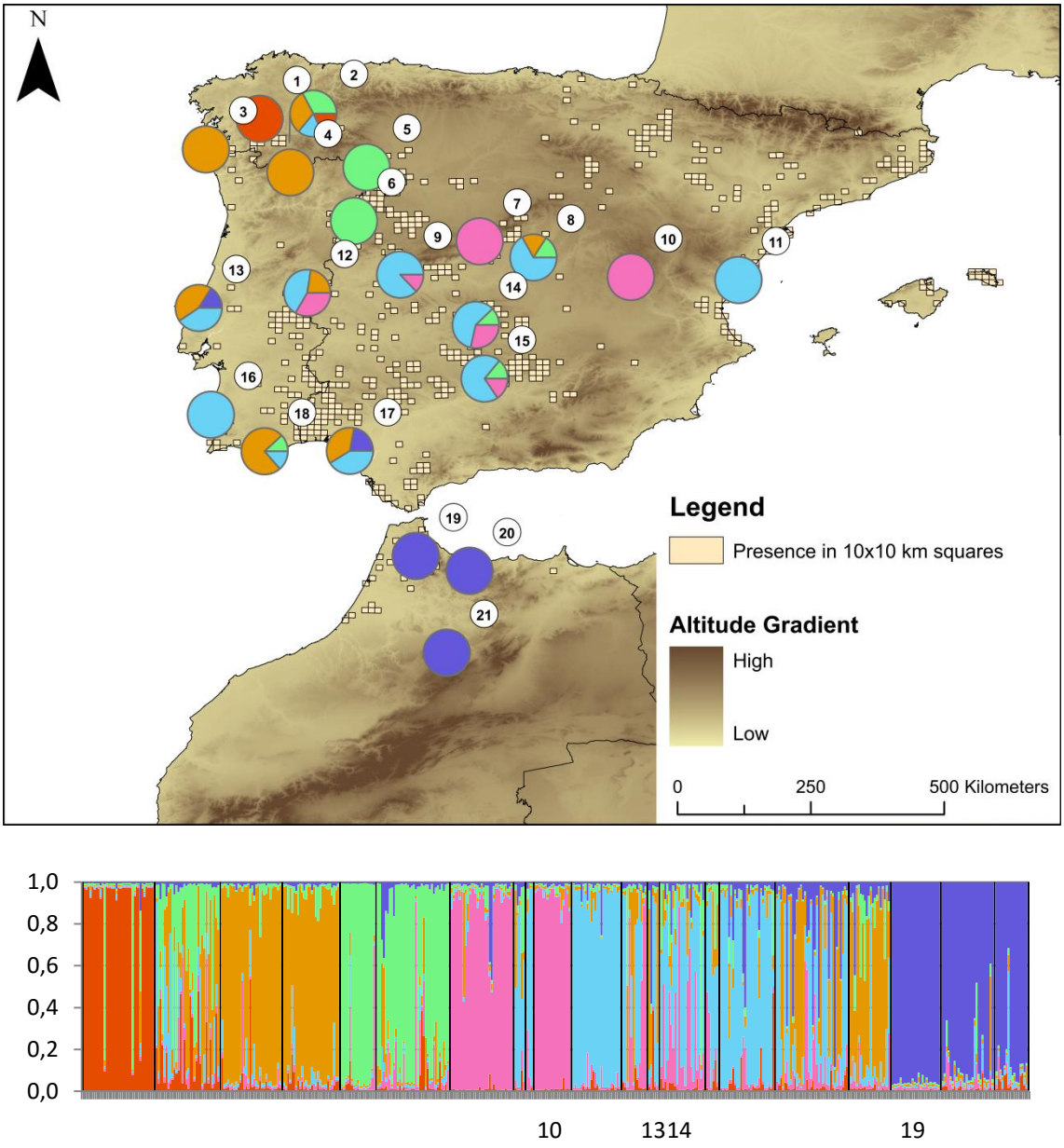

Figure 3.5: Top: Spatial representation of the genetic structure found for the entire distribution of *Emys orbicularis occidentalis* for k=6. Each pie chart corresponds to a sampling location, identified by a number that corresponds to the population in the structure graphic below. **Bottom:** Structure output for k=6, each horizontal line corresponds to a individual, where the proportion of the colours indicates its probability of assignment to the corresponding cluster. Black lines separate sampling locations.

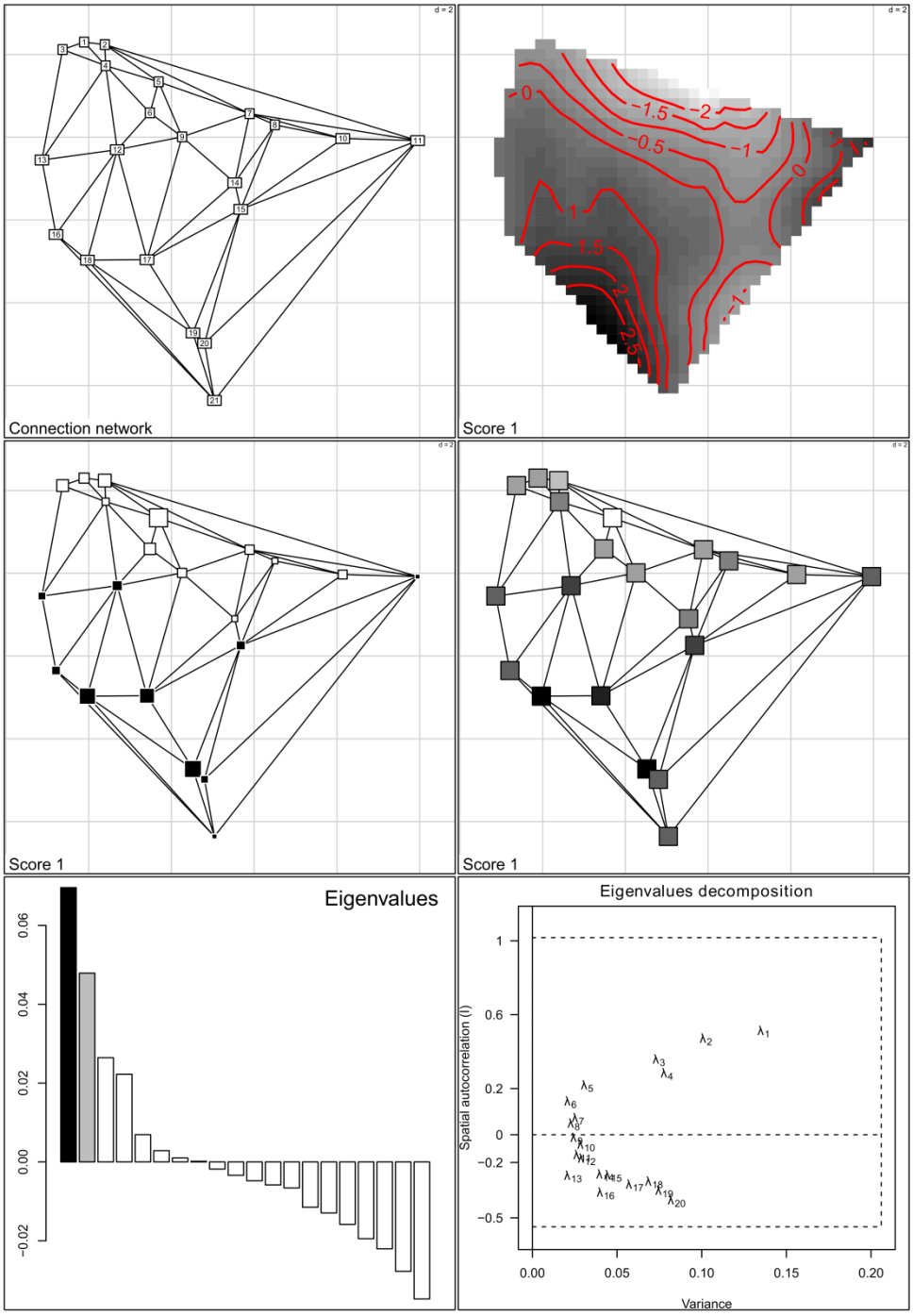

Figure 3.6: Summary display of the sPCA; **Top left:** The resulting connection networker after applying the Delaunay triangulation. This type of network was the only that could detect some type of connection between the Easter Iberian Coast and the South of the Iberian Peninsula; **Top right:** Spatial Interpolation of the scores of the first eigenvalue of the sPCA. The red lines denote regions of abrupt change in allelic frequencies; **Middle Left:** Assignment of the populations to one of the two groups recovered by the method, the method show a segregation between populations at North of the Central System to the others; **Middle Right:** A somewhat similar approach to the one in the left but in this case it uses the lagged scores to interpolate the assignment; **Bottom Left:** This graphic shows how much each eigenvector explains of the data. The eigenvector being observed in all previous graphics is underlined in black, while other selected eigenvectors for the sPCA are shown in light gray. **Bottom Right:** Decomposition of the eigenvalues in accordance in their spatial autocorrelation and Moran's I. We can observe here that, by the fact that component one and two are isolated from other eigenvalues, they are good predictors to infer global structure

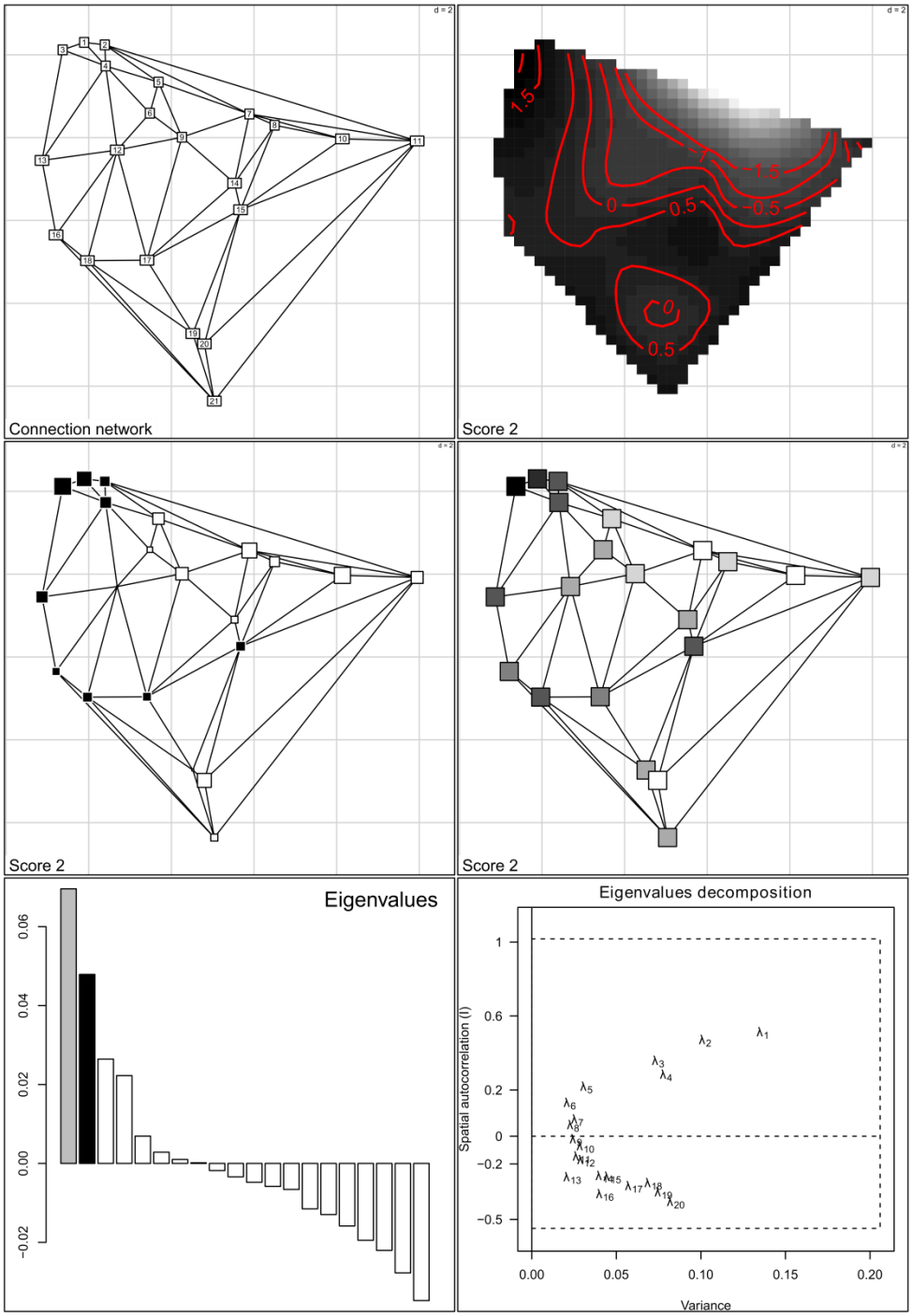

Figure 3.7: Summary display of the sPCA; **Top left:** The resulting connection networker after applying the Delaunay triangulation. This type of network was the only that could detect some type of connection between the Easter Iberian Coast and the South of the Iberian Peninsula; **Top right:** Spatial Interpolation of the scores of the second eigenvalue of the sPCA. The red lines denote regions of abrupt change in allelic frequencies; **Middle Left:** Assignment of the populations to one of the two groups recovered by the method, the division between a group in the Western Coast from other populations might be indicative of a possible colonization route; **Middle Right:** A somewhat similar approach to the one in the left but in this case it uses the lagged scores to interpolate the assignment; **Bottom Left:** This graphic shows how much each eigenvector explains of the data. The eigenvector being observed in all previous graphics is underlined in black, while other selected eigenvectors for the sPCA are shown in light gray. **Bottom Right:** Decomposition of the eigenvalues in accordance in their spatial autocorrelation and Moran's I. We can observe here that, by the fact that components one and two are isolated from other eigenvalues, they are good predictors to infer global structure.

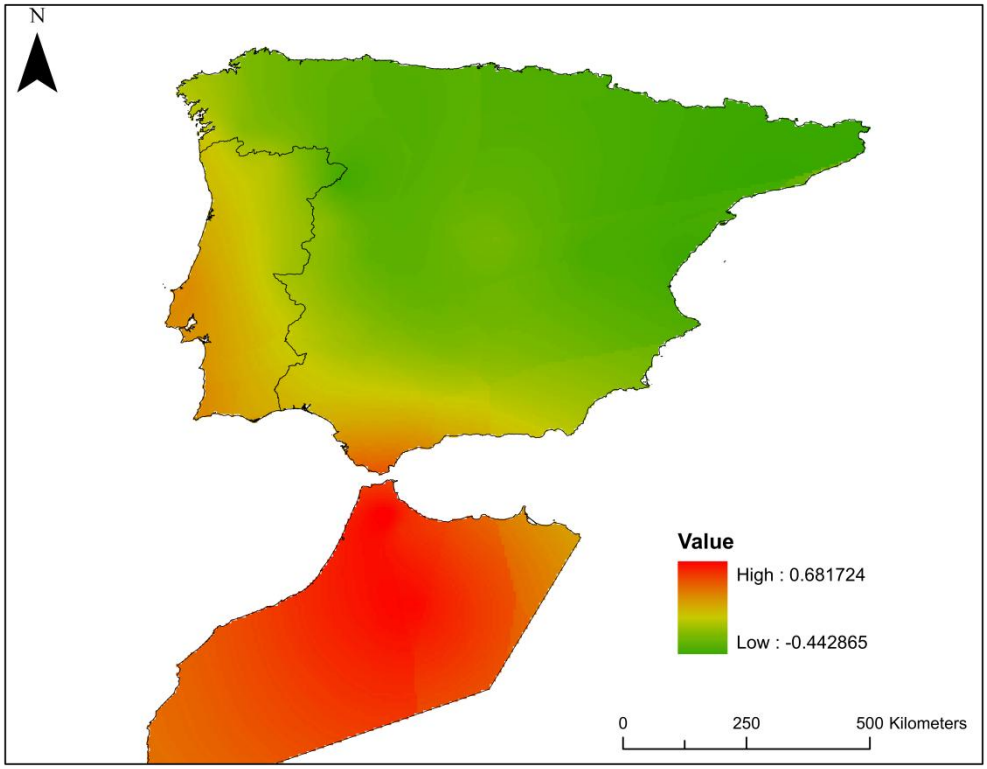

Figure 3.8: First component of the PCA of genetic distances. A strong relationship is found between the western Coast and the Southwestern Peninsula.

3.4.3 The West Coast as a colonization route for the Northwestern populations

The observed pattern in the second component and in the genetic distance interpolation seems to be indicative of a coastal corridor for the colonization of northwestern populations. A reminiscence of this pattern is also observed in the Structure results for k=6 where a close relationship is found between the coastal populations of Porriño, Paul da Tornada, Almograve and Doñana. A recent climate reconstruction for the Iberian Peninsula, from 15 000 until 3 000 years ago (Tarroso *et al.* 2014), shows that the western coast of the Iberian Peninsula underwent the same climatic evolution and was the most stable region in the Iberian Peninsula across millennia. Furthermore, the authors indicate that this region might have been associated with potential dispersal or isolation events, as in fact, several areas of this region have been indicated as potential glacial *refugia* (Weiss & Ferrand 2007). Such genetic structure pattern is depicted in several species such as *Lacerta schreiberi* (Godinho *et al.* 2006) and *Lissotriton boscai* (Martínez-Solano *et al.* 2006). It is then plausible that *Emys orbicularis* could have used this corridor to expand northwestwards.

Chapter 4: Manuscript III

A short note on the geographic allocation of European pond turtle (*Emys orbicularis*) individuals with unknown origin

Authors:

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

Affiliations:

[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]CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Rua dos Bragas, 289. 4050-123 Porto, Portugal

4.1 Abstract

The commercialization of exotic turtles, in order to satisfy increasing demand, had lead to the heavily exploitation of this group during the past years and, as result, around 40% of the recognized turtle species are now considered as Threatened. Furthermore, the increased numbers of exotic individuals liberated in the wild, resulted in several introductions of species that are now considered highly invasive, competing directly with native populations for food.

The European pond turtle, *Emys orbicularis*, is considered a Near Threatened species by the IUCN. Individuals of this species are often illegally captured or reared to be maintained as pets, directly impacting the wild populations that have suffered demographic contractions in the past years. The use of animals located in recovery centres to reinforce wild populations might be a viable course of action to halt the current pattern. Nonetheless, allocating individuals with very different genetic backgrounds might be harmful for the receiving population due to the outbreeding depression.

A genetic tool capable of allocating individuals with unknown origins to their population, would be a good mechanism to avoid genetic homogenization of populations. Here we aim to increase the current baseline knowledge about the genetic patterns observed in the Peninsula and to extend that knowledge to Morocco. We tested the effectiveness of our dataset to identify the population of origin, and obtained a 76% success rate. Furthermore, with this baseline, several individuals of *Emys orbicularis occidentalis* have been re-alocated to their most likely population of origin.

Keyword: *Emys orbicularis*, conservation genetics, assignment tools

4.2 Introduction

Turtles are one of the most threatened vertebrate groups, with *ca.* 40% of the recognized species considered Threatened by the 2012 IUCN Red List (Van Dijk *et al.* 2012). Turtles' populations have been facing several threats, not only regarding their habitat (e.g. habitat destruction and/or fragmentation), but also they are being used as food, in traditional medicine and as pets across several cultures (Van Dijk *et al.* 2000, 2012; Moll & Moll 2004). With the increased exotic turtles pet trade market (Gong *et al.* 2009), when these animals become large, owners often release them in non-native regions (Velo-Antón *et al.* 2007, 2011b). The introduction of exotic species, some of which might display invasive behaviour, such as *Trachemys scripta*, have strong impact on *Emys orbicularis* populations and may act as vectors for the introduction of new pathogens (Servan & Arvy 1997; Cadi & Joly 2003, 2004; Polo-Cavia *et al.* 2010; Hidalgo-Vila *et al.* 2011). Several native species are also often collected from wild populations or reared to be maintained as pets. When these native animals are collected from captivity by supervisory authorities or delivered in official reception centres, they are normally kept and displayed in zoos, once no information about its region of origin is available. Relocation of these animals to native populations can reinforce their population size, but a careful planning should be used to identify the probable region of origin of all individuals, as the introduction of animals from distinct regions may lead to the genetic homogenization of populations, constraining local adaptations and resulting in outbreeding depression and lower short-term population fitness (Storfer 1999). Nonetheless, other key questions for reintroducting animals in the wild, focusing namely in the sanitary state of animals, population and ecosystem issues should be also taken into account (Armstrong & Seddon 2008).

The European pond turtle, *Emys orbicularis* (Linneus, 1758) is one of the two native freshwater turtles native to the Iberian Peninsula (Keller & Andreu 2002; Loureiro *et al.* 2008). It has a very widespread but patchy distribution, ranging from the Eastern Europe to the European Mediterranean peninsulas and North Africa, with nine distinct mitochondrial DNA lineages throughout its range (Stuckas *et al.* 2014). This species is in clear regression throughout its distribution due to multiple threats (Cadi & Joly 2003, 2004; Andueza & Alcayde 2004; Matson *et al.* 2005; Velo-Antón *et al.* 2007, 2011a, unpublished data; Trakimas & Sidaravi 2008; Fahd *et al.* 2009), which rendered the status of *Near Threatened* on the IUCN red list. Even though *Emys orbicularis* is now a

protected species, translocations and illegal trading were common in the last century, and it is common to find translocated specimens both in natural habitats and recovery centres.

In a previous work of Velo-Antón *et al.* (2007) a first attempt to develop a genetic tool to allocate individuals from recovery centres to their probable population of origin was sucessfully acomplished. Nevertheless, the dataset used was small and covered only a limited number of populations. In this work, we focus on the *Emys orbicularis occidentalis* subspecies (corresponding to the lineage VI as described in Lenk et al. 1999; Fritz et al. 2007), which occurs in the Iberian Peninsula and Northern Morocco, where we aim to comprehensively expand the baseline information to use for the allocation of individuals from recovery centres, while testing the effectiveness of the previous method with a larger dataset, in an attempt to provide decision makers a tool to identify the region of origin of a given specimen.

4.3 Material and Methods

4.3.1 Sampling and lab procedures

The microsatellite dataset obtained in Chapter 3 was here used as baseline for the genetic allocation of individuals to their population of origin. Furthermore, 30 samples from recovery centres or with unknown origin were also extracted and genotyped following the Chapter 3 methodology (see subchapter 3.3.1 for extraction methods and 3.3.2 for the microsatellite amplification). Also, all individuals with unknown origin were sequenced for the mitochondrial cytochrome *b* region, following the procedure described in Chapter 2 (see subchapter 2.3.2 for details), in order to identify their mitochondrial lineage. This step is necessary as individuals from non-native lineages were found in the Iberian Peninsula (Velo-Antón *et al.* 2011b).

4.3.2 Assignment tests

STRUCTURE (Pritchard *et al.* 2000) was run to explain the observable genetic structure in the data, with a burn-in of 1000000 steps with 4000000 iterations after the initial burn-in, for each k four independent replicates were run. Our parameter choice accounted for correlated allelic frequencies and genetic admixture between genetic

groups. We chose k=16 (Figure 4.1), the second best solution to explain the observable genetic structure, based on the previous results (see chapter 3.3.3). We discarded the best solution, k=2, as it wouldn't be informative enough to allow a correct genetic allocation of individuals. The individual probabilities of assignment were then used to estimate the most probable genetic cluster of each individual.

In order to verify the capability of this method to identify the probable putative population of origin, a first STRUCTURE run was performed with only individuals with a known population of origin. The individual's probability of assignment (q) to each cluster was verified and those with less than 0.5 probability of assignment to any cluster were deemed as misidentifications. Furthermore, individuals captured from a population but assigned to a different genetic cluster than that of its population of origin were also classified as misidentifications. Afterwards, a second STRUCTURE run was performed adding the individual with unknown origin to the dataset in an attempt to identify their population of origin.

4.4 Results

4.4.1 Proportion of assignment of each putative population

Of the 21 populations here analyzed, 15 populations have been assigned to only one cluster with a proportion of assignment higher than 0.5 (Table 4.1).

The remaining 6 populations, with the exception of Orense, had a very small sample size for a more viable assignment to a single cluster and had their proportion of assignments split between 2 to 3 genetic clusters (considering a minimum of proportion of assignment of 0.1). The Orense population has been assigned to a single cluster (with $q > 0.1$); however, the proportion of assignment to this cluster was below our threshold of 0.5.

The populations of Cuenca, Ciudad Real, Castelo Branco, Cardeña and Robledillo, located in the central area of the Iberian Peninsula, show a high degree of admixture. In fact, all these populations seem to have a proportion of their assignment

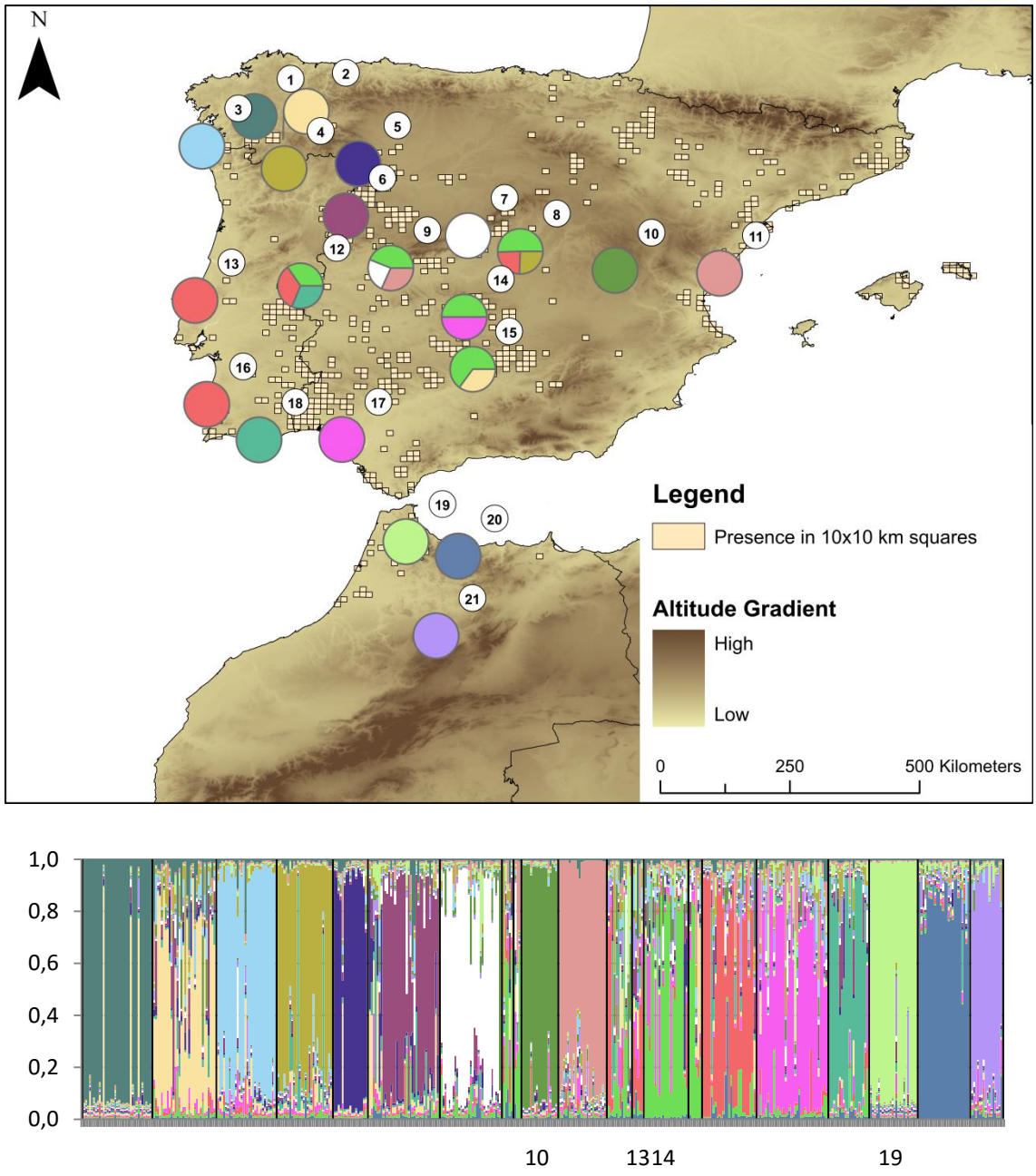

Figure 4.1: **Top:** Spatial representation of the genetic structure found for the entire distribution of *Emys orbicularis occidentalis* for k=16. Each pie chart corresponds to a sampling location, identified by a number that corresponds to the population in the structure graphic below. **Bottom:** Structure output for k=16, each horizontal line corresponds to a individual, where the proportion of the colours indicates its probability of assignment to the corresponding cluster. Black lines separate sampling locations.

to a genetic cluster only present in the area, but always admixed with neighbouring genetic clusters.

4.4.2 Accuracy of the test

When using a threshold of 0.5 for the probability of assignment to a certain cluster, we correctly identified the assigned genetic cluster of 369 out of 453 samples, revealing a 76% accuracy of the test to identify the most probable population of origin. The probabilities of assignment to the correct cluster ranged from 0.512 to 0.942 (Table 4.1).

4.4.3 Estimations of the population of origin of animals found in Recovery Centres

From the 30 individuals with unknown origin, we were able to successfully attribute probabilities of assignment to one genetic cluster (ranging from 0.506 to 0.832) to 16 samples and we failed to detect a probability of assignment of at least 0.5 for the remaining 14 individuals. The majority of samples with unknown origin (n=5) are assigned to the genetic cluster common to Cuenca, Ciudad Real, Castelo Branco, Cardeña and Robledillo. One sample was assigned to the Fifi population in Morocco; two individuals were assigned to the Orense region, one to Algarve, two to Zamora, one to Salamanca, two to Boticas and one to Valencia (Table 4.2).

4.5 Discussion

The translocation of individuals to reinforce populations is a recurring strategy to safeguard the population size in the wild, and thus the maintenance of its genetic viability (Griffith et al. 1989; see Fritz & Chiari 2013 for several example on *Emys orbicularis*). However, only in few cases, the effort to measure the genetic distance between the source population and the population of interest for conservation is made (Witzenberger & Hochkirch 2008). The usage of individuals to maintain the genetic viability of a certain population, also known as genetic rescue (Tallmon et al. 2004), was popularized by the Florida Panther case (Johnson et al. 2010) where Texan panthers, the closest genetic group to the Florida panther, were introduced in the region in an attempt to decrease the already strong effect of inbreeding.

Table 4.1: Proportion of membership of each putative population in each of the 16 clusters. Populations that failed to reach at least 0.5 of proportion of membership are marked with a *.

Population	N	GC1	GC2	GC3	GC4	GC5	GC6	GC7	GC8	GC9	GC10	GC11	GC12	GC13	GC14	GC15	GC16
Porriño	30	0.01	0.01	0.02	0.03	0.04	0.01	0.01	0.01	0.02	0.02	0.01	0.76	0.02	0.01	0.01	0.02
Orense *	32	0.01	0.02	0.02	0.02	0.49	0.03	0.04	0.08	0.02	0.01	0.04	0.04	0.06	0.01	0.03	0.08
Ribadavia	35	0.01	0.01	0.01	0.01	0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.85
Boticas	28	0.01	0.01	0.01	0.02	0.03	0.09	0.01	0.02	0.01	0.01	0.01	0.03	0.72	0.01	0.01	0.01
Zamora	17	0.01	0.01	0.01	0.01	0.03	0.01	0.82	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.02	0.02
Salamanca	36	0.02	0.03	0.02	0.02	0.07	0.05	0.09	0.54	0.04	0.02	0.01	0.02	0.02	0.02	0.01	0.03
Madrid	31	0.02	0.03	0.01	0.02	0.01	0.02	0.01	0.02	0.71	0.01	0.07	0.02	0.01	0.03	0.02	0.01
Argana del Rey *	5	0.05	0.37	0.10	0.04	0.05	0.02	0.03	0.08	0.02	0.02	0.01	0.01	0.12	0.02	0.03	0.02
Ciudad Real	22	0.01	0.56	0.03	0.10	0.01	0.01	0.05	0.03	0.03	0.03	0.05	0.03	0.01	0.02	0.02	0.02
Cuenca	18	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.87	0.01	0.01	0.01	0.01	0.01
Valencia	24	0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.82	0.01
Doñana	36	0.02	0.03	0.02	0.60	0.02	0.01	0.01	0.02	0.02	0.05	0.02	0.03	0.04	0.05	0.06	0.02
Almograve	27	0.04	0.04	0.62	0.01	0.02	0.05	0.01	0.03	0.01	0.03	0.02	0.02	0.02	0.01	0.04	0.03
Castelo Branco *	12	0.02	0.16	0.16	0.06	0.05	0.13	0.01	0.04	0.06	0.03	0.10	0.03	0.06	0.06	0.02	0.02
Paul da Tornada*	5	0.04	0.02	0.30	0.09	0.07	0.03	0.01	0.07	0.01	0.07	0.02	0.05	0.05	0.05	0.05	0.07
Algarve	20	0.02	0.03	0.02	0.04	0.02	0.65	0.03	0.06	0.02	0.02	0.01	0.02	0.03	0.02	0.02	0.01
Fifi	26	0.82	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01
Sidi Mimoun	16	0.03	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.79	0.01	0.01	0.02	0.03	0.01	0.01
Moulay Abdesalam	24	0.02	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.01	0.03	0.01	0.01	0.01	0.88	0.00	0.00
Cardeña *	6	0.01	0.47	0.03	0.04	0.17	0.01	0.01	0.03	0.03	0.01	0.01	0.03	0.03	0.01	0.08	0.01
Robledillo *	3	0.02	0.35	0.02	0.07	0.04	0.05	0.02	0.01	0.15	0.01	0.01	0.01	0.01	0.00	0.23	0.01

Table 4.2: Summary table for the genetic assignment results of individuals with unknown origin to a putative population. Note that to facilitate reading, Genetic Clusters have inherited the name of the putative population with a higher proportion of membership.

Samples	Fifi	Ciudad Real	Almograve	Doñana	Orense	Algarve	Zamora	Salamanca	Madrid	Sidi Mimoun	Cuenca	Porriño	Boticas	Moulay Abdesalam	Valencia	Ribadavia
UO1	0.06	0.18	0.06	0.39	0.05	0.07	0.01	0.01	0.01	0.02	0.01	0.07	0.01	0.01	0.04	0.01
UO2	0.01	0.02	0.01	0.01	0.03	0.03	0.01	0.01	0.01	0.01	0.02	0.02	0.70	0.01	0.12	0.01
UO3	0.07	0.12	0.01	0.03	0.01	0.01	0.04	0.03	0.20	0.08	0.36	0.01	0.01	0.01	0.01	0.01
UO4	0.02	0.08	0.01	0.02	0.01	0.02	0.39	0.25	0.03	0.06	0.06	0.01	0.02	0.01	0.01	0.01
UO5	0.03	0.55	0.05	0.04	0.01	0.01	0.02	0.04	0.03	0.01	0.01	0.02	0.02	0.01	0.06	0.07
UO6	0.32	0.31	0.07	0.04	0.02	0.06	0.01	0.04	0.03	0.01	0.03	0.02	0.02	0.00	0.02	0.00
UO7	0.32	0.31	0.07	0.04	0.02	0.07	0.01	0.04	0.03	0.01	0.03	0.02	0.02	0.00	0.02	0.00
UO8	0.02	0.02	0.01	0.03	0.01	0.03	0.01	0.36	0.11	0.10	0.05	0.03	0.14	0.06	0.01	0.01
UO9	0.02	0.02	0.10	0.01	0.13	0.06	0.08	0.34	0.01	0.09	0.01	0.01	0.01	0.01	0.06	0.04
UO10	0.10	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.23	0.01	0.02	0.01	0.52	0.01
UO11	0.01	0.02	0.01	0.01	0.07	0.53	0.04	0.01	0.03	0.01	0.09	0.01	0.02	0.01	0.12	0.02
UO12	0.37	0.02	0.01	0.01	0.01	0.03	0.01	0.03	0.05	0.02	0.01	0.00	0.01	0.10	0.32	0.01
UO13	0.01	0.01	0.01	0.01	0.03	0.01	0.48	0.27	0.00	0.02	0.07	0.01	0.03	0.01	0.01	0.05
UO14	0.01	0.72	0.06	0.01	0.01	0.01	0.01	0.01	0.08	0.01	0.01	0.02	0.02	0.01	0.03	0.01
UO15	0.01	0.80	0.02	0.01	0.01	0.03	0.00	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.04	0.00
UO16	0.01	0.01	0.01	0.01	0.03	0.01	0.64	0.19	0.01	0.02	0.01	0.01	0.03	0.02	0.00	0.01
UO17	0.01	0.02	0.04	0.01	0.61	0.01	0.05	0.07	0.01	0.01	0.01	0.03	0.03	0.03	0.01	0.06
UO18	0.01	0.01	0.01	0.01	0.01	0.04	0.12	0.53	0.04	0.01	0.13	0.06	0.01	0.01	0.01	0.02
UO19	0.00	0.04	0.01	0.01	0.25	0.04	0.54	0.01	0.01	0.01	0.01	0.04	0.01	0.01	0.01	0.02
UO20	0.01	0.13	0.28	0.03	0.04	0.04	0.11	0.09	0.06	0.02	0.01	0.02	0.03	0.01	0.11	0.01
UO21	0.01	0.05	0.01	0.09	0.01	0.02	0.01	0.03	0.06	0.46	0.01	0.11	0.02	0.01	0.12	0.01
UO22	0.03	0.01	0.01	0.03	0.52	0.02	0.01	0.11	0.01	0.01	0.01	0.01	0.13	0.01	0.07	0.02
UO23	0.01	0.01	0.04	0.01	0.34	0.03	0.01	0.07	0.00	0.02	0.01	0.05	0.38	0.01	0.01	0.01
UO24	0.01	0.83	0.01	0.01	0.01	0.05	0.00	0.01	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01

UO25	0.02	0.01	0.06	0.02	0.05	0.08	0.00	0.02	0.01	0.03	0.02	0.01	0.51	0.01	0.13	0.01
UO26	0.01	0.08	0.32	0.16	0.01	0.02	0.01	0.05	0.02	0.14	0.09	0.05	0.01	0.02	0.01	0.01
UO27	0.01	0.10	0.17	0.05	0.02	0.34	0.05	0.03	0.01	0.01	0.03	0.04	0.12	0.01	0.01	0.00
UO28	0.01	0.71	0.01	0.04	0.02	0.01	0.01	0.02	0.03	0.02	0.02	0.03	0.01	0.02	0.02	0.03
UO29	0.78	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.04	0.01	0.01	0.02	0.01	0.08	0.00
UO30	0.03	0.05	0.02	0.18	0.02	0.01	0.04	0.03	0.06	0.18	0.24	0.05	0.02	0.03	0.02	0.02

In *Emys orbicularis* several breeding centres have been built and reintroductions have been made in an attempt to increase the population size of several critical populations (see Fritz & Chiari 2013 for a review). Nonetheless, the lack of attention to the genetic origin of the introduced individuals has lead to several problematic situations. For example, one Germanic population of *Emys orbicularis* is now threatened due to the introductions of allochthonous individuals from different *E. orbicularis* genetic lineages (Fritz & Chiari 2013).

The results obtained here show a reasonable success rate for the individual assignment to a putative population of origin. The lower probabilities of assignment when compared to those of Velo-Antón *et al.* (2007) are expectable due to the comprehensive increase in both sample size and number of sampled locations.

Overall, the use of hypervariable markers, such as microsatellites (reviewed by Manel *et al.* 2005), is a good strategy to tackle the problem of the assignment of individuals to a probable population of origin.

Chapter 5: Final Remarks

The dynamic and complex paleogeographic history of the western Mediterranean basin alongside the past climatic fluctuations had an important impact in the biogeographical history 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.* 2012a; Husemann *et al.* 2014). Based on ecological and biological constrains, different species coped in distinct ways to these dynamics, allowing different evolutionary histories to shape diverse biogeographic and genetic patterns. With this study, we were able reinforce the previous knowledge on the biogeographical history of the Ibero-Maghrebian lineage of the Mediterranean Pond Turtle (*Emys orbicularis occidentalis*).

The phylogeographic patterns observed in this subspecies denotes a complex biogeographical history. The presence of few fossil records dating from the Holocene in North Africa (de Lapparent de Broin 2000) and the presence of older fossil records in the Iberian Peninsula (Blasco *et al.* 2011) seem to indicate a longer presence of *E. o. occidentalis* in the Iberian Peninsula than in North Africa. These results are contrasting with the genetic patterns observed. The higher genetic diversity and structure in North Africa (see Chapter 2 and 3), allied to its basal role on the phylogeny of the subspecies are indicative of a colonization of the Iberian Peninsula from North Africa. In light of these results, one can say that either the fossils present in Iberian Peninsula might be reminiscent from a previous lineage that inhabited the Iberian Peninsula and over went a massive extinction, or they represent the ancestor of the current Iberian lineage.

While the current available genetic data is enough to pinpoint the direction of the colonization, it is still insufficient to determine the ancestor of the African populations that later colonized the Iberian Peninsula. Two possible explanations are discussed by Stuckas *et al.* (2014). The first pinpoints the origin of North African populations in a possible lineage that previously inhabited the Iberian Peninsula that underwent massive extinction. The second attributes the origin of the African populations to a colonization event from the Apennine Peninsula and Sicily. While the latter is easily refuted by the fact that the current pond turtles inhabiting this region are from a genetically distinct species, *Emys trinacris*, the first hypothesis could be further explored by the use of ancient DNA to evaluate the relationship of older fossils present in the Iberian Peninsula to the current lineage presence in the area.

In this work it was also possible to explore the impacts that such expansion from North Africa had on the genetic patterns observed in the lineage. The current

knowledge on natural range expansions is still limited and is given by only a handful of examples (e.g. Damselflies, Swaegers *et al.* 2013). Our work on the dynamics of the range expansion of this lineage represents, to our knowledge, the first comprehensive study that covers the entire Iberian Peninsula. Even though previous studies on *Testudo graeca* (Graciá *et al.* 2013a) tackled this issue, the distribution of this species in the Iberian Peninsula is very limited.

The explicit consideration of range expansions when interpreting current patterns of genetic diversity is of major importance, as the dynamics of such events often create patterns commonly attributed to other phenomena such as the effect of adaptation (Excoffier & Ray 2008). Furthermore, with the ongoing climatic changes several species are undergoing range expansions and contractions (Hickling *et al.* 2005, 2006), so it is highly relevant to understand the dynamics of such events as it will allow us to predict the ability of newly founded populations to locally adapt to new environments, as well as the capacity of such events to promote advantageous mutations that surf with the wave of expansion promoting, for example, an increase in dispersal capacity as found in the Cane toad (Phillips *et al.* 2006).

Additionally, with the considerable increase of genetic data resultant from this work, with several new sampled populations, covering now the full extent of the subspecies distribution, we can provide more accurate data for decision makers to base conservation actions, namely the genetic allocation tool presented in Chapter 4. It is worth mention that it has already been used to provide the probable population of origin of a few animals from the Gaia Biological Park and that one of such individuals was found in the facilities of CIBIO and had its probable population of origin in Boticas where it was released.

After this thesis, several questions are still worth of attention. For example, it would be worthwhile to evaluate how the species range shifted during the Glacial-Interglacial periods. An Ecological Niche Model based analysis would be rather interesting to determine suitable habitats for this species during both Glacial and Interglacial periods, in order to shed new light into several of the current observable patterns (for example the possible coastal corridor in the western coast of the Iberian Peninsula) and could be used to predict further range shifts promoted by the current climate changes.

Chapter 6: References

Allem A (2000) The terms genetic resource, biological resource, and biodiversity examined. *The Environmentalist*, **20**, 335–341.

Álvarez Y, Mateo J, Andreu A *et al.* (2000) Brief communication. Mitochondrial DNA haplotyping of *Testudo graeca* on both continental sides of the Straits of Gibraltar. *Journal of Heredity*, **91**, 39–41.

Andueza IL, Alcayde VS (2004) Advances in the action plan for *Emys orbicularis* in the Valencia region, Spain. *Biologia*, **54**, 173–176.

Armstrong DP, Seddon PJ (2008) Directions in reintroduction biology. *Trends in Ecology & Evolution*, **23**, 20–25.

Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H (1997) Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, **51**, 148–164.

Avise JC, Bowen BW, Lamb T, Meylan a B, 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.

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

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations.

Berthouly-Salazar C, Hui C, Blackburn TM *et al.* (2013) Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion. *Molecular Ecology*, **22**, 5793–5804.

Beukema W, De Pous P, Donaire D *et al.* (2010) Biogeography and contemporary climatic differentiation among Moroccan *Salamandra algira*. *Biological Journal of the Linnean Society*, **101**, 626–641.

Biek R, Henderson JC, Waller L a, Rupprecht CE, Real L a (2007) A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 7993–7998.

Bilton D, Freeland J, Okamura B (2001) Dispersal in freshwater invertebrates. *Annual Review of Ecology and Systematics*, **32**, 159–181.

Blasco R, Blain H-A, Rosell J *et al.* (2011) Earliest evidence for human consumption of tortoises in the European Early Pleistocene from Sima del Elefante, Sierra de Atapuerca, Spain. *Journal of human evolution*, **61**, 503–509.

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**, 14237–14252.

Bridle JR, Vines TH (2007) Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution*, **22**, 140–147.

Brito J, Fahd S, Martínez-Freiría F *et al.* (2011) Climate change and peripheral populations: predictions for a relict Mediterranean viper. *Acta Herpetologica*, **6**, 105–118.

Brito JC, Santos X, Pleguezuelos JM, Sillero N (2008) Inferring evolutionary scenarios with geostatistics and geographical information systems for the viperid snakes *Vipera latastei* and *Vipera monticola*. *Biological Journal of the Linnean Society*, **95**, 790–806.

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.

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

Caccone A, Gibbs JP, Ketmaier V, Suatoni E, Powell JR (1999) Origin and evolutionary relationships of giant Galápagos tortoises. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 13223–13228.

Cadi A, Joly P (2003) Competition for basking places between the endangered European pond turtle (*Emys orbicularis galloitalica*) and the introduced red-eared slider (*Trachemys scripta*). *Canadian Journal of Zoology*, **81**, 1392–1398.

Cadi A, Joly P (2004) Impact of the introduction of the red-eared slider (*Trachemys scripta elegans*) on survival rates of the European pond turtle (*Emys orbicularis*). *Biodiversity and Conservation*, **13**, 2511–2518.

Campbell Grant EH, Nichols JD, Lowe WH, Fagan WF (2010) Use of multiple dispersal pathways facilitates amphibian persistence in stream networks. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 6936–6940.

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, Geniez P, Roca J, Mateo J a (2008) Radiation, multiple dispersal and parallelism in the skinks, *Chalcides* and *Sphenops* (Squamata: Scincidae), with comments on *Scincus* and *Scincopus* and the age of the Sahara Desert. *Molecular Phylogenetics and Evolution*, **46**, 1071–1094.

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.

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

Collen B, Whitton F, Dyer EE *et al.* (2014) Global patterns of freshwater species diversity, threat and endemism. *Global Ecology and Biogeography*, 23, 40–51.

Cordero A, Ayres C (2004) A management plan for the European pond turtle (Emys orbicularis) populations of the Louro river basin (Northwest Spain). *Biologia*, **59**, 161–171.

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.

Costanza R, d'Arge R, de Groot R *et al.* (1997) The value of the world's ecosystem services and natural capital. *Nature*, **387**, 253–260.

Currat M, Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 679–688.

Cuttelod A, García N, Malak DA, Temple H, Katariya V (2008) The Mediterranean: A biodiversity Hotspot under Threat. In: *The 2008 Review of The IUCN Red List of Threatened Species* (eds Vié J-C, Hilton-Taylo C, Stuart SN). IUCN Gland, Switzerland.

Díaz S, Fargione J, Chapin FS, Tilman D (2006) Biodiversity loss threatens human well-being. *PLoS biology*, **4**, e277.

Van Dijk PP, 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*

Van Dijk PP, 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.

Dudgeon D, Arthington AH, Gessner MO *et al.* (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews of the Cambridge Philosophical Society*, **81**, 163–182.

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

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

Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 975–979.

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, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.

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.

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, **23**, 347–351.

Fahd S, El Marnisi B, Mediani M, Fritz U (2009) Zur Verbreitung und zum Bedrohungsstatus der Europäischen Sumpfschildkröte (*Emys orbicularis*) in Marokko. *Elaphe*, **17**, 30–33.

Ficetola G, Padoa-Schioppa E, Monti A *et al.* (2004) The importance of aquatic and terrestrial habitat for the European pond turtle (*Emys orbicularis*): implications for conservation planning and management. *Canadian Journal of Zoology*, **82**, 1704–1712.

Frankham R (2003) Genetics and conservation biology. *Comptes Rendus Biologies*, **326**, 22–29.

Fritz U (2001) *Emys orbicularis* (Linnaeus, 1758) – Europäische Sumpfschildkröte. In: *Handbuch der Reptilien und Amphibien Europas. Band 3/IIIA: Schildkröten I* (ed Fritz U), pp. 343–515. Aula- Verlag, Wiebelsheim.

Fritz U, Ayaz D, Hundsdörfer AK *et al.* (2009) Mitochondrial diversity of European pond turtles (*Emys orbicularis*) in Anatolia and the Ponto-Caspian Region: Multiple old refuges, hotspot of extant diversification and critically endangered endemics. *Organisms Diversity & Evolution*, **9**, 100–114.

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, Chiari Y (2013) Conservation actions for European pond turtles–a summary of current efforts in distinct European countries. *Herpetology Notes*, **6**, 105.

Fritz U, Fattizzo T, Guicking D *et al.* (2005) A new cryptic species of pond turtle from southern Italy, the hottest spot in the range of the genus *Emys* (Reptilia, Testudines, Emydidae). *Zoologica Scripta*, **34**, 351–371.

Fritz U, Guicking D, Kami H *et al.* (2007) Mitochondrial phylogeography of European pond turtles (*Emys orbicularis*, *Emys trinacris*) an update. *Amphibia-Reptilia*, **28**, 418–426.

Fritz U, Havaš P (2007) Chelonians of the World 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.

García-París M, Arano B, Herrero P (2001) Molecular characterization of the contact zone between *Triturus pygmaeus* and *T. marmoratus* (Caudata: Salamandridae) in Central Spain and their taxonomic assessment. *Revista Española de Herpetología*, **15**, 115–126.

Gerlach J, Muir C, Richmond MD (2006) The first substantiated case of trans-oceanic tortoise dispersal. *Journal of Natural History*, **40**, 2403–2408.

Gleick PH (Ed.) (1993) *Water in crisis : a guide to the world's fresh water resources*. Oxford University Press, New York.

Gleick P (1998) The human right to water. *Water Policy*, **1**, 487–503.

Godinho R, Domingues V, Crespo EG, Ferrand N (2006) Extensive intraspecific polymorphism detected by SSCP at the nuclear C-mos gene in the endemic Iberian lizard *Lacerta schreiberi*. *Molecular Ecology*, **15**, 731–738.

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, Botella F, Anadón J *et al.* (2013a) Surfing in tortoises? Empirical signs of genetic structuring owing to range expansion. *Biology Letters*, **9**, 1–5.

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

Griffith B, Scott J, Carpenter J, Reed C (1989) Translocation as a species conservation tool: status and strategy. *Science*, **245**, 477–480.

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

Hallatschek O, Hersen P, Ramanathan S, Nelson DR (2007) Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 19926–19930.

Hallatschek O, Nelson DR (2008) Gene surfing in expanding populations. *Theoretical Population Biology*, **73**, 158–70.

Harris DJ, Batista V, Carretero MA (2004) Assessment of genetic diversity within *Acanthodactylus erythrurus* (Reptilia : Lacertidae) in Morocco and the Iberian Peninsula using mitochondrial DNA sequence data. *Amphibia-Reptilia*, **25**, 227–232.

Hedges SB, Kumar S (2009) *The timetree of life*. Oxford University Press.

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

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

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

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

Hickling R, Roy DB, Hill JK, Fox R, Thomas CD (2006) The distributions of a wide range of taxonomic groups are expanding polewards. *Global Change Biology*, **12**, 450–455.

Hickling R, Roy DB, Hill JK, Thomas CD (2005) A northward shift of range margins in British Odonata. *Global Change Biology*, **11**, 502–506.

Hidalgo-Vila J, Florencio M, Díaz-Paniagua C, Pérez-Santigosa N (2011) Does the exotic invader turtle, *Trachemys scripta elegans*, compete for food with coexisting native turtles? *Amphibia-Reptilia*, **32**, 167–175.

Hijmans RJ (2015) raster: Geographic data analysis and modeling. R package version 2.5-2.

Hsü K, Montadert L, Bernoulli D *et al.* (1977) History of the Mediterranean salinity crisis. *Nature*, **267**, 399–403.

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

Jiménez-Moreno G, Fauquette S, Suc J-P (2010) Miocene to Pliocene vegetation reconstruction and climate estimates in the Iberian Peninsula from pollen data. *Review of Palaeobotany and Palynology*, **162**, 403–415.

Johnson M (1996) Water, animals and agricultural technology: a study of settlement patterns and economic change in Neolithic Southern Greece. *Oxford Journal of Archaeology*, **15**, 267–295.

Johnson WE, Onorato DP, Roelke ME *et al.* (2010) Genetic restoration of the Florida panther. *Science*, **329**, 1641–1645.

Jombart T (2008) A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Devillard S, Dufour A-B, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

De Jong H (1998) In search of historical biogeographic patterns in the western Mediterranean terrestrial fauna. *Biological Journal of the Linnean Society*, **65**, 99–164.

Joyce WG, Gauthier J a (2004) Palaeoecology of triassic stem turtles sheds new light on turtle origins. *Proceedings of the Royal Society of London B: Biological Sciences*, **271**, 1–5.

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 M a. (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, Andreu AC (2002) Emys orbicularis (Linnaeus, 1758). Galápago europeo. In: *Atlas y Libro Rojo de los anfibios y reptiles de España* (eds Pleguezuelos J, Márquez R, Lizana M), pp. 181–186. Ministerio de Medio Ambiente, Madrid.

King TL, Julian SE (2004) Conservation of microsatellite DNA flanking sequence across 13 Emydid genera assayed with novel bog turtle (*Glyptemys muhlenbergii*) loci. *Conservation Genetics*, **5**, 719–725.

Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, **23**, 482–490.

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

Labrousse A, Romero L (2001) *Rapport sur la situation du cannabis dans le Rif marocain.*

De Lapparent de Broin F (2000) African chelonians from the Jurassic to the present: phases of development and preliminary catalogue of the fossil record. *Palaeontologia Africana*, **36**, 82.

Leakey RE, Lewin R (1996) *The Sixth Extinction: Biodiversity and Its Survival*. Phoenix, London.

Lehe R, Hallatschek O, Peliti L (2012) The rate of beneficial mutations surfing on the wave of a range expansion. *PLoS Computational Biology*, **8**, e1002447.

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.

Loureiro A, Ferrand de Almeida N, Carretero MA, Paulo OS (Eds.) (2008) *Atlas dos Anfíbios e Répteis de Portugal*. Instituto da Conservação da Natureza e da Biodiversidade, Lisboa.

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.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, **20**, 136–142.

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

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**, 3375–3388.

Matson CW, Palatnikov G, Islamzadeh A *et al.* (2005) Chromosomal damage in two species of aquatic turtles (*Emys orbicularis* and *Mauremys caspica*) inhabiting contaminated sites in Azerbaijan. *Ecotoxicology*, **14**, 513–525.

Meylan PA (2002) Testudines (Tortoises , Turtles and Terrapins). *Encyclopedia of Life Sciences*.

Michaux JR, Libois R, Filippucci M-G (2005) So close and so different: comparative phylogeography of two small mammal species, the yellow-necked fieldmouse (*Apodemus flavicollis*) and the woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Heredity*, **94**, 52–63.

Millennium Ecosystem Assessment (2005) *Ecosystems and Human Well-Being - Health Synthesis*. Island Press, Washington, DC.

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, USA.

Murphy LN, Kirk-Davidoff DB, Mahowald N, Otto-Bliesner BL (2009) A numerical study of the climate response to lowered Mediterranean Sea level during the Messinian Salinity Crisis. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **279**, 41–59.

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

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

Nunes PALD, van den Bergh JCJ. (2001) Economic valuation of biodiversity: sense or nonsense? *Ecological Economics*, **39**, 203–222.

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.

Oki T, Kanae S (2006) Global hydrological cycles and world water resources. *Science (New York, N.Y.)*, **313**, 1068–1072.

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

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Paulo O, Pinto I, Bruford MW, Jordan WC, Nichols RA (2002) The double origin of Iberian peninsular chameleons. *Biological Journal of the Linnean Society*, **75**, 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.

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

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.

Petit RJ (2011) Early insights into the genetic consequences of range expansions. *Heredity*, **106**, 203.

Phillips BL, Brown GP, Webb JK, Shine R (2006) Invasion and the evolution of speed in toads. *Nature*, **439**, 803.

Pleguezuelos JM, Brito JC, Fahd S *et al.* (2010) Setting conservation priorities for the Moroccan herpetofauna: the utility of regional red lists. *Oryx*, **44**, 501–508.

Polo-Cavia N, López P, Martín J (2010) Competitive interactions during basking between native and invasive freshwater turtle species. *Biological Invasions*, **12**, 2141–2152.

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

R Core Team (2012) R: A Language and Environment for Statistical Computing.

Rands MRW, Adams WM, Bennun L *et al.* (2010) Biodiversity conservation: challenges beyond 2010. *Science*, **329**, 1298–1303.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity*, **86**, 248–249.

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.

Ricciardi A, Rasmussen JB (1999) Extinction Rates of North American Freshwater Fauna. *Conservation Biology*, **13**, 1220–1222.

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**, 103–106.

Servan J, Arvy C (1997) The introduction of Trachemys scripta in France. A new competitor for the European pond turtles. *Bulletin Francais de la Peche et de la Pisciculture*, **344/345**, 173–177.

Shahin M (2002) *Hydrology and Water Resources of Africa, Volume 1*. Springer, Netherlands.

Sommer RS, Lindqvist C, Persson A *et al.* (2009) Unexpected early extinction of the European pond turtle (Emys orbicularis) in Sweden and climatic impact on its Holocene range. *Molecular Ecology*, **18**, 1252–1262.

Sommer RS, Persson A, Wieseke N, Fritz U (2007) Holocene recolonization and extinction of the pond turtle, *Emys orbicularis* (L., 1758), in Europe. *Quaternary Science Reviews*, **26**, 3099–3107.

Sousa P, Harris DJ, Froufe E, van der Meijden A (2012) Phylogeographic patterns of *Buthus* scorpions (Scorpiones: Buthidae) in the Maghreb and South-Western Europe based on CO1 mtDNA sequences (M-O Rödel, Ed,). *Journal of Zoology*, **288**, 66–75.

Spinks PQ, Shaffer HB (2009) Conflicting mitochondrial and nuclear phylogenies for the widely disjunct Emys (Testudines: Emydidae) species complex, and what they tell us about biogeography and hybridization. *Systematic Biology*, **58**, 1–20.

Storfer A (1999) Gene flow and endangered species translocations: a topic revisited. *Biological Conservation*, **87**, 173–180.

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*, **14**, 295–306.

Swaegers J, Mergeay J, Therry L *et al.* (2013) Rapid range expansion increases genetic differentiation while causing limited reduction in genetic diversity in a damselfly. *Heredity*, **111**, 422–429.

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

Tallmon D a, Luikart G, Waples RS (2004) The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution*, **19**, 489–496.

Tarroso P, Carrión J, Dorado-Valiño M *et al.* (2014) Spatial climate dynamics in the Iberian Peninsula since 15 000 Yr BP. *Climate of the Past Discussions*, **10**, 3901–3930.

Trakimas G, Sidaravi J (2008) Road mortality threatens small northern populations of the European pond turtle , *Emys orbicularis*. *Acta Herpetologica*, **3**, 161–166.

Travis JMJ, Münkemüller T, Burton OJ *et al.* (2007) Deleterious mutations can surf to high densities on the wave front of an expanding population. *Molecular Biology and evolution*, **24**, 2334–2343.

Vamberger M, Stuckas H, Ayaz D *et al.* (2014) Massive transoceanic gene flow in a freshwater turtle (Testudines: Geoemydidae: *Mauremys rivulata*). *Zoologica Scripta*, **43**, 313–322.

Veith M, Mayer C, Samraoui B, Barroso D, 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, 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, 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, Becker CG, Cordero-Rivera A (2011a) Turtle carapace anomalies: the roles of genetic diversity and environment. *PloS one*, **6**, e18714.

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.

Velo-Antón G, Godinho R, Harris DJ *et al.* (2012a) 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**, 965–973.

Velo-Antón G, Rodríguez D, Savage AE *et al.* (2012b) Amphibian-killing fungus loses genetic diversity as it spreads across the New World. *Biological Conservation*, **146**, 213–218.

Velo-Antón G, Godinho R, Campos JC, Brito JC (2014) Should I stay or should I go? Dispersal and population structure in small, isolated desert populations of West African crocodiles. *PloS one*, **9**, e94626.

Veríssimo J (2014) Phylogeography and current patterns of genetic diversity and structure of the Mediterranean pond turtle. University of Porto, Master thesis.

Weiss S, Ferrand N (Eds.) (2007) *Phylogeography of Southern European Refugia*. Springer, Dordrecht, The Netherlands.

White TA, Perkins SE, Heckel G, Searle JB (2013) Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology*, **22**, 2971–2985.

Widmer A, Lexer C (2001) Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology & Evolution*, **16**, 267–269.

Witzenberger K a., Hochkirch A (2008) Genetic consequences of animal translocations: A case study using the field cricket, *Gryllus campestris* L. *Biological Conservation*, **141**, 3059–3068.

Wood PJ, Greenwood MT, Agnew MD (2003) Pond biodiversity and habitat loss in the UK. *Area*, **35**, 206–216.

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

Chapter 7: Supplementary Material

Table 7.1: Characteristics of the microsatellite loci used in *Emys orbicularis*. Primer (µL) = quantity of primer (µL) of a 10 µM primer solution; NA = Number of alleles sampled; primers marked with * are common for all three manuscripts

Multiplex	Locus	Primer (µL)	Allelic Range	NA	Original Reference
1	GmuD114*	0.4	104-140	14	King & Julian 2004
1	GmuD93*	0.45	240-280	16	King & Julian 2004
1	GmuA32	0.4	166-168	2	King & Julian 2004
1	GmuD51*	1.5	235-299	19	King & Julian 2004
1	GmuD87*	0.5	226-254	18	King & Julian 2004
1	GmuD88*	1	136-180	15	King & Julian 2004
2	GmuD121	0.4	144-148	2	King & Julian 2004
2	GmuD16*	0.6	165-233	21	King & Julian 2004
2	GmuB08*	0.3	194-203	5	King & Julian 2004
2	GmuD28	0.35	206-226	6	King & Julian 2004
2	GmuD40	1.2	173-233	15	King & Julian 2004

Table 7.2: Summary table of all sequenced samples in the study area. Locality: Sampling location; Lat: Latitude; Long: Longitude. ND: information not available.

Sample	Country	Locality	Lat	Long	Lineage	Haplotype	Reference
GVA4155	Morocco	Darrati	35,14	-5,62	VI	VIf	This study
GVA1252	Morocco	Fifi	35,02	-5,21	VI	VIf	This study
GVA1255	Morocco	Fifi	35,02	-5,21	VI	VIf	This study
GVA1262	Morocco	Fifi	35,02	-5,21	VI	VIf	This study
GVA1269	Morocco	Fifi	35,02	-5,21	VI	VIf	This study
GVA1271	Morocco	Fifi	35,02	-5,21	VI	VIf	This study
GVA1274	Morocco	Fifi	35,02	-5,21	VI	VIf	This study
GVA1278	Morocco	Fifi	35,02	-5,21	VI	VIf	This study
GVA2579	Morocco	Fifi	35,02	-5,21	VI	VIf	This study
GVA2580	Morocco	Fifi	35,02	-5,21	VI	VIf	This study
GVA2525	Morocco	Moulay Abdesalam	35,27	-5,49	VI	VIf	This study
GVA2556	Morocco	Moulay Abdesalam	35,27	-5,49	VI	VIf	This study
GVA2564	Morocco	Moulay Abdesalam	35,27	-5,49	VI	VIf	This study
GVA2566	Morocco	Moulay Abdesalam	35,27	-5,49	VI	VIf	This study
GVA2568	Morocco	Moulay Abdesalam	35,27	-5,49	VI	VIf	This study
GVA2524	Morocco	Moulay Abdessalam	35,27	-5,49	VI	VIf	This study
GVA2530	Morocco	Moulay Abdesalam	35,27	-5,49	VI	VIm	This study
GVA2533	Morocco	Moulay Abdesalam	35,27	-5,49	VI	VIl	This study
EoMarro	Morocco	Ouezan: Jeramena	34,69	-5,57	VI	VIn	This study
GVA1282	Morocco	Sidi Mimoun	33,65	-4,97	VI	VIc	This study
GVA1284	Morocco	Sidi Mimoun	33,65	-4,97	VI	VIc	This study
GVA1290	Morocco	Sidi Mimoun	33,65	-4,97	VI	VIc	This study
GVA1292	Morocco	Sidi Mimoun	33,65	-4,97	VI	VIc	This study
GVA1295	Morocco	Sidi Mimoun	33,65	-4,97	VI	VIc	This study
GVA1283	Morocco	Sidi Mimoun	33,65	-4,97	VI	VIi	This study
6420	Morocco	Kenitra Province	ND	ND	VI	VIc	Stuckas et al. 2014
6421	Morocco	Kenitra Province	ND	ND	VI	VIc	Stuckas et al. 2014
5103	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
5104	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
5105	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
5367	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
5873	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
5874	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
5875	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
5876	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
6422	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
9084	Morocco	Middle Atlas	ND	ND	VI	VIi	Stuckas et al. 2014
9085	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
9086	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
9087	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
9088	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014

Eo84	Morocco	Middle Atlas	ND	ND	VI	VIc	Stuckas et al. 2014
2456	Morocco	Rif Mountains	ND	ND	VI	VIf	Stuckas et al. 2014
2748	Morocco	Rif Mountains	ND	ND	VI	VIf	Stuckas et al. 2014
2749	Morocco	Rif Mountains	ND	ND	VI	VIf	Stuckas et al. 2014
3816	Morocco	Rif Mountains	ND	ND	VI	VIh	Stuckas et al. 2014
7306	Morocco	Rif Mountains	ND	ND	VI	VIf	Stuckas et al. 2014
9079	Morocco	Rif Mountains	ND	ND	VI	VIf	Stuckas et al. 2014
9080	Morocco	Rif Mountains	ND	ND	VI	VIf	Stuckas et al. 2014
9081	Morocco	Rif Mountains	ND	ND	VI	VIf	Stuckas et al. 2014
9082	Morocco	Rif Mountains	ND	ND	VI	VIf	Stuckas et al. 2014
9083	Morocco	Rif Mountains	ND	ND	VI	VIf	Stuckas et al. 2014
GVA3450	Portugal	Alentejo: Almograve	37,65	-8,80	VI	VIa	Velo-Antón et al. 2011
GVA3459	Portugal	Alentejo: Almograve	37,65	-8,80	VI	VIa	Velo-Antón et al. 2011
GVA3460	Portugal	Alentejo: Almograve	37,65	-8,80	VI	VIa	Velo-Antón et al. 2011
GVA3462	Portugal	Alentejo: Almograve	37,65	-8,80	VI	VIa	Velo-Antón et al. 2011
GVA2377	Portugal	Alentejo: Castro Verde	37,69	-8,04	VI	VIb	This study
Eo0101	Portugal	Algarve	37,03	-8,03	VI	VId	This study
Eo0104	Portugal	Algarve	37,03	-8,03	VI	VId	This study
Eo0105	Portugal	Algarve	37,03	-8,03	VI	VId	This study
BO1	Portugal	Boticas: Vila Real	41,74	-7,59	VI	VIa	Velo-Antón et al. 2008
BO2	Portugal	Boticas: Vila Real	41,74	-7,59	VI	VIa	Velo-Antón et al. 2008
BO3	Portugal	Boticas: Vila Real	41,74	-7,59	VI	VIa	Velo-Antón et al. 2008
BO4	Portugal	Boticas: Vila Real	41,74	-7,59	VI	VIa	Velo-Antón et al. 2008
BO5	Portugal	Boticas: Vila Real	41,74	-7,59	VI	VIa	Velo-Antón et al. 2008
Eo0301	Portugal	Caldas da Rainha	39,45	-9,13	VI	VIa	This study
Eo0305	Portugal	Caldas da Rainha	39,45	-9,13	VI	VIa	This study
Eo0206	Portugal	Castelo Branco: Monte Barata	39,70	-7,31	VI	VId	This study
Eo0207	Portugal	Castelo Branco: Monte Barata	39,70	-7,31	VI	VId	This study
GVA1424	Spain	Albacete: Laguna de los Ojos de Villaverde	38,81	-2,37	VI	VIa	This study
Eo0601	Spain	Madrid: Arganda del Rey	40,30	-3,50	VI	VIa	This study
Eo0602	Spain	Madrid: Arganda del Rey	40,30	-3,50	VI	VIa	This study
Eo0603	Spain	Madrid: Arganda del Rey	40,30	-3,50	VI	VIa	This study
Eo0604	Spain	Madrid: Arganda del Rey	40,30	-3,50	VI	VIa	This study
Eo0605	Spain	Madrid: Arganda del Rey	40,30	-3,50	VI	VId	This study
CA1	Spain	Cáceres: Jaraiz de la Vera	40,01	-5,74	VI	VId	Velo-Antón et al. 2011
CA2	Spain	Cáceres: Jaraiz de la Vera	40,01	-5,74	VI	VId	Velo-Antón et al. 2011
GVA4205	Spain	Cádiz: Campo de Gibraltar	36,23	-5,39	VI	VIn	This study
GVA4203	Spain	Cádiz: Grazalema	36,75	-5,37	VI	VIn	This study
GVA4202	Spain	Cádiz, río Roche	36,33	-6,13	VI	VIn	This study
CR1	Spain	Ciudad Real: Abenójar	38,91	-4,47	VI	VIa	Velo-Antón et al.

							2008
CR2	Spain	Ciudad Real: Abenójar	38,91	-4,47	VI	VIa	Velo-Antón et al. 2008
CR3	Spain	Ciudad Real: Abenójar	38,91	-4,47	VI	VIa	Velo-Antón et al. 2008
CR4	Spain	Ciudad Real: Abenójar	38,91	-4,47	VI	VIa	Velo-Antón et al. 2008
GVA3130	Spain	Córdoba: Cardeña	38,25	-4,32	VI	VIa	This study
GVA3128	Spain	Córdoba: Cardeña	38,25	-4,32	VI	VIk	This study
Eo0403	Spain	Cuenca: Cañadas del Hoyo	39,98	-1,86	VI	VIa	This study
Eo0407	Spain	Cuenca: Cañadas del Hoyo	39,98	-1,86	VI	VIa	This study
Eo0411	Spain	Cuenca: Cañadas del Hoyo	39,98	-1,86	VI	VIa	This study
Eo0413	Spain	Cuenca: Cañadas del Hoyo	39,98	-1,86	VI	VIa	This study
Eo0418	Spain	Cuenca: Cañadas del Hoyo	39,98	-1,86	VI	VIa	This study
GVA2426	Spain	Girona: Can Prats	41,84	2,71	II	IIa	Velo-Antón et al. 2011
GVA2427	Spain	Girona: Can Prats	41,84	2,71	II	IIa	Velo-Antón et al. 2011
GVA2435	Spain	Girona: Can Cunill	41,83	2,67	II	IIa	Velo-Antón et al. 2011
GVA2436	Spain	Girona: Can Cunill	41,83	2,67	V	Va	Velo-Antón et al. 2011
GVA2443	Spain	Girona: Can Barrots	41,84	2,68	II	IIa	Velo-Antón et al. 2011
GVA2444	Spain	Girona: Can Barrots	41,84	2,68	II	IIa	Velo-Antón et al. 2011
GVA2445	Spain	Girona: Cades de Malabella	41,82	2,78	II	IIa	Velo-Antón et al. 2011
GVA2446	Spain	Girona: Cades de Malabella	41,82	2,78	II	IIa	Velo-Antón et al. 2011
GVA2448	Spain	Girona: Cades de Malabella	41,82	2,78	II	IIa	Velo-Antón et al. 2011
not filed	Spain	Girona: La Bisbal	ND	ND	II	IIa	Velo-Antón et al. 2011
not filed	Spain	Barcelona: 30 km north of Barcelona	ND	ND	VI	VIa	Velo-Antón et al. 2011
not filed	Spain	Barcelona: Foix reservoir	ND	ND	VI	VIa	Velo-Antón et al. 2011
4879	Spain	Girona	41,98	2,81	II	IIa	Pedall et al. 2011
4880	Spain	Girona	41,98	2,81	II	IIa	Pedall et al. 2011
4881	Spain	Girona	41,98	2,81	II	IIa	Pedall et al. 2011
4882	Spain	Girona	41,98	2,81	II	IIa	Pedall et al. 2011
4883	Spain	Girona	41,98	2,81	II	IIa	Pedall et al. 2011
4884	Spain	Girona	41,98	2,81	II	IIa	Pedall et al. 2011
4885	Spain	Girona	41,98	2,81	II	IIa	Pedall et al. 2011
4886	Spain	Girona	41,98	2,81	II	IIa	Pedall et al. 2011
4812	Spain	Ebro	40,70	0,77	II	IIa	Pedall et al. 2011
4813	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4814	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4815	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4816	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4817	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4818	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4819	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011

4820	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4821	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4822	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4823	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4824	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4825	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4826	Spain	Ebro	40,70	0,77	V	Va	Pedall et al. 2011
4828	Spain	Ebro	40,70	0,77	VI	VIa	Pedall et al. 2011
4829	Spain	Ebro	40,70	0,77	VI	VIa	Pedall et al. 2011
4830	Spain	Ebro	40,70	0,77	VI	VIa	Pedall et al. 2011
4831	Spain	Ebro	40,70	0,77	VI	VId	Pedall et al. 2011
4777	Spain	Huelva: Doñana	37,03	-6,38	VI	VIb	Pedall et al. 2011
4778	Spain	Huelva: Doñana	37,03	-6,38	VI	VId	Pedall et al. 2011
4779	Spain	Huelva: Doñana	37,03	-6,38	VI	VId	Pedall et al. 2011
4798	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4799	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4800	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4801	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4802	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4803	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4804	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4805	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4806	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4807	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4808	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4809	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4810	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
4811	Spain	Huelva: Doñana	37,03	-6,38	VI	VIa	Pedall et al. 2011
GVA1425	Spain	Jaén: Aldeaquemada	38,44	-3,38	VI	VIg	This study
MA2	Spain	Madrid: El Escorial	40,57	-4,12	VI	VIa	Velo-Antón et al. 2008
MA3	Spain	Madrid: El Escorial	40,57	-4,12	VI	VIa	Velo-Antón et al. 2008
MA4	Spain	Madrid: El Escorial	40,57	-4,12	VI	VIa	Velo-Antón et al. 2008
MA5	Spain	Madrid: El Escorial	40,57	-4,12	VI	VIa	Velo-Antón et al. 2008
MA6	Spain	Madrid: El Escorial	40,57	-4,12	VI	VIa	Velo-Antón et al. 2008
Eo0601	Spain	Madrid: Arganda del Rey	40,32	-3,49	VI	VIa	This study
Eo0602	Spain	Madrid: Arganda del Rey	40,32	-3,49	VI	VIa	This study
Eo0603	Spain	Madrid: Arganda del Rey	40,32	-3,49	VI	VIa	This study
Eo0604	Spain	Madrid: Arganda del Rey	40,32	-3,49	VI	VIa	This study
Eo0605	Spain	Madrid: Arganda del Rey	40,32	-3,49	VI	VId	This study
GVA3518	Spain	Málaga	36,75	-5,29	VI	VIh	This study
GVA3519	Spain	Málaga	36,75	-5,29	VI	VIo	This study
OU50	Spain	Ourense: Baños de Molgas	42,25	-7,62	VI	VIa	Velo-Antón et al. 2008

OU55	Spain	Ourense: Baños de Molgas	42,25	-7,62	VI	VIa	Velo-Antón et al. 2008
OU56	Spain	Ourense: Baños de Molgas	42,25	-7,62	VI	VIa	Velo-Antón et al. 2008
OU57	Spain	Ourense: Baños de Molgas	42,25	-7,62	VI	VIa	Velo-Antón et al. 2008
RI1	Spain	Ourense: Ribadavia	42,31	-8,12	VI	VIc	This study
RI2	Spain	Ourense: Ribadavia	42,31	-8,12	VI	VIc	This study
RI3	Spain	Ourense: Ribadavia	42,31	-8,12	VI	VIc	This study
GVA2372	Spain	País Vasco: Hondarribia	43,36	-1,79	VI	VIa	Velo-Antón et al. 2011
GVA2373	Spain	País Vasco: Hondarribia	43,36	-1,79	VI	VIa	Velo-Antón et al. 2011
12538	Spain	Pontevedra: Gándaras de Budiño	42,16	-8,61	VI	VIa	Pedall et al. 2011
12539	Spain	Pontevedra: Gándaras de Budiño	42,16	-8,61	VI	VIa	Pedall et al. 2011
12542	Spain	Pontevedra: Gándaras de Budiño	42,16	-8,61	VI	VIa	Pedall et al. 2011
12543	Spain	Pontevedra: Gándaras de Budiño	42,16	-8,61	VI	VIa	Pedall et al. 2011
12544	Spain	Pontevedra: Gándaras de Budiño	42,16	-8,61	VI	VIa	Pedall et al. 2011
12545	Spain	Pontevedra: Gándaras de Budiño	42,16	-8,61	VI	VIa	Pedall et al. 2011
12540	Spain	Pontevedra: Gándaras de Budiño	42,16	-8,61	VI	VIe	Pedall et al. 2011
12541	Spain	Pontevedra: Gándaras de Budiño	42,16	-8,61	VI	VIe	Pedall et al. 2011
44373	Spain	Pontevedra: Gándaras de Budiño	42,16	-8,61	VI	VIa	Pedall et al. 2011
GVA3182	Spain	Robledillo	39,49	-4,35	VI	VIa	This study
GVA3183	Spain	Robledillo	39,49	-4,35	VI	VIa	This study
GVA3184	Spain	Robledillo	39,49	-4,35	VI	VIa	This study
GVA705	Spain	Salamanca: Olmedo de Camaces	40,88	-6,62	VI	VId	This study
GVA711	Spain	Salamanca: Palomares	40,39	-5,74	VI	VIa	This study
GVA2365	Spain	Zamora: Fornillos de Fermoselle	41,35	-6,32	VI	VIa	Velo-Antón et al. 2008
GVA2366	Spain	Zamora: Fornillos de Fermoselle	41,35	-6,32	VI	VIa	Velo-Antón et al. 2008
GVA2367	Spain	Zamora: Fornillos de Fermoselle	41,35	-6,32	VI	VIa	Velo-Antón et al. 2008
GVA2368	Spain	Zamora: Fornillos de Fermoselle	41,35	-6,32	VI	VIa	Velo-Antón et al. 2008
GVA682	Spain	Zamora: Ciudad Rodrigo	40,60	-6,52	VI	VIa	Velo-Antón et al. 2008
GVA683	Spain	Zamora: Ciudad Rodrigo	40,60	-6,52	VI	VIa	Velo-Antón et al. 2008
GVA684	Spain	Zamora: Ciudad Rodrigo	40,60	-6,52	VI	VIa	Velo-Antón et al. 2008
GVA685	Spain	Zamora: Ciudad Rodrigo	40,60	-6,52	VI	VIa	Velo-Antón et al. 2008
GVA689	Spain	Salamanca: Zorita	41,16	-6,04	VI	VIa	This study
GVA692	Spain	Salamanca: Zorita	41,16	-6,04	VI	VIa	This study
GVA695	Spain	Salamanca: Zorita	41,16	-6,04	VI	VIa	This study
Eo0714	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	V	Va	This study
Eo0768	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	V	Va	This study
Eo0705	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	VI	VIa	This study
Eo0722	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	VI	VIa	This study
Eo0734	Spain	Valencia: Sagunto, Burriana &	39,92	-0,05	VI	VIa	This study

		Moro					
Eo0743	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	VI	VIa	This study
Eo0756	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	VI	VIa	This study
Eo0759	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	VI	VIa	This study
Eo0770	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	VI	VIa	This study
Eo0778	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	VI	VIa	This study
Eo0782	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	VI	VIa	This study
Eo0774	Spain	Valencia: Sagunto, Burriana & Moro	39,92	-0,05	VI	VIj	This study
4878	Spain	Valencia	39,33	-0,36	VI	VIa	Pedall et al. 2011
4893	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4894	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4895	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4896	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4897	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4898	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4899	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4900	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4901	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4902	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4903	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4904	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4905	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4906	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4907	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4908	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4909	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
4910	Spain	Valencia	39,33	-0,36	VI	VIa	Pedall et al. 2011
4911	Spain	Valencia	39,33	-0,36	VI	VIa	Pedall et al. 2011
4912	Spain	Valencia	39,33	-0,36	VI	VIa	Pedall et al. 2011
4913	Spain	Valencia	39,33	-0,36	VI	VIa	Pedall et al. 2011
4914	Spain	Valencia	39,33	-0,36	VI	VIa	Pedall et al. 2011
4915	Spain	Valencia	39,33	-0,36	VI	VIa	Pedall et al. 2011
4916	Spain	Valencia	39,33	-0,36	VI	VIa	Pedall et al. 2011
4917	Spain	Valencia	39,33	-0,36	V	Va	Pedall et al. 2011
44358	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44359	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44360	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44361	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44362	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44363	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44364	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44365	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011

44366	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44367	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44368	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44369	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44370	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44371	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
44372	Spain	Navarra	42,28	-1,75	II	IIa	Pedall et al. 2011
4832	Spain	Estaña	42,05	0,55	II	IIa	Pedall et al. 2011
4833	Spain	Estaña	42,05	0,55	II	IIa	Pedall et al. 2011
4834	Spain	Estaña	42,05	0,55	II	IIa	Pedall et al. 2011
4835	Spain	Estaña	42,05	0,55	II	IIa	Pedall et al. 2011
4836	Spain	Estaña	42,05	0,55	II	IIa	Pedall et al. 2011
4837	Spain	Estaña	42,05	0,55	II	IIa	Pedall et al. 2011
4838	Spain	Estaña	42,05	0,55	VI	VId	Pedall et al. 2011
4887	Spain	Tarragona	41,12	1,24	V	Va	Pedall et al. 2011
4888	Spain	Tarragona	41,12	1,24	V	Va	Pedall et al. 2011
4889	Spain	Tarragona	41,12	1,24	V	Va	Pedall et al. 2011
4890	Spain	Tarragona	41,12	1,24	V	Va	Pedall et al. 2011
4891	Spain	Tarragona	41,12	1,24	V	Va	Pedall et al. 2011
4892	Spain	Tarragona	41,12	1,24	V	Va	Pedall et al. 2011
4784	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4785	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4786	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4787	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4788	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4789	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4790	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4791	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4792	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4793	Spain	Girona: Sils	41,80	2,75	V	Va	Pedall et al. 2011
4794	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4795	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4796	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011
4797	Spain	Girona: Sils	41,80	2,75	II	IIa	Pedall et al. 2011

Table 7.3: Summary table of all samples used for the microsatellite data used in this work. Sample: sample code; Country: country from where the sample was collected; Locality: sampling locality; Lat: Latitude; Long: Longitude; Reference: Source of the genotype.

Sample	Country	Population	Lat	Long	Reference
R1	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R2	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R3	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R4	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R5	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R6	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R7	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R8	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R9	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R10	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R13	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R14	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R15	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R16	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R17	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R18	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R19	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R21	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R23	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R24	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R25	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R26	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R27	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R29	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R30	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R31	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R32	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R33	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R34	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R35	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R36	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R40	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R41	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R42	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
R43	Spain	Ribadavia	42.3	-8.11	Velo-Antón et al. (2011)
OA	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O5	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O10	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O11	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O17	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O19	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)

O20	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O24	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O25	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O33	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O37	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O41	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O45	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O46	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O48	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O49	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O50	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O53	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O54	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O55	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O56	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O57	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O58	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O59	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O60	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O61	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O62	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O63	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O64	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O65	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O66	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
O67	Spain	Ourense	42.24	-7.61	Velo-Antón *et al.* (2008)
P17	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P23	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P24	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P41	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P43	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P44	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P53	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P58	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P62	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P63	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P66	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P70	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P71	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P72	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P73	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P74	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P81	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P86	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)

P88	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P90	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P91	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P92	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P93	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P94	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P104	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P106	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P107	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P109	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P110	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
P111	Spain	Porriño	42.12	-8.63	Velo-Antón *et al.* (2008)
B1	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B2	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B3	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B4	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B5	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B6	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B7	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B8	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B9	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B10	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B11	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B12	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B13	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B14	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B15	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B16	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B17	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B18	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B19	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B20	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B21	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B22	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B23	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B24	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B25	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B26	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B27	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
B28	Portugal	Boticas	41.73	-7.59	Velo-Antón et al. (2008)
ZA1	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA2	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA3	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA4	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)

ZA5	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA6	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA7	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA8	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA9	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA10	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA11	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA12	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA13	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA14	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA15	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA16	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
ZA17	Spain	Zamora	41.34	-6.31	Velo-Antón et al. (2008)
GVA689	Spain	Salamanca	41.16	-6.04	This study
GVA690	Spain	Salamanca	41.16	-6.04	This study
GVA691	Spain	Salamanca	41.16	-6.04	This study
GVA692	Spain	Salamanca	41.16	-6.04	This study
GVA693	Spain	Salamanca	41.16	-6.04	This study
GVA694	Spain	Salamanca	41.16	-6.04	This study
GVA695	Spain	Salamanca	41.16	-6.04	This study
GVA696	Spain	Salamanca	41.16	-6.04	This study
GVA697	Spain	Salamanca	41.16	-6.04	This study
GVA698	Spain	Salamanca	41.16	-6.04	This study
GVA699	Spain	Salamanca	41.16	-6.04	This study
GVA700	Spain	Salamanca	41.16	-6.04	This study
GVA701	Spain	Salamanca	41.16	-6.04	This study
GVA702	Spain	Salamanca	41.16	-6.04	This study
GVA703	Spain	Salamanca	40.88	-6.62	This study
GVA704	Spain	Salamanca	40.88	-6.62	This study
GVA705	Spain	Salamanca	40.88	-6.62	This study
S1	Spain	Salamanca	40.59	-6.52	Velo-Antón et al. (2008)
S2	Spain	Salamanca	40.59	-6.52	Velo-Antón et al. (2008)
S3	Spain	Salamanca	40.59	-6.52	Velo-Antón et al. (2008)
S4	Spain	Salamanca	40.59	-6.52	Velo-Antón et al. (2008)
S5	Spain	Salamanca	40.59	-6.52	Velo-Antón et al. (2008)
S6	Spain	Salamanca	40.59	-6.52	Velo-Antón et al. (2008)
S7	Spain	Salamanca	40.59	-6.52	Velo-Antón et al. (2008)
GVA706	Spain	Salamanca	40.39	-5.74	This study
GVA707	Spain	Salamanca	40.39	-5.74	This study
GVA708	Spain	Salamanca	40.39	-5.74	This study
GVA709	Spain	Salamanca	40.39	-5.74	This study
GVA710	Spain	Salamanca	40.39	-5.74	This study
GVA711	Spain	Salamanca	40.39	-5.74	This study
GVA712	Spain	Salamanca	40.39	-5.74	This study

GVA713	Spain	Salamanca	40.39	-5.74	This study
GVA714	Spain	Salamanca	40.39	-5.74	This study
GVA715	Spain	Salamanca	40.39	-5.74	This study
GVA716	Spain	Salamanca	40.39	-5.74	This study
GVA717	Spain	Salamanca	40.39	-5.74	This study
M1	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M2	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M3	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M4	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M5	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M6	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M7	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M8	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M9	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M10	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M11	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M12	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M13	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M14	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M15	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M16	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M17	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M18	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M19	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M20	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M21	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M22	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M23	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M24	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M25	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M26	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M27	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M28	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M29	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M30	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
M32	Spain	Madrid	40.57	-4.11	Velo-Antón et al. (2008)
Eo0601	Spain	Argana del Rey	40.3	-3.5	This study
Eo0603	Spain	Argana del Rey	40.3	-3.5	This study
Eo0604	Spain	Argana del Rey	40.3	-3.5	This study
Eo0605	Spain	Argana del Rey	40.3	-3.5	This study
Eo0602	Spain	Argana del Rey	40.3	-3.5	This study
GVA3182	Spain	Robledillo	40.01	-5.74	This study
GVA3183	Spain	Robledillo	40.01	-5.74	This study
GVA3184	Spain	Robledillo	40.01	-5.74	This study

Eo0401	Spain	Cuenca	39.97	-1.86	This study
Eo0406	Spain	Cuenca	39.97	-1.86	This study
Eo0407	Spain	Cuenca	39.97	-1.86	This study
Eo0408	Spain	Cuenca	39.97	-1.86	This study
Eo0409	Spain	Cuenca	39.97	-1.86	This study
Eo0402	Spain	Cuenca	39.97	-1.86	This study
Eo0410	Spain	Cuenca	39.97	-1.86	This study
Eo0411	Spain	Cuenca	39.97	-1.86	This study
Eo0412	Spain	Cuenca	39.97	-1.86	This study
Eo0413	Spain	Cuenca	39.97	-1.86	This study
Eo0414	Spain	Cuenca	39.97	-1.86	This study
Eo0415	Spain	Cuenca	39.97	-1.86	This study
Eo0416	Spain	Cuenca	39.97	-1.86	This study
Eo0417	Spain	Cuenca	39.97	-1.86	This study
Eo0418	Spain	Cuenca	39.97	-1.86	This study
Eo0403	Spain	Cuenca	39.97	-1.86	This study
Eo0404	Spain	Cuenca	39.97	-1.86	This study
Eo0405	Spain	Cuenca	39.97	-1.86	This study
V6	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V7	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V8	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V9	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V10	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V11	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V12	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V13	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V14	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V15	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V16	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V31	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V32	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V33	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V34	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V35	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V36	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V37	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V38	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V39	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V41	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V42	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V43	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
V44	Spain	Valencia	39.92	-0.05	Velo-Antón et al. (2008)
Eo0201	Portugal	Castelo Branco	39.7	-7.31	This study
Eo0202	Portugal	Castelo Branco	39.7	-7.31	This study

Eo0203	Portugal	Castelo Branco	39.7	-7.31	This study
Eo0204	Portugal	Castelo Branco	39.7	-7.31	This study
Eo0205	Portugal	Castelo Branco	39.7	-7.31	This study
Eo0206	Portugal	Castelo Branco	39.7	-7.31	This study
Eo0207	Portugal	Castelo Branco	39.7	-7.31	This study
Eo0208	Portugal	Castelo Branco	39.7	-7.31	This study
Eo0209	Portugal	Castelo Branco	39.7	-7.31	This study
Eo0210	Portugal	Castelo Branco	39.7	-7.31	This study
Eo0301	Portugal	Paul da Tornada	39.45	-9.13	This study
Eo0302	Portugal	Paul da Tornada	39.45	-9.13	This study
Eo0303	Portugal	Paul da Tornada	39.45	-9.13	This study
Eo0304	Portugal	Paul da Tornada	39.45	-9.13	This study
Eo0305	Portugal	Paul da Tornada	39.45	-9.13	This study
C1	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C2	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C3	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C4	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C6	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C7	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C8	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C9	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C10	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C11	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C12	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C13	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C14	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C16	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C17	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C18	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C19	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C20	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C21	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C22	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C23	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
C24	Spain	Ciudad Real	38.9	-4.47	Velo-Antón et al. (2008)
GVA3106	Spain	Cardeña	38.26	-4.32	This study
GVA3118	Spain	Cardeña	38.26	-4.32	This study
GVA3119	Spain	Cardeña	38.26	-4.32	This study
GVA3128	Spain	Cardeña	38.26	-4.32	This study
GVA3129	Spain	Cardeña	38.26	-4.32	This study
GVA3130	Spain	Cardeña	38.26	-4.32	This study
GVA2376	Portugal	Castroverde	37.69	-8.04	This study
GVA2377	Portugal	Castroverde	37.69	-8.04	This study
A1	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)

A2	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A3	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A4	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A5	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A6	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A7	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A8	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A9	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A10	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A11	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A12	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A13	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A14	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A15	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A16	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A17	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A18	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A19	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A20	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A21	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A22	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A23	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A24	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A25	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A26	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
A27	Portugal	Almograve	37.65	-8.79	Velo-Antón et al. (2008)
D1	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D2	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D3	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D4	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D5	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D6	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D7	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D8	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D9	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D10	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D11	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D12	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D13	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D14	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D15	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D16	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D17	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D18	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)

D19	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D20	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D21	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D22	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D23	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D24	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D25	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D26	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D27	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D28	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D29	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D30	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D31	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D32	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D33	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D34	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D35	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
D36	Spain	Doñana	37.04	-6.59	Velo-Antón et al. (2008)
Eo0115	Portugal	Algarve	37.03	-8.03	This study
Eo0116	Portugal	Algarve	37.03	-8.03	This study
Eo0117	Portugal	Algarve	37.03	-8.03	This study
Eo0112	Portugal	Algarve	37.03	-8.03	This study
Eo0113	Portugal	Algarve	37.03	-8.03	This study
Eo0118	Portugal	Algarve	37.03	-8.03	This study
Eo0119	Portugal	Algarve	37.03	-8.03	This study
Eo0120	Portugal	Algarve	37.03	-8.03	This study
Eo0121	Portugal	Algarve	37.03	-8.03	This study
Eo0122	Portugal	Algarve	37.03	-8.03	This study
Eo0123	Portugal	Algarve	37.03	-8.03	This study
Eo0124	Portugal	Algarve	37.03	-8.03	This study
Eo0125	Portugal	Algarve	37.03	-8.03	This study
Eo0101	Portugal	Algarve	37.03	-8.03	This study
Eo0102	Portugal	Algarve	37.03	-8.03	This study
Eo0103	Portugal	Algarve	37.03	-8.03	This study
Eo0104	Portugal	Algarve	37.03	-8.03	This study
Eo0105	Portugal	Algarve	37.03	-8.03	This study
Eo0108	Portugal	Algarve	37.03	-8.03	This study
Eo0107	Portugal	Algarve	37.03	-8.03	This study
GVA2524	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2525	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2526	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2527	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2528	Morocco	Moulay	35.27	-5.48	This study

		Abdesalam			
GVA2529	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2530	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2532	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2533	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2553	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2554	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2555	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2556	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2557	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2558	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2559	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2561	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2562	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2563	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2564	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2565	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2566	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2567	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA2568	Morocco	Moulay Abdesalam	35.27	-5.48	This study
GVA1252	Morocco	Fifi	35.02	-5.2	This study
GVA1253	Morocco	Fifi	35.02	-5.2	This study
GVA1254	Morocco	Fifi	35.02	-5.2	This study
GVA1255	Morocco	Fifi	35.02	-5.2	This study
GVA1259	Morocco	Fifi	35.02	-5.2	This study
GVA1261	Morocco	Fifi	35.02	-5.2	This study
GVA1262	Morocco	Fifi	35.02	-5.2	This study
GVA1263	Morocco	Fifi	35.02	-5.2	This study
GVA1264	Morocco	Fifi	35.02	-5.2	This study
GVA1265	Morocco	Fifi	35.02	-5.2	This study
GVA1266	Morocco	Fifi	35.02	-5.2	This study
GVA1267	Morocco	Fifi	35.02	-5.2	This study
GVA1268	Morocco	Fifi	35.02	-5.2	This study
GVA1269	Morocco	Fifi	35.02	-5.2	This study
GVA1271	Morocco	Fifi	35.02	-5.2	This study
GVA1272	Morocco	Fifi	35.02	-5.2	This study
GVA1273	Morocco	Fifi	35.02	-5.2	This study

GVA1274	Morocco	Fifi	35.02	-5.2	This study
GVA1275	Morocco	Fifi	35.02	-5.2	This study
GVA1276	Morocco	Fifi	35.02	-5.2	This study
GVA1277	Morocco	Fifi	35.02	-5.2	This study
GVA1278	Morocco	Fifi	35.02	-5.2	This study
GVA1280	Morocco	Fifi	35.02	-5.2	This study
GVA1281	Morocco	Fifi	35.02	-5.2	This study
GVA2579	Morocco	Fifi	35.02	-5.2	This study
GVA2580	Morocco	Fifi	35.02	-5.2	This study
GVA1282	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1283	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1284	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1285	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1286	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1287	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1288	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1289	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1290	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1291	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1292	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1293	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1294	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1295	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1296	Morocco	Sidi Mimoun	33.64	-4.96	This study
GVA1297	Morocco	Sidi Mimoun	33.64	-4.96	This study
UO1	----	unknown	----	----	
UO2	----	unknown	----	----	
UO3	----	unknown	----	----	
UO4	----	unknown	----	----	
UO5	----	unknown	----	----	
UO6	----	unknown	----	----	
UO7	----	unknown	----	----	
UO8	----	unknown	----	----	
UO9	----	unknown	----	----	
UO10	----	unknown	----	----	
UO11	----	unknown	----	----	
UO12	----	unknown	----	----	
UO13	----	unknown	----	----	
UO14	----	unknown	----	----	
UO15	----	unknown	----	----	
UO16	----	unknown	----	----	
UO17	----	unknown	----	----	
UO18	----	unknown	----	----	
UO19	----	unknown	----	----	

UO20	----	unknown	----	----
UO21	----	unknown	----	----
UO22	----	unknown	----	----
UO23	----	unknown	----	----
UO24	----	unknown	----	----
UO25	----	unknown	----	----
UO26	----	unknown	----	----
UO27	----	unknown	----	----
UO28	----	unknown	----	----
UO29	----	unknown	----	----
UO30	----	unknown	----	----

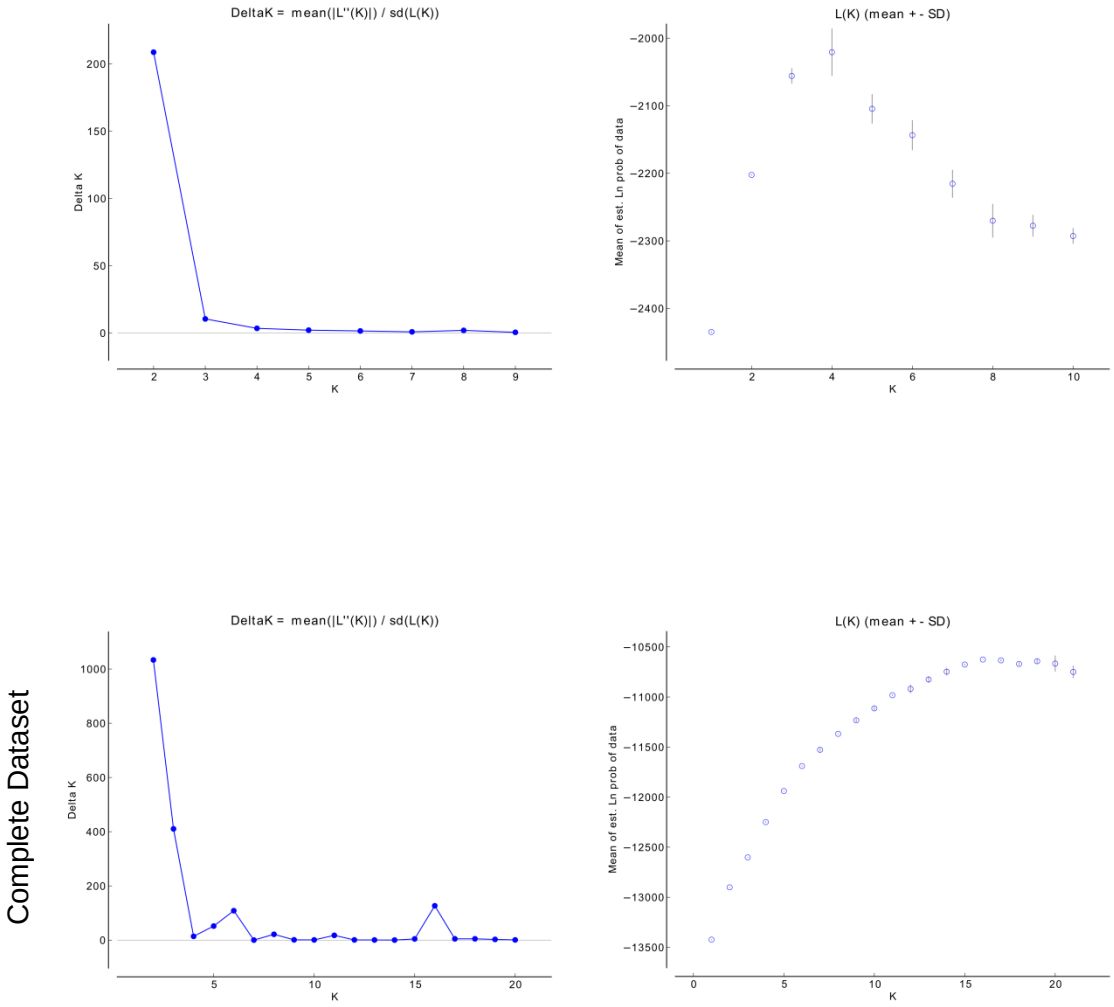

Figure 7.1: Graphical output of the Structure Harvester results for Delta K and L(k); **Top:** Results for the dataset used in Chapter 2; **Bottom:** Results for the dataset used in Chapters 3 and 4.

Script 1: performing a PCA using the raster package for R

```
require(raster)

dir <- Directory
names <- list()
label <- list.files(dir)
for (i in label) {
    temp <- paste(i, sep = "")
    names <- append(names, temp)
}

rBrick <- brick(names)


PCA <- prcomp(rBrick[])
Predic <- predict(PCA, rBrick[])

Results <- raster(label[1])
Results[] <- Predic[, 1]

plot(Results)
WR <- writeRaster(Results, filename = "allelfre.asc", format =
"ascii"))
```

Script 2: Sliding window analysis for the expected heterozygosity and on the allelic frequencies

Script 2.1: Allelic frequencies

```
### total ####

# load the file: INPUT file should be in the following format (it works
# better without header as it allows the for routine to be used for
# automaticaly explore your data): POP Latitude freq_Allele1 freq_Allele2
# ... freq_Allele n

# example

# Pop1 37 0.5 0.1 n


# Load the data
al <- read.csv("allfreq.txt", sep = " ", header = FALSE, )

# #load a label file #the label file should contain a single collum with
# the variable names (in order to name saved images and graphic labels)
label <- read.csv("lab.txt", header = FALSE, stringsAsFactors = FALSE)
# #count the number of collumns in the data (it will be necessary further
# on)

nCollumns <- length(al)
listNames <- list()


lmp <- function(modelobject) {
    if (class(modelobject) != "lm")
        stop("Not an object of class 'lm' ")
    f <- summary(modelobject)$fstatistic
    p <- pf(f[1], f[2], f[3], lower.tail = F)
    attributes(p) <- NULL
    return(p)
}

pVals <- list()
rSqs <- list()
allName <- list()
minAllFreq <- list()
maxAllFreq <- list()
dFreq <- list()
```

```
for (i in 1:(nCollumns - 3)) {
    df <- data.frame()
    data <- data.frame(al[, 2], al[, i + 3])
    windowWidth <- 1
    position <- min(al[, 2])
    posList <- list()
    posAvg <- list()

    if (length(data[data[, 2] > 0, 2]) <= 1) {

        next
    }


    while (TRUE) {
        d <- data[data[, 1] < position + windowWidth & data[, 1] > position -
            windowWidth, ]

        posList <- append(posList, position)
        posAvg <- append(posAvg, mean(d[, 2]))
        position <- position + 0.2

        if (position > max(al[, 2]) + 1) {
            break()
        }
    }
    lLA <- list(Lat = posList, Avg = posAvg)
    df <- data.frame(matrix(unlist(lLA), nrow = length(posList), byrow = F))
    alleleName <- label[i + 3, ]
    savePath <- paste("a/", alleleName, ".png", sep = "")



    coordinate <- df[, 1]
    value <- df[, 2]
    try(lmfit <- lm(value ~ coordinate))
    pVal <- lmp(lmfit)
    rSq <- summary(lmfit)$r.squared

    pVals <- append(pVals, pVal)
    rSqs <- append(rSqs, rSq)
    allName <- append(allName, alleleName)
    minAllFreq <- append(minAllFreq, min(value))
    maxAllFreq <- append(maxAllFreq, max(value))
    dFreq <- append(dFreq, max(value) - min(value))



    # Plot and save
    png(file = savePath)
    lo <- loess(value ~ coordinate)
    xl <- seq(min(coordinate), max(coordinate), (max(coordinate) -
min(coordinate))/1000)

    plot(coordinate, value, ylab = alleleName, xlab = "Latitude", pch = 19)
    lines(xl, predict(lo, xl), col = "red", lwd = 2)
    abline(lmfit, col = "green")
    dev.off()

}
temp <- list(Names = allName, `p-values` = pVals, `r-squared` = rSqs, minAllFreq =
minAllFreq,
    maxAllFreq = maxAllFreq, dFreq = dFreq)
stats <- data.frame(matrix(unlist(temp), nrow = length(pVals), byrow = F))
```

Script 2.2: Heterozygosity

```
# #Dist he


# Sliding
data <- read.csv("degreegenetic.txt", sep = "\t", header = T)
windowWidth <- 1
```

```
position <- 37
posList <- list()
posAvg <- list()

lmp <- function(modelobject) {
    if (class(modelobject) != "lm")
        stop("Not an object of class 'lm' ")
    f <- summary(modelobject)$fstatistic
    p <- pf(f[1], f[2], f[3], lower.tail = F)
    attributes(p) <- NULL
    return(p)
}


while (TRUE) {
    d <- data[data[, 2] < position + windowWidth & data[, 2] > position - windowWidth,
        ]

    posList <- append(posList, position)
    posAvg <- append(posAvg, mean(d[, 9]))
    position <- position + 0.2

    if (position > max(data[, 2]) + 1) {
        break()
    }
}
lLA <- list(Lat = posList, Avg = posAvg)
df <- data.frame(matrix(unlist(lLA), nrow = length(posList), byrow = F))
coordinate <- df[, 1]
value <- df[, 2]
try(lmfit <- lm(value ~ coordinate))
pVal <- lmp(lmfit)
rSq <- summary(lmfit)$r.squared


pdf(file = "Heterozygosity.pdf", paper = "a4r", width = 20)
plot(df[, 1], df[, 2], xlab = "Latitude", ylab = "Mean Heterozygosity", col = "red")
abline(lmfit, col = "red", lwd = 1)
neww <- data.frame(x = seq(from = min(coordinate), to = max(coordinate), length.out =
32))
c.lim <- as.data.frame(predict(lmfit, neww, level = 0.95, interval = "confidence"))
lines(cbind(neww, c.lim$lwr), col = "blue", lty = "dashed")
lines(cbind(neww, c.lim$upr), col = "blue", lty = "dashed")
dev.off()
```

Table 7.4: List of non private alleles that were tested for signs of allelic frequency clines. Highlighted in red are those alleles that had a p-value > 0.001 and a Amplitude of frequency of at least 0.2. Allele: List of non-private allele here tested; p-val: significance value of the linear regression; r^2: measures if a good fit to the linear model was found or not; minFreq: minimum observed frequency of a allele in all populations; maxFreq: maximum observed frequency of a allele in all populations; Amplitude: Absolute difference between minFreq and maxFreq.

Allele	p-val	r^2	minFreq	maxFreq	Amplitude
D114.104	3.44E-04	0.24	0.00	0.17	0.17
D114.108	3.09E-05	0.31	0.00	0.03	0.03
D114.112	1.30E-05	0.34	0.00	0.32	0.32
D114.116	4.87E-10	0.56	0.00	0.30	0.30
D114.120	6.69E-06	0.35	0.14	0.76	0.61
D114.124	8.13E-01	0.00	0.05	0.31	0.26
D114.128	7.42E-01	0.00	0.00	0.25	0.25
D114.132	2.86E-07	0.43	0.00	0.15	0.15
D114.136	7.86E-02	0.06	0.00	0.14	0.14
D114.140	8.50E-01	0.00	0.00	0.07	0.07
D114.144	8.44E-01	0.00	0.00	0.08	0.08
D114.152	1.47E-02	0.12	0.00	0.04	0.04
D16.153	3.28E-01	0.02	0.00	0.08	0.08
D16.157	2.46E-01	0.03	0.00	0.01	0.01
D16.161	2.57E-03	0.18	0.00	0.01	0.01
D16.165	5.22E-08	0.47	0.00	0.29	0.29
D16.169	1.31E-01	0.05	0.00	0.22	0.22
D16.173	2.68E-02	0.10	0.00	0.32	0.32
D16.177	1.07E-03	0.21	0.00	0.18	0.18
D16.181	2.65E-05	0.32	0.00	0.06	0.06
D16.185	1.05E-04	0.28	0.00	0.37	0.37
D16.189	7.37E-02	0.07	0.02	0.11	0.09
D16.193	1.58E-04	0.26	0.00	0.22	0.22
D16.197	6.34E-01	0.00	0.00	0.16	0.16
D16.201	2.16E-02	0.11	0.02	0.46	0.43
D16.205	3.22E-09	0.53	0.00	0.27	0.27
D16.209	3.21E-07	0.43	0.00	0.09	0.09
D16.213	9.72E-04	0.21	0.00	0.15	0.15
D16.217	3.86E-03	0.16	0.00	0.07	0.07
D16.221	1.57E-06	0.39	0.00	0.04	0.04
D16.229	4.32E-06	0.36	0.00	0.03	0.03
D87.218	2.22E-11	0.62	0.00	0.01	0.01
D87.222	7.67E-02	0.07	0.00	0.08	0.08
D87.226	3.02E-03	0.17	0.00	0.22	0.22
D87.230	3.70E-13	0.68	0.00	0.41	0.41
D87.234	1.91E-12	0.66	0.02	0.59	0.57
D87.238	5.25E-03	0.15	0.02	0.16	0.14
D87.242	1.33E-16	0.77	0.03	0.59	0.56
D87.245	6.83E-01	0.00	0.00	0.08	0.08

D87.246	5.83E-05	0.29	0.03	0.27	0.24
D87.250	5.67E-07	0.42	0.00	0.15	0.15
D87.254	6.51E-01	0.00	0.00	0.16	0.16
D87.258	3.04E-11	0.61	0.00	0.24	0.24
D87.262	2.81E-15	0.74	0.00	0.02	0.02
D87.270	5.80E-01	0.01	0.00	0.01	0.01
D93.224	9.16E-01	0.00	0.00	0.14	0.14
D93.228	2.01E-01	0.03	0.00	0.07	0.07
D93.232	3.66E-20	0.84	0.00	0.52	0.52
D93.236	6.87E-06	0.35	0.00	0.31	0.31
D93.240	1.82E-04	0.26	0.01	0.59	0.57
D93.244	4.99E-09	0.52	0.00	0.15	0.15
D93.248	4.52E-01	0.01	0.00	0.12	0.12
D93.252	9.41E-06	0.34	0.00	0.41	0.41
D93.256	8.10E-08	0.46	0.00	0.30	0.30
D93.260	3.54E-08	0.48	0.00	0.15	0.15
D93.264	3.27E-15	0.74	0.00	0.10	0.10
D93.272	1.08E-05	0.34	0.00	0.10	0.10
B08.194	1.50E-12	0.66	0.00	0.45	0.45
B08.197	2.69E-09	0.53	0.00	0.59	0.59
B08.200	1.61E-10	0.58	0.00	0.37	0.37
B08.203	2.23E-08	0.49	0.17	0.61	0.44
D88.136	2.42E-04	0.25	0.00	0.04	0.04
D88.140	1.64E-02	0.12	0.00	0.33	0.33
D88.144	1.52E-04	0.27	0.00	0.25	0.25
D88.148	2.72E-04	0.25	0.00	0.19	0.19
D88.152	7.09E-03	0.14	0.00	0.31	0.31
D88.156	6.83E-05	0.29	0.03	0.21	0.19
D88.160	1.03E-02	0.13	0.03	0.19	0.16
D88.164	1.36E-02	0.12	0.04	0.71	0.67
D88.168	4.27E-01	0.01	0.00	0.21	0.21
D88.172	8.13E-01	0.00	0.00	0.16	0.16
D88.176	1.83E-09	0.54	0.00	0.21	0.21
D88.180	1.02E-02	0.13	0.00	0.09	0.09
D88.184	4.84E-01	0.01	0.00	0.02	0.02
D88.188	1.36E-01	0.05	0.00	0.02	0.02
D51.231	3.82E-01	0.02	0.00	0.07	0.07
D51.235	7.26E-01	0.00	0.00	0.04	0.04
D51.239	1.98E-06	0.38	0.00	0.05	0.05
D51.243	1.59E-03	0.19	0.00	0.11	0.11
D51.247	1.49E-01	0.04	0.00	0.12	0.12
D51.251	6.85E-01	0.00	0.00	0.21	0.21
D51.255	5.51E-02	0.08	0.00	0.41	0.41
D51.259	1.71E-01	0.04	0.00	0.16	0.16

D51.263	6.82E-05	0.29	0.00	0.20	0.20
D51.267	8.04E-03	0.14	0.00	0.25	0.25
D51.271	7.37E-06	0.35	0.01	0.40	0.39
D51.275	8.86E-06	0.35	0.03	0.37	0.34
D51.279	1.59E-01	0.04	0.01	0.10	0.09
D51.283	2.66E-03	0.18	0.00	0.25	0.25
D51.287	7.07E-07	0.41	0.00	0.07	0.07
D51.291	2.57E-01	0.03	0.00	0.02	0.02
D51.295	2.05E-01	0.03	0.00	0.03	0.03
D51.299	1.40E-09	0.55	0.00	0.03	0.03

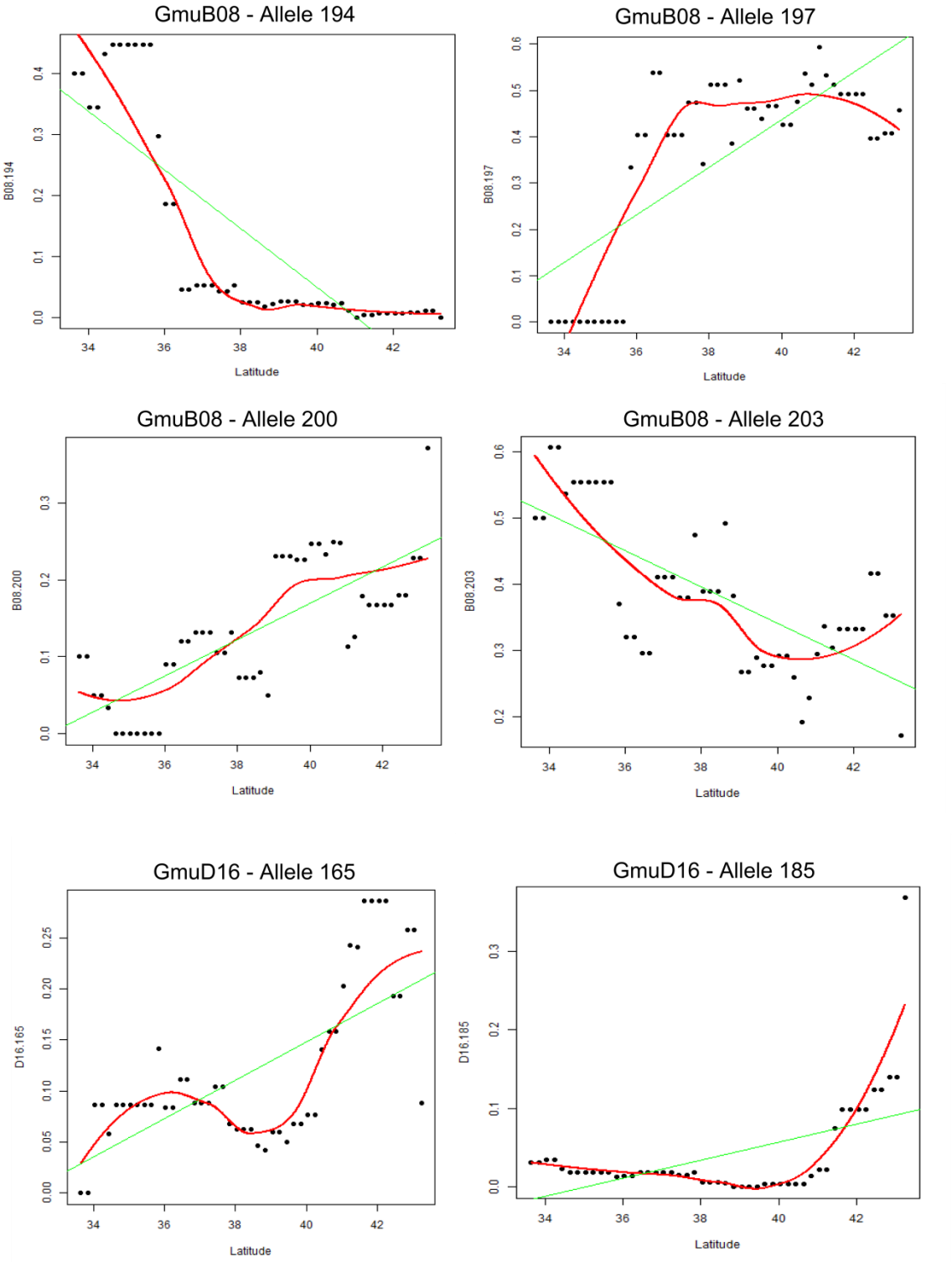

Figure 7.2: Allelic patterns of all alleles showing signs of allelic frequency clines.

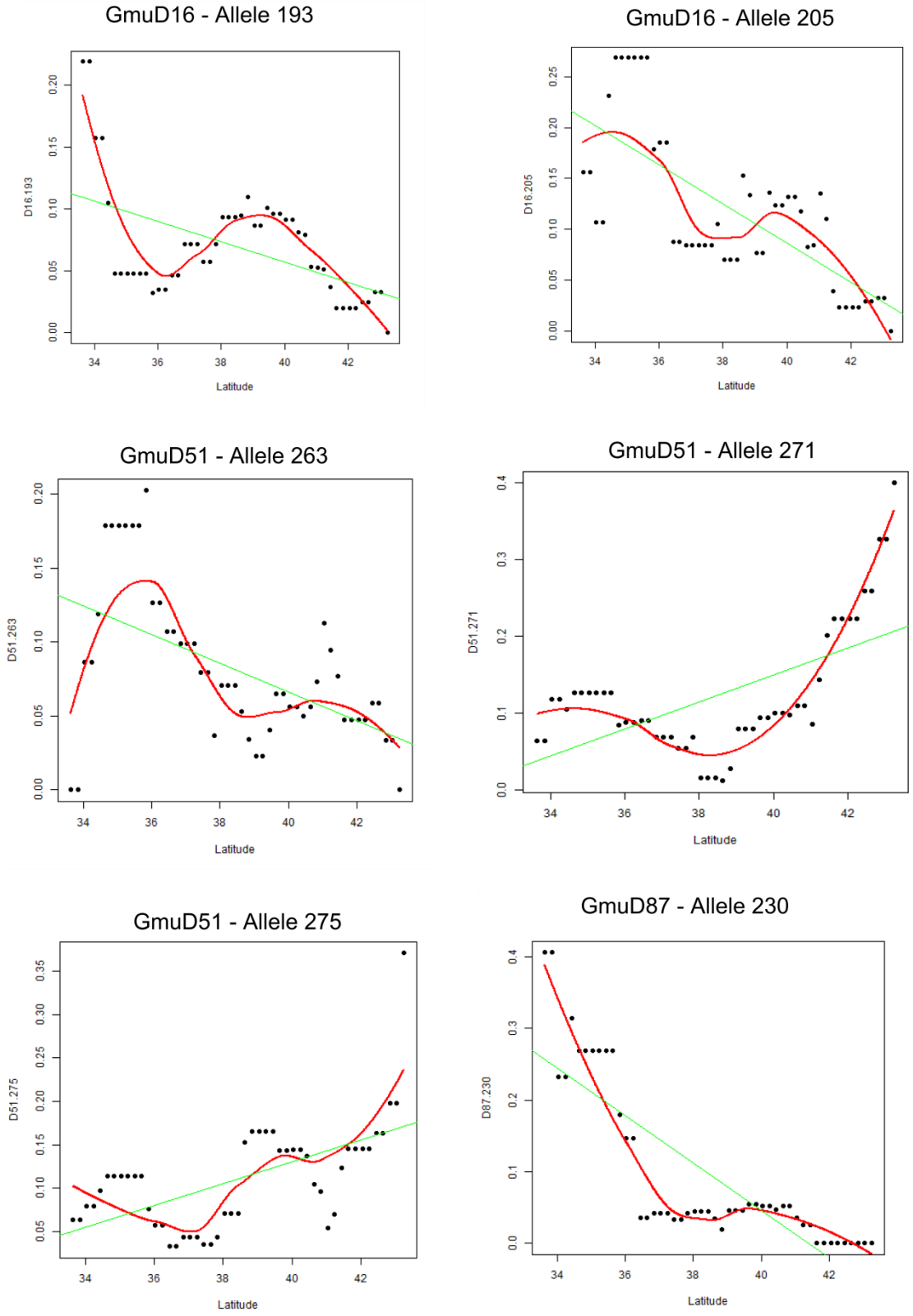

Figure 7.3: Allelic patterns of all alleles showing signs of allelic frequency clines.

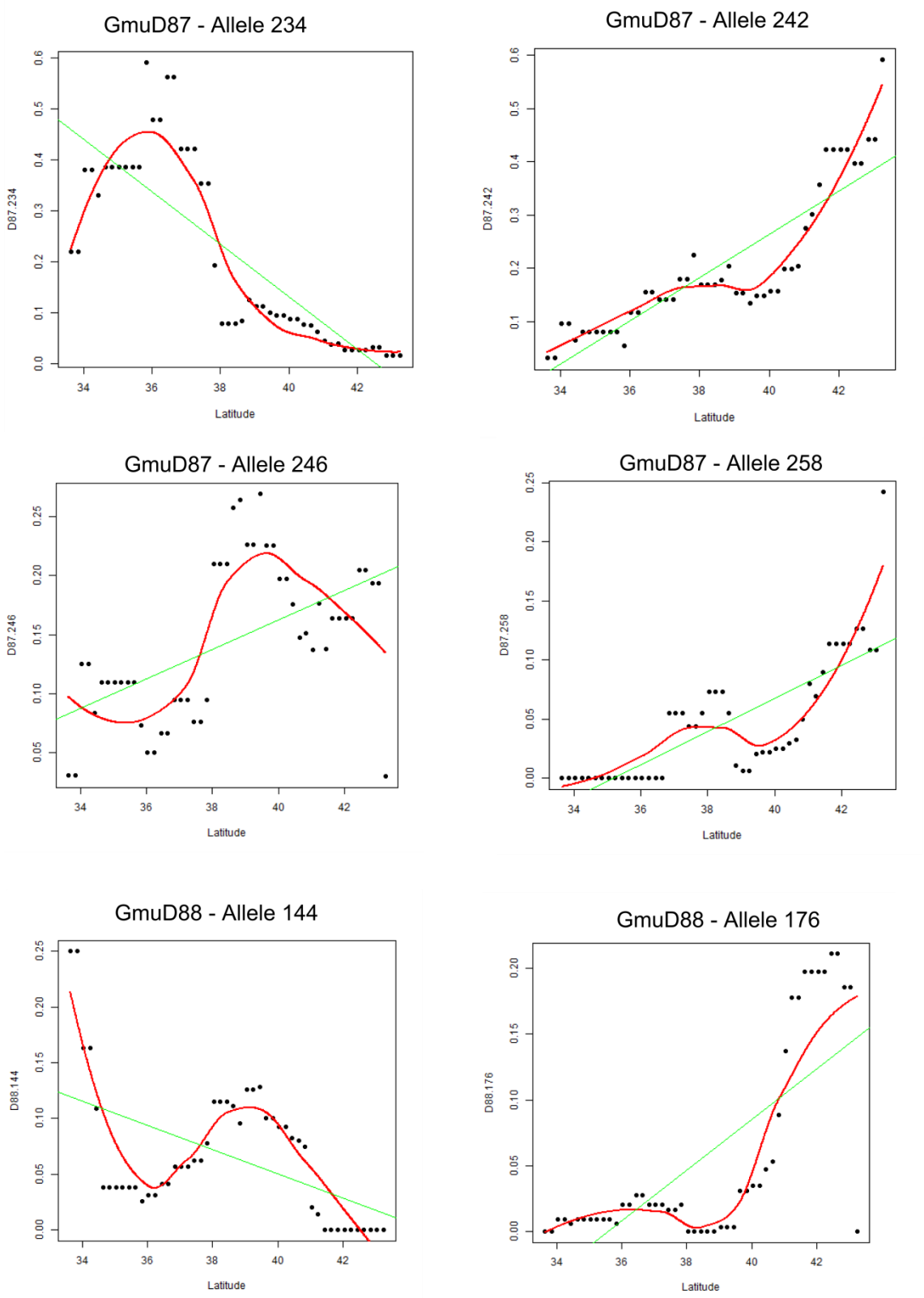

Figure 7.4: Allelic patterns of all alleles showing signs of allelic frequency clines.

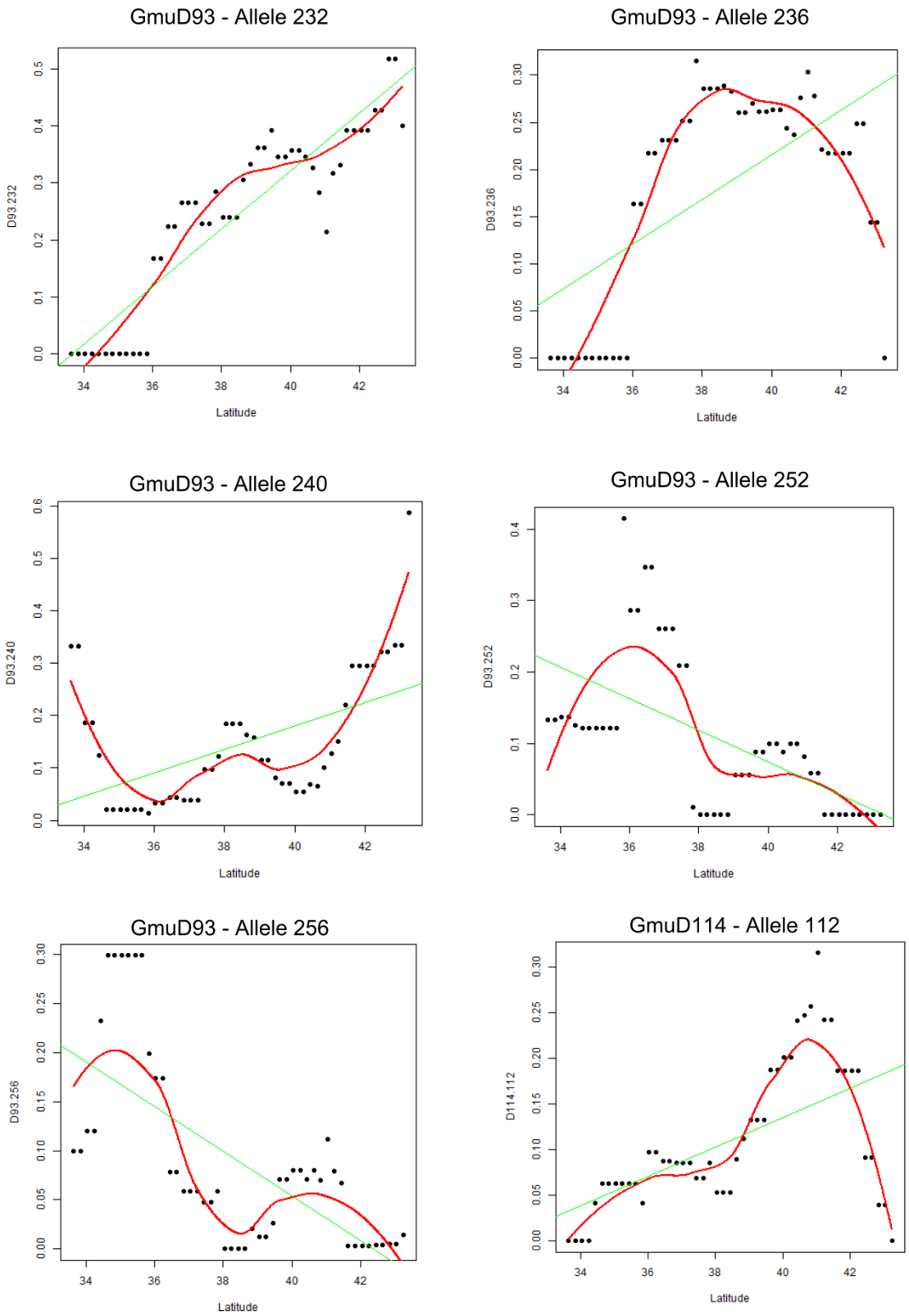

Figure 7.5: Allelic patterns of all alleles showing signs of allelic frequency clines.

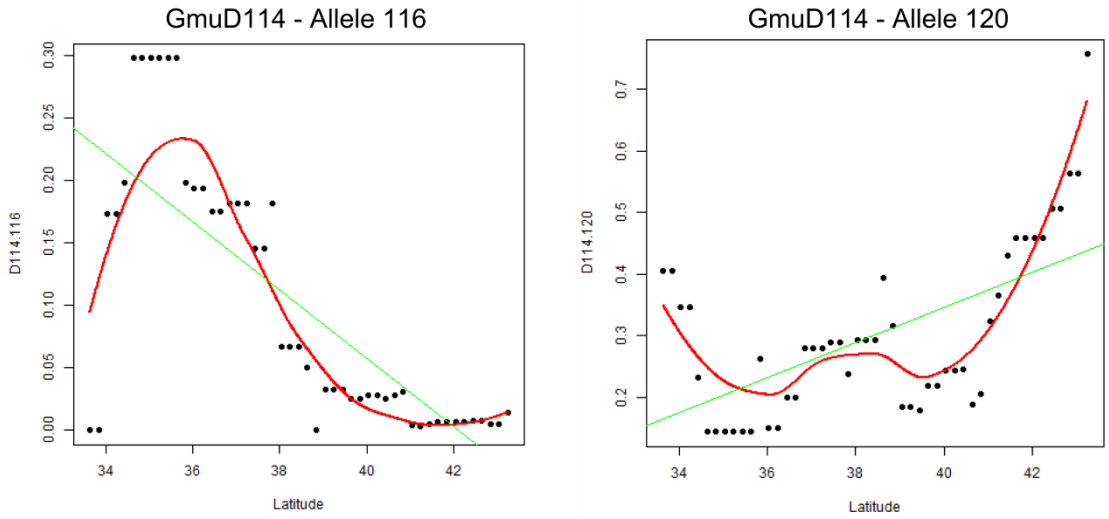

Figure 7.6: Allelic patterns of all alleles showing signs of allelic frequency clines.

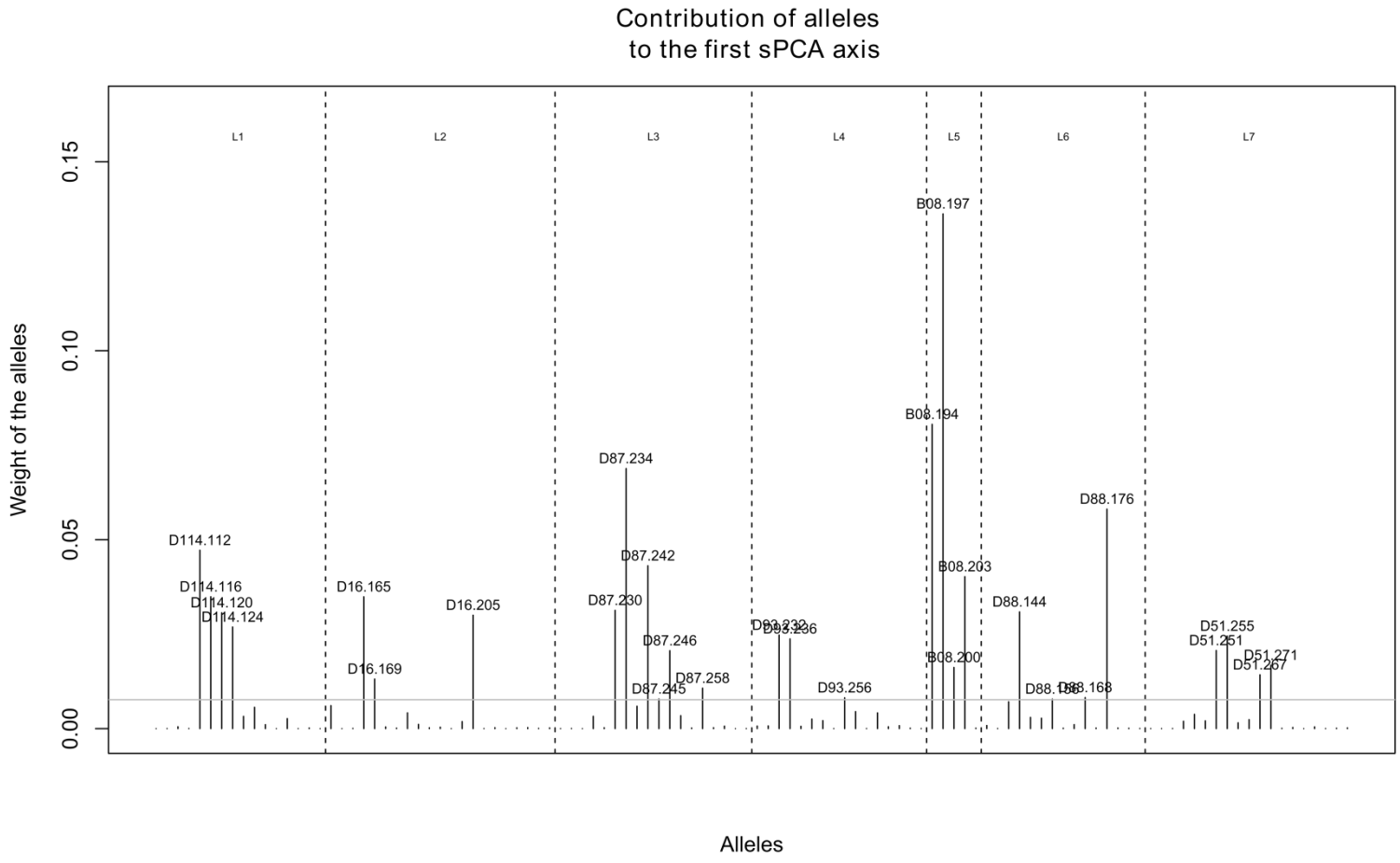

Figure 7.7: Most important alleles contributing to the patterns observed in the first eigenvalue of the sPCA.

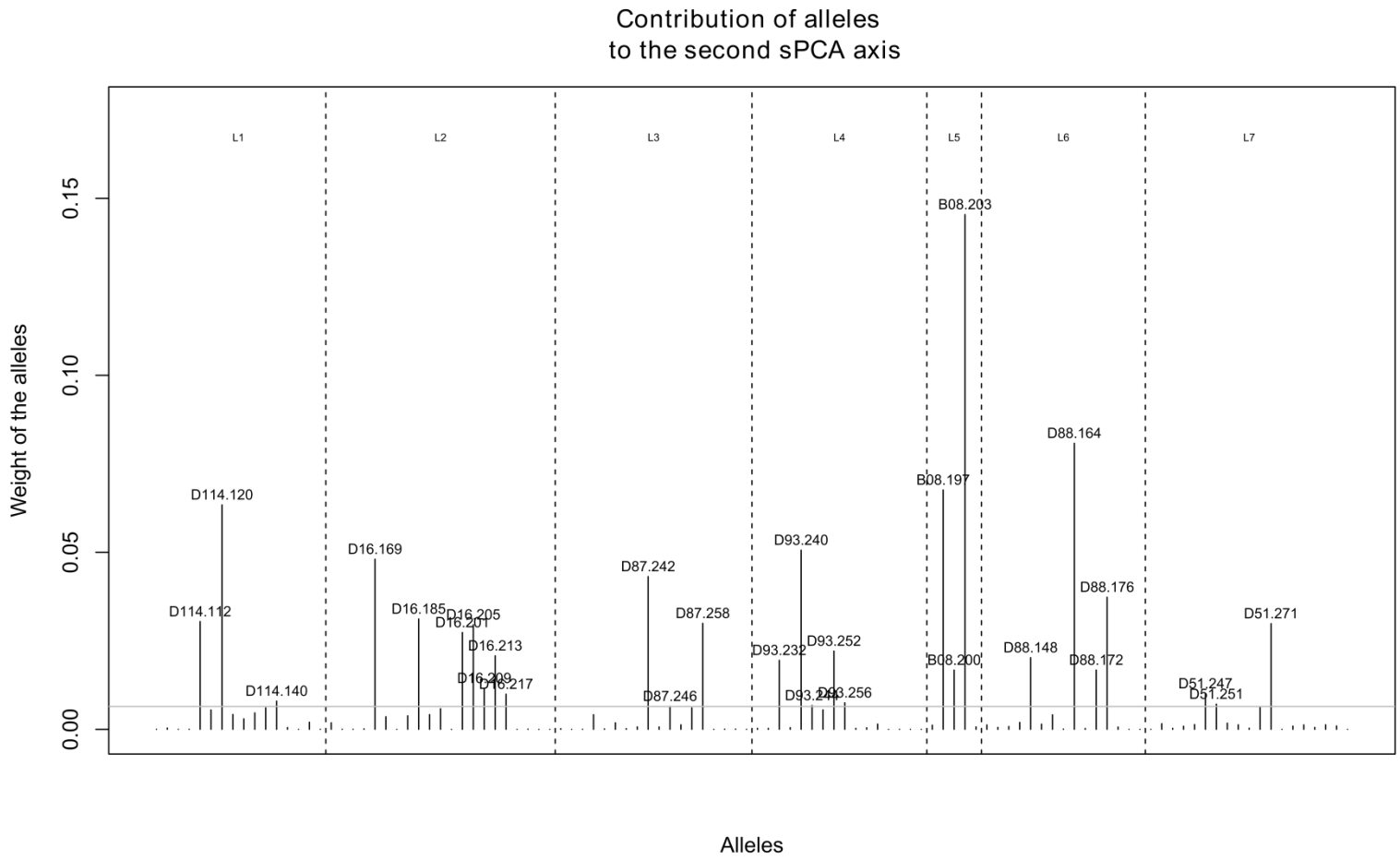

Figure 7.8: Most important alleles contributing to the patterns observed in the second eigenvalue of the sPCA