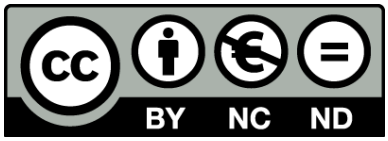
UNIVERSITAT DE BARCELONA

Evolutionary history of two Iberian Soricomorpha: genomics, phylogeography and dispersal patterns

Historia evolutiva de dos soricomorfos ibéricos: genómica, filogeografía y patrones de dispersión

Marina Querejeta Coma



Evolutionary history of two Iberian Soricomorpha: genomics, phylogeography and dispersal patterns

Marina Querejeta Coma

Front cover: Virginia Gutiérrez Río
Graphic design: Laia Vinsac
Barcelona 2017

“Nature is not a place to visit. It is home.”

Gary Snyder

A mis padres.

EVOLUTIONARY HISTORY OF TWO IBERIAN SORICOMORPHA: GENOMICS, PHYLOGEOGRAPHY AND DISPERSAL PATTERNS.

HISTORIA EVOLUTIVA DE DOS SORICOMORFOS IBÉRICOS: GENÓMICA, FILOGEOGRAFÍA Y PATRONES DE DISPERSIÓN.

TESIS DOCTORAL

PROGRAMA DE DOCTORADO EN GENÉTICA

MEMORIA PRESENTADA POR

MARINA QUEREJETA COMA

PARA OPTAR AL GRADO DE DOCTOR PARA LA UNIVERSITAT DE BARCELONA

TRABAJO REALIZADO EN EL INSTITUT DE BIOLOGIA EVOLUTIVA (CSIC-UNIVERSITAT POMPEU FABRA)

BARCELONA 2017

DOCTORANDO	DIRECTOR DE LA TESIS	TUTOR DE LA TESIS
MARINA QUEREJETA COMA	JOSÉ CASTRESANA VILLAMOR INVESTIGADOR CIENTÍFICO DEL CSIC INSTITUT DE BIOLOGIA EVOLUTIVA (CSIC-UNIVERSITAT POMPEU FABRA)	MARTA RIUTORT LEÓN PROFESOR TITULAR UNIVERSITAT DE BARCELONA

AGRADECIMIENTOS

En primer lugar, me gustaría agradecer a José Castresana el haberme ofrecido la oportunidad de realizar la tesis en su grupo en el Instituto de Biología Evolutiva. Además, quiero agradecer a Marta Riutort su papel de tutora.

Quiero dar las gracias al primer P59 por acogerme como una más: a Ana, por llevarme a esa terraza tan bonita, a Javi, por enseñarme con tanta paciencia a trabajar en el laboratorio y por haberte convertido en un gran amigo a pesar de la distancia, a Guifré, por lo que nos hemos reído y a Joan, por supuesto, aunque de ti hablaré más adelante. También quiero agradecer al nuevo P59, por todas las horas y horas que hemos pasado juntos: a Lidia, a Karla, a Oliver y a Alfonso. No hubiese sido lo mismo trabajar en el laboratorio sin los cafés y las cervezas. Gracias al laboratorio de las mariposas, a Gerard y a Vlad, y sobretodo a Raluca, por nuestro viaje a Cádiz y por todas las fiestas, y las que nos quedan.

Muchas gracias al resto de compañeros del IBE: a Adrián, porque estaba claro que necesitábamos uno más del Norte en el grupo, a Marc, por nuestros gins y lo bien que lo pasamos en O Porto, a David, por enseñarnos los mejores bares de Barcelona y por los vermús que se convierten como por arte de magia en cenas, a Ana Fernández, porque los últimos años de tesis no hubiesen sido lo mismo sin tu risa contagiosa. A Luís y a Ana Trindade, por ser tan buenos amigos. Gracias a Pedro, porque, aunque no hayas estado mucho tiempo, me has ayudado un

montón y siempre con paciencia y una sonrisa. Y no me puedo olvidar de Helena, de Duarte, de Joao, de Susana, de Claudia, de Adriana y, por supuesto, de David Sánchez. Muchas gracias a todos.

Gracias a Aína y a Andy, por ser los mejores compañeros de piso y a Paloma, por acogerme tan bien nada más llegar a Barcelona. Y también a Laia (por maquetarme la tesis aunque tengas tantas y tantas cosas qué hacer), a Pere, y a Gael por ser ya de la familia.

No me puedo olvidar de mis amigos de Gijón, porque cada vez que vuelvo parece como si nunca me hubiese ido. En especial a Ochoa, a Laura, a Noemí, a Chema, a Marta, a Miguel y, por supuesto, a Viky.

Muchas gracias a María Atencia y a Ina Klaumann porque fue con vosotras dónde me empezó a interesar la conservación, siempre me acordaré de los visones y de Berschweiller.

Muchas gracias a Joan, por todo lo que me has ayudado en esta tesis, en cada pequeña parte. Eres un científico excelente y serás un supervisor magnífico. Pero sobretodo, gracias por Borneo, por Australia, por Escocia, por Suiza, por Londres, por París y, por todos los sitios que nos quedan por conocer juntos. Gracias por haber aparecido, enseñarme cómo son las tormentas del Pirineo y cambiar mi vida.

Gracias a mi hermano Agus por “our great English conversations” y gracias a Pepín, por dejarme un hueco en su habitación cada vez que vuelvo a casa.

Por último y lo más importante, gracias a mis padres, porque son los únicos responsables de que haya llegado hasta aquí y de que todos podáis a leer esta tesis. Gracias por todo.

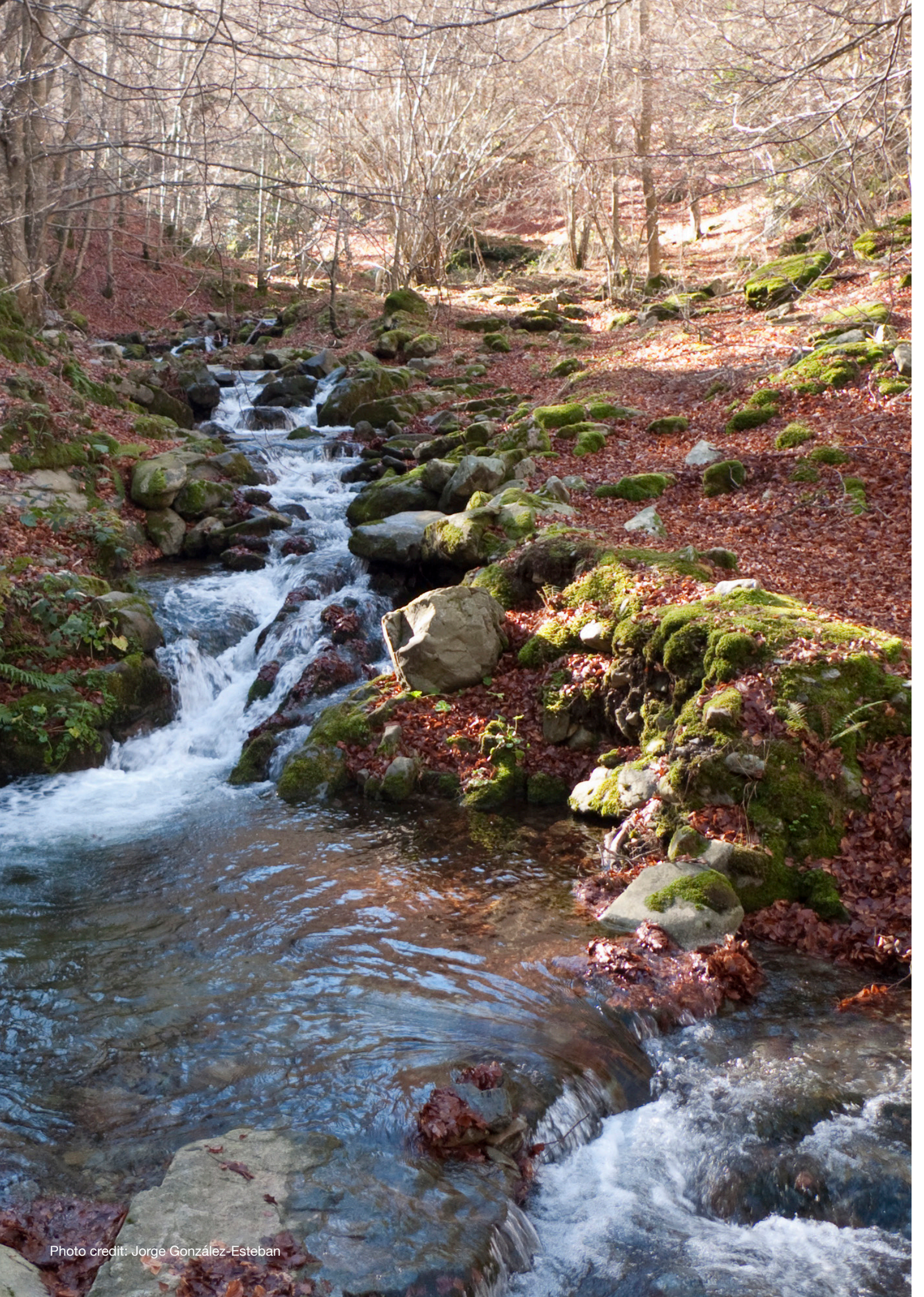


Photo credit: Jorge González-Esteban

CONTENTS

Photo credit: Jorge González-Esteban

OBJECTIVES

The Iberian Peninsula contains a large number of endemic species with complex evolutionary histories. There are still many histories and patterns to be revealed within this region. The study of the evolution of these species could help in the conservation and preservation of intraspecific diversity.

The main aim of this thesis is to shed light on the evolutionary history of two Iberian endemic semi-aquatic mammals: the Pyrenean desman *(Galemys pyrenaicus)* and the Mediterranean water shrew *(Neomys anomalus)*.

There are three specific objectives in this thesis:

1. Study the population structure of the Pyrenean desman using a genomic approach, modifying and optimizing different protocols for their use with samples from endangered species. More especifically, the aim was to evaluate and filter the genomic libraries obtained, determine the sex of the individuals, estimate the individual heterozygous rate and infer the genomic structure using SNP data.

2. Delimit the contact zone between the main mitochondrial clades of the Pyrenean desman in the Cantabrian Mountains and study the dispersal patterns of one of these clades using an extensive and fine-scale sampling. It was also intended to study the structure of genetic diversity within this clade and reveal if river basins affected the genetic structure.

3. Study the evolutionary history of the Mediterranean water shrew. Specifically, the objective was to infer the genetic structure and explore if it was correlated with geography. Also, the structure of genetic diversity and the species distribution modelling would help to infer the main glacial refugia during the Last Glacial Maximum.

Photo credit: Peter Porta

INTRODUCTION

"A thing is right when it tends to preserve the integrity, stability and beauty of the biotic community. It is wrong when it tends otherwise" Aldo Leopold. **A Sand County Almanac: And Sketches Here and There.** 1949

Conservation Biology as a discipline appeared in the 80's as an answer of many scientist to the great loss of biodiversity (Simberloff, 1988). However, by that time many scientists and naturalists had already dedicated entire lives to preserve species and ecosystems.

This is the case of Aldo Leopold (1887-1948), an American environmentalist, scientist and writer, who is considered one of the greatest conservation thinkers of all time. He developed the idea of humans needed to achieve a responsible relationship with the land they inhabit, what he later named the "land ethics". In his book **"A Sand County Almanac: And Sketches Here and There",** he embraces the idea that people should treat the environment with appreciation and understanding, not only for its own sake, also for ethical reasons and for the good of future generations. Leopold changed many ways of thinking and inspired many environmentalists movements in the 20th century (Cohen, 1988).

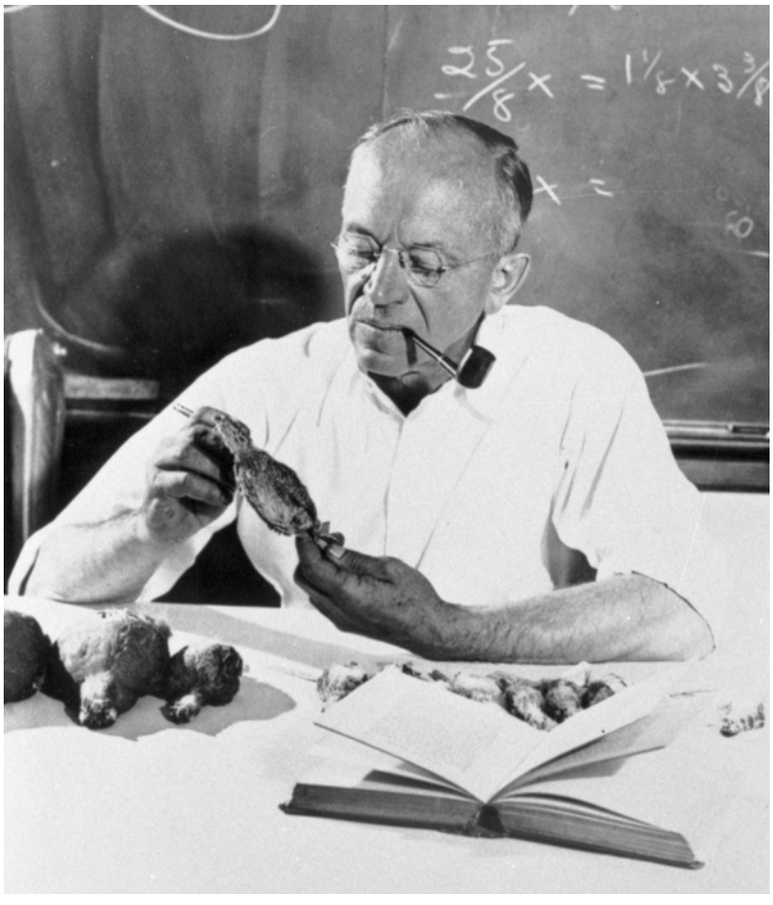


Figure 1. Aldo Leopold cataloguing birds. Photo Credit: The Aldo Leopold Foundation. https://www.aldoleopold.org

Although more than half a century has gone by since Leopold died, a lot of research in conservation biology is still waiting to be done. In the last decades, evolutionary biology in general and genetics in particular have added new theoretical and methodological tools to conservation and preservation of species and ecosystems. Knowing the species helps to protect them, and particularly knowing the evolutionary history of a species is crucial to develop sound conservation plans (Crandall *et al.*, 2000; Carvalho *et al.*, 2011). For instance, the genetic structure of a species helps in the delimitation of conservation units, and studies of gene flow and genetic diversity allow the possibility to prioritize resources on the populations more susceptible to enter extinction vortices (Frankham *et al.*, 2002).

With this conservation goal in mind, the aim of this thesis is to understand the evolutionary history of two Iberian endemisms: the Pyrenean desman *(Galemys pyrenaicus)* and the Mediterranean water shrew *(Neomys anomalus)*, using several approaches such as genomics, phylogeography and species distribution modelling.

1. INTRASPECIFIC EVOLUTIONARY HISTORIES

The study of population genetics and the structure of genetic diversity within species are of great interest in ecology and evolutionary biology. First, these studies deal mainly with fine-scale patterns, which in broader-scale or macroecological studies could hardly be detected. Also, the findings and results concerning intraspecific genetics are essential for conservation planning. For these purposes, the science of phylogeography, first conceived in 1987 by John C. Avise, is essential (Avise, 2000). More recently, next-generation sequencing (NGS) has resulted in a great leap in evolutionary biology, making possible the acquisition of large amounts of genome-wide molecular markers in short periods of time in an inexpensive way (Mardis, 2008, 2013; Davey et al., 2011; Glenn, 2011).

1.1. PHYLOGEOGRAPHY IN THE STUDY OF INTRASPECIFIC EVOLUTIONARY HISTORIES

Phylogeography is a multidisciplinary field in which ecology, genetics, phylogenetics, geography, natural history and statistics converge. The main aim of phylogeography is to study the spatial structure of genetic lineages in space and time, mainly among closely related species and within the same species (Avise et al., 1987; Avise, 2000).

Initially, phylogeographic studies used only mitochondrial (mt) DNA to study the relationships of individuals within the same species through a common ancestor and to understand how genetic structure was correlated with geographic structure. Mitochondrial DNA was, and still is, used in phylogeographic studies because it is easy to isolate and amplify in the laboratory, being a great advantage especially when dealing with a large amount of specimens or with low quality samples,

such as non-invasive samples. Also, mtDNA has fast evolutionary rates in animals (Brown et al., 1979), which is essential to detect genetic variability in populations and closely related species. Moreover, mtDNA is of maternal inheritance and lacks recombination (Hutchison et al., 1974).

A large amount of mitochondrial phylogeographic studies have been extremely helpful to identify evolutionary distinct populations among and within species, reveal spatial patterns of species richness, identify evolutionary isolated areas and infer dispersal patterns which took place after glaciations (Bermingham & Moritz, 1998). Although mtDNA is still broadly used in phylogeographic studies, nuclear genes, of both maternal and paternal inheritance, are necessary to study gene flow and detect sex biased genetic structure (Hare, 2001).

1.2. THE ERA OF GENOMICS

The appearance of Next Generation Sequencing (NGS) methods to perform genome-wide studies has revolutionized the field of phylogenetics and phylogeography (McCormack et al., 2013). This new methodology produces massive amounts of genetic data, which constitutes a great computational challenge (Emerson et al., 2010; Turmelle et al., 2011; McCormack et al., 2012). New computational methods and bioinformatic pipelines must be developed. Thus, instead of being in the era of Genomics, it can be said that scientist are living in what could be called “the era of post-genomics” (Agrafiotis et al., 2002), where bioinformaticians work against the clock to be able to analyse all the sequences generated.

1.2.1. NEXT GENERATION SEQUENCING LIBRARY PREPARATION METHODS AND PLATFORMS

Next Generation Sequencing (NGS) is also called high-throughput sequencing and it is used to describe several modern sequencing technologies which allow a large amount of samples to be sequenced at once. However, this requires library preparation before sequencing. Among the “wet-lab” methods of library preparation currently available are the following (McCormack et al., 2013):

· **Multiplex PCR and amplicon sequencing:** in this method, samples are amplified by traditional PCR before sequencing in a NGS platform (Mayer-Scholl et al., 2010).

· **Restriction digest-based methods:** this method uses restriction enzymes to digest the genome in fragments. After, these fragments are selected by size to obtain a reduced representation library, that is, a set of variable genome-wide fragments. There is a specific restriction digest-based method of preparing NGS sequencing libraries, which is widely used in phylogeny and phylogeography and it is called Restriction-site Associated DNA (RAD) sequencing (Baird et al., 2008). This technique is different from other restriction digest-based methods in the control of the

digested fragments, which are marked with adaptors (short sequences of nucleotides to identify the specimens) before sequencing. In this thesis, double-digest Restriction-site Associated DNA (ddRAD) (Fig.2) is used (Peterson et al., 2012), being the DNA digested into short fragments by two different restriction enzymes, EcoRI and Msp II (See **Chapter 1**).

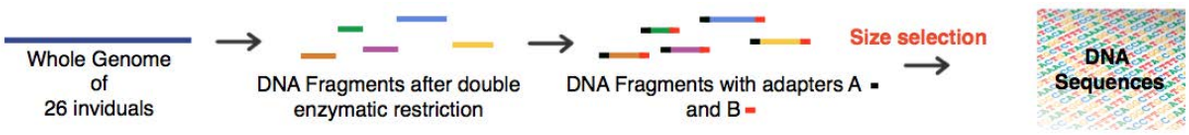


Figure 2. Scheme from ddRADSeq laboratory protocol.

· **Target enrichment:** this method captures selected genomic regions before NGS sequencing (Mamanova et al., 2010).

· **Transcriptome sequencing (RNA-seq):** in this method a population of RNA (total or fractioned) is transformed into cDNA fragments and linked to adaptors. Every molecule is then sequenced (Marioni et al., 2008).

Regarding sequencing platforms, Illumina, Roche, SOLiD and Ion Torrent are the most widely used nowadays. These platforms mainly differ in the length of reads sequenced and, also, in the sequencing method (Mardis, 2008, 2013; Davey et al., 2011; Glenn, 2011).

The reason to use one or other platform lies in the type of scientific question. For instance, long-reads can be used for initial genome assembly. By contrast, short-reads are widely used for population genetics or gene expression studies (Glenn, 2011).

1.2.2. DATA ANALYSIS

Next Generation Sequencing (NGS) data is different from traditional Sanger sequencing data mainly in the quantity of the data to be analysed, as NGS platforms can produce large amount of markers per specimen. Also, the quality of the final sequences obtained is extremely important. This quality is measured by a parameter, the coverage or depth, which is the number of sequences which contain a given nucleotide in the final reconstructed sequence (Sims et al., 2014).

The first step to analyse NGS data is filtering and removing the sequences with low coverage (Oliver et al., 2010). Then, sequences have to be detected by their tag o adatptor. Lastly, the homologous sequences are aligned for genotyping or, what is called, SNPs calling. These SNPs could be used to infer phylogeographies performing analyses such as Principal Component Analysis (PCA), Structure and distance-based phylogenetic trees (Nielsen et al., 2011).

With NGS data and the subsequent analyses, evolutionary histories can be inferred more accurately than with traditional markers (Falush et al., 2003), as genetic divergences between populations can be detected with a high level of detail. There are several evolutionary histories, previously studied with traditional markers, which have been revisited using genomic markers to obtain more detailed information. For instance, the population structure of the genera of North American *Lycaeides* butterflies was studied using genomic markers, revealing its evolutionary history and its Pleistocene glacial refugia, demonstrating the potential of genomic studies in ecology and evolution (Carr et al., 2008).

Still, traditional mitochondrial phylogeography is very useful for non-invasive samples and research concerning endangered species. This was indeed the case of the endangered South-American marine otter (*Lontra felina*), study in which this population was tracked using scats. This study provides information for reassessing the conservation status of this species, proving the potential of non-invasive sampling for conservation management (Valqui et al., 2010)

2. INFERRING GLACIAL REFUGIA AND DISPERSAL PATTERNS WITHIN SPECIES

Glacial refugia and past dispersal patterns are essential to understand the evolution of population throughout time. Hence, there is a need of developing methodologies that could reveal the most plausible scenarios.

2.1. GLACIAL REFUGIA DURING THE LAST GLACIAL MAXIMUM (LGM)

During the Quaternary (2.6 Myr ago to present), several species distribution ranges were shaped (Provan & Bennett, 2008). They were the result of changes in climate and, consequently, multiple dispersal processes. Within the Quaternary, all along the Last Glacial Maximum (LGM) (Fig.3), 25,000 to 17,000 years ago (Yokoyama et al., 2000), the worldwide ice-sheets reached their maximum volume and extension in the recent history, mainly due to a significant change in climate, in which mean temperature decreased and snow storms increased (Clark et al., 2009). Hence, several species distribution ranges were reduced or displaced (Provan & Bennett, 2008) and were isolated in glacial refugia (Petit et al., 2003), where the climatic and vegetation conditions enabled to survive (Brewer et al., 2002; Gordillo et al., 2005; Cheddadi et al., 2006; Médail & Diadema, 2009; Tian et al., 2009; Igea et al., 2013). This resulted in long-term isolated and differentiated populations. After this glaciation period, some of these populations became extinct or, conversely, expanded through postglacial colonizations. These colonizations of differentiated populations have given rise to most of the current Palearctic intraspecific diversity (Petit et al., 2003). Understanding why these past

glacial refugia were located in certain spots is key to predicting the effects of the climate change the Earth is currently experiencing on species and ecosystems (Provan & Bennett, 2008).

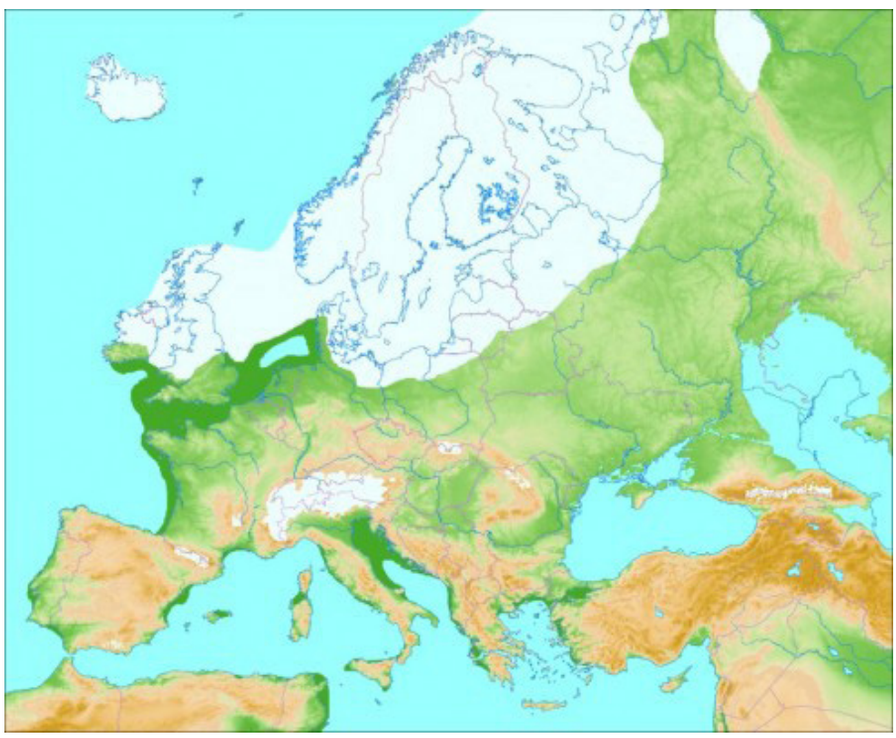


Figure 3. Last glaciation in Europe.
Photo credit: Wikimedia Commons (user Ulamm)

2.1.1. THE IBERIAN PENINSULA: A SPECIES "RESERVOIR" DURING THE LAST GLACIAL MAXIMUM

Many studies agree that Southern Europe harboured most of the species in glacial refugia during the Last Glacial Maximum (LGM), recolonizing during the late Pleistocene the north part of Europe (Hewitt, 1999, 2000, 2001, 2004; Feliner, 2011). Hence, the Iberian Peninsula, together with other southern European peninsulas, such as Italy and Greece, was one of the most important glacial refugia in Europe, principally because of its wide range of different climates and habitats in a relative small area. It is therefore a great model to study and compare phylogeographic patterns (Hewitt, 1999, 2000; Gómez & Lunt, 2007). As a matter of fact, the Iberian Peninsula shows a large number of endemic and native species, which is consistent with population isolation, survival, differentiation and speciation during long periods of time (Gómez-Campo et al., 1984). Within the Iberian Peninsula, several locations consistent with multiple glacial refugia are known to have existed. This hypothetical scenario with more than one glacial refugia within the Iberian Peninsula is called "refugia within refugia" (Gómez & Lunt, 2007). This hypothesis suggests that the Iberian Peninsula was not only a big reservoir of species that were able to recolonize northern Europe in the warmer postglacial periods but also a group of multiple refugia, which resulted in a great instraspecific diversification, a complex postglacial scenario and, eventually, several extant endemic species. However, how the rest of Europe was recolonized from Iberia is still a matter of debate (Feliner, 2011).

This hypothesis of “refugia within refugia” has opened a promising path of study regarding evolution and comparative phylogeography. Although several research have focused on phylogeography of Iberian endemic and native species (Consuegra et al., 2002; Martínez-Solano et al., 2006; Pinho et al., 2007; Gonçalves et al., 2009; Stamatis et al., 2009; Igea et al., 2013), there are still many unknown evolutionary histories whose study could lead to a better understanding of the Paleartic biogeography. One of the unknown histories is the Mediterranean water shrew *(Neomys anomalus)* phylogeography, which is a topic of this dissertation (See **Chapter 3**). Also, it is an objective of this thesis to study in depth the northwestern glacial refugia for the Pyrenean desman and the mode of postglacial dispersal (**Chapter 2**).

2.1.2. INFERRING GLACIAL REFUGIA AND POSTGLACIAL COLONIZATION ROUTES

The inference of intraspecific genetic structure through phylogeographic studies has helped to discover new glacial refugia and postglacial colonization routes within Europe. There are many examples to illustrate the importance of these approaches. One noted study is the one regarding the Scots pine *(Pinus sylvestris)* along Europe, which revealed that the distribution of genetic diversity is consistent with long periods of isolation within the glacial refugia (Cheddadi et al., 2006).

Based on the premise that potential glacial refugia locations are consistent with high levels of nucleotide diversity and the areas further away from them are consistent with low levels of genetic diversity (Provan & Bennett, 2008), it is possible to infer glacial refugia interpolating genetic diversity (Igea et al., 2013). Moreover, the distribution of lineages within the distribution range using phylogenetic inference can delimit the potential postglacial colonization routes. Although traditional phylogeography is essential to infer glacial refugia, it has several limitations when used on its own. For instance, phylogeography restricts the spatial context of the glacial refugia to geographic distances, loosing the perspective of the multivariate space (Graham et al., 2004). Also, traditional phylogeography cannot identify glacial refugia outside the boundaries of the distribution range (Willerslev et al., 2007). To overcome these limitations, combining traditional phylogeography with species distribution modelling (SDM) provides a broader picture of more plausible evolutionary scenarios (Guisan & Zimmermann, 2000).

SDM uses environmental variables and field observations (occurrences and/or absences) to predict past, current and future species distribution ranges. The environmental variables can provide landscape, climate or abiotic information, among others. This ecological and spatial data is incorporated in one or more modelling methods and the result is used to predict the distribution range of the species under the chosen climate scenario. When projecting the distribution models into past climate scenarios, areas suggested to have been optimal climates are consistent with potential locations of ancient glacial refugia, for instance, during the Last Glacial Maximum (Guisan & Zimmermann, 2000; Richards et al., 2007; Kozak et al., 2008; Pearman et al., 2008; Elith

& Leathwick, 2009; Gavin et al., 2014). Thus, SDM, together with phylogeographic and genetic diversity information, enables the inference of glacial refugia and evolutionary histories. In fact, many studies have worked at the interface between phylogeography and SDM to perform biogeographical studies, obtaining highly interesting evolutionary and ecological results (Benito Garzón et al., 2007; Hoarau et al., 2007; Cordellier & Pfenninger, 2009; Wielstra et al., 2013).

2.2. PAST DISPERSAL PATTERNS

Dispersal is any movement that could lead to gene flow between populations (Ronce, 2007) and it plays an essential role in evolution, as it is responsible for the distribution of genetic diversity across space (Wright, 1969). Moreover, dispersal can keep populations from local extinction, when the source habitat or climatic conditions are not optimal (Brown & Kodric-Brown, 1977) or, in contrast, it can also increase the global extinction rate, for instance, when exotic or invasive species present high levels of dispersal, colonizing areas and displace native species. Several evolutionary processes, such as speciation (Barton, 2001), inbreeding depression (Roze & Rousset, 2003) or a large number of historical traits (Pen & Weissing, 2000) are affected by the past dispersal of species or populations. Understanding dispersal patterns is key, not only in evolution, but also in conservation management of natural populations (Macdonald & Johnson, 2001), as inferring dispersal routes of endangered species could determine corridors or habitats to preserve and restore.

Semi-aquatic animals can swim as a means of dispersing, feeding, escaping or foraging. Although they present several adaptations to aquatic environments, they still maintain the ability to disperse across or acquire food on land. This switch from one dispersal mode to another makes the study of the relative importance of each dispersal mode in these species extremely interesting from an ecological perspective. In fact, in semi-aquatic animals, depending on the way the animal disperse, river networks can act as a barrier to dispersal and, consequently to gene flow (Chaput-Bardy et al., 2008; Byrne et al., 2015; Paz-Vinas et al., 2015) or, in contrast, they can connect populations (Raeymaekers et al., 2008; Bartáková et al., 2015; Byrne et al., 2015).

Although dispersal is key in shaping evolutionary patterns, there are few studies regarding dispersal patterns in semi-aquatic animals (Furlan et al., 2013; Senn et al., 2014; Pagacz, 2016). This is because it is a problematic pattern to estimate from direct observations, as these animals are difficult to detect in the field and, also, the observations can be biased. Another methodology is the use of measures of stable isotopes and mark-recapture techniques to identify dispersal patterns and barriers to dispersal (Macneale et al., 2005). For instance, in the case study of the stonefly *(Leuctra ferruginea)*, the dispersal patterns were determined using this method, suggesting that forests did not represent a real barrier to dispersal. This method is useful but mainly for invertebrates. Another direct method useful for semi-aquatic vertebrates is radio-tracking, which is an accurate method that enables the detection of dispersal routes. It was used to study dispersal of the Pyrenean

desman, among others (Stone & Gorman, 1985; Melero et al., 2012, 2014). All these methods are extremely useful to identify current dispersal patterns in species. However, to infer past dispersal patterns and the evolution of dispersal throughout time, indirect methods should be used.

One of the indirect methods is based on the genetic structure. This is to say that geography can shape dispersal patterns of species and these dispersal patterns can be reflected in the genetic structure of populations (Manel et al., 2003). Based on this premise, one widely used approach derives from the well-known isolation-by-distance model, which is based on the assumption that genetic isolation increases with geographical distance and predicts genetic similarity at shorter distances as well as differentiation at greater distances due to limited dispersal (Wright, 1943). Using this approach, dispersal modes can be inferred through genetic data and it is possible to make inferences to the past, which are essential to understand the evolution of species throughout time.

3. THE PYRENEAN DESMAN (*Galemys pyrenaicus*) AND THE MEDITERRANEAN WATER SHREW (*Neomys anomalus*): TWO EXAMPLES OF SEMI-AQUATIC IBERIAN ENDEMISMS

This thesis uses two models of study: the Pyrenean desman *(Galemys pyrenaicus)* and the Mediterranean water shrew *(Neomys anomalus)*. Both species are semi-aquatic and endemic to the Iberian Peninsula. Hence, they are both excellent models to study their evolutionary history due to their singularity and, also, to the particular characteristics of the Iberian Peninsula.

3.1. THE PYRENEAN DESMAN *(Galemys pyrenaicus)*

The Pyrenean desman *(Galemys pyrenaicus)* is a unique mammal, not only from an ecological perspective, but also from an evolutionary and conservation point of view. This uniqueness makes it a one-of-a-kind model to explore the insights of intraspecific evolution and ecological patterns, which can be potentially applied to future conservation actions, such as translocations or reintroductions.

As it is one of the two Iberian mammals used as a model in this thesis, a review of the main features of the Pyrenean desman known so far is presented in the following section.

3.1.1. HISTORY OF THE RESEARCH ABOUT THE PYRENEAN DESMAN

The Pyrenean desman *(Galemys pyrenaicus)* (Fig. 4) is the only extant species of the genus *Galemys* and it is one out of the two members of the subfamily Desmaninae (Family Talpidae). The other species of this subfamily Desmaninae (Family Talpidae) is the Russian desman *(Desmana moschata)*

(Palmeirim & Hoffmann, 1983; Nores et al., 2007).

According to the revision of fossil and recent Desmaninae published by Cornelia Geradina Rümke (Rümke, 1985), the first author known to have described a desman was Clusius in 1605,

Figure 4. The Pyrenean desman *(Galemys pyrenaicus)*. Photo Credit: Jorge González-Esteban

naming a Russian species as a water-mole, whose scientific name was *Mus aquaticus exotica.* But it was not until 1663 when Daubentian used the name desman for the first time. Ten years later, in 1673, Charleston classifies this water-mole within Insectivora, giving it the scientific name of *Sorex moscovite*. However, in 1758 Linnaeus still considered it as a rodent. Finally, in 1777, Gueldenstaed named it as *Desmana*, including a detailed description of the animal. In 1811, Geoffroy described another similar but smaller water-mole and named it *Mygale pyrenaica* but, in 1829, Kaup distinguished this species as a different genus: *Desmana*. Finally, in 1912, Miller named the species as we know them nowadays: the Pyrenean species was *Galemys pyrenaicus* and the Russian species was *Desmana moschata* (Schreuder, 1940; Rümke, 1985).

3.1.2. GENERAL CHARACTERISTICS OF THE PYRENEAN DESMAN

The Pyrenean desman is a small semi-aquatic mammal, strongly adapted to riverine ecosystems. For instance, one of the most noticeable adaptations is its long and flexible snout (Fig.5) used to explore the river surroundings while looking for food. Another important adaptation is its posterior legs, which are much longer than the anterior legs and adapted to swimming (Palmeirim & Hoffmann, 1983; Nores et al., 2007)

Figure 5. Detail of the flexible and long snout of the Pyrenean desman. Photo credit: Biosfera Consultoría Ambiental

It is distributed from the French watershed of the Pyrenees to the northern half of Portugal, showing a patchy and discontinuous distribution (Fig.6). It inhabits mountain rivers of unpolluted waters within this distribution (Fig.7). Moreover, the Pyrenean desman is endemic to the Iberian Peninsula and, furthermore, it is considered as Vulnerable (V) by the IUCN (IUCN, 2001) and it is legally protected in the four countries where it inhabits: France, Spain, Portugal and Andorra. This status is mainly because it is a very specialized mammal, with strong ecological requirements, that makes it highly vulnerable to changes in the environment. Therefore, several threats have been described. For instance, the fragmentation of the desman populations is remarkable, as it inhabits mainly in the mountain river headwaters, which are highly disconnected between them. This makes the different populations highly isolated, reducing their dispersal capabilities (Reed & Frankham, 2003). Also, there are several anthropogenic threats, such as the dams and hydro-electric power plants within the rivers and, also, the water contamination or the river dams (Nores et al., 2007).

The Pyrenean desman is predated mainly by the Eurasian otter *(Lutra lutra)*, although some other

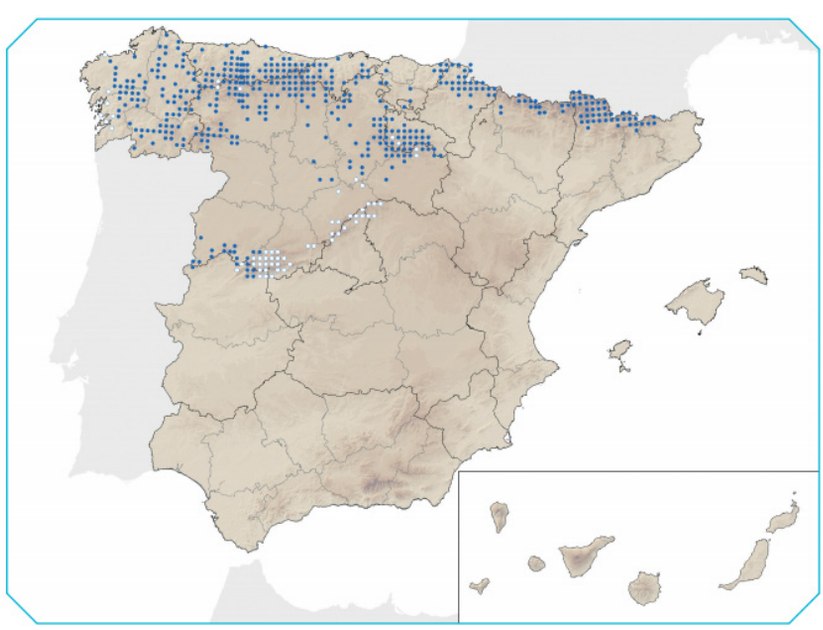


Figure 6. Distribution range of the Pyrenean desman. (Nores et al., 2007)

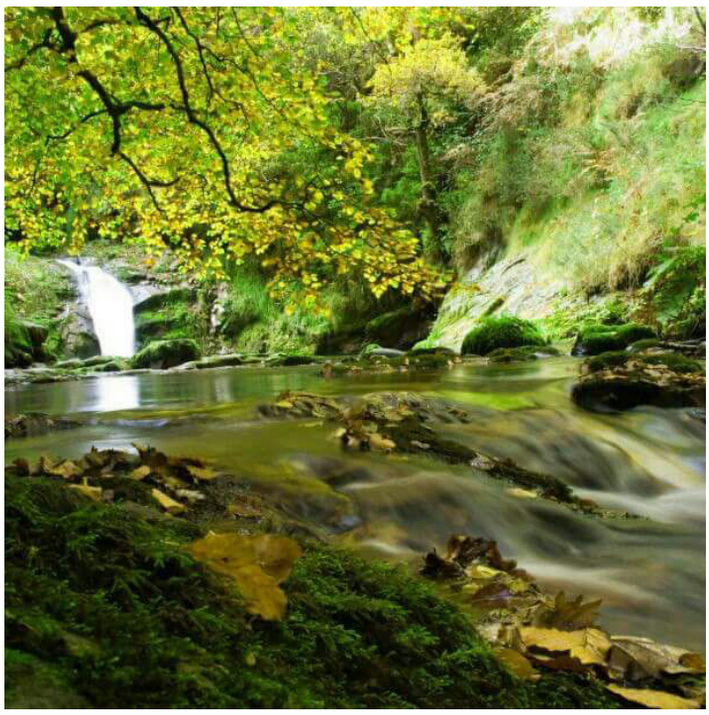


Figure 7. Typical habitat of the Pyrenean desman. Photo Credit: Pablo González

species such as the grey heron *(Ardea cinerea)* or the American mink *(Neovison vison)* can also feed on desmans (Nores et al., 2007). With respect to its diet, several studies have been conducted. First, Castién and Gosálbez performed a detailed exploration of digestive tracks of Pyrenean desmans, identifying diet species by direct analysis via binocular microscope. They concluded that the Pyrenean desman feeds mainly on larvae from Trichoptera, Diptera, Ephemeroptera and Plecoptera (Castién & Gosálbez, 1992). Later, a non-invasive approach was used to determine the diet of the desman. Desman faecal samples were analysed using NGS technologies (Gillet et al., 2015). Although they obtained similar results as in the previous study, the optimization of this approach is relevant as non-invasive techniques are preferable in endangered species.

Regarding reproduction, Pyrenean desmans have a heat period that goes from January to March and deliveries that take place from March to July. Their litters are composed of one to five individuals (Palmeirim & Hoffmann, 1983; Nores et al., 2007) In relation to life expectancy, desmans can live over five years. Although many ecological studies require age determination of the individuals, it is very difficult to verify their age accurately using external biometric parameters (length and body mass), due to the similarity in these measures between juveniles and adults (González-Esteban et al., 2002). Nevertheless, it has been proven that the age is correlated with the teeth ware, this being a useful method to determine the age of the individuals (Richard, 1976; González-Esteban et al., 2002).

G. pyrenaicus is polymorphic and divided into two subspecies, *pyrenaicus and rufulus*, although this still matter of debate (Juckwer, 1990; González-Esteban et al., 1999; López-Fuster et al., 2006). Fur coloration shows a clear geographical variability. For instance, the individuals from the Basque country and the Iberian Range show a blackish coloration, different from the individuals from the

Western Region (Asturias, Galicia and Portugal), whose coloration is light brownish or reddish (González-Esteban et al., 1999).

Hair types from the fur are a distinguishable feature in the Pyrenean desman. Three different types of hairs can be distinguished: i) the woody hairs, showing no special features, ii) the guard hairs, showing a sword shape and iii) the "Grannen", which are the most characteristic hair type in the Pyrenean desman, showing a lancet shape (Fig.8). In fact, the "Grannen" is the hair type that is used to distinguish the desman samples from other mammals (Poduschka & Richard, 1985). For instance, in faeces from predators of the desman, such as the Eurasian otter *(Lutra lutra)* or the America mink (*Neovison vison*), identifying this hair type with the optical microscope is an effective technique to identify desmans as prey.

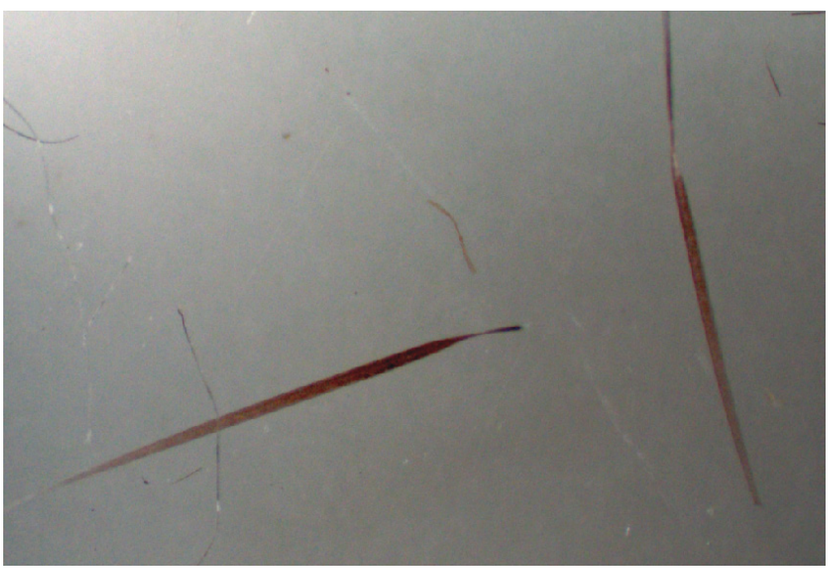


Figure 8. The "Grannen" hair type of a Pyrenean desman. Photo credit: Marina Querejeta Coma

3.1.3. SOCIAL ORGANIZATION AND ACTIVITY PATTERNS OF THE PYRENEAN DESMAN

The social organization and activity patterns of insectivorous mammals are difficult to study because these species are small and crepuscular. However, there are some studies regarding these patterns in the Pyrenean desman (Stone & Gorman, 1985; Stone, 1986, 1987a, 1987b; Melero et al., 2012, 2014).

Concerning movements and resting sites, it was suggested that they use permanent resting sites during their entire lives (Stone, 1987b) but more recent studies have arrived at the opposite hypothesis, in which the resting sites keep changing during their entire life (Melero et al., 2012). Regarding behaviour, some studies have proven their aggressiveness and solitary behaviour (Stone, 1987a) but, in contrast, others do not agree and claim that they are not aggressive and do sometimes coexist sometimes in their lives (Melero et al., 2012). Both studies agree in the location of their resting sites, being submerged or semi-submerged (Stone, 1987b; Melero et al., 2012).
The Pyrenean desman is endangered and endemic, being extremely valuable from a conservation

point of view. Moreover, some of its population is known to be decreasing over recent years and in high risk of extinction, specially the southern populations located in the Central System (Nores et al., 2007). Hence, more research must be carried out with live-camera traps or indirect genetic methods to describe or detect movement or dispersal patterns. This way, corridors and optimal areas could be detected.

3.1.4. SYSTEMATICS AND INTRASPECIFIC MITOCHONDRIAL PHYLOGEOGRAPHY OF THE PYRENEAN DESMAN

As stated before, the Pyrenean desman is one of the two extant species of the subfamily Desmaninae, together with the Russian desman *(Desmana moschata)*. From a systematic perspective, the subfamily Desmaninae belongs to the family Talpidae which, in turn, is included in the order Eulipotyphla (insectivorous mammals). Aside from *Desmana* and *Galemys*, during the Miocene, Pliocene and Pleistocene, there was another genus belonging to the subfamily Demaninae, which was *Archaeodesmana*. Moreover, the oldest fossil record from Desmaninae subfamily is 8.2 Myr old (Fortelius, 2015).

Based on previous phylogenetic studies, the monophyly of the subfamily Desmaninae within the

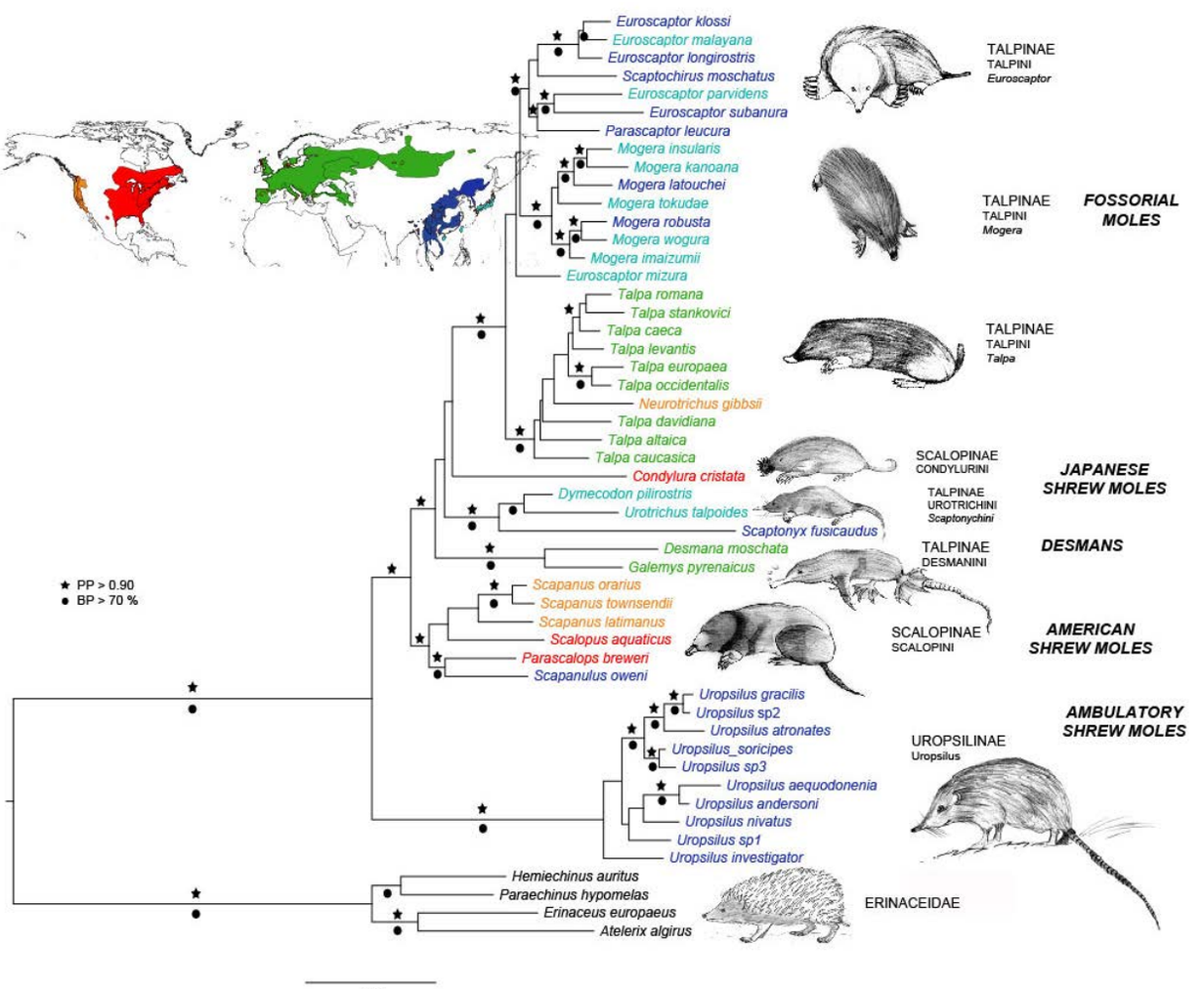


Figure 9. Phylogenetic relationships within the family Talpidae. (Bernet García-Santesmases, 2016)

order Eulipotyphla was highly supported. Moreover, the Russian desman was also confirmed as a sister group of the Pyrenean desman with high phylogenetic support (Fig. 9) (Cabria et al., 2006). Another study placed the divergence of these two species at 13.9 Myr ago (Igea et al., 2013).

Concerning mitochondrial phylogeography, a previous study divided the Pyrenean desman in two main clades, A and B, which, in turn, were also divided into two subclades each (A1 and A2; B1 and B2) (Igea et al., 2013). These clades are correlated with the geographic structure, being these mainly allopatric. Two main contact zones were detected between the two main clades (Fig. 10). One is in the Iberian System and the other one in the Cantabrian Range, although the latter was not clearly delimited due to the shortage of sampling in this area. Regarding dating estimates, it was estimated that the two mitochondrial clades diverged 0.32 Myr ago, putting an upper limit to the divergence of the two main populations. Finally, the main glacial refugium was inferred to be in the northwestern area of the distribution, being consistent with the areas of high genetic diversity and, also with the optimal areas for the Pyrenean desman, according to the SDM. Other glacial refugia, likely to be less important, were inferred within this distribution range (Fig. 11). From these glacial refugia, the postglacial expansions gave rise to the current distribution (Igea et al., 2013).

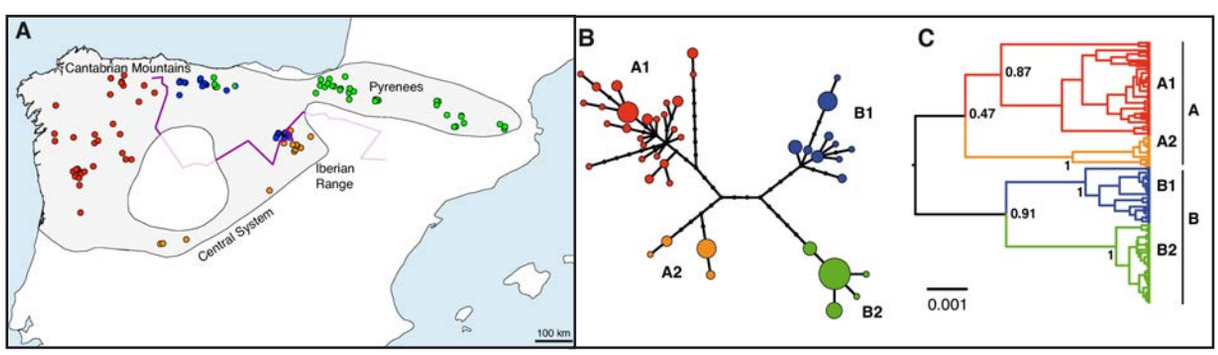


Figure 10. Phylogeography of the Pyrenean desman *(Galemys pyrenaicus).* A shows a map of the distribution range of the Pyrenean desman, correlated with the clades and subclades showed in B (haplotype genealogy based in a maximum-likelihood tree) and in C (Bayesian inference tree). (Igea et al., 2013). *Reprinted with the permission of the authors*

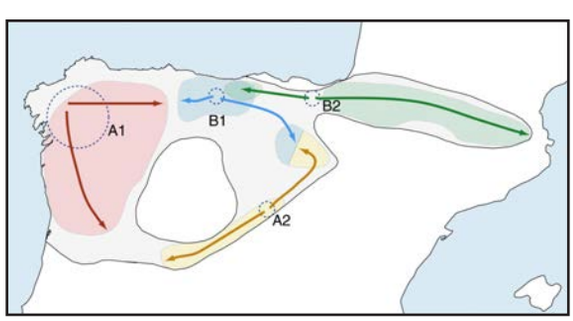


Figure 11. Scheme of the glacial refugia for the Pyrenean desman in the Last Glacial Maximum (LGM). (Igea et al., 2013). *Reprinted with the permission of the authors*

3.2 THE MEDITERRANEAN WATER SHREW *(Neomys anomalus)*

The Mediterranean water shrew *(Neomys anomalus)* is a semi-aquatic mammal endemic of the Iberian Peninsula (Fig. 12). This species used to be considered as the subspecies *Neomys anomalus anomalus*, belonging together with the subspecies *Neomys anomalus milleri* to the species *Neomys anomalus*. But a recent species delimitation project considered them as separate species (Igea et al., 2015). This work paved the way to study the phylogeography of this Iberian endemism.

Figure 12. The Mediterranean water shrew *(Neomys anomalus)*.
Picture credit: Virginia Gutiérrez Río

For this reason, the Mediterranean water shrew was used as a model of study in this thesis. In addition, it is interesting to compare its phylogeography with that of the Pyrenean desman. Hence, in the following sections a review of what is known so far about this small mammal is described.

3.2.1. *Neomys anomalus* WITHIN THE FAMILY SORICIDAE

The Mediterranean water shrew belongs to the family Soricidae, the subfamily Soricinae and the tribe Nectogalini (Repenning, 1967). Despite their primitive appearance, such as the long nose, shrews (family Soricidae) have never grabbed the attention of many zoologists. This may be due to the fact that they are so difficult to sample. In any case, they had not been studied, properly classified and described until midway through the 20th century (Reumer, 1998).

Two paleontologists, Simpson (1945) and Friant (1947) were the first to classify the shrews (family Soricidae) (Simpson & others, 1945; Friant, 1947; Reumer, 1998). But it was not until 1967 when Reppening published "The Subfamilies and Genera of Soricidae" that the classification is similar

to what we use nowadays was established, although minor changes have since been included (Repenning, 1967; Reumer, 1998).

Until 2015, when it was proposed that *Neomys anomalus anomalus* changed its status from subspecies to species (Igea et al., 2015), *N. anomalus* was considered a Paleartic species with a wide-spread distribution, occupying central, southern and part of eastern Europe. After the species delimitation study, the *Neomys* genus was divided into four different species, *N. teres*, which is restricted to a patch in eastern Europe (Armenia, Azerbayan and Georgia), *N. fodiens*, which is widely distributed in the Palearctic from the north of the Iberian Peninsula to the Baikal lake and *N. milleri,* which occupies central and eastern Europe, including Italy, Greece and Turkey and also part of Asia. Finally, *N. anomalus* is endemic to the Iberian Peninsula. *N. anomalus, N. milleri* and *N. teres* are allopatric species as shown in the maps in Fig. 13, but, in contrast, *N. fodiens* overlaps with the three others.

3.2.2. GENERAL CHARACTERISTICS OF THE MEDITERRANEAN WATER SHREW

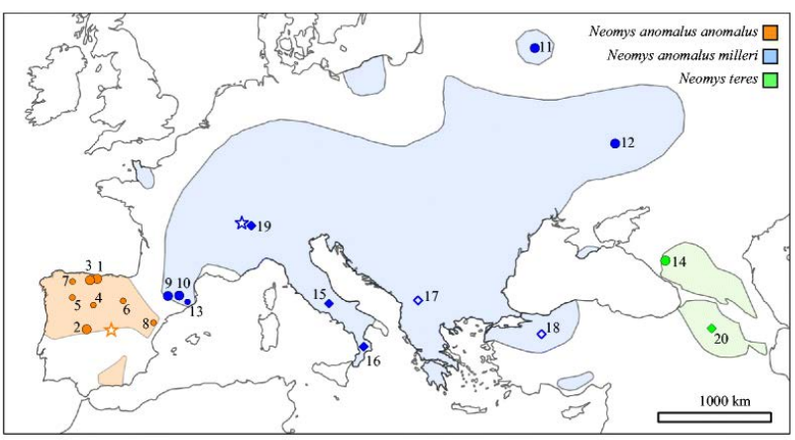


Figure 13. *Distribution ranges of the Neomys anomalus anomalus, Neomys anomalus milleri and Neomys teres* (Igea et al., 2015). *Reprinted with the permission of the authors*

The Mediterranean water shrew is a small semi-aquatic mammal with a long nose that it uses to feed on both river and terrestrial macroinvertebrates. They mainly forage in shallow waters, in muddy areas and in drier terrestrial habitats (Rychlik, 1997). Their legs are adapted to aquatic environments and it has a gland that secrets venomous substances. Its fur coloration ranges from black to grey (Ventura, 2007). As stated before, it is endemic to the Iberian Peninsula.

Although it is considered as Least Concern by the IUCN (Categories, 2001), *N. anomalus* and *N. milleri* populations are known to be decreasing due to habitat fragmentation (IUCN, 2001). Their conservation status, especially now that they have been split, may have to be revised.

Regarding habitat preferences, this mammal lives in riverine ecosystems, although it is known to have colonized areas far away from the rivers and it is known to swim and dive less than its sister species, the Eurasian water shrew *(Neomys fodiens)* (Rychlik, 1997). It is distributed from the sea level to 1600 meters of altitude. Concerning reproduction, only studies regarding *Neomys milleri*

have been published (when *Neomys anomalus* and *Neomys milleri* were considered subspecies of *Neomys anomalus*), which indicates a reproduction period from spring to autumn. However, a revision of this reproduction patterns should be performed for *N. anomalus*. There is a similar lack of knowledge regarding population abundance and behaviour.

3.2.3. PHYLOGENETIC POSITION OF THE MEDITERRANEAN WATER SHREW *(Neomys anomalus)* WITHIN THE FAMILY SORICIDAE

According to previous phylogenetic studies, the monophyly of the subfamily Soricinae within the family Soricidae is clear and well supported. Moreover, the tribe Nectogalini is highly supported in the phylogenetic tree and its split from its "sister" tribe Notiosociorini was 13.9 Myr ago (Dubey et al., 2007). Within the tribe Nectogalini, the genus *Neomys* is clearly the sister clade of the rest of genera. Regarding the origin of this family, it was inferred to be in Eurasia, expanding from there during the Miocene (Repenning, 1967; Rzebik-Kowalska, 1998; Storch et al., 1998).

Based on a Bayesian species delimitation method, the tree with the four differentiated lineages within the genus *Neomys (N. teres, N. fodiens, N. anomalus* and *N. milleri*) resulted in a posterior probability over 0.95 (Igea et al 2015). These and another analyses that indicated very low gene flow between lineages led to the conclusion that these four lineages deserve the status of species, elevating *N. anomalus* to species status. Moreover, the *N. anomalus* and *N*. milleri split was estimated to have occurred in the Middle Pleistocene, indicating a large period of isolation and genetic divergence (Igea et al., 2015). As a matter of fact, these two species were originally described as different species (Cabrera, 1907; Mottaz, 1907). Morphological differences between them included the skull shape, which in *N. milleri* is round and more angular in *N. anomalus*. Moreover, the tail and the hind feet are longer in *N. anomalus* (Miller, 1912; Cabrera, 1914).

4. THE ROLE OF GENETICS IN CONSERVATION

Currently, we are living through what is called the "sixth extinction", where the magnitude of loss of biodiversity is similar to the last massive extinction, according to geological records (Leakey, 1995). Although extinctions are natural evolutionary processes where the equilibrium is maintained by the appearance of new species, this time the number of extinctions exceeds the number of new speciation processes. This is because human actions are accelerating the extinction, reducing considerably the planet biodiversity (Frankham et al., 2002). Thus, it is necessary to take action to restore and conserve habitats and species. And the first step to preserve species and habitats is to understand their biology. Genetic research can help in obtaining useful information on many

aspects that can be essential for conservation (Frankham et al., 2002; Schwartz et al., 2007). In this thesis, I try to understand the evolutionary history of two endemic species, the Pyrenean desman and the Mediterranean water shrew, with the purpose of obtaining useful scientific information and the final aim of preserving and conserving them.

The main threats of biodiversity are habitat loss, over-exploitation, invasive species, pollution and the current climate change, as stated before, mainly due to anthropogenic causes (Wilson, 1989; Dudgeon et al., 2006; Groom, 2006). But, why is it so important to conserve biodiversity? The first reason is mainly ethical and philosophical. Living organisms have been evolving over thousands of years and they should be able to continue to exist following the natural equilibrium of extinction and speciation. That is to say that the man should interfere the minimum possible in the natural environment, following the idea of "the land ethics" that Aldo Leopold developed. But, aside from this ethical justification, biodiversity has a huge economical value. For instance, organisms provide bioresources such as food or pharmaceutical drugs (Harvey, 2008). Also, the ecosystem services available, such as the oxygen produced by plants or climate regulation, are key for the functioning of all organisms. In fact, the pollution of the air is causing considerable economic losses in the health industry, aside from the high levels of lung cancer and other associated diseases (Lave & Seskin, 1973). Another reason to preserve biodiversity is what is called the "aesthetic" value, referring to the pleasure that biodiversity gives to humans. This is, for example, the enjoyment of natural parks and reserves, the ornamental use of plants, ecotourism and wild animal safaris, among others. Hence, the decrease of biodiversity would result in a remarkable loss, both economic and patrimonial (Frankham et al., 2002). This makes biodiversity extremely important to preserve, and indeed this need is recognized by the IUCN.

Genetics is recognized as essential for species fitness and capacity of evolving (Frankham et al., 2002). Genetic studies can be applied to conservation in several ways, such as reducing the extinction risk due to loss of genetic diversity, resolving taxonomic uncertainties that would determine if a population is unique and needs special conservation treatments or defining sites for reintroductions or translocations (Frankham et al., 2002). It is widely known that population with low genetic diversity have less fitness and high risk of extinction (Reed & Frankham, 2003). For example, a genetic study regarding genetic diversity of the Tasmanian devil *(Sarcophilus harrisii)* determined that the low genetic diversity of their populations is the reason why an epidemic facial tumour is leading the species to extinction. This finding has improved the conservation management, reintroducing specimens and isolating healthy populations (Jones et al., 2007), proving that genetics plays an essential role in conservation.

Phylogeographic studies are also useful for conservation, as it helps to describe the primary motors of evolution, allowing the detection of important evolutionary processes, such as dispersal (Bonin et al., 2007; Gugerli et al., 2008). Together with protecting species and habitats, evolutionary processes

should be included in conservation programmes, as these are the drivers of diversification, speciation and, consequently, the persistence of populations and species (Carvalho et al., 2011; Crandall et al., 2000).

DIRECTOR'S REPORT

ADVISOR'S REPORT ON THE PUBLICATION STATUS OF THE RESULTS AND THE IMPACT FACTOR OF THE PUBLISHED PAPERS

Dr. José Castresana, as advisor of the PhD thesis of Marina Querejeta Coma entitled “Evolutionary history of two Iberian Soricomorpha: Genomics, phylogeography and dispersal patterns”, reports that the Thesis is formed by three chapters consisting in respectively two published papers and one complete manuscript ready for submission.

PAPER 1

Querejeta, M., González-Esteban, J., Gómez, A., Fernández-González, A., Aymerich, P., Gosálbez, J., Escoda, L., Igea, J., and Castresana, J. (2016). Genomic diversity and geographical structure of the Pyrenean desman. *Conservation Genetics* 17, 1333-1344.

Conservation Genetics is an international journal included in the Web of Science Core Collection. It has in the latest edition of the Journal Citations Reports (2015) an impact factor of 2.04 and is in quartile 2 of the Biodiversity Conservation category (15 of 49).

PAPER 2

Querejeta, M., Fernández-González, A., Romero, R., and Castresana, J. (2017). Postglacial dispersal patterns and mitochondrial genetic structure of the Pyrenean desman *(Galemys pyrenaicus)* in the north-western region of the Iberian Peninsula. *Ecology and Evolution, in press.*

Ecology and Evolution is an international journal included in the Web of Science Core Collection. It has in the latest edition of the Journal Citations Reports (2015) an impact factor of 2.537 and is in quartile 2 of the Ecology category (54 of 150).

PAPER 3

Querejeta, M. and Castresana, J. Evolutionary history of the endemic Mediterranean water shrew *Neomys anomalus*: glacial refugia within the Iberian Peninsula.

This manuscript is complete and ready to be submitted to an international journal of the Web of Science Core Collection.

Barcelona, 5th April 2017

José Castresana

ADVISOR'S REPORT ON THE CONTRIBUTION OF THE PHD CANDIDATE IN THE PAPERS FORMING THE THESIS

Dr. José Castresana, as advisor of the PhD thesis of Marina Querejeta Coma entitled “Evolutionary history of two Iberian Soricomorpha: Genomics, phylogeography and dispersal patterns”, reports that the contribution of the PhD candidate in each one of the presented papers was as follows:

PAPER 1

J.C. conceived the ideas; J.G.-E., A.G., A.F.-G., P.A., and J.G. collected samples; M.Q. performed the laboratory work; L.E. and J.I. performed some laboratory work; J.C. and M.Q. analysed the data; and J.C. led the writing with input from the other authors.

PAPER 2

J.C. and M.Q. conceived the ideas; A.F.-G. and R.R. collected the samples; M.Q. performed the laboratory work and analysed the data; and M.Q. and J.C. led the writing with input from the other authors.

PAPER 3

M.Q. and J.C. conceived the ideas; M.Q. performed the laboratory work and analysed the data; and M.Q. and J.C. led the writing.

I also confirm that none of these publications or the presented data has been used, or will be used, in a different PhD thesis.

Barcelona, 5th April 2017
José Castresana

ABSTRACT

The study of intraspecific diversity is essential in ecology and evolution as they reveal fine-scale patterns and it is crucial in conservation planning. The main aim of this thesis is to study the evolutionary history of two Iberian Soricomorpha: The Pyrenean desman (*Galemys pyrenaicus*) and the Mediterranean water shrew *(Neomys anomalus)*. Both mammals are semi-aquatic and endemic to the Iberian Peninsula. To this end, different approaches were applied, such as genomics, phylogeography and species distribution modelling (SDM), in the three chapters of this thesis. In **Chapter 1**, the main aim was to study the population structure of the Pyrenean desman using genomic markers. In order to achieve this objective, we optimized a genomic reduction protocol, double-digestion RAD sequencing (ddRAD), for the low DNA quantities needed when working with endangered species, like the Pyrenean desman. Moreover, we filtered and evaluated the library and determined the sex of the specimens used. Then, we estimated the heterozygosity rate and we determined population structure and evolutionary units performing a Principal Component Analysis (PCA), a Structure analysis and a genomic phylogenetic tree. As result, the heterozygosity rate turned out to be one of the lowest in mammals and five clades were delimited in the population structure. These results have important conservation implications. In **Chapter 2**, our main aim was to infer the mitochondrial genetic structure and the postglacial dispersal patterns of the north-western clade of the Pyrenean desman, using non-invasive samples. We delimited a contact zone in the Cantabrian Mountains and we studied the structure of nucleotide diversity through an interpolation map, the genetic structure through an Analysis of Molecular Variance (AMOVA) and the dispersal patterns through an Isolation-by-Distance approach. The results revealed areas of very high genetic diversity and areas of extremely low genetic diversity. Moreover, half of the genetic structure can be explained by the isolation of river basin, meaning that the desmans were not completely isolated in these basins, which is consistent with postglacial overland dispersal between the basins. This result was confirmed by the Isolation-by-Distance approach, which also revealed a high relative importance of overland dispersal during the Holocene but a similiar relative importance of overland and river dispersal for short distances. This study has also conservation implications, as it suggests that terrestrial corridors within the desman distribution range should also be restored and preserved. Finally, in **Chapter 3**, we studied the phylogeography of the Mediterranean water shrew, which was recently delimited as a species. We studied its phylogenetic structure using a Maximum Likelihood and Bayesian Inference phylogenetic trees. Moreover, we interpolated the nucleotide diversity and performed a species distribution modelling. The results revealed two main clades correlated with geographic structure. Moreover, the study suggests certain coincidence of high genetic diversity areas with optimal areas in the Last Glacial Maximum. Which is is consistent with the location of the glacial refugia.

RESÚMEN (IN SPANISH)

La diversidad dentro de especies es clave desde el punto de vista ecológico y evolutivo y, además, para la conservación. El principal objetivo de esta tesis es estudiar la historia evolutiva de dos mamíferos ibéricos: el Desmán de los Pirineos (*Galemys pyrenaicus*) y el Musgaño de Cabrera *(Neomys anomalus)*. Para ello, se han utilizado diversas metodologías, entre ellas genómica, filogeografía y modelos de distribución de especies, a lo largo de los tres capítulos que conforman esta tesis. En el capítulo 1, se busca estudiar la estructura genética del desmán usando marcadores genómicos. Para ello, se optimizó un protocolo de reducción genómica (ddRADSeq) y se realizaron análisis genómicos con los datos resultantes de la librería. Los resultados revelaron una heterozigosidad extremadamente baja y cinco clados como unidades evolutivas. Estos resultados tienen repercusión en la conservación de la especie. En el capítulo 2, se estudia la estructura genética mitocondrial y los patrones de dispersión postglacial del desmán en el clado nor-occidental. Los resultados sugieren que los desmanes no estuvieron completamente aislados en las cuencas de los ríos y se dispersaron durante el Holoceno por tierra para realizar desplazamientos largos. Por el contrario, para desplazamientos cortos, los resultados sugieren que se pudieron desplazar tanto por el río como por tierra. Finalmente, en el capítulo 3, se estudia la filogeografía e historia evolutiva del Musgaño de Cabrera y se ha descubierto una estructura genética dividida en dos clados correlacionados con la estructura geográfica. Además, los resultados sugieren la existencia de refugios glaciares y una compleja historia postglacial.

CHAPTER 1

GENOMIC DIVERSITY AND GEOGRAPHICAL STRUCTURE OF THE PYRENEAN DESMAN

Photo Credit: Jorge González-Esteban

Marina Querejeta [1], Jorge González-Esteban [2], Asunción Gómez [3], Angel Fernández-González [4], Pere Aymerich [5], Joaquim Gosálbez 5, Lídia Escoda [1], Javier Igea [1, 6], Jose Castresana [1]*

[1] Institut de Biologia Evolutiva (CSIC-UPF), Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain

[2] Desma Estudios Ambientales S.L., Ukulu 11, 31791 Sunbilla (Navarra), Spain

[3] Tragsatec, Área de Biodiversidad, Calle Julián Camarillo 6, 28037 Madrid, Spain

[4] Biosfera Consultoría Medioambiental S.L., Calle Candamo 5, 33012 Oviedo, Spain

[5] Departament de Biologia Animal, Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain

[6] Department of Plant Sciences, University of Cambridge, Cambridge, CB2 3EA, United Kingdom

Conserv Genet
DOI 10.1007/s10592-016-0865-y

RESEARCH ARTICLE

Genomic diversity and geographical structure of the Pyrenean desman

Marina Querejeta[1] · Jorge González-Esteban[2] · Asunción Gómez[3] · Angel Fernández-González[4] · Pere Aymerich[5] · Joaquim Gosálbez[5] · Lídia Escoda[1] · Javier Igea[1,6] · Jose Castresana[1]

Received: 19 February 2016 / Accepted: 7 July 2016

Abstract The Pyrenean desman (*Galemys pyrenaicus*) is a small semi-aquatic mammal endemic to the Iberian Peninsula. The species has recently experienced a strong decline and some of its populations are severely threatened with extinction. To help in the preservation of this species, it is critical to understand its genetic structure and main evolutionary units, as these may have specific local adaptations and could be of great conservation value. Sequencing reduced representation libraries (ddRAD) from 26 specimens selected from across its entire range, we obtained around 45,000 loci per specimen and 1185 single nucleotide polymorphisms. Heterozygosity varied substantially among individuals from different areas. Interestingly, specimens from the southeastern Pyrenees had some of the lowest proportions of heterozygous positions inferred from genome-wide data in mammals so far. In addition, we estimated a tree reflecting genomic divergence, performed a principal component analysis, and carried out a Bayesian analysis of the population structure. Combined evidence supported the existence of five distinct genomic clusters largely coincident with the main mountain ranges where the species occurs, with few specimens presenting relevant admixture levels. There was good correspondence between these populations and the mitochondrial lineages detected in a previous study, yet substantial differences in some areas demonstrate the importance of performing genomic analysis to reveal the whole population history. Although the analysis of further specimens is necessary to better characterize the distribution of the different evolutionary units, the distinctive geographical structure of this species revealed by the genomic data should be considered in future conservation plans.

Keywords *Galemys pyrenaicus* · ddRAD · SNPs · Conservation genomics · Genetic structure

Electronic supplementary material The online version of this article (doi:10.1007/s10592-016-0865-y) contains supplementary material, which is available to authorized users.

✉ Jose Castresana
jose.castresana@csic.es

[1] Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain

[2] Desma Estudios Ambientales S.L, Ukulu 11, 31791 Sunbilla, Navarra, Spain

[3] Tragsatec, Área de Biodiversidad, Calle Julián Camarillo 6, 28037 Madrid, Spain

[4] Biosfera Consultoría Medioambiental S.L, Calle Candamo 5, 33012 Oviedo, Spain

[5] Departament de Biologia Animal, Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain

[6] Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK

Introduction

Genomic analyses have enormous potential in the conservation of threatened species (Ouborg et al. 2010; Steiner et al. 2013; Shafer et al. 2015). In particular, the study of large numbers of genetic markers allows robust conclusions to be drawn regarding population structure and gene flow, which is essential knowledge for delimiting conservation units (Crandall et al. 2000; Funk et al. 2012). The identification of different genetic clusters within species is especially important because these clusters may possess singular characteristics and adaptations to their particular

Published online: 11 July 2016

habitats (Orsini et al. 2013; Lanier et al. 2015). In addition, the detection of low genetic diversity in specific populations is crucial because it may be related to loss of adaptive variation and reduced fitness (Tallmon et al. 2004; Allendorf et al. 2010). Although mitochondrial sequences and microsatellites have long been used for these purposes, SNPs and sequences obtained through next-generation sequencing (NGS) techniques have the advantage of providing massive quantities of genetic data, thus revealing more precise details of the population structure and variability within a species and admixture levels of specimens (Morin et al. 2009; Vonholdt et al. 2011; Xue et al. 2015; Kjeldsen et al. 2015). However, genome-wide studies used to address conservation problems are still very scarce and the actual potential of these new sequencing techniques for improving the management of endangered species is still to be determined.

The Pyrenean desman (*Galemys pyrenaicus*) is a small semi-aquatic mammal endemic to the northern part of the Iberian Peninsula. It lives in clean and oxygenated rivers, a habitat generally found in mountain areas (Palmeirim and Hoffmann 1983). Some populations have experienced a strong decline over recent decades, with the most southerly populations, in the Central System, having suffered the most significant decreases (Nores et al. 2007; Gisbert and Garcia-Perea 2014). There are several factors that have been suggested to explain these population decreases such as water pollution, water extraction, habitat fragmentation caused by hydroelectric plants and reservoirs, and predation by the invasive American mink (*Neovison vison*) escaped from fur farms (Fernandes et al. 2011). The Pyrenean desman is legally protected in the four countries where it is present (Spain, Portugal, France and Andorra) and currently appears as "Vulnerable" on the IUCN Red List (Fernandes et al. 2011). Furthermore, the population of the Central System was recently catalogued by the Spanish Government as "In danger of extinction", which is the highest protection category. If reinforcement of some populations becomes necessary in the future, it is important to have sufficient information on the population structure and conservation units of the species in order to do so without altering the genetic identity of the different populations. This knowledge is essential because, even if the Pyrenean desman occupies a relatively small geographic range, its populations are not homogeneous. First, two subspecies, *pyrenaicus* and *rufulus*, were described according to differences in coloration and size (González-Esteban et al. 1999; Lopez-Fuster et al. 2006). Furthermore, a strong phylogeographic structure was revealed with mitochondrial data, as explained below (Igea et al. 2013). Finally, the distribution of the Pyrenean desman encompasses a large range of climatic and hydrological regimes, and the existence of local adaptations cannot be discarded. Therefore, outbreeding depression after admixture of divergent populations might occur (Lynch 1991; Tallmon et al. 2004; Allendorf et al. 2010).

In a previous work, Igea et al. (2013) studied the phylogeography of the Pyrenean desman using mitochondrial markers. One of the principal conclusions was the existence of a marked phylogeographic structure in which two main groups (A and B) were further divided into two subgroups. The resulting four mitochondrial lineages (A1, A2, B1 and B2) had an allopatric distribution and a likely glacial origin. The populations with largest genetic diversity were detected in the most occidental part of the distribution, suggesting that one of the most important glacial refugia was in this area. Eight nuclear introns permitted corroboration of these conclusions about genetic diversity although their variability was too low to add complementary information about population structure (Igea et al. 2013). The contact zones between the main mitochondrial groups, one located in the Cantabrian Mountains and the other one in the Iberian Range, indicated an almost complete absence of spatial mixing between these maternal lineages after postglacial recolonization (Igea et al. 2013). This is a very peculiar situation in both contact zones, where no apparent barriers to the dispersal of desmans seem to exist. Since this phylogeographic analysis was based on mitochondrial genes, of exclusive maternal inheritance, it is possible that analyzing large numbers of nuclear loci could reveal additional details of the overall genetic structure, as demonstrated with other species (Godinho et al. 2008; Nater et al. 2011).

In this work, we aimed to study the population structure of the Pyrenean desman using genomic sequences and SNPs obtained through NGS techniques. Genomic studies of threatened species such as the Pyrenean desman present several difficulties arising from the shortage of samples, the low quality of some non-invasive samples and the small quantity from each sample usually available for DNA extraction. We tried to overcome these difficulties by improving certain laboratory protocols and bioinformatic analyses. Genomic libraries were generated using a double digest restriction associated DNA (ddRAD) protocol (Peterson et al. 2012), a reduced representation library approach that allows a large number of specific genomic fragments to be sequenced, but with several modifications to be able to start with small quantities of DNA. We also used full sequence data to perform essential quality tests of the library sequences, determine the sex of individuals, and estimate the proportion of heterozygous positions. Finally, we studied the population structure of the species using SNP data. The number of samples available for this work was small and some important geographical locations could not be covered with enough sample size. However,

we were able to test these NGS techniques and uncover the global population structure of the species, which could be useful for developing conservation strategies.

Materials and methods

Samples

We used 26 samples of the Pyrenean desman from different geographical locations and mitochondrial clades (Table S1; Fig. S1), 22 of which came from DNA extractions performed in a previous phylogeographic study of the species (Igea et al. 2013). Most of these 22 samples came from very small amounts of fresh tissue obtained during previous survey works of the species but three of them belonged to dry tissue from different biological collections (Table S1). From the areas where a higher number of samples was available (Occident and Pyrenees), a maximum of 8 samples were used to avoid overrepresentation of these populations. From other areas, all samples from which enough DNA was available were used. Unfortunately, for a relevant area like the Central System only one DNA sample could be used. Four additional samples were obtained from specimens captured in a survey commissioned by the Regional Government of La Rioja (Spain), in 2011. These were used to complete the sampling in the Iberian Range, which included few samples from the previous study. A small piece of tail tip was taken from the captured specimens and stored in ethanol, and the animals were released back in the wild. The work to obtain samples for genetic analyses was performed with permit number A/2011/52 issued by the Regional Government of La Rioja; it was carried out following national and international regulations, and all necessary steps were taken to prevent any damage to the specimens. DNA was extracted from these samples as in Igea et al. (2013).

Library construction and sequencing

The DNA library was constructed using the ddRAD protocol (Peterson et al. 2012), with several modifications to allow for smaller quantities of initial DNA. The experiments were performed in series of 12 samples each time, repeating samples with low sequence yield in subsequent experiments. To avoid bias that could be related to specific sequencing runs, samples were randomly selected from different geographical areas for each experiment. In order to identify each specimen in the analyses, we used 12 different P1 adapters, each with a different 5 nucleotide barcode, and one common P2 adapter. Using the single-stranded oligonucleotides and the corresponding complementary molecules provided in Peterson et al. (2012), an annealing reaction was carried out in order to produce double-stranded adapters.

The DNA concentration was measured either by absorbance at 260 nm (with Nanodrop) or using qPCR (see supplementary information). From each sample, a quantity of genomic DNA varying between 200 and 250 ng was double digested using restriction enzymes EcoRI and MspI, in a 20 µl volume. After incubating overnight, the enzymatic reaction was heat inactivated over a 30-min period at 80 °C. We then added (with no previous cleaning step in order to avoid DNA loss) a 80 µl ligation mix, including T4 DNA ligase, P1 adapter (which binds to the EcoRI overhangs), and P2 adapter (which binds to the MspI overhangs), and the solution was incubated overnight at 23 °C. The ligation reaction was heat inactivated at 65 °C over a 10-min period.

We then mixed all the ligation reactions in a single tube. This mix was concentrated in 30 µl using the MinElute PCR Purification Kit (QIAGEN). The entire 30 µl pool was run on a 1 % low-melting agarose gel for 1 h and was then selected by size (300–400 bp) using a fresh razor blade. The gel purification was eluted in 30 µl using the MinElute Gel Extraction Kit (QIAGEN).

Subsequently, a PCR amplification of the size-select sample was performed to add Illumina adapter sequences, using the oligonucleotides described in Peterson et al. (2012). Phusion high-fidelity DNA polymerase (New England Biolabs) amplifications with a number of cycles ranging from 14 to 24, depending on the intensity of initial PCR products, were carried out. A total of 4 µl of the size-select samples pool was used as template. To increase the concentration of the libraries and to minimize bias, a total of 4–7 PCRs were performed. Then, we combined and concentrated the PCR products into a 30 µl volume using the MinElute PCR Purification Kit. Finally, to get rid of small fragments, the library was a run in a precast EX 2 % agarose gel using an E-Gel system (Invitrogen). The band corresponding to the library was then extracted with the Qiaquick Gel Extraction Kit (QIAGEN) in 20 µl.

The library size was checked with a Bioanalyzer 2100 (Agilent) and sequenced using a MiSeq (Illumina) in the Genomics Core Facility at the Pompeu Fabra University. Sequencing was carried out using the MiSeq Reagent Kit v3 in a single read length of 150 cycles.

Sequence processing

We used the Stacks 1.10 package (Catchen et al. 2011, 2013) to process the sequences obtained. First, the *process_radtags* program was used to filter out reads with low-quality sequences or malformed restriction sites or barcodes, and to separate reads belonging to different

samples according to the barcodes. After this step, sequences of the same sample from different runs were combined. Then *ustacks* was used to assemble loci in each sample, with a minimum number of sequences for each sample at a locus (minimum stack depth or m) of 3 and a maximum number of differences between reads (M) of 2. The mean sequence coverage of each specimen was then calculated as the number of assembled reads divided by the number of assembled loci. The loci of all the specimens were subsequently merged with *cstacks*, allowing for the maximum number of differences between reads (n) of 2, and a catalog of loci and sequences was then created with *sstacks*. Then the *populations* program was used to save two different output files for downstream analyses. First, the SNPs were saved from complete loci, i.e., those represented in all samples ($r = 1$). In addition, assembled loci represented in at least 14 samples ($r = 0.51$) were saved in FASTA format.

Using a series of custom scripts, different quality checks and processing steps were performed on the sequences obtained in FASTA format. First, incompatible loci (i.e., those with more than two alleles) were eliminated from the FASTA file. In addition, loci with less than 140 bp were also removed.

It has been suggested that the Stacks algorithm may allow for the presence of duplicate loci, which could cause problems for correct assembling (Sovic et al. 2015). To detect these, we performed a Blast search (Altschul et al. 1997) of all the loci against themselves, using one sequence per locus and an e-value of $1e^{-30}$. Different e-values may lead to different stringencies in the search. In this case, a stringent value was used so that only duplicate loci with a high similarity, which would be the most problematic, were detected.

Since there is no *G. pyrenaicus* reference genome available, we needed to detect possible contaminant sequences such as those coming from bacteria present in the samples. To do so, we performed an additional Blast search of one sequence per locus against the GenBank nucleotides database. In this case, a less stringent e-value of $1e^{-10}$ was used to ensure that any contamination was detected.

Sequences detected by Blast searches as duplicates or contaminants were eliminated and this set of filtered loci was used for further downstream analyses.

In addition, from this set of filtered loci, a new set of complete SNPs was constructed for comparison with the initial set saved by Stacks, using only loci previously genotyped by Stacks. For the new set, we selected from each locus the SNP with the largest minimum allele frequency.

Sex determination

Sixteen samples were initially sexed by qPCR amplification of a Y-linked fragment (see supplementary information). Using the nine males and seven females with sex determined in this way, we searched, within the assembled loci, for putative Y chromosome loci in order to sex the rest of specimens. For this analysis, the Stacks *populations* program was run with $r = 0.25$ to ensure the assembly of Y-chromosome loci from males. We then selected all loci present in the seven known males and in none of the nine females, and in which the two sequences assembled by Stacks were identical. The application of these conditions expected for Y-chromosome genes resulted in the initial identification of 50 loci. One of the loci was inconsistent with the others (it was found in two specimens that did not have any of the other loci) and was consequently eliminated. The discriminating value between the sexes of the remaining 49 loci was tested using the sequenced reads (just filtered with *process_radtags*). For this purpose, we performed a Blast search of the reads of each sample against the likely Y-linked loci using very stringent criteria to avoid finding duplicates in the X chromosome ($1e^{-74}$).

Proportion of heterozygous positions in each specimen

To ensure adequate coverage for individual heterozygosity rate estimation (Buerkle and Gompert 2013), a new dataset of assembled loci was generated with Stacks using a minimum stack depth for each sample of 9. This new set of sequences was processed as described above, and loci with duplicate and contaminant sequences were eliminated. The proportion of heterozygous positions for each sample was estimated from the sequences in FASTA format as the number of variable positions for the sample in all loci divided by the total length of the loci.

Genomic tree

We estimated a genomic tree from the variable loci using two different phylogenetic methods, as in Igea et al. (2015). First, we reconstructed a distance tree from the matrix of the average genomic divergence between specimens (Gronau et al. 2011; Freedman et al. 2014). To summarize the divergence of the two separate alleles, we calculated pairwise distances between all specimens using formula 8.2 in Freedman et al. (2014). Distances were corrected for multiple substitutions using the Jukes-Cantor formula. The resulting matrix of pairwise distances was used to reconstruct a distance tree using the Fitch program in the Phylip package (Felsenstein 1989). Another tree was obtained from the distances calculated using formula 8.1 in Freedman et al. (2014), where a more conservative estimate of the differences at each position is computed, but the results were very similar (not shown).

In addition, we reconstructed a nuclear genomic tree by concatenating all loci. We performed several concatenations where the order of each allele pair was randomly changed. From each concatenation set, a maximum-likelihood tree was estimated using RaxML version 7.4.2 (Stamatakis 2006) with a general time reversible plus invariable sites model.

Tree figures were prepared with SplitsTree (Huson and Bryant 2006). In these figures, the tree branches were oriented to maximize the match with the geographical location of the populations without altering the tree topology.

Principal component analysis and population structure

Principal component analysis (PCA) was performed with SNPrelate (Zheng et al. 2012) from the genetic covariance matrix of the SNP genotype data. The axes of the plot were oriented to maximize the match of the positions of the samples in the plot and their geographical locations.

Population structure and admixture were estimated from the SNP data sets with the program Structure 2.3.4, which implements a Bayesian model-based clustering method (Pritchard et al. 2000) using the admixture and correlated allele frequency models, and with no prior information on population origin. A total of 500,000 generations were run after a burn-in of 50,000 generations with a number of clusters (K) ranging from 2 to 7 ($K = 1$ was not used because it gave unrealistically low posterior probability values). We performed 10 different runs for each K value in order to plot the trend of the estimated posterior probability of the data, Ln P(D) (Pritchard et al. 2000). In addition, the optimal K value was estimated using the ΔK method (Evanno et al. 2005) as the inflection point of the Ln P(D) curve before reaching an horizontal or slightly increasing plateau. All 10 runs gave very similar results for meaningful K values and only the first run was shown in the table of admixture proportions and figures.

Results

Genome fragments obtained

A total of 62,551,837 sequence reads passed the initial quality filters, with an average of 2,405,840 reads for each of the 26 *G. pyrenaicus* specimens analyzed (Table S2). Different strategies were used to assemble these reads, as indicated in Fig. S2. Under a default minimum stack depth in the Stacks program ($m = 3$), 2,130,583 reads and 69,111 loci were assembled on average per sample, resulting in a mean depth of coverage of 28. When loci of different specimens were merged and only loci with at least 14 samples were selected, 57,212 loci remained, with a total of 8,326,212 bp potentially sequenced for each sample. However, the samples differed in the resulting number of assembled loci, ranging from 56,378 (99 % of the total loci) to 20,256 (35 %), with an average of 45,446 (79 %) (Table S2; Fig. S3). The average sequenced length per sample was 6,201,426 bp. Samples represented in few loci may have slightly degraded DNA but this did not prevent their use in further analyses.

The number of variable loci in this set was 16,742. The plot of the frequency of SNPs at each position along the sequence read provides a measure of the sequencing quality (Sovic et al. 2015). There is a slight trend in increasing polymorphisms towards the end of the read (Fig. S4), which could be related to the decrease in quality characteristics of Illumina reads. However, given the small proportion of positions affected, these errors are not likely to influence the results.

In order to evaluate the quality of the assembled loci, we applied further tests and filters. First, one sequence of each locus was selected to perform a Blast search against all other sequences. This way, 3563 duplicate loci were detected.

Next, a Blast search against the GenBank database was performed to estimate the amount of contaminants and other problematic sequences in our libraries. Of the 57,212 assembled loci, 6655 gave a significant hit to some sequence in GenBank. Of these, 21 belonged to non-Craniates and were considered contaminants: 16 sequences were found to belong to various bacterial species, 4 to nematodes and there was 1 unclassified sequence. There was also a hit to one mitochondrial fragment from *G. pyrenaicus*, not appropriate to be analyzed with nuclear data. All other hits corresponded to nuclear genes of Craniates (mostly mammals) and were considered endogenous because Craniate species are unlikely to be degrading organisms or parasites in these tissues (and it is expected that a few divergent sequences give a Blast hit outside mammals). The rest of sequences, giving no hit, probably belonged to non-coding DNA and other fast-evolving regions of the genome. When duplicate and other problematic loci were removed (some loci were detected in both analyses), we obtained a new data set with 53,637 loci, 15,242 of which were variable.

Sex determination with sequenced reads

Using 49 probable Y-chromosome loci identified with a subset of specimens of known sex, we determined the sex of all samples by Blast, searching the sequenced reads against these loci. We found that the number of hits per million reads was either less than 3 or more than 326 in the

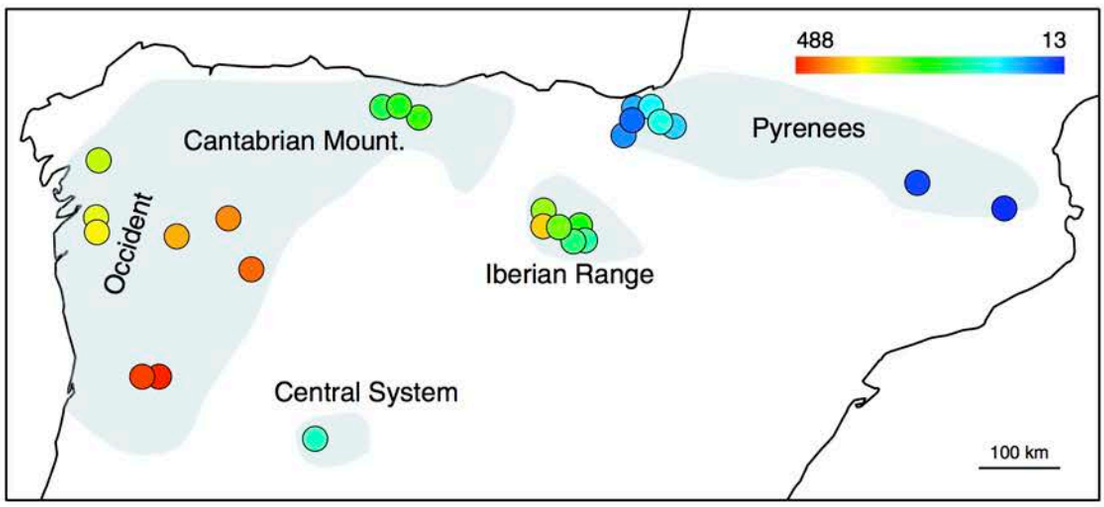


Fig. 1 Map plotting *color*-coded heterozygosity rates in different specimens of the Pyrenean desman. The *color scale* is in number of heterozygous positions per million bases. The *shadowed* area approximately represents the current species distribution according to different sources. (Color figure online)

different samples. Therefore, these 49 loci were perfectly discriminated between the sexes and clearly Y-linked, resulting in 13 males and 13 females (Table S3).

Proportion of heterozygous positions

To estimate the individual heterozygosity rate with loci of adequate coverage for accurate genotyping, we considered the assembled loci with a minimum stack depth for each sample of 9. This rendered 32,955 loci with at least 14 samples, totaling 4,796,424 bp of sequence length (Fig. S2). The heterozygosity rate estimated from these sequences revealed large differences between specimens, with values ranging from 0.000013 to 0.000488, that is, from 13 to 488 heterozygous positions per million bases, which is more than an order of magnitude of difference (Table S4). Furthermore, when the heterozygosity was plotted on the map (Fig. 1), it was observed that the specimens with the highest values were those from the occidental zone whereas much lower values came from the Pyrenees. In particular, the southeasternmost extent of the Pyrenees showed the lowest heterozygosity rates.

Genomic tree

Genomic tree reconstruction was performed using two sets of variable loci: without filtering duplicates and contaminants, and with filtered loci (Fig. S2). A genomic tree was first obtained from the 16,742 unfiltered variable loci (2,436,857 bp) previously determined with a default minimum stack depth (Fig. 2). This tree clearly revealed the existence of five clades that grouped specimens from the same geographical region: Pyrenees, Occident, Cantabrian Mountains, Iberian Range and Central System. A similar phylogeny was obtained by maximum likelihood of the concatenated loci (Figure S5). The trees obtained after filtering duplicates and contaminants were also similar (not shown).

PCA and population structure

Two sets of SNPs were also used for PCA and population structure analysis (Fig. S2). First, the output from the Stacks program rendered 1185 SNPs present in all samples. Second, after filtering duplicate and contaminant loci, a set of 1053 SNPs remained. The results were similar for the two sets and only those for the largest set are shown.

PCA based on these SNPs distinguished the same five groups as the genome tree, with the former more clearly showing the separation of the Central System population due to a large discrimination in the second PCA axis (Fig. 3).

We also used the Structure program with an admixture model. A plot of the estimated posterior probability of the data versus the number of clusters, K, shows that, after $K = 5$, this likelihood does not notably increase and does not even converge in some runs (Fig. S6). In addition, according to the Evanno method (Evanno et al. 2005), the optimal K value is 5. Figure 4 shows the map of the different specimens and the admixture proportions of each of them as determined with $K = 5$ (Table S5). The five clusters basically coincide with those obtained in the genome tree and the PCA analysis. A strong population structure is observed, with different populations occupying

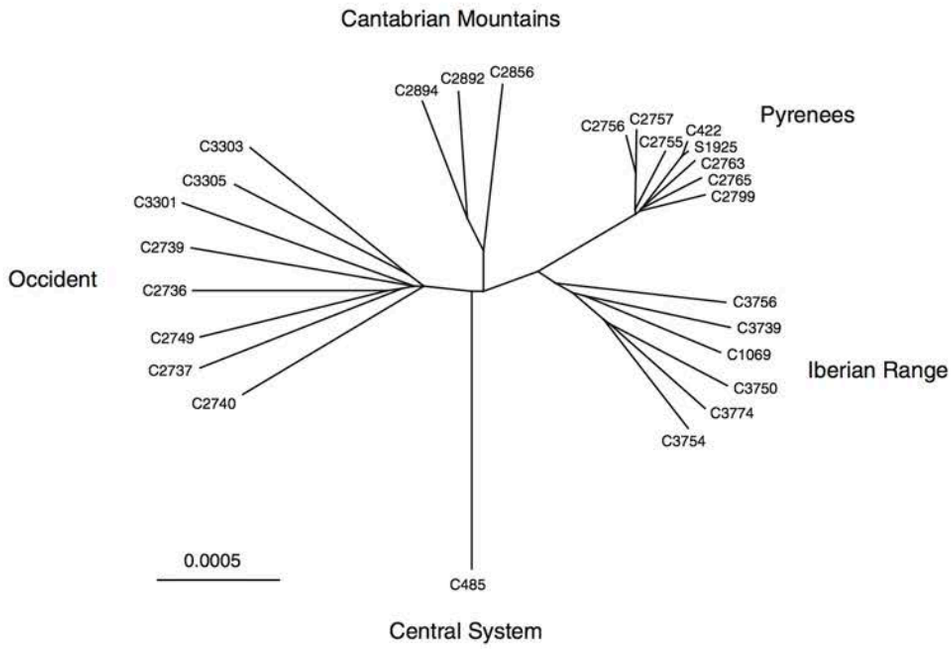


Fig. 2 Genomic tree obtained from the genetic distances of the 16,742 variable loci (2,436,857 bp). The *scale* is in substitutions per position

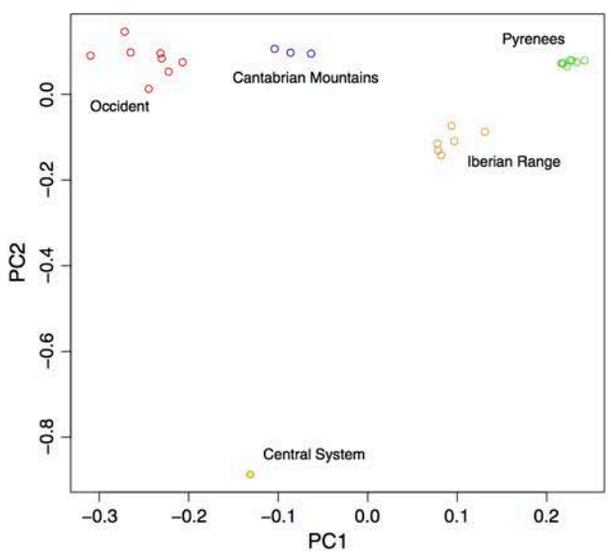


Fig. 3 Principal component analysis (PCA) based on the SNP data

separate areas of the distribution range of the species. In addition, only four specimens showed relevant admixture levels (>5 %). Three specimens from the northwestern part of the Iberian Range had components of the Pyrenean genome and, two of them, also had elements of the Cantabrian Mountains genome. In addition, one specimen from the Cantabrian Mountains showed components from both Pyrenean and Occidental clusters (Table S5).

Since determining the optimal K value is usually difficult (Pritchard et al. 2000), we also examined different aspects of the genetic structure with K values ranging from 2 to 5 (Fig. S7). When only 2 clusters are considered, there is a basic subdivision between Pyrenean specimens (with components present in adjacent populations) and the other populations. With 3 clusters, a new population is distinguished in the Iberian Range. With 4 clusters, samples from the Cantabrian Mountains give rise to a new population. Finally, with 5 clusters, the only sample of the Central System becomes distinct.

Discussion

NGS libraries and quality tests performed

The quantity and quality of initial DNA material is usually a limitation when applying genetic techniques, including NGS, to endangered species. This makes developing protocols that require low amounts of starting material an important first step in these studies. By removing some cleaning steps, we were able to prepare libraries starting with as little as 200 ng of DNA, in the low range of initial material recommended in a previously developed protocol (Peterson et al. 2012). However, it should be taken into account that working with low amounts of starting DNA, and particularly if it is of low quality, may affect the representation of all loci in the final library. For this reason, we evaluated the quality of the library sequences by calculating the proportion of total loci sequenced and properly assembled for each specimen (Table S2; Fig. S3). For this estimation, we used loci present in at least 14 of the 26

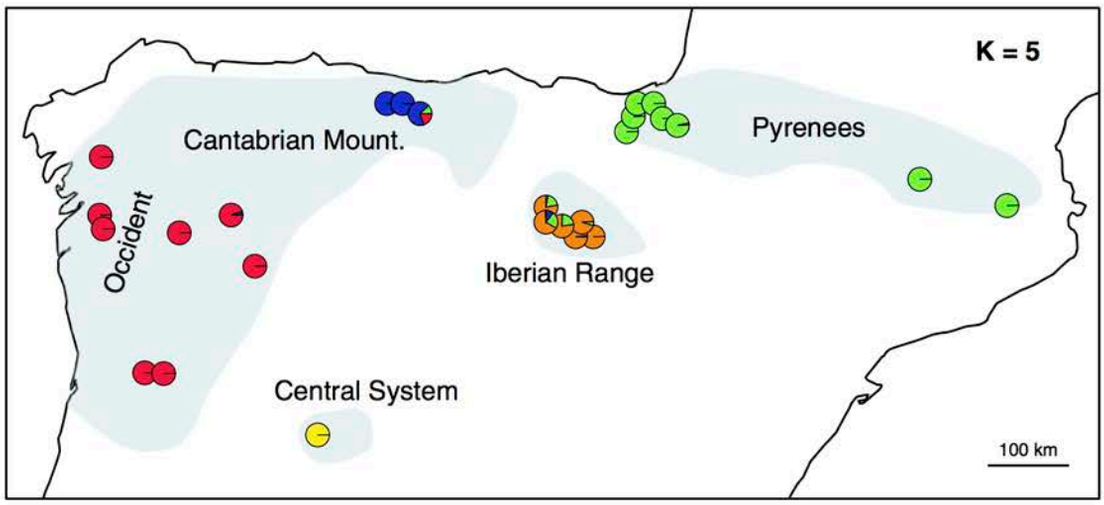


Fig. 4 *Map* plotting *color*-coded admixture proportions of each specimen as determined with structure and K = 5. (Color figure online)

samples processed (57,212 loci). A few specimens did not even have 50 % of these loci. The two samples with lowest percentages of loci corresponded to dry skins from old collections (IBE-C2892 and IBE-C2894, prepared in 1989 and 1988, respectively) and therefore some DNA degradation may have caused this problem. Specimen IBE-C1069 belonged to an individual found dead in the field. Individual sequencing runs of this sample always gave a low number of reads, but the combination of three runs allowed the assembly of most loci. In the end, all the specimens in our final set, even those with the smallest number of assembled loci, provided reliable and coherent information in the different analyses carried out.

As with many other non-model species, no *G. pyrenaicus* genome is available for reference, making it necessary to assess additional aspects of the quality of the anonymous loci obtained. There are two main sources of potential problems with the sequences obtained: duplicate sequences and contaminants. Duplicate loci, if very similar, may present problems when assigning different alleles to the correct locus (Sovic et al. 2015). We therefore conducted a Blast search of all sequences against themselves and detected that 6 % of all loci were duplicate, although not all of these might necessarily present assembling problems. The genetic structure and variability results were similar when filtering and not filtering out these loci, indicating that duplications were not an important issue in this case. However, this test using full sequence data should still be performed when a new species is studied, or a novel protocol that selects for different loci is used.

We also used full sequence information of loci to explore if contamination with sequences from bacteria, parasites or other non-endogenous sources could be problematic. A Blast search of our final assembled loci against GenBank produced 6655 significant hits and, of these, we only detected 16 sequences belonging to bacteria and 4 to nematodes. This was a small number of contaminant sequences, as expected due to the low number of suboptimal samples we used.

The results obtained after performing a Blast search of the sequenced reads of each sample against a set of 49 putative Y-chromosome loci indicated that these loci were perfectly discriminating between the sexes. Detected Y-chromosome loci may also be used to evaluate cross-contamination between samples, since female samples are expected to have no hits for any of these loci (Green et al. 2009). This contamination may arise during the handling of samples or library preparation, especially after pooling samples. All females in our work showed very low levels of hits for Y-linked genes (less than 3 per million reads; Table S3), indicating the suitability of our approach to library preparation. The sexing results, with equal numbers of males and females, also showed a lack of sex bias in the selected specimens.

Extremely low genomic heterozygosity rate in the Pyrenean desman

The individual heterozygosity rate estimated from the genomic data has been shown to be a relevant parameter for assessing the genetic diversity of species (Fig. S8). For example, the heterozygosity rate in primates varies from 800 heterozygous positions per million bases in non-African humans, eastern lowland gorillas, bonobos and western chimpanzees to around 1000 in African humans and 2000 in central chimpanzees, western lowland gorillas and both orangutan species (Prado-Martinez et al. 2013). Some highly endangered taxa in which this parameter has been

estimated from their complete genomes include the giant panda, with 1350 heterozygous positions per million bases (Li et al. 2010), the Amur tiger, with 490, the snow leopard, with 230 (Cho et al. 2013), and the cheetah, with 200 (Dobrynin et al. 2015). More recently, heterozygosity rates estimated in seven specimens of the Channel Island fox, a species that inhabits isolated islands with extremely small population sizes, varied between 14 and 408 (Robinson et al. 2016). In comparison, the Pyrenean desman has one of the lowest heterozygosity rates reported in any mammal so far, with a mean of 246 heterozygous positions per million bases (Table S4). Furthermore, there are substantial differences among the individual samples. We find values as low as 13 and 20 heterozygous positions per million bases in two samples from the southeastern area of the Pyrenees. All other specimens from the Pyrenean mountains have heterozygosity rates lower than 140. The only sample from the Central System also has a very low rate, with 147 heterozygous positions per million bases. The highest values are found in the occidental population, with one sample having 488 heterozygous positions per million bases. Interestingly, the range of heterozygosity rates found in the Pyrenean desman (13–488) is very similar to that of the Channel Island fox despite the larger geographical distribution of the desman. Extreme bottlenecks in small refugia during the last glacial maximum and before the postglacial recolonization may have created in the Pyrenean desman a pattern of genomic diversity similar to that found in oceanic islands.

A note of caution is necessary, however, when comparing the heterozygosity rates of *G. pyrenaicus* with other mammals, since this value was estimated in the present work from sequence fragments obtained from a double digestion of the genome and a size selection of 300–400 bp for the library construction. Although these loci should be a random representation of the genome, we cannot rule out that another factor may cause that the heterozygosity rates from ddRAD fragments and whole genomes are not totally comparable. Therefore, a firm conclusion on the heterozygosity rates of *G. pyrenaicus* compared with other mammals should await its complete genome sequencing. However, this parameter is very important for comparing specimens and populations within this species, showing how full sequence information obtained from the ddRAD data can also be exploited to achieve crucial knowledge on genetic variation.

The low heterozygosity rates found in the Pyrenees agree with the nucleotide diversity values estimated from mitochondrial and intronic data in a previous study (Igea et al. 2013). Another study of the northern part of the Pyrenean populations based on microsatellites also detected low levels of genetic diversity in this area (Gillet 2015). These populations have, until recently, been relatively well conserved (Aymerich et al. 2001; Aymerich and Gosálbez 2014), suggesting that long-term demographic decline does not explain the low genetic diversity in the Pyrenees. Instead, postglacial recolonization after a tight bottleneck seems to be the most likely explanation (Igea et al. 2013). However, recent data show some regression of the species in the Pyrenees (Charbonnel 2015). The isolation of subpopulations in some basins and the consequent increase of inbreeding or future unforeseen environmental changes may be more detrimental for populations starting with very low genetic diversity values (Allendorf et al. 2013).

Population structure of the Pyrenean desman: novel insights from genomic data

In a previous work, we unveiled a strong phylogeographic structure in the Pyrenean desman, with an almost total absence of spatial mixing of the determined clades (Igea et al. 2013). These conclusions were obtained with mitochondrial DNA but, given the importance of analyzing this structure with nuclear markers, we also developed several introns and sequenced them in the main populations. Unfortunately, the nuclear introns did not show enough variability for these aspects to be analyzed in depth (Igea et al. 2013). It was therefore essential to develop a higher number of variable nuclear markers with genomic techniques to obtain more robust conclusions about the geographical structure of the Pyrenean desman.

A method based on nuclear genetic distances from concatenated loci (genomic tree) and another based on allele frequencies of SNPs (PCA) both revealed five clusters largely coincident with the main geographical regions and mountain ranges: Pyrenees, Occident, Cantabrian Mountains, Iberian Range and Central System (Figs. 2, 3). As previously shown (Igea et al. 2013), and contrary to expectations for a species that is highly adapted to aquatic ecosystems, this structure is not related to river basins. This indicates that, at least during the postglacial recolonization of these areas, boundaries between river basins did not act as important barriers to the dispersal of the Pyrenean desman populations. However, neither the genome tree nor the PCA analysis determines whether any of these groups comprise admixed specimens. The Structure analysis based on an admixture model showed the same five clusters and few specimens with genomic components belonging to different populations. In fact, only four specimens showed relevant admixture levels: three from the northwestern part of the Iberian Range and one from the Cantabrian Mountains (Fig. 4). The estimated admixture is relatively small and may be an indication of the low dispersal of this species. However, the lack of sampling in contact zones prevents any conclusions being drawn at this time about the true degree of interbreeding between the different

populations. In addition, the Central System population, with only one individual analyzed here, should be studied using a larger sample size.

The clusters found with genomic data corresponded fairly well with the mitochondrial lineages detected in a previous study (Igea et al. 2013) at the initial subdivisions. Thus, the major mitochondrial groups, A and B, are largely coincident with the two populations found here with Structure at $K = 2$. However, at higher subdivisions the mitochondrial clades and the nuclear clusters showed some remarkable differences. One of the major differences was found in the Iberian Range. The previous study showed that, in the Iberian Range, there are specimens belonging to the two main mitochondrial clades, A2 and B1, but they seemed to be strictly separated into two different areas (southeast and northwest, respectively), suggesting a highly restricted dispersal of female desmans despite the existence of any obvious geographic barrier to gene flow. However, the genomic data analyzed in the present work reveals a different scenario in this area. In principle, the six specimens from the Iberian Range (three from each area and mitochondrial clade) can be assigned to the same genomic cluster, despite three of them having some components of other populations. This suggests a homogenization of the two Iberian Range phylogroups through certain degree of gene flow between the areas of the two main mitochondrial clades. This discrepancy between the mitochondrial and nuclear geographical structures may be due to dispersal differences between the sexes, although current data on movements of the Pyrenean desman prevent confirmation of any dispersal sex bias (Melero et al. 2012; Gillet 2015). Alternatively, reduced dispersal together with the smaller mitochondrial population size has been shown to enable the maintenance of a mitochondrial phylogeographic break in the absence of nuclear structure, but further work is necessary to discern different possibilities (Irwin 2002; Toews and Brelsford 2012). The Iberian Range provides a clear example of how the mitochondrial phylogeography of the Pyrenean desman showed only one part of the population history, that of females, while genomic data was crucial to reveal that the whole population in this mountain range is more homogeneous than previously suspected.

The other major difference between the mitochondrial and nuclear geographical structures was found in the Central System. While the mitochondrial phylogeography showed that specimens from this area belong to the A2 clade, also present in the Iberian Range (Igea et al. 2013), the single specimen from the Central System studied here appears clearly separate in the genomic tree and, particularly, the PCA analysis (Figs. 2, 3). The results of the Structure analysis with $K = 5$ also indicate that this individual may be part of a differentiated population (Fig. 4). However, it is obvious that assessing the genetic status of the Central System population using only a single specimen is difficult. Furthermore, other specimens may have a different genomic background and, therefore, genomic data from further samples from the Central System is critical to understand more fully the evolutionary history of *G. pyrenaicus* in this area. The robust characterization of the Central System population will be particularly important because it is the most endangered population according to current data, and an in-depth understanding of its genetic distinctiveness will be necessary for delineating informed conservation plans.

Despite the small number of samples we could analyze in this work, the use of massive information from nuclear loci and SNPs generated with a genomic approach allowed us to complement the genetic structure analysis of the Pyrenean desman initiated with mitochondrial DNA and nuclear introns. Genomic data corroborated the existence of a marked genetic structure in this species and supported a scenario of glacial refugia that gave rise to genetically distinct populations associated to the main mountain ranges. According to the distribution of the mitochondrial clades, a scenario of four refugia (NW of the Iberian Peninsula, Cantabrian Mountains, Central System and Basque Mountains) was proposed in Igea et al. (2013), but the current results would be compatible with the existence of another refugium situated in the Iberian Range to help explain the genomic cluster found in this area. The Iberian Peninsula was classically considered a single homogeneous refugium in Europe (Hewitt 2000) but several phylogeographic studies in the last few years evidenced that the Iberian Peninsula (and other South European peninsulas) contained multiple refugia that gave rise to distinct evolutionary lineages, supporting the "refugia within refugia" hypothesis (Gomez and Lunt 2007 and references therein). Our mitochondrial and genomic studies with the Pyrenean desman, not only provide strong evidence for the importance of the Iberian Peninsula as the origin of distinct evolutionary lineages, but also show that some of them may be of critical conservation interest.

Conclusions

Our results show the importance of performing genomic studies for properly defining evolutionary units of conservation value. This knowledge is crucial for species with endangered populations like the Pyrenean desman. The approach we used here included the sequencing by NGS techniques of specific short genomic fragments (around 45,000 fragments on average per specimen) and the generation of 1185 SNPs present in all samples. The genomic data obtained from 26 Pyrenean desmans selected from across their entire range revealed extremely low levels of

individual heterozygosity rate, with specimens from the southeastern Pyrenees having some of the lowest values inferred from genome-wide data in mammals so far. In addition, the analysis of genetic structure showed a strong geographical structure, with five main clusters that presented low levels of admixture among them and were largely coincident with the main mountain ranges. Preliminarily, and pending further analysis of more specimens from the whole distribution area, the populations found in each main mountain range should be considered as different conservation units. There is no perfect coincidence with the previously determined mitochondrial clades, particularly in the Iberian Range, demonstrating the importance of performing genomic analysis to reveal the whole population history of the species. The laboratory and bioinformatic techniques tested and improved here open the door to perform more specific analyses with additional specimens to cover poorly sampled regions. Among these, the Central System holds the most endangered population of the species according to current data therefore necessitating detailed genomic studies. In addition, the study of contact zones between different clusters would be of great help in defining the boundaries between them. These further analyses should help us better understand the dispersal patterns of the species and the true degree of isolation of the different populations. However, the overall geographical structure already uncovered here using genomic data should be taken into account in future management plans involving the Pyrenean desman.

Acknowledgments We thank the BTVS-ICNF collection (Banco de Tecidos de Vertebrados Selvagens—Instituto da Conservação da Natureza e das Florestas), Xunta de Galicia, Gobierno de Navarra, Diputación Foral de Gipuzkoa, Gobierno de La Rioja, Parque Nacional de Picos de Europa, Julio Gisbert and Rosa García-Perea (Proyecto Galemia) for *Galemys pyrenaicus* samples from their respective biological collections, and Henrique Carvalho, Carla Marisa Quaresma and Carlos P. Santos (Instituto da Conservação da Natureza e das Florestas) for their help with the access to the BTVS-ICNF collection. We also thank all the people of the Genomics Core Facility at the Pompeu Fabra University for help with the library preparation and sequencing. This work was financially supported by research Projects CGL2011-22640 and CGL2014-53968-P of the "Plan Nacional I + D+I del Ministerio de Economía y Competitividad" (Spain) to J.C. with contribution of project 014/2008 of the "Convocatoria de ayudas a proyectos de investigación en la Red de Parques Nacionales" (Spain) to J.G. M.Q. was supported by fellowship BES-2012-057135 of the Ministerio de Economía y Competitividad (Spain).

References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11:697–709. doi:10.1038/nrg2844

Allendorf FW, Luikart G, Aitken SN (2013) Conservation and the genetics of populations, 2nd edn. Wiley, Hoboken

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 25:3389–3402

Aymerich P, Gosálbez J (2014) El desmán ibérico *Galemys pyrenaicus* (É. Geoffroy Saint-Hilaire, 1811) en los Pirineos meridionales. In: Conservation and Management of semi-aquatic mammals of South-western Europe. Munibe Monographs. Nature Series 3. Aranzadi Society of Sciences, San Sebastian, pp 37–77

Aymerich P, Casadesús F, Gosálbez J (2001) Distribució de *Galemys pyrenaicus* (Insectivora, Talpidae) a Catalunya. Orsis 16:93–110

Buerkle CA, Gompert Z (2013) Population genomics based on low coverage sequencing: how low should we go? Mol Ecol 22:3028–3035. doi:10.1111/mec.12105

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) *Stacks*: building and genotyping loci *de novo* from short-read sequences. Genes Genom Genet 1:171–182. doi:10.1534/g3.111.000240

Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. doi:10.1111/mec.12354

Charbonnel A (2015) Influence multi-échelle des facteurs environnementaux dans la répartition du Desman des Pyrénées (*Galemys pyrenaicus*) en France. Dissertation, Université de Toulouse

Cho YS, Hu L, Hou H et al (2013) The tiger genome and comparative analysis with lion and snow leopard genomes. Nat Commun 4:2433. doi:10.1038/ncomms3433

Crandall KA, Bininda-Emonds ORP, Mace G, Wayne RK (2000) Considering evolutionary processes in conservation biology. Trends Ecol Evol 15:290–295

Dobrynin P, Liu S, Tamazian G et al (2015) Genomic legacy of the African cheetah, *Acinonyx jubatus*. Genome Biol 16:277. doi:10.1186/s13059-015-0837-4

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Felsenstein J (1989) PHYLIP-phylogeny inference package (version 3.4). Cladistics 5:164–166

Fernandes M, Herrero J, Aulagnier S, Amori G (2011) *Galemys pyrenaicus*. IUCN Red List of Threatened Species Version 20112 1–3

Freedman AH, Gronau I, Schweizer RM et al (2014) Genome sequencing highlights the dynamic early history of dogs. PLoS Genet 10:e1004016. doi:10.1371/journal.pgen.1004016

Funk WC, Mckay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends Ecol Evol 27:489–496. doi:10.1016/j.tree.2012.05.012

Gillet F (2015) Génétique et biologie de la conservation du desman des Pyrénées (*Galemys pyrenaicus*) en France. Dissertation, Université de Liège

Gisbert J, Garcia-Perea R (2014) Historia de la regresión del desmán ibérico *Galemys pyrenaicus* (É. Geoffroy Saint-Hilaire, 1811) en el Sistema Central (Península Ibérica). In: Conservation and Management of semi-aquatic mammals of South-western Europe. Munibe Monographs. Nature Series 3. Aranzadi Society of Sciences, San Sebastian, pp 19–35

Godinho R, Crespo EG, Ferrand N (2008) The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. Mol Ecol 17:4670–4683. doi:10.1111/j.1365-294X.2008.03929.x

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

González-Esteban J, Castién E, Gosálbez J (1999) Morphological and colour variation in the Pyrenean desman *Galemys pyrenaicus* (Geoffroy, 1811). Z Säugetierkunde 64:1–11

Green RE, Briggs AW, Krause J, Prüfer K, Burbano HA, Siebauer M, Lachmann M, Pääbo S (2009) The Neandertal genome and ancient DNA authenticity. EMBO J 28:2494–2502. doi:10.1038/emboj.2009.222

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. Nat Genet 43:1031–1034. doi:10.1038/ng.937

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. Nature 405:907–913. doi:10.1038/35016000

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267. doi:10.1093/molbev/msj030

Igea J, Aymerich P, Fernández-González A, González-Esteban J, Gómez A, Alonso R, Gosálbez J, Castresana J (2013) Phylogeography and postglacial expansion of the endangered semi-aquatic mammal *Galemys pyrenaicus*. BMC Evol Biol 13:115. doi:10.1186/1471-2148-13-115

Igea J, Aymerich P, Bannikova AA, Gosálbez J, Castresana J (2015) Multilocus species trees and species delimitation in a temporal context: application to the water shrews of the genus *Neomys*. BMC Evol Biol 15:209. doi:10.1186/s12862-015-0485-z

Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. Evolution 56:2383–2394

Kjeldsen SR, Zenger KR, Leigh K, Ellis W, Tobey J, Phalen D, Melzer A, FitzGibbon S, Raadsma HW (2015) Genome-wide SNP loci reveal novel insights into koala (*Phascolarctos cinereus*) population variability across its range. Conserv Genet 17:337–353. doi:10.1007/s10592-015-0784-3

Lanier HC, Massatti R, He Q, Olson LE, Knowles LL (2015) Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). Mol Ecol 24:3688–3705. doi:10.1111/mec.13270

Li R, Fan W, Tian G et al (2010) The sequence and *de novo* assembly of the giant panda genome. Nature 463:311–317. doi:10.1038/nature08696

Lopez-Fuster MJ, Garcia-Perea R, Fernández-Salvador R, Gisbert J, Ventura J (2006) Craniometric variability of the Iberian desman, *Galemys pyrenaicus* (Mammalia: Erinaceomorpha: Talpidae). Folia Zool 55:29–42

Lynch M (1991) The genetic interpretation of inbreeding depression and outbreeding depression. Evolution 45:622–629

Melero Y, Aymerich P, Luque-Larena JJ, Gosálbez J (2012) New insights into social and space use behaviour of the endangered Pyrenean desman (*Galemys pyrenaicus*). Eur J Wildl Res 58:185–193. doi:10.1007/s10344-011-0561-7

Morin PA, Martien KK, Taylor BL (2009) Assessing statistical power of SNPs for population structure and conservation studies. Mol Ecol Resour 9:66–73. doi:10.1111/j.1755-0998.2008.02392.x

Nater A, Nietlisbach P, Arora N, van Schaik CP, van Noordwijk MA, Willems EP, Singleton I, Wich SA, Goossens B, Warren KS, Verschoor EJ, Perwitasari-Farajallah D, Pamungkas J, Krützen M (2011) Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant Orangutans (genus: *Pongo*). Mol Biol Evol 28:2275–2288. doi:10.1093/molbev/msr042

Nores C, Queiroz AI, Gisbert J (2007) *Galemys pyrenaicus*. Atlas y libro rojo de los mamíferos terrestres de España, pp 92–98

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol 22:5983–5999. doi:10.1111/mec.12561

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. Trends Genet 26:177–187. doi:10.1016/j.tig.2010.01.001

Palmeirim JM, Hoffmann RS (1983) *Galemys pyrenaicus*. Mamm Species 207:1–5

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135. doi:10.1371/journal.pone.0037135

Prado-Martinez J, Sudmant PH, Kidd JM et al (2013) Great ape genetic diversity and population history. Nature 499:471–475. doi:10.1038/nature12228

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

Robinson JA, Ortega-Del Vecchyo D, Fan Z, Kim BY, Vonholdt BM, Marsden CD, Lohmueller KE, Wayne RK (2016) Genomic flatlining in the endangered Island Fox. Curr Biol 26:1183–1189. doi:10.1016/j.cub.2016.02.062

Shafer ABA, Wolf JBW, Alves PC et al (2015) Genomics and the challenging translation into conservation practice. Trends Ecol Evol 30:78–87. doi:10.1016/j.tree.2014.11.009

Sovic MG, Fries AC, Gibbs HL (2015) AftrRAD: a pipeline for accurate and efficient *de novo* assembly of RADseq data. Mol Ecol Resour 15:1163–1171. doi:10.1111/1755-0998.12378

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690. doi:10.1093/bioinformatics/btl446

Steiner CC, Putnam AS, Hoeck PEA, Ryder OA (2013) Conservation genomics of threatened animal species. Annu Rev Anim Biosci 1:261–281. doi:10.1146/annurev-animal-031412-103636

Tallmon DA, Luikart G, Waples RS (2004) The alluring simplicity and complex reality of genetic rescue. Trends Ecol Evol 19:489–496. doi:10.1016/j.tree.2004.07.003

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol 21:3907–3930. doi:10.1111/j.1365-294X.2012.05664.x

Vonholdt BM, Pollinger JP, Earl DA, Knowles JC, Boyko AR, Parker H, Geffen E, Pilot M, Jedrzejewski W, Jedrzejewska B, Sidorovich V, Greco C, Randi E, Musiani M, Kays R, Bustamante CD, Ostrander EA, Novembre J, Wayne RK (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Res 21:1294–1305. doi:10.1101/gr.116301.110

Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, Frandsen P, Chen Y, Yngvadottir B, Cooper DN, de Manuel M, Hernandez-Rodriguez J, Lobon I, Siegismund HR, Pagani L, Quail MA, Hvilsom C, Mudakikwa A, Eichler EE, Cranfield MR, Marques-Bonet T, Tyler-Smith C, Scally A (2015) Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science 348:242–245. doi:10.1126/science.aaa3952

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328. doi:10.1093/bioinformatics/bts606

Supplementary Methods, Tables and Figures

Next-generation sequencing in conservation: genomic diversity and geographical structure of the Pyrenean desman

Marina Querejeta, Jorge González Esteban, Asunción Gómez, Angel Fernández-González, Pere Aymerich, Joaquim Gosálbez, Lídia Escoda, Javier Igea, Jose Castresana

Supplementary Methods

Quantification of DNA concentration by qPCR

We selected an ultraconserved element among mammals (Chr2_23668) (Faircloth et al. 2012) for qPCR amplification. The primers (AAATGCAGCGATCAGCAGT / ACGGGTGCCACATGTTAAG) were designed with the software Primer3 v. 0.4.0 (Koressaar and Remm 2007; Untergasser et al. 2012) to amplify a 76 bp fragment. The qPCR with these primers was successfully tested in different species of insectivores, rodents and carnivores.

qPCR reactions were performed in a final volume of 20 µl, containing 10 µl of IQ SYBR Green Supermix (Bio-Rad), 300 nM of each primer, and 2 µl of genomic DNA (prediluted 100 fold). The reactions were set in triplicates and run on a MyiQ Single-Color Real-Time PCR Detection System (Bio-Rad) with a thermocycling protocol of 2 min at 95ºC for the initial denaturation followed by 40 cycles of denaturation (15 s at 95 ºC), and annealing/extension (30 s at 60 ºC). The standard curve for absolute quantification was prepared using a commercial *Bos taurus* DNA sample of known concentration (Sigma-Aldrich) consisting of a ten-fold dilution series ranging from 2 ng to 0.02 ng and a non-template control.

Sex determination by qPCR

In order to sex *Galemys pyrenaicus* samples we needed to design primers for amplification of a Y chromosome fragment. Since few Y chromosome sequences of mammals were available in databases for primer design, we used a heterologous gene with copies present in both X and Y chromosomes. Following the approach used in Igea et al. (2010) based on the comparison of several mammalian genomes, we selected a gene located in the X-chromosome, USP9X, with an heterologous copy in the Y-chromosome, USP9Y. Alignments of the adjacent exon sequences were used to design primers in the conserved regions that simultaneously amplified intron 42 of both heterologous copies of the gene. Then, intron 42 of USP9X and USP9Y was amplified in several male *G. pyrenaicus* samples (with sex morphologically determined). The resulting PCR product was subcloned into a pstBlue-1 vector (Invitrogen) to obtain the sequences belonging to the X- and Y-specific introns. With these Y intron sequences of *G. pyrenaicus*, we could design new highly specific primers for qPCR amplification of a smaller fragment within the intron. These primers (GACAGCTTCCAAAATAAAGAATTT / GAACTGGCAGTAATTTTCAAAGTG) were designed with the software Primer3 v. 0.4.0 (Koressaar and Remm 2007; Untergasser et al. 2012) and amplified a 77 bp region of intron 42 of USP9Y only in males. qPCR reactions were performed as described above. The standard curve for Y-chromosome quantification was prepared using a male *G. pyrenaicus* DNA sample of known concentration consisting of a ten-fold dilution series ranging from 8.5 ng to 0.085 ng and a non-template control.

References

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst Biol 61:717–726. doi: 10.1093/sysbio/sys004

Igea J, Juste J, Castresana J (2010) Novel intron markers to study the phylogeny of closely related mammalian species. BMC Evol Biol 10:369. doi: 10.1186/1471-2148-10-369

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–1291. doi: 10.1093/bioinformatics/btm091

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3-new capabilities and interfaces. Nucleic Acids Res 40:e115. doi: 10.1093/nar/gks596

Table S1. Specimens used in this study.

Specimen code	Geographical location	Locality	Latitude	Longitude	Reference
IBE-C422	Pyrenees	Tor	42.6	1.4	Igea el al. 2013
IBE-C2756	Pyrenees	Rio Aiaiturrieta-Ataun	43.0	-2.1	Igea el al. 2013
IBE-C2763	Pyrenees	Rio Olazar-Eugi	43.0	-1.5	Igea el al. 2013
IBE-S1925	Pyrenees	Riu Ritort-Camprodon	42.3	2.4	Igea el al. 2013
IBE-C2755	Pyrenees	Rio Leitzaran-Berastegi	43.2	-2.0	Igea el al. 2013
IBE-C2799	Pyrenees	Rio Elama-Artikutza	43.2	-1.8	Igea el al. 2013
IBE-C2757	Pyrenees	Amundarain-Zaldibia	43.1	-2.0	Igea el al. 2013
IBE-C2765	Pyrenees	Rio Ezpelura-Urrotz	43.1	-1.7	Igea el al. 2013
IBE-C2739	Occidental	Rio Cabras-Laza	42.1	-7.4	Igea el al. 2013
IBE-C3305	Occidental	Paiva-Fraguas	40.8	-7.8	Igea el al. 2013
IBE-C2736	Occidental	Rio Caranyo-Covelo	42.3	-8.4	Igea el al. 2013
IBE-C2737	Occidental	Rio Riobo-A Estrada	42.7	-8.4	Igea el al. 2013
IBE-C3301	Occidental	Paiva-Fraguas	40.8	-7.8	Igea el al. 2013
IBE-C2740	Occidental	Rio Meladas-Carballeda	42.2	-6.8	Igea el al. 2013
IBE-C2749	Occidental	Rio Termes-As Neves	42.1	-8.4	Igea el al. 2013
IBE-C3303	Occidental	Sabor-Macas-Quintanilla	41.8	-6.6	Igea el al. 2013
IBE-C2856	Cantabr. Mount.	Rio Lamedo	43.1	-4.6	Igea el al. 2013
IBE-C2892	Cantabr. Mount.	Rio Cares-Cain	43.2	-4.9	Igea el al. 2013
IBE-C2894	Cantabr. Mount.	Rio Cares-Casiellos	43.2	-4.9	Igea el al. 2013
IBE-C3739	Iberian Range	Ciloria	42.3	-3.1	This work
IBE-C3750	Iberian Range	Iregua	42.1	-2.7	This work
IBE-C3756	Iberian Range	La Soledad	42.2	-3.1	Igea el al. 2013
IBE-C3754	Iberian Range	La Vieja	42.0	-2.6	This work
IBE-C3774	Iberian Range	Iregua-Achichuelo	42.1	-2.7	This work
IBE-C1069	Iberian Range	Rio Calamantio	42.2	-2.9	Igea el al. 2013
IBE-C485	Central System	Rio Ambroz-Hervas	40.3	-5.8	Igea el al. 2013

Igea, J., P. Aymerich, A. Fernández-González, J. González-Esteban, A. Gómez, R. Alonso, J. Gosálbez, and J. Castresana (2013). Phylogeography and postglacial expansion of the endangered semi-aquatic mammal *Galemys pyrenaicus*. BMC Evol. Biol. 13:115.

Table S2. Basic statistics of the obtained sequences.

Specimen code	Total reads	Assembled reads	Assembled loci	Coverage	Loci (r=0.51)	% Loci
IBE-C422	3,340,793	3,168,467	73,542	43.1	52,557	92
IBE-C2756	5,488,309	4,870,237	95,669	50.9	54,436	95
IBE-C2763	1,131,290	981,430	68,875	14.2	50,954	89
IBE-S1925	1,479,918	1,358,615	63,422	21.4	49,356	86
IBE-C2755	1,621,373	1,508,894	56,349	26.8	45,188	79
IBE-C2799	1,026,682	926,741	37,741	24.6	29,730	52
IBE-C2757	4,267,442	4,025,554	86,240	46.7	54,874	96
IBE-C2765	977,876	797,482	55,921	14.3	43,749	76
IBE-C2739	2,933,094	2,807,102	70,351	39.9	50,841	89
IBE-C3305	719,617	485,621	39,441	12.3	28,433	50
IBE-C2736	1,104,564	958,648	66,385	14.4	50,602	88
IBE-C2737	2,131,404	1,953,263	78,034	25.0	54,517	95
IBE-C3301	2,889,024	2,599,493	89,172	29.2	54,918	96
IBE-C2740	2,918,438	2,502,729	73,618	34.0	37,494	66
IBE-C2749	2,271,888	1,950,688	62,010	31.5	35,048	61
IBE-C3303	875,577	406,220	48,293	8.4	30,946	54
IBE-C2856	2,457,624	2,279,458	78,800	28.9	55,170	96
IBE-C2892	223,377	141,154	22,687	6.2	20,256	35
IBE-C2894	241,988	145,557	22,639	6.4	20,401	36
IBE-C3739	1,420,150	1,290,934	68,342	18.9	52,690	92
IBE-C3750	2,004,518	1,828,753	77,181	23.7	54,373	95
IBE-C3756	6,180,801	5,587,283	99,216	56.3	56,378	99
IBE-C3754	7,932,104	7,367,368	98,576	74.7	54,898	96
IBE-C3774	3,134,533	2,866,594	82,863	34.6	48,612	85
IBE-C1069	2,054,961	974,520	107,903	9.0	45,519	80
IBE-C485	1,724,492	1,612,349	73,615	21.9	49,651	87
Average	2,405,840	2,130,583	69,111	28	45,446	79
Total	62,551,837	55,395,154	1,796,885			

Table S3. Sex determination by Blast against Y-linked loci.

Specimen code	Geographical location	Number of hits per million reads	Sex
IBE-C422	Pyrenees	0.0	Female
IBE-C2756	Pyrenees	0.2	Female
IBE-C2763	Pyrenees	563.2	Male
IBE-S1925	Pyrenees	819.9	Male
IBE-C2755	Pyrenees	924.9	Male
IBE-C2799	Pyrenees	898.4	Male
IBE-C2757	Pyrenees	0.2	Female
IBE-C2765	Pyrenees	646.5	Male
IBE-C2739	Occidental	779.6	Male
IBE-C3305	Occidental	615.7	Male
IBE-C2736	Occidental	0.0	Female
IBE-C2737	Occidental	486.6	Male
IBE-C3301	Occidental	0.3	Female
IBE-C2740	Occidental	2.4	Female
IBE-C2749	Occidental	0.9	Female
IBE-C3303	Occidental	326.7	Male
IBE-C2856	Cantabr. Mount.	524.6	Male
IBE-C2892	Cantabr. Mount.	752.3	Male
IBE-C2894	Cantabr. Mount.	0.0	Female
IBE-C3739	Iberian Range	0.0	Female
IBE-C3750	Iberian Range	527.4	Male
IBE-C3756	Iberian Range	653.6	Male
IBE-C3754	Iberian Range	0.3	Female
IBE-C3774	Iberian Range	0.0	Female
IBE-C1069	Iberian Range	0.5	Female
IBE-C485	Central System	1.2	Female

Table S4. Proportion of heterozygous positions of each specimen (per position).

Specimen code	Geographical location	Heterozygosity rate
IBE-C422	Pyrenees	0.000020
IBE-C2756	Pyrenees	0.000037
IBE-C2763	Pyrenees	0.000127
IBE-S1925	Pyrenees	0.000013
IBE-C2755	Pyrenees	0.000114
IBE-C2799	Pyrenees	0.000132
IBE-C2757	Pyrenees	0.000024
IBE-C2765	Pyrenees	0.000138
IBE-C2739	Occidental	0.000389
IBE-C3305	Occidental	0.000488
IBE-C2736	Occidental	0.000351
IBE-C2737	Occidental	0.000348
IBE-C3301	Occidental	0.000425
IBE-C2740	Occidental	0.000390
IBE-C2749	Occidental	0.000359
IBE-C3303	Occidental	0.000407
IBE-C2856	Cantabr. Mount.	0.000301
IBE-C2892	Cantabr. Mount.	0.000257
IBE-C2894	Cantabr. Mount.	0.000271
IBE-C3739	Iberian Range	0.000329
IBE-C3750	Iberian Range	0.000278
IBE-C3756	Iberian Range	0.000368
IBE-C3754	Iberian Range	0.000159
IBE-C3774	Iberian Range	0.000208
IBE-C1069	Iberian Range	0.000303
IBE-C485	Central System	0.000147
Average		0.000246

Table S5. Admixture proportions of each specimen estimated with Structure and K = 5.

Specimen code	Geographical location	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
IBE-C422	Pyrenees	0.0001	0.9996	0.0001	0.0001	0.0001
IBE-C2756	Pyrenees	0.0002	0.9988	0.0001	0.0006	0.0001
IBE-C2763	Pyrenees	0.0017	0.9702	0.0003	0.0273	0.0005
IBE-S1925	Pyrenees	0.0001	0.9996	0.0001	0.0001	0.0001
IBE-C2755	Pyrenees	0.0003	0.9984	0.0002	0.0010	0.0002
IBE-C2799	Pyrenees	0.0002	0.9986	0.0001	0.0009	0.0002
IBE-C2757	Pyrenees	0.0001	0.9996	0.0001	0.0001	0.0001
IBE-C2765	Pyrenees	0.0004	0.9974	0.0001	0.0019	0.0002
IBE-C2739	Occidental	0.0014	0.0005	0.0002	0.0004	0.9974
IBE-C3305	Occidental	0.0073	0.0002	0.0003	0.0003	0.9919
IBE-C2736	Occidental	0.0003	0.0003	0.0002	0.0003	0.9989
IBE-C2737	Occidental	0.0008	0.0002	0.0006	0.0004	0.9980
IBE-C3301	Occidental	0.0008	0.0005	0.0075	0.0004	0.9908
IBE-C2740	Occidental	0.0408	0.0006	0.0004	0.0007	0.9576
IBE-C2749	Occidental	0.0003	0.0002	0.0002	0.0002	0.9991
IBE-C3303	Occidental	0.0013	0.0003	0.0004	0.0003	0.9976
IBE-C2856	Cantabr. Mount.	0.6772	0.1141	0.0002	0.0061	0.2024
IBE-C2892	Cantabr. Mount.	0.9984	0.0006	0.0002	0.0004	0.0004
IBE-C2894	Cantabr. Mount.	0.9987	0.0004	0.0004	0.0002	0.0003
IBE-C3739	Iberian Range	0.0311	0.1980	0.0089	0.7606	0.0013
IBE-C3750	Iberian Range	0.0012	0.0675	0.0028	0.9282	0.0003
IBE-C3756	Iberian Range	0.1047	0.2262	0.0014	0.6566	0.0111
IBE-C3754	Iberian Range	0.0002	0.0026	0.0002	0.9967	0.0002
IBE-C3774	Iberian Range	0.0004	0.0201	0.0003	0.9789	0.0003
IBE-C1069	Iberian Range	0.0006	0.2126	0.0003	0.7863	0.0002
IBE-C485	Central System	0.0005	0.0014	0.9956	0.0019	0.0007

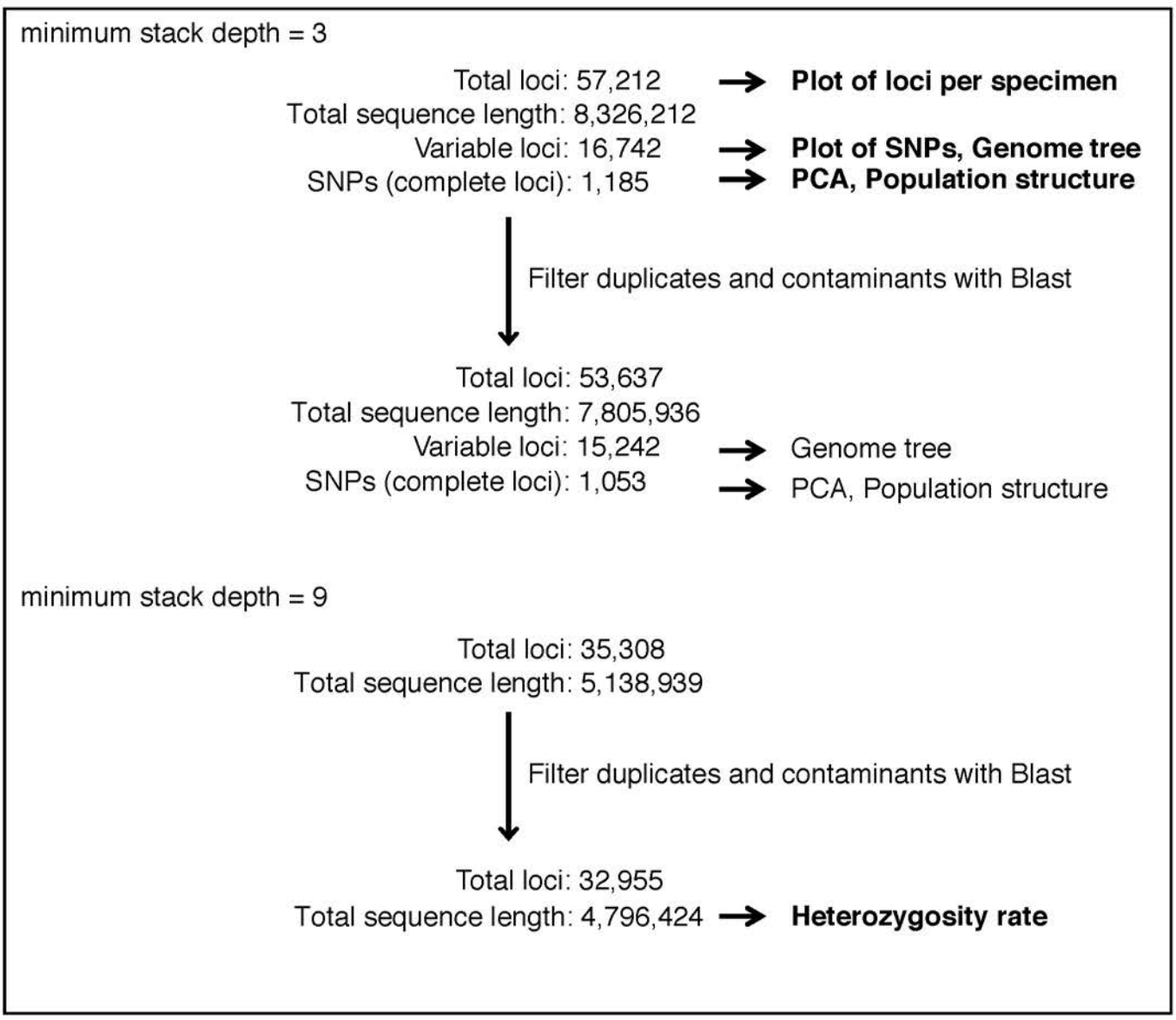


Figure S1. Scheme of the steps followed to assemble loci using two different minimum stack depths and analyses performed with different datasets. Main analyses are shown with bold font.

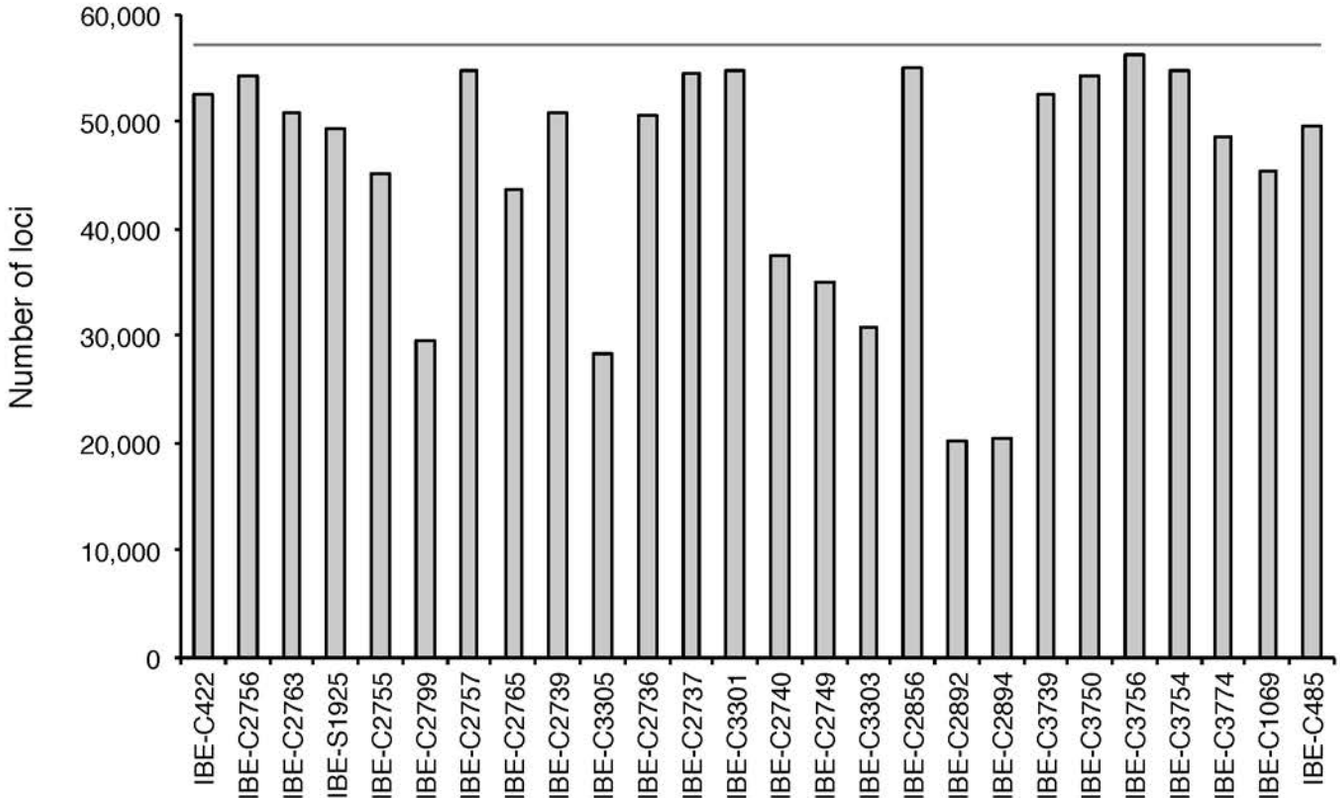


Figure S2. Plot of loci per specimen: Number of loci represented in each specimen after merging all loci. Only loci assembled in more than half the number of samples (r = 0.51 in the *populations* program), are used. The horizontal bar represents total loci, which were 57,212.

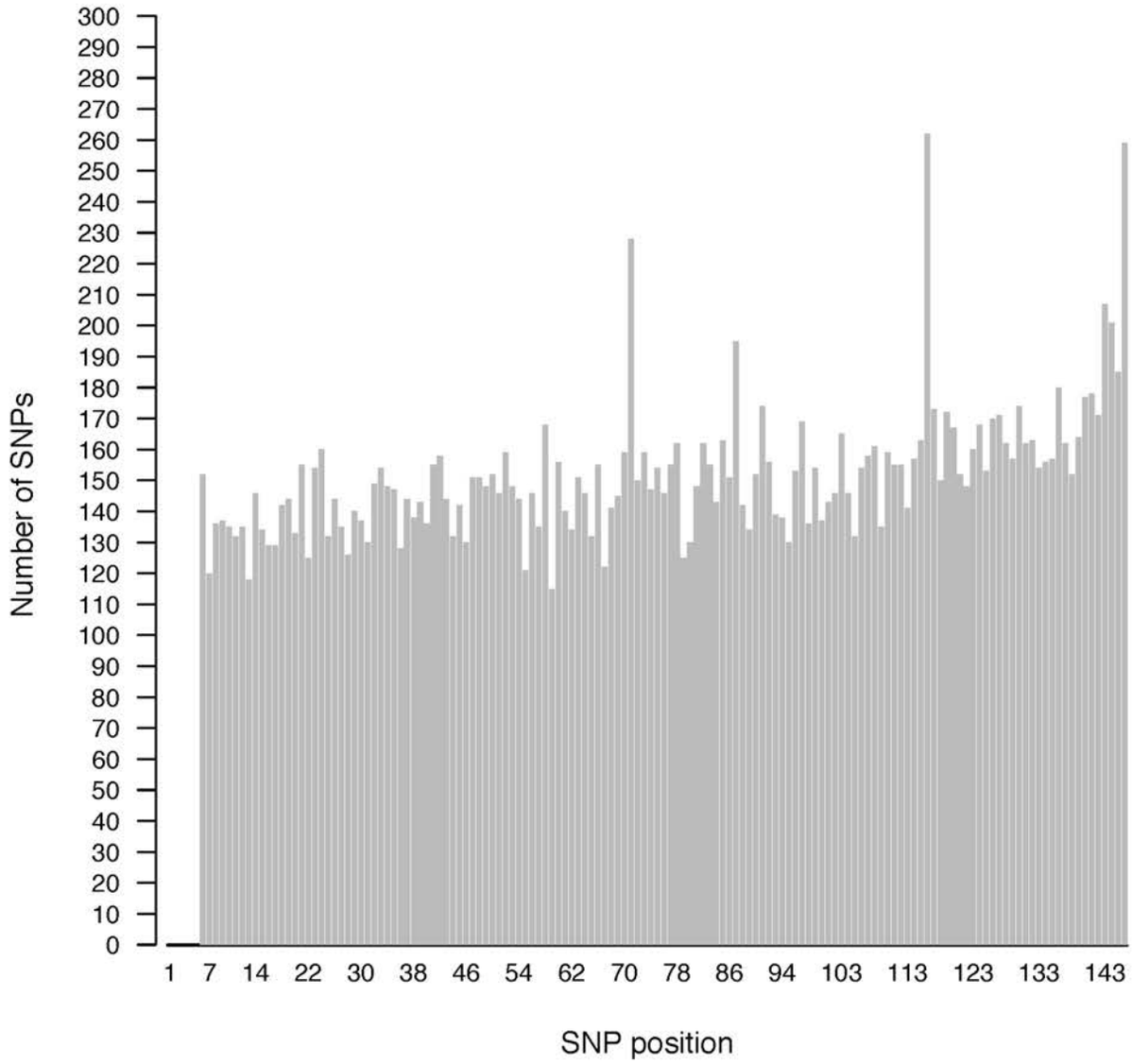


Figure S3. Plot of SNPs: Number of SNPs found along the sequence reads. The first five bases correspond to the restriction site and do not have any SNP.

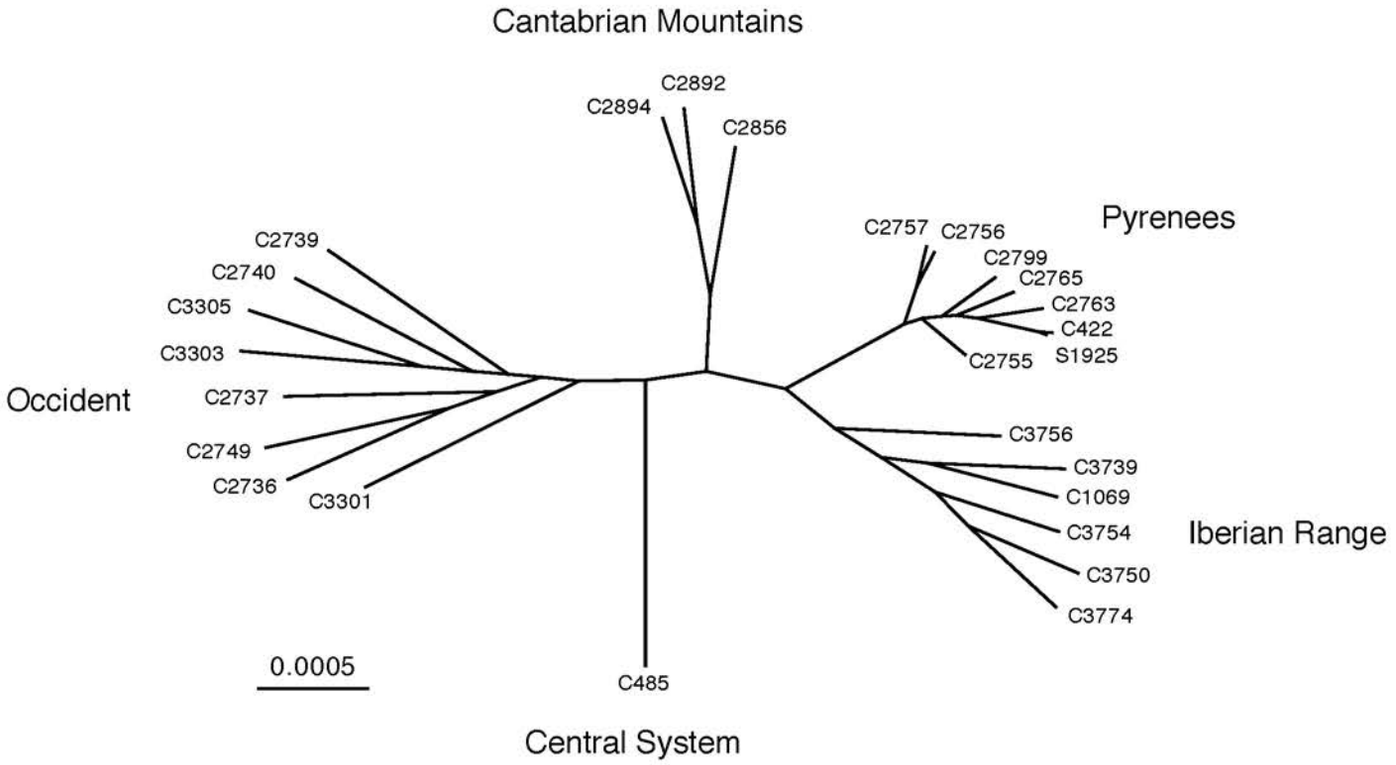


Figure S4. Genomic tree obtained by maximum likelihood of one of the possible concatenations of the two alleles of the 16,742 variable loci. The scale is in substitutions per position. Several concatenations in which the order of each allele pair was randomly changed yielded very similar trees.

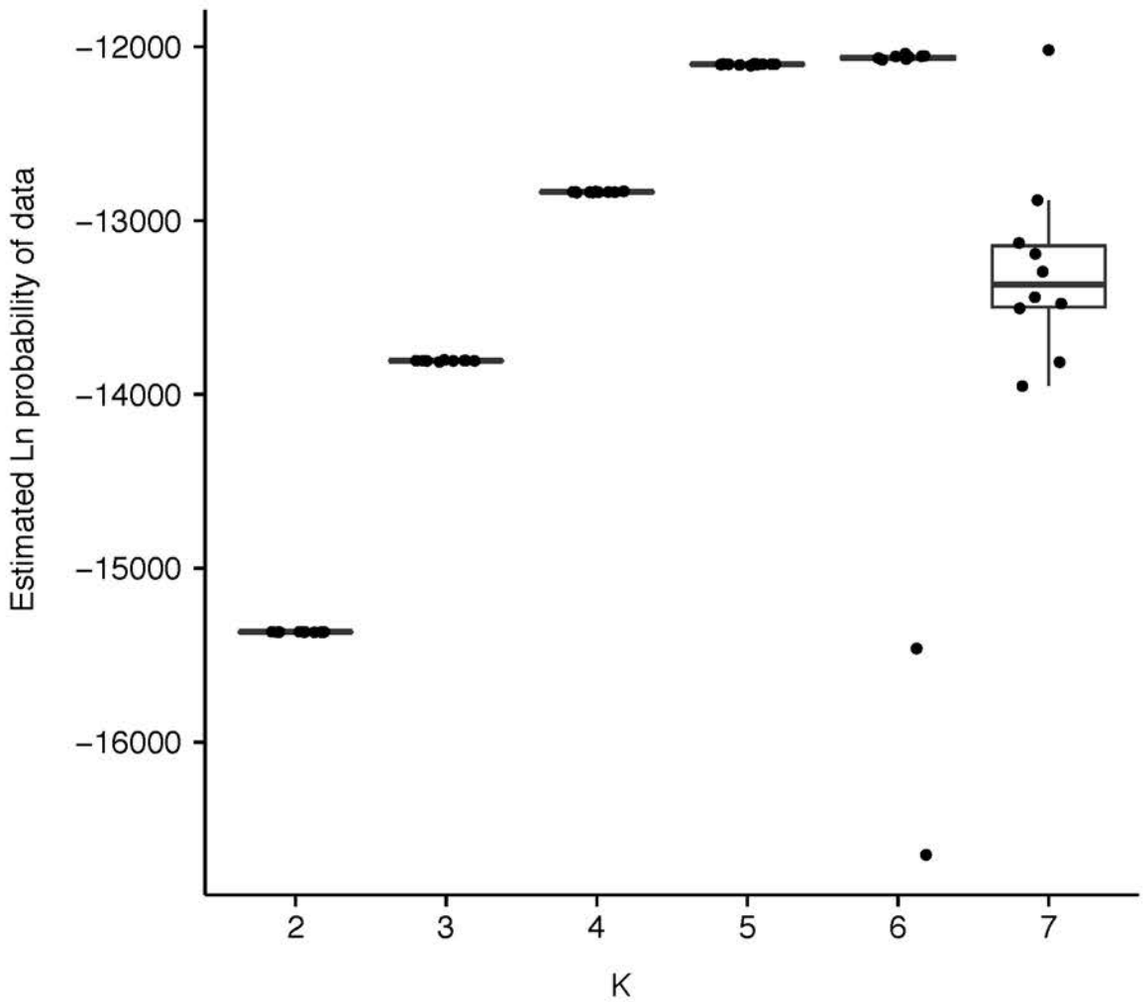


Figure S5. Box plots of the estimated posterior probability of the data for different K values in Structure. Values for each of 10 different runs are jittered for better visibility.

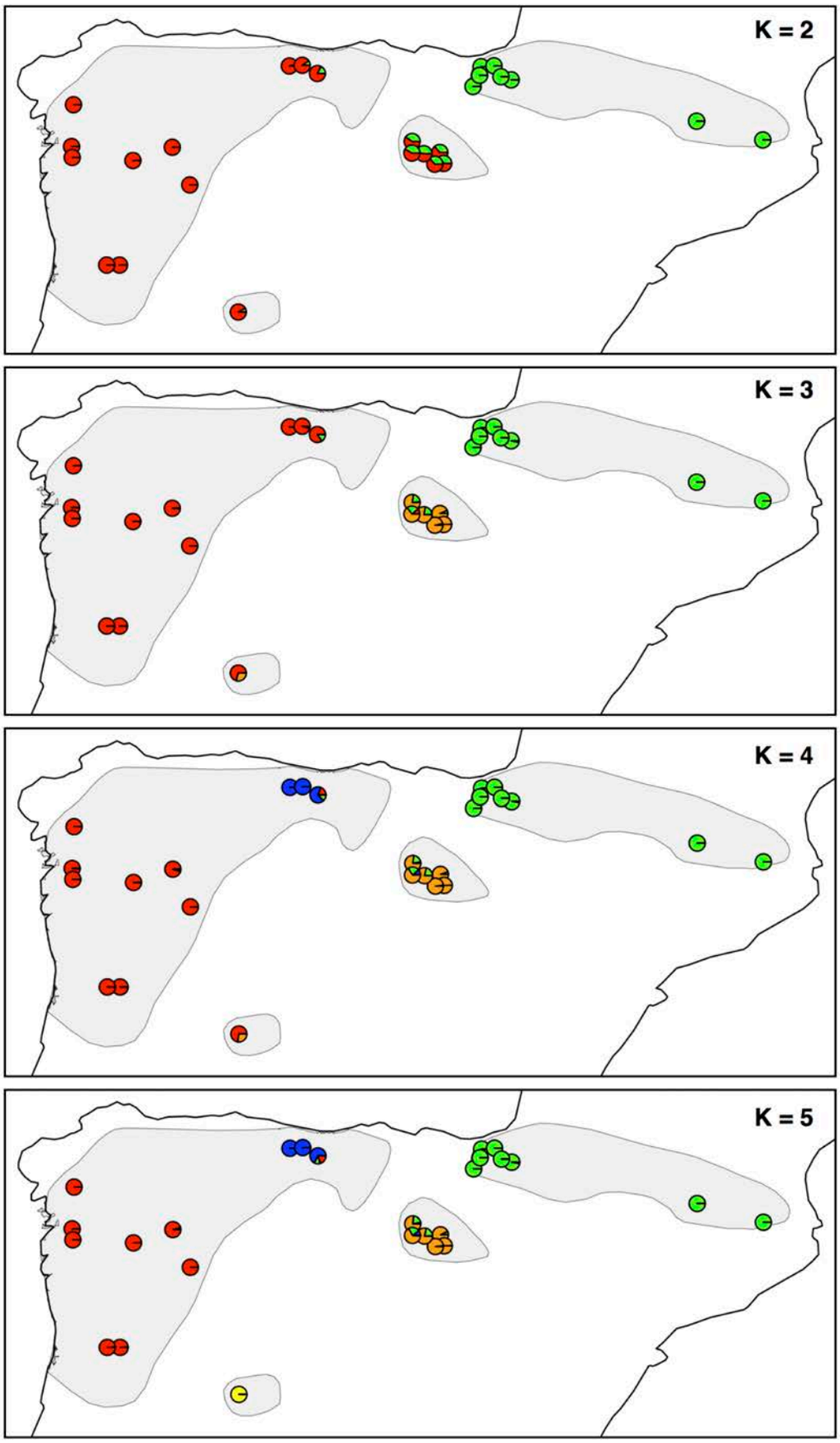


Figure S6. Maps plotting color-coded admixture proportions of each specimen as determined with Structure and different K values.

CHAPTER 2

POSTGLACIAL DISPERSAL PATTERNS AND MITOCHONDRIAL GENETIC STRUCTURE OF THE PYRENEAN DESMAN (*Galemys pyrenaicus*) IN THE NORTH-WESTERN REGION OF THE IBERIAN PENINSULA

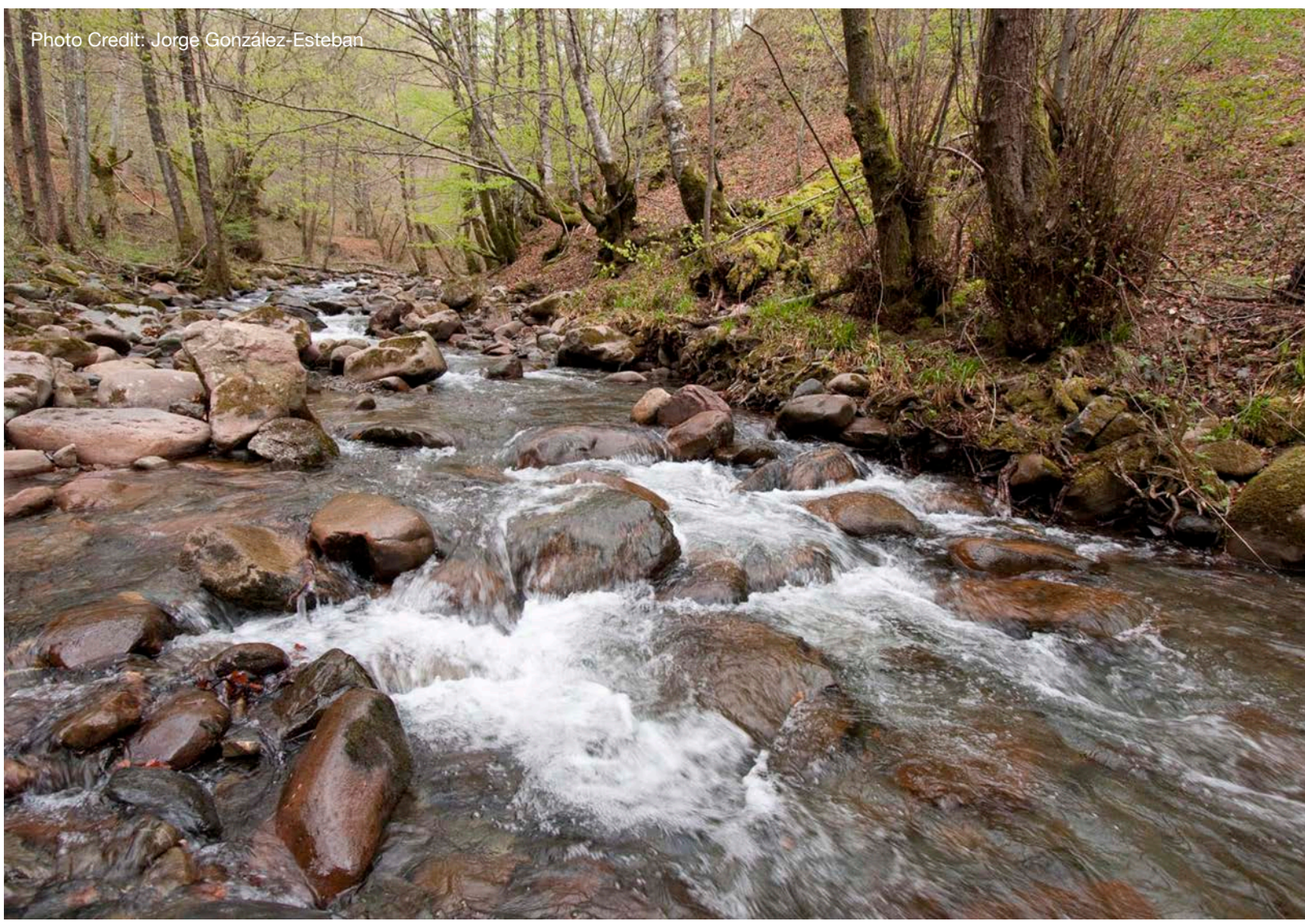


Photo Credit: Jorge González-Esteban

Marina Querejeta [1], Angel Fernández-González [2], Rafael Romero [3] and Jose Castresana [1]

[1] Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain

[2] Biosfera Consultoría Medioambiental S.L., Calle Candamo 5, 33012 Oviedo, Spain

[3] Calle Presidente Salvador Allende 13, 15705 Santiago de Compostela, Spain

ABSTRACT

The genetic structure of small semiaquatic animals may be influenced by dispersal across both rivers and land. The relative importance of these two modes of dispersal may vary across different species and with ecological conditions and evolutionary periods. The Pyrenean desman (*Galemys pyrenaicus*) is an endemic mammal of the Iberian Peninsula with a strong phylogeographic structure and semiaquatic habits, thus making it an ideal model to study the effect of river and overland dispersal on its genetic structure. Thanks to different types of non-invasive samples, we obtained an extensive sampling of the Pyrenean desman from the north-western region of the Iberian Peninsula and sequenced two mitochondrial DNA fragments. We then analysed, using an isolation-by-distance approach, the correlation between phylogenetic distances and geographical distances measured along both river networks and land to infer the relative importance of river and overland dispersal. We found that the correlations in the whole area and in a large basin were consistent with an effect of overland dispersal, which may be due to the frequent postglacial colonization of new territories using terrestrial corridors, likely facilitated by a more extensive fluvial network that may have been present in the Holocene. However, in a small basin, likely to be less influenced by the impact of ancient postglacial dispersal, the correlations suggested a significant effect of both overland and river dispersal, as expected for a semi-aquatic mammal. Therefore, different scales reflect different aspects of the evolutionary history and ecology of this semiaquatic species. The results we obtained may have crucial implications for the conservation of the Pyrenean desman because they reinforce the importance of inter-basin dispersal for this species and the need to protect, not only the riverine ecosystems, but also the upper parts of rivers and riparian corridors between basins, which may have a great potential for the dispersal of the species.

Keywords: dispersal, *Galemys pyrenaicus*, genetic diversity, Iberian Peninsula, isolation-by-distance, mitochondrial DNA

INTRODUCTION

One of the greatest challenges in phylogeography is to understand how geography and ecology shape the genetic structure of species (Avise 2009). Indeed, geography in particular can affect genetic structure in many ways, such as by inducing the fragmentation of populations and shaping the dispersal patterns among them (Manel et al. 2003). This is particularly important for riverine species, which usually have strict ecological requirements. When the river networks have suitable physical and ecological conditions for the species, they can connect populations from different locations of the network (Chaput-Bardy et al. 2008; Paz-Vinas et al. 2015; Byrne et al. 2015). Rivers can also act as barriers to dispersal of aquatic species when their width, depth, water regimes or the presence of water infrastraures are inappropriate for the species and thus have an extreme influence on genetic structure (Raeymaekers et al. 2008; Bartáková et al. 2015; Byrne et al. 2015).

Despite its interest, there are still few studies that have analysed the effects of river systems on the genetic structure of semiaquatic mammals, in which their degree of dependency on the aquatic medium is crucial in understanding when rivers act as barriers and when they are fundamental for their dispersal (Furlan et al. 2013; Senn et al. 2014; Pagacz 2016).

In this study, we focus on the Pyrenean desman [*Galemys pyrenaicus* (Geoffroy, 1811)], a small semiaquatic mammal endemic to the Iberian Peninsula. The Pyrenean desman is highly dependent on specific features of the rivers where it lives, generally preferring small rivers and streams with constant flow through the year and certain slope to favour oxygenation of the waters (Palmeirim & Hoffmann 1983; Charbonnel et al. 2015). It presents strong adaptations to the aquatic ecosystems such as a highly-mobile protracted snout, large hindfeet and a long tail with stiff hairs (Palmeirim & Hoffmann 1983). Due the fragility of its habitat, this species is endangered in part of its distribution range and is considered as Vulnerable at the global scale by the IUCN red list (Fernandes et al. 2011). According to the IUCN, major threats include water pollution, increased water extraction for irrigation, and habitat fragmentation due to the construction of hydroelectric plants and dams. In addition, predation by American mink *(Neovison vison)* might be affecting populations where this invasive species is abundant. Climate change is anticipated to be a threat to the desman in the future, as the species tends to occur in areas with annual rainfall superior to 1000 mm and, given climate scenarios for the Iberian Peninsula, the species may decrease its distribution in the most vulnerable areas (Fernandes et al. 2011). Previous studies on the genetic structure and diversity of this species revealed a strong phylogeographic structure (Igea et al. 2013; Querejeta et al. 2016). According to mitochondrial data, there are two main groups (A and B), which are further subdivided into four subgroups (A1, A2, B1 and B2) of allopatric distribution and likely glacial origin (Igea et al. 2013). The populations with the largest genetic diversity were detected in the occidental part of the distribution, thereby suggesting that one of the most important glacial refugia was in this area. Interestingly, the contact zones between the main mitochondrial groups (one located in the Cantabrian Mountains and the other in the Iberian Range) indicated an almost complete absence of spatial mixing between these maternal lineages after postglacial recolonization (Igea et al. 2013). A genomic analysis based on ddRAD sequences and SNPs corroborated the existence of a strong structure (Querejeta et al. 2016). The high genetic diversity of the occidental population compared to other populations was confirmed by the same genomic analysis (Querejeta et al. 2016), thus making this population of great interest for studying the interaction between genetic structure and dispersal patterns across rivers and land.

Previous studies showed that the overall genetic structure of the Pyrenean desman was basically associated with the main mountain ranges, whereas the influence of river basins on the partition of the genetic diversity was smaller than previously believed (Igea et al. 2013; Querejeta et al. 2016). However, the sampling density in these studies was low and the study of the effects that rivers may have on the genetic structure at smaller scales was not possible. In this work, we obtained an

extensive sampling of the Pyrenean desman in the occidental population thanks to different types of non-invasive samples and studied how the river networks shaped the phylogeographic pattern of this semiaquatic species. We aimed to answer two basic questions: (1) what is the genetic structure and the variability of the genetic diversity in the Pyrenean desman in the north-western region of the Iberian Peninsula, and (2) whether this genetic structure is more influenced by overland or river dispersal.

Our approach derives from the well-known isolation-by-distance model (Wright 1943), which is based on the premise that genetic isolation increases with geographical distance and predicts genetic similarity at shorter distances as well as differentiation at greater distances due to limited dispersal. In principle, Euclidean distances can be used to infer isolation by distance for terrestrial organisms. However, it has been observed that distances that account for the landscape structure can better predict isolation by distance (Coulon et al. 2004). Furthermore, for aquatic and semiaquatic organisms, distances across the river network should be taken into account (Vignieri 2005). Following these previous works, we studied for the occidental clade of the Pyrenean desman the correlation between genetic distance and three kinds of geographic distances. Firstly, we calculated the Euclidean distance, as a simple measure of the geographical distance between two samples. We also calculated a least-cost path distance (Adriaensen et al. 2003) that accounts for altitude as only covariate; this distance complements the Euclidean distance as a better approximation for the geographical distance in mountain areas. Finally, we calculated the river distance along the river networks, which we hypothesized to be a highly important variable for a semi-aquatic species (Figure 1). A high correlation between genetic distance and Euclidean and/or altitudinal distance would be consistent with an important effect of an isolation-by-distance relationship and dispersal occurring overland, whereas a high correlation between genetic distance and river distance would suggest isolation by distance with dispersal taking place along rivers. We also discuss the interest of this approach to discern the evolutionary and ecological factors involved in the dispersal of semiaquatic species and the importance that this knowledge may have for conservation planning of the Pyrenean desman.

METHODS

SAMPLING, DNA EXTRACTION AND AMPLIFICATION

A total of 192 samples from Pyrenean desmans of the north-western part of the Iberian Peninsula (Spain and Portugal) were used (see Table S1 and Figure S1 of Supporting Information). Most of these samples (131) were collected as part of two LIFE+ projects dedicated to the study of the Pyrenean desman: DESMANIA (LIFE11 NAT/ES/000691) and MARGALULLA (LIFE09 NAT/ES/000514). Of these samples, 84 were obtained from faeces that desmans deposit on emerged rocks in the rivers. A further 47 samples were from depredated desmans found in the excrements of carnivores,

likely Eurasian otters, which were collected in rivers (Callejo & Delibes 1987). The dataset was complemented with the sequences from 61 samples of a previous study of the Pyrenean desman (Igea et al. 2013).

DNA of the new samples was extracted using the DNeasy Blood and Tissue kit from QIAGEN, following the manufacturer's instructions, in a separated UV-irradiated area with dedicated equipment. Desman faeces were treated as described in Igea et al. (2013). Carnivore faeces containing depredated desmans were cleaned, and hair and tissue of desman origin were separated for extraction. This material was then processed as for the other samples.

The last 724 bp of the mitochondrial cytochrome b gene, which is the fragment with the highest number of variable positions, was amplified using two PCR reactions, and the first 342 bp of the D-loop fragment was amplified in another PCR. PCR conditions, primers and sequencing procedures can be found in Igea et al. (2013). The sequences were edited and assembled using GENIOUS PRO (Biomatters Ltd). The final sequences of both genes obtained in this and previous work were concatenated, resulting in an alignment of 1066 bp for 192 sequences. New sequences were deposited in GenBank under accession numbers X-X.

PHYLOGENETIC ANALYSES

The concatenated alignment of cytochrome b and the D-loop was used to reconstruct a maximum-likelihood phylogenetic tree using PHYML 3.0 (Guindon et al. 2010). An HKY model was used following the work of Igea et al. (2013). A haplotype genealogy was obtained based on the reconstructed tree using HAPLOVIEWER 1.0 (Salzburger et al. 2011).

Our analysis confirmed the existence of two main mitochondrial groups in the area, A and B, with a mostly allopatric distribution. Subsequent analyses were performed with clade A, which consisted of 157 samples, to avoid the confounding effects of the more divergent sequences from clade B.

GENETIC DIVERSITY

To explore the genetic diversity of the Pyrenean desman in the studied area, we used the alignment and the sample coordinates to compute, for each sampling point, the nucleotide diversity (π) values for all samples found in a radius of 10 km around the sampling point (Igea et al. 2013). We then used an inverse distance weight algorithm to interpolate the values of π in a grid of 1 km x 1 km of the distribution studied, using the *idw* function of the R package GSTAT (Pebesma & Graeler 2016). The surface obtained was superimposed on a map of the distribution studied. In addition, we obtained the unique haplotypes within the alignment using the *FindHaplo* function of the R package SIDIER (Muñoz-Pajares 2013) and also plotted them on the map.

ANALYSIS OF MOLECULAR VARIANCE

To study the extent to which river basins determine the genetic structure we conducted a hierarchical analysis of molecular variance (AMOVA) (Excoffier et al. 1992) using a matrix of pairwise genetic distances and the isolated river basins in which the samples were found as unique factor. This analysis was performed using the R package PEGAS (Paradis 2010).

The matrix of genetic distances between all samples was calculated as the sum of the branch lengths separating them in the maximum-likelihood tree (patristic or phylogenetic distances), which should account for all the divergence among sequences. We used the *cophenetic.phylo* function of the R package APE for this calculation (Paradis 2004).

GEOGRAPHICAL DISTANCES

Before computing the geographic distances, sample coordinates were moved to the nearest point in the river using the *snapPointsToLines* function of the R package MAPTOOLS (Bivand et al. 2016). Euclidean pairwise distances were calculated using the *pointDistance* function of the R package *Raster* (Hijmans et al. 2016), which takes into account the curvature of the Earth.

The altitude-based least-cost path distance was computed using the R package GDISTANCE (van Etten 2015). We first calculated the transition matrix from a *raster* of altitudes taken from WorldClim (http://www.worldclim.org) using the *transition* function and then applied a correction to take into account the curvature of the Earth with the *geoCorrection* function. Finally, the least-cost path distance was calculated using the *costDistance* function.

River distances were computed using the R package SECRLINEAR (Efford 2015). First, a shapefile of the river network from the distribution studied was obtained by merging the river shapefiles from Spain (Ministerio de Agricultura, Alimentación y Medio Ambiente) and Portugal (Atlas do Ambiente Digital - Instituto do Ambiente). This shapefile was converted to an R object of *linearmask* type using the *read.linearmask* function. The pairwise distance matrix was then obtained using the *netwotkdistance* function, which computes the shortest paths along the network. When two points belonged to different basins, and therefore unconnected river networks, the pair was treated as missing data in the correlation analyses.

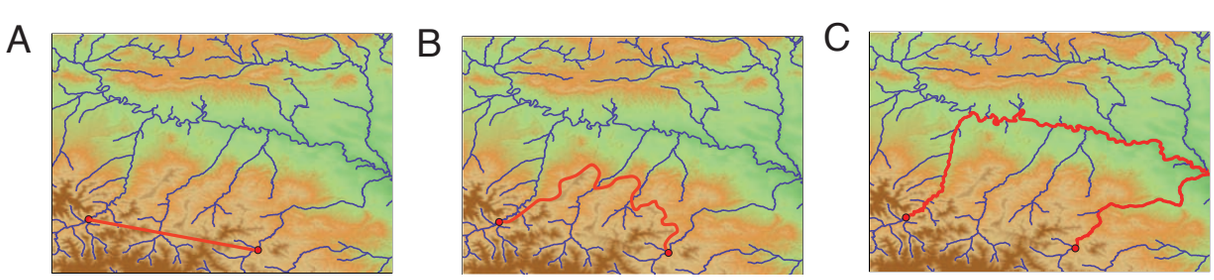


Figure 1. Scheme of the computation of Euclidean (A), altitude-based least-cost path (B) and river (C) distances between two sampling points.

Correlations between phylogenetic distance and the three types of geographic distances were calculated by means of a Mantel test using the *mantel.test* function of the R package NCF (Bjornstad 2016), for a total of 9999 permutations.

RESULTS

PHYLOGENETIC ANALYSIS

The haplotype genealogy based on the maximum likelihood tree (Figure 2A) showed a clear separation between clades A (with samples belonging to subgroup A1) and B (with samples from subgroups B1 and B2). The visualization of the main clades in the map (Figure 2B) confirmed a strict geographical separation between them despite the large number of samples used. Thus, clade A1 occupies the occidental part of the studied area and clade B1 the north-eastern part, with almost no mixing between them. Specimens from both clades were found in only two rivers (Curueño and Porma; Figure S2). A few samples from clade B2 are intermixed with B1 samples, as reported previously (Igea et al. 2013).

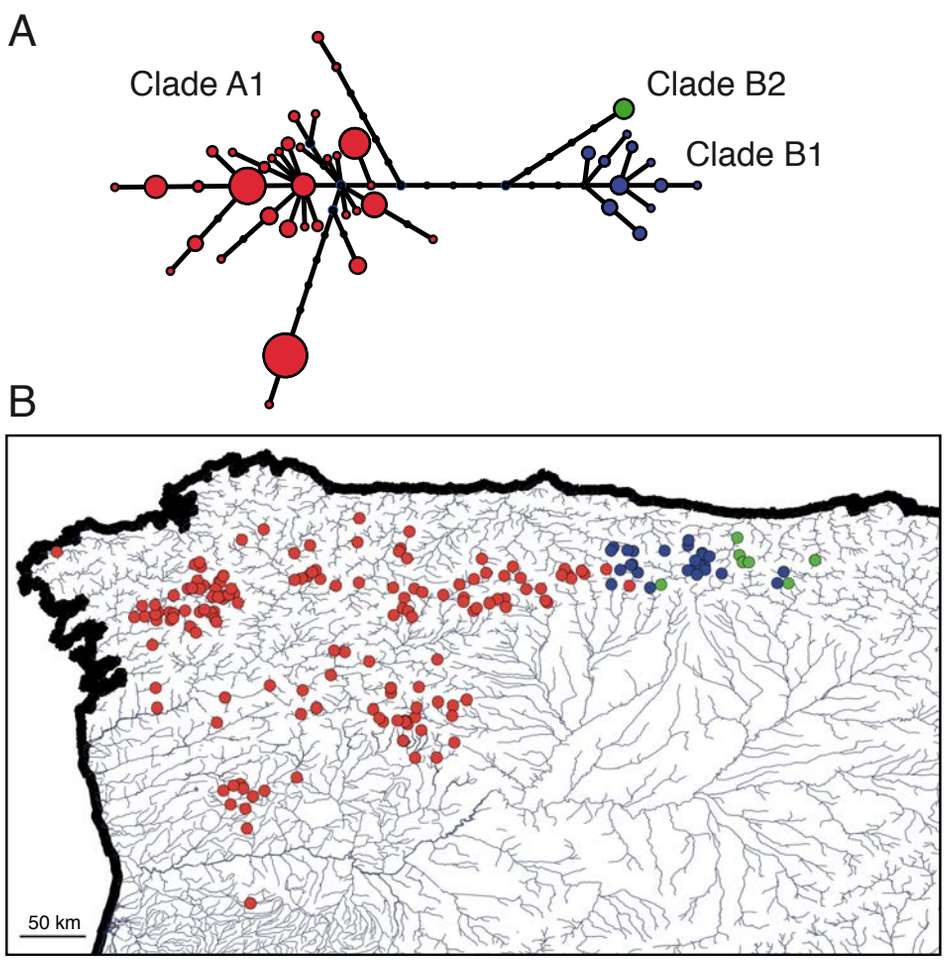


Figure 2. Haplotype genealogy of the mitochondrial sequences based on a maximum-likelihood tree in which circles represent haplotypes, size is proportional to the number of individuals, and black dots represent intermediate, unsampled haplotypes (A), and map of the samples coloured according to the mitochondrial clade (B)

GENETIC DIVERSITY

Genetic diversity was analysed using 157 samples from clade A1. Both the plot of interpolated nucleotide diversity values and the distribution of different haplotypes (Figure 3) revealed that genetic diversity is not homogeneous across the area, with spots of low and high diversity being found in the distribution studied. When averages per basin were determined, nucleotide diversity showed the highest value in the Duero basin (0.00547). The lowest average was found in the Ulla basin (0.00019), where a large number of identical haplotypes were present.

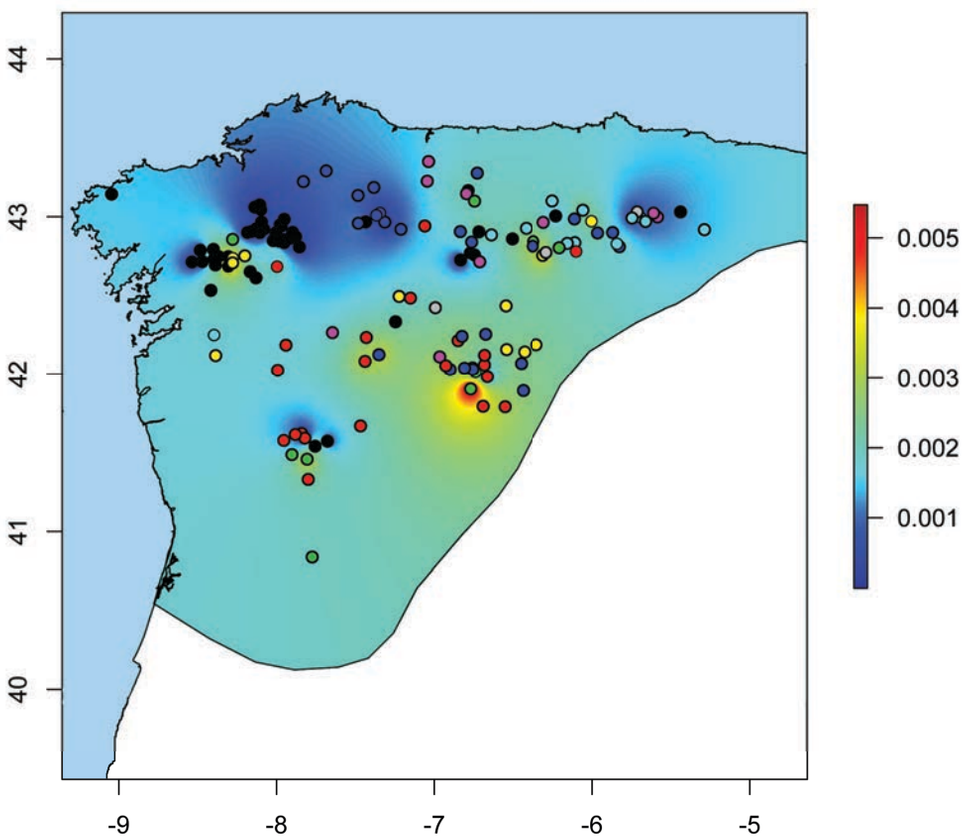


Figure 3. Plot of interpolated nucleotide diversity values, with the colour scale indicating nucleotide diversity. Different haplotypes are shown with different colours. A different random colour is used for each haplotype and therefore a greater variety of colours in an area also indicates greater genetic diversity.

EFFECT OF RIVER BASIN ON GENETIC STRUCTURE

The one-way AMOVA performed using samples from clade A1 and the 11 isolated river basins as a factor (Table S1 and Figure S1) revealed that 53.8 % of the genetic variation was due to the partition between river basins (Table S2; $p > 0.05$).

DISPERSAL PATTERNS

We then performed the Mantel tests for all samples from clade A1 and found a correlation of 0.32 between the phylogenetic and Euclidean distances, and 0.48 between the phylogenetic and

altitude-based least-cost path distances, with both having significant p-values ($p < 0.05$; Table 1). The correlation between phylogenetic and river distances was not significant.

We also performed the correlation analyses in the two basins with the largest number of samples. The correlation between the phylogenetic and Euclidean distances in the Duero basin (Figure S1), with 55 samples, was 0.11 and that between the phylogenetic and altitude-based least-cost path distances was 0.16, although only the latter was significant (Table 1). The correlation between the phylogenetic and river distances was not significant.

Distances compared	Basin	Correlation	P-value
Euclidean vs. Phylogenetic	All	0.32	0.0002
Altitude vs. Phylogenetic	All	0.48	0.0002
River vs. Phylogenetic	All	-0.06	0.0533
Euclidean vs. Phylogenetic	Duero	0.11	0.0568
Altitude vs. Phylogenetic	Duero	0.16	0.0290
River vs. Phylogenetic	Duero	0.02	0.2578
Euclidean vs. Phylogenetic	Miño	0.40	0.0002
Altitude vs. Phylogenetic	Miño	0.41	0.0002
River vs. Phylogenetic	Miño	0.38	0.0002

Table 1. Results obtained from the Mantel tests for all the distance matrices. Altitude refers to altitude-based least-cost path distance.

In contrast, the three Mantel tests for the Miño basin (Figure S1), with 44 samples, were all significant (Table 1), with a value of 0.40 for the correlation between phylogenetic and Euclidean distances, 0.41 when altitude-based least-cost path distances were considered, and 0.38 for river distances.

DISCUSSION

PHYLOGEOGRAPHIC STRUCTURE AND STRICT CONTACT ZONE IN THE CANTABRIAN MOUNTAINS

The Pyrenean desman shows a strong genetic structure, probably due to the past existence of several isolated glacial refugia in which different populations persisted during Pleistocene glaciations and accumulated genetic divergence (Igea et al. 2013; Querejeta et al. 2016). The existence of multiple glacial refugia within South European peninsulas, classically considered to be homogeneous refugia, has been proven for many other endemic and semi-endemic species (Alexandrino et al.

2000; Martinez-Solano et al. 2006; Gomez & Lunt 2007; Godinho et al. 2008; Centeno-Cuadros et al. 2009; Gonçalves et al. 2009; Razgour et al. 2015). This list includes species associated to aquatic ecosystems such as the golden-striped salamander *Chioglossa lusitanica* (Alexandrino et al. 2000) or the southern water vole *Arvicola sapidus* (Centeno-Cuadros et al. 2009). In this work, we used a large sample set of the Pyrenean desman from the occidental part of the Iberian Peninsula, which includes the two main mitochondrial clades of this species and the contact zone between them, to determine whether a larger dataset could help to reveal some spatial intermixing of clades and to study how dispersal across land and rivers may have shaped the genetic structure of this semiaquatic species. Specifically, the mitochondrial clades A1 and B1 likely entered into contact in the Cantabrian Mountains after the postglacial recolonization from their respective refugia (Igea et al. 2013). The detailed genetic characterization of these populations carried out in this work allowed us to detect the contact zone between these clades and revealed that there is indeed a small degree of spatial mixing between them (Figure S2). Specimens belonging to the two clades were only found in two rivers, namely the Porma and its tributary the Curueño. The fact that no other specimen of clade A1 can be found in the area of clade B1, and vice versa, supports a strong phylogeographic structure for this species (Igea et al. 2013). The distribution pattern of both mitochondrial clades is intriguing because there is no apparent physical barrier to dispersal in this contact zone. Future in-depth analyses in this contact zone, including genomic analysis, may reveal crucial aspects concerning contemporary dispersal and gene flow, thereby helping to better delimit conservation units (Crandall et al. 2000).

VARIABILITY OF THE GENETIC DIVERSITY AND GLACIAL HISTORY

The spatial analysis of the genetic diversity of clade A1 showed several areas of greater genetic diversity (Figure 3). In particular, the area with highest values (> 0.004) is located in the centre of the studied area. The species-distribution modelling performed in a previous work revealed that the areas of maximum probability of potential distribution during the Last Glacial Maximum were largely coincident with areas of maximum genetic diversity (Igea et al. 2013). However, according to such study the optimal glacial refugium in the nor-occidental area was closer to the coast than the high genetic spot observed. Mixing of haplotypes from different areas might explain areas with high mitochondrial diversity displaced from original refugia. Additionally, the study of these details about the glacial history of the Pyrenean desman will require an exhaustive sampling in this area and, particularly, the use of genomic data to summarize genetic diversity from more loci.

Spots of low genetic diversity include the contact zone of clades A1 and B1 in the Cantabrian Mountains and another area in the most north-western part (Ulla basin), which in principle may correspond to areas recently colonized via a bottleneck (Taberlet et al. 1998; Hewitt 2000; Gomez & Lunt 2007). This recent colonization is likely to have happened in the contact zone between the mitochondrial clades, which may have been subjected to harsh conditions during the Last Glacial

Maximum (Igea et al. 2013). However, the Ulla basin is close to areas of potential distribution during the Last Glacial Maximum (Igea et al. 2013). Therefore, a more complex hypothesis that includes the existence of a glacial refugium, local extinctions and subsequent recolonization via a bottleneck, cannot be ruled out to explain the current low genetic diversity in the Ulla basin. These complex scenarios, also suggested for other species of the Iberian Peninsula, would give further support the intricate evolution of populations associated to the glacial history of South European peninsulas (Gomez & Lunt 2007; Godinho et al. 2008).

RELATIVE IMPORTANCE OF DISPERSAL MODES IN SEMIAQUATIC SPECIES

The AMOVA revealed that only 53.8 % of the genetic variability was due to the isolated river basins, thus indicating that, although these basins may have played an important role in structuring the genetic diversity, the desmans are not completely isolated within them.

The Mantel tests revealed a high correlation between overland distances (both Euclidean and least-cost path) and phylogenetic distances for the whole distribution studied, thus indicating a strong effect of isolation by distance in the Pyrenean desman. The main implication of this result is that dispersal is low in this species, creating an isolation-by-distance pattern. This conclusion is in agreement with data known from radiotracking studies, which showed that desmans have reduced vital areas, usually including only hundreds of meters in a stretch of river, and limited dispersal capability (Stone 1987; Melero et al. 2012; 2014). Moreover, long-range movement are exceptional in this species, as shown by genotyping of faeces (Gillet et al. 2016).

The correlation found with overland distances but not river distances indicates that the isolation-by-distance pattern can be better modelled, in principle, through terrestrial dispersal in the Pyrenean desman. In addition, the lack of correlation between river and phylogenetic distances suggests that the effects of dispersal along the river network were eclipsed by overland dispersal. These results are in apparent contradiction with the known dependency of the Pyrenean desman on the aquatic medium but they may be explained by historical factors. One possibility is that desmans colonized new basins from the glacial refugia using terrestrial corridors (Igea et al. 2013). This mode of overland dispersal may have been facilitated during the Holocene deglaciation, in which large glaciers were subdivided into smaller ones (Serrano et al. 2016), likely giving rise to a greater abundance of streams and humid habitats. In other words, it is possible that the river network was larger than it is today, facilitating connectivity between basins. This way, overland dispersal of the Pyrenean desman would have been reduced to short stretches between basins. Therefore, the Holocene may have been highly favourable for the dispersal and colonization of large areas by a semiaquatic species like the Pyrenean desman. Additional connections between some specific basins during glacial periods invoked to explain the distribution of some Iberian fishes (Aboim et al. 2013) cannot be totally discarded although their effects on the overall dispersal of the Pyrenean desman may not have been as important as the existence of a larger fluvial network.

Similar results were found for the Duero basin, which is the largest one in the area studied. A significant correlation was found between phylogenetic distances and altitude-based least-cost path distances but not river distances. The correlations were low in this basin, implying a relatively marginal isolation-by-distance pattern at this level. These weak correlations may be caused by the particular distribution of samples in this basin, which only covers its north-western edge and does not include most of the basin area (Figure S1). In any case, these results are again consistent with the colonization of this large basin via an important component of overland dispersal. The large and continuous forest areas found near the headwaters of the mountain rivers in this basin are likely to have contributed to an increased connectivity between rivers.

In contrast, a significant correlation between phylogenetic distances and both overland and river distances was found for the Miño basin, as expected for a semiaquatic mammal. The small mountains in this basin and the presence of corridors at low altitudes may have favoured overland interconnections between rivers, thus explaining the correlation between phylogenetic and overland distances. However, the presence of a significant correlation between phylogenetic and river distances indicates a prominent effect of river dispersal. The smaller size of this basin makes it likely that these correlations reflect overall more recent dispersal rather than postglacial colonization, as compared to larger basins and the whole area.

To explain the different dispersal modes found for the Pyrenean desman in the Duero and Miño basins, important aspects other than differences in basin size should be taken into account. Within the Duero basin, the Pyrenean desman is present mainly in the headwaters of rivers, where it can find these suitable habitats, whereas large tributaries and rivers present in the central part of this basin are not adequate for the species, as indicated by its absence in this area (Figure S1). This central area in the Duero basin interrupts dispersal and breaks the correlation between phylogenetic and river distances and, probably, also contributes to the low correlation observed between phylogenetic and terrestrial distances. The Miño basin, however, has suitable riverine habitats for the desman in the majority of the basin, without any important area in which the species is absent, thus facilitating interconnection of desmans through a larger part of the river network and explaining the positive correlation between phylogenetic and river distances.

It is also important to consider that rivers inhabited by the Pyrenean desman are interrupted by artificial barriers that include hydroelectric plants and dams, but also by ecological barriers caused by water pollution and desiccation due to water extractions (Fernandes et al. 2011). This is creating fragmented populations with reduced connectivity, which, in turn, may have affected the calculation of the correlations between phylogenetic and river distances; this is a problem that we could not circumvent in this work. Specific studies aimed to investigate these water infrastructures will be necessary to understand how they affect the dispersal of the Pyrenean desman.

Taken together, these results demonstrate a strong isolation-by-distance effect in the Pyrenean desman that can in principle be modelled through overland dispersal when large areas area considered, and through both river and overland dispersal when smaller basins with highly appropriate habitats are analysed; when a large proportion of the analysed area is unsuitable for the species, the isolation-by-distance effect may be smaller. These results also suggest that the postglacial colonization of new areas was quite successful during the deglaciation phase, in which favourable habitats for the desman may have been highly abundant, thus leading to the rapid colonization of large areas that constitute the current species range. This dispersal may have occurred through overland corridors between basins or through a more extensive fluvial network that may have been present during the Holocene. It is likely that contemporary inter-river dispersal became more reduced after this period. This restricted dispersal would explain the lack of current spatial mixing of mitochondrial clades in their contact zone despite the absence of any clear geographical barrier (Figure S2).

The methodology used here to study these dispersal modes, and the results obtained, open up a promising avenue of research regarding the study of evolutionary and ecological factors that affect dispersal in different semiaquatic species. Thus, these methods can complement previous studies on semiaquatic species that also disperse along rivers or land depending on different ecological factors, such as the Eurasian otter (Pagacz 2016), the Eurasian beaver (Senn et al. 2014) or the platypus (Furlan et al. 2013), to name a few, and determine to what extent they depend on river and overland dispersal modes. As we have shown here, it is important to note that the results of this isolation-by-distance method may be dependent on the size and ecological features of the area analysed and, consequently, different scales may reflect different aspects of the evolutionary history and ecology of semiaquatic species.

It should also be taken into account that the divergence levels in mitochondrial DNA analysed herein arose during ancient periods, and therefore the dispersal patterns inferred with these data apply mainly to them (Avise 2009). The study of contemporary dispersal patterns would require an analysis of a higher number of nuclear loci to reveal fine-scale genetic structure. The computation of genetic distances derived from the nuclear genome may also help to avoid problems with dispersal differences between sexes because mitochondrial DNA is only inherited by females and would not allow different patterns of male dispersal to be determined (Waits et al. 2000; Nater et al. 2011). However, both mitochondrial and genomic studies provide complementary data and are necessary to understand both the evolutionary history of the species and its dispersal behaviour.

CONSERVATION IMPLICATIONS

Studies of dispersal patterns in semiaquatic species are especially relevant for endangered species like the Pyrenean desman. Without doubt, preserving and restoring the aquatic habitat is the most

important aspect for the conservation of this species. However, our results suggest that certain degree of overland dispersal was important in the recent evolutionary past of the species and therefore reinforce the idea that riparian corridors between isolated river basins should also be protected because they have a great potential for the dispersal of the species. Specifically, an adequate plan for the conservation of a semiaquatic species like the Pyrenean desman should consider the delimitation and protection of potential riparian corridors in addition to the conservation of rivers. In particular, riverine habitats of the upper parts of rivers, especially those situated at low altitudes and with tributaries close to the watershed divides, may facilitate genetic interchange between populations, and therefore deserve special protection. Forest areas in watershed divides may also favour inter-river dispersal. Ultimately, an effective protection of the whole riverine ecosystem that includes the upper parts of rivers and potential corridors between them may be essential for the long-term survival of the Pyrenean desman populations.

REFERENCES

Aboim MA, Mesquita N, Drago M, Coelho MM, Alves MJ (2013) Assessing inter-drainage connections: patterns of genetic diversity in an Iberian cyprinid fish. *Biological Journal of the Linnean Society*, 109, 656–669.

Adriaensen F, Chardon JP, De Blust G, Swinnen E, Villalba S, Gulinck H, Matthysen E (2003) The application of “least-cost” modelling as a functional landscape model. *Landscape and Urban Planning*, 64, 233–247.

Alexandrino J, Froufe E, Arntzen JW, Ferrand N (2000) Genetic subdivision, glacial refugia and postglacial recolonization in the golden-striped salamander, *Chioglossa lusitanica* (Amphibia: urodela). *Molecular Ecology*, 9, 771–781.

Avise JC (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography*, 36, 3–15.

Bartáková V, Reichard M, Blažek R, Polačik M, Bryja J (2015) Terrestrial fishes: rivers are barriers to gene flow in annual fishes from the African savanna. *Journal of Biogeography*, 42, 1832–1844.

Bivand R, Lewin-Koh N, Pebesma E *et al*. (2016) *maptools*: Tools for Reading and Handling Spatial Objects. R package available at CRAN.

Bjornstad ON (2016) *ncf*: Spatial Nonparametric Covariance Functions. R package available at CRAN.

Byrne MS, Quintana RD, Bolkovic ML, Cassini MH, Túnez JI (2015) The role of river drainages in shaping the genetic structure of capybara populations. *Genetica*, 143, 645–656.

Callejo A, Delibes M (1987) Dieta de la nutria *Lutra lutra* (Linnaeus, 1758) en la cuenca del Alto Ebro, norte de España. *Misc. Zool.*, 11, 353–362.

Centeno-Cuadros A, Delibes M, Godoy JA (2009) Phylogeography of Southern Water Vole (*Arvicola sapidus*): evidence for refugia within the Iberian glacial refugium? *Molecular Ecology*, 18, 3652–3667.

Chaput-Bardy A, Lemaire C, Picard D, Secondi J (2008) In-stream and overland dispersal across a river network influences gene flow in a freshwater insect, *Calopteryx splendens. Molecular Ecology*, 17, 3496–3505.

Charbonnel A, Buisson L, Biffi M, DAmico F, Besnard A, Aulagnier S, Blanc F, Gillet F, Lacaze V, Michaux JR, Némoz M, Pagé C, Sanchez-Perez JM, Sauvage S, Laffaille P (2015) Integrating hydrological features and genetically validated occurrence data in occupancy modelling of an endemic and endangered semi-aquatic mammal, *Galemys pyrenaicus*, in a Pyrenean catchment. *Biological Conservation*, 184, 182–192.

Coulon A, Cosson J-F, Angibault JM, Cargnelutti B, Galan M, Morellet N, Petit E, Aulagnier S, Hewison AJM (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. *Molecular Ecology*, 13, 2841–2850.

Crandall KA, Bininda-Emonds ORP, Mace G, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, 15, 290–295.

Efford M (2015) *secrlinear*: Spatially Explicit Capture-Recapture for Linear Habitats. R package available at CRAN.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479–491.

Fernandes M, Herrero J, Aulagnier S, Amori G (2011) *Galemys pyrenaicus*. IUCN Red List of *Threatened Species. Version 2011*. 2, 1–3.

Furlan EM, Griffiths J, Gust N, Handasyde KA, Grant TR, Gruber B, Weeks AR (2013) Dispersal patterns and population structuring among platypuses, *Ornithorhynchus anatinus*, throughout south-eastern Australia. *Conservation Genetics*, 14, 837–853.

Gillet F, Le Roux B, Blanc F, Bodo A, Fournier-Chambrillon C, Fournier P, Jacob F, Lacaze V, Nemoz M, Aulagnier S, Michaux JR (2016) Genetic monitoring of the endangered Pyrenean desman (*Galemys pyrenaicus*) in the Aude River, France. *Belgian Journal of Zoology*, 146, 44–52.

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

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

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

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

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

Hijmans RJ, Cheng J, Mattiuzzi M, Sumner M, Greenberg JA, Racine EB, Shortridge A (2016) *raster*: Geographic Data Analysis and Modeling. R package available at CRAN.

Igea J, Aymerich P, Fernández-González A, González-Esteban J, Gómez A, Alonso R, Gosálbez J, Castresana J (2013) Phylogeography and postglacial expansion of the endangered semi-aquatic mammal *Galemys pyrenaicus*. BMC *Evolutionary Biology*, 13, 115.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, 18, 189–197.

Martinez-Solano I, Teixeira J, Buckley D, García-París M (2006) Mitochondrial DNA phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple refugia in an Iberian endemic. *Molecular Ecology*, 15, 3375–3388.

Melero Y, Aymerich P, Luque-Larena JJ, Gosálbez J (2012) New insights into social and space use behaviour of the endangered Pyrenean desman *(Galemys pyrenaicus). European Journal of Wildlife Research*, 58, 185–193.

Melero Y, Aymerich P, Santulli G, Gosálbez J (2014) Activity and space patterns of Pyrenean desman (*Galemys pyrenaicus*) suggest non-aggressive and non-territorial behaviour. *European Journal of Wildlife Research*, 60, 707–715.

Muñoz-Pajares AJ (2013) SIDIER: substitution and indel distances to infer evolutionary relationships. *Methods in Ecology and Evolution*, 4, 1195–1200.

Nater A, Nietlisbach P, Arora N, van Schaik CP, van Noordwijk MA, Willems EP, Singleton I, Wich SA, Goossens B, Warren KS, Verschoor EJ, Perwitasari-Farajallah D, Pamungkas J, Krützen M (2011) Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant Orangutans (genus: *Pongo*). *Molecular Biology and Evolution*, 28, 2275–2288.

Pagacz S (2016) The effect of a major drainage divide on the gene flow of a semiaquatic carnivore, the Eurasian otter. *Journal of Mammalogy*, 97, 1164–1176.

Palmeirim JM, Hoffmann RS (1983) *Galemys pyrenaicus. Mammalian Species*, 207, 1–5.

Paradis E (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20, 289–290.

Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26, 419–420.

Paz-Vinas I, Loot G, Stevens VM, Blanchet S (2015) Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology*, 24, 4586–4604.

Pebesma E, Graeler B (2016) *gstat*: Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation. R package available at CRAN.

Querejeta M, González-Esteban J, Gómez A, Fernández-González A, Aymerich P, Gosálbez J, Escoda L, Igea J, Castresana J (2016) Genomic diversity and geographical structure of the Pyrenean desman. *Conservation Genetics*, 17, 1333–1344.

Raeymaekers JAM, Maes GE, Geldof S, Hontis I, Nackaerts K, Volckaert FAM (2008) Modeling genetic connectivity in sticklebacks as a guideline for river restoration. *Evolutionary Applications*, 1, 475–488.

Razgour O, Salicini I, Ibáñez C, Randi E, Juste J (2015) Unravelling the evolutionary history and future prospects of endemic species restricted to former glacial refugia. *Molecular Ecology*, 24, 5267–5283.

Salzburger W, Ewing GB, Haeseler von A (2011) The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. *Molecular Ecology*, 20, 1952–1963.

Senn H, Ogden R, Frosch C, Syrůčková A, Campbell-Palmer R, Munclinger P, Durka W, Kraus RHS, Saveljev AP, Nowak C, Stubbe A, Stubbe M, Michaux JR, Lavrov V, Samiya R, Ulevicius A, Rosell F (2014) Nuclear and mitochondrial genetic structure in the Eurasian beaver (*Castor fiber*) - implications for future reintroductions. *Evolutionary Applications*, 7, 645–662.

Serrano E, González Trueba JJ, Pellitero R, Gómez-Lende M (2016) Quaternary glacial history of the Cantabrian Mountains of northern Spain: a new synthesis. *Geological Society, London, Special Publications.*

Stone R (1987) The activity patterns of the Pyrenean desman (*Galemys pyrenaicus*) (Insectivora: Talpidae), as determined under natural conditions. *Journal of Zoology*, 213, 95–106.

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

van Etten J (2015) *gdistance*: distances and routes on geographical grids. R package available at CRAN.

Vignieri SN (2005) Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse *(Zapus trinotatus). Molecular Ecology*, 14, 1925–1937.

Waits LP, Taberlet P, Swenson JE, Sandegren F, Franzén R (2000) Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear *(Ursus arctos). Molecular Ecology*, 9, 421–431.

Wright S (1943) Isolation by Distance. *Genetics*, 28, 114–138.

ACKNOWLEDGEMENTS

We thank the following institutions for permits for collecting samples or sending samples for our work: Junta de Castilla y León, Fundación Patrimonio Natural de Castilla y León, Xunta de Galicia, Principado de Asturias, and ICNF (Instituto da Conservação da Natureza e das Florestas). We thank Rosario Pérez Redondo, of the Instituto de Biotecnología de León (INBIOTEC), for providing DNA extracts of samples they analysed for the DESMANIA LIFE+ project. We also thank Joan Garcia-Porta for help with analysis. This work was financially supported by research project CGL2014-53968-P of the "Plan Nacional I+D+I del Ministerio de Economía y Competitividad" (Spain) to J.C. M.Q. was supported by fellowship BES-2012-057135 of the Ministerio de Economía y Competitividad (Spain).

AUTHOR CONTRIBUTIONS

J.C. and M.Q. conceived the ideas; A.F.-G. and R.R. collected the samples; M.Q. performed the laboratory work and analysed the data; and M.Q. and J.C. led the writing with input from the other authors.

SUPPORTING INFORMATION

ECOLOGY AND EVOLUTION - POSTGLACIAL DISPERSAL PATTERNS AND MITOCHONDRIAL GENETIC STRUCTURE OF THE PYRENEAN DESMAN (GALEMYS PYRENAICUS) IN THE NORTH-WESTERN REGION OF THE IBERIAN PENINSULA

Marina Querejeta, Angel Fernández-González, Rafael Romero and Jose Castresana

Specimen code	Latitude	Longitude	Basin	Sample type
IBE-BC0064	42.69	-8.31	Ulla	Otter faeces
IBE-BC0093	42.01	-6.74	Duero	Desman faeces
IBE-BC0123	42.48	-7.15	Miño	Otter faeces
IBE-BC0129	42.01	-6.74	Duero	Desman faeces
IBE-BC0204	42.26	-6.68	Miño	Desman faeces
IBE-BC0256	43.00	-5.69	Duero	Desman faeces
IBE-BC0259	42.92	-5.29	Duero	Desman faeces
IBE-BC0269	42.22	-6.85	Miño	Desman faeces
IBE-BC0311	42.06	-6.69	Duero	Desman faeces
IBE-BC0370	42.04	-6.76	Duero	Desman faeces
IBE-BC0377	43.02	-5.61	Duero	Desman faeces
IBE-BC0384	42.11	-6.97	Miño	Desman faeces
IBE-BC0388	42.03	-6.90	Duero	Desman faeces
IBE-BC0389	42.06	-6.93	Duero	Desman faeces
IBE-BC0401	42.03	-6.76	Duero	Desman faeces
IBE-BC0472	42.90	-5.87	Duero	Desman faeces
IBE-BC0476	41.90	-6.44	Duero	Desman faeces
IBE-BC0483	42.43	-6.55	Miño	Desman faeces
IBE-BC0497	42.84	-6.11	Duero	Desman faeces
IBE-BC0505	42.12	-6.68	Duero	Desman faeces
IBE-BC0508	42.14	-6.43	Duero	Desman faeces
IBE-BC0541	41.99	-6.67	Duero	Desman faeces
IBE-BC0542	42.16	-6.54	Duero	Desman faeces
IBE-BC0567	42.19	-6.35	Duero	Desman faeces
IBE-BC0586	42.93	-6.42	Miño	Desman faeces
IBE-BC0605	42.07	-6.45	Duero	Desman faeces
IBE-BC0615	43.03	-5.71	Duero	Desman faeces
IBE-BC0618	43.00	-5.58	Duero	Desman faeces
IBE-BC0623	42.90	-5.96	Duero	Desman faeces
IBE-BC0648	43.05	-4.85	Duero	Desman faeces

Table S1. Samples used in this study, including those also used in a previous study.

IBE-BC0652	43.05	-4.77	Duero	Desman faeces
IBE-BC0653	43.03	-5.44	Duero	Desman faeces
IBE-BC0692	43.16	-5.04	Sella	Desman faeces
IBE-BC0702	42.97	-5.66	Duero	Desman faeces
IBE-BC0718	43.02	-5.61	Duero	Desman faeces
IBE-BC0723	43.00	-6.23	Miño	Desman faeces
IBE-BC0725	42.99	-5.74	Duero	Desman faeces
IBE-BC0734	43.05	-4.82	Duero	Desman faeces
IBE-BC0780	43.03	-4.77	Duero	Desman faeces
IBE-BC0800	43.02	-5.35	Duero	Desman faeces
IBE-BC0803	42.83	-6.15	Duero	Desman faeces
IBE-BC0807	42.78	-6.10	Duero	Desman faeces
IBE-BC0809	42.80	-6.21	Duero	Desman faeces
IBE-BC0816	42.81	-5.82	Duero	Desman faeces
IBE-BC0830	42.84	-6.37	Miño	Desman faeces
IBE-BC0839	42.91	-6.84	Navia	Desman faeces
IBE-BC0843	42.84	-6.77	Miño	Desman faeces
IBE-BC0854	42.94	-4.24	Duero	Desman faeces
IBE-BC0858	42.90	-6.72	Miño	Desman faeces
IBE-BC0859	42.77	-6.78	Miño	Desman faeces
IBE-BC0869	42.89	-6.64	Miño	Desman faeces
IBE-BC0871	42.73	-6.84	Miño	Desman faeces
IBE-BC0887	43.00	-4.69	Duero	Desman faeces
IBE-BC0891	42.96	-6.31	Miño	Desman faeces
IBE-BC0893	42.83	-5.84	Duero	Desman faeces
IBE-BC0900	42.76	-6.31	Miño	Desman faeces
IBE-BC0945	42.92	-8.13	Ulla	Desman faeces
IBE-BC0948	43.04	-6.06	Duero	Desman faeces
IBE-BC0951	42.94	-4.32	Duero	Desman faeces
IBE-BC0956	42.99	-6.11	Duero	Desman faeces
IBE-BC0959	42.76	-6.75	Miño	Desman faeces
IBE-BC0961	42.77	-6.29	Miño	Desman faeces
IBE-BC0980	42.71	-6.71	Miño	Desman faeces

IBE-BC1008	42.86	-6.51	Miño	Desman faeces
IBE-BC1055	42.91	-5.15	Duero	Desman faeces
IBE-BC1077	42.93	-5.08	Duero	Desman faeces
IBE-C1154	43.17	-6.79	Navia	Igea et al. 2013
IBE-C1174	43.15	-6.80	Navia	Igea et al. 2013
IBE-C1177	43.15	-6.80	Navia	Igea et al. 2013
IBE-C1729	43.01	-4.27	Ebro	Igea et al. 2013
IBE-C1843	43.10	-6.75	Nalon	Igea et al. 2013
IBE-C1851	43.14	-4.81	Deva	Igea et al. 2013
IBE-C1869	43.28	-6.73	Navia	Igea et al. 2013
IBE-C2517	43.15	-5.41	Nalón	Igea et al. 2013
IBE-C2607	43.15	-5.26	Nalón	Igea et al. 2013
IBE-C2736	42.25	-8.40	Miño	Igea et al. 2013
IBE-C2737	42.75	-8.37	Ulla	Igea et al. 2013
IBE-C2738	42.75	-8.37	Ulla	Igea et al. 2013
IBE-C2739	42.08	-7.44	Duero	Igea et al. 2013
IBE-C2740	42.24	-6.83	Miño	Igea et al. 2013
IBE-C2741	42.24	-6.83	Miño	Igea et al. 2013
IBE-C2742	42.19	-7.94	Miño	Igea et al. 2013
IBE-C2743	42.19	-7.94	Miño	Igea et al. 2013
IBE-C2745	43.23	-7.05	Navia	Igea et al. 2013
IBE-C2747	42.24	-7.43	Miño	Igea et al. 2013
IBE-C2750	42.12	-8.39	Miño	Igea et al. 2013
IBE-C2751	42.13	-7.35	Miño	Igea et al. 2013
IBE-C2752	42.13	-7.35	Miño	Igea et al. 2013
IBE-C2753	42.94	-7.06	Navia	Igea et al. 2013
IBE-C2754	42.42	-7.00	Miño	Igea et al. 2013
IBE-C2856	43.12	-4.56	Deva	Igea et al. 2013
IBE-C2893	43.18	-4.90	Deva	Igea et al. 2013
IBE-C2894	43.22	-4.90	Deva	Igea et al. 2013
IBE-C2950	43.17	-5.39	Nalón	Igea et al. 2013
IBE-C2975	41.57	-7.68	Duero	Igea et al. 2013
IBE-C2990	41.62	-7.84	Duero	Igea et al. 2013

IBE-C3301	41.79	-6.55	Duero	Igea et al. 2013
IBE-C3303	40.84	-7.78	Duero	Igea et al. 2013
IBE-C3305	40.84	-7.78	Duero	Igea et al. 2013
IBE-C3338	43.08	-4.54	Deva	Igea et al. 2013
IBE-C3342	43.07	-4.50	Deva	Igea et al. 2013
IBE-C3522	41.91	-6.77	Duero	Igea et al. 2013
IBE-C3541	41.80	-6.69	Duero	Igea et al. 2013
IBE-C3542	42.92	-8.08	Ulla	Desman faeces
IBE-C3545	41.91	-6.77	Duero	Igea et al. 2013
IBE-C3557	43.35	-7.04	Eo	Desman faeces
IBE-C3568	42.04	-6.81	Duero	Desman faeces
IBE-C3602	41.33	-7.80	Duero	Igea et al. 2013
IBE-C3620	42.93	-5.40	Duero	Igea et al. 2013
IBE-C3623	43.15	-5.41	Nalón	Igea et al. 2013
IBE-C3638	41.91	-6.77	Duero	Igea et al. 2013
IBE-C3641	43.15	-5.30	Nalón	Igea et al. 2013
IBE-C3647	41.91	-6.77	Duero	Igea et al. 2013
IBE-C3795	42.69	-8.39	Ulla	Desman faeces
IBE-C3812	42.72	-8.47	Ulla	Desman faeces
IBE-C3813	42.87	-7.88	Ulla	Desman faeces
IBE-C3822	42.73	-8.40	Ulla	Desman faeces
IBE-C3828	42.75	-8.01	Ulla	Desman faeces
IBE-C3833	42.97	-8.10	Ulla	Desman faeces
IBE-C3834	42.90	-8.11	Ulla	Desman faeces
IBE-C3843	42.86	-8.28	Ulla	Desman faeces
IBE-C3848	42.79	-8.40	Ulla	Desman faeces
IBE-C3849	42.75	-8.20	Ulla	Desman faeces
IBE-C3852	42.74	-8.03	Ulla	Desman faeces
IBE-C3868	42.75	-8.11	Ulla	Desman faeces
IBE-C438	42.97	-6.00	Duero	Igea et al. 2013
IBE-C4402	43.22	-7.83	Miño	Otter faeces
IBE-C4403	42.53	-8.42	Lerez	Otter faeces
IBE-C4404	42.33	-7.25	Miño	Otter faeces

IBE-C4405	42.92	-7.21	Miño	Otter faeces
IBE-C4407	42.03	-7.99	Limia	Otter faeces
IBE-C4409	42.27	-7.65	Miño	Otter faeces
IBE-C4410	42.83	-7.95	Ulla	Otter faeces
IBE-C4411	43.14	-9.05	Grande	Otter faeces
IBE-C4412	42.50	-7.22	Miño	Otter faeces
IBE-C4413	43.04	-8.10	Tambre	Otter faeces
IBE-C4414	43.06	-8.14	Tambre	Otter faeces
IBE-C4415	43.07	-8.11	Tambre	Otter faeces
IBE-C4417	42.85	-8.00	Ulla	Otter faeces
IBE-C4418	43.13	-7.49	Miño	Otter faeces
IBE-C4420	43.19	-7.39	Miño	Otter faeces
IBE-C4422	42.99	-7.95	Ulla	Otter faeces
IBE-C4423	42.71	-8.54	Ulla	Otter faeces
IBE-C4424	42.90	-8.19	Ulla	Otter faeces
IBE-C4428	42.75	-8.31	Ulla	Otter faeces
IBE-C4429	42.98	-7.40	Miño	Otter faeces
IBE-C4431	42.97	-7.32	Miño	Otter faeces
IBE-C4433	42.97	-7.44	Miño	Otter faeces
IBE-C4435	43.29	-7.69	Miño	Otter faeces
IBE-C4437	43.02	-7.35	Miño	Otter faeces
IBE-C4438	43.01	-7.37	Miño	Otter faeces
IBE-C4440	42.96	-7.48	Miño	Otter faeces
IBE-C4441	42.81	-7.86	Ulla	Otter faeces
IBE-C4442	42.86	-7.96	Ulla	Otter faeces
IBE-C4449	42.95	-7.98	Ulla	Otter faeces
IBE-C4451	42.76	-8.08	Ulla	Otter faeces
IBE-C4452	42.70	-8.01	Ulla	Otter faeces
IBE-C4453	42.68	-8.00	Ulla	Otter faeces
IBE-C4454	42.74	-8.28	Ulla	Otter faeces
IBE-C4459	42.65	-8.17	Ulla	Otter faeces
IBE-C4460	42.61	-8.13	Ulla	Otter faeces
IBE-C4461	42.94	-7.97	Ulla	Otter faeces

IBE-C4462	42.90	-8.01	Ulla	Otter faeces
IBE-C4465	42.98	-7.95	Ulla	Otter faeces
IBE-C4466	42.91	-8.16	Ulla	Otter faeces
IBE-C4470	42.79	-8.48	Ulla	Otter faeces
IBE-C4473	42.71	-8.28	Ulla	Otter faeces
IBE-C4474	42.85	-8.02	Ulla	Otter faeces
IBE-C4476	42.85	-7.99	Ulla	Otter faeces
IBE-C4477	42.86	-7.92	Ulla	Otter faeces
IBE-C4478	42.90	-7.90	Ulla	Otter faeces
IBE-C448	42.81	-6.38	Miño	Igea et al. 2013
IBE-C449	43.10	-6.25	Nalon	Igea et al. 2013
IBE-C539	43.24	-4.58	Deva	Igea et al. 2013
IBE-C618	43.09	-4.07	Duero	Igea et al. 2013
IBE-C683	43.09	-4.07	Saja	Igea et al. 2013
IBE-C720	43.13	-4.81	Deva	Igea et al. 2013
IBE-C723	43.11	-4.76	Deva	Igea et al. 2013
IBE-C779	41.58	-7.96	Duero	Igea et al. 2013
IBE-C802	41.59	-7.82	Duero	Igea et al. 2013
IBE-C831	41.62	-7.88	Duero	Igea et al. 2013
IBE-C844	41.67	-7.47	Duero	Igea et al. 2013
IBE-C894	41.54	-7.76	Duero	Igea et al. 2013
IBE-C895	41.46	-7.81	Duero	Igea et al. 2013
IBE-C919	41.54	-7.76	Duero	Igea et al. 2013
IBE-C922	41.49	-7.90	Duero	Igea et al. 2013
IBE-C975	43.09	-4.86	Duero	Igea et al. 2013
IBE-LE0402	43.01	-5.23	Duero	Desman faeces
IBE-LE0406	43.07	-5.27	Duero	Desman faeces
IBE-LE0411	43.06	-5.29	Duero	Desman faeces
IBE-LE0529	43.02	-4.90	Duero	Desman faeces
IBE-LE0536	42.99	-4.84	Duero	Desman faeces

Igea, J., P. Aymerich, A. Fernández-González, J. González-Esteban, A. Gómez, R. Alonso, J. Gosálbez, and J. Castresana. 2013. Phylogeography and postglacial expansion of the endangered semi-aquatic mammal *Galemys pyrenaicus*. BMC Evol. Biol. 13:115.

Source of variation	d.f.	SS	MS	Est. Variance	%	P-value
Among basins	9	0.0012429	0.0001381	0.0000102	53.8	***
Within basins	147	0.0012894	0.0000087	0.0000087	46.2	***
Total	156	0.0025324	0.0000162	0.0000189		

Table S2. Analysis of Molecular Variance (AMOVA) for the phylogenetic distances between desmans inhabiting isolated river basins and within desmans inhabiting isolated river basins. % is the percentage of variance explained by the isolation of river basins.

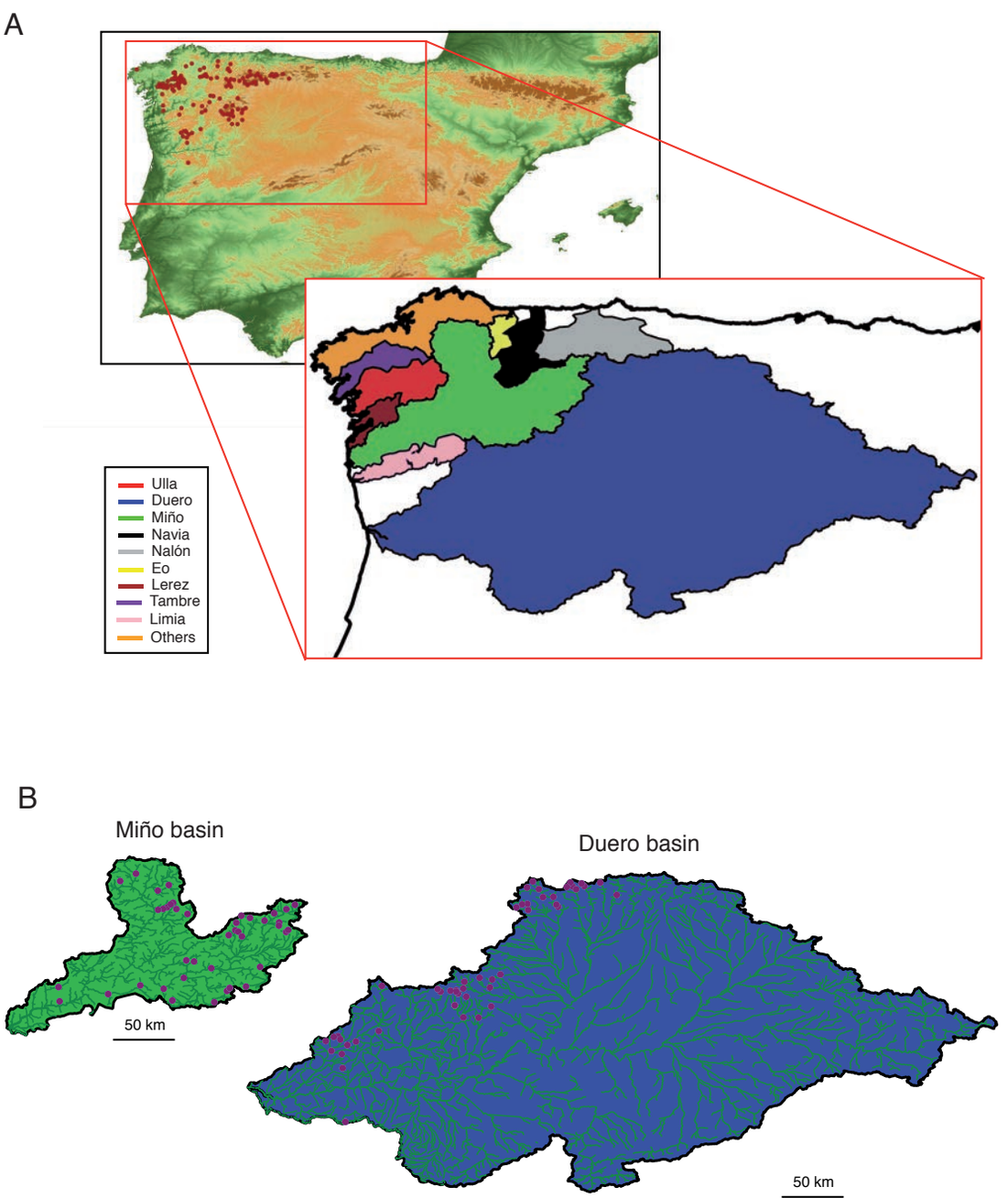


Figure S1. Map of the study area showing the main basins (A), and enlargement of the Miño and Duero basins (B). A number of smaller basins were grouped as "Others"

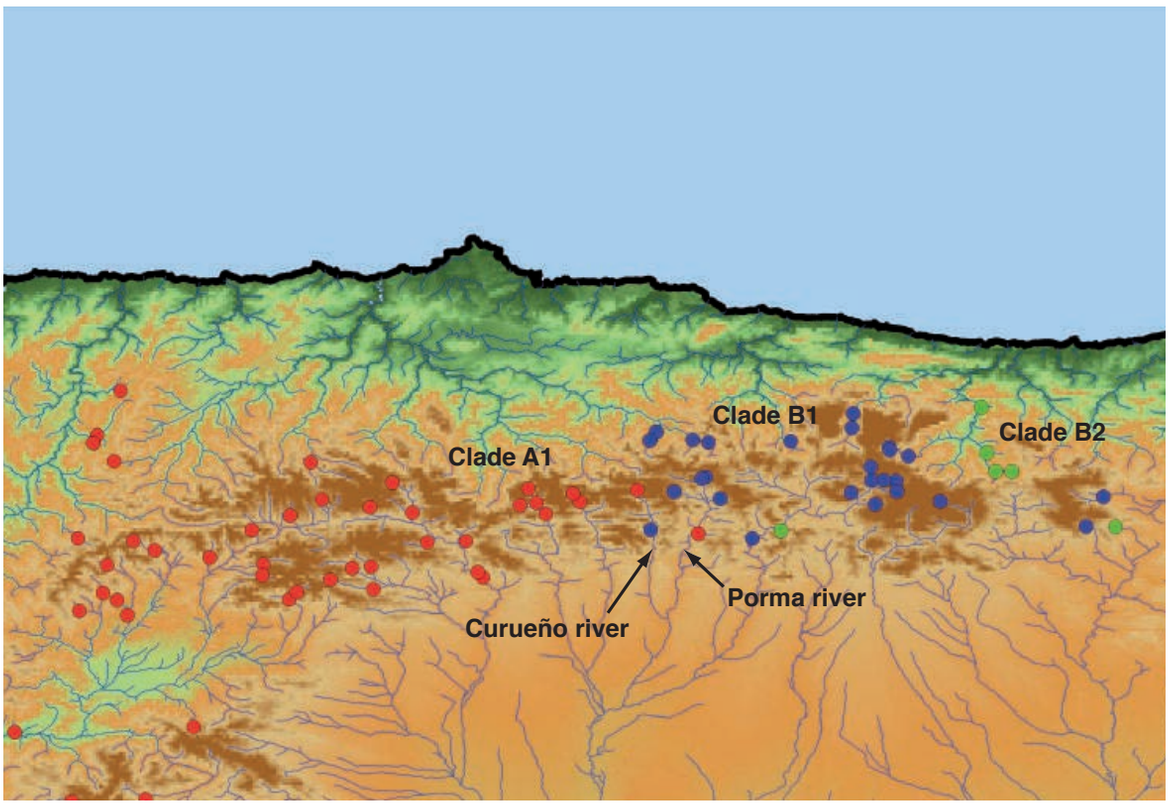


Figure S2. Enlargement of the contact zone between the main mitochondrial clades in the Cantabrian Mountains.

CHAPTER 3

EVOLUTIONARY HISTORY OF THE ENDEMIC MEDITERRANEAN WATER SHREW (*Neomys anomalus)*: GLACIAL REFUGIA WITHIN THE IBERIAN PENINSULA


Photo Credit: kleinsaeuger.at

Marina Querejeta and Jose Castresana

Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain

ABSTRACT

Inferring glacial refugia is key to understanding the origin of extant Paleartic species. The Iberian Peninsula was likely one of the most important glacial refugia during the Last Glacial Maximum, giving rise to different endemic species. This is the case of the Mediterranean water shrew *(Neomys anomalus)*, which is a recently delimited species endemic to the Iberian Peninsula and inhabits riparian ecosystems. Here, we inferred its evolutionary history and glacial refugia during the Last Glacial Maximum combining phylogeography, spatial distribution of genetic diversity and species distribution modelling. To perform these analyses, we used 133 non-invasive samples covering most of its distribution and sequenced partial mitochondrial cytochrome *b* and D-loop genes. Using these sequences, we constructed maximum-likelihood and Bayesian phylogenetic trees, which indicated that *N. anomalus* was divided into two main phylogroups (A and B), strongly correlated with geography. The nucleotide diversity interpolation maps showed the highest genetic diversity in the Cantabrian Mountains, in the north of the Iberian Peninsula. Finally, we performed a species distribution modelling using the Global Biodiversity Information Facility and our sampling localities as occurrences together with bioclimatic variables, and projected this model into the current and past (Last Interglacial and Last Glacial Maximum) climate scenarios. Combined results allowed inferring multiple glacial refugia during the Last Glacial Maximum and more ancient refugia in the Last Interglacial, restricted to the north of the Iberian Peninsula. This is the first attempt to infer the evolutionary history of this newly delimited species, whose restricted distribution and vulnerable aquatic habitat might make it subject of conservation concern.

KEY WORDS: phylogeography, nucleotide diversity, species distribution modelling, *Neomys anomalus*, Iberian Peninsula, endemism

INTRODUCTION

Evolutionary processes drive intraspecific phylogeographic patterns, making the interface between genetic and geographic structure a key tool to detect past glacial refugia (Avise et al., 1987; Avise, 2000), study the diversification of lineages in a geographic space (Lansman et al., 1983), interpret genetic data in a spatial context (Avise et al., 1987; Avise, 1989, 2000) and, finally, understand the evolutionary history of species. Moreover, evolutionary histories become especially interesting in endemic species due to their uniqueness from a conservation perspective (Myers et al., 2000).

During the Quaternary glaciations, and particularly the Last Glacial Maximum, worldwide ice-sheets reached considerable high volumes (Yokoyama et al., 2000), reducing or making disappear species habitats within the Paleartic region. Populations that were able to resist these disadvantageous climate conditions remained isolated in glacial refugia during thousands of years, giving rise to

new lineages that colonized the Paleartic region during the Holocene. Therefore, glacial refugia are known to have preserved a large proportion of intraspecific diversity whose identification is essential to infer the evolutionary history of species (Petit et al., 2003; Provan & Bennett, 2008; Clark et al., 2009).

The Iberian Peninsula was one of the most important glacial refugia in Europe during the Pleistocene (Hewitt, 1999, 2000). This is mainly because this region harbours a wide range of different climates and habitats in a relatively small region (Gómez & Lunt, 2007). This explains the large amount of species that remained isolated during the Last Glacial Maximum in the Iberian Peninsula, giving rise to a large number of endemisms (Gómez-Campo et al., 1984). The current distributions of these species is consistent with long periods of isolation in several locations, followed by several postglacial dispersal events. This plausible scenario of multiple glacial refugia within the Iberian Peninsula and other south European peninsulas, called "refugia-within-refugia" (Gómez & Lunt, 2007), has opened a promising path to perform comparative phylogeographic studies aiming to reveal common refugia and postglacial dispersal patterns. Although this scenario has been confirmed for several Iberian endemics (Consuegra et al., 2002; Martínez-Solano et al., 2006; Gonçalves et al., 2009; Stamatis et al., 2009; Igea et al., 2013; Barbosa et al., 2016), there are still many of these species whose evolutionary patterns remain unknown.

The Mediterranean water shrew *(Neomys anomalus)* is a small semi-aquatic mammal belonging to the family Soricidae. It inhabits riverine ecosystems although it is also able to reach locations far from riverbanks (Rychlik, 1997). They have several morphological adaptations to aquatic environment such as large hind feet and the presence of short stark hairs in the tail and along the side of their toes, which are used to improve their swimming capacity (Krystufek et al., 2000). They additionally have a short snout to look for food in the river environment (Rychlik, 1997).

This species used to be considered as the subspecies *Neomys anomalus anomalus*, but a recent genetic species delimitation study on the genus considered that it deserved species status, separated from *N. milleri*, of mainly European distribution (Igea et al., 2015). This study dated the split between *N. milleri* and *N. anomalus* at 0.4 Myr ago. *N. anomalus* is then endemic to the Iberian Peninsula, inhabiting its northern half and two isolated patches in the southern half (Ventura, 2007). Although it is considered as Least Concern by the IUCN, its fragmented and reduced distribution makes it susceptible to local extinctions (Hutterer et al., 2008). Despite the importance of this shrew for the Iberian fauna, there are no previous studies about its evolutionary history. Being a semi-aquatic species, its study becomes more challenging, because factors such as the structure of river networks could play an important role in shaping its genetic structure (Vignieri, 2005; Furlan et al., 2013; Igea et al., 2013; Byrne et al., 2015, Querejeta et al., in press).

In this work, we aim to study the evolutionary history of the Mediterranean water shrew and infer

the main potential glacial refugia for this species. Our main questions are: 1) Is the genetic structure of *N. anomalus* correlated with geography and did river basins play a role in shaping this genetic structure? 2) Is there a structure in the genetic diversity of this species? 3) What are the potential refugia during the Last Glacial Maximum and, also, in the Last Interglacial?. To achieve these aims, we combined a phylogeographic and genetic diversity approach together with a species distribution modelling, as the combination of these techniques can provide a better picture of the potential evolutionary scenarios (Guisan & Zimmermann, 2000) (Graham et al., 2004) (Willerslev et al., 2007).

MATERIAL AND METHODS

SAMPLING, DNA EXTRACTION AND AMPLIFICATION

We used 133 samples of *Neomys anomalus* from across most of its distribution range (Table_S1). A total of 126 samples were obtained from excrements that water shrews deposit on emerged rocks of rivers (Igea et al., 2015). One sample was obtained from the Mammals collections of the Estación Biológica de Doñana (EBD, CSIC). The sampling was complemented with 6 additional samples from a previous study (Igea et al., 2015).

DNA was extracted using DNEasy Blood and Tissue Kit from QIAGEN, following the manufacturer's instructions with the exception that the elution volume was 50 µl.

We then sequenced 752 bp of the mitochondrial cytochrome *b* gene using primers from a previous study (Igea et al., 2015). In addition, we designed forward and reversed primers to sequence a mitochondrial D-loop fragment of 500 bp with the help of Primer3Plus (Untergasser et al., 2007). The primers used were: ATCAACACCCAAAGCTGATATTCTA and TTAATGTGCCTTGTCCGATT. Due to the complexity of the D-loop, which showed several repeat regions, the fragment was trimmed to a 272 bp fragment, from position 128 to 400, which was common to all the sequences. PCR conditions to amplify both genes were as in Igea et al. (Igea et al., 2015). PCR products were checked in an agarose gel and purified using ExoSAP-IT. They were then sequenced in Macrogen Inc (Seoul, South Korea).

The sequences obtained were edited and assembled with Genious Pro (Biomatters Ltd). MAFFT v7.130b (Katoh et al., 2002) was used to produce an alignment for the D-loop fragment, choosing the FFT-NS-i strategy (iterative refinement method). Gap positions within this D-loop alignment were eliminated using Gblocks (Castresana, 2000). Finally, both the cytochrome b and D-loop alignments were concatenated.

SUBSTITUTION MODEL SELECTION AND PHYLOGENETIC ANALYSES

Before performing the phylogenetic analyses, we chose the best partition and substitution model scheme for the concatenated genes using the software PartitionFinder version 1.1.1 (Lanfear et al.,

2012) with the options for RaxML (Stamatakis, 2006) and BEAST (Drummond & Rambaut, 2007). We defined four possible partitions: one for each codon position of the cytochrome *b* fragment and a fourth one for the D-loop fragment. We selected the Bayesian Information Criterion, branch lengths linked and all partition schemes analysed. In the case of BEAST, PartitionFinder defined four partitions: K80+I was selected for the first codon position of cytochrome *b*, HKY for the second position, HKY+G for the third positions and HKY+I+G for the D-loop. For RaxML, three partitions were defined: GTR+G for the first and second codon positions, GTR+G for the third codon position and GTR+I+G for the D-loop.

A maximum-likelihood tree was generated with RaxML (Stamatakis, 2006), using the three partitions defined by PartitionFinder. Since RaxML uses a single model for all partitions, we selected GTR+I+G, but the same result was obtained with GTR+G (not shown).

A Bayesian inference phylogenetic tree was conducted using BEAST v1.8.4. (Drummond & Rambaut, 2007). The substitutions models selected by PartitionFinder and a strict molecular clock were applied to each of the four partitions. The analysis was based on three independent Markov Chain Monte Carlo runs of 10,000,000 generations sampled every 1,000 generations. Convergence and mixing of each run was checked by means of Tracer v.1.6. (Rambaut et al., 2014). The three runs converged to similar posterior estimates and their parameters and trees were combined with LogCombiner v1.8.4. (part of BEAST package), after removing 1,000,000 generations as burning in each run. Finally, we calculated a summary tree as the maximum clade credibility tree with median node heights using TreeAnnotator v.1.8.4. (also included in BEAST package).

EXPLORING GENETIC DIVERSITY

We computed, in each sampling point, the values of nucleotide diversity (π) for all samples in a radius of 75 km around it, using the alignment of cytochrome *b* and the sample coordinates (Igea et al., 2013). We then used an inverse distance weighted spatial interpolation to estimate the values of the nucleotide diversity (π) in a grid of 1x1 km of the distribution range. To do this, we used the function *idw* of the R package GSTAT (Pebesma et al., 2016). The obtained *raster* layer with the surface interpolated was superimposed on a map of the distribution range of *N. anomalus*. This procedure was repeated for all sequences and for the different detected clades separately.

HIERARCHICAL ANALYSIS OF THE MOLECULAR VARIANCE (AMOVA)

We performed a one-way hierarchical analysis of the molecular variance (AMOVA) to shed light on which amount of variability within the Mediterranean water shrew can be explained by river basins. We first computed a matrix of pairwise phylogenetic distances of the complete sequences based on the maximum-likelihood tree, using the function *cophenetic.phylo* from the R package *ape* (Paradis et al., 2004). The river basins were treated as a unique factor and the AMOVA was performed using the R package *pegas* (Paradis, 2010).

SPECIES DISTRIBUTION MODELLING FOR *Neomys anomalus*

We compiled an occurrence dataset from the Global Biodiversity Information Facility (Flemons et al., 2007) within its recently delimited distribution (Igea et al., 2015) and, also, the sampling obtained for this study. The initial dataset had 911 georreferenced occurrence localities. To avoid biases, we selected one sample for each 1x1 km grid cell (Veloz, 2009; Anderson, 2012; Hijmans, 2012) using the function *gridSample* from the R package RASTER (Hijmans & Van Etten, 2014). After the filtering process, the final occurrence dataset contained 858 presences.

The environmental data was obtained from WorldClim 1.4 (Hijmans et al., 2006), using a resolution of 2.5 arc-minutes, for both current and past data (Last Glacial Maximum and Last Interglacial). Three different general atmospheric circulation models of the Last Glacial Maximum were used: CCSM, MIROC and MSPI. All the layers were cropped to the extent of the Iberian Peninsula (35.95833, 43.95833, -9.708333, 3.625) using the R package RASTER (Hijmans & Van Etten, 2014). From the 19 bioclimatic variables and the altitude layer also available in WorldClim 1.4 (Hijmans et al., 2006), a selection was performed to reduce the multivariate complexity and avoid correlated variables. For this purpose, we computed the Pearson's correlation matrix between all the bioclimatic variables (Dormann et al. 2007; Dormann et al. 2013) and a cladogram from this matrix using the function *hclust* of the R package STATS (Hijmans & van Etten, 2014). A total of 11 bioclimatic variables with a correlation lower than 0.8 were included in the models: Annual Mean Temperature (BIO1), Mean Diurnal Range (BIO2), Isothermality (BIO3), Minimum Temperature of Coldest Month (BIO6), Temperature Annual Range (BIO7), Mean Temperature of Wettest Quarter (BIO8), Mean Temperature of Driest Quarter (BIO9), Mean Temperature of Warmest Quarter (BIO10), Precipitation of Wettest Month (BIO13), Precipitation of Driest Month (BIO14) and Precipitation Seasonality (BIO15).
We generated the species distribution model using the maximum entropy algorithm implemented in Maxent 3.3.3k (Phillips et al., 2012). The model was replicated 50 times, selecting training and test samples randomly. Other settings were left as default. We then chose the area under the curve (AUC) of the receiver operating characteristics as evaluator. The average model obtained was projected onto the past climate conditions (Last Glacial Maximum and Last Interglacial). To show a consensus of the three Last Glacial Maximum projections, the output of each atmospheric circulation model was set to logistic and we created a binary layer using the 10^{th} percentile of the value of each model. Finally, the consensus of the three binary layers was obtained using the *Raster* Calculator tool included in the software Quantum GIS (QGIS) version 2.0.1 (QGIS Development Team, 2009). Since the use of a standard 10^{th} percentile to generate the consensus (Garcia-Porta et al., 2012; Valbuena-Ureña et al., 2013) generated an optimal distribution covering most of the studied area (not shown), we also obtained a more restricted projection that would be more informative at the scale of the Iberian Peninsula, using the 90^{th} percentile.

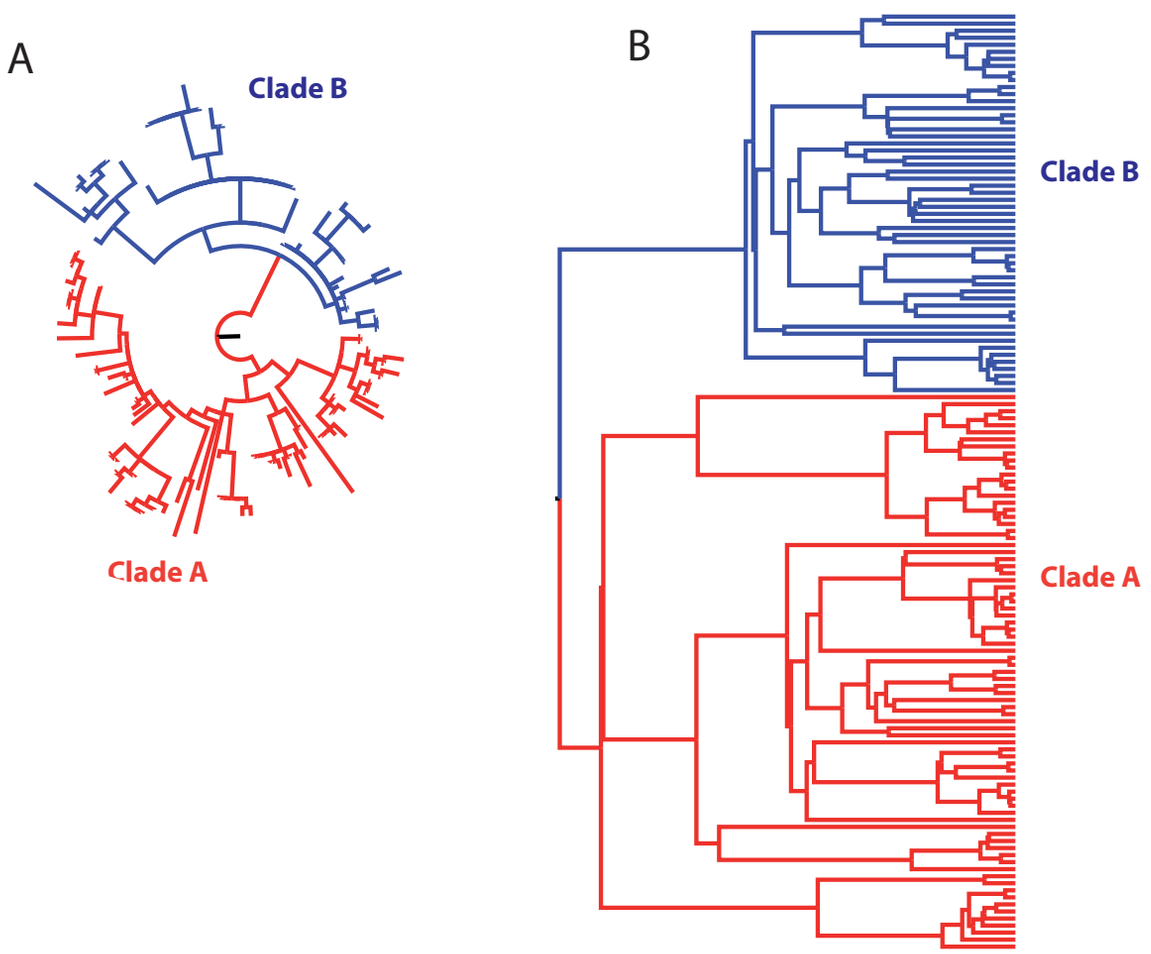
Assuming that high genetic diversity and Last Glacial Maximum optimal habitat areas are consistent with location of potential glacial refugia, we inferred the location of the glacial refugia for *Neomys*

anomalus. To do this, we first extracted the 90th percentile values of each atmospheric circulation model (CCSM, MIROC and MSPI) and we produced a *raster* with this value for each of these models. Then, we summed the three *raster* layers up in one binary layer using the *Raster* Calculator tool in Quantum GIS (QGIS) version 2.0.1. (QGIS Development Team, 2009). This binary layers has 4 values: 0 for areas where there was no 90th percentile projection for the LGM and 1, 2 and 3, for areas where one, two and 3 of the 90th projections was present, respectively. The areas with value 3 would be the ones with the highest probability of being the most optimal habitat for *N. anomalus* in the LGM. Moreover, we overlapped these areas of value 3 and another polygon shapefile with areas of high nucleotide diversity An intersection was obtained when we selected a threshold of nucleotide diversity. An intersection was obtained when we selected a threshold of nucleotide diversity of of 0.0075. Finally, we intersected both shapefiles, assuming these intersected areas to be the ones with highest probability to have been a glacial refugia.

RESULTS

PHYLOGENETIC ANALYSES

Both the maximum-likelihood (Figure 1A) and the Bayesian inference (Figure 1B) trees performed with the concatenated sequences showed two main phylogroups, A and B, highly correlated with geographic structure (Figure 1C). Two contact zones can be detected in the map between both phylogroups, one located in the Cantabrian Mountains and the other in the Central System (Figure



C

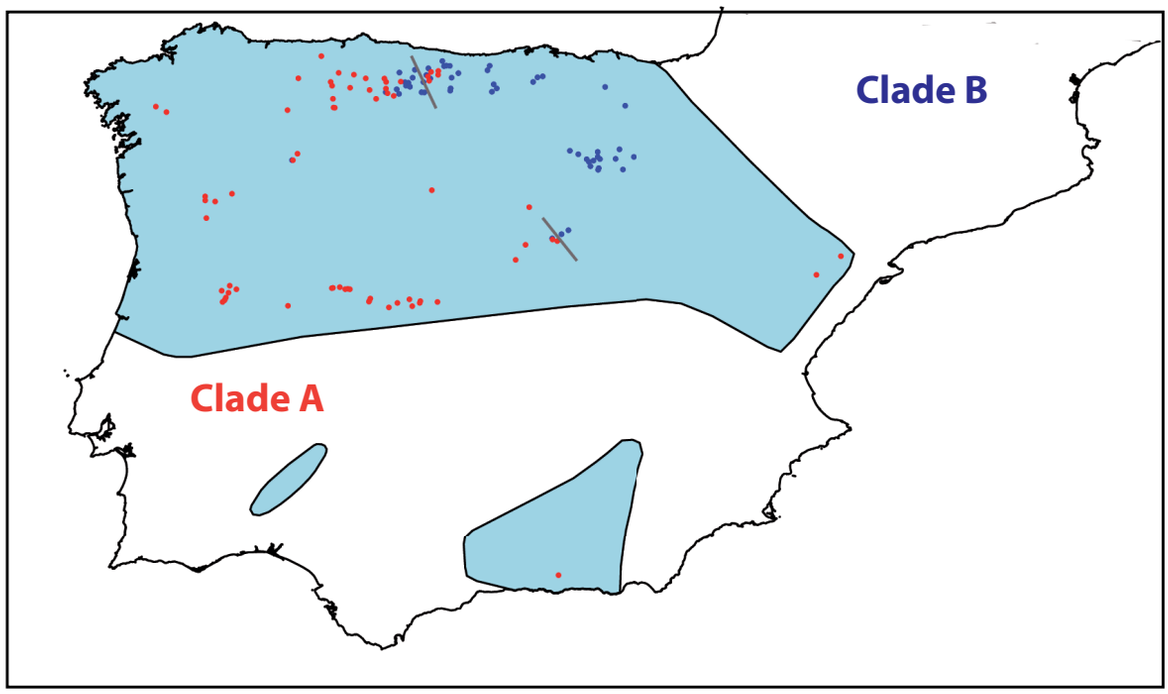


Figure 1. Phylogeography of the Mediterranean water shrew. A) Maximum likelihood phylogenetic tree showing clades A and B. B) Bayesian Inference phylogenetic tree showing the two clades. C) Map showing the geographic structure of the species

1A), both presenting some spatial mixing between the clades. The level of mixing in the contact zone is higher in the Cantabrian Mountains.

GENETIC DIVERSITY

Three plots of the interpolated nucleotide diversity showed strong differences in genetic diversity within the distribution range, with the areas of highest genetic diversity located in the north of the Iberian Peninsula (Figure 2). The highest value of nucleotide diversity was 0.0092 and the lowest value was 0.0027. When clades A and B were analysed separately, spots of high genetic diversity were also detected in the northern area (Figure S1).

The one-way AMOVA using the 10 river basins as a factor (Table S1) showed that only 13 % of the genetic variation was due to isolation by river basins (Table S2).

SPECIES DISTRIBUTION MODELLING

The average model calibrated with current occurrences of *Neomys anomalus* showed an AUC value of 0.841 with a mean standard deviation of 0.004, indicating a good performance of the model (Elith, 2000). The bioclimatic variables Annual Mean Temperature (BIO1) and Mean Temperature of Warmest Quarter (BIO10) showed the highest contribution to the SDM, with 45.3% and 21.9%, respectively. The current potential distribution matches the actual distribution, except that the southeastern area shows a low potential distribution (Figure S2A).

The prediction map for the Last Interglacial (Figure 3A) shows a potential distribution restricted to

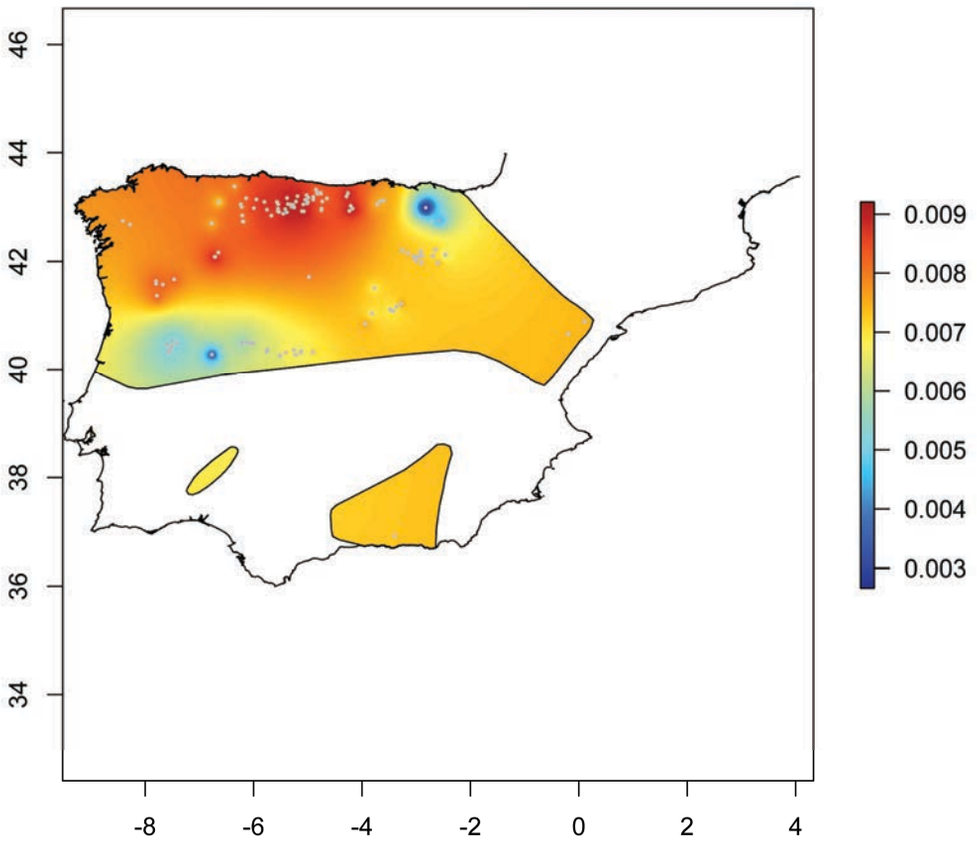


Figure 2. Interpolation of the nucleotide diversity using the inverse distance weight (IDW) method. This figure shows spots of high and low genetic diversity within the distribution of the Mediterranean water shrew.

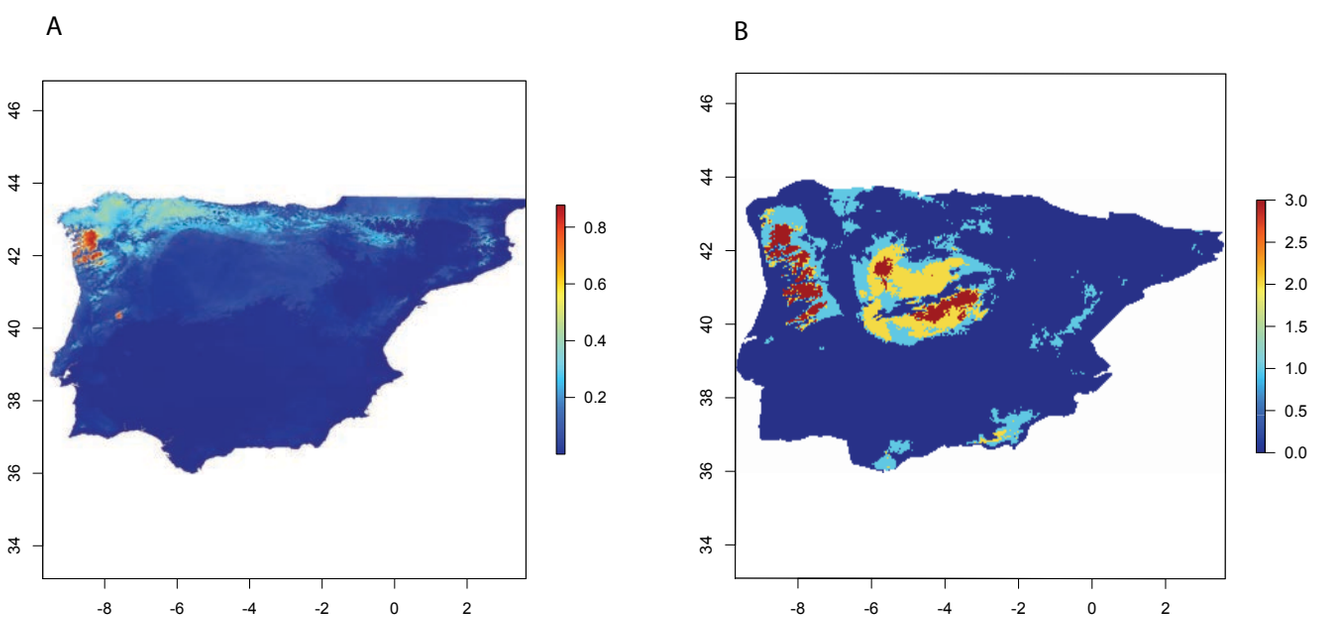


Figure 3. A) Projection of the SDM to the Last Interglacial. B) Assembly binary layer of the three atmospheric circulation models for the projection to the Last Glacial Maximum, using the 90th percentile model values.

the north of the Iberian Peninsula. In contrast, in the Last Glacial Maximum, the optimal surface increases substantially for the three climatic scenarios (Figure S2B), even when using the 10% percentile for generating the consensus of the three circulation models (Figure 3B). This layer shows main regions, one in the northwestern region of the Iberian Peninsula and another one in the Central

System and the Central Plateau.

INFERENCE OF GLACIAL REFUGIA

The intersection of the polygons of the outline of the genetic diversity and the optimal regions resulted in regions located in the northwestern region and the Central Plateau (Figure 4).

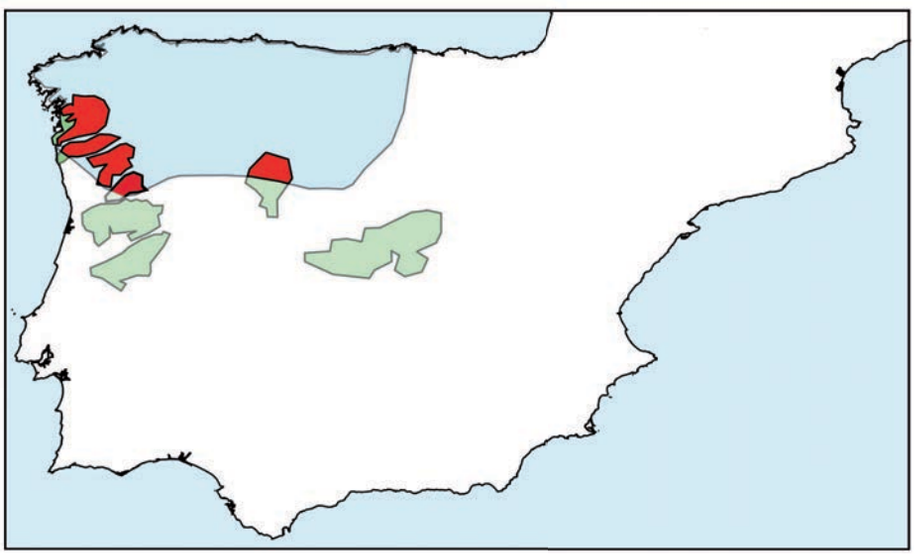


Figure 4. Intersection (red) between the area of the highest nucleotide diversity (blue) and the area with the highest probability of being the optimal region for *N.anomalus* in the Last Glacial Maximum.

DISCUSSION

INTRASPECIFIC PHYLOGEOGRAPHY OF THE MEDITERRANEAN WATER SHREW *(Neomys anomalus)*

This study is the first attempt to explain the glacial history of this endemic mammal from the Iberian Peninsula after it was recently described as a species separated from *N. milleri* (Igea et al., 2015).

Regarding the intraspecific phylogeography, the two main phylogroups detected, A and B, are mainly allopatric (Fig.1). The two main contact zones between these clades, situated in the Cantabrian Mountains and the Central System, respectively, are well defined, although there is spatial mixing in some localities of both clades. For instance, the phylogroups A and B coexist in the same area of Picos de Europa in the Cantabrian Mountains, which is consistent with the high nucleotide diversity in that area (Fig. 2). However, when both clades A and B were analysed separately, they also showed a high diversity region in this Cantabrian Mountains region, meaning that the contact zone is not the only reason of this high diversity. Interestingly, this region was an optimal habitat for *Neomys anomalus* in the Last Interglacial, which is consistent with the continuous existence of the species for a long period of time and the high genetic diversity observed in this area. In contrast, the contact zone in the Central System has low levels of genetic diversity, indicating that *N. anomalus* could have more recently colonized this region.

Although the Mediterranean water shrew shows semi-aquatic habits and is associated to riparian

ecosystems, there is a low contribution of the river basins to its genetic structure (13%) according to the one-way AMOVA performed. This result suggests that this species is able to move, disperse and expand its populations overland. That is, although the species uses the riverine habitat to feed or perform short dispersals, long-distance overland dispersals must have occurred to explain the AMOVA.

The data from different endemic species and the comparison of the phylogeography of these species may lead to new perspectives to understand the origin of Iberian fauna (Consuegra et al., 2002; Martínez-Solano et al., 2006; Gonçalves et al., 2009; Stamatis et al., 2009; Igea et al., 2015; Barbosa et al., 2016). In this sense, there is a considerable correspondence of the contact zones between the two mitochondrial clades of *N. anomalus* that we show here and the ones described for the Pyrenean desman (*Galemys pyrenaicus*), another semi-aquatic mammal endemic to the Iberian Peninsula. In the Pyrenean desman , the mitochondrial phylogenetic tree revealed also two major clades and two contact zones between them (Igea et al. 2013). One of the contact zones is situated in the Cantabrian Mountains, in a very similar position to that of *N. anomalus*. The other one is in the Iberian System, to the north of the equivalent contact zone in *N. anomalus*. In the case of *G. pyrenaicus*, both contact zones are stricter, with basically no spatial mixing between clades. Therefore, it seems that glacial cycles have exerted similar effects into two these semi-aquatic habitats. However, the differences in the degree of mixing that takes places in the contact zones of the two species may be due to the different dispersal patterns, with *G. pyrenaicus* having a more reduced dispersal potential than *N. anomalus*.

GLACIAL REFUGIA

The species distribution models are consistent in showing the northwestern area of the Iberian Peninsula as an optimal habitat for *N. anomalus* during the Last Interglacial and the Last Glacial Maximum. Also, this region overlaps with the area of highest genetic diversity and therefore this area can be considered as the main potential glacial refugium during the Last Glacial Maximum. There is another optimal area situated in the Central Plateau, which could have also harboured a refugium. Other areas with either high genetic diversity or optimal habitat during the Last Glacial Maximum cannot be discarded, although the scenarios to explain them may be more complex. The existence of areas of high genetic diversity without being optimal habitats during the glaciation could be due to recent migration of haplotypes to these areas. On the other hand, potential refugia according to the species distribution models that do not currently have high genetic diversity may have corresponded to minor refugia.

After the split of *N. anomalus* and *N. milleri* 0.40 Myr ago (Igea et al., 2015), *N. anomalus* may have become isolated likely in the northwestern region of the Iberian Peninsula, as suggested by the large optimal area in the Last Interglacial projection (Fig.3). Then, the existence of at least two potential glacial refugia during the Last Glacial Maximum could have differentiated clades A and B. The occidental refugium would have been the most important one for the species. The position of this refugium in the occidental area, far from the distribution of *N. milleri*, gives further support to the

large period of isolation of these two species (Igea et al. 2015).

CONSERVATION IMPLICATIONS

In this study, we inferred the evolutionary history of the Mediterranean water shrew, whose current conservation status is Least Concern by the IUCN (Hutterer et al., 2008). However, this evaluation was performed for the European distribution of both *N. anomalus* and *N. milleri*. Also, it is known that riverine ecosystems and wetland areas are decreasing due to human activities and climate change, making its distribution range much more fragmented than it used to be, (Hutterer et al., 2008). Thus, the results from this study, which gives further support to the species status of *N. anomalus*, reinforce the need to improve the knowledge of the populations of this endemic shrew and eventually reassess its conservation status.

ACKNOWLEDGMENTS

We thank all persons who provided samples, especially Angel Fernández González, Pere Aymerich, Joaquim Gosálbez, Javier Igea, Julio Gisbert, Rosa García Perea and Jacinto Román. We acknowledge a sample from the Mammals collections of the Estación Biológica de Doñana (EBD, CSIC). We also thank Pedro Tarroso, Luis Machado and Joan Garcia-Porta and Javier Igea for their advice and help in performing the analyses. This work was financially supported by research project CGL2014-53968-P of the "Plan Nacional I+D+I del Ministerio de Economía y Competitividad" (Spain) to J.C, cofinanced with FEDER funds. M.Q. was supported by fellowship BES-2012-057135 of the Ministerio de Economía y Competitividad (Spain).

REFERENCES

Anderson R.P. (2012) Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*, 1260, 66–80.

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

Avise J.C. (2000) *Phylogeography: the history and formation of species*. Harvard university press,

Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A., & Saunders N.C. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual review of ecology and systematics*, 489–522.

Barbosa S., Paupério J., Herman J.S., Ferreira C.M., Pita R., Vale-Gonçalves H.M., Cabral J.A., Garrido-García J.A., Soriguer R.C., Beja P., & others (2016) Endemic species may have complex histories: within-refugium phylogeography of an endangered Iberian vole. *Molecular Ecology* , 26, 951-967.

Byrne M.S., Quintana R.D., Bolkovic M.L., Cassini M.H., & Túnez J.I. (2015) The role of river drainages in shaping the genetic structure of capybara populations. *Genetica*, 143, 645–656.

Castresana J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, 17, 540–552.

Clark P.U., Dyke A.S., Shakun J.D., Carlson A.E., Clark J., Wohlfarth B., Mitrovica J.X., Hostetler S.W., & McCabe A.M. (2009) The last glacial maximum. *Science*, 325, 710–714.

Consuegra S., de Leániz C., Serdio A., González Morales M., Straus L.G., Knox D., & Verspoor E. (2002) Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular ecology*, 11, 2037–2048.

Dormann C.F., Elith J., Bacher S., Buchmann C., Carl G., Carré G., Marquéz J.R.G., Gruber B., Lafourcade B., Leitão P.J., & others (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27–46.

Dormann C.F., M McPherson J., B Araújo M., Bivand R., Bolliger J., Carl G., G Davies R., Hirzel A., Jetz W., Daniel Kissling W., & others (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609–628.

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

Elith J. (2000) Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. *Quantitative methods for conservation biology* pp. 39–58.

Flemons P., Guralnick R., Krieger J., Ranipeta A., & Neufeld D. (2007) A web-based GIS tool for exploring the world's biodiversity: The Global Biodiversity Information Facility Mapping and Analysis Portal Application (GBIF-MAPA). *Ecological Informatics*, 2, 49–60.

Furlan E.M., Griffiths J., Gust N., Handasyde K.A., Grant T.R., Gruber B., & Weeks A.R. (2013) Dispersal patterns and population structuring among platypuses, *Ornithorhynchus anatinus*, throughout south-eastern Australia. *Conservation Genetics*, 14, 837–853.

Garcia-Porta J., Litvinchuk S.N., Crochet P.-A., Romano A., Geniez P.H., Lo-Valvo M., Lymberakis P., & Carranza S. (2012) Molecular phylogenetics and historical biogeography of the west-palearctic common toads (*Bufo bufo* species complex). *Molecular Phylogenetics and Evolution*, 63, 113–130.

Gómez A. & Lunt D.H. (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia* pp. 155–188. Springer,

Gómez-Campo C., Bermudez-de-Castro L., Cagiga M.J., & Sánchez-Yélamo M.D. (1984) Endemism in the Iberian peninsula and Balearic Islands. *Webbia*, 38, 709–714.

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

Graham C.H., Ron S.R., Santos J.C., Schneider C.J., & Moritz C. (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, 58, 1781–1793.

Guisan A. & Zimmermann N.E. (2000) Predictive habitat distribution models in ecology. *Ecological modelling*, 135, 147–186.

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

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

Hijmans R. & van Etten J. (2014) *raster*: *raster*: Geographic data analysis and modeling. *R package version*, 517, 2.

Hijmans R.J. (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, 93, 679–688.

Hijmans R.J., Cameron S.E., & Parra J.L. (2006) WorldClim Global Climate Layers Version 1.4.

Hutterer R., Amori G., Kryštufek B., Yigit N., Mitsain G., Meinig H., Bertolino S., & Palomo L.J. (2008) *Neomys anomalus*.

Igea J., Aymerich P., Bannikova A.A., Gosálbez J., & Castresana J. (2015) Multilocus species trees and species delimitation in a temporal context: application to the water shrews of the genus Neomys. *BMC evolutionary biology,* 15, 1.

Igea J., Aymerich P., Fernández-González A., González-Esteban J., Gómez A., Alonso R., Gosálbez J., & Castresana J. (2013) Phylogeography and postglacial expansion of the endangered semi-aquatic mammal *Galemys pyrenaicus*. *BMC evolutionary biology*, 13, 115.

Katoh K., Misawa K., Kuma K., & Miyata T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, 30, 3059–3066.

Krystufek B., Davison A., & Griffiths H.I. (2000) Evolutionary biogeography of water shrews (*Neomys* spp.) in the western Palaearctic Region. *Canadian journal of zoology*, 78, 1616–1625.

Lanfear R., Calcott B., Ho S.Y.W., & Guindon S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution*, 29, 1695–1701.

Lansman R.A., Avise J.C., Aquadro C.F., Shapira J.F., & Daniel S.W. (1983) Extensive genetic variation in mitochondrial DNA's among geographic populations of the deer mouse, *Peromyscus maniculatus. Evolution*, 37, 1-16.

Martínez-Solano I., Teixeira J., Buckley D., & GARCÍA-PARÍS M. (2006) Mitochondrial DNA phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple refugia in an Iberian endemic. *Molecular Ecology*, 15, 3375–3388.

Myers N., Mittermeier R.A., Mittermeier C.G., Da Fonseca G.A.B., & Kent J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, 403, 853–858.

Paradis E. (2010) pegas: an R package for population genetics with an integrated--modular approach. *Bioinformatics*, 26, 419–420.

Paradis E., Claude J., & Strimmer K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.

Pebesma E., Graeler B., & Pebesma M.E. (2016) Package "gstat." .

Petit R.J., Aguinagalde I., de Beaulieu J.-L., Bittkau C., Brewer S., Cheddadi R., Ennos R., Fineschi S., Grivet D., Lascoux M., & others (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, 300, 1563–1565.

Phillips S., Dudik M., & Schapire R. (2012) MaxEnt 3.3. 3k. *AT & T Labs-Research, Princeton University*.

Provan J. & Bennett K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in ecology & evolution*, 23, 564–571.

Querejeta M., Fernández-González A., Romero R & Castresana J. (in press). Postglacial dispersal patterns and mitochondrial genetic structure of the Pyrenean desman (*Galemys pyrenaicus*) in the north-western region of the Iberian Peninsula. *Ecology and Evolution.*

QGIS Development Team (2009) QGIS Geographic Information System.

Rambaut A., Suchard M.A., Xie D., & Drummond A.J. (2014) Tracer v1. 6.

Rychlik L. (1997) Differences in foraging behaviour between water shrews: *Neomys anomalus and Neomys fodiens. Acta Theriologica*, 42, 351–386.

Stamatakis A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* (Oxford, England), 22, 2688–90.

Stamatis C., Suchentrunk F., Moutou K.A., Giacometti M., Haerer G., Djan M., Vapa L., Vukovic M., Tvrtković N., Sert H., & others (2009) Phylogeography of the brown hare (*Lepus europaeus*) in Europe: a legacy of south-eastern Mediterranean refugia? *Journal of Biogeography,* 36, 515–528.

Untergasser A., Nijveen H., Rao X., Bisseling T., Geurts R., & Leunissen J.A.M. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic acids research,* 35, W71–W74.

Valbuena-Ureña E., Amat F., & Carranza S. (2013) Integrative phylogeography of *Calotriton* newts (Amphibia, Salamandridae), with special remarks on the conservation of the endangered Montseny brook newt (*Calotriton arnoldi*). PloS *one*, 8, e62542.

Veloz S.D. (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36, 2290–2299.

Ventura J. (2007) *Neomys anomalus* Cabrera, 1907. *Atlas y libro rojo de los mamíferos terrestres de España. Madrid, Spain: Dirección General para la Biodiversidad-SECEMSECEMU,* 114–116.

Vignieri S.N. (2005) Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse *(Zapus trinotatus). Molecular Ecology*, 14, 1925–1937.

Willerslev E., Cappellini E., Boomsma W., Nielsen R., Hebsgaard M.B., Brand T.B., Hofreiter M., Bunce M., Poinar H.N., Dahl-Jensen D., & others (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, 317, 111–114.

Yokoyama Y., Lambeck K., De Deckker P., Johnston P., & Fifield L.K. (2000) Timing of the Last Glacial Maximum from observed sea-level minima. *Nature*, 406, 713–716.

SUPPORTING INFORMATION

EVOLUTIONARY HISTORY OF THE ENDEMIC MEDITERRANEAN WATER SHREW (*Neomys anomalus*): GLACIAL REFUGIA WITHIN THE IBERIAN PENINSULA

Marina Querejeta & Jose Castresana

Specimen Code	Latitude	Longitude	Basin	Clade	Sample Type
EBD-3530	36.92	-3.42	Guadalfeo	A	Nail
IBE-BC0112	42.71	-6.78	Miño	A	Faeces
IBE-BC0121	40.33	-4.92	Tajo	A	Faeces
IBE-BC0201	40.50	-6.22	Duero	A	Faeces
IBE-BC0209	42.85	-5.68	Duero	A	Faeces
IBE-BC0246	40.28	-6.77	Duero	A	Faeces
IBE-BC0254	40.33	-5.77	Tajo	A	Faeces
IBE-BC0304	43.01	-5.31	Duero	B	Faeces
IBE-BC0357	40.50	-6.22	Duero	A	Faeces
IBE-BC0359	42.85	-5.68	Duero	A	Faeces
IBE-BC0374	40.34	-5.76	Tajo	A	Faeces
IBE-BC0396	42.96	-5.76	Duero	A	Faeces
IBE-BC0407	40.50	-6.23	Duero	A	Faeces
IBE-BC0416	40.49	-6.06	Tajo	A	Faeces
IBE-BC0530	40.49	-6.03	Tajo	A	Faeces
IBE-BC0540	42.85	-5.68	Duero	A	Faeces
IBE-BC0589	40.32	-5.14	Duero	A	Faeces
IBE-BC0592	42.85	-5.68	Duero	A	Faeces
IBE-BC0632	42.93	-5.56	Duero	B	Faeces
IBE-BC0642	42.96	-5.76	Duero	A	Faeces
IBE-BC0662	42.97	-5.42	Duero	A	Faeces
IBE-BC0694	43.05	-5.32	Duero	B	Faeces
IBE-BC0708	42.85	-5.68	Duero	A	Faeces
IBE-BC0715	42.92	-5.54	Duero	A	Faeces
IBE-BC0716	42.97	-5.42	Duero	B	Faeces
IBE-BC0720	43.06	-5.37	Duero	A	Faeces
IBE-BC0726	42.88	-5.46	Duero	A	Faeces
IBE-BC0750	43.00	-5.26	Duero	B	Faeces
IBE-BC0781	42.85	-5.68	Duero	A	Faeces
IBE-BC0874	42.94	-4.24	Duero	B	Faeces
IBE-BC0907	42.85	-6.22	Duero	A	Faeces
IBE-BC0908	42.74	-6.21	Miño	A	Faeces
IBE-BC0949	42.74	-6.19	Miño	A	Faeces
IBE-BC1027	42.93	-5.08	Duero	B	Faeces
IBE-BC1072	42.93	-5.12	Duero	B	Faeces
IBE-BC1330	40.49	-7.41	Mondego	A	Faeces
IBE-BC1339	40.38	-7.55	Tajo	A	Faeces
IBE-BC1382	40.53	-7.50	Mondego	A	Faeces
IBE-BC1399	41.08	-3.43	Tajo	A	Faeces
IBE-BC1405	40.47	-7.60	Mondego	A	Faeces
IBE-BC1414	41.12	-3.49	Tajo	B	Faeces
IBE-BC1416	40.53	-7.50	Mondego	A	Faeces
IBE-BC1417	41.22	-3.29	Tajo	B	Faeces

Table S1. Samples used in this study, including those also used in a previous study.

IBE-BC1432	41.10	-3.49	Tajo	A	Faeces
IBE-BC1436	41.17	-3.38	Tajo	B	Faeces
IBE-BC1442	40.33	-7.59	Tajo	A	Faeces
IBE-BC1480	40.35	-7.56	Tajo	A	Faeces
IBE-BC1507	40.44	-7.51	Mondego	A	Faeces
IBE-BC1543	40.36	-7.56	Tajo	A	Faeces
IBE-C1029	42.98	-4.75	Duero	B	Faeces
IBE-C1054	42.10	-2.90	Ebro	B	Faeces
IBE-C1061	42.13	-2.93	Ebro	B	Faeces
IBE-C1114	42.95	-4.76	Duero	B	Faeces
IBE-C1144	43.10	-6.65	Nalón	A	Igea et al. 2015
IBE-C1435	41.71	-4.99	Duero	A	Igea et al. 2015
IBE-C1510	40.34	-5.77	Tajo	A	Faeces
IBE-C1515	40.37	-5.75	Tajo	A	Faeces
IBE-C1516	42.09	-6.72	Duero	B	Faeces
IBE-C1518	40.49	-6.01	Tajo	A	Faeces
IBE-C1520	40.51	-6.13	Tajo	A	Faeces
IBE-C1662	41.97	-2.61	Duero	B	Igea et al. 2015
IBE-C1683	42.17	-6.66	Duero	A	Igea et al. 2015
IBE-C1750	40.34	-5.76	Tajo	A	Faeces
IBE-C1789	40.26	-5.52	Duero	A	Igea et al. 2015
IBE-C1856	43.21	-5.20	Sella	B	Faeces
IBE-C2520	43.10	-5.58	Nalón	A	Faeces
IBE-C2604	40.89	0.11	Ebro	A	Faeces
IBE-C2605	43.11	-5.23	Nalón	B	Faeces
IBE-C2664	40.66	-0.19	Ebro	A	Igea et al. 2015
IBE-C2843	40.28	-5.23	Duero	A	Faeces
IBE-C3230	41.67	-7.47	Duero	A	Faeces
IBE-C3237	43.17	-5.39	Nalón	B	Faeces
IBE-C3244	41.37	-7.79	Duero	A	Faeces
IBE-C3247	43.06	-5.56	Nalón	A	Faeces
IBE-C326	43.10	-5.01	Duero	B	Faeces
IBE-C3262	40.34	-5.13	Duero	A	Faeces
IBE-C3271	41.57	-7.68	Duero	A	Faeces
IBE-C3273	43.05	-5.28	Duero	B	Faeces
IBE-C331	43.19	-4.91	Deva	A	Faeces
IBE-C333	43.32	-4.86	Deva	B	Faeces
IBE-C3330	40.32	-5.42	Duero	A	Faeces
IBE-C3350	40.36	-5.27	Duero	A	Faeces
IBE-C3394	43.17	-4.66	Deva	B	Faeces
IBE-C3430	42.19	-2.93	Ebro	B	Faeces
IBE-C3456	42.22	-2.66	Ebro	B	Faeces
IBE-C3458	42.10	-2.71	Ebro	B	Faeces
IBE-C3463	42.13	-2.48	Ebro	B	Faeces
IBE-C3493	43.17	-6.14	Nalón	A	Faeces
IBE-C360	43.26	-4.76	Deva	B	Faeces

IBE-C3611	42.91	-5.40	Duero	B	Faeces
IBE-C3617	42.97	-5.56	Duero	A	Faeces
IBE-C3690	43.15	-5.96	Nalón	A	Faeces
IBE-C370	43.11	-4.74	Deva	B	Faeces
IBE-C3794	42.68	-8.28	Ulla	A	Faeces
IBE-C385	43.23	-5.03	Sella	B	Faeces
IBE-C3946	43.07	-5.02	Duero	A	Faeces
IBE-C3947	43.06	-5.09	Duero	B	Faeces
IBE-C4023	43.06	-3.73	Ebro	B	Faeces
IBE-C4042	43.13	-3.61	Ebro	B	Faeces
IBE-C4090	41.50	-3.78	Duero	A	Faeces
IBE-C4337	43.00	-2.84	Ebro	B	Faeces
IBE-C4339	43.11	-3.68	Ebro	B	Faeces
IBE-C4347	43.00	-2.84	Ebro	B	Faeces
IBE-C4487	42.16	-3.17	Duero	B	Faeces
IBE-C4499	42.10	-3.06	Duero	B	Faeces
IBE-C4505	42.01	-3.02	Duero	B	Faeces
IBE-C4511	42.07	-3.04	Ebro	B	Faeces
IBE-C4515	43.38	-6.36	Nalón	A	Faeces
IBE-C4618	41.04	-3.82	Duero	A	Faeces
IBE-C524	43.06	-6.24	Nalón	A	Faeces
IBE-C529	42.09	-6.71	Duero	A	Faeces
IBE-C544	43.10	-5.81	Nalón	A	Faeces
IBE-C568	42.99	-6.00	Duero	A	Faeces
IBE-C573	43.01	-6.23	Miño	A	Faeces
IBE-C597	42.98	-4.18	Ebro	B	Faeces
IBE-C630	43.26	-4.27	Saja	B	Faeces
IBE-C637	43.20	-4.30	Saja	B	Faeces
IBE-C688	43.14	-5.06	Sella	B	Faeces
IBE-C690	43.13	-5.04	Sella	A	Faeces
IBE-C692	43.26	-4.81	Deva	B	Faeces
IBE-C698	43.15	-4.91	Deva	A	Faeces
IBE-C740	43.25	-4.84	Deva	B	Faeces
IBE-C748	43.18	-5.00	Sella	A	Faeces
IBE-C761	43.03	-4.22	Ebro	B	Faeces
IBE-C785	42.75	-8.42	Ulla	A	Faeces
IBE-C825	41.59	-7.80	Duero	A	Faeces
IBE-C843	41.64	-7.80	Duero	A	Faeces
IBE-S1920	42.76	-2.59	Ebro	B	Faeces
VM77-31-B	42.20	-3.27	Duero	B	Faeces
WM04-SO61-B	41.97	-2.92	Duero	B	Faeces
WM04-SO63-B	41.99	-2.92	Duero	B	Faeces
WM05-01-A	42.08	-2.98	Ebro	B	Faeces
XX23-02-A	40.85	-3.95	Tajo	A	Faeces

Source of variation	d.f.	SS	MS	Est. Variance	%	P-value
Among basins	9	0.006316627	0.0007895784	5.1116e-05	13	***
Within basins	147	0.031430801	0.0003343702	3.3437e-04	87	***
Total	156	0.037747428	0.0003700728	0.000385486		

% is the percentage of variance explained by the isolation of river basins.

Table S2. Analysis of Molecular Variance (AMOVA) for the phylogenetic distances between Mediterranean water shrews inhabiting isolated river basins and within desmans inhabiting isolated river basins. (*** means significant p-value).

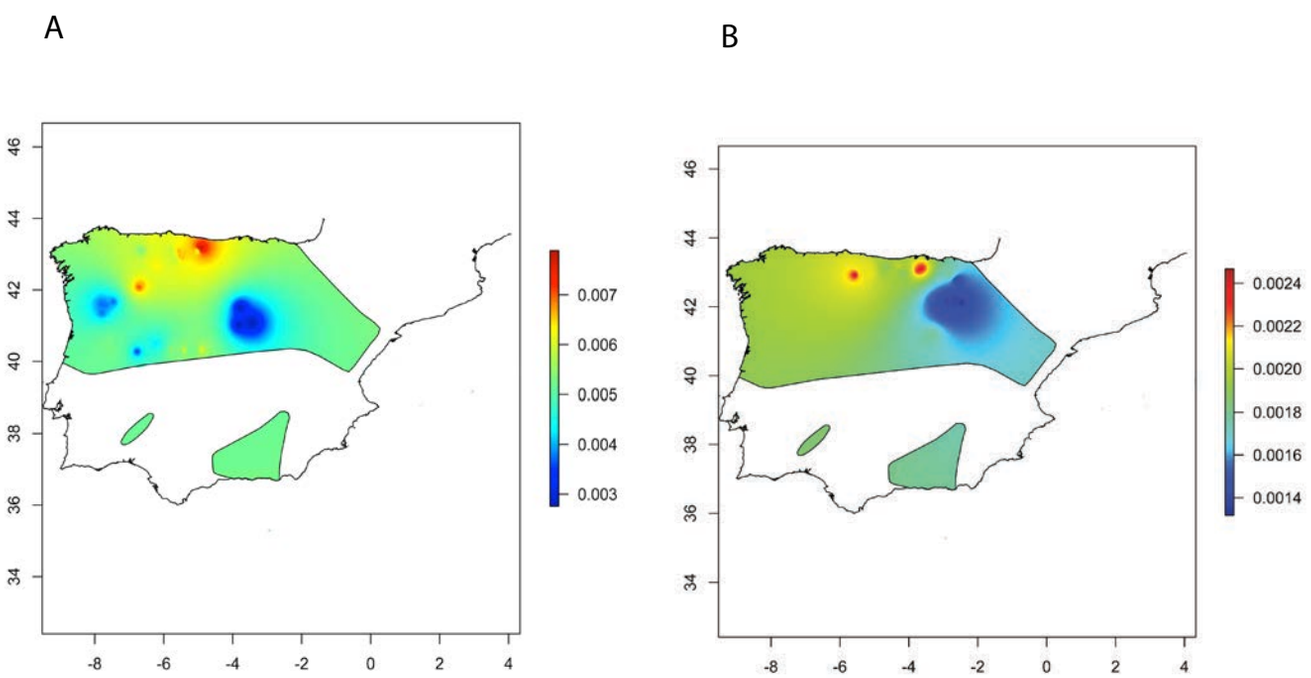


Figure S1. A) Interpolation of the nucleotide diversity for clade A. B) Interpolation of the nucleotide diversity for clade B.

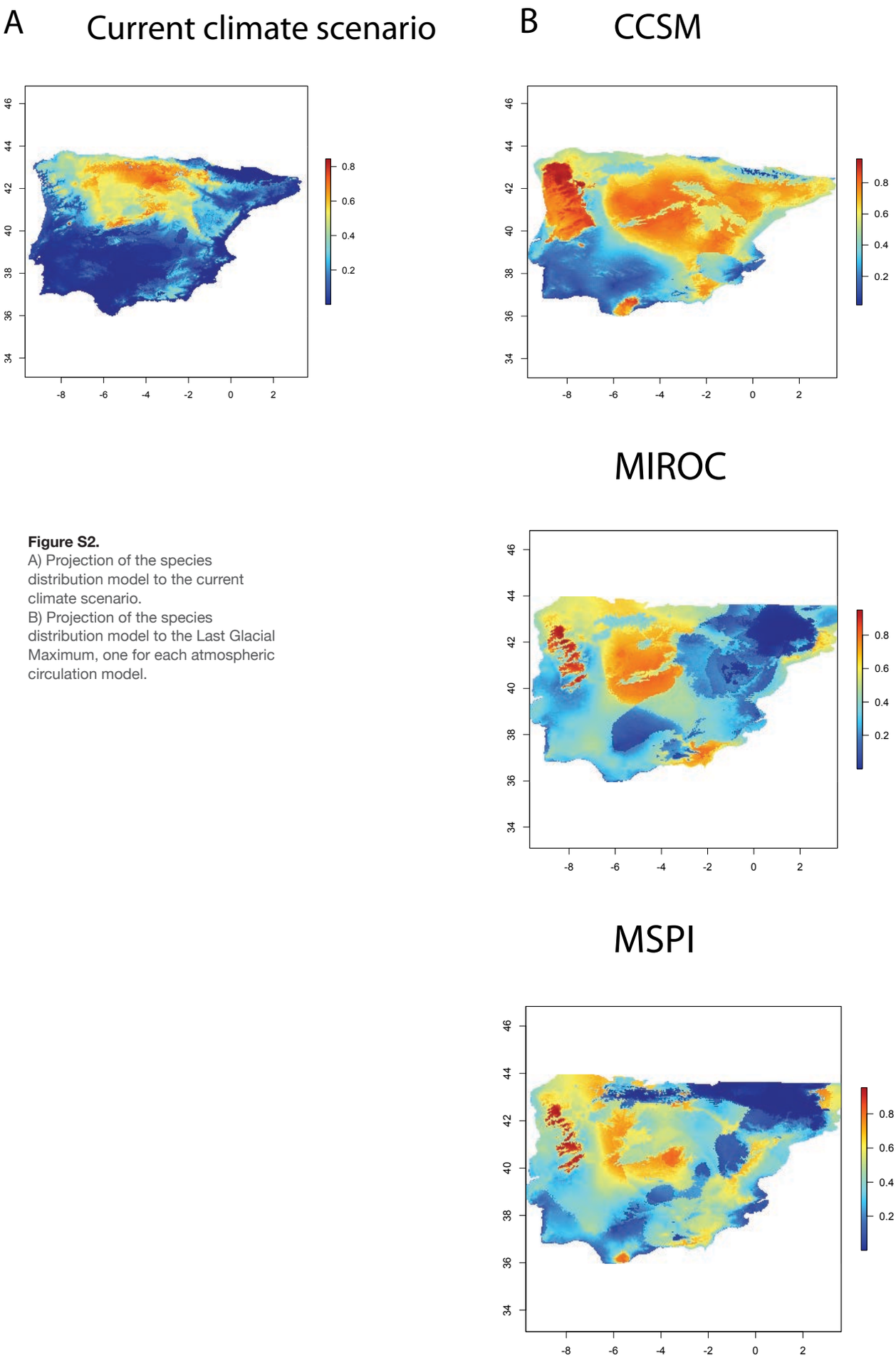


Figure S2.
A) Projection of the species distribution model to the current climate scenario.
B) Projection of the species distribution model to the Last Glacial Maximum, one for each atmospheric circulation model.

GENERAL DISCUSSION

Across the three chapters of this thesis, I have explored the evolutionary history of two Iberian endemic semi-aquatic mammals: the Pyrenean desman and the Mediterranean water shrew. Using different approaches, this work fills a gap in the knowledge of these two species. In the following section, I extract several generalities from the dissertation and contextualize them in a wider and up-to-date context, also addressing potential future research.

EVOLUTIONARY GENOMICS AS A CONSERVATION TOOL FOR ENDANGERED SPECIES

As stated before, the Pyrenean desman is endangered and its populations are known to have experienced a strong decline in the last few decades (Fernandes et al., 2011). Although the phylogeographic structure of this mammal was previously studied using mitochondrial data (Igea et al., 2013), more genetic markers were needed to understand its genetic structure. That is the reason why, in **Chapter 1**, we performed a population genomics study of this species with the objective of understanding the evolutionary history of this interesting mammal and helping to preserve it by quantifying levels of genetic variation and delimiting preliminary conservation units.

The estimation of individual heterozygosity rate is an indicator for genetic diversity. This parameter has been estimated in several studies regarding endangered species, such as the giant panda *(Ailuropoda melanoleuca)* (Li et al., 2010), the Amur tiger *(Panthera tigris altaica)* (Cho et al., 2013) , the Channel Island fox *(Urocyon littoralis)* (Robinson et al., 2016) or the Iberian lynx *(Lynx pardinus)* (Abascal et al., 2016). In **Chapter 1**, we estimated this rate for the Pyrenean desman, finding

extremely low levels of heterozygosity (a mean of 246 heterozygous position per million bases). These rates are similar to the ones for the Channel Island fox, a species with a smaller distribution range and population size than the Pyrenean desman, taking into account that species with small population sizes are known to present low genetic diversity levels (Frankham, 1996) and might indicate a risk of extinction (Reed & Frankham, 2003). Within the Iberian Peninsula, the lowest levels of the heterozygosity were found in the Pyrenees, in agreement with the low levels of mitochondrial genetic diversity found in previous studies (Igea et al., 2013; Gillet et al., 2014). These low levels were possibly due to postglacial colonization after a bottleneck and a long isolation period in the glacial refugia. With regard to conservation, the populations in this area are known to be declining and some of them remain in isolated valleys, increasing the risk of a local extinction. Thus, regular monitoring of the populations in the Pyrenees would be crucial as, due to the possible existence of inbreeding depression, these populations may show deterioration of fitness traits such as growth, survival and fecundity (Gall, 1987; Leberg & others, 1990).

The genomic population analyses of the Pyrenean desman allowed us to obtain a large number of variable markers to shed light on the geographical structure of this small mammal. The principal component analysis, the genomic tree and the Structure analysis agree in delimiting five different genetic clades correlated with the geographic structure, showing low levels of admixture. These clades or evolutionary pounits should be taken into account in conservation planning. The number of samples that could be studied was small, especially in the Central System and the Cantabrian Mountains, and therefore further studies will be necessary to corroborate these clades. The number of clades found in **Chapter 1** differ from the four clades revealed by the mitochondrial phylogeography of the Pyrenean desman (Igea et al., 2013) but the overall structure detected in both studies was similar. Although the Pyrenean desman is a riverine mammal, the geographical clades do not correspond with isolated river basins but rather with mountain ranges. This is consistent with the results of **Chapter 2**, which suggest that the isolated river basins are responsible for 53.8% of the genetic structure and, thus, the desmans were not completely isolated in the basins; consequently they must have dispersed overland from basin to basin during Holocene. Hence, the preservation and restoration of terrestrial corridors between the basins is as important as preserving the riverine ecosystems.

Genomic studies require high concentrations and high quantities of starting DNA. This is a problem for endangered species as usually the sampling is scarce for them. Thus, optimizing genomic protocols for low DNA starting quantities is necessary. In **Chapter 1**, aside from studying the population structure of the Pyrenean desman using genomic markers, we also modified and optimized an existing genomic protocol of library preparation by the ddRAD technique (Peterson et al., 2012), which can also be applied to other species.

We suggest that our findings regarding the genomic structure of the Pyrenean desman should be taken into account in future conservation programmes. Moreover, we believe that our efforts

may have opened a promising path to perform studies applying the same methodology to other endangered mammals.

DISPERSAL PATTERNS IN SEMI-AQUATIC MAMMALS

Semi-aquatic mammals live at the interface between aquatic and terrestrial environments and, from an ecological perspective, it is interesting to study their dispersal patterns, determining whether the river network acts as a barrier or, conversely, it connects populations. Also, how the river network affects these dispersal patterns and, consequently, the genetic structure, is interesting from evolutionary and ecological perspectives. In **Chapter 2**, we inferred dispersal patterns from mitochondrial genetic data, we studied the role of the river network in the genetic structure and the structure of genetic diversity in the north-western clade of the Pyrenean desman. The results of the study were unexpected, bearing in mind that the desman is known to inhabit rivers. In fact, both AMOVA and the isolation-by-distance approach agreed in that terrestrial dispersal was more important than previously believed. At least during humid periods of the Holocene, the desmans likely dispersed overland from basin to basin in order to colonize new territories from the glacial refugia (Igea et al., 2013). A previous study on the genetic structure of the Pyrenean desman also suggested a small influence of the river basins in the genetic structure for the whole distribution (Igea et al., 2013). These conclusions could mean that the Pyrenean desman is able to use terrestrial corridors to disperse. Thus, it would be advantageous for its conservation to restore and preserve terrestrial corridors within its distribution range.

In the comparison of the Pyrenean desman with other semi-aquatic mammals, it is interesting to remark the similarities found so far. For instance, in **Chapter 3**, the phylogeographic study of the Mediterranean water shrew showed that only a 13% of the genetic structure could be explained by the structuration of populations in river basins. In this case, the importance of the river network is even smaller than for the Pyrenean desman, meaning that they must have dispersed overland to colonize their current distribution range and that these shrews are not so closely dependent on the aquatic environments, as expected. The case of the Mediterranean water shrew is not the only one (Vignieri, 2005; Furlan et al., 2013; Byrne et al., 2015). For instance, in the platypus *(Ornithorhynchus anatinus)*, a semi-aquatic mammal with similar ecological requirements in Australia to the desman in the Iberian Peninsula, river basins play a role in shaping the genetic structure, but the isolation is not complete, so they also combine overland dispersal with river dispersal, depending on ecological factors (Furlan et al., 2013). There is another semi-aquatic mammal, the capybara *(Hydrochoerus hydrochaeris),* which is an herbivorous rodent that inhabits large regions of South America. The genetic structure of the capybara in some regions of Venezuela and Argentina is not affected by the structure of the river network, meaning that they also disperse overland in some regions (Byrne et al., 2015). In conclusion, semi-aquatic mammals are always influenced in some way by aquatic environments but their relative importance is not always as high as previously thought. In fact, most studies reveal a strong component of overland dispersal that could be due to many reasons,

especially the need to colonize new basins. However, there are still many semi-aquatic mammals for which the dispersal patterns have not been studied. This work opens a promising avenue of study by comparing and extracting generalities on dispersal behaviour as more semi-aquatic species are studied.

EVOLUTIONARY HISTORIES WITHIN THE IBERIAN PENINSULA

Intraspecific phylogeographies are essential in ecology and evolution as they reveal fine-scale patterns and provide crucial information for efficient conservation planning (Avise et al., 1987). Especially, the Iberian Peninsula is an interesting model to study evolutionary histories as it has a large amount of endemisms. This is because the wide range of climates and habitats that were present in the Iberian Peninsula during the Last Glacial Maximum, making this region optimal for species to take shelter during this glaciation (Gómez-Campo et al., 1984; Gómez & Lunt, 2007). The three chapters of this thesis deal with intraspecific patterns of Iberian species.

In **Chapter 3**, we present the first phylogeography of the Mediterranean water shrew since it was delimited as a species (Igea et al., 2015). As stated before, its genetic structure is divided into two main clades correlated with the geography, showing two contact zones, one in the Cantabrian Mountains and another one in the Central System, with the clades mainly allopatric and with low levels of spatial mixing. A main glacial refugium in the occidental region was inferred, according to the projection of the species distribution modelling to the Last Glacial Maximum and to the structure of genetic diversity. In the Pyrenean desman (Igea et al., 2013), there are also two main clades but in this case these clades could be subdivided into two subclades each, meaning that this species is more genetically structured than the Mediterranean water shrew. The contact zones coincide approximately with the ones described in the shrew, but in the desman, the mixing in the mitochondrial phylogeography is almost non-existent and the clades are more clearly allopatric. The Cantabrian contact zone of the Pyrenean desman is described in **Chapter 2**, showing that only two tributaries exhibit mixing of individuals of the two main clades. In the genomic approach of **Chapter 1** we delimited five genomic clades and described a slightly higher level of mixing thanks to the use of additional markers. Regarding the glacial refugia, four different ones were inferred in the mitochondrial study (Igea et al., 2013), but five were suggested in **Chapter 1** of this thesis. These discordances between mitochondrial and genomic inferences are not rare in wild populations (Springer et al., 2001; Toews & Brelsford, 2012). In any case, postglacial colonization routes are different for both Iberian mammals but both are consistent with the “refugia within refugia” hypothesis (Gómez & Lunt, 2007). The Iberian Peninsula is known to have harboured a large number of species during the Last Glacial Maximum together with other south European Peninsulas, such as the Italian Peninsula or the Balcans. These large refugia are known to have had many small refugia (Gómez-Campo et al., 1984; Gómez & Lunt, 2007). The cases of the Pyrenean desman and the Mediterranean water shrew are not isolated; there are several species whose glacial refugia

were inside the Iberian Peninsula, such as the European hare *(Lepus europeaus)* (Stamatis et al., 2009) or the beetle *Timarcha* (Gómez-Zurita et al., 2000).

The populations of Mediterranean water shrew are known to be decreasing (Hutterer et al., 2008). This, together, with the new species status suggests that further studies should be undertaken to specifically reassess its conservation status, as several populations could be under threat.

FUTURE RESEARCH

Although this thesis fills several gaps of knowledge regarding two Iberian endemisms, there is still much additional research that would be necessary. For instance, the development of genomic protocols for non-invasive samples would be advantageous, as these species are extremely difficult to live-trap, especially the Pyrenean desman. Moreover, a genomic analysis of dispersal patterns would be essential, as more markers could give information from a fine-scale perspective. Also, a species distribution modelling projected into the future could give insights into the future of the species and could give essential information for conservation programmes.

Regarding the Mediterranean water shrew, a genomic population analysis to delimit evolutionary units could be an interesting project. The inference of dispersal patterns and a comparative phylogeographic study with the Pyrenean desman would also be of great interest.

CONCLUSIONS

1. A genomic protocol (ddRAD) was modified and optimized for working with low amounts of starting material for the endangered Pyrenean desman. The quality of the genomic library was evaluated, revealing that this protocol produced reliable and consistent information that can potentially be applied to other case studies in non-model species. (**Chapter 1**)

2. The sex determination in the Pyrenean desman revealed equal numbers of males and females, indicating no sex bias in the sampling. (**Chapter 1**)

3. The individual heterozygosity rate was extremely low in the Pyrenean desman in comparison with other mammals. The lowest rate was located in the Pyrenees, being consistent with the lowest levels of genetic diversity in this area. (**Chapter 1**)

4. The genomic analyses performed, principal component analysis, genomic tree and Structure analysis, agreed with the existence of five geographically structured clades coinciding with main mountain ranges. These clades should in principle be considered as conservation units. (**Chapter 1**)

5. The delimitation of conservation units should be included in future conservation planning involving the Pyrenean desman. (**Chapter 1**)

6. The contact zone between clades A and B of the Pyrenean desman located in the Cantabrian Mountains has been clearly delimited, showing spatial mixing only in two tributaries, Curueño and Porma, despite no apparent physical barrier between these two clades. (**Chapter 2**)

7. Regarding the structure of genetic diversity in the northwestern area of the distribution range of the Pyrenean desman, there are spots of high diversity in the middle of the distribution of clade A1, in the Duero basin, coinciding with areas of potential distribution during the Last Glacial Maximum. This is consistent with the location of a glacial refugium in this area. (**Chapter 2**)

8. There are spots of low diversity in the in the Cantabrian Mountains and in the Ulla basin, which could be due to a recent colonization of these areas. (**Chapter 2**)

9. A total of 53.8% of the genetic structure of the clade A1 of the Pyrenean desman can be explained by the isolation in river basins. This means that desmans had to disperse overland between basins at some point during their evolutionary history. (**Chapter 2**)

10. Concerning the relative importance of the dispersal patterns for the whole clade A1 of the Pyrenean desman, it is suggested that they had an important component of overland dispersal during the Holocene. The same pattern was revealed for large basins (Duero basin). In contrast, for small basins and, probably, recent dispersal, important patterns of both overland and river dispersal were observed. (**Chapter 2**)

11. Preserving and restoring riverine ecosystems is of great importance for the Pyrenean desman but, due to the importance of overland dispersal, terrestrial corridors should also be an objective for preservation and restoration. (**Chapter 2**)

12. The genetic structure of the Mediterranean water shrew is divided into two mitochondrial clades, A and B, strongly correlated with geographic structure. The clades are mainly allopatric, with a small degree of mixing in the two contact zones detected in the Cantabrian Mountains and the Central System, respectively. (**Chapter 3**)

13. Only 13% of the genetic structure of the Mediterranean water shrew can be explained by the isolation of the shrews in river basins, suggesting overland dispersal from basin to basin and a weak relationship to riverine ecosystems. (**Chapter 3**)

14. The spots of high nucleotide diversity and the species distribution model suggest the existence of the main glacial refugium during the Last Glacial Maximum in the north-western area of the distribution range. (**Chapter 3**)

15. The new species status of the Mediterranean water shrew is further supported with these results and argues the need for further studies to reassess the conservation status of this species. (**Chapter 3**)

REFERENCES

Abascal F., Corvelo A., Cruz F., Villanueva-Cañas J.L., Vlasova A., Marcet-Houben M., Martínez-Cruz B., Cheng J.Y., Prieto P., Quesada V., & others (2016) Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology*, 17, 251.

Agrafiotis D.K., Lobanov V.S., & Salemme F.R. (2002) Combinatorial informatics in the post-genomics era. *Nature Reviews Drug Discovery,* 1, 337–346.

Avise J.C. (2000) *Phylogeography: the history and formation of species.* Harvard university press.

Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A., & Saunders N.C. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual review of ecology and systematics*, 489–522.

Baird N.A., Etter P.D., Atwood T.S., Currey M.C., Shiver A.L., Lewis Z.A., Selker E.U., Cresko W.A., & Johnson E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, 3, e3376.

Bartáková V., Reichard M., Blažek R., Polačik M., & Bryja J. (2015) Terrestrial fishes: rivers are barriers to gene flow in annual fishes from the African savanna. *Journal of Biogeography*, 42, 1832–1844.

Barton N.H. (2001) The evolutionary consequences of gene flow and local adaptation: future approaches. *Dispersal. Oxford University Press, New York*, 329–340.

Benito Garzón M., de Dios R., & Sáinz Ollero H. (2007) Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and Mid-Holocene. *Ecography*, 30, 120–134.

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

Bernet García-Santesmases C. (2016) Phylogenetic relationships, global biogeography and life-style diversification of talpids (Lipotyphla: Talpidae). (Master thesis)

Bonin A., Nicole F., Pompanon F., Miaud C., & Taberlet P. (2007) Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conservation Biology*, 21, 697–708.

Brewer S., Cheddadi R., De Beaulieu J.L., & Reille M. (2002) The spread of deciduous Quercus throughout Europe since the last glacial period. *Forest Ecology and Management*, 156, 27–48.

Brown J.H. & Kodric-Brown A. (1977) Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology*, 58, 445–449.

Brown W.M., George M., & Wilson A.C. (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, 76, 1967–1971.

Byrne M.S., Quintana R.D., Bolkovic M.L., Cassini M.H., & Túnez J.I. (2015) The role of river drainages in shaping the genetic structure of capybara populations. *Genetica*, 143, 645–656.

Cabrera A. (1907) XXXII.—Three new Spanish insectivores. *Journal of Natural History*, 20, 212–215.

Cabrera A. (1914) Fauna ibérica. Mamíferos. Junta para la ampliación de estudios e investigaciones científicas.

Cabria M.T., Rubines J., Gómez-Moliner B., & Zardoya R. (2006) On the phylogenetic position of a rare Iberian endemic mammal, the Pyrenean desman (*Galemys pyrenaicus*). *Gene*, 375, 1–13.

Carr S.M., Marshall H.D., Duggan A.T., Flynn S.M.C., Johnstone K.A., Pope A.M., & Wilkerson C.D. (2008) Phylogeographic genomics of mitochondrial DNA: highly-resolved patterns of intraspecific evolution and a multi-species, microarray-based DNA sequencing strategy for biodiversity studies. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 3, 1–11.

Carvalho J.C. & Gomes P. (2004) Feeding resource partitioning among four sympatric carnivores in the Peneda-Gerês National Park (Portugal). *Journal of zoology*, 263, 275–283.

Carvalho S.B., Brito J.C., Crespo E.J., & Possingham H.P. (2011) Incorporating evolutionary processes into conservation planning using species distribution data: a case study with the western Mediterranean herpetofauna. *Diversity and Distributions*, 17, 408–421.

Castién E. & Gosálbez J. (1992) Distribución geográfica y hábitats ocupados por *Galemys pyrenaicus* (Geoffroy, 1811 Insectivora: Talpidae) en los Pirineos Occidentales. *Doñana, Acta Vertebrata*, 19, 37–44.

Categories I.R.L. (2001) Criteria: Version 3.1. *IUCN Species Survival Commission, Gland, Switzerland.*

Chaput-Bardy A., Lemaire C., Picard D., & Secondi J. (2008) In-stream and overland dispersal across a river network influences gene flow in a freshwater insect, *Calopteryx splendens. Molecular Ecology*, 17, 3496–3505.

Cheddadi R., Vendramin G.G., Litt T., François L., Kageyama M., Lorentz S., Laurent J.-M., De Beaulieu J.-L., Sadori L., Jost A., & others (2006) Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography*, 15, 271–282.

Cho Y.S., Hu L., Hou H., Lee H., Xu J., Kwon S., Oh S., Kim H.-M., Jho S., Kim S., & others (2013) The tiger genome and comparative analysis with lion and snow leopard genomes. *Nature communications*, 4.

Clark P.U., Dyke A.S., Shakun J.D., Carlson A.E., Clark J., Wohlfarth B., Mitrovica J.X., Hostetler S.W., & McCabe A.M. (2009) The last glacial maximum. *Science*, 325, 710–714.

Cohen M.P. (1988) Aldo Leopold: His Life and Work. By Kurt Meine. University of Wisconsin Press.

Consuegra S., de Leániz C., Serdio A., González Morales M., Straus L.G., Knox D., & Verspoor E. (2002) Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular Ecology*, 11, 2037–2048.

Cordellier M. & Pfenninger M. (2009) Inferring the past to predict the future: climate modelling predictions and phylogeography for the freshwater gastropod *Radix balthica* (Pulmonata, Basommatophora). *Molecular Ecology*, 18, 534–544.

Crandall K.A., Bininda-Emonds O.R.P., Mace G.M., & Wayne R.K. (2000) Considering evolutionary processes in conservation biology. *Trends in ecology & evolution*, 15, 290–295.

Davey J.W., Hohenlohe P.A., Etter P.D., Boone J.Q., Catchen J.M., & Blaxter M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, 499–510.

Dubey S., Salamin N., Ohdachi S.D., Barrière P., & Vogel P. (2007) Molecular phylogenetics of shrews (Mammalia: Soricidae) reveal timing of transcontinental colonizations. *Molecular phylogenetics and evolution*, 44, 126–137.

Dudgeon D., Arthington A.H., Gessner M.O., Kawabata Z.-I., Knowler D.J., Lévêque C., Naiman R.J., Prieur-Richard A.-H., Soto D., Stiassny M.L.J., & others (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews*, 81, 163–182.

Elith J. & Leathwick J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677.

Emerson K.J., Merz C.R., Catchen J.M., Hohenlohe P.A., Cresko W.A., Bradshaw W.E., & Holzapfel C.M. (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the national academy of sciences*, 107, 16196–16200.

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

Feliner G.N. (2011) Southern European glacial refugia: a tale of tales. *Taxon*, 60, 365–372.

Fernandes M., Herrero J., Aulagnier S., & Amori G. (2011) *Galemys pyrenaicus*. *IUCN Red List of Threatened Species Version*, 20112, 1–3.

Fortelius M. (2015) New and Old Worlds database of fossil mammals (NOW).

Frankham R. (1996) Relationship of genetic variation to population size in wildlife. *Conservation Biology*, 10, 1500–1508.

Frankham R., Briscoe D.A., & Ballou J.D. (2002) *Introduction to conservation genetics.* Cambridge university press.

Friant M. (1947) Les musaraignes (Soricidae) quaternaires et actuelles de Europe occidental. Characteres dentaires et osteologiques, base de leur classification: Soc. Geol. *Nord Annales*, 67, 222–269.

Furlan E.M., Griffiths J., Gust N., Handasyde K.A., Grant T.R., Gruber B., & Weeks A.R. (2013) Dispersal patterns and population structuring among platypuses, *Ornithorhynchus anatinus*, throughout south-eastern Australia. *Conservation Genetics*, 14, 837–853.

Gall G.A.E. (1987) Inbreeding. *Population genetics and fishery management*, 47–87.

Gavin D.G., Fitzpatrick M.C., Gugger P.F., Heath K.D., Rodríguez-Sánchez F., Dobrowski S.Z., Hampe A., Hu F.S., Ashcroft M.B., Bartlein P.J., & others (2014) Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, 204, 37–54.

Gillet F., Garrido C., Teresa M., Blanc F., Némoz M., Fournier-Chambrillon C., Sourp E., Vial-Novella C., Zardoya R., Aulagnier S., & others (2017) Evidence of a fine-scale genetic structure for the endangered Pyrenean desman (*Galemys pyrenaicus*) in the French Pyrenees. *J Mammal,* 98 (2): 523-532.

Gillet F., Tiouchichine M.-L., Galan M., Blanc F., Némoz M., Aulagnier S., & Michaux J.R. (2015) A new method to identify the endangered Pyrenean desman (*Galemys pyrenaicus*) and to study its diet, using next generation sequencing from faeces. *Mammalian Biology-Zeitschrift für Säugetierkunde*, 80, 505–509.

Glenn T.C. (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology resources*, 11, 759–769.

Gómez A. & Lunt D.H. (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia* pp. 155–188.

Gómez-Campo C., Bermudez-de-Castro L., Cagiga M.J., & Sánchez-Yélamo M.D. (1984) Endemism in the Iberian peninsula and Balearic Islands. *Webbia*, 38, 709–714.

Gómez-Zurita J., Petitpierre E., & Juan C. (2000) Nested cladistic analysis, phylogeography and speciation in the *Timarcha goettingensis* complex (Coleoptera, Chrysomelidae). *Molecular Ecology*, 9, 557–570.

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

González-Esteban J., Castién E., & Gosálbez J. (1999) Morphological and colour variation in the Pyrenean desman *Galemys pyrenaicus* (Geoffroy, 1811). *Zeitschrift fur Saugetierkunde*, 64, 1–11.

González-Esteban J., Villate I., Castién E., Rey I., & Gosálbez J. (2002) Age determination of *Galemys pyrenaicus*. *Acta theriologica*, 47, 107–112.

Gordillo S., Coronato A.M.J., & Rabassa J.O. (2005) Quaternary molluscan faunas from the island of Tierra del Fuego after the Last Glacial Maximum. *Scientia Marina*, 69, 337–348.

Graham C.H., Ron S.R., Santos J.C., Schneider C.J., & Moritz C. (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, 58, 1781–1793.

Groom M.J. (2006) Threats to biodiversity. *Principles of conservation biology*, 3, 63–109.

Gugerli F., Englisch T., Niklfeld H., Tribsch A., Mirek Z., Ronikier M., Zimmermann N.E., Holderegger R., Taberlet P., Consortium I., & others (2008) Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation--a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics*, 10, 259–281.

Guisan A. & Zimmermann N.E. (2000) Predictive habitat distribution models in ecology. *Ecological modelling*, 135, 147–186.

Hare M.P. (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution*, 16, 700–706.

Harvey A.L. (2008) Natural products in drug discovery. *Drug discovery today*, 13, 894–901.

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

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

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

Hewitt G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 359, 183–195.

Hoarau G., Coyer J.A., Veldsink J.H., Stam W.T., & Olsen J.L. (2007) Glacial refugia and recolonization pathways in the brown seaweed Fucus serratus. *Molecular Ecology*, 16, 3606–3616.

Hutchison C.A., Newbold J.E., Potter S.S., & Edgell M.H. (1974) Maternal inheritance of mammalian mitochondrial DNA. *Nature*, 251, 536–538.

Hutterer R., Amori G., Kryštufek B., Yigit N., Mitsain G., Meinig H., Bertolino S., & Palomo L.J. (2008) *Neomys anomalus*.

Igea J., Aymerich P., Bannikova A.A., Gosálbez J., & Castresana J. (2015) Multilocus species trees and species delimitation in a temporal context: application to the water shrews of the genus *Neomys. BMC evolutionary biology*, 15, 1.

Igea J., Aymerich P., Fernández-González A., González-Esteban J., Gómez A., Alonso R., Gosálbez J., & Castresana J. (2013) Phylogeography and postglacial expansion of the endangered semi-aquatic mammal *Galemys pyrenaicus*. *BMC evolutionary biology*, 13, 115.

IUCN S.S.C. (2001) IUCN red list categories and criteria: version 3.1. *Prepared by the IUCN Species Survival Commission*.

Jaksic F.M., Seib R.L., & Herrera C.M. (1982) Predation by the barn owl *(Tyto alba)* in Mediterranean habitats of Chile, Spain and California: A comparative approach. *American Midland Naturalist*, 107, 151-162.

Jones M.E., Jarman P.J., Lees C.M., Hesterman H., Hamede R.K., Mooney N.J., Mann D., Pukk C.E., Bergfeld J., & McCallum H. (2007) Conservation management of Tasmanian devils in the context of an emerging, extinction-threatening disease: devil facial tumor disease. *EcoHealth*, 4, 326–337.

Juckwer E.A. (1990) *Galemys pyrenaicus* (Geoffroy, 1811)--Pyrenäen-Desman. *Handbuch der Säugetiere Europas: Insektenfresser, Herrentiere*, 79–92.

Kozak K.H., Graham C.H., & Wiens J.J. (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution*, 23, 141–148.

Lave L.B. & Seskin E. (1973) Air pollution and human health. *Readings in Biology and Man*, 169, 294.

Leakey R.E. (1995) *The sixth extinction: patterns of life and the future of humankind*. Doubleday Books.

Leberg P.L. & others (1990) Influence of genetic variability on population growth: implications for conservation. *Journal of Fish Biology*, 37, 193–195.

Li R., Fan W., Tian G., Zhu H., He L., Cai J., Huang Q., Cai Q., Li B., Bai Y., & others (2010) The sequence and de novo assembly of the giant panda genome. *Nature*, 463, 311–317.

López-Fuster M.J., García-Perea R., Fernández-Salvador R., Gisbert J., & Ventura J. (2006) Craniometric variability of the Iberian desman, *Galemys pyrenaicus* (Mammalia: Erinaceomorpha: Talpidae). *Folia Zoologica*, 55, 29.

Macdonald D.W. & Johnson D.D.P. (2001) 25. Dispersal in theory and practice: consequences for conservation biology.

Macneale K.H., Peckarsky B.L., & Likens G.E. (2005) Stable isotopes identify dispersal patterns of stonefly populations living along stream corridors. *Freshwater Biology*, 50, 1117–1130.

Mamanova L., Coffey A.J., Scott C.E., Kozarewa I., Turner E.H., Kumar A., Howard E., Shendure J., & Turner D.J. (2010) Target-enrichment strategies for next-generation sequencing. *Nature methods,* 7, 111–118.

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

Mardis E.R. (2008) Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.,* 9, 387–402.

Mardis E.R. (2013) Next-generation sequencing platforms. *Annual review of analytical chemistry*, 6, 287–303.

Marioni J.C., Mason C.E., Mane S.M., Stephens M., & Gilad Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research*, 18, 1509–1517.

Martínez-Solano I., Teixeira J., Buckley D., & García-París M. (2006) Mitochondrial DNA

phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple refugia in an Iberian endemic. *Molecular Ecology*, 15, 3375–3388.

Mayer-Scholl A., Draeger A., Göllner C., Scholz H.C., & Nöckler K. (2010) Advancement of a multiplex PCR for the differentiation of all currently described Brucella species. *Journal of microbiological methods*, 80, 112–114.

McCormack J.E., Hird S.M., Zellmer A.J., Carstens B.C., & Brumfield R.T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, 66, 526–538.

McCormack J.E., Maley J.M., Hird S.M., Derryberry E.P., Graves G.R., & Brumfield R.T. (2012) Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Molecular Phylogenetics and Evolution*, 62, 397–406.

Médail F. & Diadema K. (2009) Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, 36, 1333–1345.

Melero Y., Aymerich P., Luque-Larena J.J., & Gosàlbez J. (2012) New insights into social and space use behaviour of the endangered Pyrenean desman (*Galemys pyrenaicus*). *European journal of wildlife research*, 58, 185–193.

Melero Y., Aymerich P., Santulli G., & Gosàlbez J. (2014) Activity and space patterns of Pyrenean desman (*Galemys pyrenaicus*) suggest non-aggressive and non-territorial behaviour. *European journal of wildlife research*, 60, 707–715.

Miller G.S. (1912) *Catalogue of the Mammals of Western Europe: (Europe Exclusive of Russia) in the Collection of the British Museum*. order of the Trustees.

Mottaz C. (1907) *Préliminaires à nos études de micromammalogie: Description du Neomys milleri, sp. nova.*

Nielsen R., Paul J.S., Albrechtsen A., & Song Y.S. (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, 12, 443–451.

Nores C., Queiroz A.I., & Gisbert J. (2007) *Galemys pyrenaicus* (E. Geoffroy Saint-Hilaire, 1811). *Atlas y libro rojo de los mamíferos terrestres de España*, 92.

Oliver T.A., Garfield D.A., Manier M.K., Haygood R., Wray G.A., & Palumbi S.R. (2010) Whole-genome positive selection and habitat-driven evolution in a shallow and a deep-sea urchin. *Genome biology and evolution*, 2, 800–814.

Pagacz S. (2016) The effect of a major drainage divide on the gene flow of a semiaquatic carnivore, the Eurasian otter. *Journal of Mammalogy*, 97, 1164–1176.

Palmeirim J.M. & Hoffmann R.S. (1983) *Galemys pyrenaicus*. *Mammalian Species*, 1–5.

Paz-Vinas I., Loot G., Stevens V.M., & Blanchet S. (2015) Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology*, 24, 4586–4604.

Pearman P.B., Guisan A., Broennimann O., & Randin C.F. (2008) Niche dynamics in space and time. *Trends in Ecology & Evolution*, 23, 149–158.

Pen I. & Weissing F.J. (2000) Towards a unified theory of cooperative breeding: the role of ecology and life history re-examined. *Proceedings of the Royal Society of London B: Biological Sciences*, 267, 2411–2418.

Peterson B.K., Weber J.N., Kay E.H., Fisher H.S., & Hoekstra H.E. (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, 7, e37135.

Petit R.J., Aguinagalde I., de Beaulieu J.-L., Bittkau C., Brewer S., Cheddadi R., Ennos R., Fineschi S., Grivet D., Lascoux M., & others (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, 300, 1563–1565.

Pinho C., Harris D.J., & Ferrand N. (2007) Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex: the case of Iberian and North African wall lizards *(Podarcis, Lacertidae). Biological Journal of the Linnean Society*, 91, 121–133.

Poduschka W. & Richard B. (1985) Hair types in the fur of the Pyrenean desman (*Galemys pyrenaicus*) Geoffroy, 1811 (Insectivora: Talpidae: Desmaninae). *Oesterr. Akad. Wiss. Math-Naturwiss. Kl. Sitzungsber*, 1–5.

Provan J. & Bennett K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in ecology & evolution*, 23, 564–571.

Raeymaekers J.A.M., Maes G.E., Geldof S., Hontis I., Nackaerts K., & Volckaert F.A.M. (2008) Modeling genetic connectivity in sticklebacks as a guideline for river restoration. *Evolutionary applications*, 1, 475–488.

Reed D.H. & Frankham R. (2003) Correlation between fitness and genetic diversity. *Conservation biology*, 17, 230–237.

Repenning C.A. (1967) Subfamilies and genera of the Soricidae.

Reumer J.W.F. (1998) A classification of the fossil and recent shrews. *Evolution of shrews*, 5–22.

Richard P.R. (1976) Détermination de l'âge et de la longévité chez le Desman des Pyrénées (*Galemys pyrenaicus*). *Société nationale de protection de la nature et d'acclimatation de France, Paris (FRA)*

Richards C.L., Carstens B.C., & Lacey Knowles L. (2007) Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, 34, 1833–1845.

Robinson J.A., Ortega-Del Vecchyo D., Fan Z., Kim B.Y., Marsden C.D., Lohmueller K.E., Wayne R.K., & others (2016) Genomic flatlining in the endangered island fox. *Current Biology*, 26, 1183–1189.

Ronce O. (2007) How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annu. Rev. Ecol. Evol. Syst.*, 38, 231–253.

Roze D. & Rousset F. (2003) Selection and drift in subdivided populations: a straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. *Genetics*, 165, 2153–2166.

Rümke C.G. (1985) *A review of fossil and recent Desmaninae (Talpidae, Insectivora)*. Utrecht University.

Rychlik L. (1997) Differences in foraging behaviour between water shrews: *Neomys anomalus and Neomys fodiens. Acta Theriologica*, 42, 351–386.

Rzebik-Kowalska B. (1998) Fossil history of shrews in Europe. *Evolution of shrews*, 23–92.

Schreuder A. (1940) A revision of the fossil water-moles (Desmaninae). *Archives néerlandaises de zoologie*, 4, 201–333.

Schwartz M.K., Luikart G., & Waples R.S. (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, 22, 25–33.

Senn H., Ogden R., Frosch C., Syr\uučková A., Campbell-Palmer R., Munclinger P., Durka W., Kraus R.H.S., Saveljev A.P., Nowak C., & others (2014) Nuclear and mitochondrial genetic structure in the Eurasian beaver (*Castor fiber*)--implications for future reintroductions. *Evolutionary applications,* 7, 645–662.

Simberloff D. (1988) The contribution of population and community biology to conservation science. *Annual review of ecology and systematics*, 19, 473–511.

Simpson G.G. & others (1945) The principles of classification and a classification of mammals. *Bull. Amer. Museum Nat. History*, 85.

Sims D., Sudbery I., Ilott N.E., Heger A., & Ponting C.P. (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15, 121–132.

Springer M.S., DeBry R.W., Douady C., Amrine H.M., Madsen O., de Jong W.W., & Stanhope M.J. (2001) Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Molecular Biology and Evolution,* 18, 132–143.

Stamatis C., Suchentrunk F., Moutou K.A., Giacometti M., Haerer G., Djan M., Vapa L., Vukovic M., Tvrtković N., Sert H., & others (2009) Phylogeography of the brown hare *(Lepus europaeus)* in Europe: a legacy of south-eastern Mediterranean refugia? *Journal of Biogeography*, 36, 515–528.

Stone R.D. (1986) *The social ecology of the European mole, Talpa europaea L. and the Pyrenean desman, Galemys pyrenaicus G.: a comparative study.* University of Aberdeen.

Stone R.D. (1987a) The activity patterns of the Pyrenean desman (*Galemys pyrenaicus*) (Insectivora: Talpidae), as determined under natural conditions. *Journal of Zoology,* 213, 95–106.

Stone R.D. (1987b) The social ecology of the Pyrenean desman (*Galemys pyrenaicus*) (Insectivora: Talpidae), as revealed by radiotelemetry. *Journal of Zoology*, 212, 117–129.

Stone R.D. & Gorman M.L. (1985) Social organization of the European mole (Talpa europaea) and the Pyrenean desman *(Galemys pyrenaicus). Mammal review,* 15, 35–42.

Storch G., Qiu Z., & Zazhigin V.S. (1998) Fossil history of shrews in Asia. *Evolution of shrews*, 93–120.

Tian B., Liu R., Wang L., Qiu Q., Chen K., & Liu J. (2009) Phylogeographic analyses suggest that a deciduous species *(Ostryopsis davidiana Decne., Betulaceae)* survived in northern China during the Last Glacial Maximum. *Journal of Biogeography,* 36, 2148–2155.

Toews D.P.L. & Brelsford A. (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21, 3907–3930.

Turmelle A.S., Kunz T.H., & Sorenson M.D. (2011) A tale of two genomes: contrasting patterns of phylogeographic structure in a widely distributed bat. *Molecular Ecology*, 20, 357–375.

Valqui J., Hartl G.B., & Zachos F.E. (2010) Non-invasive genetic analysis reveals high levels of mtDNA variability in the endangered South-American marine otter *(Lontra felina). Conservation Genetics,* 11, 2067–2072.

Ventura J. (2007) *Neomys anomalus* Cabrera, 1907. *Atlas y libro rojo de los mamíferos terrestres de España. Madrid, Spain: Dirección General para la Biodiversidad-SECEMSECEMU,* 114–116.

Vignieri S.N. (2005) Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse *(Zapus trinotatus). Molecular Ecology*, 14, 1925–1937.

Wielstra B., Crnobrnja-Isailović J., Litvinchuk S.N., Reijnen B.T., Skidmore A.K., Sotiropoulos K., Toxopeus A.G., Tzankov N., Vukov T., & Arntzen J.W. (2013) Tracing glacial refugia of *Triturus* newts based on mitochondrial DNA phylogeography and species distribution modeling. *Frontiers in zoology*, 10, 13.

Willerslev E., Cappellini E., Boomsma W., Nielsen R., Hebsgaard M.B., Brand T.B., Hofreiter M., Bunce M., Poinar H.N., Dahl-Jensen D., & others (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, 317, 111–114.

Wilson E.O. (1989) Threats to biodiversity. *Scientific American*, 261, 108–116.

Wright S. (1943) Isolation by distance. *Genetics*, 28, 114.

Wright S. (1969) The theory of gene frequencies. Vol. 2. *Evolution and the Genetics of Populations,* 294–324.

Yokoyama Y., Lambeck K., De Deckker P., Johnston P., & Fifield L.K. (2000) Timing of the Last Glacial Maximum from observed sea-level minima. *Nature*, 406, 713–716.

Ecology and Evolution

Preview

From: ecoevo@wiley.com

To: marina.querejeta@ibe.upf-csic.es, angelfg@biosfera.es, rromerosuances@gmail.com, jose.castresana@csic.es

CC: allen.moore@ecolevol.org

Subject: Ecology and Evolution - Decision on Manuscript ID ECE-2016-12-01294.R1 [email ref: DL-RW-1-a]

Body: 10-Apr-2017

Dear Dr. Castresana:

It is a pleasure to accept your manuscript entitled "Postglacial dispersal patterns and mitochondrial genetic structure of the Pyrenean desman (Galemys pyrenaicus) in the north-western region of the Iberian Peninsula" in its current form for publication in Ecology and Evolution.

We appreciate your efforts and attention to detail raised by the referees and associate editor. You have treated them with the care and attention we expect.

Thank you for your fine contribution. On behalf of the Editors of Ecology and Evolution, we look forward to your continued contributions to the Journal. Enjoy the benefits of an OA paper.

Sincerely,
Prof. Allen Moore
Editor in Chief, Ecology and Evolution
allen.moore@ecolevol.org

--Instruction Summary—
Please note that although your manuscript has been accepted, the files will now be checked to ensure that everything is ready for publication and you may be contacted if final versions of the files are required. You will also receive an email from the editorial office confirming the next steps towards publication of your manuscript, which includes information about the Article Publication Fee. Once your article has been received by the Production Team, you will then receive an invitation from Wiley's Author Services system, which will ask you to log in and sign the appropriate license.

Associate Editor Comments to Author:

Associate Editor
Comments to the Author:
The original submission of the manuscript was in very could condition. The revision is an improvement. The authors did an excellent job of addressing all of the reviewers' concerns. I have not issue recommending this amended manuscript for publication.

Reviewer(s)' Comments to Author:

Date Sent: 10-Apr-2017