UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Spatial patterns of genetic diversity in *Hyla molleri*

Patrícia Cruz Gonçalves Guedes

Mestrado em Biologia da Conservação

Dissertação orientada por:
Doutora Ana Veríssimo
Doutor Carlos Fernandes

2019

Agradecimentos

À Ana, à Sílvia e ao Guillermo, os meus orientadores no CIBIO e os investigadores que tornaram possível o desenvolvimento desta tese. Obrigada por toda a paciência que demonstraram neste caminho cheio de curvas, e por todos os ensinamentos que me deram ao longo destes meses. Agradeço imenso a oportunidade que tive de expor o meu trabalho sob formato de poster no "Evolution" em Agosto passado, foi uma experiência incrível. Sinto que aprendi muito com esta tese, não só a nível de desenvolvimento de trabalhos científicos mas também sobre mim mesma e devo-vos isso.

Ao Professor Carlos, por ter sido sempre um orientador interno interessado no progresso da minha tese e por me ter ajudado em diversas ocasiões, mostrando-se sempre disponível.

Às quatro raparigas que foram a minha rocha neste percurso: a Carolina, a Patxó, a Sara e a Tânia. Sem vocês não sei o que tinha sido de mim ou desta tese. Obrigada por terem sido o melhor grupo de apoio que alguma vez poderia ter pedido, celebrando as minhas vitórias e ajudando incansavelmente nos momentos difíceis. Esta tese é minha, mas não a teria feito sem vocês. Não conheço palavras que expressem o quanto gosto de vocês e o impacto que têm na minha vida. Obrigada.

À Cata, por me ter acompanhado em todas as alturas desta viagem. És das melhores pessoas que conheço e agradeço muito estares na minha vida. Mereces uma medalha por me aturares.

Ao José e à Joana (Zé e Jú, vá), pelo apoio incondicional que têm demonstrado ao longo dos anos de amizade (e já lá vão muitos!), reforçado ao longo deste percurso. Sem o vosso incentivo e incansável fé em mim não sei onde estaria. A vida sem vocês não seria a mesma.

Por fim, quero agradecer à minha família, sem a qual nada disto teria sido possível. Não só me apoiam incondicionalmente, como acreditam em mim quando eu não o faço. Obrigada por tudo o que têm feito por mim. Não poderia ter família melhor.

Esta tese fez-me crescer muito. A nível profissional, claro, mas também a nível pessoal, permitindo-me compreender as minhas capacidades, força e limites. Mas acima de tudo mostrou-me que estou rodeada de pessoas incríveis e que estão presentes nos bons e maus momentos. Foi uma viagem acidentada, mas valeu a pena. Estou ansiosa pelo futuro.

Este trabalho foi financiado por Fundos FEDER através do Programa Operacional Factores de Competitividade – COMPETE e por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito do projeto PTDC/BIA-BIC/3545/2014.

Resumo alargado

A diminuição acentuada das populações de anfíbios a nível global é uma das maiores preocupações de conservação da Natureza da actualidade. De facto, a União Internacional para a Conservação da Natureza (UICN) classifica 41% das espécies de anfíbios como ameaçadas de extinção, o que torna este grupo num dos mais ameaçados do planeta. Apesar de se conhecer os factores de ameaça dos anfíbios (p.e. perda de habitat, alterações climáticas, quitridiomicose, etc.), ameaças como baixa diversidade genética, aumento da deriva genética e endogamia tem sido largamente ignoradas.

A diversidade genética é produto da diversidade alélica e genotípica encontrada na espécie e, de acordo com a UICN, é um dos três níveis de biodiversidade que necessita de medidas de conservação. Por ser a base do potencial evolutivo das espécies, a diversidade genética é fulcral para a capacidade que uma espécie tem de ultrapassar alterações no *fitness* populacional, actuando como uma salvaguarda no caso de redução significativa da população face a um evento catastrófico.

A diversidade genética pode ser classificada em duas categorias: adaptativa ou neutral. A primeira está sob a influência de forças selectivas e confere maior fitness nas condições em que a espécie se encontra; a segunda não está sob a influência de forças selectivas e permite inferências sobre processos demográficos. Em estudos de conservação, o foco tem sido sob a diversidade genética neutral, maioritariamente para a definição de Unidades de Conservação dentro das espécies.

De um ponto de vista genético, as populações de anfíbios têm reduzido efectivo populacional, que se traduz num baixo número de indivíduos reprodutores, e estão mais propensas a fenómenos de deriva genética e redução de diversidade genética. Várias características do habitat, tais como a elevação ou distância geográfica, podem também impactar a diversidade genética dos anfíbios, actuando como barreiras ao fluxo de genes devido aos gastos energéticos associados à deslocação entre locais. A baixa capacidade de dispersão aliada às necessidades de habitat (como elevada humidade atmosférica), limitam a conectividade entre populações e levam a um aumento da deriva genética. Adicionalmente, a história biogeográfica da espécie pode também desempenhar um papel importante na diversidade genética: em anfíbios, os padrões de diversidade genética aparentam ser mais fortemente moldados por acontecimentos históricos do que por fenómenos recentes, com a maioria dos *taxa* mostrando múltiplas linhagens de ADN mitocondrial estruturado geograficamente. Mais especificamente, foi o período de glaciações do Pleistoceno que profundamente alterou a distribuição das espécies, através de repetidos acontecimentos de extinções locais, concentração em áreas de refúgio, e nova expansão a partir desses locais, levando a uma forte estrutura genética e divergência entre populações.

A genética da paisagem invoca um design de amostragem focado em características da paisagem e no uso de ferramentas genéticas e estatísticas para identificar padrões de diversidade genética que possam ser explicados por uma ou várias características ambientais/da paisagem. A paisagem pode influenciar os padrões genéticos de uma espécie por principalmente dois processos: isolamento-por-distância e isolamento-por-ambiente. No isolamento-por-distância, é o aumento da distância geográfica que impulsiona a diferenciação genética, através da redução de fluxo genético entre as populações. No isolamento-por-ambiente são as diferenças ambientais entre locais que incitam a diferenciação genética, limitando o fluxo genético por meios de selecção natural, independentemente das distâncias geográficas.

A *Hyla molleri* é um endemismo ibérico, pertencente à família Hylidae, cuja origem e diversidade está localizada nos neotrópicos. Na Península Ibérica existem duas espécies de

Hyla: *H. molleri* e *Hyla meridionalis* (Boettger, 1874). A *H. molleri* é principalmente encontrada na parte centro e norte da Península Ibérica, e a *H. meridionalis* está maioritariamente presente no sul, sendo que a parte centro da península actua como zona de simpatria.

Apenas em 2008 é que um estudo conduzido por Stock *et. al* demonstrou que a população de *H. molleri* é, de facto, geneticamente diferenciada de *H. arborea*. Anteriormente a esta separação oficial de espécies, Rosa & Oliveira (1994) averiguaram a diferenciação entre *H. meridionalis* e *H. arborea molleri* (como era anteriormente conhecida *H. molleri*), e descobriram valores muito baixo de diversidade genética, sugerindo até que a população de *H. molleri* poderia ser vista como uma única população, na qual os acasalamentos se davam aleatoriamente, independentemente da distância geográfica.

Desde da separação oficial da *H. arborea*, foram poucos os estudos que se focaram na estrutura populacional e diversidade genética de *H. molleri.* Além disso, nenhum destes estudos conseguiu avaliar toda a distribuição da espécie ou teve uma amostragem que permitisse tirar conclusões robustas sobre os padrões espaciais de diversidade genética, assim como possíveis zonas de contacto, desta espécie. A falta de dados sobre a influência das distâncias geográficas e ambientais em anfíbios levam também a uma necessidade de explorar este tema.

Assim, os objectivos deste trabalho foram: 1) inferir os padrões de estrutura populacional ao longo de toda a distribuição da Península Ibérica e 2) analisar a influência de distâncias ambientais e geográficas na distribuição da diversidade genética de *H. molleri*. Para este efeito, foram utilizados marcadores moleculares do tipo Single Nucleotide Polymorphims (SNPs), uma vez que estes marcadores já mostraram ter grande poder de detecção de diferenciação genética, são mais abundantes no genoma, e apresentam melhor relação preço-eficácia em estudos deste tipo relativamente a marcadores como microsatélites nucleares.

Os resultados obtidos apontaram para um gradiente de diferenciação genética das populações de *H. molleri* do norte para o sul da Península Ibérica, assim como do centro da distribuição para a periferia. A presença de quatro populações ancestrais foi também identificada, contribuindo para a variabilidade na composição genética das populações atuais segundo o referido padrão de diferenciação. Tendo em conta os níveis de diversidade genética encontrados para outros anfíbios da Península Ibérica (com distribuição e hábitos semelhantes), os resultados mostraram baixa diversidade genética para esta espécie, com os valores mais altos localizados nas populações do sul. Aqui, tanto o número de alelos privados como valores de heterozigotia observada foram dos mais altos. Tanto a distância geográfica como a distância ambiental mostraram correlações positivas com a distância genética entre populações.

Os padrões observados sugerem que a *H. molleri* se refugiou principalmente no sul da Península Ibérica durante as glaciações do Pleistoceno, tendo desde então vindo a expandir a sua área de distribuição. No entanto, num estudo recente de Sánchez-Montes *et. al* (em impressão) usando microsatélites foram encontrados níveis mais elevados de diversidade genética no norte da Península Ibérica, sugerindo esta área como refúgio glaciar. O esforço de amostragem estará na origem destes resultados contraditórios: o nosso estudo incluiu 85 indivíduos de 27 localidades, enquanto que o de Sánchez-Montes *et. al* se baseou em 248 indivíduos provenientes de 60 localidades. Adicionalmente, ao contrário do presente estudo, Sánchez-Montes *et. al* incluíram na sua amostragem indivíduos de localidades do sul de França. Há ainda que referir o uso de diferentes marcadores genéticos entre os estudos: Sanchéz-Montes *et. al* recorreram a uma combinação de DNA mitocondrial e de microssatélites específicos para esta espécie, sendo que os microssatélites apresentam uma taxa de mutação mais elevada que os SNPs, o que pode resultar numa diversidade genética mais elevada. Assim, o maior esforço de amostragem por

parte de Sanchéz-Montes *et. al*, aliado ao uso de microssatélites, poderá explicar a diferença de resultados obtidos.

Apesar de termos obtido correlações significativas entre distâncias genéticas e distâncias geográficas e ambientais, é de ressalvar que o nosso modelo teve um coeficiente de determinação relativamente baixo, o que indica que há mais variáveis explicativas para a distância genética entre populações, que não foram tidas em contas neste trabalho.

Para melhorar futuros trabalhos recomendamos uma maior amostragem tanto em termos de locais amostrados como de indivíduos amostrados, acompanhada com uma maior resolução variáveis ambientais (1 x 1 Km, por exemplo) de forma a incluir possíveis microhabitats, e ainda a inclusão de um maior número de variáveis explicativas nos modelos de genética da paisagem, tais como: topografia, uso do território, massas de água e rodovias, um vez que outros estudos (tanto focados em *H. arborea* como em outros anfíbios já mostraram influência destas variáveis na dispersão dos mesmos).

Palavras-chave: *Hyla molleri*, Single Nucleotide Polymorphism, Diversidade genética, Estrutura populacional, Genética de paisagem

Summary

Global decline of amphibian population has become a major concern for the scientific and conservation communities. With 41% of amphibian species currently classified has threatened with extinction, amphibians are one of the most threatened group in the planet.

From a genetic point of view, amphibian populations generally have low effective size, which translates into a small amount of active breeders. Small populations are more prone to have low genetic diversity due to arbitrary genetic drift. Thus, being more likely to be affected by genetic drift and genetic diversity reduction. Habitat characteristics, such as elevation or geographical distance, can also impact amphibian genetic diversity by acting as a barrier to gene dispersal given the energy required to move between places. Landscape genetics encompass a sampling design focused in landscape characteristics and using a range of genetic and statistical tools to find patterns of genetic diversity that can be explained by one or several landscape/environmental characteristics

Hyla molleri is an Iberian endemism, belonging to the Hylidae family whose origin and diversity is located in the neotropics. It was only in 2008 that Stock *et al.* showed that the Iberian population was distinct from the rest of the European populations. Before this official separation from *H. arborea*, Rosa & Oliveira (1994) studied the genetic differentiation between *Hyla meridionalis* and "*H. arborea molleri*", and found very low values of genetic diversity, even suggesting that the samples could be perceived as the result of a single population with random mating, regardless of their distance.

Since the official separation of *H. molleri* from *H. arborea*, few studies have been conducted on the species' population structure and none of these studies has comprehensively studied the genetic population structure of the species across its entire range, or has had a sampling design that allowed for more robust conclusions. There is also a knowledge gap in environmental and geographical distance influence on genetic distances in this species. Therefore, this study aimed to: 1) Infer the spatial genetic population structure, across the whole range of *H. molleri*; 2) Analyse the influence of geographic and environmental distances on the distribution of the genetic diversity of *H. molleri*.

Our results point to a genetic differentiation gradient between individuals from northern and southern populations of the Iberian Peninsula, as well as a between individuals in the center and peripheral areas of the species distribution. This pattern is further corroborated by the uncover of four ancestral populations. Greater genetic diversity was found in southern populations. Genetic distance was positively correlated with both geographical and environmental distances.

This suggests that *H. molleri* took refuge in the southern part of the Iberian Peninsula and has expanded its range from there, with the northern range being the last to be occupied. However, in a recent study by Sánchez-Montes *et. al* (2018, unpublished) using microsatellites, higher genetic diversity was found in northern populations, suggesting that *H. molleri* glacial refugia were in fact located in the north part of Iberia. Sample size and genetic marker choice are the main suspects for these contradictory results.

We suggest more sampling (both more individuals and localities) and adding other explanatory variables (*e.g.* topography, land cover, hydrologic map, road traffic, *etc.*), which have been found to affect similar amphibians' distribution, in future works for a more complete analysis.

Key-words: *Hyla molleri*, Single Nucleotide Polymorphism, Genetic diversity, Population structure, Landscape genetic

INDEX

TABLE INDEX

FIGURE INDEX

1. INTRODUCTION

1.1 GLOBAL AMPHIBIAN DECLINE

Global decline of amphibian population has become a major concern for the scientific and conservation communities[1–4]. Although the threats are known (*e.g.* habitat loss, over-exploitation, agricultural expansion, invasive species, anthropogenic climate change, increased UV-radiation, road mortality and the harvest for human consumption[5,6]) the exact drivers of the declines are not fully understood[4]. Additionally, the International Union for Conservation of Nature (IUCN) Red List of Threatened Species shows that 41% of amphibian species are currently threatened with extinction. This places amphibians at a greater conservation concern compared to other vertebrates[6].

The main environmental factor shaping amphibian distributions is water availability, as it impacts both phases of amphibian life cycle due to desiccation risk[7]. Eggs and larvae have a mortality risk associated with the water depth of ponds (or other lentic water sources) not being enough to counterbalance water evaporation[7]. Additionally, reduced water availability also leads to smaller pond sizes, which will have many indirect effects, such as reduced food supply, alterations in tadpole density and smaller size at metamorphosis[7]. Adult amphibians, although not completely water dependent, can also be prone to desiccation if faced with conditions in which water loss through the respiratory system and skin is elevated[7]. A decrease in environmental moisture can also lead to limited periods of activity, diminished mobility, less capability to evade predators and a decrease in food supply[7]. Therefore, environmental conditions highly influence amphibian colonization capability by having a direct impact in amphibian distribution.

Possible shifts in local environment due to climate change can also alter the distribution and abundance of some amphibian species by turning previously unfit locations into habitable places and vice versa[8]. The causes of these behavioural and demographic modifications are a result of direct (*e.g.* changes in phenology, alterations to movement and new physiological stress) and indirect effects (*e.g.* different predators, competitors, habitat modifications and changes in food supply), which highly influence population dynamics[8].

In light of the various environmental alterations felt worldwide attributed to climate change, several studies have been focused on climate change predictions and the effects they would have on various *taxa*. For amphibians, studies indicate that climate changes might intensify population fragmentation, diminish distribution areas, increase extinction rates and cause multiple alterations in biotic interactions[9,10]. Therefore, there is a need to study amphibian population distribution with more detail and understand what factors are limiting and driving it.

1.2. THE IMPORTANCE OF GENETIC DIVERSITY FOR AMPHIBIAN CONSERVATION

Other threats which have been largely overlooked are genetic factors, including low genetic diversity, increased genetic drift and inbreeding[6]. Genetic diversity results from the diversity of alleles and genotypes found in species and, according to the IUCN[11], constitutes a level of biodiversity in need of conservation actions. By being the foundation of evolutionary potential, genetic diversity is crucial for a species' capacity to overcome changes in population fitness, acting as a safeguard in the event of unexpected and catastrophic events that can strongly

diminish population size[6]. Essentially, genetic diversity is what allows for the development of adaptive responses that species need for their long term survival[12].

Genetic diversity can be classified into two categories: adaptive or neutral[13]. The first is under selection and has an effect on individual fitness; the second is not under selection and provides insights into population dynamics and evolutionary forces (*i.e.* genetic drift, mutation, migration)[13]. Conservation biologists have been focusing on neutral diversity[13], mostly to define Conservation Units (CUs) within species[14], which are essential in the development and management of conservation efforts. Conservation priority is usually given to CUs that have genetic and ecological uniqueness (Evolutionarily Significant Units) and which populations are more influenced by their own dynamics than by immigration (Management Units)[14]. However, adaptive variants are also crucial for species' survival[15] as they improve the population fitness for different environmental conditions[16]. In fact, adaptive differences in relation to climate gradients have been found among populations of the same species[17].

From a genetic point of view, amphibian populations are generally small, which translates into a small amount of active breeders[6] and are more prone to have low genetic diversity due to arbitrary genetic drift[18]. Additionally, their reproductive output, reproductive success and mortality rates can vary greatly from one year to another due to their dependence on water availability[6]. This high sensitivity to external factors has lead Allentoft & O'Brien (2010) to hypothesize that in the event of severe predatory pressure or extreme heat (leading to desiccation), it is possible that only a few egg clutches survive leading to a population entirely constituted by siblings[6], thus showing how an amphibian population can be affected by inbreeding in just one reproductive cycle. Thus, amphibian populations are likely to have higher levels of homozygosity and low genetic diversity enhancing the risk of inbreeding depression in populations, which can result in lowered fitness[6], increasing extinction risk[18]. Additionally, the impact that genetic diversity loss has in populations is a long term impact, meaning that reversing its effects is more difficult[19]. Therefore, maintaining genetic diversity within species populations is vital for their survival.

Scientist tend to regard amphibian populations as metapopulations: during the reproductive season, most species of amphibians aggregate in ponds to mate and lay their eggs[8], with ponds differing greatly among each other (*e.g.* in diameter, depth, vegetation, coverage, etc.), granting each pond the status of a population[8]. In metapopulations, migrant exchange (*i.e*. gene flow) and colonization dynamics (*i.e.* local extinctions and recolonizations) ensure a balance between events of colonization and extinction[8]. However, if amphibian metapopulations become isolated (*e.g.* due to habitat fragmentation), gene flow will diminish and populations will be more affected by inbreeding, genetic drift and selection, therefore becoming more prone to loss of genetic diversity, reduced fitness, and higher extinction rate[6]. For instance, human-induced habitat fragmentation in the Lolland island populations of *H. arborea* (Linnaeus, 1758) led to population bottlenecks, with most populations suffering severe genetic diversity loss[7]. Some populations were even at risk of extinction due to high levels of inbreeding[7].

Habitat characteristics, such as elevation or geographical distance, can also impact amphibian genetic diversity by acting as a barrier to dispersal given the energy required to move between places[13,20]. The combination of low dispersal ability and specific habitat needs (*e.g*. high moisture levels) can also limit population connectivity and lead to high levels of genetic drift[6]. Moreover, the species' biogeographical history can also play a major role in genetic diversity[21]. Garner *et al*. (2004) verified this in a study that revealed that *Rana latastei* (Boulenger, 1879) genetic diversity followed an east-west gradient, caused by several founder events during the

species expansion from the Balkan area, which served as a glacial refugium[21]. Hence, understanding whether low genetic diversity is caused by human action, by biological features, biogeographic history or a combination of factors, is important from a conservation point of view since it will influence conservation actions[6].

Spatial patterns of genetic diversity in amphibian taxa appear to be strongly influenced by historical processes and less so by current events[18, 21], with most *taxa* presenting multiple geographically structured mitochondrial DNA lineages[21].

1.3. THE IBERIAN PENINSULA AND AMPHIBIAN DISTRIBUTION

The Iberian Peninsula is considered a biodiversity hotspot, because it holds a high number of endemic species. In addition, many species present high intra-specific diversity, which is often spatially structured[10]. Several factors have contributed to turning the Iberian Peninsula in a biodiversity hotspot, such as geological events (land connection with the African continent allowing for African species to colonize), geographical characteristics (large mountain ranges with an east-west orientation that allows for microclimates to develop, providing a refuge for populations when climate shifts, allowing for their survival[18, 21]), climatic influence (from Atlantic to Mediterranean and Desert climate) and climatic history (*e.g.* Pleistocene Ice Ages)[18,21]. The latter played an important role in the number of endemism's found in Iberia since this area functioned as an important glacial refugia[18].

In fact, the Pleistocene Ice Ages deeply altered Iberian species distribution through repeated local extinctions, dispersal to new locations, concentration in refugia and expansion from there, *etc*.[18], resulting in strong genetic structure and divergence among populations[21]. This genetic patterns across several species range point to several refugia within the Iberian Peninsula during the Pleistocene: a refugia-within-refugia hypothesis[21]. Suggested by Gómez & Lunt (2007), the refugia-within-refugia hypothesis amounts to seven refugia within the Iberian Peninsula, being the majority found in the south. Gómez & Lunt (2007) also compiled data showing a higher genetic diversity in the south of the Peninsula since southern populations appear to be more demographically stable[21]. However, not all refugia appears to have been adequate throughout the Ice Ages, meaning that the same species might have taken refuge in different refugia in different glaciations[21].

Except a few exceptions, Iberian amphibians were also more likely to have taken refuge in the south, more specifically in the Betic Range and in Central Portugal (Serra da Estrela)[21].

1.4. LANDSCAPE GENETICS

Landscape genetics encompass a sampling design focused in landscape characteristics and using a range of genetic and statistical tools to find patterns of genetic diversity that can be explained by one or several landscape/environmental characteristics[22]. Landscape can influence the genetic patterns of a species mainly through two processes: isolation-by-distance and isolation-by-environment[23]. In isolation-by-distance it is the increasing geographical distance and barriers that drive genetic differentiation due to reducing gene flow among populations[20]; in isolation-by-environment it is the environmental differences that push adaptive genetic differentiation limiting gene flow by means of natural selection, independently of geographical distances[23]. Landscape genetics tries to disentangle the effect of both of these processes to shed some light in the species' genetic patterns.[23]

With such goals in view, the first step in landscape genetics is to identify genetic patterns across the landscape and the second step is to associate those patterns with landscape composition[22]. In order to identify the genetic patterns, researchers must collect genetic data from as many individuals as possible and register the exact geographical location of the sampling[22]. Here, the individual is the preferable study unit, as it provides more detailed result[22]. Nevertheless, if enough populations are sampled, through the use of allele frequencies, each population can be the study unit[22].

Since the main threats to the focus species of this thesis are landscape features (that can lead to isolation), it was important to incorporate this analysis in an attempt to better understand which drivers can be affecting *H. molleri* population structure and genetic diversity.

1.5. SINGLE NUCLEOTIDE POLYMORPHISM (SNP's)

In this work we focused on using SNPs and DArTseq technology to study *H. molleri* (Bedriaga, 1890) population structure and genetic diversity.

When a mutation affects a single nucleotide position at a locus, creating an allele with an alternative basis it originates a SNP – Single Nucleotide Polymorphism-[24]. Therefore, SNPs can be considered as the final cause for genetic differences between individuals[24].

Despite microsatellites and simple sequence repeats being the most used genetic markers in genetic diversity studies[25], they are quickly being replaced by SNPs, for SNPs are more abundant and stable[26], amenable to automation, efficient and gradually more cost-effecient[25]. Additionally, the development of restriction site-associated DNA sequencing (RADseq) methods[27], has allowed for SNP development with simultaneous discovery and genotyping for non-model species[28]. RADseq methods create DNA libraries and are a fast, robust and cost-effective high-throughput method for genetic diversity and population structure analysis in non-model species[28]. Diversity Arrays Technology – DArT – is a RADseq method that first uses restriction enzymes to reduce genome complexity and then employs hybridization to microarrays to discover several hundred polymorphic loci across the entire genome, without requiring *a priori* information of the genome[29].This method actively selects portions of the genome with active gene, which is an advantage when working with species with large genomes (such as amphibians). Additionally,by determination of the most fitting method for complexity reduction, this technology is enhanced for both the organism and application chosen, providing several thousand markers at a relatively low cost per sample[30]

1.6. *HYLA MOLLERI* AS A CASE-STUDY SPECIES

Hyla molleri is an Iberian endemism, belonging to the Hylidae family whose origin and diversity is located in the neotropics[31, 32]. In fact, within the hylidae, only the genus *Hyla* extends into the Palearctic region, including four species groups: *arborea*, *cinerea*, *versicolor* and *eximia*[33], with the *arborea* group dispersing into Europe[34].

In the Iberian Peninsula, two species of *Hyla* can be found: *H. molleri* and *H. meridionalis* (Boettger, 1874)[33]. *Hyla molleri* is mainly found in the North and Central part of the Iberian Peninsula[44] and *H. meridionalis* in the Mediterranean coastal zone and the South[34], with the central part of the Peninsula acting as a sympatry zone where hybridization can occur, originating unfertile hybrids[34].

Hyla molleri was considered to be *H. arborea* or a subspecies of *H. arborea* (*H.a.molleri*) until recently[31]. The *H.a.molleri* designation was based in morphological criteria (such as body length and the length of the posterior limbs) which did not provide enough scientific support to be accepted as a separate species by the scientific community[34]. In 2008, Stock *et al.* showed that the Iberian population was distinct from the rest of the European populations[31].

The Spanish populations of *H. arborea* were classified as "Almost Threatened", prior to revision of its taxonomic status, due to population isolation in the south-eastern and south-western regions[35]. These populations may still be declining in the more arid regions due to loss of sites suitable for reproduction[35]. In Portugal, *H. molleri* has also several populations that appear to be isolated, such as those near the Douro and Minho rivers, Serra da Padrela and Alvão[36]. In addition, Rosa & Oliveira (1994) found *H. molleri* to have lower genetic diversity than expected in a genus already shown to have low genetic diversity[345].

The putative declining population sizes and increased isolation raise concerns regarding its long term survival and the need to critical evaluate the current conservation status.

Despite the above mentioned alarming signals of declining population sizes and connectivity, data regarding the species´ current population structure and levels of gene flow among populations are scarce.

Before the official separation from *H. arborea*, Rosa & Oliveira (1994) studied the genetic differentiation between *H. meridionalis* and "*H. arborea molleri*", and found very low values of genetic diversity, even suggesting that the samples could be perceived as the result of a single population with random mating, regardless of their distance[35].

Since the official separation of *H. molleri* from *H. arborea*, few studies have been conducted on the species' population structure. The following studies are, to the extent of my knowledge, the ones that have done so:

- Barth *et. al* (2011)[38]: In this study, researchers studied genetic diversity at mitochondrial genes in populations across the Iberian Peninsula. The sampling efforts were focused in Galicia, due to the combination of Mediterranean climate in the southeast and Atlantic climate in the north. Here preliminary data had indicated weak genetic differentiation in populations located in the northern coast of Galicia. The results showed: i) low mitochondrial differentiation of populations across the Iberian Peninsula; ii) no significant correlation between genetic distance and geographical distance; iii) weak genetic differentiation between populations located in the coastal area of Galicia and populations in central Spain; and iv) possible areas of admixture in inland Galicia and in northwestern Spain and northern Portugal. In light of these results, the authors concluded that there seems to be a considerable amount of gene flow or recent population expansion in Iberia. However, due to the uneven geographical sampling (several populations and multiple specimens per population in Galicia and smaller sample sizes from central Spain), the results might have lead to erroneous interpretation of population differentiation and isolation by distance. Yet, it is important to acknowledge the research developed by Gvozdik *et. al* (2015)[39], on speciation history and introgression of several European *Hyla* species, where the authors also found a distinct haplotype in Galicia. This suggests that this region might have played a role as a glacial refugium.

- Stock *et. al* (2012)[33]: Here the authors used mitochondrial and nuclear markers to define the range of three species of *Hyla*: *H. arborea, H. molleri* and *H. orientalis*. Based on

knowledge of anuran dispersal capability, they expected to find that altitude zones (*e.g.* mountains such as the Alps, Pyrenees and Carpathians), are important barriers to gene flow, keeping gene flow restricted to low altitude areas. They also expected to find varying amounts of geographic genetic structuring among *H. molleri*'s distribution range, with lower diversity in the northern regions and higher endemism in the southern ones (due to more stable climate during the last glaciation). The results showed: i) low genetic structure within species; ii) range overlap and hybridization of *H. molleri* and *H. arborea* in the southwest of France, and iii) little mtDNA diversity throughout the *H. molleri* range. However, this study had limited sampling for *H. molleri* (*i.e.* only 37 out of the 462 individuals sampled-, and only one locality was sampled in France *versus* two localities in northern Spain and the rest in central Spain. Therefore, this sampling may have biased the results, which may have led to unreliable conclusions.

- Moreira, C. (2012)[40]: While studying from a molecular and bioacoustics approach the populations of *H. molleri* and *H. meridionalis* in Portugal, this researcher found two divergent groups of mtDNA haplotypes in *H. molleri*, namely on group occurring in sites located south of the Mondego River and another group occurring in sites north of the Mondego River and northwest Spain. Additionally, within each group, high levels of haplotype diversity were detected, indicating a high level of genetic diversity, contrary to previous studies.

- Sánchez-Montes *et. al* (2019)[41]: Through the use of mtDNA and microsatellites specific for this species, the authors reconstructed the historical biogeography of *H. molleri*. Sampling included 248 individuals from 60 localities, covering the species whole distribution. Their results showed 1) higher genetic diversity in the northern Iberian mountains and western areas, 2) a concentration of private alleles in the extremes of this species distribution and 3) genetic structure was better explain when K=4 or K=7.

So, despite the important insights into *H. molleri* population structure these studies provided, with exception to Sánchez-Montes *et. al* (2019), none of these studies has comprehensively studied the genetic population structure of the species across its entire range, or has had a sampling design that allowed for more robust conclusions. Additionally, these studies have been using mtDNA, microsatellites or a combination of both as their chosen genetic marker. Therefore, there is a need for further studies regarding this species' spatial genetic diversity patterns.

Environmental and geographical distance influence on genetic distances is also unclear for amphibians. Species such as *Alytes obstetricans* have suggested that species-characteristic genetic diversity drivers are the main factor the spatial patterns observed and not the environment itself[42]. However, Reino *et. al* (2017)[43]found that this species was more prone to be present in areas with time-concentrated precipitation. Therefore, research on environmental influence in *H. molleri* genetic distance is required. As for geographical distance influence in genetic distance, a positive correlation between geographical and genetic distances for *H. molleri* has been found by Barth *et. al* (2011), and Reino *et. al* (2017) found that this species was more abundant in areas with lower slope. This points to geographical distance as a possible barrier to dispersal, meaning that further analysis should be conducted on *H. molleri*.

2. OBJECTIVES

The objectives of this thesis are:

1) Infer the spatial genetic population structure, across the whole range of *H. molleri*;

2) Analyse the influence of geographic and environmental distances on the distribution of the genetic diversity of *H. molleri*.

These objectives will be accomplished by sampling tens of individuals across the species' range and using of DArTseq technology for simultaneous calling and genotyping of several thousands of SNPs distributed across the genome

3. METHODS

3.1 STUDY AREA

The study area included the whole distribution range of *H. molleri* in the Iberian Peninsula. The Iberian Peninsula is located in the southwest corner of Europe and it includes Portugal and Spain's continental territories, as well as Andorra, Gibraltar and a small portion of French territory in the north-eastern part. It is mainly influenced by two types of climate: Mediterranean climate – the most influential climate due to the influence of the Mediterranean Sea, which is predominant in the southern part of the Peninsula-, and is characterized by very dry summers and high precipitation during the winter; and Atlantic climate –which predominates in the north and northwest of the Peninsula, as well as in the major mountain systems, which is characterized by having cool temperatures year round, with little oscillation in the annual temperature range. *H. molleri* distribution does not include the whole Iberian Peninsula as this species mainly occurs in the North, Center and Western part of the Peninsula[34](Figure 3.1).

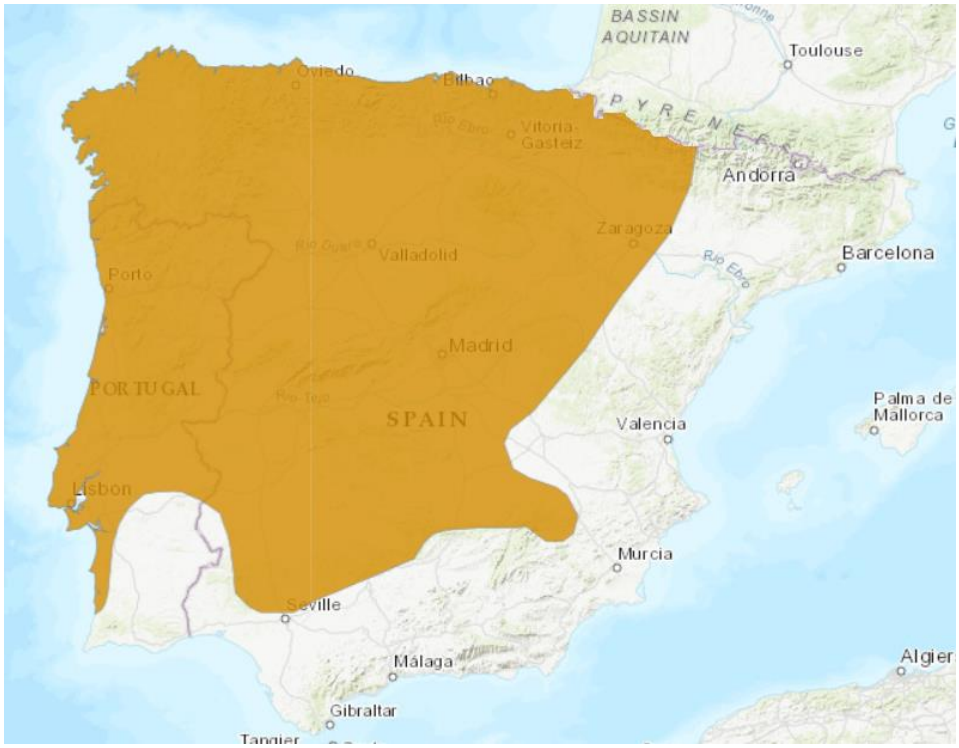


Figure 3.1- *Hyla molleri* distribution range and climate influence on the Iberian Peninsula. Map adapted from the IUCN List of Threatened species.

3.2. ENVIRONMENTAL STRATIFICATION FOR TISSUE SAMPLING

Sampling sites for tissue collection were chosen in order to cover the whole species' distribution range and the whole climatic variability throughout the species range. This strategy was selected to allow testing for the effect of environmental heterogeneity on the distribution of genetic diversity of *H. molleri*.

Species' distribution records were obtained from the Portuguese and Spanish atlases of amphibians and reptiles[35,36] and from the database of the Spanish Herpetological Society (http://siare.herpetologica.es/). Both atlases are referenced to a 10x10 km resolution UTM grid.

To ensure coverage of the whole climatic variability, the species' range was assessed for its spatial heterogeneity in a set of climatic variables which were subjected to a Principal Component Analysis (PCA) to reduce data dimensionality, followed by a model-based clustering method to identify the most likely number of climatic clusters.

Twenty variables were retrieved from the WorldClim database[44] covering the species' range (at 2.5 minutes resolution). The variables included annual mean temperature, mean diurnal range, isothermality, temperature seasonality, maximum temperature of warmest month, minimum temperature of coldest month, temperature annual range, mean temperature of wettest quarter, mean temperature of driest quarter, mean temperature of warmest quarter, mean temperature of coldest quarter, annual precipitation, precipitation of wettest month, precipitation of driest month, precipitation seasonality, precipitation of wettest quarter, precipitation of driest quarter, precipitation of warmest quarter, precipitation of coldest quarter and altitude.

The first three principal components (PCs) were obtained from the PCA, using the *princomp* function in R, and used to estimate the most likely number of environmental clusters throughout the species' distribution using model-based clustering implemented in the R package *mclust*[45]. The *mclust* package uses an Expectation-Maximization (EM) algorithm (which finds the maximum likelihood parameters) to perform the clustering analysis, ensuring that each cluster includes locations with similar climatic conditions. We ran *Mclust* with G = 1:16, being G the number of clusters for which the Bayesian Information Criterion (BIC) is calculated. The BIC is an index used to compare two or more alternative models, by valuing model fitness and reduced model complexity. The model with the lowest BIC is considered the best[45]. For the clusters identified in the best model, we analyzed similarity between clusters using the mahalanobis distance. This distance was calculated based on the first 3 PCs using the *pairwise.mahalanobis function* from the *HDMD* R package[46]. Mahalanobis distances were then used to perform a hierarchical cluster analysis using the complete linkage method (function *hclust* in R).

In addition to this, we also identified the variables with highest loadings for each PC and created raster layers for each PC for spatial visualization of the corresponding patterns. We also analyzed each climatic variable value distribution for each climatic clusters using boxplots.

3.3. FIELDWORK

Sampling included 2-3 different locations per environmental cluster and 4-5 individuals per location. This sampling design was meant to ensure representation in the final dataset, and guarantee sufficient genetic data per cluster. Tissue collection was carried in the late Spring/early Summer of 2017 and included tadpoles (*i.e.* tip of the tail) and adults (*i.e.* first phalange of one of the posterior members). We registered the corresponding GPS coordinates of each sampling site. Sampling gaps were filled with museum samples collected (between 2013 and 2015) by Iñigo Martinez-Solano).

3.4. MOLECULAR DATA COLLECTION

3.4.1. DNA extraction

We conducted genomic DNA (gDNA) extractions at CIBIO-InBIO laboratory, using the *ExtractMe* Genomic DNA 96-Well kit (DNA GDAŃSK) and QIAamp DNA Micro Kit (QIAGEN GmbtH), depending on the amount of tissue sample available. We followed the manufacturer's instructions in all extraction procedures. We used agarose gel electrophoresis (0.5% w/v) run at 300 V in 0.5X TAE buffer to assess the extracted gDNA quality and quantity, and used PicoGreen™ fluorometry in VICTOR™ (Perkin Elmer) to determine its concentration.

3.4.2. SNP development in *Hyla molleri*

We calculated the quantity of gDNA needed from each extraction so that every sample had 500 ng of gDNA. The samples were evaporated and reconstructed with 10 µL of pure water so that every sample had equal concentration: 50 ng gDNA/ µL. This step maximizes the probability of equal representation of reads across samples, and thus of the number of SNPs called and genotyped per individual. The gDNA samples were sent to Diversity Arrays Technology Pty Ltd, where simultaneous SNP calling and genotyping was done using proprietary DArTseq technology (Diversity Arrays Technology)[44].

3.5. RAW DATA TREATMENT

DArTseq output presented genotypes coded with "0", "1", "2" or "-", indicating whether the individual was a homozygote for the reference allele, a homozygote for the alternative allele, an heterozygote or whether the genotype was missing, respectively. To facilitate data analyses in the Rstudio environment[46], we transformed the raw data matrix into a *genind* object (which stores individual genotypes) using the function *df2genind* from the *adegene*t package, to allow computations using several packages, including *adegenet*[48] and *poppr*[49]. For that purpose the raw matrix was transposed so that individuals were in rows and loci in columns, and replaced the genotype codes "0", "1", "2" and "-" with "AA", "TA" ,"TT" and "NA", respectively.

We performed a preliminary analysis to assess how many SNPs had a call rate higher than 0.8, *i.e.* with genotype informations found in at least 80% of sampled individuals, which corresponds to SNPs with less than 20% missing data.

We then performed a Principal Coordinate Analysis (PCoA) on the genotype matrix. The PCoA allows data exploring and visual representation based on genetic distances among data points[50]. The individual genotypes were used to estimate the proportion of shared alleles (function *propShared* from *adegenet* package[48]), and this individual genetic distance was used to perform the PCoAs as implemented in the *dudi.pco* function of *ade4* package. We first used PCoA on the whole data set (no loci or individuals removed) to detect putative outlier individuals, and then on several datasets with different cut-offs for missing data at the locus- (*i.e.* loci missing genotypes at some samples) as well as at the individual-level (*i.e.* individuals missing genotypes at some loci), obtained through the *missingno* function (*poppr* package).

Upon removal of putative outlier individuals, loci identified as monomorphic were removed.

3.6. GENETIC DIVERSITY

In order to understand how genetic diversity is distributed in *H. molleri* we calculated two different genetic diversity indices: average number of private alleles per locality and observed heterozygosity. These metrics allowed having different perspectives on genetic diversity and hence, have a better understanding of genetic similarities and differences among sample localities.

Private alleles are unique alleles found in each population, which provides a good, yet easily interpretable, measure of genetic differentiation among populations[51]. If gene flow is high among populations, the fixation of distinct alleles is difficult, meaning this population will show lower values of private alleles[51]. To calculate the number of private alleles, we used the *private_alleles* function (*poppr* package[49]).To avoid biased results due to differences in the number of individuals sampled in each locality, we divided the number of private alleles per locality by the number of sampled individuals in each locality.

We calculated the observed heterozygosity for each locality of the whole dataset, using the summary function of the *genind* object as implemented in adegenet.

3.7. DETECTION OF PUTATIVE NON-NEUTRAL LOCI

In order to assess the distribution of neutral vs. adaptive diversity in *H. molleri*, we used the Moran Eigenvector Maps (MEM) approach described in Wagner *et. al* (2017) to detect putative outlier loci[52]. This approach relies on the assumption that adaptive loci behave as outliers, by showing a distinct spatial signature compared to neutral loci, due to the result of selection rather than of gene flow[52].

First, we used a Gabriel Graph to obtain a neighbour network and obtained a spatial weights matrix, which is needed to calculate the MEM axes. In the next step we obtained the power spectrum for each locus by calculating the squared of the correlations of each MEM axis with a matrix of allelic frequencies. The power spectrum shows the amount of variance that each locus has linked with each MEM. The average power spectrum across loci is then subtracted from each locus to obtain the standardized z-scores (*i.e.* standard deviation from the mean of each locus). We then used three cutoff values to identify outliers: 0.05, 0.01 and 0.001.

3.8. POPULATION STRUCTURE ANALYSIS

3.8.1. Spatial Analysis of Principal Components (sPCA)

We conducted a sPCA to identify the spatial genetic pattern within our sampling area and provide further insights about the species' genetic population structure[50].

The sPCA aims at finding independent synthetic variables (the principal components) which optimize the product of genetic variance and spatial autocorrelation of the haplotype frequencies of multiple loci. The spatial autocorrelation is measured using the Moran's I statistic based on the samples' geographical position and their allelic frequencies[50]. When allelic frequencies at neighboring sites are more similar than expected at random, there is positive spatial autocorrelation (*i.e.* global structure)[50]. If allelic frequencies are more genetically distinct than randomly expected, the spatial autocorrelation is negative (*i.e.* local structure) [50]. Since the variance of allelic frequencies term is always positive, the signal (positive or negative) of the obtained sPCA eigenvalues define if the spatial autocorrelation is positive (global structures) or negative (local structures).

We used the spca function implemented in the *adegenet* package with the Delaunay's triangulation as the connection network to establish neighboring sites, since this method requires that if a circle is drawn through three nodes it ensures that any point on the surface is as close as possible to a node.

We performed Monte Carlo simulations to statistically test the significance of observed global and local structures (*global.rtest* and *local.rtest* functions), using 999 permutations.

We evaluated which eigenvalues should be further analysed - based on whether they contained enough variability and spatial structure - through the *screeplot* function, in which the eigenvalues of the sPCA (λ_χ) are represented according to their variance and Moran's I value. Only the eigenvalues that show the highest variance and spatial autocorrelation should be used[50]. We then plotted a two dimensional scatter plot based on each individual score on the selected eigenvalues, onto an Iberian Peninsula map (*s.value* function from the *adegraphics* package), to visualize the genetic differences/similarities amongst individuals.

3.8.2. Sparse Non-Negative Matrix Factorization (snmf) analysis

In order to gain further insights into *H. molleri* current population structure, we used a snmf analysis which provides an independent perspective of genetic differentiation and allows comparing results among analytical approaches.

We used the snmf function of the *LEA* package[54], that uses sparse Non-Negative Matrix Factorization algorithms, to estimate individual ancestry coefficients and ancestral allele frequencies[54]. Sparse Non-Negative Matrix Factorization is an unsupervised statistical method, meaning that it uses likelihood methods to infer ancestral gene pools instead of predefined populations[55]. This algorithms reduces data dimensionality and allows for hidden data structure to become known[54].

We chose this approach in place of other more often used methods, such as STRUCTURE software[55-59], because snmf is faster[55], allows for more efficient data exploring, is more suitable for large datasets with many loci, and the choice of number of genetic clusters is based on a cross-validation criterion, which may be more reliable than those used in other methods[53]. In addition, the output has a very easy interpretation as the percentage of each ancestry lineage found in each individual can be displayed using a barplot.

To use this function we had to create a *geno* object. The *geno* object is a file format that stores the data with one row for each SNP[53] For the number of ancestral populations for which the snmf algorithm estimates have to be calculated (K)[53], we choose from 1 to 14, given that there were 14 environmental clusters and the sampling strategy aimed at collecting individuals on a per-cluster basis. We ran the snmf function with 10 repetitions for each value of K with 1 000 iterations and for 3 alpha values (being alpha the regularization parameter, whose value can alter the results): 10, 100 and 500, to assess congruence of results. We then calculated the corresponding cross-entropy criterion for each run. This criterion assesses the fit of a model with K populations: a smaller value of cross-entropy equals a better prediction capability for that run[53].

We used the most likely K value, *i.e.* with lower cross-entropy values, in the cross.entropy function to create a Q-matrix (*i.e.* individual admixture coefficient matrix) and the resulting barplots. For easier visualization of the spatial distribution of individual genetic admixtures, we used *QGIS* software (2.14.20 version) to plot individual pie charts on a map.

3.9. LANDSCAPE GENETICS

For the landscape genetics analysis we tested IBD and IBE simultaneously by performing a multiple matrix regression with randomization (MMRR) analysis on matrices of genetic (response variable), geographical and environmental distances (explanatory variables)[61]. This analysis allowed us to test if spatial distances and environmental heterogeneity had an effect on the genetic distribution of *H. molleri*. The MMRR output is a multiple regression equation, that tests if the dependent variable changes with respect to the different independent variables[61], and if so, how is that change. Thus, the regression coefficients quantify how the genetic distances respond to variation in environmental and geographical distances; the coefficient of determination evaluates the overall fit of the model (R^2) and the *p-value* allows to infer statistical significance of the coefficients[61].

Since this analysis runs at the population level, we used pairwise F_{ST} as the genetic distance. As for the environmental distance, since the sampling approach adopted for this study aimed to include individuals from the 14 environmental clusters defined using a PCA (see above for details), the loadings of the first 3 PCs of each locality were used to estimate multivaried euclidean environmental distance among pairs of localities. Finally, the geographical euclidean distances (in kilometres) were calculated using the *GeoDistanceInMetresMatrix* function, which uses geographical coordinates to calculate the euclidean distance between two points.

Because F_{ST} was used as the genetic distance, only populations with more than one individual were used in this analysis. The pairwise F_{ST} values were estimated using the *pairwise.fst* function implemented in the *hierfstat* package[62].

Prior to implementing the MMRR we assessed the data for linearity by fitting linear models of the genetic distance (dependent variable) in function of the geographical distance (independent variable). We also scaled and centered the geographical distance matrix, to reduce bias introduced by different absolute values among distance matrices.

The MMRR was performed with the *lgrMMRR* function as implemented in the *PopGenReport* package[63], with 999 permutations to assess statistical significance.

4. RESULTS

4.1. ENVIRONMENTAL STRATIFICATION FOR SAMPLE COLLECTION

The first three PCs explained 88.6% of the variance in the climatic data set (Figure 4.1).

Table 4.1 - List of the most important climatic variables for each of the first three PCs. Loadings of each variable in parentheses.

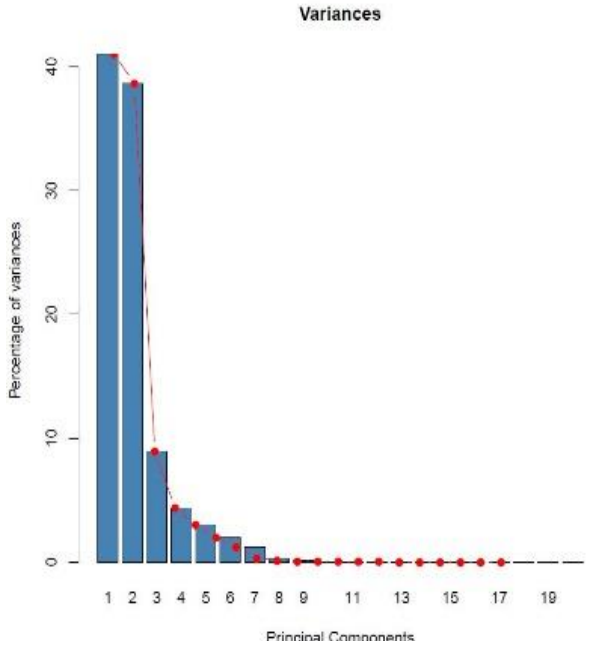


Figure 4.1- Percentage of variance explained by each PC.

PC	Most important variables
PC 1	Mean Temperature of Warmest Quarter (0.33)
	Precipitation of Driest Quarter (-0.32)
	Precipitation of Warmest Quarter (-0.32)
	Max Temperature of Warmest Month (0.31)
PC 2	Temperature Annual Range (-0.32)
	Temperature Seasonality (-0.32)
	Precipitation of Coldest Quarter (0.31)
	Mean Diurnal Range (-0.31)
PC 3	Mean Temperature of Wettest Quarter (0.47)
	Isothermality (0.38)
	Precipitation of Coldest Quarter (-0.32)

The loadings show how each variable contributes and behaves in each PC: when the variable signal is positive it means that when PC values increase so do the values of the variable; if the signal is negative it means that the variable value decreases as PC value increase. Taking that into account, PC 1 will have higher values where the mean temperature of the warmest quarter and the maximum temperature of the warmest month are higher and where the precipitation of the driest quarter and of the warmest quarter are lower, thus characterizing a warmer and drier

climate, as it is expected in areas with the Mediterranean climate. When observing the spatial distribution of the loadings (Figure 4.2a), PC 1 shows a north-south gradient, thus separating the two most felt climates in the Iberian Peninsula: the Atlantic climate, in the north, and the Mediterranean climate, in the south.

As for PC 2, the loadings indicate that this PC will have higher values where temperature annual range, temperature seasonality and mean diurnal range are lower and precipitation of coldest quarter is higher. This characterizes a climate with few fluctuations in temperature range and high precipitation during the winter. Figure 4.2b, shows that this PC has higher values in the coastal areas of the species range, indicating a separation between inland and coastal climates.

Finally, for PC 3, the loadings display higher positive values for the mean temperature of the wettest quarter and isothermality and a negative value for the precipitation of the coldest quarter. This implies higher values for areas where temperatures are more constant, the rainy season has higher temperature and winters are drier. Figure 4.2c shows that PC 3 displays a periphery-center gradient with higher values in the periphery of the species range.

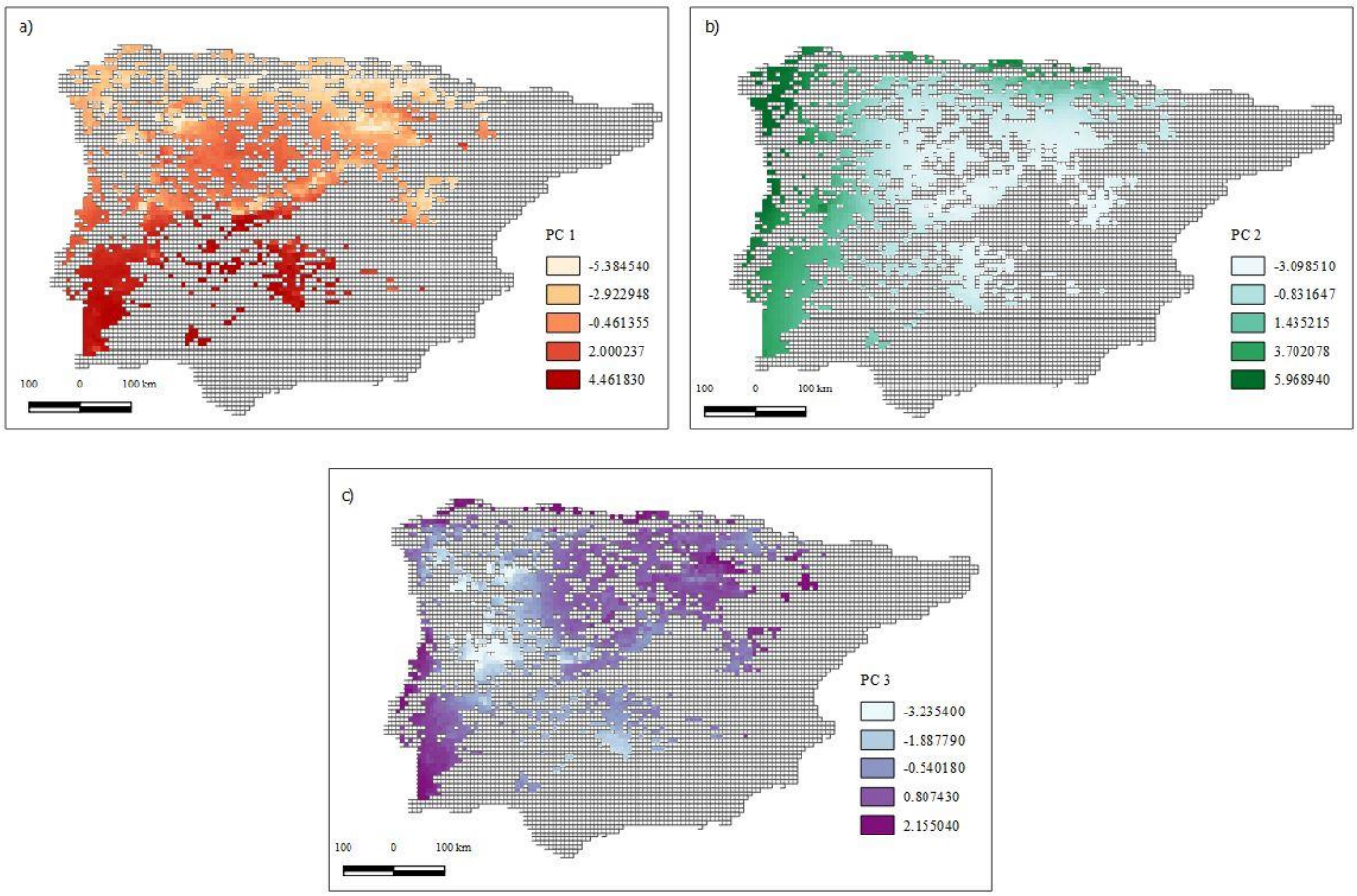


Figure 4.2 - First three PCs values across the species distribution. Values for PC 1 are represented in a); values for PC 2 are represented in b) and for PC 3 the values are represented in c). The colour gradient represents the range of values for each PC, being lighter colours lower values and darker colours higher values.

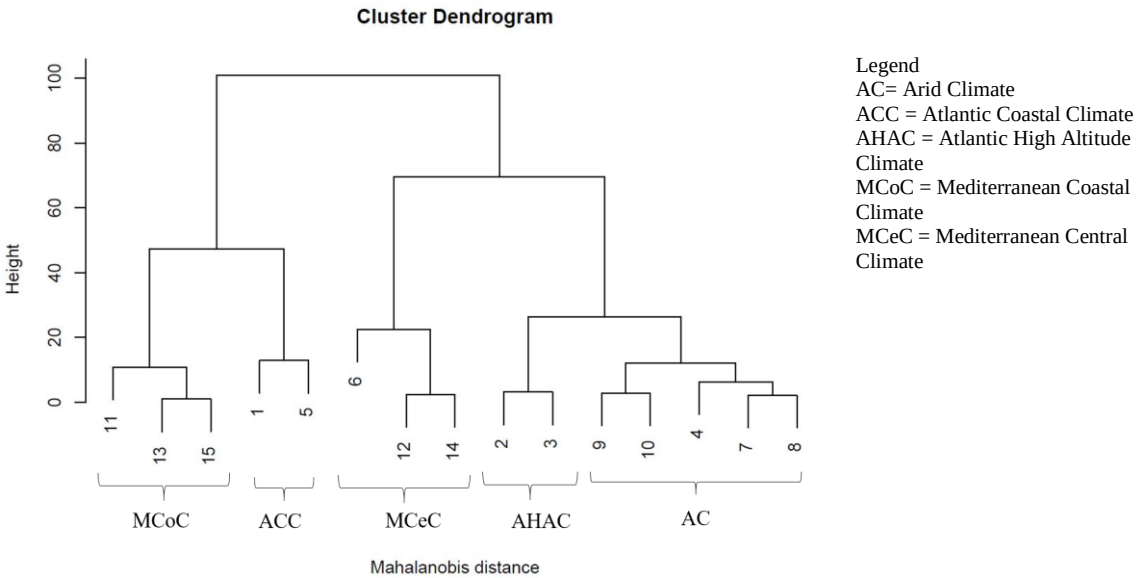


Figure 4.3 - Cluster Dendogram based on Mahalabonis distances showing relationships among climatic clusters

The Model-based clustering estimated 15 climatic clusters within the specie's range (Figure 4.3). Based on the hierarchical cluster analysis, five main climatic areas where defined by their Mahalanobis distances proximity, hereon referred to as Atlantic Coastal Climate (ACC, clusters 1 and 5); Atlantic High Altitude Climate (AHAC, clusters 2-3); Arid Climate (AC, clusters 4, 7-10); Mediterranean Central Climate (MCeC, clusters 6, 12 and 14), and Mediterranean Coastal Climate (MCoC, clusters , 11, 13 and 15) (Figure 4.3 and 4.4), due to their geographical position and climatic influence. Detailed characterization of each climatic cluster and main climatic areas are shown in Appendix Figures 1-20, Appendices.

The Atlantic Coastal Climate is characterized by temperatures between 15-22ºC, with low seasonality and high amounts of precipitation throughout the year. Cluster 1 has a smaller temperature range and higher precipitation during the warmest quarter than cluster 5, as well as higher precipitation during the driest month.

The Atlantic High Altitude Climate also shows high precipitation, but lower temperatures. Cluster 3 has lower precipitation during both the warmest quarter and the driest month and shows lower temperatures during the wettest quarter than cluster 2.

The Arid climate shows a high annual temperature range and low precipitation during the whole year. Cluster 7 shows higher precipitation in warmest quarter, than cluster 10 and 9. Cluster 9 also shows the highest precipitation for the coldest quarter.

As for the Mediterranean Central Climate, it is characterized by high temperatures during the warmest quarter, very little precipitation, and a high annual temperature range. Cluster 14 shows a bigger temperature range, having high temperatures during the warmest quarter and low temperatures during the wettest quarter. Cluster 12 appears to be drier, having lower precipitation during the coldest quarter.

Finally, the Mediterranean Coastal Climate shows high annual temperature range with low precipitation. All clusters within this climatic area were mainly uniform in their characteristics, except for cluster 6 which had higher ranges in all variables and showed higher annual temperature range.

4.2. FIELDWORK

A total of 34 localities were sampled throughout the species range, including 14 of the 15 climatic clusters (Figure 4.4). However, specimens of *H. molleri* were obtained from only 27 of the 34 sampled locations. Cluster 4 was not sampled due to its more remote geographical location and sampling time limitations. However, this cluster covers a relatively small area, it is quite similar to clusters 9 and 10, and several localities were sampled in the neighbouring clusters. It should also be noted that samples from cluster 10 (belonging to the Arid Climate Region, Figure 4.3) refer to the location of Villa Verde (southeast Spain), which is very distant from the area where the majority of the grid cells of this cluster are located (*i.e.* north-central Iberia, Figure 4.4).

One hundred individuals were genotyped, of which 11 were sampled during field work in 2018 (cluster 6 and cluster 15), and the remainder were obtained from the tissue collection at Spanish National Museum of Natural Sciences (Museo Nacional de Ciencias Naturales).

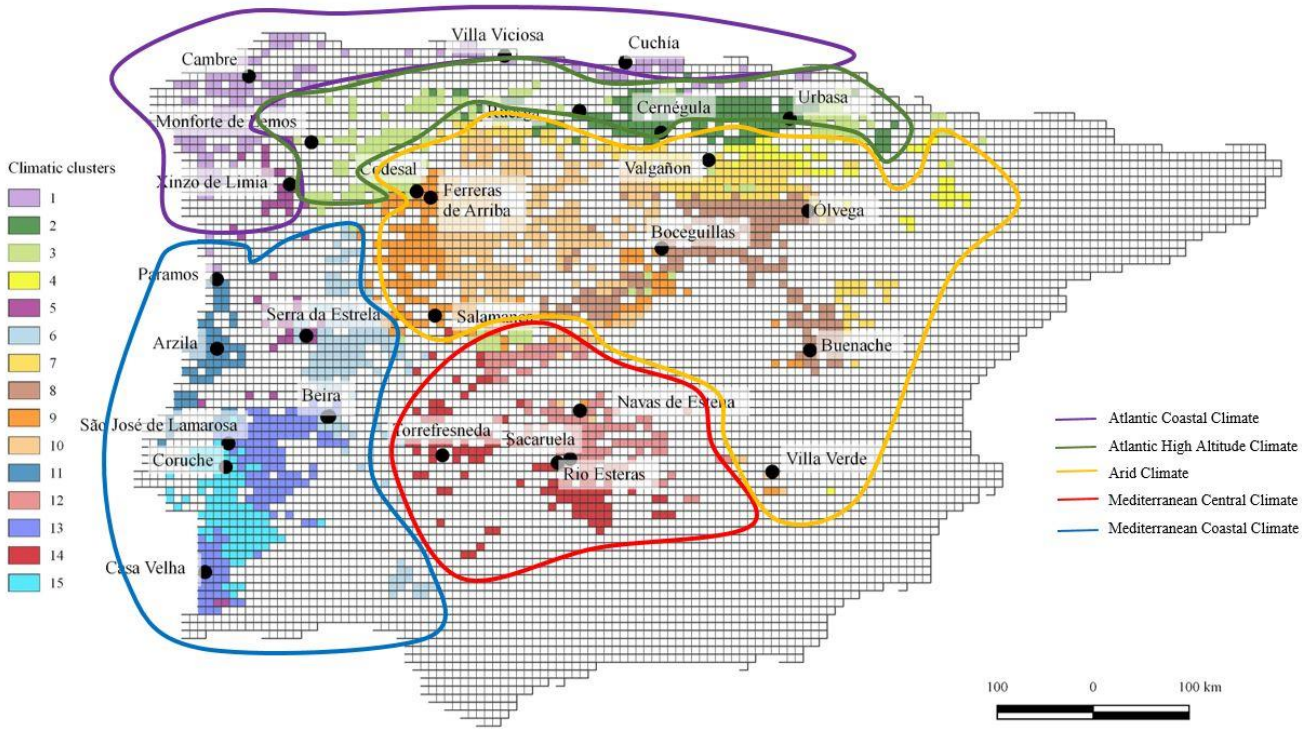


Figure 4.4 - Sampling locations within each of the climatic clusters

The number of sampling locations per environmental cluster varied between one and four, and the number of individuals per location varied between one and five (Table 4.2).

Table 4.2- Summary table of total sampled locations, total number of samples, number of samples genotyped and number of samples used in the final dataset.

Cluster	Location	Total number of samples	Nr of samples genotyped	Nr of samples used in the final dataset
1	Cuchía	5	5	4
	Villa Viciosa	4	3	3
	Cambre	5	1	1
2	Ruesga	5	5	5
	Cernégula	5	1	1
	Urbasa	9	4	4
3	Puerto de La Cubilla	5	1	0
	Valgañon	5	4	3
	Monforte de Lemos	6	1	1
5	Serra da Estrela	6	4	4
	Xinzo de Limia	6	2	2
6	Beira	5	5	4
7	Ólvega	5	5	4
8	Buenache	5	5	5
9	Boceguillas	5	5	5
	Codesal	2	2	2
	Salamanca	4	4	4
	Ferreras de Arriba	5	2	2
	Ciudad Rodrigo	2	1	0

10	Villa Verde	7	6	5
	Albires	5	Unsuccessful DNA Extractions	
11	Paramos	4	4	4
	Arzila	5	5	4
12	Navas de Estena	5	4	3
	Saceruela	4	3	3
	Navalcán	5	1	0
13	Casa Velha	8	5	5
	São José de Lamarosa	2	1	1
	Monteclaro	1	Unsuccessful DNA Extraction	
14	Torresfreneda	5	2	1
	Rio Esteras	4	3	2
	Fontanosas	1	1	0
	Cabezarrubias	5	Unsuccessful DNA Extractions	
15	Coruche	6	6	3
Total		161	100	85

4.3. RAW DATA TREATMENT

Single Nucleotide Polymorphism calling and genotyping using DArTseq in 100 individuals of *H. molleri* yield a total of 31 957 SNPs, of which 8 191 had a call rate >80% of the individuals. One sample was excluded from the analyses due to missing data in over 95% of the loci, referring to the sampling locality of "Fontanosas" belonging to cluster 14. Since this was the only sample available from this locality, all downstream analyses did not include data from Fontanosas.

Preliminary assessment of the effect of missing data on the overall pattern of genetic distances among individuals using PCoA suggested robustness of the signal to noise ratio (Table 4.3). Across datasets, the main pattern uncovered was a North-South split of the individuals and a general tendency of increased genetic distance among more distantly located individuals (Figure 4.5) (For additional graphics see Appendix Figure 21, Appendices).

Table 4.3 - SNP genotype datasets based on different cutoff levels of missing data at the locus and individual levels, as used in the PCoAs.

Loci missing data (%)	Genotype missing data (%)	No. of loci kept	No. of individuals kept
5	5	689	85
	10		92
	15		93
	20		96
10	5	2 972	70
	10		85
	15		88
	20		92
15	5	5 911	56
	10		73

	15		84
	20		87
20	5	8 703	31
	10		66
	15		78
	20		85
25	20	12 418	78
30	20	16 108	72
40	20	22 761	60

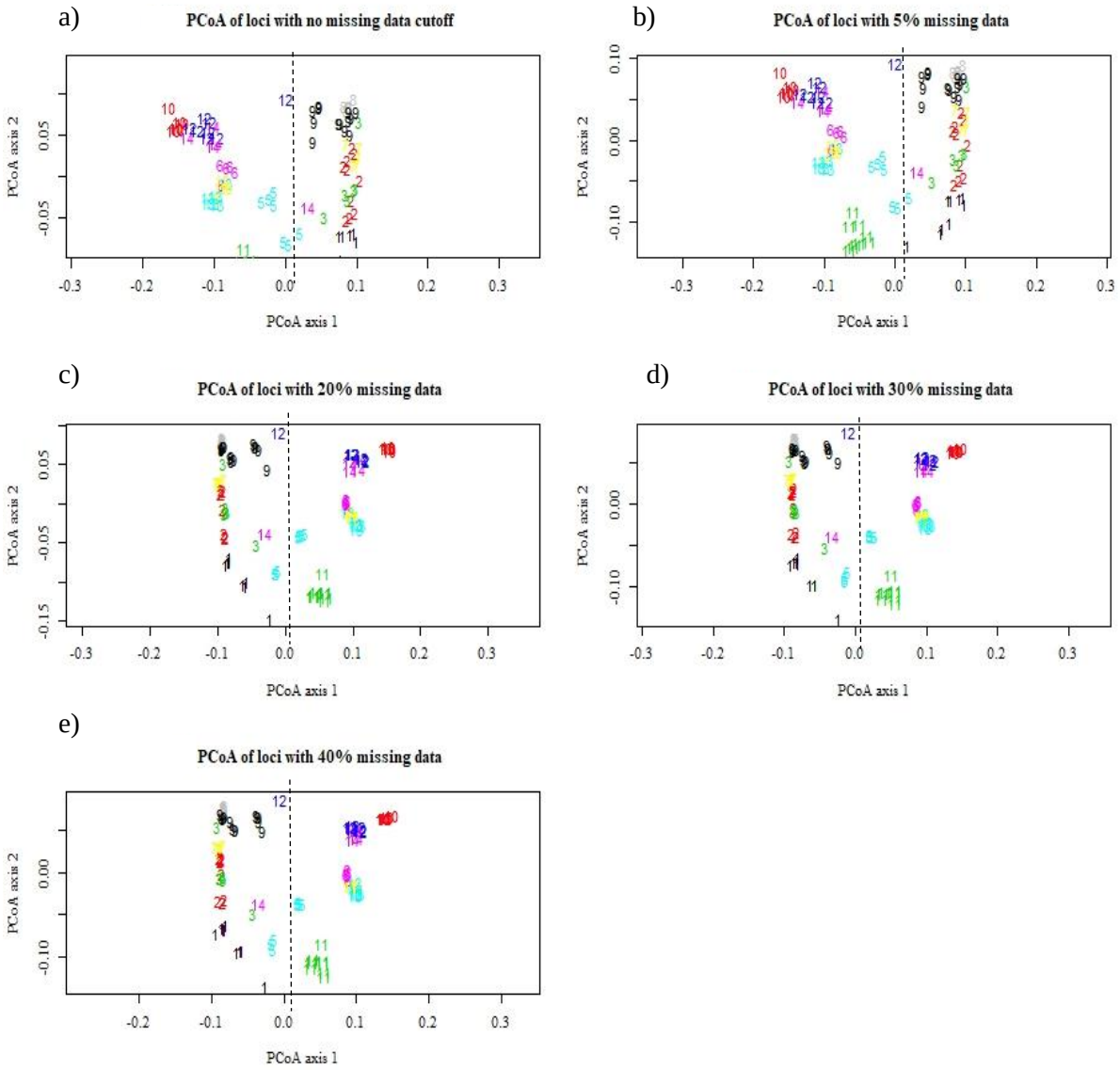


Figure 4.5 - PCoA results for the datasets with different cutoff levels of missing data at the locus and individual levels. Numbers represent each individual cluster. Only the first two dimensions of the PCoA were plotted.

Given the consistency of the above pattern across datasets, and aiming at the best compromise between the number of individuals and number of loci included in the analyses, the dataset with

a 20% level of missing data at the locus- and individual-levels was used in all subsequent analyses, *i.e.* including 85 individuals and 8 532 SNP loci.

It should be noted that one individual from cluster 14 and another from cluster 12 are consistently separated from the rest of the individuals from the same group.

4.3. GENETIC DIVERSITY

Private alleles

When considering the full dataset, the mean number of private alleles per individual varied between 0 (Cernégula, Cluster 2, and Ólvega, Cluster 7) and 161 (Casa Velha, Cluster 13). Southern populations showed higher average of private alleles per individual than northern populations (Figure 4.6).

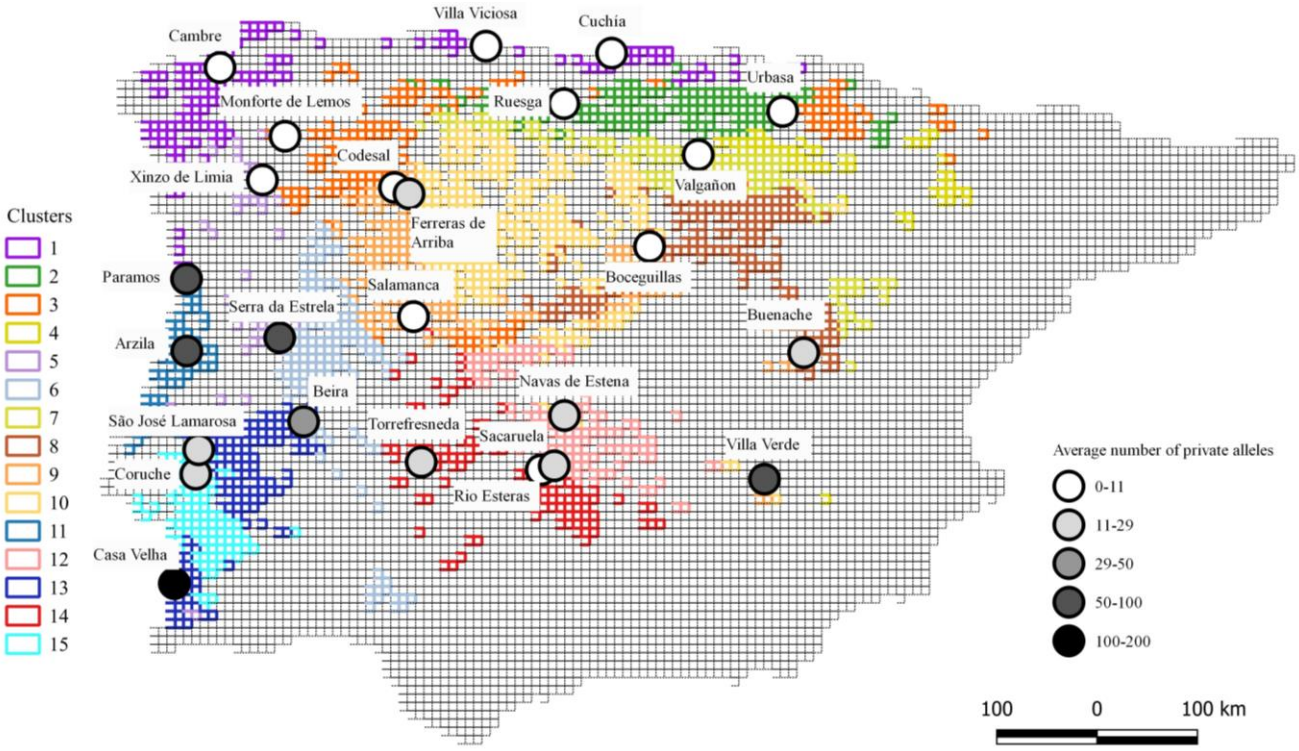


Figure 4.6 - Number of private alleles per location.

Observed Heterozygosity

Heterozygosity values ranged from 0.05 to 0.132. The populations with higher observed heterozygosity were Saceruela, Navas de Estena, Torresfresneda and Rio Esteras, all located in the south center of Iberia, and belonging to clusters 12 and 14, both influenced by the Mediterranean Central Climate (Figure 4.7). Populations with lower observed heterozygosity were Villa Verde, Buenache, Cambre, Villa Viciosa and Cuchía.

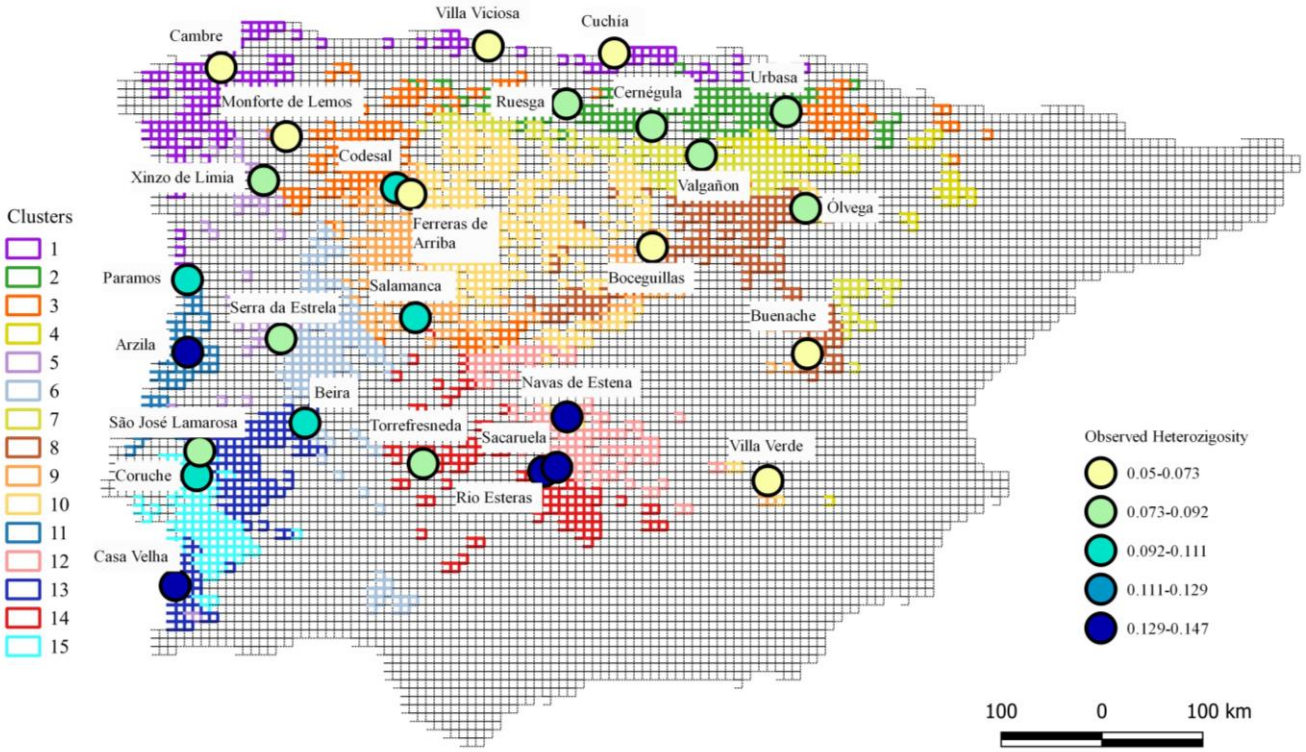


Figure 4.7 – Observed Heterozygosity for each population

4.4. PUTATIVE NON-NEUTRAL LOCI

In the MEM analysis few loci presented a z-score higher that the three tested cutoffs (Figure 4.8). The number of loci classified as outliers varied between 110 loci, 4 loci and 0 loci for the

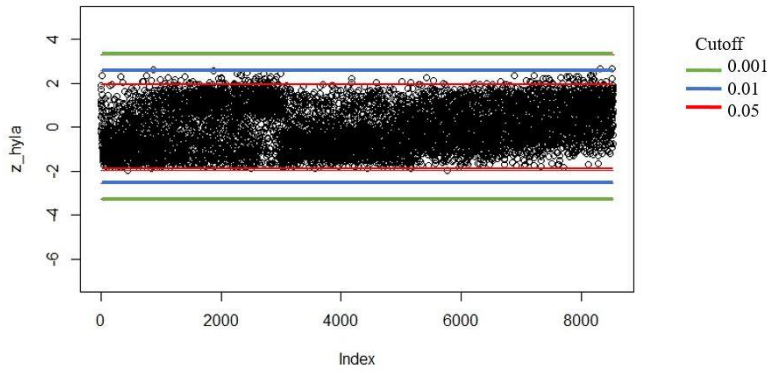


Figure 4.8 - z-scores for the power spectrum.

0.05, 0.01 and 0.001 cutoffs, respectively.

Since we want to ensure that no false positives are being detected, we chose to pursue a more conservative approach and keep the 0.001 cutoff, meaning that no loci was considered an outlier.

4.5. POPULATION STRUCTURE ANALYSIS

4.5.1 Spatial Analysis of Principal Components (sPCA)

The sPCA showed larger eigenvalues in the positive axis, suggesting the existence of global structure in the full dataset and positive spatial autocorrelation (Figure 4.9). The Monte-Carlo test confirmed the presence of global structure (p-value < 0.05) but not of local structure (p-value > 0.05). Taking the screeplot into consideration, we chose to explore the first two eigenvalues (λ_1 and λ_2) further, since they held the largest variance components of the data

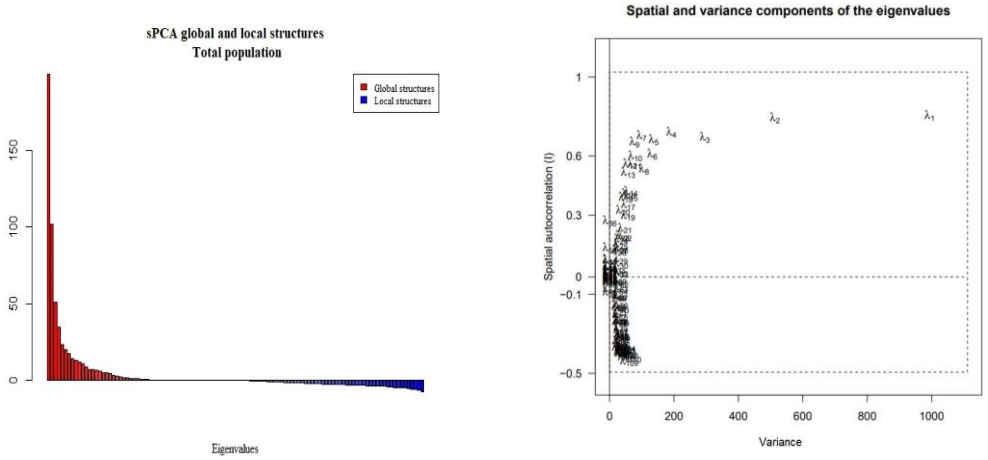


Figure 4.9 - Global and local structures obtained by *sPCA* eigenvalues (left) and screeplot (right)

(Figure 4.9).

The global structure detected in the sPCA refers to a) a north/south gradient between the samples, clearly visible in the first positive axis; and b) a gradient between coastal and inland zones in the second positive axis (Figure 4.10).

One of the two samples from Rio Esteras presented a low scores of both the first and second global eigenvalue, in contrast to the large values of its counterpart (see Appendix Table 1 and Appendix Table 2, Appendices) suggesting it is genetically closer to other individuals with low scores (such as those found in Cambre and Monforte de Lemos – first eigenvalue-, and Ólvega – second eigenvalue-) than individuals from geographically closer locations. This individual is also separate from those of the same cluster in the PCoA results suggesting that despite its southern geographical location, it has a high percentage of northern genetic composition that is not found in individuals from the same area.

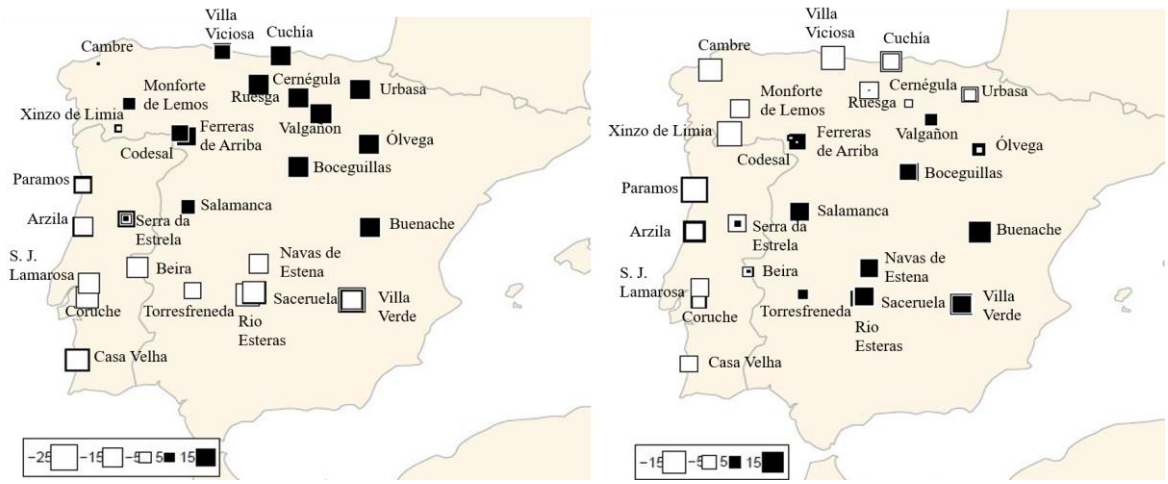


Figure 4.10- Two dimensional scatter plot of *sPCA* eigenvalue first (left) and second (right) axis.

Since northern populations were identified by positive values and southern ones by negative values on the first global axis, it would be expected that Xinzo de Limia showed positive values, due to its northern geographical location. However, the two individuals from Xinzo de Limia show low negative values.

4.5.2. Sparse Non-Negative Matrix Factorization (snmf) analysis

The graphical representation of the cross-entropy values for all runs showed the most likely number for ancestral populations (K) to be 4 (Figure 4.11). However, there is an abrupt decrease in cross-entropy values between K=1 and K=2. For this reason, both K=2 and K=4 were evaluated as the number of ancestral populations.

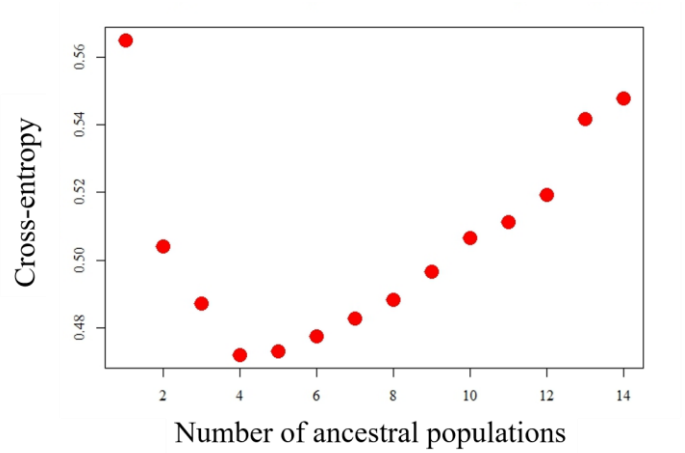


Figure 4.11 - Cross-entropy values for the run with the data set with 20% missing data for both for loci and genotypes.

To facilitate visualization of the results below, the clusters were ordered by their position in the PCoA graphics (Figure 4.5), *i.e.* from top to bottom and from left to right. When closely examined, this pattern reveals that the first 6 clusters correspond to the northern part of the Iberian Peninsula moving from the more inland clusters to the coastal ones. Cluster 5, is located between the northern and southern groups of *H. molleri*, and its followed by clusters found in the south, this time moving from the coastal to inland clusters.

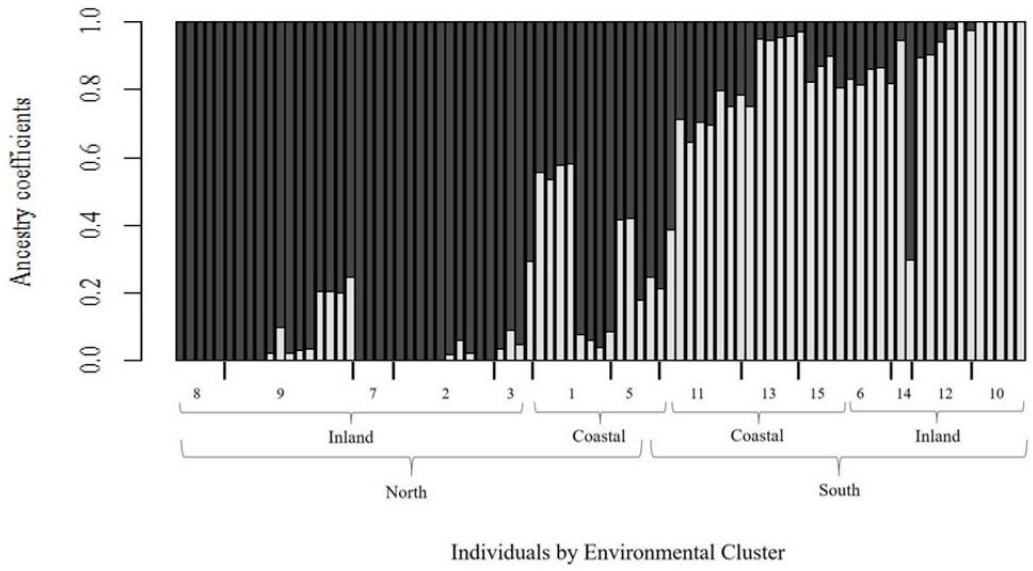


Figure 4.12 - Bar plot of ancestry coefficients for each individual when K = 2. Each colour represents one ancestral group.

For K=2, the results showed that clusters 8, 9, 7, 2, 3, and 1, had distinct ancestry coefficients from clusters 11, 13, 15, 6, 14, 12 and 10, highlighted in black and white respectively (Figure 4.12). However, one individual from cluster 14 showed genetic composition very distinct from its counterparts, being more similar to those of cluster 1 and 9. Cluster 5 displayed similar proportions of the two ancestral populations. This result is coherent with the North-South gradient suggested by the PCoA and by the sPCA.

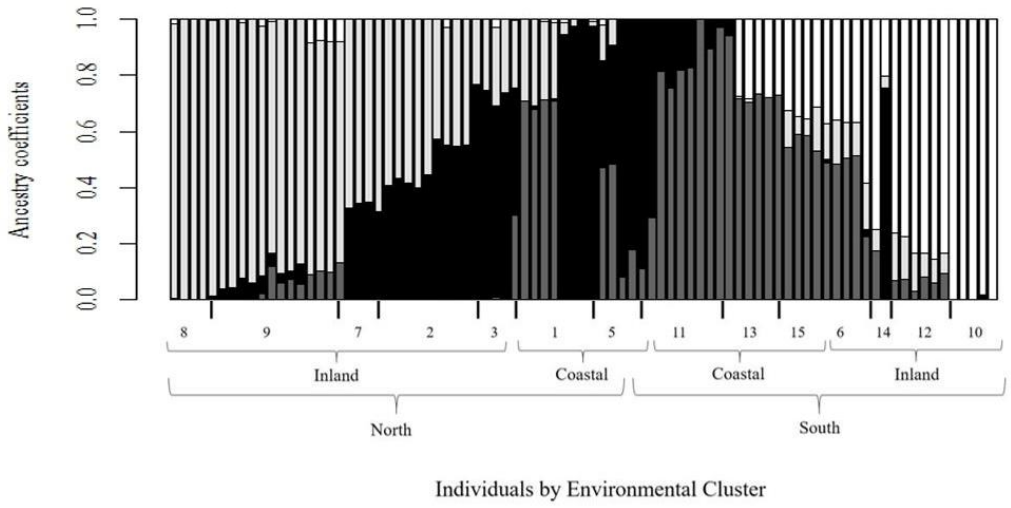


Figure 4.13- Bar plot of ancestry coefficients for each individual when K= 4. Each colour represents one ancestral group.

For K=4, the barplot of individual ancestry coefficients showed the north-south separation obtained by K=2 and a further separation within each group as shown by the presence of two predominantly northern ancestral populations and two predominantly southern ancestral populations (Figure 4.13). Within each of the northern and southern groups, the coastal and inland clusters showed distinct admixture proportions of each of the two dominant ancestral populations. This divide is also congruent with sPCA results.

Figure 4.14 shows how the coefficients of ancestry change across the sampled localities for both values of K. When K=4 there is almost a compatible pattern among the ancestry groups and the climatic areas previously defined: the ancestry group represented in black has a higher influence in locations from the northern coast (cluster 1) and the transition zone between coastal and inland northern populations (clusters 2, 3 and 5) – areas characterized by Atlantic Coastal Climate and Atlantic High Altitude Climate-; being replaced by the ancestry group represented in light grey as we move inland (mainly comprised of cluster 8, 9 and 10) - influenced by the Arid Climate; across the western coast the group represented in darker grey dominates, including mainly clusters 11, 13 and 15 – characterized by the Mediterranean Coastal Climate; and finally the forth ancestry group, represented in white, has a bigger contribution in inland southern populations (cluster 12 and 14) – characterized by Mediterranean Central Climate.

The separate ancestry groups found with snmf are congruent with the north/south and coastal/inland gradient found in the sPCA analysis, suggesting genetic differentiation among populations of these areas. In both cases it is possible to observe that clusters located on the southeastern tip of the species' range (clusters 8 and 10, populations of Buenache and Villa Verde), show distinctive exclusive ancestry groups. It is also interesting that the central area of the Peninsula (Salamanca, Torresfreneda, Navas de Estena, Saceruela and Rio Esteras) show genetic signature of northern ancestry.

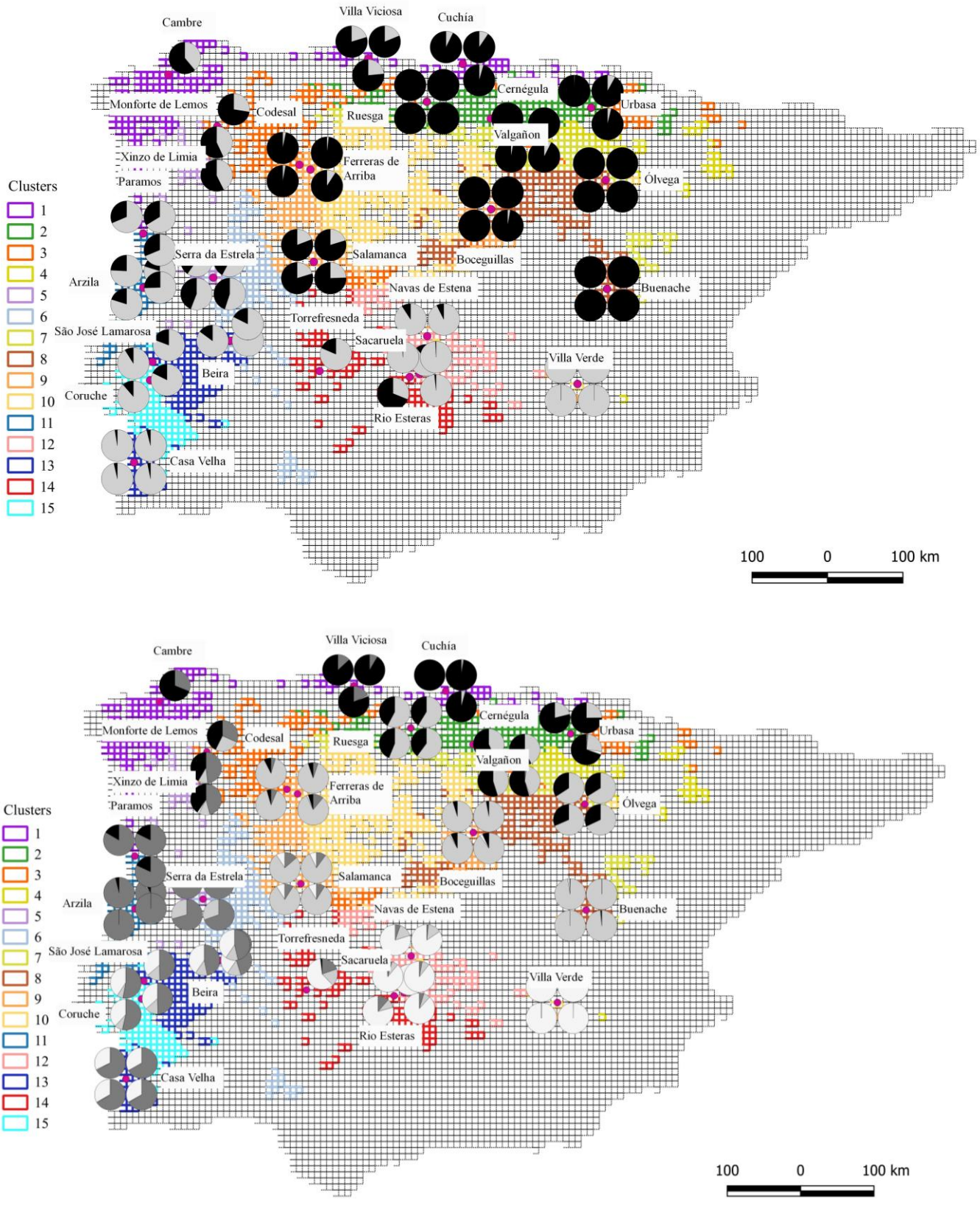


Figure 4.14.- Pie charts showing the average admixture coefficient of individuals per sampled locality for K = 2 (top) and K=4 (bottom).

4.6. LANDSCAPE GENETICS

The MMRR analysis for the whole dataset showed a significant (*P*-value < 0.05) positive correlation between genetic distance (F_{ST}) and both geographical distance and environmental distances (Table 4.4). This result indicates that when the geographical and environmental distance between individuals increases, so does the genetic distance. However, regression coefficients are low, meaning that although statistically significant, the relation is fairly low. The coefficient of determination is not very high if considering that R^2=0.7 is commonly considered good.

Table 4.4 - MMRR results.

Predictors	Regression coefficient	*p-value*	R^2
Geographical distance	0.0179	0.005	0.196
Environmental Distance	0.0059	0.044	

5. DISCUSSION

5.1. *HYLA MOLLERI*'S GENETIC DIVERSITY AND POPULATION STRUCTURE

Through our work we were able to infer the spatial genetic structure of *H. molleri* across the species range and analyse if geographic and environmental distances influence the distribution of this species genetic diversity, thus fulfilling our goals.

Our results point to a differentiation between individuals from northern and southern populations of the Iberian Peninsula, as well as a differentiation between individuals in the center and peripheral areas of the species distribution, being individuals from the northern populations more similar between each other than with individuals from the southern group, and individuals from the center region also genetically more similar amongst themselves than with individuals from the periphery, and *vice versa*. The four ancestral populations coefficient distribution across individuals further reinforces this pattern of separation between northern-southern and center-periphery population. Greater genetic diversity was found in southern populations, in an area currently dominated by Mediterranean Climate, with both average of private alleles per individual and observed heterozygosity displaying higher values in this area. Our results also suggested Xinzo de Limia as an area of admixture between the differentiated population groups. Genetic distance was positively correlated with both geographical and environmental distances.

Low genetic diversity in *H. molleri* contradicts Moreira (2012) results but is in accordance with Rosa & Oliveira (1994), Bart *et. al* (2011) and a recent study by Sanchéz-Montes *et. al* (2019).

We found some inconsistent results with Sánchez-Montes *et. al* (2019), such as their findings of higher allelic diversity (allelic richness and observed heterozygosity) in northern Iberian mountains and in the western part of Iberia and higher value of private alleles in the southwestern and north-eastern areas of the species range[41]. Their results are more congruent

with Gvozdik *et. al* (2015), where the Galician populations presented higher genetic diversity and seemed to have a distinct haplotype, suggesting it might have been a glacial refugia[40]. Studies focused in other amphibians have shown results congruent with ours, *e.g*. *Discoglossus galganoi*[64], *Pleurodeles waltl*[65-674], *Alytes cisternasii*[43] and *Pelobates cultripes*[67, 68], showing higher genetic diversity in the southern range of its distribution, indicating that northern populations result from a recent colonization[65] after a rapid post-glacial expansion from a more southern refugia[43]. This differences might be explained by the use of different genetic markers and sample size. Microsatellites have higher mutation rates which can lead to higher levels in genetic diversity, and larger sample sizes can produce more detailed information.

For *Pelobates cultripes* lower genetic diversity was found in Monforte de Lemos, Limia and in the Galician populations, suggesting a common origin for this populations from a single expansion event[69]. A higher number of private alleles and genetic diversity was also found in the southwestern part of *Pelobates cultripes* range, congruent with the southern part of Iberia acting as a possible glacial refugia, from where the species has expanded[68].For both *Alytes* species, Reino *et. al* (2017) found evidence of several independent glacial refugia in western Portugal during the Pleistocene, from which the species' appear to have colonized the northern and eastern areas of its current range, leading to a progressive loss of genetic diversity along these geographical axes of expansion[43]. Therefore, considering that several studies have pointed out that as geographical distance from glacial refugia increases, genetic diversity appears to decrease[70], and bearing in mind that low genetic diversity has been consistently linked to repeated extinction and colonization effects[61], our results suggest that *H. molleri* took refuge in the southern part of the Iberian Peninsula and has expanded its range from there, with the northern range being the last to be occupied. This hypothesis is further corroborated by the extreme low values of observed heterozygosity in Villa Verde, Cambre, Villa Viciosa and Cuchía. However, Barth *et. al* (2011) and Sánchez-Montes *et. al* (2018) results indicate the opposite: higher genetic diversity in northern populations, suggesting that *H. molleri* glacial refugia were in fact located in the north part of Iberia[76].

A north-south and center-periphery differentiation has also been found in other studies focused in population structure of other Iberian amphibians such as *Pleurodeles waltl*[65], *Rana iberica*[72], *Alytes cisternasii*[43] and *Pelobates cultripes*[68], as well as in Sánchez-Montes *et. al* (2019) study. They further hypothesised that the two major clades (which correspond to the north-south group) had their origin in the Pleistocene, and that K=4 was the number of cluster that best explained the observed population structure, with similar display as our groups, supporting our results.

Environmental and geographical influence on genetic distances was expected as it had already been suggest by previous studies[38,43]. Once again our results are not completely compatible with Sánchez-Montes *et. al* (2018), as they did not find a correlation between environmental and genetic diversity[41]. However, the MMRR model had a very low coefficient of determination, implying that variables that were not taken into account are influencing *H. molleri* genetic distances the most, and not the environmental and geographical distances.

Finally, we must address one of our Rio Esteras sample, which throughout the study consistently showed signs of being different from other individuals of the same cluster. This sample showed higher genetic similarity with individuals from the northern area, especially those from cluster 3, and later showed a mix of ancestral contribution different from samples in close proximity, with a high influence of ancestral populations mainly found in the north. Therefore this sample is not representative of the Rio Esteras population as this curious

individual appears to be the result of a bizarre mating among a specimen brought from the norther populations and an individual from the southern populations.

5.2 IMPLICATIONS FOR CONSERVATION

. Due to central southern populations showing a higher value of private alleles and higher values of observed heterozygosity, this grants them an intrinsic value for conservation value. However, due to the contradicting results between our study and Sánchez-Montes *et. al* (2019), further studies should be conducted in order to better pin point where higher genetic diversity is found.

Finally, Sánchez-Montes *et. al* (2019) found what might be a correlation between climatic stable areas since the Pleistocene and *H. molleri* distribution. Therefore, future studies should aim to predict which areas within the species range are capable of keeping climatic stability in the future and assess whether these sites should be made into a conservation priority[68].

5.3 STUDY LIMITATIONS AND SUGGESTIONS FOR FUTURE STUDIES

My smaller sample size and genetic marker choice might have led to the discrepancies between my results and those of Sánchez-Montes *et. al* (2019). Ability to infer population structure is directly linked to the number and type of genetic loci, sample size collected from each population and the differentiation level among populations, which is directly impacted by the species effective population size, generation time and its dispersal rate[73]. Since both Sánchez-Montes *et. al* (2019) and our work was focused in *H. molleri*, generation time and dispersal rate cannot be considered as an explanatory factor for the differences found between studies. As for genetic markers, several studies have shown that SNPs have the same power as microsatellites in detecting fine-scale population structure[74-76] and their ability to detect geographical influence in genetic distances even in populations with very low differentiation ($F_{ST} < 0.01$), has also been prooved[71]. However, when it comes to uncovering genetic diversity patterns, microsatellites' higher mutation rate can result in different genetic diversity patterns than those produced by SNPs. Additionally, when it is possible to obtain a high number of SNPs (1000s), the power to detected genetic differentiation is kept, even with small sample sizes. Sample size in Sánchez-Montes *et. al* (2019) was almost three times larger than mine (85 *versus* 248 individual samples; 27 *versus* 60 populations). The distribution range they considered for *H. molleri* was also slightly different than ours, including a south-western part of France.

Although providing interesting insights into *H. molleri*'s population structure, genetic diversity and environmental features that are influencing gene flow, this study had some limitations. Even though great sampling efforts were made, we were not able to cover the whole species' distribution evenly. Cluster 10, as mentioned before, was not sampled in its main area of occurrence, being the only sample Villa Verde, which is located in the South part of the North/South divide of Iberia, while the main area of cluster 10 is located in the North. Therefore, I may not have a good representation of the cluster genetic diversity. As for cluster 4, on the contrary, no sampling might not have had an important impact in the obtained results since this cluster occupies a very small area and had several locations from nearby clusters sampled.

As for suggestions to future work, adding to the sampling design coverage suggestions, we also believe the following should be considered in future studies:

- Use higher resolution for environmental variables, such as 1 x 1 km, which would allow to account for microhabitats (such as ponds that the species can use to reproduce) that are dismissed when a 10 x 10 km resolution is used;

- Add more explanatory variables when testing for IBE, such as:

- Topography: it has been proven that topography as an important effect in population structure for *H. molleri*[77]and other amphibian species'[78-80], hence this variable should be taken into account when doing this type of analysis;

- Raw environmental variables: although the PCs provided a good notion of how different environmental conditions affect genetic diversity, it would be interesting to know which variables specifically are affecting population structure in order to better predict how climatic change can affect this species;

- Land cover: since different habitats affect amphibian distribution in distinct ways, land cover should be taken into account when assessing the drivers of genetic differentiation in these species. *Hyla arborea*, for example, is very influenced by wetlands presence as a result of being a pond breeder[81]. Angelone *et. al* (2011) showed that *H. arborea* uses habitat edges and a series of herbaceous habitats as dispersal corridors. Due to the similar behaviour expected between *H. arborea* and *H. molleri*, it may be safe to assume that *H. molleri* might prefer/need the same habitat features for dispersal. Therefore, the presence or absence of these habitats can impact *H. molleri*'s ability to disperse and, consequently, the genetic patterns observed.

- Hydrologic map: some water bodies, such as wide rivers might act as a barrier to dispersal - the Tagus as Douro rivers acted as a barrier to gene flow for *Discoglossus galganoi* until recently[64]-.

It would also be more informative if the study was conducted at different spatial scales: Angelone *et. al* (2011) found that different landscape elements influence *H. arborea* gene flow depending on the spatial scale. In a 0-2 km scale, only the proportion of rivers and lakes seemed to constrain gene flow, but in a 2-4 km scale it was a combination of the geographical distance and wetlands proportions that inhibit gene flow[81]. Consequently, a finer scale approach might allow for better understanding of what drives and restrains dispersal and gene flow in this species.

Finally, we also suggest incorporating road traffic in future studies, as a strong correlation was found between high traffic indices and *H. arborea* absence, indicating that roads negatively impact its distribution, and act as a barrier[82]. Due to the similarity between *H. arborea* and *H. molleri*, a similar effect may also be felt in this species.

REFERENCES[1]

1. Wake, D. B. Declining amphibian populations. *Science* **253,** 860 (1991).
2. Houlahan, J. E., Findlay, C. S., Schmidt, B. R., Meyer, A. H. & Kuzmin, S. L. Quantitative evidence for global amphibian population declines. *Nature* **404,** 752–755 (2000).
3. Gardner, T. Declining amphibian populations: a global phenomenon in conservation biology. *Anim. Biodivers. Conserv.* **24,** 25–44 (2001).
4. Stuart, S. N. *et al.* Status and Trends of Amphibian Declines and Extinctions Worldwide. *Science* **306,** 1783–1787 (2004).
5. Hoffmann, M. *et al.* The Impact of Conservation on the Status of the World 's Vertebrates. *Science* **330,** 1503–1509 (2010).
6. Allentoft, M. E. & O'Brien, J. Global amphibian declines, loss of genetic diversity and fitness: A review. *Diversity* **2,** 47–71 (2010).
7. Carey, C. & Alexander, M. a. Climate change and amphibian declines : is there a link ? *Divers. Distrib.* **9,** 111–121 (2003).
8. Corn, P. S. Climate change and amphibians. *Anim. Biodivers. Conserv.* **28,** 59–67 (2005).
9. Duan, R., Kong, X., Huang, M. & Varela, S. The potential effects of climate change on amphibian distribution, range fragmentation and turnover in China. *PeerJ* 1–17 (2016).
10. Carvalho, S. B. *et al.* From climate change predictions to actions – conserving vulnerable animal groups in hotspots at a regional scale. *Glob. Chang. Biol.* **16,** 3257–3301 (2010).
11. Frankham, R., Ballou, J. D. & Briscoe, D. A. A Primer of Conservation Genetics. *Cambridge University Press* (2004).
12. Rao V. R. &Hodgkin T. Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell. Tissue Organ Cult.* **68,** 1–19 (2002).
13. Sgrò, C. M., Lowe, A. J. & Hoffmann, A. A. Building evolutionary resilience for conserving biodiversity under climate change. *Evol. Appl.* **4,** 326–337 (2011).
14. Funk, C., McKay, J., Hohenlohe, P. A. & Allendorf, F. W. Harnessing genomics for delineating conservation units. *Trends Ecol. Evol.* **27,** 489–496 (2012).
15. Toro, M. A. & Caballero, A. Characterization and conservation of genetic diversity in subdivided populations. *Philos. Trans. R. Soc. B Biol. Sci.* **360,** 1367–1378 (2005).
16. Hedrick, P. W. Conservation genetics: where are we now? *Trends Ecol. Evol.* **16,** 629–636 (2001).
17. Manel, S. *et al.* Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Mol. Ecol.* **19,** 3760–3772 (2010).
18. Hewitt, G. The genetic legacy of the quaternary ice ages. *Nature* **405,** 907–913 (2000).
19. Frankham, R. Genetics and extinction. *Biol. Conserv.* **126,** 131–140 (2005).
20. Wright, S. Isolation by Distance. *Genetics* **28,** 114–138 (1943).

[1] The references in this dissertation follow the rules of *Nature* journal

21. Gomez, A. & Lunt, D. H. Refugia within refugia: patterns of phylogeographic concordance in Iberian Peninsula. *Phylogeography of Southern European Refugia* 155–188 (2007).

22. Manel, S., Schwartz, M. K., Luikart, G. & Taberlet, P. Landscape genetics: Combining landscape ecology and population genetics. *Trends Ecol. Evol.* **18,** 189–197 (2003).

23. Wang, I. J., Glor, R. E. & Losos, J. B. Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecol. Lett.* **16,** 175–182 (2013).

24. Zhanjiang, L. Aquaculture genome technologies. *Blackwell Publishing* (2007).

25. Morin, P. A., Luikart, G. & Wayne, R. K. SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* **19,** 208–216 (2004).

26. Brookes, A. J. The essence of SNPs. *Gene* **234,** 177–186 (1999).

27. Lowry, D. B. *et al.* Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Resour.* **17,** 142–152 (2017).

28. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6,** 1–10 (2011).

29. Kilian, A., Sanewski, G. & Ko, L. The application of DArTseq technology to pineapple. *Acta Hortic.* **1111,** 181–188 (2016).

30. Nguyen, N. H., Premachandra, H. K. A., Kilian, A. & Knibb, W. Genomic prediction using DArT-Seq technology for yellowtail kingfish Seriola lalandi. *BMC Genomics* **19,** 1–9 (2018).

31. Stöck, M. *et al.* Mitochondrial and nuclear phylogeny of circum-Mediterranean tree frogs from the *Hyla arborea* group. *Mol. Phylogenet. Evol.* **49,** 1019–1024 (2008).

32. Blair Hedges, S. An electrophoretic analysis of holarctic hylid frog evolution. *Syst. Biol.* **35,** 1–21 (1986).

33. Stöck, M. *et al.* Cryptic diversity among Western Palearctic tree frogs: Postglacial range expansion, range limits, and secondary contacts of three European tree frog lineages (*Hyla arborea* group). *Mol. Phylogenet. Evol.* **65,** 1–9 (2012).

34. Rosa, H. D. & Oliveira, M. E. Genetic differentiation of the Iberian tree frogs *Hyla arborea molleri* and *Hyla meridionalis* (Amphibia: Anura). *J. Zool. Syst. Evol. Res.* **32,** 117–128 (1994).

35. Pleguezuelos, J. M., Márquez, R. & Lizana, M. (eds. . *Atlas y Libro Rojo de los Anfibios y Reptiles de España*. *Dirección General de Conservación de la Naturaleza-Asociación Herpetologica Española (2ª impresión), Madrid* (2002).

36. Loureiro, A., Ferrand de Almeira, N., Carretero, M. A. & S. Paulo, O. *Atlas dos Anfíbios e Répteis de Portugal*. (Instituto da Conservação da Natureza e da Biodiversidade, 2008).

37. Gonçalves, H., Sequeira, F. & Ferrand, N. Multilocus phylogeography of the common midwife toad, *Alytes obstetricans* (Anura, Alytidae): contrasting patterns of lineage diversification and genetic structure in the Iberian refugium. *Mol. Phylogenet. Evol.* **(93),** 363-379 (2015)

38 Barth, A. *et al.* Mitochondrial uniformity in populations of the treefrog *Hyla molleri* across the Iberian Peninsula. *Amphibia-Reptilia* **32,** 557–564 (2011).

39. Gvoždík, V. *et al.* Speciation history and widespread introgression in the European short-call tree frogs (*Hyla arborea sensu lato, H. intermedia* and *H. sarda*). *Mol. Phylogenet. Evol.* **83,** 143–155 (2015).

40 Moreira, C. Population Analysis of *Hyla arborea* and *Hyla meridionalis* (Amphibia, Anura) in Portugal: A Molecular Genetic and Bioacoustic Approach. (Lisbon University, 2012).

41. Sánchez-Montes, G., Recuero, E., Barbosa, A. M. & Marínez-Solano, I. Complementing the Pleistocene biogeography of European amphibians: testimony from a southern Atlantic species. *J. Biogeogr.* **66** (2018).

42. Gonçalves, H., Matínez-Solano, I., Pereira, R.J., Carvalho, B. High levels of population subdivision in a morphologically conserved Mediterranean toad (*Alytes cisternasii*) result from recent , multiple refugia: evidence from mtDNA , microsatellites and nuclear genealogies. *Molecular Ecology* **18(24)**, 5143–5160 (2009).

43. Reino, L. *et al.* Favourable areas for co-occurrence of parapatric species: niche conservatism and niche divergence in Iberian tree frogs and midwife toads. *J. Biogeogr.* **44,** 88–98 (2017).

44. Fick, S. E. & Hijmans, R. J. Worldclim 2: New 1-km spatial resolution climate surfaces

45. Scrucca L., Fop M., Murphy T.B. & Raftery A.E. mclust 5: clustering, classiification and density estimation usin Gaussian finite mixture models. *The R journal* **(8/1)**, pp. 205-233 (2016)

46. McFerrin, L. Package 'HDMD': Statistical Analysis Tools for High Dimension Molecular Data. *Bioinformatics* **21** (2015).

47. RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA (2015)

48. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**,1403-1405 (2008)

49. Kamvar ZN, Tabima JF, Grünwald NJ. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281. doi: 10.7717/peerj.281

50. Jombart, T. A tutorial for the spatial Analysis of Principal Components using *adegenet* 1.3-4. Imperial College London.

51. Slatkin, M. Rare alleles as indicators of gene flow. *Evolution* **39,** 53–65 (1985).

52. Wagner, H. H., Chávez-Pesqueira, M. & Forester, B. R. Spatial detection of outlier loci with Moran eigenvector maps. *Mol. Ecol. Resour.* **17,** 1122–1135 (2017).

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

54. Frichot, M. E. & Francois, O. LEA: an R package for Landscape and Ecological Association Studies. *Methods in Ecology and Evolution* **6(8)** (2015).

55 Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196,** 973–983 (2014).

56. Hoyer, P. O. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research* **5,** 1457–1469 (2004).

58. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155,** 945–959 (2000).

59. Falush, D., Stephens, M. & Pritchard, J. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* **164,** 1567–1587 (2003).

60. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Mol. Ecol. Notes* **7,** 574–578 (2007).

61. Wang, I. J. Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution (N. Y).* **67,** 3403–3411 (2013).

62. Goudet, J. & Jombart, T. hierfstat: Estimation and Tests of Hierarchical F-Statistics. *R Core Team* **58** (2015).

63. Adamack, A. & Gruber, B. PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecologu and Evolution* **5(4)**, 384-387 (2014)

64. Martínez-Solano, Í. Phylogeography of Iberian *Discoglossus* (Anura: Discoglossidae). *J. Zool. Syst. Evol. Res.* **42,** 298–305 (2004).

65. Gutiérrez-rodríguez, J., Barbosa, A. M. & Martínez-solano, Í. Integrative inference of population history in the Ibero-Maghrebian endemic *Pleurodeles waltl* (Salamandridae). *Molecular Phylogenetics and Evolution* **112,** 122–137 (2017).

66. Carranza, S. & Arnold, E. N. History of West Mediterranean newts , *Pleurodeles* (Amphibia : History of West Mediterranean newts , *Pleurodeles* (Amphibia : Salamandridae), inferred from old and recent DNA. *Syst. Biodivers.* **1,** 327–337 (2004).

67. Veith, M. *et al.* From Europe to Africa and vice versa : evidence for multiple intercontinental dispersal in ribbed salamanders (Genus *Pleurodeles*) *Journal of Biogeography* **31(1)**, 159–171 (2004).

68. Gutiérrez-Rodríguez, J., Barbosa, A.M., Martínez-Solano. Present and past climatic effects on the current distribution and genetic diversity of the Iberian spadefoot toad (*Pelobates cultripes*): an integrative approach. *Journal of Biogeography* **44(2)**, (2016).

69. Crottini, A., Galán, P. & Vences, M. Mitochondrial diversity of Western spadefoot toads, *Pelobates cultripes*, in northwestern Spain. *Amphibia-Reptilia* **31,** 443–448 (2010).

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

71. Edenhamn, P., Höggren, M. & Carlson, A. Genetic Diversity and Fitness in Peripheral and Central Populations of the European Tree Frog *Hyla Arborea*. *Hereditas* **133,** 115–122 (2000).

72.. Teixeira, J., Gonçalves, H., Ferrand, N., García-París, M. & Recuero, E. Mitochondrial phylogeography of the Iberian endemic frog *Rana iberica* , with implications for its conservation. *Curr. Zool.* **64,** 755–764 (2018).

73. Morin, P. A. *et al.* An empirical comparison of SNPs and microsatellites for population structure , assignment , and demographic analyses of bowhead whale populations. *Endanger. Species Res.* **19,** 129–147 (2012).

74. Morin, P., Martien, K. K. & Taylor, B. Assessing statistical power of SNPs for population structure and conservation studies. *Mol. Ecol. Resour.* **9,** 66–73 (2009).

75. Haasl, R. J. & Payseur, B. A. Multi-locus inference of population structure : a comparison between single nucleotide polymorphisms and microsatellites. *Heredity (Edinb).* **106,** 158–171 (2011).

76. Garke, C. *et al.* Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Anim. Genet.* **43,** 419–428 (2011).

77. Sánchez-Montes, G., Wang, J., Ariño, A. H. & Martínez-Solano, Í. Mountains as barriers to gene flow in amphibians: Quantifying the differential effect of a major mountain ridge on the genetic structure of four sympatric species with different life history traits. *J. Biogeogr.* **45,** 318–331 (2018).

78. Wang, I. J. Fine-scale population structure in a desert amphibian: Landscape genetics of the black toad (*Bufo exsul*). *Mol. Ecol.* **18,** 3847–3856 (2009).

79. Emaresi, G., Pellet, J., Dubey, S., Hirzel, A. H. & Fumagalli, L. Landscape genetics of the Alpine newt (*Mesotriton alpestris*) inferred from a strip-based approach. *Conserv. Genet.* **12,** 41–50 (2011).

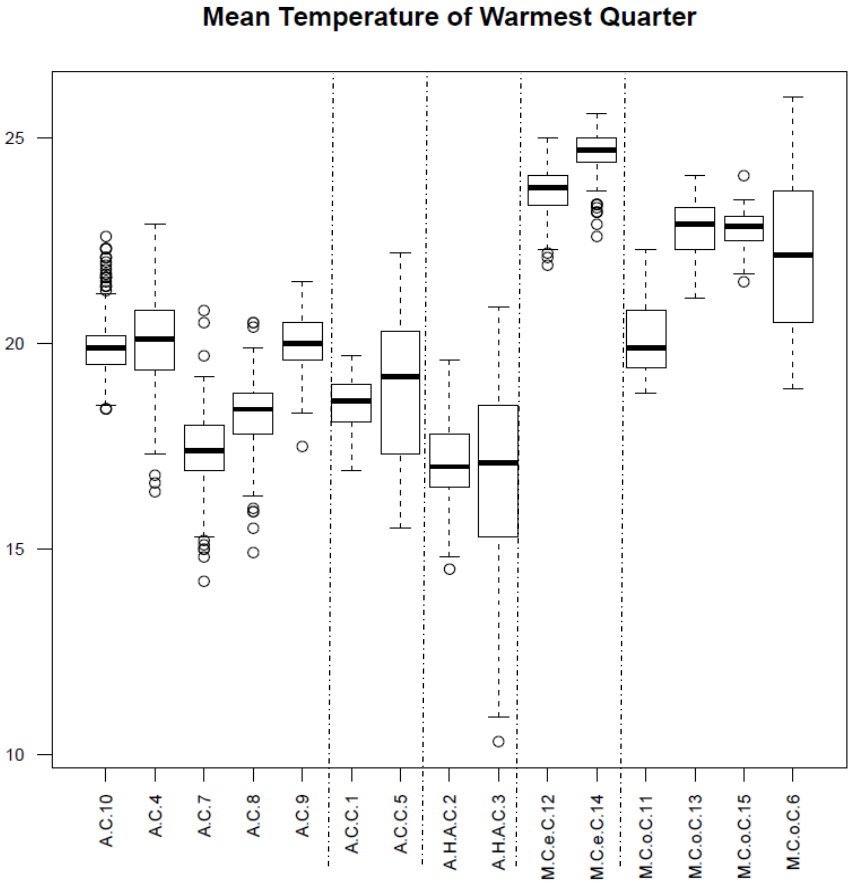
80. Spear, S. F., Peterson, C. R., Matocq, M. D. & Storfer, A. Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Mol. Ecol.* **14,** 2553–2564 (2005).

81. Angelone, S., Kienast, F. & Holderegger, R. Where movement happens: Scale-dependent landscape effects on genetic differentiation in the European tree frog. *Ecography (Cop.).* **34,** 714–722 (2011).

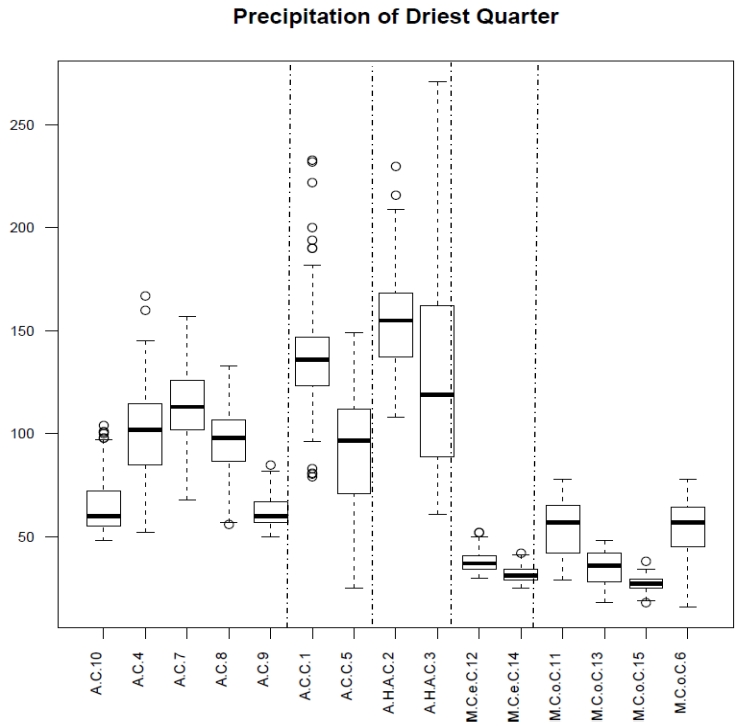
82. Pellet, J., Guisan, A. & Perrin, N. A concentric analysis of the impact of urbanization on the threatened European tree frog in an agricultural landscape. *Conserv. Biol.* **18,** 1599–1606 (2004).

APPENDICES

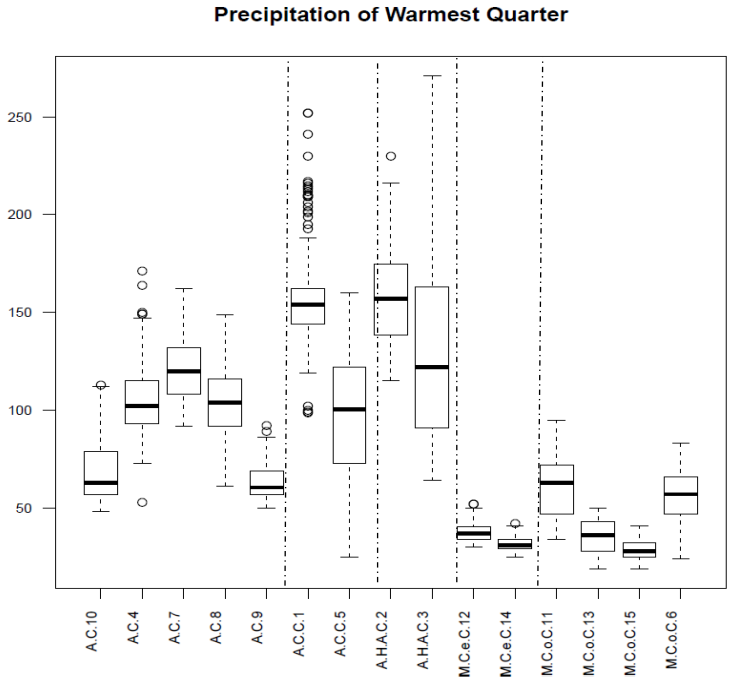
1. CHARACTERIZATION OF EACH CLIMATIC CLUSTER AND MAIN CLIMATE AREAS



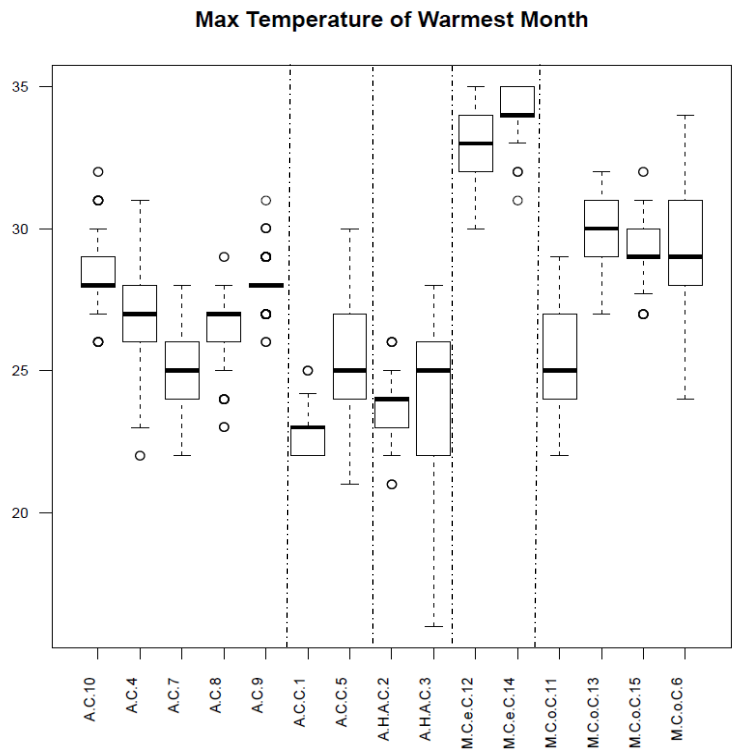
Appendix Figure 1- Mean Temperature of Warmest Quarter for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



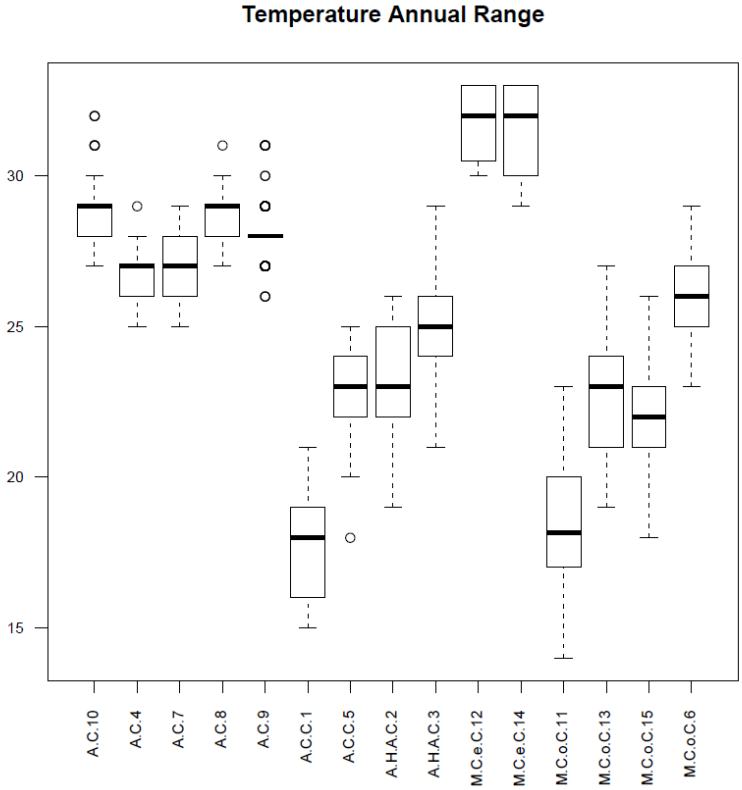
Appendix Figure 2- Precipitation of Driest Quarter for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



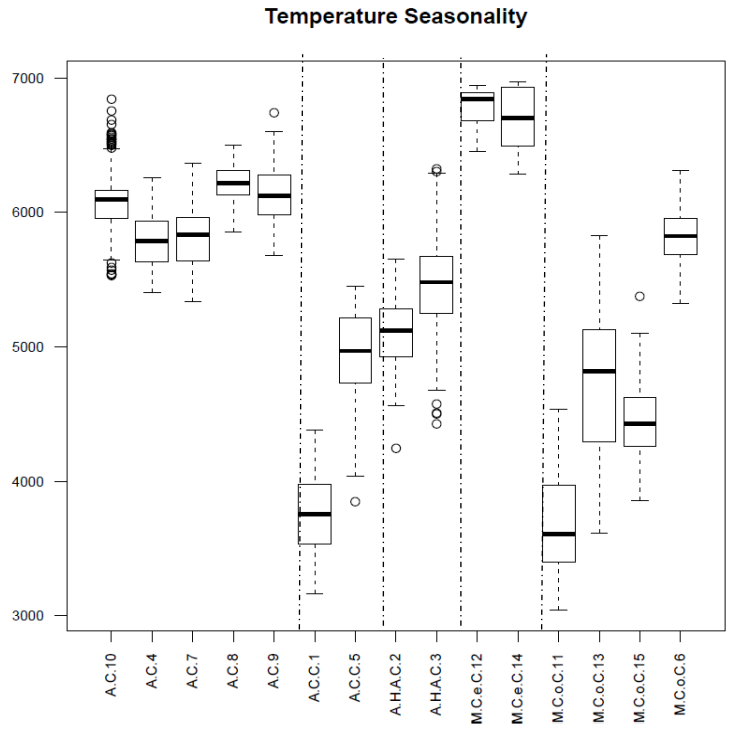
Appendix Figure 3- Precipitation of Warmest Quarter for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



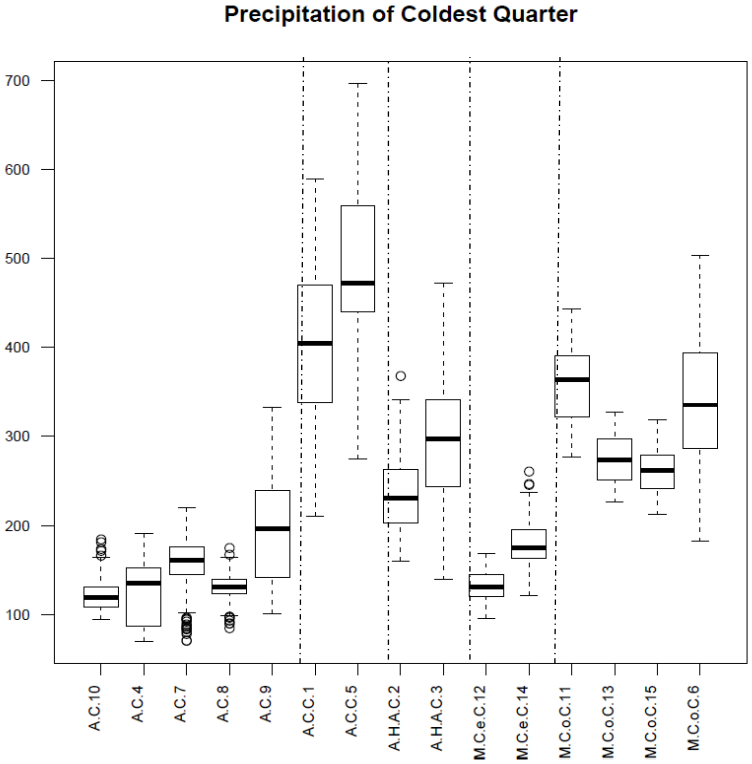
Appendix Figure 4- Maximum temperature of warmest month for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



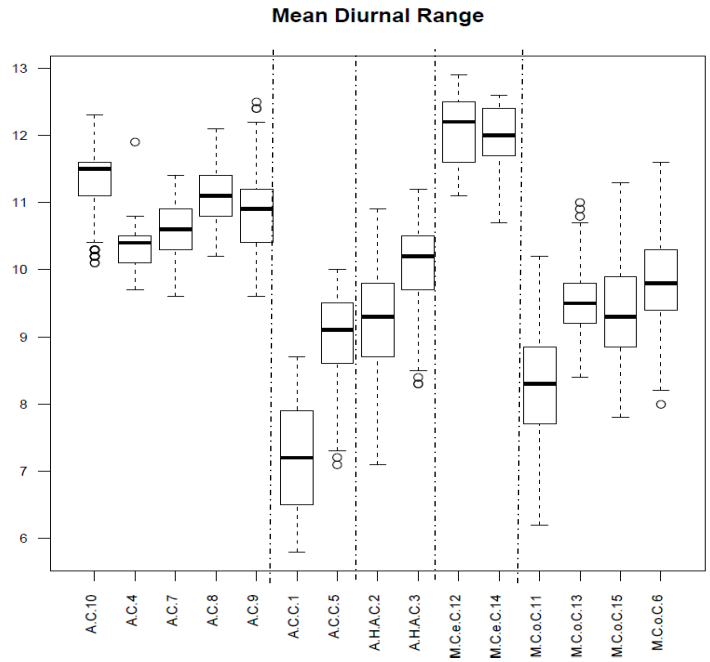
Appendix Figure 5- Temperature annual range for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



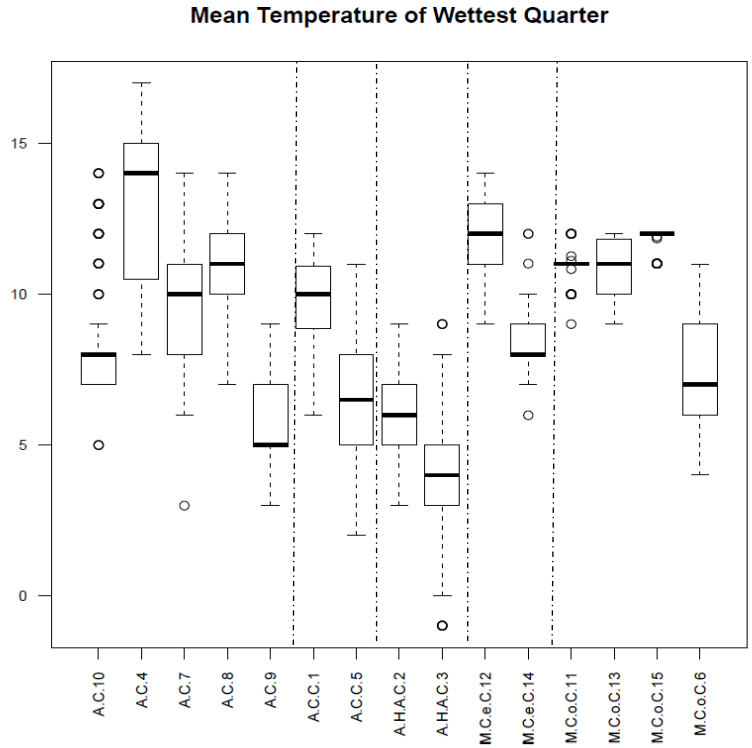
Appendix Figure 6- Temperature seasonality for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



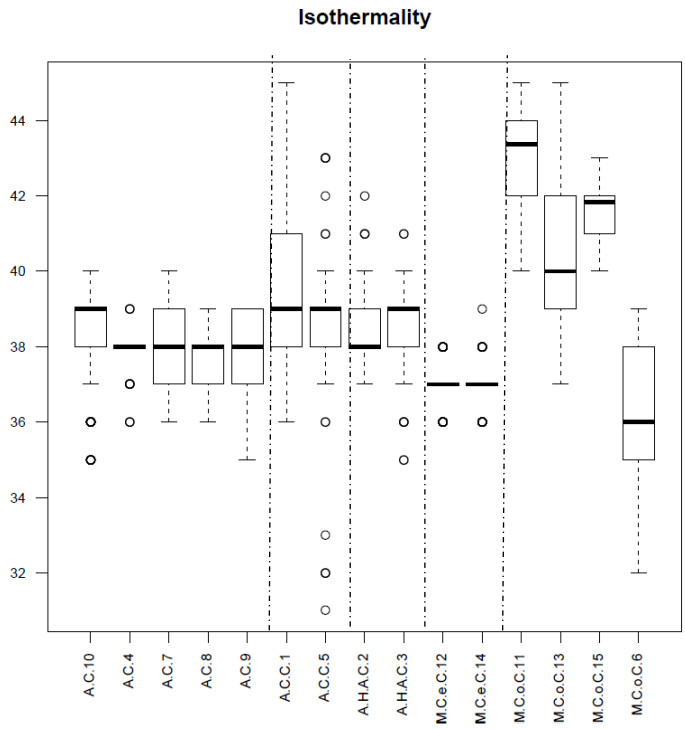
Appendix Figure 7- Precipitation of coldest quarter for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



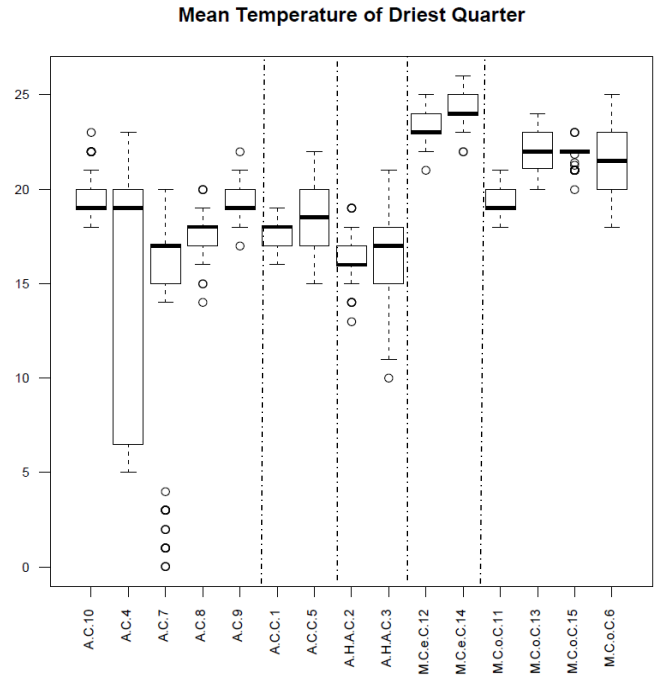
Appendix Figure 8- Mean diurnal range for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



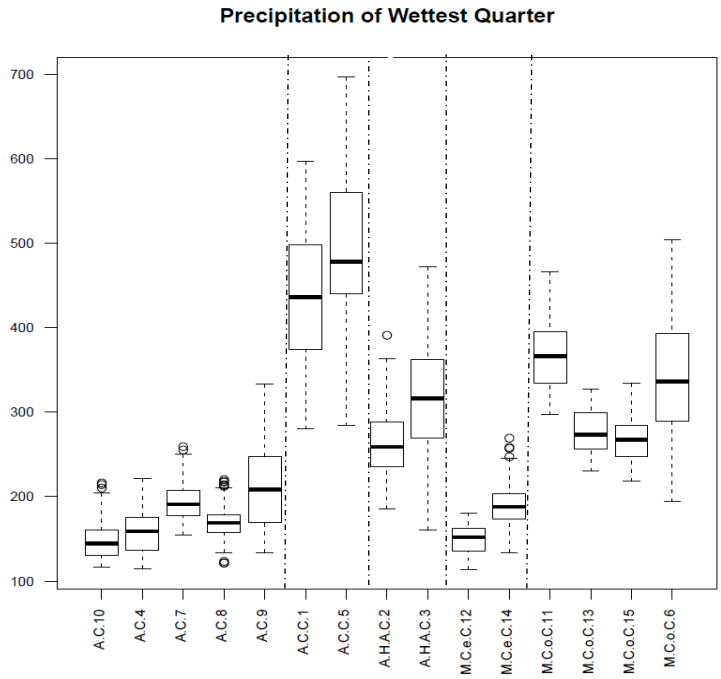
Appendix Figure 9- Mean temperature of wettest quarter for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



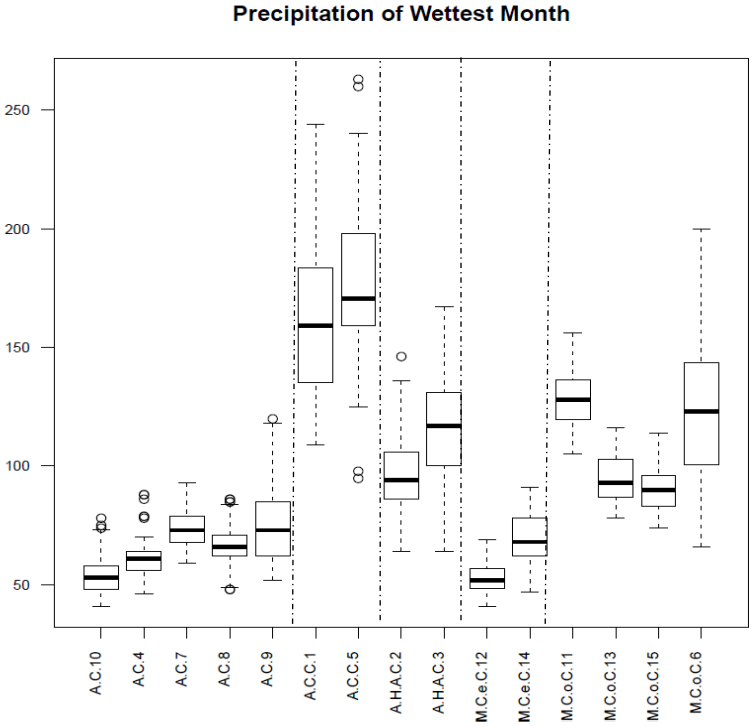
Appendix Figure 10- Isothermality for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



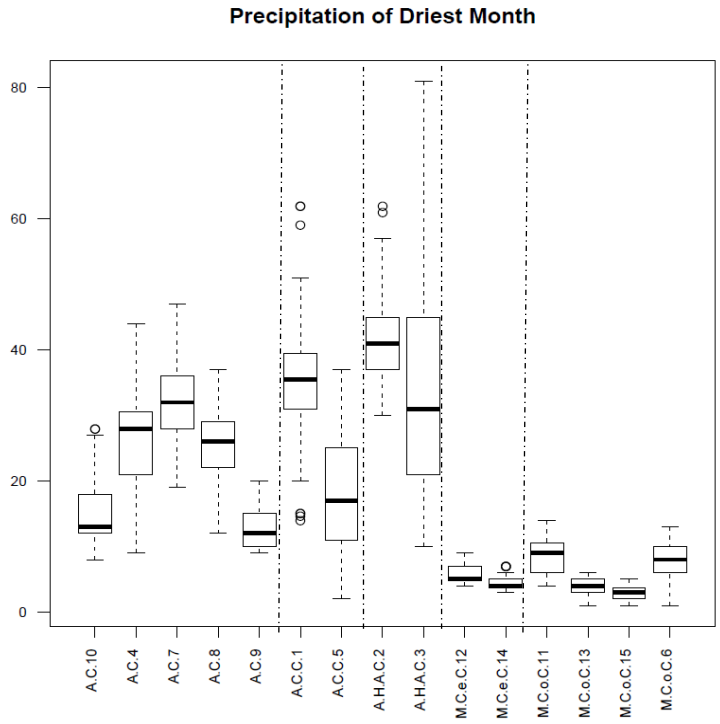
Appendix Figure 11- Mean temperature of driest quarter for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



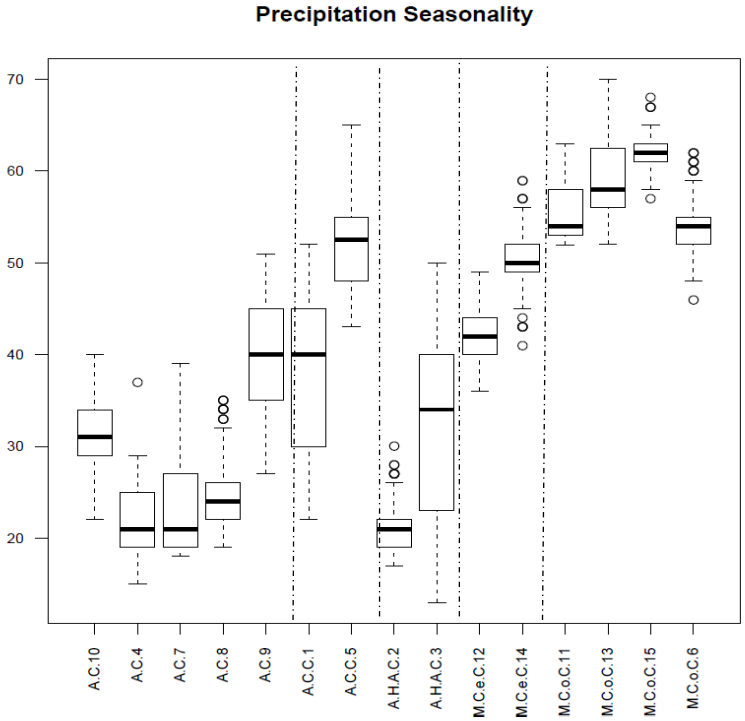
Appendix Figure 12 – Precipitation of wettest quarter for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



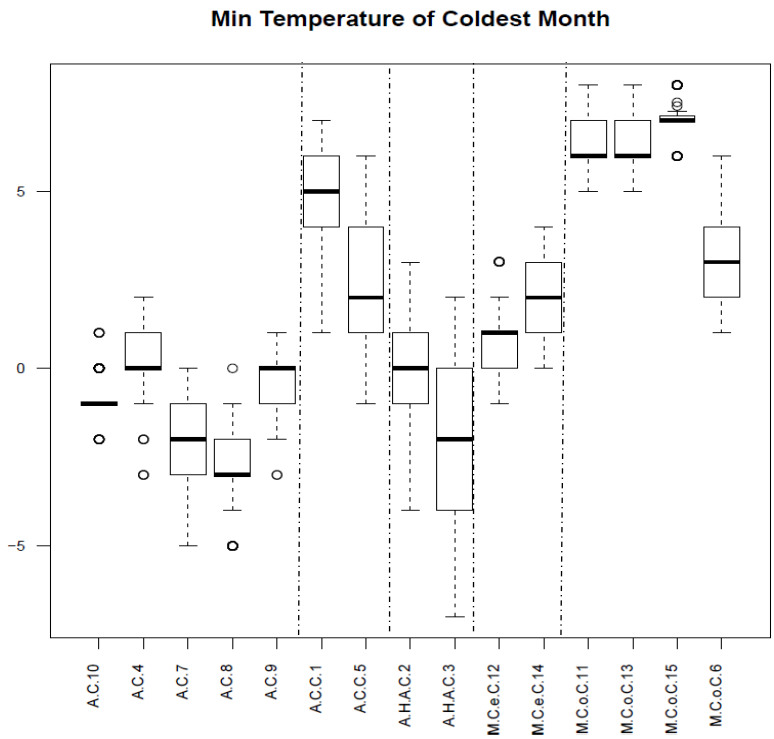
Appendix Figure 13 – Precipitation of wettest month for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



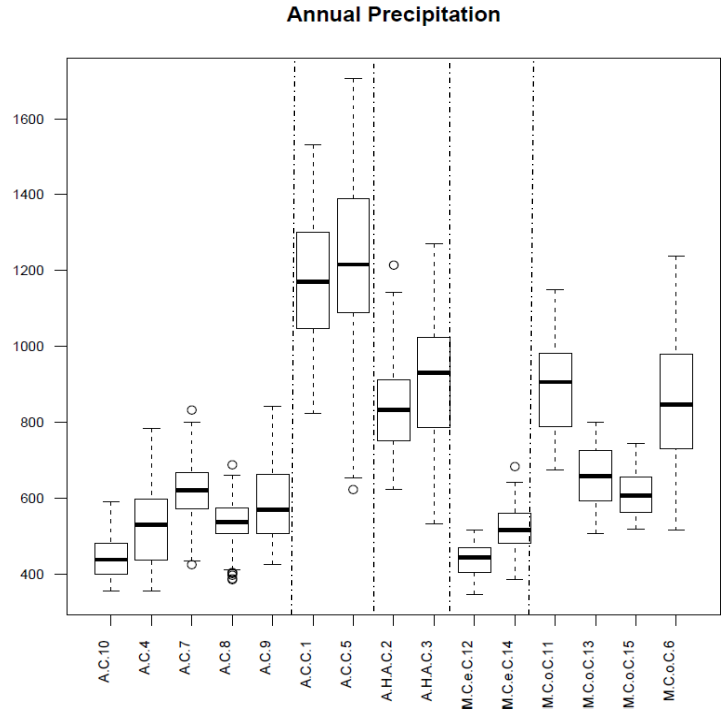
Appendix Figure 14 – Precipitation of driest month for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



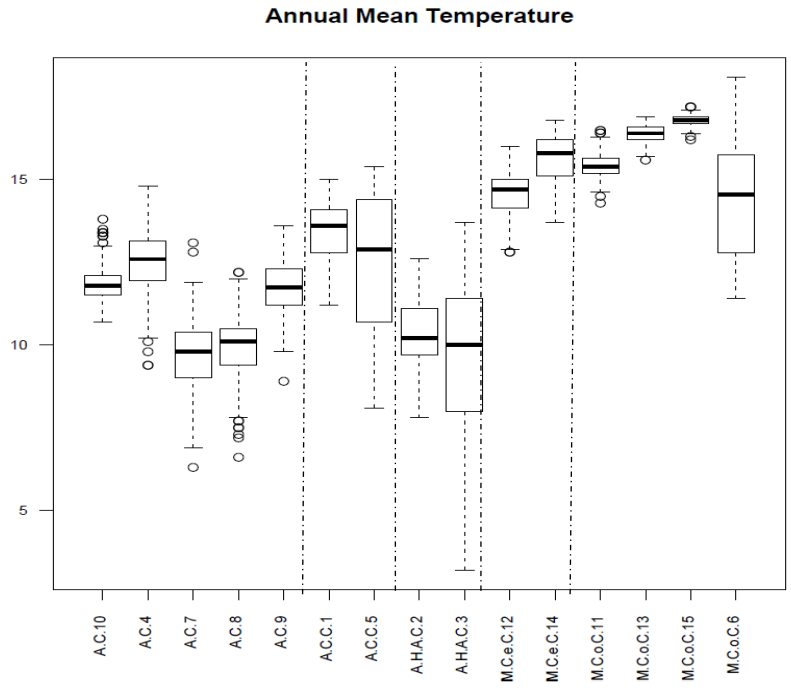
Appendix Figure 15 – Precipitation of seasonality for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



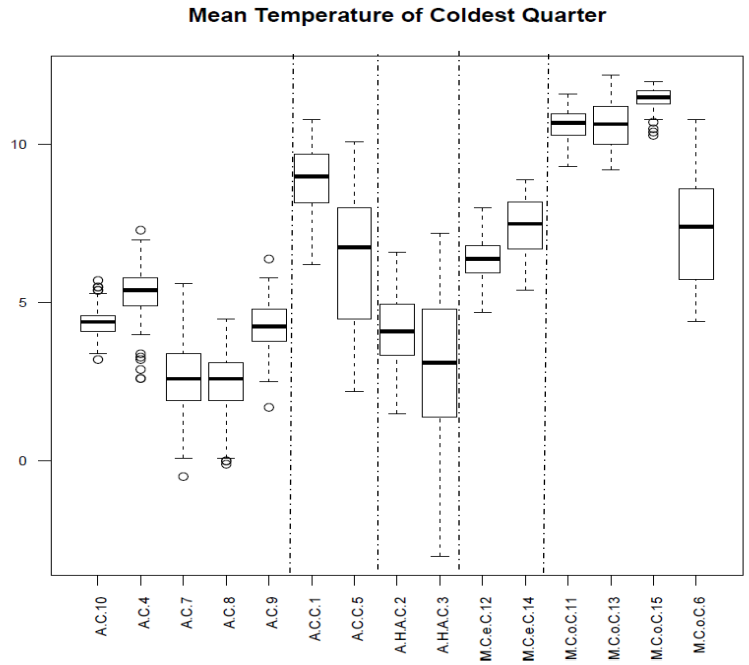
Appendix Figure 16 –Minimum temperature of coldest month for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



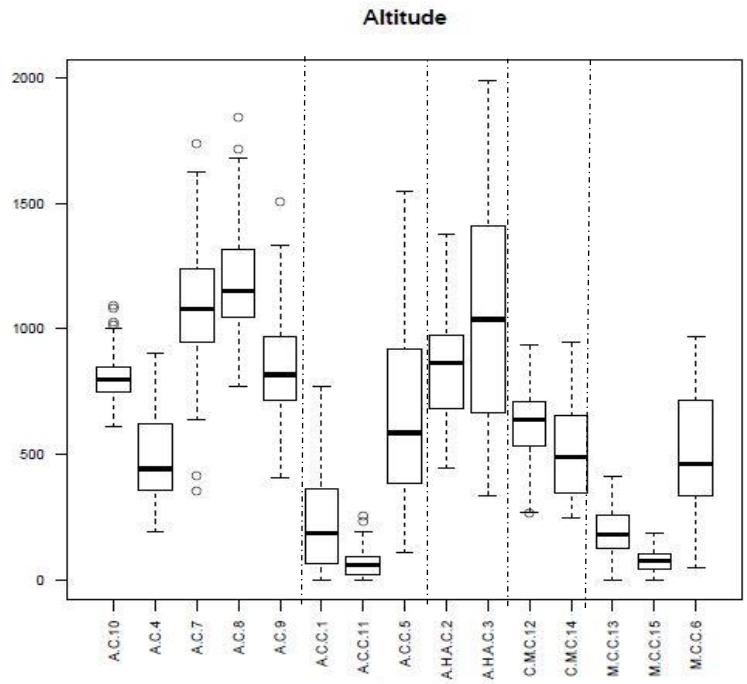
Appendix Figure 17 –Annual precipitation for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



Appendix Figure 18 –Annual mean temperature for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.

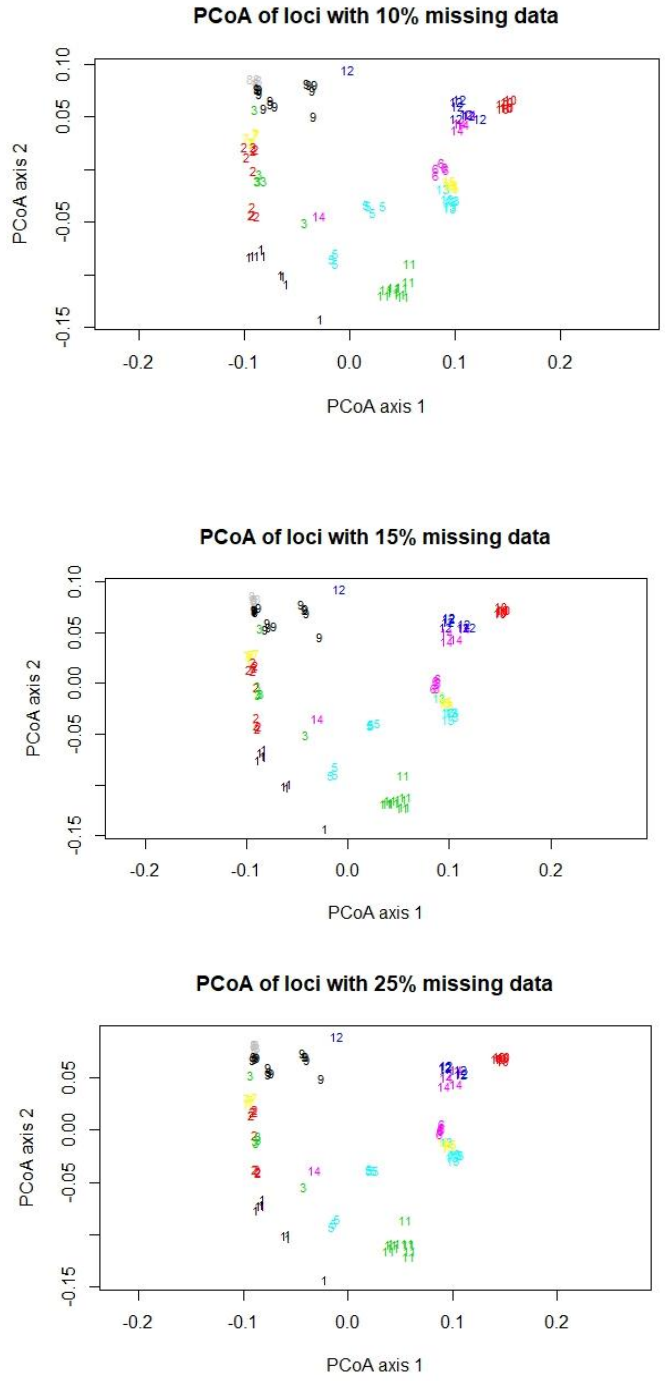


Appendix Figure 19 –Annual mean temperature for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.



Appendix Figure 20 – Altitude for each climatic cluster. Each cluster is identified by the climatic area it belongs to followed by its number. The dotted line separated the climatic areas. A.C. = Atlantic Climate; A.C.C. = Atlantic Coastal Climate; A.H.A.C. = Atlantic High Altitude Climate; C.M.C = Coastal Mediterranean Climate; C.M.C. = Central Mediterranean Climate; M.C.C. = Mediterranean Coastal Climate.

2. PCoA Additional graphics



Appendix Figure 21 – Additional PCoA results for the datasets with different cutoff levels of missing data at the locus and individual levels. Numbers represent each individual cluster. Only the first two dimensions of the PCoA were plotted

Appendix Figure 21 – Additional PCoA results for the datasets with different cutoff levels of missing data at the locus and individual levels. Numbers represent each individual cluster. Only the first two dimensions of the PCoA were plotted

3. sPCA AXIS SCORES

Appendix Table 1 - First global axis score for each sample of *H. molleri*. The sample from Rio Esteras is highlighted. See chapter "4.4.2. Spatial Analysis of Principal Components (sPCA)" for details.

Sample ID	Locality	Axis 1	Sample ID	Locality	Axis 1
68544	Buenache	15.81709	69017	Cuchía	13.66563
68545	Buenache	16.731	80737	Xinzo de Limia	-0.45112
68546	Buenache	17.14358	80741	Xinzo de Limia	-0.54468
68547	Buenache	17.50603	IMS 4702	VillaViciosa	9.766795
68548	Buenache	16.90742	69027	VillaViciosa	8.099631
68457	Boceguillas	17.54005	69028	VillaViciosa	9.264067
68458	Boceguillas	18.07669	11069	Cambre	1.731203
68454	Boceguillas	16.85915	80729	Paramos	-11.0284
68455	Boceguillas	17.28822	80730	Paramos	-10.1507
68456	Boceguillas	15.5653	80731	Paramos	-10.9152
68486	Ferreras de Arriba	12.91448	80732	Paramos	-10.993
68488	Ferreras de Arriba	14.96193	80374	Arzila	-13.7062
IMS 4217	Codesal	13.45538	80375	Arzila	-12.6195
IMS 4223	Codesal	14.15052	80376	Arzila	-14.1857
GVA6459	Salamanca	8.395311	80377	Arzila	-13.1439
GVA6457	Salamanca	7.109705	68479	Casa Velha	-20.5534
GVA6460	Salamanca	7.458233	68472	Casa Velha	-20.2242

GVA6458	Salamanca	6.316642	68475	Casa Velha	-20.3371
68520	Ólvega	18.01611	68477	Casa Velha	-20.1039
68521	Ólvega	18.77251	68478	Casa Velha	-20.3972
68522	Ólvega	18.2549	GVA7411	Coruche	-15.9511
68523	Ólvega	17.18078	GVA7412	Coruche	-17.2968
69049	Ruesga	17.24525	GVA7413	Coruche	-18.1214
69050	Ruesga	17.00669	68484	S. J. de Lamarosa	-15.3291
69051	Ruesga	16.75228	GVA7394	Beira	-15.489
69052	Ruesga	17.16304	GVA7395	Beira	-15.2558
69053	Ruesga	17.17578	GVA7396	Beira	-17.0983
68332	Valgañon	15.91632	GVA7392	Beira	-16.2312
68333	Valgañon	14.52882	80391	Torrefresneda	-15.3553
68334	Valgañon	15.37102	68988	Rio Esteras	-18.8729
11078	Cernégula	16.49279	68991	Rio Esteras	3.437604
IMS 3929	Urbasa	15.72498	80383	Navas de Estena	-16.199
68377	Urbasa	15.27307	80384	Navas de Estena	-16.4343
68378	Urbasa	14.23313	80379	Navas de Estena	-17.1339
68379	Urbasa	15.29894	68982	Saceruela	-19.7926
80723	Monforte de Lemos	5.496121	68983	Saceruela	-20.4074
80743	Serra da Estrela	-6.3842	68984	Saceruela	-20.257
80744	Serra da Estrela	-5.96778	68958	Villaverde	-25.5602

80746	Serra da Estrela	-7.11774	68960	Villaverde	-25.877
80742	Serra da Estrela	-6.19458	68961	Villaverde	-26.5245
69014	Cuchía	13.77426	68955	Villaverde	-25.3232
69015	Cuchía	14.1015	68956	Villaverde	-25.3887
69016	Cuchía	14.97707			

Appendix Table 2 - Second global axis score for each sample of *H. molleri*. The sample from Rio Esteras is highlighted. See chapter "4.4.2. Spatial Analysis of Principal Components (sPCA)" for details.

Sample ID	Locality	Axis 1	Sample ID	Locality	Axis 1
68544	Buenache	14.02751	69017	Cuchía	-12.6853
68545	Buenache	14.46042	80737	Xinzo de Limia	-13.165
68546	Buenache	14.11715	80741	Xinzo de Limia	-14.6825
68547	Buenache	14.55413	IMS 4702	VillaViciosa	-18.1888
68548	Buenache	13.92157	69027	VillaViciosa	-15.7239
68457	Boceguillas	12.1145	69028	VillaViciosa	-17.5291
68458	Boceguillas	11.97092	11069	Cambre	-24.3471
68454	Boceguillas	11.8348	80729	Paramos	-18.9327
68455	Boceguillas	11.33519	80730	Paramos	-19.0245
68456	Boceguillas	10.79459	80731	Paramos	-19.6020
68486	Ferreras de Arriba	8.571843	80732	Paramos	-20.0683
68488	Ferreras de Arriba	8.825717	80374	Arzila	-17.0273
IMS 4217	Codesal	8.257043	80375	Arzila	-15.0212

IMS 4223	Codesal	7.493977	80376	Arzila	-17.1884
GVA6459	Salamanca	12.11973	80377	Arzila	-16.0437
GVA6457	Salamanca	11.9425	68479	Casa Velha	-3.546
GVA6460	Salamanca	11.2497	68472	Casa Velha	-3.62318
GVA6458	Salamanca	10.75547	68475	Casa Velha	-4.01654
68520	Ólvega	4.013214	68477	Casa Velha	-3.3239
68521	Ólvega	3.612004	68478	Casa Velha	-4.53916
68522	Ólvega	3.633747	GVA7411	Coruche	-1.24434
68523	Ólvega	4.024227	GVA7412	Coruche	-2.34947
69049	Ruesga	1.579503	GVA7413	Coruche	-2.33416
69050	Ruesga	0.753068	68484	S. J. de Lamarosa	-2.43607
69051	Ruesga	0.926906	GVA7394	Beira	0.388397
69052	Ruesga	1.086307	GVA7395	Beira	1.63622
69053	Ruesga	0.708783	GVA7396	Beira	-0.16685
68332	Valgañon	-2.7211	GVA7392	Beira	0.793497
68333	Valgañon	-2.45912	80391	Torrefresneda	7.367907
68334	Valgañon	-1.29527	68988	Rio Esteras	8.468941
11078	Cernégula	-2.40476	68991	Rio Esteras	-4.9057
IMS 3929	Urbasa	-8.33439	80383	Navas de Estena	12.75937
68377	Urbasa	-7.99998	80384	Navas de Estena	12.11897

68378	Urbasa	-6.63373
68379	Urbasa	-8.00428
80723	Monforte de Lemos	-9.61719
80743	Serra da Estrela	-7.05722
80744	Serra da Estrela	-6.30091
80746	Serra da Estrela	-6.54653
80742	Serra da Estrela	-6.06483
69014	Cuchía	-13.6026
69015	Cuchía	-12.7807
69016	Cuchía	-14.0655

80379	Navas de Estena	12.731
68982	Saceruela	11.47113
68983	Saceruela	11.66471
68984	Saceruela	11.1402
68958	Villaverde	18.08927
68960	Villaverde	17.97227
68961	Villaverde	17.7732
68955	Villaverde	17.95621
68956	Villaverde	16.58767