The evolutionary history and conservation of an endemic and threatened Iberian rodent: the Cabrera vole (*Microtus cabrerae*)

Soraia Isabel Moreira Barbosa

Doutoramento em Biodiversidade, Genética e Evolução
Departamento de Biologia
2017

Orientador
Paulo Célio Alves, PhD, Professor Associado
CIBIO/InBIO, Laboratório Associado, Universidade do Porto

Coorientador
Jeremy B Searle, PhD, Professor
Department of Ecology and Evolutionary Biology, Cornell University

"…the study of life becomes a hollow, rarefied pursuit if the very animals and plants that fired our imaginations as children and triggered our curiosity as students should perish."

Michael E. Soulé and Bruce A. Wilcox

Foreword

In compliance with the no. 2 of article 4 of the General Regulation of Third Cycles of the University of Porto and with article 31 of the Decree-Law no. 74/2006, of 24 March, with the alteration introduced by the Decree-Law no. 230/2009, of 14 September, the results of already published works were totally used and included in some of the chapters of this dissertation. As there works were performed in collaboration with other authors, the candidate clarifies that, in all these works, participated in obtaining, interpreting, analysing and discussing the results, as well as in the writing of the published forms.

This research was supported by Programa Operacional Potencial Humano – Quadro de Referência Estratégico National funds from the European Social fund and the Portuguese Ministério de Educação e Ciência through a PhD fellowship (Fundação para a Ciência e a Tecnologia, FCT - SFRH/BD/77726/2011).

This thesis should be cited as:
Barbosa S (2017) The evolutionary history and conservation of an endemic and threatened Iberian rodent: the Cabrera vole (*Microtus cabrerae*). PhD Thesis, University of Porto, Porto, Portugal.

Agradecimentos / Acknowledgments

Em primeiro lugar quero agradecer aos meus orientadores e instituições associadas, CIBIO/InBIO e Cornell University, por todo o apoio que me foi dado no decorrer desta tese.

Ao meu orientador principal, Paulo Célio Alves* quero agradecer a oportunidade de desenvolver este projeto e toda a educação científica que me proporcionou. Eu faço exatamente o que gosto e espero continuar a fazê-lo toda a minha vida, sendo que foi com o Paulo Célio que aprendi não só a importância da conservação, mas também o quão gratificante é trabalhar neste tema. Acima de tudo, com o Paulo Célio aprendi a importância de um orientador se preocupar com o bem estar dos seus estudantes, sendo essa talvez a lição mais importante que irei tirar da sua orientação.

To my co-supervisor, Jeremy Searle*, I would like to thank having accepted me as his student and the amazing work hosting me in Ithaca. Jeremy has an incredible crew of people working with him, which reflects his great scientific abilities, but I think mostly his kind-heartedness. With Jeremy I learned how to be extremely critical of my own work, organized and punctual (he's British), and I will definitely hold on to these great assets for my future academic career. I really appreciate the care and friendship you provided me with during these last five years.

**these authors contributed equally to my academic development*

Ao grupo dos *Small Mammals* (*ConGen*): tenho tanto a agradecer à Joana Paupério que nem sem muito bem por onde começar. A Joana esteve envolvida em todos os projetos em que participei, desde o mestrado até ao fim do doutoramento. Ela faz malabarismo com o impossível: trabalha com 327 pessoas em 85 projetos (números ao calhas, mas certamente perto da realidade), plus she's *super nice*! Com a Joana aprendi que é possível equilibrar a vida profissional com a vida pessoal. Só é pena ela agora ter saltado do *small* barco para trabalhar na água (ahem… EcoMetaEnvApplGen).

Por isso, Clara (Lechucha) Ferreira, cai-te em cima o peso dos small mammals (que não deve ser muito porque afinal de contas são pequenos). À Clara tenho a agradecer os grandes momentos nas nossas odisseias de armadilhagem e a colaboração em muito do trabalho de laboratório desta tese, mais os abraços aleatórios. Com a Clara aprendi a ser uma pessoa mais altruísta e otimista, mesmo face ao pior (Lirón Careto)!

Estas pessoas foram as que contribuíram diretamente para esta tese, no entanto há muitas outras às quais também quero agradecer, porque de uma forma ou de outra, contribuíram para o meu desenvolvimento profissional e pessoal.

No CTM encontram-se um conjunto extraordinário de pessoas que contribuem para que muito do nosso trabalho seja possível e por isso quero deixar desde já expresso a importância que estas pessoas tiveram em me ajudarem a chegar onde estou hoje.

At Cornell, I must thank the Searle Lab for all your friendship and welcoming, all the discussions, and for making me wanting to be a better scientist with your dedication to all things evolution and knowledge on the most random wildlife things (so random)!

Do LNIV e das churrascadas (não devia escrever isto, pois não?), da favela ao aquário, nunca houve um minuto aborrecido e agradeço a todos pelos momentos de chorar a rir. Nos pequenos-almoços com a Guida, a Fabiana e a Teresa partilhamos pequenas derrotas e grandes vitórias, e com a Freitas e o Guilherme, discutimos e desabafamos a vida (são uns filósofos…).

Aos meus pais, tenho a agradecer… bem tudo! Mas relativamente a esta tese, tenho que agradecer à minha mãe pelos excelentes momentos a fazer trabalho de campo, onde aprendi que consigo ser… qual a palavra… ah, insuportável!, e ao meu pai pelo apoio logístico (bandeirinhas fluorescentes) e aparvalhanço que ajudaram em muito durante todo o meu percurso cá e lá. É especialmente gratificante para mim ver que vocês se prestaram a olhar para a natureza pelos meus olhos, e ver que os ratos também podem ser especiais! Aos dois, sei que não deve ser fácil ver o cuco a sair do ninho, mas para onde quer que o futuro me leve, nunca estarei verdadeiramente longe porque *home is where your heart is*!

Last but not least, quero agradecer ao gajo da minha vida por ter aceitado entrar nesta aventura comigo, apesar de não termos noção do grande sacrifício pessoal que seria. Não me vou estender muito porque já sabes: *to be continued*…

A todos, MUITO OBRIGADA!

Summary

Species and populations across the globe are currently subjected to dramatic climatic changes and rapid habitat loss due to anthropogenic pressures. As a consequence, their ability to persist is limited by their ability to track suitable habitat and to adapt to new conditions. For species or populations to evolve and adapt to new conditions, whether they respond *in situ* or not, genetic variability is necessary for reducing the extinction probability. Species that have undergone extensive population declines are particularly likely to have insufficient genetic variability and many may already be threatened with extinction: as populations become smaller, genetic drift becomes higher, inbreeding increases and genetic diversity decreases, leading to reduced fitness and adaptive potential. Thus, to conserve species in the face of current climate and habitat changes, especially for those which are already threatened, it is important to gain deeper understanding of the species evolutionary history, evaluate the amount of current genetic diversity and assess how it is spatially distributed.

This thesis focuses on a near-threatened rodent endemic to the Iberian Peninsula, the Cabrera vole (*Microtus cabrerae*). This species is specialised on fast disappearing damp grassy areas, which are often subject to drainage and conversion as part of agricultural land use change. Thus the current main threat for the Cabrera vole is human induced habitat loss, which is leading to a decrease in the numbers of individuals, promoting high levels of inbreeding and subsequent decreased fitness.

To increase sampling opportunities and avoid disturbing susceptible species, this thesis first describes the development of non-invasive methodologies that allow genetic data to be obtained without requiring capture and handling of animals. First, we developed a barcoding protocol that confirms the identity of tissue samples, but also that identifies non-invasive samples from the field, both faeces and bones from owl (*Tyto alba*) pellets. Due to the low quantity and quality of the DNA found in such samples, amplification of nuclear fragments using traditional Sanger sequencing or microsatellite genotyping is very low (< 30%), and thus more efficient and reliable methodologies were needed for more detailed studies.

We used both non-invasive and traditional tissue samples for different parts of this study. Tissue samples from a variety of vole species were used to examine one of the most critical and controversial aspects of Cabrera vole biology: its phylogenetic position. The Cabrera vole is attributed to the subgenus *Iberomys*, of which it is the only living representative. We used mitogenomes and genome wide SNPs and show that the

Cabrera vole is closely related to the subgenus *Agricola*, where it was previously placed, and for which the field vole (*M. agrestis*) is currently the only representative. The phylogeny that we obtained suggests that these two species (or subgenera) represent one of the first expansions of *Microtus* that extended to Europe from an origin in Asia, at an earlier time than the expansions and radiations of the *Microtus* and *Terricola* subgenera. The claim that *M. cabrerae* should be elevated to its own genus based on the uniqueness of some karyotypic, morphologic and life-history traits was refuted given that the species has an internal position to the *Microtus* phylogeny.

The fossil record shows the Cabrera vole's first appearance around 100 kya, and that its distribution fluctuated with the climatic oscillations of the Quaternary. Our phylogeographic study using mitochondrial and single nuclear loci data estimates the time to the most recent common ancestor around 17 kya, suggesting that the last major contraction of extant populations occurred during the Last Glacial Maximum (LGM), possibly to the south-east of the Iberian Peninsula. After the LGM, the refugial population expanded to other areas of Iberia, but were soon interrupted by the subsequent impact of the Younger Dryas (YD) cold spell. This event represented another time of contraction for the Cabrera vole, with the species apparently restricted to at least two separate locations in western and eastern Iberia. We believe that this geographic pattern is the basis of the divergence of the two main mitochondrial lineages found today, named *west* and *east*, as well as the main population subdivision observed with the SNP dataset. The fossil record supports this hypothesis by revealing an increased presence of the Cabrera vole in peripheral areas of the Iberian Peninsula, both to the west and east. With the beginning of the Holocene, we suggest that there was a rapid expansion of the YD refugial populations to most areas of the Iberian Peninsula. Currently, however, the populations have increasingly limited ranges and low genetic diversity, likely associated to habitat loss. The western populations are genetically more homogeneous than those found in the east, which likely got separated and gradually diverged during the last major expansion of the Cabrera vole, possibly due to the higher topographic complexity of eastern Iberia. The Cabrera vole's current population is thus divided into four main geographic nuclei: in the west it is represented by the *Lusocarpetan* geographic nucleus, composed of all populations from Portugal and the Spanish Central System mountain range, and in the east by the *pre-Pyrenean* (pre-Pyrenean massifs), *Montiberic* (Iberian System mountain range) and *Betic* (Betic System mountain range) geographic nuclei. Our phylogeographic analysis further identified a secondary contact between the western and eastern groups, which appears to represent a demographic expansion rather than selection based on mitochondrial sequence, given that signals of population

expansion were also obtained with the single nuclear loci. This secondary contact precedes the Cabrera vole decline, *c.* 3 kya, also observed in the fossil record, although genetic exchange between the western and eastern groups may have still occurred for a long period of time since the beginning of the decline.

Aiming at developing a conservation framework, we used the SNPs dataset to determine conservation units (*CUs*) for the Cabrera vole and their functional connectivity. By using different sets of SNPs we identified three evolutionarily significant units (*ESUs*, using all loci), five management units (*MUs*, using neutral loci) and two adaptive units (*AUs*, using outlier loci). *ESU1* (*Lusocarpetan*), is the most widespread *CU* and has high levels of connectivity between most of its populations, especially in the south; *ESU2* (*Montiberic* and *Betic*) has the highest levels of genetic diversity, but very low connectivity between most of its populations, especially in the *Betic* geographic nucleus; and *ESU3* (*pre-Pyrenean*) is the most isolated and least diverse, but interestingly the most divergent of all three, being the one that contributes the most to the Cabrera vole's overall genetic diversity. Only *ESU2* is subdivided into three *MUs*, possibly as a result of lack of gene flow between populations, and subsequent divergence of allele frequencies. The *MUs* in the southernmost regions of *ESU2* (especially in the *Betic* region), are very distinct genetically and could have characteristics important for adaptation to future environmental conditions. However as habitat is increasingly lost in these regions, these populations will likely disappear due to high levels of inbreeding and the inability to exchange with other populations. As there is mounting evidence that the Cabrera vole's distribution is overestimated in the eastern part of the species range, new assessments are needed to ensure that the remaining populations are able to persist, either by the protection of suitable habitat or by providing dispersal corridors to more suitable areas.

Given the importance of monitoring species for conservation, our final aim was to use a restriction enzyme-based genotyping-by-sequencing (GBS) method to detect SNPs from faecal and bone DNA samples. Although we are still not able to apply HTS to individual non-invasive samples, with this preliminary study we were able to successfully genotype over 3 000 SNPs on faecal and bone DNA pools and perform population assignment, as well as determine various genetic parameters on Cabrera vole non-invasive samples. This study shows that non-invasive HTS is possible for non-model organisms, without the need of genomes or transcriptomes and expensive probe design, at a relatively low cost and with reduced laboratory work.

Resumo

As espécies e populações de todo o mundo estão atualmente sujeitas a alterações climáticas dramáticas e rápida perda de habitat devido a pressões antropogénicas. Como consequência, a capacidade destes organismos de persistir é limitada pela sua capacidade de seguir o seu habitat adequado e de se adaptar a novas condições. Para as espécies e populações evoluírem e se adaptarem a novas condições, quer respondam *in situ* ou não, é necessária variabilidade genética para reduzir a probabilidade de extinção. Espécies que sofreram declínios populacionais extremos são particularmente suscetíveis de terem variabilidade genética insuficiente e muitas estarem já ameaçadas de extinção: à medida que as populações diminuem, a deriva genética torna-se mais forte, aumenta a endogamia e a diversidade genética diminui, diminuindo o fitness e o potencial adaptativo. Assim, para conservar as espécies dadas as atuais alterações climáticas e de habitats, especialmente para aquelas espécies que já estão ameaçadas, é importante obter um entendimento mais detalhado da história evolutiva da espécie, avaliar a quantidade de diversidade genética atual e determinar a sua distribuição espacial.

Esta tese baseia-se num roedor quase-ameaçado e endémico da Península Ibérica, o rato de Cabrera *(Microtus cabrerae).* Esta espécie é especializada em habitats húmidos em rápido desaparecimento, que são muitas vezes sujeitos a drenagem e conversão para uso agrícola. Assim, as principais ameaças atuais para o rato de Cabrera são a perda de habitat induzida pelo homem, que está a levar a um acentuado decréscimo populacional, elevados níveis de endogamia e à diminuição da aptidão adaptativa.

Para aumentar o número de amostras e evitar perturbar uma espécie já suscetível de si, esta tese descreve, em primeiro lugar, o desenvolvimento de metodologias não-invasivas para obter dados genéticos que não requerem a captura e manipulação de animais. Assim, foi desenvolvido um protocolo de barcoding que confirma a identidade de amostras de tecido, mas também que identifica as amostras não-invasivas recolhidas no campo, tanto fezes como ossos obtidos de egagrópilas de coruja das torres *(Tyto alba)*. Devido à baixa quantidade e qualidade do DNA encontrado nestas amostras, a amplificação de fragmentos nucleares usando a sequenciação tradicional de Sanger ou a genotipagem de microssatélites é muito baixa (<30%) e, portanto, metodologias mais eficientes e confiáveis foram necessárias de forma a podermos realizar estudos mais detalhados.

Utilizamos amostras não-invasivas e de tecido para diferentes partes desta tese. Amostras de tecido de várias espécies de *Microtus* foram usadas para examinar um dos aspetos mais críticos e controversos da biologia do rato de Cabrera: a sua posição filogenética. O rato de Cabrera é atribuído ao subgénero *Iberomys,* do qual é o único representante atual. Utilizou-se mitogenomas e SNPs distribuídos pelo genoma para mostrar que o rato de Cabrera está intimamente relacionado com o subgénero *Agricola,* onde estava colocado previamente e para o qual o rato do campo de cauda curta *(M. agrestis*) é atualmente o único representante. A filogenia que obtivemos sugere que estas duas espécies (ou subgéneros) representam uma das primeiras expansões dos *Microtus* para a Europa a partir de uma origem na Ásia, numa altura anterior às expansões e radiações dos subgéneros *Microtus* e *Terricola.* A sugestão de que *M. cabrerae* deve ser elevado ao seu próprio género baseada na singularidade de alguns traços cariotípicos, morfológicos e da sua história de vida foi refutada, dado que a espécie tem uma posição interna na filogenia do género *Microtus.*

O registro fóssil denota o aparecimento do primeiro rato de Cabrera há cerca de 100 mil anos, sendo que a sua distribuição foi contraindo e expandindo com as oscilações climáticas do Quaternário. O nosso estudo filogeográfico com base em dados mitocondriais e nucleares estima que o tempo para o ancestral comum mais recente das populações atuais é de cerca de 17 mil anos, sugerindo que a maior contração das populações atuais terá ocorrido durante o Último Máximo Glacial. Depois do LGM, esta população refugial expandiu-se para outras áreas da península, mas foi rapidamente interrompida pelo subsequente impacto das temperaturas baixas do Dryas recente. Este evento representou um novo momento de contração para o rato de Cabrera, com a espécie aparentemente restrita a pelo menos dois locais distintos no oeste e este da Península Ibérica. Acreditamos que este evento tenha resultado na divergência das duas linhagens mitocondriais principais encontradas atualmente, chamadas *oeste* e *este,* bem como na principal subdivisão populacional observada com os dados genómicos. O registro fóssil apoia esta hipótese ao revelar uma maior presença do rato de Cabrera nas áreas periféricas da Península Ibérica, incluindo Portugal e este da Espanha. Com o início do Holoceno, este trabalho sugere que houve uma rápida expansão dessas populações refugiais da periferia para áreas mais centrais da Península Ibérica. No entanto, atualmente as populações tem distribuições cada vez mais restritas e baixa diversidade genética, provavelmente associado a perda do habitat. As populações ocidentais são geneticamente mais homogéneas do que as do este, que acabaram por gradualmente se diferenciar durante a última grande expansão do rato de Cabrera, possivelmente devido à maior complexidade topográfica do este da Península

Ibérica. A população atual do rato de Cabrera está assim dividida em quatro principais núcleos geográficos: no oeste é representado pelo núcleo geográfico *Lusocarpetano,* composta por todas as populações de Portugal e espanholas do Sistema Central e a este pelos núcleos *pré-Pirenaico* (nos maciços pré-Pirenaicos), *Montibérico* (no Sistema Ibérico) e *Bético* (no Sistema Bético). A nossa análise filogeográfica identificou ainda um contacto secundário destes dois grupos ocidental e oriental, em que a linhagem mitocondrial *oeste*, que parece representar uma expansão demográfica ao invés de uma seleção sobre os haplótipos mitocondriais da linhagem *oeste*, dado que os sinais de expansão da população foram também observados a nível dos genes nucleares estudados. Este contacto secundário precede o declínio do rato de Cabrera, há cerca de 3 mil anos, também suportado pelo registo fóssil, apesar de a troca de genes entre os grupos do oeste e este se possa ter mantido por um longo período após o início do declínio.

Com o objetivo de desenvolver um plano de conservação, utilizamos o conjunto de dados genómicos (SNPs) para definir unidades de conservação *(UCs)* para o rato de Cabrera, assim como a sua conectividade funcional. Usando diferentes conjuntos de SNPs foram identificados três unidades evolutivamente significativas *(ESUs,* usando todos os loci), cinco unidades de gestão *(MUs,* usando loci neutrais) e duas unidades adaptativas *(AUS,* usando loci outlier). A *ESU1 (*núcleo *Lusocarpetano)* é a *UC* mais extensa geograficamente, possuindo níveis de conectividade elevada entre a maioria de suas populações, especialmente no sul; a *ESU2 (*núcleos *Montibérico* e *Bético)* tem níveis mais altos de diversidade genética, mas muito baixa conectividade entre a maioria das suas populações, especialmente no núcleo geográfica *Bético;* a *ESU3 (*núcleo *pré-Pirenaico)* é a mais isolada e menos diversa, mas curiosamente a mais divergente de todos as três *ESUs*, sendo ainda a que mais contribui para a diversidade genética global do rato de Cabrera. Apenas a *ESU2* é subdividida em três *MUs,* possivelmente como resultado da falta de fluxo génico entre as suas populações, e subsequente divergência das respetivas frequências alélicas. As *MUs* periféricas da *ESU2* (especialmente no núcleo *Bético*) são muito distintas geneticamente e podem ter características importantes para a adaptação a futuras condições ambientais mais adversas. No entanto, como o habitat favorável à ocorrência do rato de Cabrera está cada vez mais ameaçado nessas regiões, estas populações provavelmente desaparecerão devido aos altos níveis de endogamia e à incapacidade de fluxo génico com outras populações. Com o crescente reconhecimento de que a presença do rato de Cabrera é sobrestimada na parte oriental da sua distribuição, são necessárias novas avaliações para assegurar que as populações remanescentes são capazes de persistir, quer pela proteção do

habitat favorável quer pela proteção de corredores de dispersão para áreas mais adequadas.

Dada a importância de realizar monitorização das espécies para a sua proteção, o nosso objetivo final foi utilizar um método de genotipagem por sequenciação (GBS) baseado em enzimas de restrição para detetar SNPs de amostras de DNA fecal e ósseo. Embora ainda não possamos aplicar HTS a amostras individuais não-invasivas, com este estudo preliminar fomos capazes de genotipar mais de 3 000 SNPs em pools de DNA fecal e ósseo e realizar a alocação às populações de origem, bem como determinar vários parâmetros genéticos em amostras não invasivas de rato de Cabrera; adicionalmente muitas outras aplicações são possíveis com dados de Pool-seq. Este estudo mostra que o HTS não-invasivo é possível para organismos não-modelo, sem a necessidade de genomas ou transcriptomas e desenho de sondas, a um custo relativamente baixo e com trabalho de laboratório reduzido.

Keywords

- Conservation genetics
- Cabrera vole (*Microtus cabrerae*)
- Non-invasive genetic sampling
- Phylogenomics
- Phylogeography
- Landscape genetics
- Conservation units

Palavras-chave

- Genética da conservação
- Rato de Cabrera (*Microtus cabrerae*)
- Amostragem genética não-invasiva
- Filogenómica
- Filogeografia
- Genética da paisagem
- Unidades de conservação

Table of Contents

Index of Tables and Figures

Abbreviations

12S: 12S ribosomal RNA gene

16S: 16S ribosomal RNA gene

aDNA: ancient DNA

ATP6/8: adenosine triphosphate subunit 6/8 gene

AU: adaptive unit

bp: base pairs

BP: before present

BRCA1: breast cancer gene 1

Bt: Betic

c.: circa

COI/II/III: cytochrome c oxidase subunit I/II/III gene

COP9: constitutive photomorphogenesis 9 signalosome gene

COPS7A: *COP9* subunit 7A gene

CU: conservation unit

cyt-*b*: cytochrome-*b*

DBX5: developing brain homeobox, intron 5

DBX7: developing brain homeobox, intron 7

D-loop: displacement loop

DNA: deoxyribonucliec acid

ENM: ecological niche modelling

ESU: evolutionarily significant unit

GBS: genotyping-by-sequencing

HTS: high throughput sequencing

IRBP: interphotoreceptor retinoid-binding protein gene

IUCN: International Union for Conservation of Nature

kya: thousands of years ago

Lc: *Lusocarpetan*

LGM: last glacial maximum

Mb: *Montiberic*

mtDNA: mitochondrial deoxyribonucliec acid

MU: management unit

Mya: millions of years ago

NADH: nicotinamide adenine dinucleotide coenzyme reduced form

ND(number): NADH-ubiquinone oxidoreductase chain (number) gene

NGS: next generation sequencing

NiGS: non-invasive genetic sampling

OSTA: organic solute transporter subunit alpha

PCR: polymerase chain reaction

PNPO: pyridoxamine 5'-phosphate oxidase gene

pP: *pre-Pyrenean*

RAD-seq: restriction site-associated DNA sequencing

RAG1: recombination activating gene 1

RNA: ribonucliec acid

RNA-seq: ribonucleic acid sequencing

rRNA: ribosomal RNA

RRS: reduced representation sequencing

SLC38A7: solute carrier family 38 member 7

SMCY7: Y-linked structural maintenance of chromosomes gene

SNP: single nucleotide polymorphism

tMRCA: time to the most recent common ancestor

tRNA: transfer RNA

WGS: whole genome sequencing

YD: younger dryas

CHAPTER I.

GENERAL INTRODUCTION

1. Conservation Biology – *a crisis discipline*

As early as the 19th century, there was already disquiet about the growing human impact on the natural world (Sodhi & Ehrlich 2010). With increasing environmental pollution, habitat disturbance and consequent depletion of biodiversity, an appreciation of the importance (and complexity) of biotic interactions and a concern about their future grew through the 20th century. This concern evolved into a new mind-set that aimed to preserve the 'biota as a whole', not only to sustain utilitarian resources but, most importantly, for maintaining the ecosystem's health (Leopold 1991; Golley 1993; Sodhi & Ehrlich 2010). From this mind-set was born the field of Conservation Biology – a discipline that uses the knowledge and tools from all biological disciplines for nature conservation (Soulé & Wilcox 1980). It was defined by Soulé (1985) as "*the application of science to conservation problems, addressing the biology of species, communities and ecosystems that are perturbed, either directly or indirectly, by human activities or other agents"*. Conservation biology is thus in the intersection of science, policy and practice, that aims for taxon survival but starts with the threat of their extinction (Frankel & Soulé 1981; Soulé 1985; Sodhi & Ehrlich 2010; Haig *et al.* 2016).

The *threatened*, the *elusive* and the *neglected*

Over the last hundred years, the rate of species loss faced by the Earth's biota entered a sixth 'mass extinction' with an average rate of vertebrate species loss over 100 times higher than pre-human background rates (Figure 1) (Mills 2013; Ceballos *et al.* 2015). However, the proportion of species currently described for each group influences the reliability of such estimates. Given that the knowledge we have on each group is a result of both opportunity and interest in describing such species, there are particularly well documented groups with detailed extinction records, in comparison with other groups that are less studied but have higher number of species; for instance vertebrates *vs* invertebrates, or mammals and birds *vs* reptiles and fish (Baillie *et al.* 2004). Threatened species are found all over the world, particularly in biodiversity hotspots, but also in areas where there are higher human densities and where centuries of human activities, either direct (e.g. persecution) or indirect (e.g. habitat destruction), have led to extensive range contractions and extinctions (Myers *et al.* 2000; Ceballos & Ehrlich 2002).

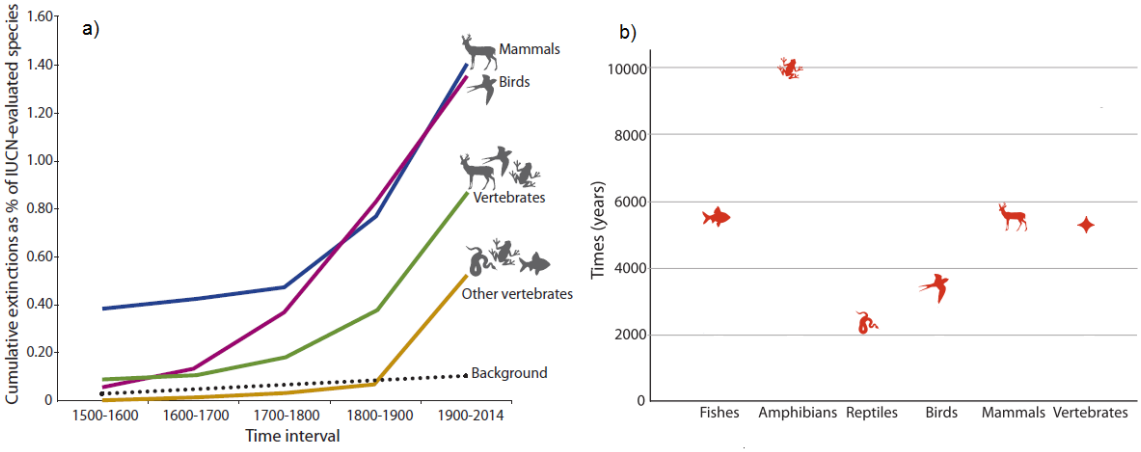


Figure 1. Observed and predicted species extinction levels: a) cumulative percentage of species extinctions during the last 500 years using a 'highly conservative' [i.e. considering only 'extinct' (EX) species] scenario for 'mammals' and 'birds' (100% of those described), 'other vertebrates' [reptiles (44%), amphibians (88%) and fishes (38%)], as well as all vertebrates combined (59%) – dashed black line represents a constant background extinction rate of 2 mammal species extinctions per 10 000 species per 100 years; b) number of years required for each group to become extinct under the background rate following a 'highly conservative' scenario [adapted from Ceballos *et al.* (2015), with permission].

Conserving biodiversity in these disturbed areas can be challenging given that favourable habitats are usually small and geographically dispersed as a consequence of human occupation; thus alternative conservation strategies in these areas are needed, such as those relying on the concept of 'umbrella species' (Lambeck 1997). Umbrella species are usually charismatic vertebrates, frequently birds and mammals, and their protection leads to the conservation of large amounts of natural habitat that include many other naturally co-occurring, and possibly threatened, species (Roberge & Angelstam 2004). Although the 'umbrella species' approach accounts for the conservation of the target and non-target species and their habitat, there are cases where particular species have higher diversity and abundance outside rather than inside protected areas. A particular case has been illustrated by Caro (2003), who showed that abundances of small mammals inside protected black rhino habitat in East Africa were lower than outside the reserve, exemplifying that different strategies may have different success rates for different groups. The taxon bias is recognised and well-illustrated in a recent study from Pérez-Espona (2017) that analysed conservation genetics studies performed on various vertebrate groups, molluscs and vascular plants from 1992 to 2014 in Europe. This study shows that some groups have a lower ratio of number of species to total number of conservation genetic studies than others, even within the same class, especially biased towards charismatic and commercially valuable species. Within mammals, most conservation genetic studies target carnivores and ungulates, which relates directly to the umbrella species concept. Bats (Chiroptera), shrews and their

relatives (Soricomorpha), and rodents (Rodentia), are mammalian orders with a small proportion of studies given the large number of extant species, despite their importance for ecosystems and frequent need for conservation.

To determine the threat status of species and their populations, the IUCN Red List Categories and Criteria (IUCN 2001) relies mostly on estimates of population size decline and geographic range reduction, disregarding that, in addition to ecosystems and species, genes are a form of biodiversity that also deserves conservation (McNeely *et al.* 1990). A recent study from Willoughby *et al.* (2015) showed that the IUCN criteria typically used to assign conservation rank tend to overlook species with reduced genetic diversity, unless their populations have undergone drastic census size decreases. Thus, although the IUCN has recently incorporated genetic diversity in the threat assessments, it still only concerns the problems of introducing genetic material from invasive and other problematic species into endangered populations (IUCN 2016). The ability of these endangered populations to adapt to increasing habitat and climatic change relates to a species' genetic diversity (Eizaguirre & Baltazar-Soares 2014), and so we cannot afford to be indifferent about conservation genetics.

2. Conservation Genetics to the rescue?

Darwin already discussed the importance of morphological diversity for maintaining healthy populations on his essays about domestication, demonstrating that humans have long been managing domestic populations based on underlying practical notions of genetic diversity and inbreeding, even before a knowledge of the mechanics of genetics (Darwin 1857). The basic principle of natural selection requires phenotypic variation to favour one form over the other and this, in essence, is a result of the expression of genetic diversity (Ghalambor *et al.* 2007; Valladares *et al.* 2014). The field of conservation genetics started with Otto Frankel's concern both for the genetic diversity of domestic species and for wildlife conservation (Frankel & Bennet 1970; Frankel 1974). Since genes are the basis for evolutionary change and adaptation, genetic variation is a crucial determinant of long term population viability; thus the more genetic diversity a population has, the higher the likelihood of it adapting to changes in the environment (Frankel & Soulé 1981; Loeschcke *et al.* 1994; Allendorf *et al.* 2012; Mills 2013; Eizaguirre & Baltazar-Soares 2014; Willoughby *et al.* 2015). Various works have highlighted the importance of genetics in maintaining populations resilient to change, however the most difficult and intrinsic part in conservation genetics is detecting

population decline before populations suffer from genetic problems, which represent some of the most important factors driving extinction (Spielman *et al.* 2004).

From declining to small populations, and then towards extinction

Two different paradigms in conservation biology have been considered by Caughley (1994). The 'declining-population paradigm' focuses on determining the causes for population decline. It is the most useful paradigm for conservation but the hardest to draw theoretical expectations from due to the great variety of factors likely involved, often requiring case-by-case analysis (Figure 2). On the other hand, the 'small-population paradigm' has a much more simple theoretical background as it deals mostly with population genetic (internal) factors and associated stochastic (natural) events, which can be more or less predicted or at least accounted for (Figure 2). The expectation is that smaller populations will have increased risk of extinction due to decrease in genetic diversity and thus in their fitness (Frankel & Soulé 1981). This correlation between fitness and population size illustrates the 'Allee effect' (Stephens *et al.* 1999). A population can be influenced by many types of Allee effects, depending on the species life history, current threats and their strength, etc. (Berec *et al.* 2007). From an ecological perspective, lower population densities are often associated with lower fitness through the impact on reproductive (e.g. mate finding) and survival (e.g. cooperative anti-predator behaviour) mechanisms, and this correlation is termed the *ecological* Allee effect (Stephens *et al.* 1999; Berec *et al.* 2007). From a genetic perspective, as populations become smaller, the loss of genetic diversity and increased genetic drift, will often result in inbreeding (and ultimately inbreeding depression) and in the accumulation of deleterious mutations (or mutational load) – the *genetic* Allee effect (Hedrick & Garcia-Dorado 2016; Luque *et al.* 2016). Thus, smaller populations will have decreased fitness and lower chances of survival when exposed to new environmental stressors, and often negative population growth rates – a phenomenon called the 'extinction vortex' (Gilpin & Soulé 1986; Amos & Balmford 2001; Charlesworth & Willis 2009; Ellegren & Galtier 2016). An excellent example of the interaction between small population size, inbreeding and relatedness on individual fitness is the study of the Isle Royale wolves (Hedrick *et al.* 2017). Wolves colonised Lake Superior's Isle Royale in the late 1940s but became isolated over the years with no new migrants arriving on the island (Peterson & Page 1988). Over time, the population grew to its maximum of around 50 individuals in the 80s, but then sharply declined to the two individuals found today (Peterson & Page 1988; Hedrick *et al.* 2017).

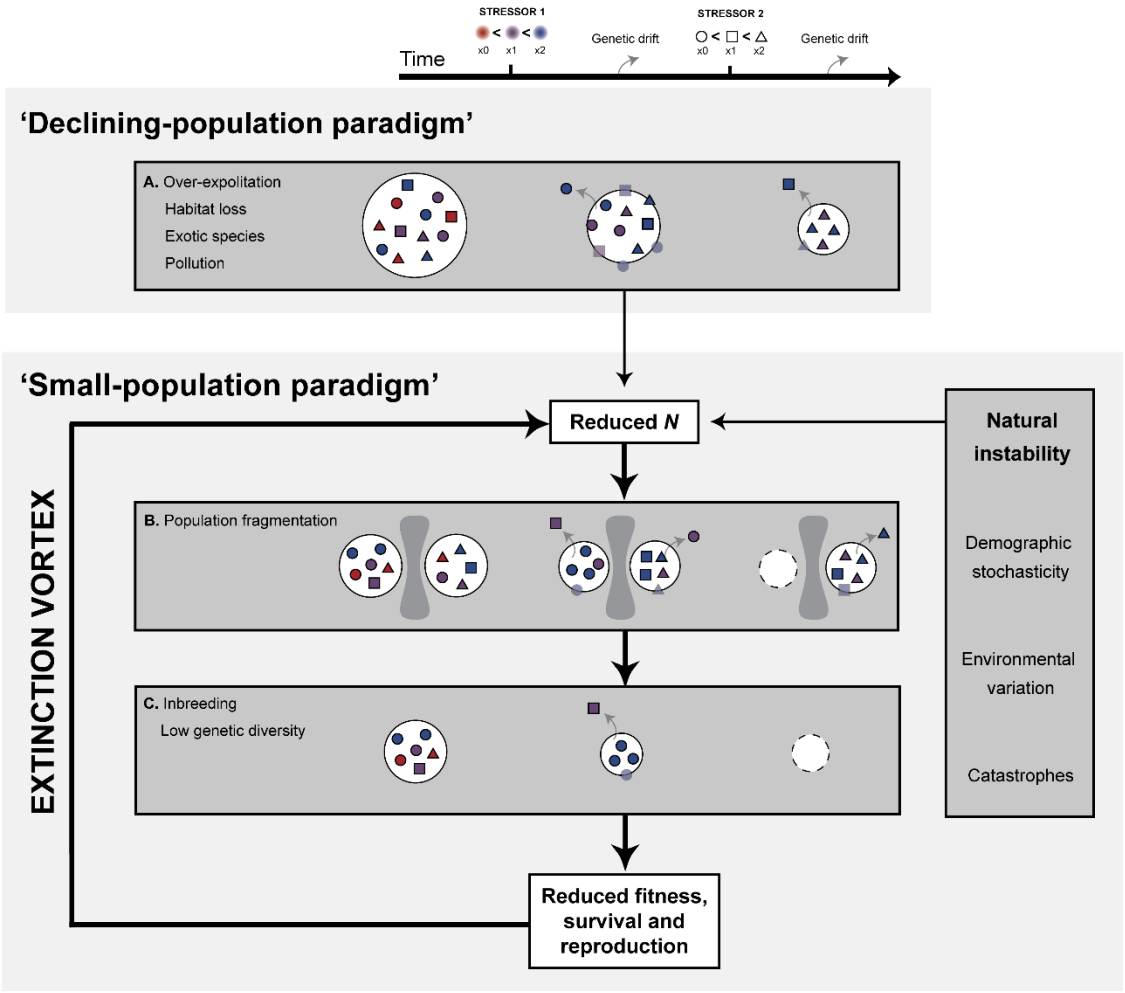


Figure 2. Schematic representation of the 'declining-population' and 'small-population' paradigms proposed by Caughley (1994), and the extinction vortex as presented by Frankham *et al.* (2002), by comparing the effects of selective pressures (Stressors) on a declining population (A), on fragmented populations (B), where A was divided into two subpopulations by a barrier (grey), and on a small population with low genetic diversity (C), represented by one of the two populations from B. Geometric forms inside circles represent individuals with different phenotypes associated with different genotypes (shape and colour); Stressors: the multiplication factor under each phenotype generates the number of each geometric form at that time point; at each of the three time points, genetic drift will remove the least common form (arrow, genetic drift) and one of the most common individuals dies (becomes transparent); the extinction vortex is a cycle represented by thick black arrows that lead to smaller population sizes (Reduced N); thin black arrows represent factors extrinsic to populations that lead to Reduced N and can be either natural instability or human mediated (A).

Most of this decline was not due to fluctuations of its prey (moose) or difficulties in finding other individuals for mating, but rather attributed to very high inbreeding and accumulation of deleterious mutations – the genetic Allee effect (Hedrick *et al.* 2014, 2017). This supports the idea that most species are driven to extinction when genetic factors such as low heterozygosity and inbreeding start affecting populations (Spielman *et al.* 2004), leading to an increased risk of genetic variability loss for species that are

increasingly threatened (Amos & Balmford 2001). Fagan & Holmes (2006) studied several vertebrate populations whose extinction was monitored over 12 years. These authors, found that when populations became closer to extinction, the rates of decline were larger, as theory predicted. However, this was associated with lower genetic diversity of the individuals remaining in the last surviving populations, rather than with a lower population size per se. Conservation genetic studies should thus promote the persistence of populations by maximising genetic diversity and positively influencing evolutionary processes (Frankel 1974; Latta 2008). However, the ecological and genetic Allee effects do not operate independently and it is important to consider their interaction to avoid false optimistic prognosis on the minimum number of individuals required to avoid the extinction vortex, given that these effects are generally additive and can even be superadditive (Fauvergue *et al.* 2012; Wittmann *et al.* 2016).

Metapopulations are a special kind of small population

In some species, small groups of individuals exist in patches that are connected by dispersal, forming a 'metapopulation', where each patch has a given probability of being colonised and suffering an extinction event (Gilpin & Hanski 1991; Hanski & Gaggiotti 2004). The difference between a metapopulation and a true small population is the fact that the first have evolved in a context of balancing recruitment and dispersal between patches in such way that the dynamics of colonisation-extinction-recolonisation is able to maintain a genetically healthy population (Hanski & Gaggiotti 2004). Despite this, a metapopulation is highly susceptible to disturbances such as habitat destruction and extinction of local nuclei. Thus the metapopulation equilibrium is often close to the extinction threshold and, coupled with decreasing patch size and increasing patch isolation, a metapopulation may quickly reach a state of non-equilibrium (Figure 3) (Nee & May 1992; Hanski 1998). Amarasekare (1998) formally tested the consequences of ecological Allee effects in metapopulations, and showed that these effects may increase metapopulation extinction risk at low habitat occupancy even more than habitat destruction alone. This is because ecological (and genetic) Allee effects can prevent a metapopulation from growing even if suitable habitat is available.

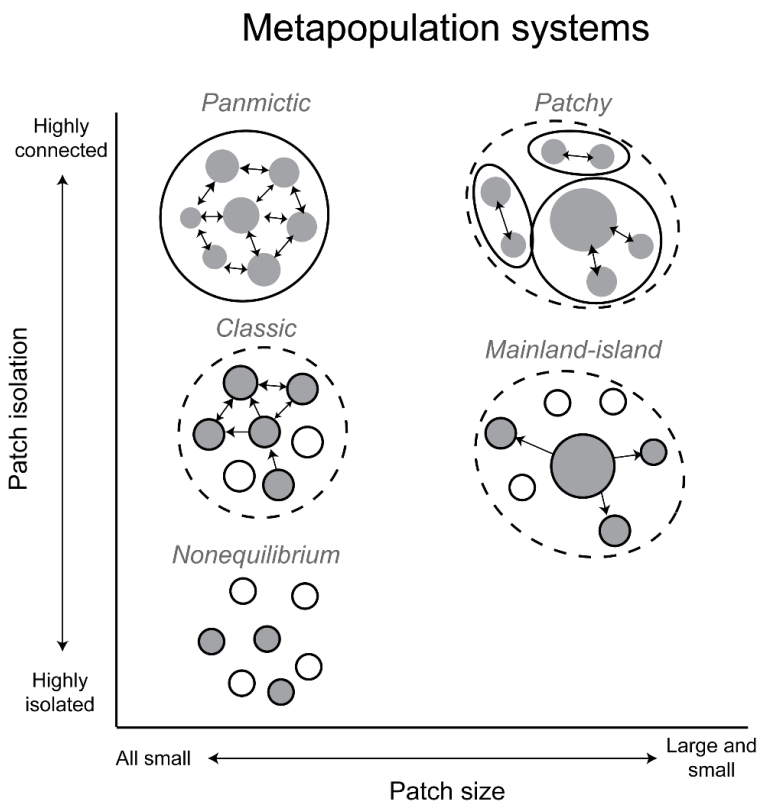


Figure 3. Schematic view of the effects of patch size and isolation on the dynamics of metapopulations. Grey and white circles represent occupied and vacant habitat patches, respectively; full and dashed outline delimit populations (except for vacant habitat patches) and metapopulations, respectively; arrows represent connected populations and the dominant direction of dispersal [adapted from Harrison & Taylor (1997) and Fullerton *et al.* (2011), with permission].

Genetic rescue

The genetic Allee effect can be counteracted through translocations or by promoting connectivity between populations. These actions may potentiate genetic diversity and increase population fitness, and generally are referred to as 'genetic rescue' (Whiteley *et al.* 2015; Hedrick & Garcia-Dorado 2016). The expectation is that a low level of immigration between small populations will result in increased fitness of the receiving population due to heterosis (masking of deleterious alleles) and/or the introduction of beneficial alleles (Tallmon *et al.* 2004; Weeks *et al.* 2016). There are various types of population rescue that, despite being categorised differently, are ultimately linked: *demographic rescue* relates to the movement of individuals from larger to smaller populations, resulting in an increase in the number of individuals; *genetic rescue* relates to the movement of alleles from a population with high genetic diversity to populations with low genetic diversity to increase fitness; and *evolutionary rescue* is related to the movement of alleles from a population of individuals adapted to given environmental

conditions to a population that has recently been exposed to those same conditions but does not have similarly well adapted alleles, resulting in increased fitness granted from the environmentally beneficial alleles (Vander Wal *et al.* 2013; Whiteley *et al.* 2015; Richardson *et al.* 2016). In contrast, immigration of genetically distinct individuals can lead to decreased fitness due to outbreeding depression in small populations, such as the loss of locally adapted genotypes through allele swamping (Smith & Wayne 1996). This might not only be problematic for cases of admixture between species (hybridisation) but also within species, leading to the loss of unique genetic diversity, often through hybrid maladaptation or inviability (Frankham *et al.* 2012). Nevertheless, the few studies that have tested absolute fitness have found mostly positive effects of low levels of migration, or a mix of positive and neutral effects (Whiteley *et al.* 2015).

Increasing evidence demonstrates that conservation genetics can help us understand population dynamics and how to preserve species and populations (DeSalle & Amato 2004; Frankham *et al.* 2010). One good example of the use of genetics to secure population resilience is the Florida panther (*Puma concolor coryi*). This subspecies of puma represented a long isolated population distributed in south-eastern United States, with a decreasing number of individuals. Inbreeding led to decreasing litter sizes and the appearance of defects not present in pumas captured west of the Mississippi river: a cowlick on the back of the neck, kinked tail, developmentally malformed sperm, heart valve defects, etc. (Johnson *et al.* 2010). A decision was made to introduce eight pumas from Texas in Florida in order to increase the genetic diversity of the Florida panther – an experiment that could compromise the original gene pool through outbreeding, but save the Florida panther that was compromised by inbreeding (Vander Wal *et al.* 2013). This genetic rescue generated an increase in heterozygosity, number of individuals and survivorship, and a decrease in the mean age of the populations, i.e. more young were becoming established in the population (Johnson *et al.* 2010). There are various examples of similar genetic rescue cases such as in bighorn sheep in the United States, and adders in Sweden, showing the same trend of enhanced population numbers after the introduction of new individuals, and suggesting that increased genetic diversity can help endangered populations to recover (Madsen *et al.* 1999; Hogg *et al.* 2006; Hedrick & Fredrickson 2010; Vander Wal *et al.* 2013). While these case studies are excellent examples of genetic rescue, it is usually advisable to perform population viability analyses before reaching such critical stages (Shaffer 1987; Beissinger & McCullough 2002; Whiteley *et al.* 2015). Many early viability analyses were based on ‘rules of thumb’ for e.g. minimum number of individuals to consider a species endangered, but what was readily understood by these studies was that a more nuanced

approach to conservation should be based on evolutionary theory and account for both neutral and adaptive variation over time through frequent monitoring (Latta 2008; Flather *et al.* 2011).

Non-invasive genetic sampling: so you can catch them all

One of the biggest issues regarding the study of threatened species is the disturbance that sampling and monitoring may cause to animals and their habitat while conducting conservation actions. For these species, which are generally elusive and rare, a widely used approach that minimises disturbance is non-invasive genetic sampling (NiGS). This methodology allows genetic information to be gathered without handling, contacting or even seeing the organisms (Waits & Paetkau 2005; Beja-Pereira *et al.* 2009). The information is obtained using DNA extracted from a variety of samples, such as, faeces, hairs, feathers, urine, egg shells, scales, skin, etc. (Smith & Wayne 1996). In the case of small vertebrates, one can also obtain samples from 'pellets' (containing indigestible hair and bones) produced by predatory birds at their roosts, in particular the cosmopolitan barn owl (*Tyto alba*). These pellet accumulations represent a cost effective way of sampling the prey from an owl's territory without human disturbance of the existing populations (Taberlet & Fumagalli 1996; Avenant 2005).

When correctly used (see below), NiGS can provide accurate and valuable information relevant to conservation such as: detection of rare species, individual and gender identification, population size estimation, determination of social structure, measurement of genetic diversity and gene flow, and detection of hybridisation, diseases and diets [for a review see Waits & Paetkau (2005)]. NiGS can thus be an excellent tool for the conservation of threatened and elusive species. This is well illustrated by Blair & Melnick (2012) who collected faecal samples from Central American squirrel monkeys (*Saimiri oerstedii*) and were able to identify over 240 individuals and obtain detailed information on population structure, long-term movements and family relationships. This work would not have been possible in the same timeframe and with similar effort with a traditional capture-mark-recapture approach. Thus, in terms of sampling, studies have shown that NiGS can outperform traditional capture-mark-recapture studies in estimating animal densities, considering not only a lower sampling effort, but also an increased 'capture' rate (Sabino-Marques *et al.* in prep). There are specific precautions one must have from sample collection to data analyses, including sterile collection conditions, dedicated laboratory areas to have especial care regarding contamination from various sources, design of specific markers for degraded DNA, replication, and the consideration

of the main sources of error when generating and analysing the data (Taberlet *et al.* 1999; Waits & Paetkau 2005; Pompanon *et al.* 2005; Broquet *et al.* 2007; Beja-Pereira *et al.* 2009). For example, in terms of sampling regime, choosing between systematic or opportunistic NiGS schemes results in either higher number of individuals detected or recaptures, respectively (Rehnus & Bollmann 2016); also, caution is needed in laboratory execution given the samples' low DNA quality and quantity results in low amplification success, genotyping error rates and varying levels of missing data (Taberlet *et al.* 1999; Beja-Pereira *et al.* 2009). Still, Smith & Wang (2014) performed a series of simulations to evaluate the effect of small sample sizes and genotyping errors on the most commonly estimated genetic parameters, and determined that reasonable estimates of genetic variation and population subdivision can be obtained from non-invasive samples. There are many other examples of the utility and feasibility of NiGS, but one of the biggest limitations is the low amplification success for nuclear markers. Current high throughput sequencing technologies are allowing cost-effective generation of a greater quantity and quality of data, which may soon revolutionise NiGS studies by overcoming many of its limitations (de Barba *et al.* 2016).

3. Transitioning to Conservation Genomics

With the advances of genomic techniques, more markers representative of the entire genome are now accessible through the use of high throughput sequencing (HTS) methods. These methods allow the generation of large amounts of sequence data, providing deeper insights into the patterns and processes involved in maintaining and creating genetic diversity (Hoffmann *et al.* 2015). However, given the generalised perception that genetic information is not necessary to conservation management, the mystique of genomics may widen this gap (Winter *et al.* 2013; Shafer *et al.* 2015; Taylor *et al.* 2017). Nevertheless, the easier access to next generation sequencing technologies and the recognition that genomics can provide better substantiated answers to previous questions and allow for new ones to be asked, should ultimately overcome the genetic/genomic-conservation gap (Taylor *et al.* 2017).

New tools for old problems

Whole genome sequencing will likely become the standard for genetic studies of natural populations, as it provides the most complete view of genetic variation, however its costs can still be prohibitive for many conservation studies (Ekblom & Wolf 2014).

There are many advantages in using a reference genome for conservation studies as it allows the assignment of reads to chromosomes and annotated loci, with which we can more accurately determine population history over time, find functional genes associated to selected candidate loci, or areas of the genome under higher or lower selection pressures (Benestan *et al.* 2016). A good example of the use of whole genomes in conservation is the detection of long runs of homozygosity (genomic flatlining) in the African cheetah, the Iberian lynx and the Channel Island fox (Dobrynin *et al.* 2015; Abascal *et al.* 2016; Robinson *et al.* 2016). These authors recognised the presence of some regions of the genome significantly more variable than others, usually associated to genes linked to key aspects of the animals' biology. This suggests that there may be a selective maintenance of diversity in 'essential' genomic regions for the persistence of these species, which is maintaining viable populations even in the face of generalised genetic depletion. However, most species of conservation concern do not benefit from well-developed genomics resources such as a well-annotated genome, and a genome is likely unnecessary for many conservation studies (Hoffmann *et al.* 2015; Shafer *et al.* 2015; Allendorf 2017). Several methods of genome complexity reduction can still provide a genome wide perspective of genetic variation – these are called reduced-representation sequencing (RRS) methods (Hoffmann *et al.* 2015) (Table 1). With these methods one can access large numbers of markers that are randomly distributed throughout the genome, being representative of the overall variation, and these may reflect both neutral and adaptive variation (Ouborg *et al.* 2010; Shafer *et al.* 2015). The choice of methodologies depends on the research question to be addressed, the sampling design, the allocation of sequencing effort and which genomic tools are already available, if any (Hoffmann *et al.* 2015; Benestan *et al.* 2016; da Fonseca *et al.* 2016). As conservation biology is in essence an applied discipline, most of the questions in conservation genetics relate to relatively recent time scales dealing with the preservation of current genetic diversity, and aiming for a high impact on management decisions with few financial resources (Benestan *et al.* 2016). The most promising tools in genomics for conservation studies include targeted DNA and RNA sequencing and restriction site-associated DNA (RAD) sequencing (RAD-seq) (Baird *et al.* 2008; Steiner *et al.* 2013). RAD-seq and other similar restriction-based techniques, such as genotyping-by-sequencing [GBS, Elshire *et al.* (2011)], have played a major role in conservation genetics given its wide range of applications in studies of species with little or no genomic resources (Table 1) (Davey *et al.* 2011; Narum *et al.* 2013; Andrews & Luikart 2014; Andrews *et al.* 2016; Benestan *et al.* 2016). Additionally, restriction-based techniques can be used not only in individual samples but also on pools of individuals (Pool-seq),

which allows cost-effective studies for addressing a wide range of questions, such as population assignment and diversity, genotype-phenotype mapping for adaptation, etc. (Schlötterer *et al.* 2014).

Still, the major limitations for most conservation studies is the need of comprehensive sampling, and here monitoring becomes an essential part of conservation genetics which can benefit from the use of HTS methods (Frankham *et al.* 2010; Schlötterer *et al.* 2014). Genetic monitoring may be especially difficult in rare, threatened and elusive species due to ethical, financial or logistic reasons, and this is where non-invasive genetics may become a valuable tool in conservation genomics (Waits & Paetkau 2005).

Table 1. The utility of different HTS methods to different questions in conservation genetics associated to a simple representation of the division of high throughput sequencing methods (HTS) into some of the most commonly used approaches [adapted from Corlett (2017), with permission].

	HTS methods					
	Complete	Reduced representation				
		Targeted			Anonymous	
		amplification		hybridisation	restriction enzyme	
	WGS	RNA-seq	DNA barcoding	UCEs	Exon capture	RAD-seq GBS
Identifying cryptic lineages	X	X		X	X	X
Delimiting conservation units	X	X				X
Optimising ex-situ conservation	X	X			X	X
Monitoring pathogens	X	X	X		X	X
Identifying pathogen resistant individuals	X	X			X	X
Selecting populations for reintroductions	X	X			X	X
Assessing past and present connectivity	X			X	X	
Assessing biodiversity	X		X			X
Detecting invasive species	X		X			X
Assessing adaptive potential	X	X			X	X
Genetic rescue of inbred populations	X	X			X	X

Non-invasive genomics

Considering the new developments of high throughput sequencing (HTS), an increasingly large array of techniques is becoming available for potentiating the application of NiGS in conservation studies. The first application of HTS to faecal samples was developed by Perry *et al.* (2010), to recover specific regions of the genome of the western chimpanzee (*Pan troglodytes*), by designing DNA probes to perform exon capture (targeted approach, Table 1). Due to the degraded nature of the DNA present in faecal samples, studies that enrich for endogenous DNA are very useful to NiGS, as it is a form of reducing genome complexity by selecting specific genomic regions. The advantage is that these endogenous genomic regions can be sequenced in multiple samples that usually have a high percentage of exogenous DNA. However they require extensive genomic resources such as genomes or transcriptomes to develop the DNA probes (Perry 2013). Due to the high costs associated with generating a genome or transcriptome, an increasing number of studies are currently attempting to use NiGS without such resources. Russello *et al.* (2015) performed genotyping-by-sequencing (GBS) on hair samples of American pika (*Ochotona princeps*) and were able to detect population structure and infer adaptation to altitude within populations. However, the collection of hair samples requires planning and monitoring the hair traps, and hair DNA is of much higher quality than the average samples used in NiGS (e.g. faeces) due to the DNA protection conferred from keratin (Henry *et al.* 2011). The application of genomic tools to more degraded samples such as faeces is still under development. Graham *et al.* (2015) performed the first study on the impacts of sample degradation levels for obtaining reliable SNPs with double-digest RAD-seq (ddRAD). They have shown that genomic studies on degraded samples are possible up to the point when samples are extensively degraded. More recently a study by Costa *et al.* (2017) demonstrated the possibility to build RAD-seq libraries from *Equus* faecal samples. This highlights the potential of non-invasive genomics in population level studies, including the definition of conservation units, detecting hybrid zones, determining the origin of illegal animal trafficking, etc.

Despite what is being achieved with genomic tools for NiGS, there are still many difficulties in generating reliable data. There are new field-based technologies being developed that show great promise for NiGS, where genetic data can be reliably obtained and analysed *in situ*, possibly resulting in direct application to urgent management decisions in conservation (Pennisi 2016; Lee 2017). Analogous genomic tools have been applied to historical natural history collections, which present similar challenges in terms

of DNA quality and quantity, and thus may prime the use of the same techniques in conservation studies (Prosser *et al.* 2016). Because these techniques for natural history collections are more substantially developed than for NiGS, they are described in the next section.

Natural history collections

Natural history collection samples have by definition been collected invasively, however they do not involve the disturbance of current populations. This greatly enhances the use of natural history collections, namely for linking Linnean names to type specimens, resolving taxonomic uncertainties, establishing boundaries of evolutionarily significant units, among other examples (Paplinska *et al.* 2011; Holmes *et al.* 2016; Prosser *et al.* 2016). One of the main advantage of museum collections is the ability to compare levels of genetic diversity between historic times and the present (Holmes *et al.* 2016). This is especially important with threatened species given that many of them have declined dramatically in the past decades and centuries, and it makes possible to quantify the rate of genetic diversity loss over time (Wandeler *et al.* 2007; Ceballos *et al.* 2015). Museum samples require special attention from sampling to data analysis similar to that given to modern degraded samples, such as those obtained non-invasively (Wandeler *et al.* 2007; Rowe *et al.* 2011). Museum samples still pose many challenges in terms of exogenous DNA content. Rowe *et al.* (2011) analysed different museum samples of *Rattus norvegicus* including toes, ankles, lips and molars, and obtained around 40% of the sequencing reads assigned to the target species with high confidence. These authors found that SNPs obtained from skin and skull samples were highly repeatable when compared to a reference genome.

Although genomes are becoming readily available for most groups and the cost of sequencing is decreasing very rapidly, there are still many taxa for which a *de novo* approach is the best alternative (Perry 2013). In the case of museum samples, there is only one study using a non-targeted approach, in this case GBS, aiming at species delimitation in golden rod (*Solanum* spp.) herbarium specimens (Beck & Semple 2015). The authors were able to routinely generate over 1700 SNPs and confirmed that the clustering of genetic variation was in line with current species delimitation. Studies like this suggest that non-targeted approaches (such as GBS and RAD-seq) can provide reliable results when applied to samples with degraded DNA. The combination of the study of natural history collections and non-invasive samples can thus result in a cost-effective strategy for looking into the past and comparing it to the present genetic

variation of threatened species. This idea of examining species genetic variation through time has been shown to be very powerful for conservation, not only as it puts numbers on what extinction is, raising awareness for the decline of many species, but it also allows us to understand the species evolutionary history and develop better conservation decisions (Crandall *et al.* 2000).

4. Understanding the past to preserve the future

As we have seen from various examples of natural history collections studies, it is important to look into the past to better protect the present and attempt to ensure the future of species. Some species are naturally more genetically diverse than others, and it is important to understand the mechanisms behind this variation to determine what is causing species decline today and if species will be able to sustain further changes. Ellegren & Galtier (2016) proposed that the life history of a given species, its mating system, demographic history, gene density across the genome and recombination rate are the main determinants of genetic diversity displayed by an individual, population or species. It is well established that animals like rodents are often *r*-strategists, resulting in high effective population sizes, and producing many germ cells per generation, leading to high mutation rates, rapid evolutionary change and adaptation, and sometimes, speciation (Nabholz *et al.* 2008; Ellegren & Galtier 2016). On the contrary, groups with the opposite traits to these (*K*-strategists) will often have lower genetic diversity on a similar evolutionary time frame and be more susceptible to changes to their environment.

Bursts of speciation and extinction through time

Evolution occurs through both slow and rapid genetic change through time, and the pace of this change can be a determinant for adapting to new environmental conditions or for becoming extinct. The Red Queen hypothesis states that "*takes all the running you can do to stay at the same place*", thus the faster the change, the faster species need to evolve (Benton 2009). But while faster evolutionary rate is associated to speciation and increased adaptive potential, lower evolutionary rates have also shown to be correlated with lower extinction rates in 'living fossils' (Bennett *et al.* 2017). There are particular time points in the earth's history when the speed of environmental change became too fast for most species to keep up, resulting in mass extinctions (Hallam & Wignall 1997). On the other hand, there were periods of time, either after harsh climatic conditions or catastrophic events, which allowed species to rapidly spread and speciate due to new suitable conditions, empty or new habitats or new advantageous adaptations

(Simões *et al.* 2016). An informative way of studying species formation and extinction through time is using phylogenetic analysis (Hey 1992), which has shed light on many cases of such rapid radiations, such as neoavian birds (Suh *et al.* 2015); the African great lakes cichlid fiches (Brawand *et al.* 2014); Philippine shrews (Giarla & Esselstyn 2015); and arvicoline rodents (Lv *et al.* 2016). What all of these and other studies have helped us to understand is that the Earth's species composition is ever-changing through bursts of speciation and extinctions. However, different species are likely to have different responses to change, so by evaluating intraspecific genetic diversity and its change through time, we will be able to predict species vulnerability to change in the future. Given that change leads to population subdivision and that spatially structured populations are genealogically linked, phylogeography emerged as a way of combining species' biogeographic and phylogenetic history (Avise 2000). To understand the processes that led to the current genetic diversity of species, and especially aimed at its protection, one must first understand the genetic history of species through time and space (Hewitt 2000; Benton 2009).

Quaternary glacial refugia (within refugia)

Before the Anthropocene, the glacial episodes of the Quaternary (2.6 million years ago–present) were the main drivers of the present day distribution of species, especially with the advance and retreat of the ice sheets through multiple glacial cycles (Provan & Bennett 2008). The most recent glaciation event was the Last Glacial Maximum (LGM) around 23 000 – 18 000 years before present (Provan & Bennett 2008), which has been thoroughly charted in Europe (Hewitt 1999, 2011). Europe is one of the most well-studied geographic areas in terms of fossils and phylogeography, where many studies now recognise different refugial areas for temperate and cold-adapted species (Sommer & Nadachowski 2006; Stewart *et al.* 2010) (Figure 4). Temperate species are the best studied group which largely persisted through glacial maxima in lower-latitude refugia, where climatic conditions were less extreme - the Mediterranean refugia (Figure 4A) (Hewitt 2000). This is shown by the higher levels of genetic diversity found in species that have recolonised mainland Europe from these areas – mostly the Iberian, Italian and Balkan peninsulas (Hewitt 1999): as temperature decreases, species distributions moved to the south and populations got separated in unconnected geographic areas.

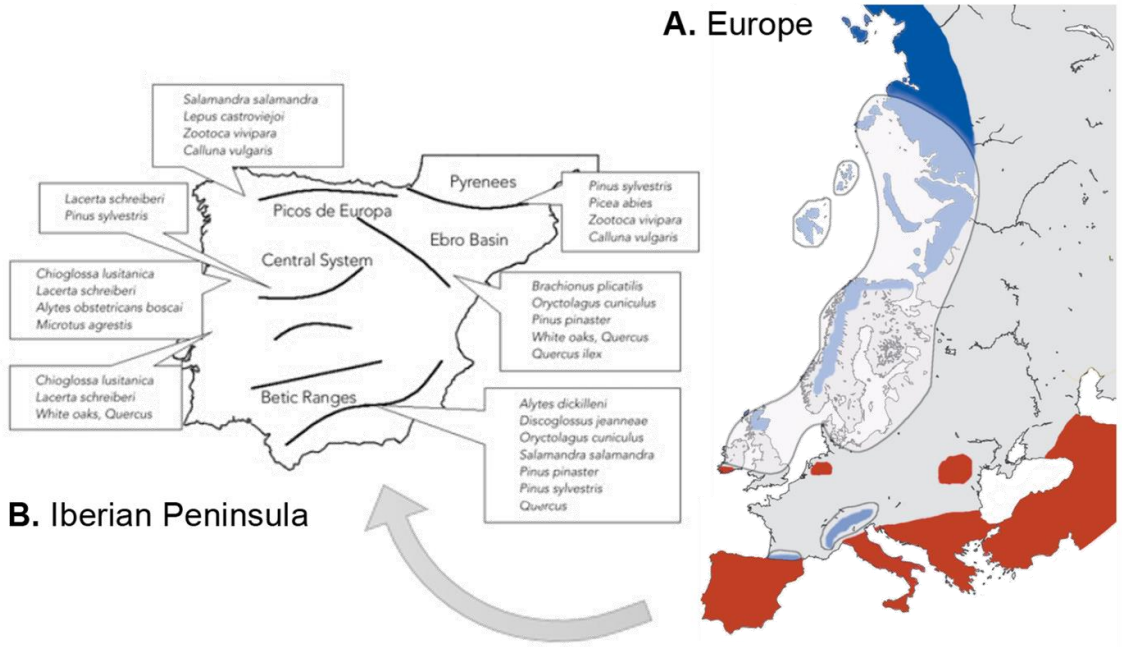


Figure 4. (A) Map showing two types of Pleistocene refugia in Europe and western Asia for temperate species (red) and interglacial refugia for cold-adapted species (blue). Polygons with transparency represent the ice sheets during the Last Glacial Maximum (LGM). (B) Detailed map of the Iberian Peninsula showing the main mountain ranges and the location of putative glacial refugia for various terrestrial animal species [adapted from Stewart *et al.* (2010) and Gómez & Lunt (2007), with permission].

Thereafter, the effects of genetic drift led to the divergence of subdivided populations that became genetically distinct over time. Once temperatures increased again, a subset of individuals recolonised central Europe; this bottleneck left a signature that can be detected by phylogeographic studies as a scenario of 'southern richness and northern purity' (Hewitt 1996, 1999, 2000, 2011; Taberlet *et al.* 1998; Provan & Bennett 2008; Stewart *et al.* 2010). There is, however, an aspect of refugial areas that has recently received more attention. For species with limited dispersal, the geographic complexity of the refugial areas often led to further subdivision of the refugial population, resulting in 'refugia within refugia' (Gómez & Lunt 2007; Abellán & Svenning 2014). It is within these isolated enclaves of suitable climate space that relict species and populations may have been formed (Hampe & Jump 2011). Specifically within the Iberian Peninsula, which is the geographic area of this thesis, we find various geographical structures such as mountain ranges and large rivers that are thought to have limited the dispersal of various species (Gómez & Lunt 2007) (Figure 4B). These are now known to be the origin of various divergent phylogenetic lineages, and for those species that were not able to expand out of the peninsula, they are also the last place for these endemic relicts.

Endemic relicts – small mammal edition

Endemic (or native) species are organisms that only naturally occur in a given geographic area (IUCN 2016). This area can be large or very restricted, and for the latter case, species are referred to as highly endemic, which seems to be the case for most rare and threatened species (Ohlemüller *et al.* 2008). There are different hypotheses to explain why rare species have restricted ranges (Stebbins 1942): the 'beginner hypothesis' – rare species are beginners that have not had the time to spread; the 'senescent hypothesis' – rare species were once common but with evolutionary time and environmental stability, species evolved from generalists to specialists, and this makes them currently unable to spread to different habitats when faced with change; and the 'genetic hypothesis' – widespread species have high genetic diversity associated to different ecotypes, while rare species are usually specialised to a single restricted ecotype. The first two hypotheses have largely been refuted and the genetic hypothesis appears to be the most accepted hypothesis considering the outcome of glaciations on temperate species (Stebbins 1942; Ohlemüller *et al.* 2008).

As species contracted their ranges to glacial refugia, and often to smaller areas within refugia, many populations had very limited connectivity, if any (Gómez & Lunt 2007). Through processes such as lineage sorting, drift, and possibly adaptation to local conditions, these populations may have diverged during these range contractions and isolation processes resulted in the high species richness and great intraspecific divergence for many taxa found currently in the Mediterranean peninsulas (Hewitt 1996, 2011). Once conditions improved, some species and lineages expanded out of the refugium, but it is currently argued that most species within Mediterranean peninsulas represent long term isolates undergoing allopatric speciation (Bilton *et al.* 1998). This is where the concept of relict species or populations arises (Bonn *et al.* 2002; Habel & Assmann 2009). These are usually highly endemic groups characterised by a small number of individuals which are restricted to small geographic areas (Habel & Assmann 2009). The Iberian Peninsula is a good example of the processes mentioned above, since we find groups of species with restrict ranges, such as within the soricomorphs (Soricomorpha) and rodents (Rodentia) (Palomo *et al.* 2007). Many lineages have been described within these groups in the Iberian Peninsula, which are very different from the lineages found elsewhere, such as in the case of the southern water vole (*Arvicola sapidus*) and the field vole (*Microtus agrestis*) (Centeno-Cuadros *et al.* 2009a; Paupério *et al.* 2012). In addition, there are five small mammal species endemic to the Iberian Peninsula: the Pyrenean desman (*Galemys pyrenaicus*), the Iberian mole (*Talpa*

occidentalis), the Iberian shrew (*Sorex granarius*), the Lusitanian pine vole (*Microtus lusitanicus*) and the Cabrera vole (*M. cabrerae*) (Randi 2007).

Recent phylogeographic data show that the Pyrenean desman is divided into two main mitochondrial lineages with ~1.0% divergence, while at the nuclear level (RAD-seq data) it is divided in five clusters (Igea *et al.* 2013; Querejeta *et al.* 2016). Given the values of nucleotide diversity and heterozygosity, the likely refugial area was inferred to be in north-western Iberia (Igea *et al.* 2013; Querejeta *et al.* 2016). The Iberian mole is distributed throughout most of Iberia with three very distinct mitochondrial lineages (divergence ranging between 2.3% and 3.3%), however no structure was found at the nuclear level (Nicolas *et al.* 2017). From distribution modelling the refugial areas during the LGM were inferred to be in western Iberia. No detailed phylogeographic studies have been carried out for the Iberian shrew so far, but analysing the cytochrome-*b* sequences available at GenBank (Fumagalli *et al.* 1999; Dubey *et al.* 2007; Yannic *et al.* 2008), there are two lineages diverging around 1.2% occurring in sympatry in central Iberia, while only one of them has been found elsewhere. The Lusitanian pine vole is endemic to the northwest quadrant of Iberia, but there is ongoing mitochondrial introgression and hybridisation with *M. duodecimcostatus*, making it difficult to determine which are the Lusitanian pine vole's genetic lineages, their divergence and phylogeography (Bastos-Silveira *et al.* 2012). As the target species of this thesis, the Cabrera's vole phylogeographic history will be discussed in detail below. However, as a species associated with the Mediterranean bioclimatic region, it is expected that its phylogeographic pattern would differ significantly from the four endemic species described above since all are associated with the Atlantic bioclimatic region (Bastos-Silveira *et al.* 2012; Igea *et al.* 2013; Suárez-Seoane *et al.* 2013; Pita *et al.* 2014).

The large number of rare species and lineages found within the various refugial areas in Iberia possibly result from long term specialisation and inability to spread due to geographic barriers such as mountain ranges (Weiss & Ferrand 2007). It is the high variability among refugial areas that makes them so important for the conservation of the genetic diversity of the species, even when considering common species. This genetic diversity can be very important for the adaptation to future climatic conditions, as illustrated by those lineages that expanded from refugia compared with others that did not. This may be a reflection of higher tolerance of the former and possibly higher specialisation to local conditions of the latter (and lower tolerance for ecological variation) as seen in *Mustela nivalis* in central Europe (Hill *et al.* 2011; McDevitt *et al.* 2012). Understanding past interactions of populations and species and evaluating niche conservationism before and after major climatic events, and current human

disturbances, allows us to better recognise species susceptibility to change and redirect conservation strategies for species protection (Levy 2013; Bolliger *et al.* 2014).

Change is ~~coming~~ already here… and it's fast!

Climate change has been naturally pervasive throughout the earth's history, either by periodic changes in the Earth's orbit and axis of rotations (Milankovitch cycles, 10-100 thousand years ago) or more randomly by catastrophic events, and these changes are usually marked by the extinction of many species and, sometimes, entire groups (Hallam & Wignall 1997; Dynesius *et al.* 2000). Apart from catastrophic events that have occurred a handful of times on the history of the Earth, most changes happened gradually, and in these cases some taxa are able to adapt to new conditions and evolve (Hallam & Wignall 1997; Mittelbach *et al.* 2007). In contrast, today we are observing a rapid change in the environment associated to human impacts (habitat destruction, pollution and direct persecution of species), and considering the last 1500 years (Figure 5), this change has been especially rapid during the last century (Mann *et al.* 2008; Neukom *et al.* 2014).

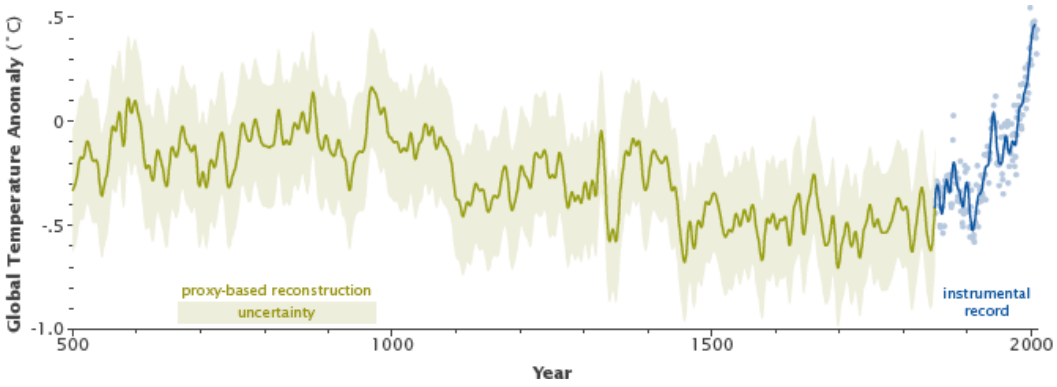


Figure 5. Temperature history during the last 1500 years. The green line was reconstructed from paleoclimate data (with shading representing uncertainty) and the blue line is based on instrument measured data [© NASA adapted from Mann *et al.* (2008), with permission].

For many species, often endemic to small areas, these climatic changes represent an extinction sentence, not only due to the fast pace of change of abiotic factors on such a short time scale, but also due to new biotic interactions, such as intra and interspecific competition for resources, habitat, etc. (Malcolm *et al.* 2006; Benton 2009; Cahill *et al.* 2012). The ability of species to adjust to environmental change is influenced by many factors, both internal and external to species (Williams *et al.* 2008): at the intrinsic level, the species dispersal ability, physiology and genetic diversity are the first line of response to a changing environment, as these will affect the sensitivity of species to change

('fundamental adaptive capacity'); at the extrinsic level, the extent of the regional climatic change, the strength of new biotic interactions, the way that species are managed, among many other factors, will lead to a higher or lower exposure of species to threats (Williams *et al.* 2008; Beever *et al.* 2016). The combination of species sensitivity and degree of exposure to new threats will result in the vulnerability of each species to extinction; i.e. their 'realised adaptive capacity' (Dawson *et al.* 2011).

Dispersal, plasticity and evolutionary change

Extinction can be avoided in populations that are able to move towards more favourable habitats (dispersal), overcome stressful new conditions via phenotypic plasticity, or undergo rapid genetic change to adapt to the new conditions depending on their adaptive potential (evolutionary change) (Hoffmann & Sgrò 2011). Although there are varied types of responses to change, it is already apparent that many species are fast shifting their ranges to higher latitudes or altitudes as a result of rapid climate warming (Chen *et al.* 2011). In cases where species are limited by dispersal, either due to their specific dispersal ability or the landscape permeability to movement, genetic diversity will play a major role in their ability to persist *in situ* (Schloss *et al.* 2012; Ghalambor *et al.* 2015). Species tend to respond to change *in situ* by adjusting within the limits of phenotypic plasticity (Ghalambor *et al.* 2007). For a given trait and environment, those traits that are more plastic have a wider tolerance to change and are thus more advantageous in heterogeneous environments, however the costs of plasticity means that the more plastic genotypes may be outcompeted by highly specialised ones, which are better adjusted when environmental conditions are very stable (Figure 6) (Chevin *et al.* 2010). The costs associated with plasticity are factors such as maintaining a physiological and developmental capacity for perceiving environmental cues that are occasionally expressed, and also producing different responses accordingly (Reed *et al.* 2011). As environmental cues shift from the long term optimum, as in the case with climate change, the most affected species will be the specialists, which is the case for most threatened species (Clavel *et al.* 2011). With environmental change, individuals belonging to specialist species will have locally decreased fitness; with increased habitat fragmentation, these individuals will no longer be able to disperse through the inhospitable matrix between populations; and with reduced gene flow, populations will lose genetic diversity and become inbred, entering the extinction vortex (Frankham *et al.* 2012; Hoffmann *et al.* 2015). Alternatively, species can adapt to the new conditions either from standing genetic variation or selection on new mutations, but the first is the

alternative likely leading to faster adaptation due to the higher speed and probability of fixation of pre-existing alleles (Barrett & Schluter 2008; Teotónio *et al.* 2009). Evolutionary conservation aims at maximising the genetic diversity of species, and thus their adaptive capacity as species, allowing evolutionary resilience to future change in these novel ecosystems (Sgrò *et al.* 2011; Eizaguirre & Baltazar-Soares 2014; Beever *et al.* 2016; Corlett 2016).

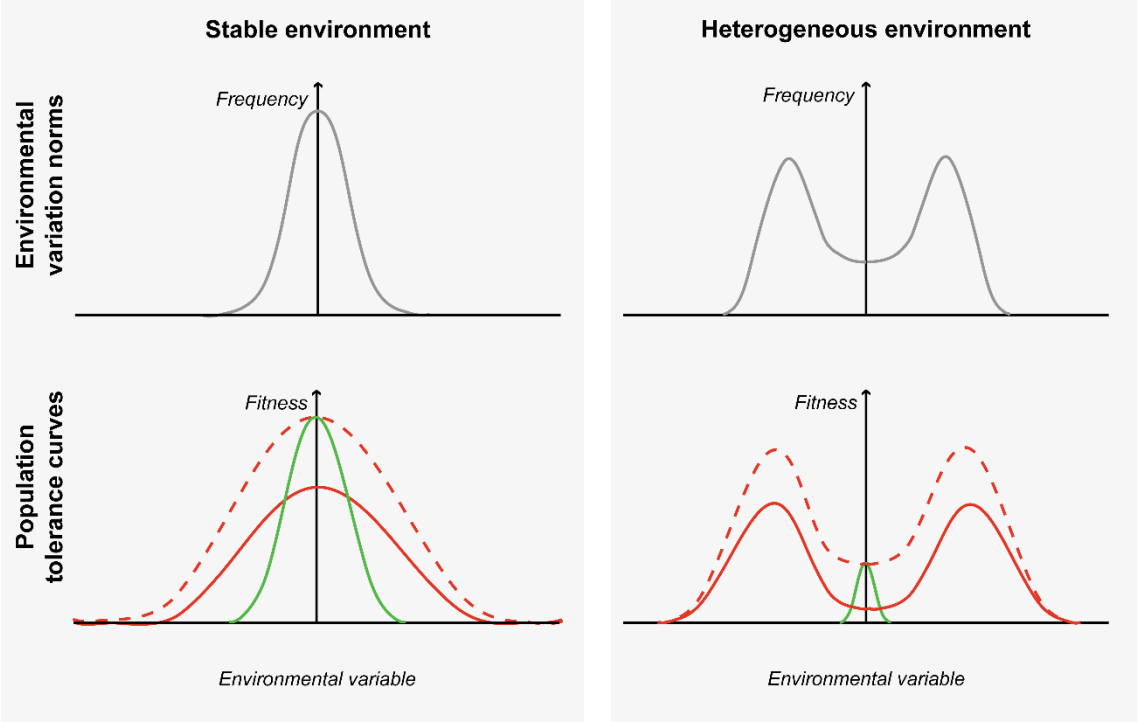


Figure 6. Phenotypic plasticity and population reaction curves to environmental variation in stable and heterogeneous environments. The top row shows patterns of variation of an environmental variable, and the frequency of its values given a time period (e.g. mean daily temperature over one year); the bottom row shows the tolerance curves, i.e. the fitness for a given value of the environmental variable (e.g. for a given temperature) for two populations where individuals have more (red) or less (green) phenotypic plasticity, and thus can tolerate more or less variation in the environmental conditions, respectively; the dashed red line represents the potential fitness of the population without the cost of plasticity, however the observed response shows lower fitness than expected due to the maintenance of plasticity to less optimal conditions [based on Chevin *et al.* (2010), with permission].

Landscape genetics: conservation in time and space

The landscape plays a crucial role in promoting or limiting species persistence. As organisms are forced to adapt to specific habitat conditions, managing species for conservation requires baseline knowledge on how genetic diversity is shaped by the landscape (Manel *et al.* 2003). Landscape genetics combines population genetics, landscape ecology and spatial analytical techniques to quantify the effects of the landscape on evolutionary processes (Manel *et al.* 2003; Balkenhol *et al.* 2016). This is

a valuable tool for conservation as it allows: genetic diversity to be mapped explicitly and to define conservation units; dispersal and movement (gene flow) between populations to be measured; and barriers to movement and matrix resistance to determine levels of isolation to be detected (Manel & Holderegger 2013). These are important data to inform conservation strategies on how to facilitate or impede gene flow, and thus how to avoid inbreeding and outbreeding depression, respectively, and to evaluate the efficacy of conservation measures (Segelbacher *et al.* 2010; Manel & Holderegger 2013; Levy 2013; Bolliger *et al.* 2014; van Strien *et al.* 2014). Depending on the question and the evolutionary processes under study, one must: define the landscape (e.g. extent, resolution and content); define the unit of study (e.g. individuals or populations); define the sampling scheme depending on the population structure; characterise the distribution of individuals in the landscape (e.g. through points, neighbourhoods, etc.); and choose an adequate method for the analysis (e.g. regression, correlation, ordination, clustering) (Balkenhol *et al.* 2016; Richardson *et al.* 2016). For a good fit to conservation purposes, Keller *et al.* (2015) proposed some guidelines within this general framework for landscape genetics studies (Figure 7).

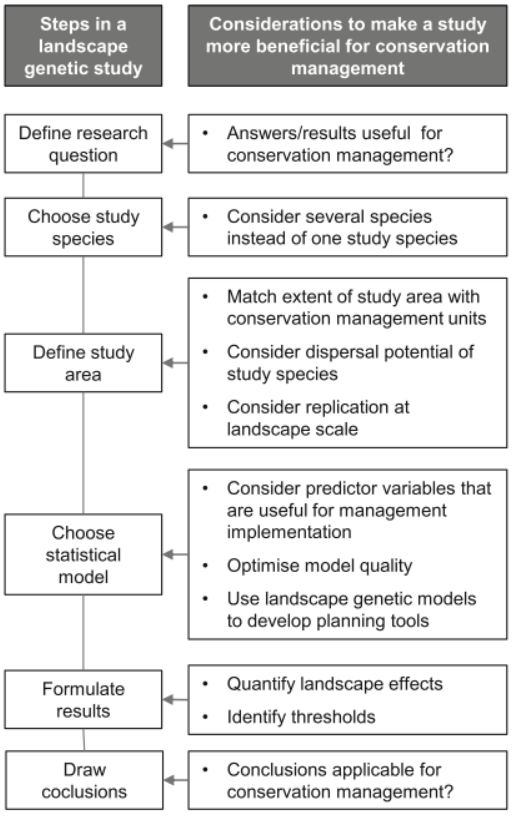


Figure 7. Flowchart of important steps and respective considerations to account for when designing a landscape genetic study to promote its inclusion in conservation planning [from Keller *et al.* (2015), with permission].

The recent use of genomic approaches in landscape genetics (also called landscape genomics) in the context of global changes is now providing substantial insights into the impact of the climate change in natural populations, as shown in alpine chipmunks (*Tamias alpinus*) from Sierra Nevada (Moritz *et al.* 2008; Rubidge *et al.* 2012; Bi *et al.* 2013; Holmes *et al.* 2016). Using both museum (from 1915) and modern day (from 2004-2008) tissue samples, the authors demonstrated an increase in population substructure over the last 100 years, indicating that habitat modifications are creating barriers to dispersal in modern populations (Figure 8) (Moritz *et al.* 2008; Rubidge *et al.* 2012; Bi *et al.* 2013; Holmes *et al.* 2016). The availability of historical and modern specimens from the same localities were fundamental for the assessment of climate induced range changes, and only by having a temporally and spatially explicit study, were the authors able to detect climate induced range shifts. One good review of the application of landscape genetics to conservation is provided by Schwartz *et al.* (2010).

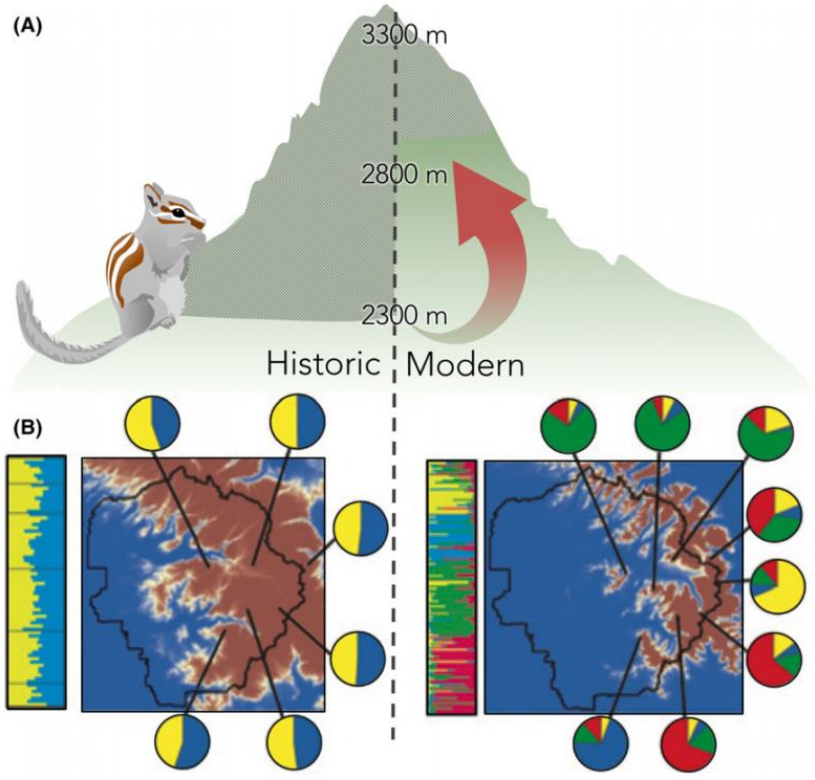


Figure 8. Detection of range shifts on alpine chipmunk, *Tamias alpinus*, through the comparison of museum and present day samples: (A) Elevational range occupied by *T. alpinus* from historical (left) to modern (right) times; (B) Maps show Yosemite National Park (black outline) coloured in blue to brown corresponding to low to high occupancy probability, also associated with altitude for *T. alpinus* in both historical (left) and modern (right) times. Pie charts correspond to STRUCTURE population assignment based on the bar plots to the left of the maps [from Holmes *et al.* (2016) based on Moritz *et al.* (2008) and Rubidge *et al.* (2012), with permission].

In the same way that landscape genetics is used as an exploratory tool for current conservation questions, it can also be used for predictive purposes, which is especially important in a context of increasing climate change and habitat reduction (van Strien *et al.* 2014). The main questions currently addressed by landscape genetics in its intersection with conservation relate to how recent changes in climate and habitats have affected the patterns of neutral and adaptive genetic variation and how species will adapt to these changes (Manel & Holderegger 2013). As noted above, with the recent advances in molecular genomics, we can now collect data on both neutral and adaptive loci without access to extensive genomic resources. By using neutral loci we can answer questions related to how past and present landscapes affect ecological factors, which in turn influence gene flow. By contrast, adaptive loci are directly affected by the landscape and climate (via selection), allowing us to answer questions relating to the influence of environmental factors on the geographic distribution of genetic variation (Balkenhol *et al.* 2016). For non-model species, most landscape genetics studies make use of SNP variation to look for the association of environmental gradients and genotypes (Sork *et al.* 2013). Given the uncertainty of what is, and especially what will be adaptive with future environmental change, most studies directly applying knowledge of adaptive variation for species conservation are restricted to plants and their association to climatic regimes, and to commercially valuable species, mainly salmonids, for e.g. the maintenance of distinct migratory life-styles (Eckert *et al.* 2010; Matala *et al.* 2014; Leitwein *et al.* 2016). A more conservative approach is looking at adaptive (non-neutral) variation as a whole, defining the species different adaptation regimes, and promoting its diversity across populations to ensure that traits under selection are as variable as possible. This is important both for phenotypic traits that are determined by a few genes with large effects, as well as for quantitative traits determined by a large number of small effect genes, which are more difficult to detect and likely the most commonly associated to adaptive responses (le Corre & Kremer 2012). However, the detection of adaptive variation in general is still surrounded by high degrees of uncertainty, thus the conservation of what is viewed as both ‘adaptive’ and ‘neutral’ genetic diversity is important to promote species persistence (le Corre & Kremer 2012; Balkenhol *et al.* 2016).

5. The Cabrera Vole

The Cabrera vole (*Microtus cabrerae*) is a microtine rodent endemic to the Iberian Peninsula and considered Near Threatened by the IUCN Red List of Threatened Species (Fernandes *et al.* 2008a). This species is also highly distinctive compared with other extant members of *Microtus*, including in its karyotype (giant sex chromosomes and multiple copies of male sex-determining genes in females), molar morphology and life-history traits (*K* reproductive strategy, small litters and large body size)(Jiménez *et al.* 1991; Bullejos *et al.* 1997; Ventura *et al.* 1998; Fernández-Salvador *et al.* 2001; Cuenca-Bescós *et al.* 2014). Ecologically, the Cabrera vole is limited to damp meadows and perennial grassland, often near small water courses and temporary ponds (Pita *et al.* 2006; Luque-Larena & López 2007).

Past and present distribution

The Cabrera vole fossil records (first detected around 1 Mya) suggest that its range has been always restricted to the Iberian Peninsula and southern France (Laplana & Sevilla 2013). The species appears to have contracted and expanded its range following the major climatic events of the Quaternary, with periods of lower and higher abundance in the glacial and interglacial periods, respectively (López-García & Cuenca-Bescós 2012; Garrido-García & Soriguer-Escofet 2012; Laplana & Sevilla 2013; Pita *et al.* 2014). The peak of Cabrera vole fossil abundance was observed during the Neolithic, where it is assumed that the transformation of the landscape into arable lands might have promoted the expansion of the species (Laplana & Sevilla 2013). This is also supported by a recent ancient DNA (aDNA) study from Varela (2016) who found signals of population expansion during that time period using partial cytochrome-*b* sequences. Since the Neolithic, the species is thought to have declined (especially since the Middle Ages) on the basis of an observed decrease in genetic diversity (from aDNA) and decrease in fossil records (Garrido-García & Soriguer-Escofet 2012; Laplana & Sevilla 2013; Varela 2016). The problem of basing inferences of population patterns on the fossil record and its DNA is illustrated by Figure 9, which shows that the Cabrera fossil record may be biased in terms of likelihood of fossilisation, being more representative of the species distribution in eastern Iberia and southern France than in western areas of the Iberian Peninsula.

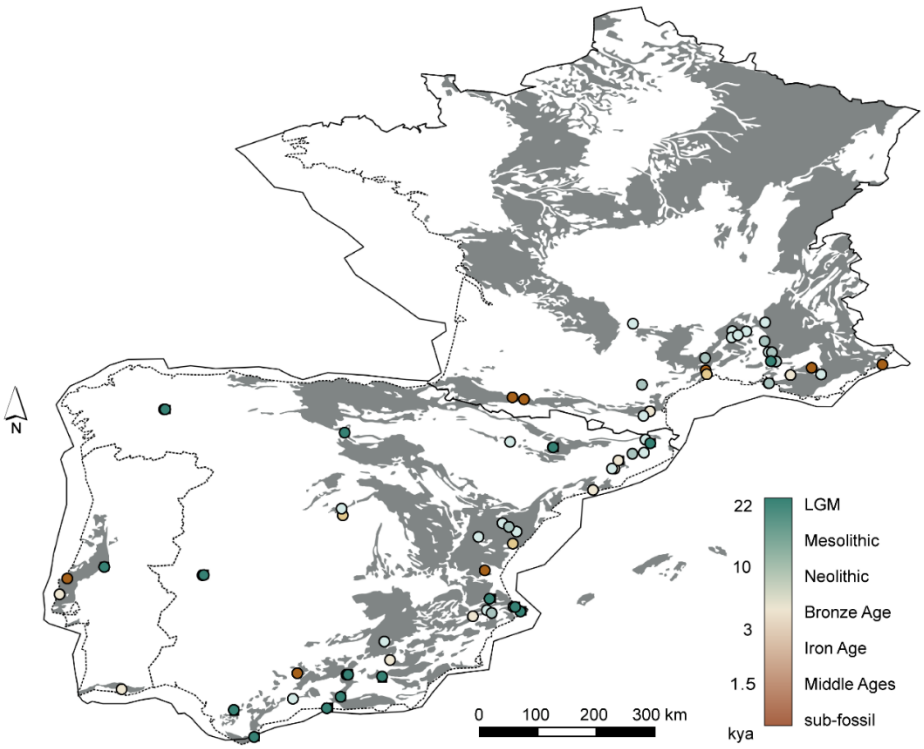


Figure 9. Map of the Iberian Peninsula and France illustrating fossil data for the Cabrera vole (circles) since the Last Glacial Maximum (LGM) to around 500 years ago. Grey shading represents karst areas known to be well-suited for fossilisation (http://web.env.auckland.ac.nz/our_research/karst/); circles represent fossils and their colour the time period to which they were assigned based on the legend on the right side of the coloured bar; the times on the left side of the bar represent estimates of the respective time periods [fossil data from Laplana & Sevilla (2013)].

Today the Cabrera vole is distributed in four main geographic nuclei restricted to the Iberian Peninsula: the *Lusocarpetan* (*Lc*), *Montiberic* (*Mb*), *Betic* (*Bt*) and *pre-Pyrenean* (*pP*) (Figure 10) (Garrido-García *et al.* 2013; Laplana & Sevilla 2013). The Cabrera vole does not have a continuous distribution within these nuclei, which rather is spatially structured and shows metapopulational dynamics, with colonisation-extinction-recolonisation of available habitat patches both through time and space (Pita *et al.* 2007). A recent study from Mestre *et al.* (2015) using ecological niche modelling and non-invasive sampling suggested that many edge populations may not yet be identified, and points out that an Iberian census is needed for an accurate assessment of the current distribution of this species. Additionally, there are other studies suggesting the existence of obsolete records, which may be upward biasing the abundance of this species in other regions, especially in eastern Iberia (Garrido-García *et al.* 2008; Escribano *et al.* 2016).

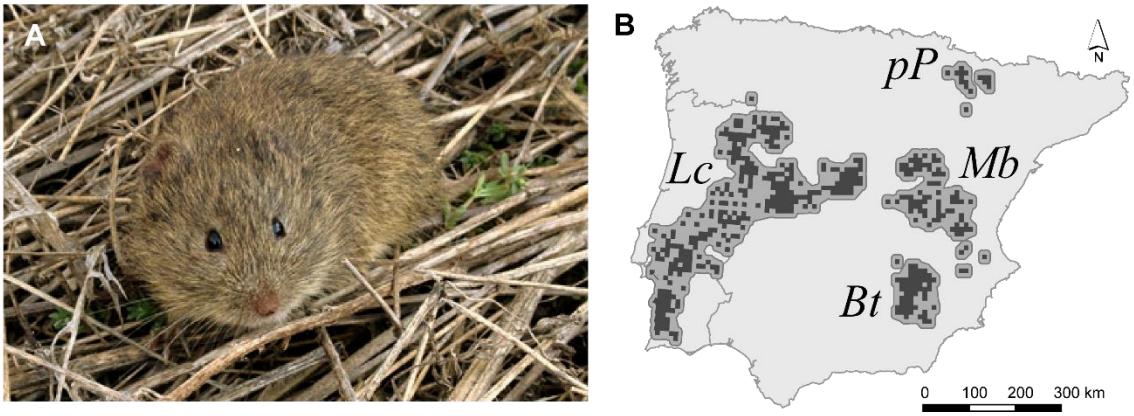


Figure 10. (A) Cabrera vole (*Microtus cabrerae*) female from Yecla, Murcia, Spain. Photograph by Soraia Barbosa. (B) Distribution of the Cabrera vole in the Iberian Peninsula: current confirmed distribution of the Cabrera vole is represented by dark grey 10x10 km UTM squares, where the species has been detected either by live trapping or non-invasive sampling; inferred distribution is represented by a grey 10 km buffer around those UTM squares. Geographic nuclei are identified according to Garrido-García *et al.* (2013): *Lusocarpetan (Lc), Montiberic (Mb), Betic (Bt)* and *pre-Pyrenean (pP)*.

Threats

The assessment of the Cabrera vole as Near-threatened by the IUCN is related to its limited distribution that is potentially in decline (Fernandes *et al.* 2008a). This species is highly restricted to damp marginal habitats, which makes it especially susceptible to habitat loss due to conversion to agriculture or pasture (Pita *et al.* 2006; Garrido-García *et al.* 2013). In addition, since its range in the Iberian Peninsula is at risk of desertification and drought under climate change, the persistence of the Cabrera vole may be even more threatened (Figure 11). The Cabrera vole has thermoregulatory mechanisms adapted to specific bioclimatic zones, which may lead to a population-specific adaptation capacity to climate change (Mathias *et al.* 2003; Castellanos-Frías *et al.* 2015). This can further influence the ability of populations to cope with climate change depending on how each of these groups adjust to new abiotic and biotic stressors (Pita *et al.* 2010, 2011).

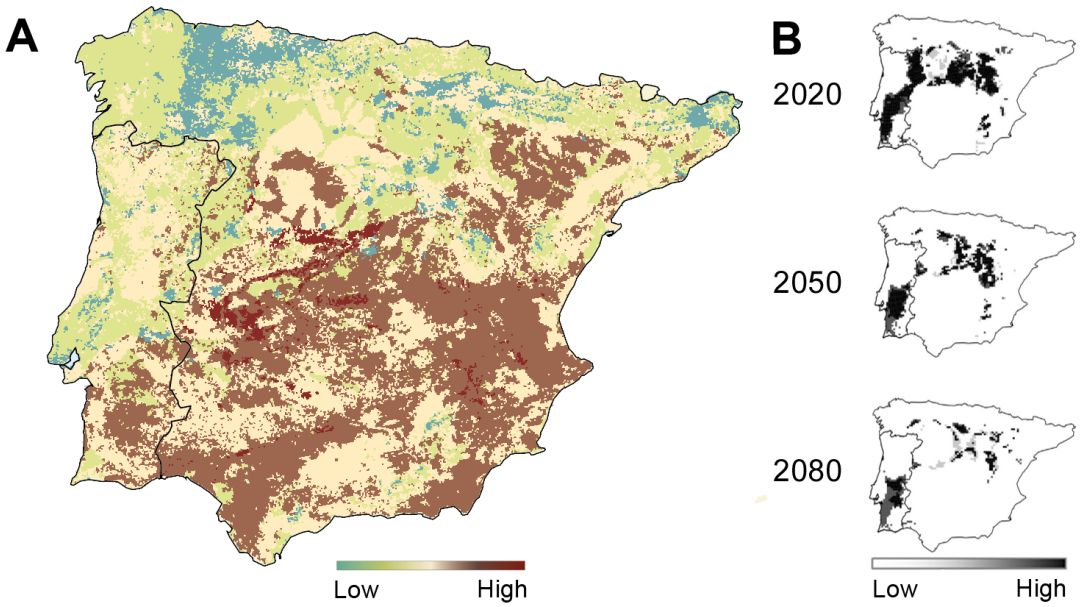


Figure 11. (A) map from the DISMED project (Desertification Information System for the Mediterranean) showing the sensitivity (from low to high) to desertification and drought based on soil quality, climate and vegetation parameters [© European Environment Agency, 2008 (http://www.eea.europa.eu/legal/copyright), with permission]. (B) projections of the Cabrera vole distribution for three 30 year intervals, and with probability of occurrence from low (white) to high (black) (Mestre *et al.* 2015).

Thus, to enhance the conservation efforts for the current populations of the Cabrera vole, below we describe the objectives of this thesis in line with several conservation oriented goals for this species.

6. Objectives and thesis framework

Genetics is widely accepted as a critical component of conservation. With increasing rates of species extinction, it is essential to evaluate the past and the current genetic diversity of species to support effective conservation planning. Although the Cabrera vole (*Microtus cabrerae*) is a near-threatened rodent, endemic to the Iberian Peninsula and has a relatively restricted area, very little is known in terms of its population genetic structure across its range. Thus, the main goal of this thesis is to better understand the evolutionary history of the Cabrera vole and to develop genetic evidence-based conservation actions for protecting the current genetic diversity and future evolutionary potential of the species. This thesis is organised into five chapters, and its workflow is presented below, also showing the main subjects and objectives of each chapter (Figure 12).

The first (present) chapter entitled *Chapter I General Introduction* presents an overview of the emergence of conservation biology and the need to conserve species, with a particular emphasis on the effects of genetic depletion in natural populations and highlighting the role of genetics in conservation. With the arrival of new sequencing technologies, we are now observing a shift to conservation genomics. These techniques provide new tools to address old questions, but they also allow us to make new enquiries, especially related with adaptation, for a better understanding of the role of evolutionary processes in species conservation. However, to preserve a species through future environmental change, it is essential to understand what processes influenced its populations in the past, so this was also addressed in this chapter. All these conceptual elements are brought into perspective for the study of the evolutionary history and conservation of the Cabrera vole.

To accomplish such a detailed study on the genetics of threatened species it is important to have a set of molecular tools that help researchers in obtaining data relevant to answer a wide range of questions, from species identification, to population structure, dispersal and kinship. There are often constraints as to how samples can be collected, and for small mammals it is critical to limit the extent to which individuals are live-trapped and handled. Thus, the second chapter, entitled *Chapter II Molecular conservation tools*, is composed of one manuscript that is aimed at using non-invasive samples for the study of small mammals, providing the opportunity for contemporary genetic sampling of species such as the Cabrera vole, without risking individuals and populations. The paper focuses on species identification, a necessary first step in conservation measures:

Paper I. Barbosa S, Paupério J, Searle JB and Alves, PC (2013) **Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to non-invasive sampling**. *Molecular Ecology Resources*, **13**, 43-56.

Conservation measures should be based on accurate information on species biology, and in *Chapter III* we deepen our knowledge of the *evolutionary history of the Cabrera vole*. For that, we reassess the phylogeny of the genus *Microtus* using genomic tools in order to i) create a well-supported phylogeny for the genus and discuss the evolutionary processes that have led to the current observed relationships, and ii) discuss the phylogenetic positioning of *M. cabrerae* which underlie controversies about its taxonomy. In the second study of this chapter we assess the species' phylogeographic history and evaluate the species genetic structure. This chapter thus comprises the following two papers:

Paper II. Barbosa S, Herman JS, White TA, Paupério J, Alves PC and Searle JB (*in prep*) **Phylogenomic analysis of the *Microtus* genus: insights into inter- and intraspecific relationships**.

Paper III. Barbosa S, Paupério J, Herman JS, Ferreira CM, Pita R, Vale-Gonçalves HM, Cabral JA, Garrido-García JA, Soriguer RC, Beja P, Mira A, Alves PC and Searle JB (2017) **Endemic species may have complex histories: within-refugium phylogeography of an endangered Iberian vole**. *Molecular Ecology*, **26**, 951-967.

With a better understanding of the Cabrera's vole evolutionary history, in *Chapter IV* we use a landscape genomic approach to support *conservation genomics planning* in the Cabrera vole. The goals of the first study were to define conservation units for this species and determine the level of connectivity observed between them, as well as to detect population isolation and to prioritise populations for conservation actions. We additionally used these new data to test the reliability of using genomic tools with non-invasive samples. Thus this chapter is formed of the following two articles:

Paper IV. Barbosa S, Mestre F, White TA, Paupério J, Alves PC and Searle JB (*submitted*) **Integrative approaches to guide conservation decisions: using genomics to define conservation units and functional corridors**.

Paper V. Barbosa S, Paupério J, Mitchell S, Alves PC and Searle JB (*in prep*) **Non-invasive population genomics: applying genotyping-by-sequencing to small mammal conservation**.

In the final chapter entitled *Chapter V General Discussion,* all the theoretical background is brought together with the research developed in this thesis, which are then considered for conservation recommendations. Finally we present the main conclusions of this thesis and future perspectives provided by this work.

This thesis was integrated into two projects supported by the Fundação para a Ciência e Tecnologia (Portugal) that aimed to understand the role of matrix permeability in the dispersal between habitat patches of the Cabrera vole and southern water vole (including the interactions between the two of them) at a local scale: “PERSIST – The role of matrix permeability for metapopulation PERSISTence in complex agricultural landscapes. The cases of Cabrera (*Microtus cabrerae*) and southern water voles (*Arvicola sapidus*)” (PTDC/BIA-BEC/105110/2008) and “NETPERSIST – NETwork analysis of critical connectivity thresholds for metapopulations PERSISTence in complex agricultural landscapes” (PTDC/AAG-MAA/3227/2012).

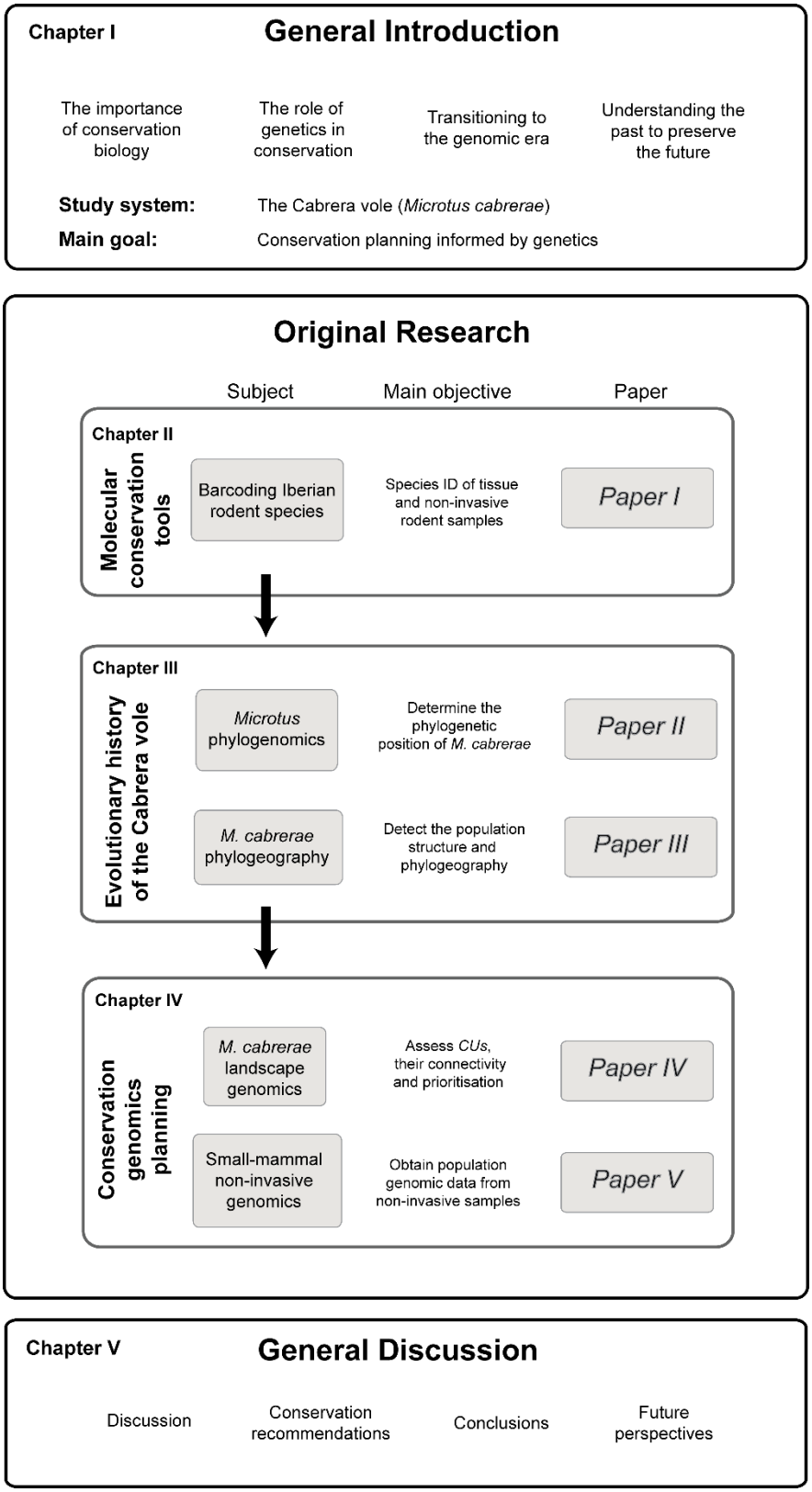


Figure 12. Thesis workflow divided in General Introduction (*Chapter I*), Original Research (*Chapters II*, *III* and *IV*), and General Discussion (*Chapter V*).

7. References

Abascal F, Corvelo A, Cruz F *et al.* (2016) Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology*, **17**, 251.

Abellán P, Svenning J-C (2014) Refugia within refugia – patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. *Biological Journal of the Linnean Society*, **113**, 13–28.

Allendorf FW (2017) Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology*, **26**, 420–430.

Allendorf FW, Luikart G, Aitken SN (2012) *Conservation and the Genetics of Populations*. Wiley-Blackwell, Oxford, UK.

Amarasekare P (1998) Allee effects in metapopulation dynamics. *The American Naturalist*, **152**, 298–302.

Amos W, Balmford A (2001) When does conservation genetics matter? *Heredity*, **87**, 257–265.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.

Andrews KR, Luikart G (2014) Recent novel approaches for population genomics data analysis. *Molecular Ecology*, **23**, 1661–1667.

Avenant NL (2005) Barn owl pellets: a useful tool for monitoring small mammal communities? *Belgian Journal of Zoology*, **135**, 39–43.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA, USA.

Baillie J, Hilton-Taylor C, Stuart SN (2004) *IUCN Red List of Threatened Species: A Global Species Assessment*. IUCN Publication Services Unit, Cambridge, UK.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, **3**, e3376.

Balkenhol N, Cushman SA, Storfer AT, Waits LP (2016) *Landscape Genetics: Concepts, Methods and Application*. Wiley-Blackwell, Oxford, UK.

de Barba M, Miquel C, Lobréaux S *et al.* (2016) High-throughput microsatellite

genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, **17**, 492–507.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**, 38–44.

Bastos-Silveira C, Santos SM, Monarca R, Mathias ML, Heckel G (2012) Deep mitochondrial introgression and hybridization among ecologically divergent vole species. *Molecular Ecology*, **21**, 5309–5323.

Beck JB, Semple JC (2015) Next-generation sampling: pairing genomics with herbarium specimens provides species-level signal in *Solidago* (Asteraceae). *Applications in Plant Sciences*, **3**, 1500014.

Beever EA, O'Leary J, Mengelt C *et al.* (2016) Improving conservation outcomes with a new paradigm for understanding species' fundamental and realized adaptive capacity. *Conservation Letters*, **9**, 131–137.

Beissinger SR, McCullough DR (2002) *Population Viability Analysis*. University of Chicago Press, Chicago, IL, USA.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, **9**, 1279–1301.

Benestan LM, Ferchaud AL, Hohenlohe PA *et al.* (2016) Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Molecular Ecology*, **25**, 2967–2977.

Bennett DJ, Sutton MD, Turvey ST (2017) Evolutionarily distinct "living fossils" require both lower speciation and lower extinction rates. *Paleobiology*, **43**, 34–48.

Benton MJ (2009) The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science*, **728**, 728–732.

Berec L, Angulo E, Courchamp F (2007) Multiple Allee effects and population management. *Trends in Ecology and Evolution*, **22**, 185–191.

Bi K, Linderoth T, Vanderpool D *et al.* (2013) Unlocking the vault: next-generation museum population genomics. *Molecular Ecology*, **22**, 6018–6032.

Bilton DT, Mirol PM, Mascheretti S *et al.* (1998) Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London B*, **265**, 1219–1226.

Blair ME, Melnick DJ (2012) Scale-dependent effects of a heterogeneous landscape on genetic differentiation in the Central American squirrel monkey (*Saimiri oerstedii*). *PLoS One*, **7**, e43027.

Bolliger J, Lander T, Balkenhol N (2014) Landscape genetics since 2003: status, challenges and future directions. *Landscape Ecology*, **29**, 361–366.

Bonn A, Rodrigues ASL, Gaston KJ (2002) Threatened and endemic species: are they good indicators of patterns of biodiversity on a national scale? *Ecology Letters*, **5**, 733–741.

Brawand D, Wagner CE, Li YI *et al.* (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375–381.

Broquet T, Ménard N, Petit EJ (2007) Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics*, **8**, 249–260.

Bullejos M, Sánchez A, Burgos M *et al.* (1997) Multiple, polymorphic copies of *SRY* in both males and females of the vole *Microtus cabrerae*. *Cytogenetics and Cell Genetics*, **79**, 167–171.

Cahill AE, Aiello-Lammens ME, Fisher-Reid MC *et al.* (2012) How does climate change cause extinction? *Proceedings of the Royal Society B*, **280**, 20121890.

Caro TM (2003) Umbrella species: critique and lessons from East Africa. *Animal Conservation*, **6**, 171–181.

Castellanos-Frías E, García-Perea R, Gisbert J, Bozinovic F, Virgós E (2015) Intraspecific variation in the energetics of the Cabrera vole. *Comparative Biochemistry and Physiology, Part A*, **190**, 32–38.

Caughley G (1994) Directions in conservation biology. *Journal of Animal Ecology*, **63**, 215–244.

Ceballos G, Ehrlich PR (2002) Mammal population losses and the extinction crisis. *Science*, **296**, 904–907.

Ceballos G, Ehrlich PR, Barnosky AD *et al.* (2015) Accelerated modern human-induced species losses: entering the sixth mass extinction. *Science Advances*, **1**, e1400253.

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.

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. *Nature Reviews Genetics*, **10**, 783–796.

Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD (2011) Rapid range shifts of species associated with high levels of climate warming. *Science*, **333**, 1024–1026.

Chevin L-M, Lande R, Mace GM (2010) Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biology*, **8**, e1000357.

Clavel J, Julliard R, Devictor V (2011) Worldwide decline of specialist species: toward a global functional homogenization? *Frontiers in Ecology and the Environment*, **9**, 222–228.

Corlett RT (2016) Restoration, reintroduction, and rewilding in a changing world. *Trends in Ecology and Evolution*, **31**, 453–462.

Corlett RT (2017) A bigger toolbox: biotechnology in biodiversity conservation. *Trends in Biotechnology*, **35**, 55–65.

le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548–1566.

Costa V, Rosenbom S, Monteiro R, O'Rourke SM, Beja-Pereira A (2017) Improving DNA quality extracted from fecal samples - a method to improve DNA yield. *European Journal of Wildlife Research*, **63**, 3.

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

Cuenca-Bescós G, López-García JM, Galindo-Pellicena MA *et al.* (2014) The Pleistocene history of *Iberomys*, an endangered endemic rodent from south western Europe. *Integrative Zoology*, **9**, 481–497.

Darwin CR (1857) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. J. Murray, London, UK.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM (2011) Beyond predictions: biodiversity conservation in a changing climate. *Science*, **332**, 53–58.

DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Reviews*

Genetics, **5**, 702–712.

Dobrynin P, Liu S, Tamazian G *et al.* (2015) Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biology*, **16**, 277.

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

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

Eckert AJ, Van Heerwaarden J, Wegrzyn JL *et al.* (2010) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*, **185**, 969–982.

Eizaguirre C, Baltazar-Soares M (2014) Evolutionary conservation - evaluating the adaptive potential of species. *Evolutionary Applications*, **7**, 963–967.

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, **7**, 1026–1042.

Ellegren H, Galtier N (2016) Determinants of genetic diversity. *Nature Reviews Genetics*, **17**, 422–433.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, e19379.

Escribano N, Ariño AH, Galicia D (2016) Biodiversity data obsolescence and land uses changes. *PeerJ*, **4**, e2743.

Fagan WF, Holmes EE (2006) Quantifying the extinction vortex. *Ecology Letters*, **9**, 51–60.

Fauvergue X, Vercken E, Malausa T, Hufbauer RA (2012) The biology of small, introduced populations, with special reference to biological control. *Evolutionary Applications*, **5**, 424–443.

Fernandes M, Pita R, Mira APP (2008) *Microtus cabrerae*. *The IUCN Red List of Threatened Species 2008: e.T13418A3915236. Downloaded on 06 May 2017. http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T13418A3915236.en.*

Fernández-Salvador R, García-Perea R, Ventura J (2001) Reproduction and postnatal growth of the Cabrera vole, *Microtus cabrerae*, in captivity. *Canadian Journal of*

Zoology, **79**, 2080–2085.

Flather CH, Hayward GD, Beissinger SR, Stephens PA (2011) Minimum viable populations: is there a "magic number" for conservation practitioners? *Trends in Ecology and Evolution*, **26**, 307–316.

da Fonseca RR, Albrechtsen A, Themudo GE *et al.* (2016) Next-generation biology: sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, **30**, 3–13.

Frankel OH (1974) Genetic conservation: our evolutionary responsibility. *Genetics*, **78**, 53–65.

Frankel OH, Bennet E (1970) *Genetic Resources in Plants - Their Exploration and Conservation*. Blackwell Scientific Publications, Oxford, UK.

Frankel O, Soulé ME (1981) *Conservation and Evolution*. Cambridge University Press, Cambridge, UK.

Frankham R, Ballou JD, Briscoe DA (2010) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK.

Frankham R, Ballou JD, Dudash MR *et al.* (2012) Implications of different species concepts for conserving biodiversity. *Biological Conservation*, **153**, 25–31.

Fullerton AH, Lindley ST, Pess GR *et al.* (2011) Human influence on the spatial structure of threatened Pacific salmon metapopulations. *Conservation Biology*, **25**, 932–944.

Fumagalli L, Taberlet P, Stewart DT *et al.* (1999) Molecular phylogeny and evolution of *Sorex* shrews (Soricidae: Insectivora) inferred from mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, **11**, 222–235.

Garrido-García JA, Rosário IT, Gisbert J *et al.* (2013) Revisión a nivel ibérico de la distribución del topillo de Cabrera o iberón, *Iberomys cabrerae* (Thomas, 1906). *Galemys*, **25**, 35–49.

Garrido-García JA, Soriguer-Escofet RC (2012) Cabrera's vole *Microtus cabrerae* Thomas, 1906 and the subgenus *Iberomys* during the Quaternary: evolutionary implications and conservation. *Geobios*, **45**, 437–444.

Garrido-García JA, Soriguer RC, Serrano DP-A, Pardávila X (2008) A revision of the distribution of Cabrera's vole (*Microtus cabrerae* Thomas 1906) in Andalusia (southern Spain). *Hystrix, Italian Journal of Mammalogy*, **19**, 121–131.

Ghalambor CK, Hoke KL, Ruell EW *et al.* (2015) Non-adaptive plasticity potentiates rapid

adaptive evolution of gene expression in nature. *Nature*, **525**, 372–375.

Ghalambor CK, McKay JK, Carroll SP, Reznick DN (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, **21**, 394–407.

Giarla TC, Esselstyn JA (2015) The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology*, **64**, 727–740.

Gilpin M, Hanski I (1991) *Metapopulation Dynamics: Empirical and Theoretical Investigations*. Academic Press, San Diego, CA, USA.

Gilpin ME, Soulé ME (1986) Minimum Viable Populations: processes of species extinction. In: *Conservation Biology: The Science of Scarcity and Diversity* (ed Soulé ME), pp. 19–34. Sinauer Associates, Sunderland, MA, USA.

Golley FB (1993) *A History of the Ecosystem Concept in Ecology: More than the Sum of the Parts*. Yale University Press, New Haven, CT, USA.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 155–188. Springer, Dordrecht, The Netherlands.

Graham CF, Glenn TC, McArthur AG *et al.* (2015) Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, **15**, 1304–1315.

Habel JC, Assmann T (2009) *Relict Species: Phylogeography and Conservation Biology*. Springer, New York, NY, USA.

Haig SM, Miller MP, Bellinger R *et al.* (2016) The conservation genetics juggling act: integrating genetics and ecology, science and policy. *Evolutionary Applications*, **9**, 181–195.

Hallam A, Wignall PB (1997) *Mass Extinctions and their Aftermath*. Oxford University Press, Oxford, UK.

Hampe A, Jump AS (2011) Climate relicts: past, present, future. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 313–333.

Hanski I (1998) Metapopulation dynamics. *Nature*, **396**, 41–49.

Hanski I, Gaggiotti OE (2004) *Ecology, Genetics, and Evolution of Metapopulations*.

Elsevier Academic Press, Burlington, MA, USA.

Harrison S, Taylor AD (1997) Empirical evidence for metapopulation dynamics. In: *Metapopulation Biology: Ecology, Genetics, and Evolution* (eds Hanski I, Gilpin M), pp. 27–42. Academic Press, San Diego, CA, USA.

Hedrick PW, Fredrickson R (2010) Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conservation Genetics*, **11**, 615–626.

Hedrick PW, Garcia-Dorado A (2016) Understanding inbreeding depression, purging, and genetic rescue. *Trends in Ecology and Evolution*, **31**, 940-952.

Hedrick PW, Kardos M, Peterson RO, Vucetich JA (2017) Genomic variation of inbreeding and ancestry in the remaining two Isle Royale wolves. *Journal of Heredity*, **108**, 120–126.

Hedrick PW, Peterson RO, Vucetich LM, Adams JR, Vucetich JA (2014) Genetic rescue in Isle Royale wolves: genetic analysis and the collapse of the population. *Conservation Genetics*, **14**, 1111–1121.

Henry P, Henry A, Russello MA (2011) A noninvasive hair sampling technique to obtain high quality DNA from elusive small mammals. *Journal of Visualized Experiments*, **49**, e2791.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

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

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

Hewitt GM (2011) Mediterranean peninsulas: the evolution of hotspots. In: *Biodiversity Hotspots* (eds Zachos FE, Habel JC), pp. 123–147. Springer, Berlin, Germany.

Hey J (1992) Using phylogenetic trees to study speciation and extinction. *Evolution*, **46**, 627–640.

Hill JK, Griffiths HM, Thomas CD (2011) Climate change and evolutionary adaptations at species' range margins. *Annual Review of Entomology*, **56**, 143–159.

Hoffmann AA, Griffin P, Dillon S *et al.* (2015) A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses*, **2**, 1–23.

Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–485.

Hogg JT, Forbes SH, Steele BM, Luikart G (2006) Genetic rescue of an insular population of large mammals. *Proceedings of the Royal Society B*, **273**, 1491–1499.

Holmes MW, Hammond TT, Wogan GOU *et al.* (2016) Natural history collections as windows on evolutionary processes. *Molecular Ecology*, **25**, 864–881.

Igea J, Aymerich P, Fernández-González A *et al.* (2013) Phylogeography and postglacial expansion of the endangered semi-aquatic mammal *Galemys pyrenaicus*. *BMC Evolutionary Biology*, **13**, 115.

IUCN (2001) *IUCN Red List Categories and Criteria: Version 3.1. IUCN Species Survival Commission*. IUCN, Gland, Switzerland.

IUCN (2016) The IUCN Red List of Threatened Species. Version 2016-3. *Downloaded on 07 Dec 2016. <http://www.iucnredlist.org>*.

Jiménez R, Carnero A, Burgos M, Sánchez A, Díaz de la Guardia R (1991) Achiasmatic giant sex chromosomes in the vole *Microtus cabrerae* (Rodentia, Microtidae). *Cytogenetics and Cell Genetics*, **57**, 56–58.

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

Keller D, Holderegger R, van Strien MJ, Bolliger J (2015) How to make landscape genetics beneficial for conservation management? *Conservation Genetics*, **16**, 503–512.

Lambeck RJ (1997) Focal species: a multi-species umbrella for nature conservation. *Conservation Biology*, **11**, 849–856.

Laplana C, Sevilla P (2013) Documenting the biogeographic history of *Microtus cabrerae* through its fossil record. *Mammal Review*, **43**, 309–322.

Latta RG (2008) Conservation genetics as applied evolution: from genetic pattern to evolutionary process. *Evolutionary Applications*, **1**, 84–94.

Lee PLM (2017) DNA amplification in the field: move over PCR, here comes LAMP. *Molecular Ecology Resources*, **17**, 138–141.

Leitwein M, Garza JC, Pearse DE (2016) Ancestry and adaptive evolution of anadromous, resident, and adfluvial rainbow trout (*Oncorhynchus mykiss*) in the San Francisco bay area: application of adaptive genomic variation to conservation

in a highly impacted landscape. *Evolutionary Applications*, **10**, 56–67.

Leopold A (1991) *The River of the Mother of God and Other Essays by Aldo Leopold* (SL Flader, JB Callicott, Eds,). University of Wisconsin Press, Madison, WI, USA.

Levy S (2013) Landscape genetics offers new insights: genetics and geography help inform species conservation. *BioScience*, **63**, 613–618.

Loeschcke V, Tomiuk J, Jain SK (1994) *Conservation Genetics*. Springer, Basel, Switzerland.

López-García JM, Cuenca-Bescós G (2012) Changes in the geographical distribution of *Microtus (Iberomys) cabrerae* (Thomas, 1906) from the Late Pleistocene to the Holocene. *Spanish Journal of Palaentology*, **27**, 117–124.

Luque-Larena JJ, López P (2007) Microhabitat use by wild-ranging Cabrera voles *Microtus cabrerae* as revealed by live trapping. *European Journal of Wildlife Research*, **53**, 221–225.

Luque GM, Vayssade C, Facon B *et al.* (2016) The genetic Allee effect: a unified framework for the genetics and demography of small populations. *Ecosphere*, **7**, e01413.

Lv X, Xia L, Ge D, Wu Y, Yang Q (2016) Climatic niche conservatism and ecological opportunity in the explosive radiation of arvicoline rodents (Arvicolinae, Cricetidae). *Evolution*, **70**, 1094–1104.

Madsen T, Shine R, Olsson M, Wittzell H (1999) Conservation biology: restoration of an inbred adder population. *Nature*, **402**, 34–35.

Malcolm JR, Liu C, Neilson RP, Hansen L, Hannah L (2006) Global warming and extinctions of endemic species from biodiversity hotspots. *Conservation Biology*, **20**, 538–548.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology and Evolution*, **28**, 614–621.

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

Mann ME, Zhang Z, Hughes MK *et al.* (2008) Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences USA*, **105**, 13252–13257.

Matala AP, Ackerman MW, Campbell MR, Narum SR (2014) Relative contributions of neutral and non-neutral genetic differentiation to inform conservation of steelhead trout across highly variable landscapes. *Evolutionary Applications*, **7**, 682–701.

Mathias ML da L, Klunder M, Santos SM (2003) Metabolism and thermoregulation in the Cabrera vole (Rodentia: *Microtus cabrerae*). *Comparative Biochemistry and Physiology Part A*, **136**, 441–446.

McDevitt AD, Zub K, Kawałko A *et al.* (2012) Climate and refugial origin influence the mitochondrial lineage distribution of weasels (*Mustela nivalis*) in a phylogeographic suture zone. *Biological Journal of the Linnean Society*, **106**, 57–69.

McNeely JA, Miller KR, Reid W V, Mittermeier RA, Werner TB (1990) *Conserving the World's Biological Diversity*. IUCN, Gland, Switzerland.

Mestre F, Pita R, Paupério J *et al.* (2015) Combining distribution modelling and non-invasive genetics to improve range shift forecasting. *Ecological Modelling*, **297**, 171–179.

Mills LS (2013) *Conservation of Wildlife Populations: Demography, Genetics, and Management*. Wiley-Blackwell, Oxford, UK.

Mittelbach GG, Schemske DW, Cornell H V *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

Moritz C, Patton JL, Conroy CJ *et al.* (2008) Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science*, **322**, 261–264.

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

Nabholz B, Glémin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals - the longevity hypothesis. *Molecular Biology and Evolution*, **25**, 120–130.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841–2847.

Nee S, May RM (1992) Dynamics of metapopulations: habitat destruction and competitive coexistence. *Journal of Animal Ecology*, **61**, 37–40.

Neukom R, Gergis J, Karoly DJ *et al.* (2014) Inter-hemispheric temperature variability over the past millennium. *Nature Climate Change*, **4**, 362–367.

Nicolas V, Martínez-Vargas J, Hugot J-P (2017) Molecular data and ecological niche modelling reveal the evolutionary history of the common and Iberian moles (Talpidae) in Europe. *Zoologica Scripta*, **46**, 12–26.

Ohlemüller R, Anderson BJ, Araújo MB *et al.* (2008) The coincidence of climatic and species rarity: high risk to small-range species from climate change. *Biology Letters*, **4**, 568–572.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. *Trends in Genetics*, **26**, 177–187.

Palomo LJ, Gisbert J, Blanco JC (2007) *Atlas y Libro Rojo de los Mamíferos Terrestres de España*. Dirección General para la Biodiversidad-SECEM-SECEM, Madrid, Spain.

Paplinska JZ, Taggart DA, Corrigan T, Eldridge MDB, Austin JJ (2011) Using DNA from museum specimens to preserve the integrity of evolutionarily significant unit boundaries in threatened species. *Biological Conservation*, **144**, 290–297.

Paupério J, Herman JS, Melo-Ferreira J *et al.* (2012) Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia. *Molecular Ecology*, **21**, 6015–6032.

Pennisi E (2016) Pocket DNA sequencers make real-time diagnostics a reality. *Science*, **351**, 800–801.

Pérez-Espona S, ConGRESS-Consortium (2017) Conservation genetics in the European Union – biases, gaps and future directions. *Biological Conservation*, **209**, 130–136.

Perry GH (2013) The promise and practicality of population genomics research with endangered species. *International Journal of Primatology*, **35**, 55–70.

Perry GH, Marioni JC, Melsted P, Gilad Y (2010) Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology*, **19**, 5332–5344.

Peterson RO, Page RE (1988) The rise and fall of Isle Royale wolves, 1975-1986. *Journal of Mammalogy*, **69**, 89–99.

Pita R, Beja P, Mira APP (2007) Spatial population structure of the Cabrera vole in Mediterranean farmland: the relative role of patch and matrix effects. *Biological*

Conservation, **134**, 383–392.

Pita R, Mira A, Beja P (2006) Conserving the Cabrera vole, *Microtus cabrerae*, in intensively used Mediterranean landscapes. *Agriculture, Ecosystems and Environment*, **115**, 1–5.

Pita R, Mira APP, Beja P (2010) Spatial segregation of two vole species (*Arvicola sapidus* and *Microtus cabrerae*) within habitat patches in a highly fragmented farmland landscape. *European Journal of Wildlife Research*, **56**, 651–662.

Pita R, Mira APP, Beja P (2011) Assessing habitat differentiation between coexisting species: the role of spatial scale. *Acta Oecologica*, **37**, 124–132.

Pita R, Mira APP, Beja P (2014) *Microtus cabrerae* (Rodentia: Cricetidae). *Mammalian Species*, **912**, 48–70.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–846.

Prosser SWJ, Dewaard JR, Miller SE, Hebert PDN (2016) DNA barcodes from century-old type specimens using next-generation sequencing. *Molecular Ecology Resources*, **16**, 487–497.

Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, **23**, 564–571.

Querejeta M, González-Esteban J, Gómez A *et al.* (2016) Genomic diversity and geographical structure of the Pyrenean desman. *Conservation Genetics*, **17**, 1333–1344.

Randi E (2007) Phylogeography of South European mammals. In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 101–126. Springer, Dordrecht, The Netherlands.

Reed TE, Schindler DE, Waples RS (2011) Interacting effects of phenotypic plasticity and evolution on population persistence in a changing climate. *Conservation Biology*, **25**, 56–63.

Rehnus M, Bollmann K (2016) Non-invasive genetic population density estimation of mountain hares (*Lepus timidus*) in the Alps: systematic or opportunistic sampling? *European Journal of Wildlife Research*, **62**, 737–747.

Richardson JL, Brady SP, Wang IJ, Spear SF (2016) Navigating the pitfalls and promise of landscape genetics. *Molecular Ecology*, **25**, 849–863.

Roberge J-M, Angelstam P (2004) Usefulness of the umbrella species concept as a conservation tool. *Conservation Biology*, **18**, 76–85.

Robinson JA, Ortega-Del Vecchyo D, Fan Z *et al.* (2016) Genomic flatlining in the endangered island fox. *Current Biology*, **26**, 1183–1189.

Rowe KC, Singhal S, Macmanes MD *et al.* (2011) Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources*, **11**, 1082–1092.

Rubidge EM, Patton JL, Lim M *et al.* (2012) Climate-induced range contraction drives genetic erosion in an alpine mammal. *Nature Climate Change*, **2**, 285–288.

Russello MA, Waterhouse MD, Etter PD, Johnson EA (2015) From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ*, **3**, e1106.

Sabino-Marques H, Ferreira CM, Paupério J *et al.* Combining noninvasive genetics with spatially explicit capture-recapture models for density estimation of a patchily distributed small mammal. *(in prep)*.

Schloss CA, Nuñez TA, Lawler JJ (2012) Dispersal will limit ability of mammals to track climate change in the Western Hemisphere. *Proceedings of the National Academy of Sciences USA*, **109**, 8606–8611.

Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, **15**, 749–763.

Schwartz MK, Luikart G, McKelvey KS, Cushman SA (2010) Landscape genomics: a brief perspective. In: *Spatial Complexity, Informatics, and Wildlife Conservation* (eds Cushman SA, Huettmann F), pp. 165–174. Springer, Tokyo, Japan.

Segelbacher G, Cushman SA, Epperson BK *et al.* (2010) Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics*, **11**, 375–385.

Sgrò CM, Lowe AJ, Hoffmann AA (2011) Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications*, **4**, 326–337.

Shafer ABA, Wolf JBW, Alves PC *et al.* (2015) Genomics and the challenging translation into conservation practice. *Trends in Ecology and Evolution*, **30**, 78–87.

Shaffer M (1987) Minimum Viable Populations: coping with uncertainty. In: *Viable Populations for Conservation* (ed Soulé ME), pp. 68–86. Cambridge University

Press, New York, NY, USA.

Simões M, Breitkreuz L, Alvarado M *et al.* (2016) The evolving theory of evolutionary radiations. *Trends in Ecology and Evolution*, **31**, 27–34.

Smith O, Wang J (2014) When can noninvasive samples provide sufficient information in conservation genetics studies? *Molecular Ecology Resources*, **14**, 1011–1023.

Smith TB, Wayne RK (1996) *Molecular Genetic Approaches in Conservation*. Oxford University Press, New York, NY, USA.

Sodhi NS, Ehrlich PR (2010) *Conservation Biology for All*. Oxford University Press, New York, NY, USA.

Sommer R, Nadachowski A (2006) Glacial refugia of mammals in Europe: evidence from fossil records. *Mammal Review*, **36**, 251–265.

Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics and Genomes*, **9**, 901–911.

Soulé ME (1985) What is conservation biology? *BioScience*, **35**, 727–734.

Soulé ME, Wilcox BA (1980) *Conservation Biology: An Evolutionary-Ecological Perspective*. Sinauer Associates, Sunderland, MA, USA.

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences USA*, **101**, 15261–15264.

Stebbins GL (1942) The genetic approach to problems of rare and endemic species. *Madroño*, **60**, 241–258.

Steiner CC, Putnam AS, Hoeck PEA, Ryder OA (2013) Conservation genomics of threatened animal species. *Annual Review of Animal Biosciences*, **1**, 261–281.

Stephens PA, Sutherland WJ, Freckleton RP (1999) What is the Allee effect? *Oikos*, **87**, 185–190.

Stewart JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B*, **277**, 661–671.

van Strien MJ, Keller D, Holderegger R *et al.* (2014) Landscape genetics as a tool for conservation planning: predicting the effects of landscape change on gene flow.

Ecological Applications, **24**, 327–339.

Suárez-Seoane S, Virgós E, Terroba O, Pardavilla X, Barea-Azcón JM (2013) Scaling of species distribution models across spatial resolutions and extents along a biogeographic gradient. The case of the Iberian mole *Talpa occidentalis*. *Ecography*, **37**, 279–292.

Suh A, Smeds L, Ellegren H (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biology*, **13**, e1002224.

Taberlet P, Fumagalli L (1996) Owl pellets as a source of DNA for genetic studies of small mammals. *Molecular Ecology*, **5**, 301–305.

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

Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution*, **14**, 323–327.

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

Taylor HR, Dussex N, Van Heezik Y (2017) Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Global Ecology and Conservation*, **10**, 231–242.

Teotónio H, Chelo IM, Bradić M, Rose MR, Long AD (2009) Experimental evolution reveals natural selection on standing genetic variation. *Nature Genetics*, **41**, 251–257.

Valladares F, Matesanz S, Guilhaumon F *et al.* (2014) The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecology Letters*, **17**, 1351–1364.

Varela RR (2016) Variaciones en la distribución y la estructura genética de los linces paleárticos y el topillo de Cabrera desde el Pleistoceno Superior. Universidad Complutense de Madrid, Madrid, Spain.

Ventura J, López-Fuster MJ, Cabrera-Millet M (1998) The Cabrera vole, *Microtus cabrerae*, in Spain: a biological and a morphometric approach. *Netherlands Journal of Zoology*, **48**, 83–100.

Waits LP, Paetkau D (2005) Noninvasive genetic sampling tools for wildlife biologists: a

review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, **69**, 1419–1433.

Vander Wal E, Garant D, Festa-Bianchet M, Pelletier F (2013) Evolutionary rescue in vertebrates: evidence, applications and uncertainty. *Philosophical Transactions of the Royal Society B*, **368**, 20120090.

Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution*, **22**, 634–642.

Weeks AR, Stoklosa J, Hoffmann AA (2016) Conservation of genetic uniqueness of populations may increase extinction likelihood of endangered species: the case of Australian mammals. *Frontiers in Zoology*, **13**, 31.

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

Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA (2015) Genetic rescue to the rescue. *Trends in Ecology and Evolution*, **30**, 42–49.

Williams SE, Shoo LP, Isaac JL, Hoffmann AA, Langham G (2008) Towards an integrated framework for assessing the vulnerability of species to climate change. *PLoS Biology*, **6**, e325.

Willoughby JR, Sundaram M, Wijayawardena BK *et al.* (2015) The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biological Conservation*, **191**, 495–503.

Winter M, Devictor V, Schweiger O (2013) Phylogenetic diversity and nature conservation: where are we? *Trends in Ecology and Evolution*, **28**, 199–204.

Wittmann MJ, Stuis H, Metzler D (2016) Genetic Allee effects and their interaction with ecological Allee effects. *Journal of Animal Ecology*, doi: 10.1111/1365-2656.12598.

Yannic G, Basset P, Hausser J (2008) A new perspective on the evolutionary history of western European *Sorex araneus* group revealed by paternal and maternal molecular markers. *Molecular Phylogenetics and Evolution*, **47**, 237–250.

CHAPTER II.

MOLECULAR CONSERVATION TOOLS

Paper I. Barbosa S, Paupério J, Searle JB and Alves, PC (2013) **Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to non-invasive sampling**. *Molecular Ecology Resources*, **13**, 43-56.

Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to non-invasive sampling

Barbosa S, Paupério J, Searle JB and Alves PC

1. Abstract

Species identification through non-invasive sampling is increasingly used in animal conservation genetics, given that it obviates the need to handle free-living individuals. Non-invasive sampling is particularly valuable for elusive and small species such as rodents. Although rodents are not usually assumed to be the most obvious target for conservation, of the 21 species or near-species present in Iberia, three are considered endangered and declining, while several others are poorly studied. Here, we develop a genetic tool for identifying all rodent species in Iberia by non-invasive genetic sampling. To achieve this purpose, we selected one mitochondrial gene [cytochrome *b* (cyt-*b*)] and one nuclear gene [interphotoreceptor retinoid-binding protein (IRBP)], which we first sequenced using tissue samples. Both genes allow for the phylogenetic distinction of all species except the sibling species *Microtus lusitanicus* and *Microtus duodecimcostatus*. Overall, cyt-*b* showed higher resolution than IRBP, revealing a clear barcoding gap. To allow these markers to be applied to non-invasive samples, we selected a short highly diagnostic fragment from each gene, which we used to obtain sequences from faeces and bones from owl pellets. Amplification success for the cyt-*b* and IRBP fragment was 85% and 43% in faecal and 88% and 64% in owl-pellet DNA extractions, respectively. The method allows the unambiguous identification of the great majority of Iberian rodent species from non-invasive samples, with application in studies of distribution, spatial ecology and population dynamics, and for conservation.

2. Introduction

The order Rodentia represents over 40% of all known mammalian species (that is, *c.* 1850 species: IUCN 2011), of which a quarter are under some sort of conservation threat. However, since rodents are better known as pests, the conservation issues relating to many endangered rodent species tend to be overlooked. There is a clear need for more studies, as is illustrated well by Iberian rodents, our study system. In the Iberian Peninsula there are 21 rodent species (Table 1), one of which is considered 'Vulnerable' (*Arvicola sapidus*) and two 'Near Threatened' (*Eliomys quercinus* and *Microtus cabrerae*) in the IUCN Red List of Threatened Species (Bertolino *et al.* 2008; Fernandes *et al.* 2008b). These are statuses which may be valid but which are clearly based on incomplete data (e.g. *E. quercinus* is listed as 'Data Deficient' in Portugal and *A. sapidus* is considered 'Least Concern' in Portugal as a result of poor species documentation, Rigaux *et al.* 2008).

The combination of small body size, elusive behaviour and rareness in rodent species under conservation threat, make monitoring and status assessment a difficult task (e.g. Álvarez-Castañeda & Ortega-Rubio 2003). Rodents are usually not easy to observe directly, and live trapping can require considerable effort and costs for some species (e.g. *Microtus cabrerae*, Alasaad *et al.* 2011) and can be harmful for captured specimens (Moncrief *et al.* 2008). It is, however, possible to reveal the presence of rodent species by non-invasive genetic sampling (NIGS, Moran *et al.* 2008) already applied to other rare and elusive animal species with difficulties in capture and handling (Beja-Pereira *et al.* 2009; Oliveira *et al.* 2010; Henry *et al.* 2011).

NIGS rests on the principle that DNA can be extracted from many sources of animal-derived material which can be obtained without capturing individuals. Faeces are the most obvious non-invasive source of rodent DNA, found scattered throughout an individual's activity range (Beja-Pereira *et al.* 2009; Centeno-Cuadros & Godoy 2010). However, amplification success with rodent faeces is likely to be affected by their age; it is known for carnivore scats that DNA has a higher probability of amplification up to one week of age, after which amplification success decreases significantly (Piggott 2004).

Owl pellets are another important source of non-invasive rodent material, providing bones that can be used for both morphological (Avenant, 2005) and genetic analysis (Poulakakis *et al.* 2005). Hair from small mammals can also be used for NIGS through various sampling strategies, such as hair tubes or duct tape webs (Harris & Nicol 2010; Henry *et al.* 2011; Pocock & Jennings 2006). Because it is difficult to distinguish rodent species from hair morphology (Teerink 1991; Bertolino *et al.* 2009), DNA extracted from

hair can be used for both species identification and population genetic analysis (Centeno-Cuadros & Godoy 2010; Oliveira *et al.* 2010; Henry *et al.* 2011). Overall, NIGS has great potential for rodent monitoring programs, including those focusing on geographic areas considered diversity hotspots, such as the Iberian Peninsula.

Table 1. Numbers of specimens examined for each rodent species, including number of tissue samples used for sequencing the long fragments of the cyt-*b* and the IRBP genes, and number of successful amplifications of shorter fragments of these genes from faecal samples and samples of bones from owl pellets. In addition to the long fragment IRBP sequences acquired, short fragment sequences were obtained from degraded tissue samples; these are shown in parentheses. For the non-invasive samples the total number of attempted extractions is given in square brackets. 'Unknown' samples are 'field' faecal samples (without species identification) for which amplification was attempted and failed.

Family	Species	Tissue samples		Non-invasive samples			
				Faeces		Bones	
		cyt-*b*	IRBP	cyt-*b*	IRBP	cyt-*b*	IRBP
Muridae	*Mus musculus*	11	9(1)	-	-	1[1]	0[1]
	Mus spretus	5	3(2)	7[8]	4[8]	2[2]	1[2]
	Apodemus flavicollis	11	10(1)	-	-	-	-
	Apodemus sylvaticus	7	27(3)	7[7]	1[1]	1[2]	2[2]
	Micromys minutus	4	3	-	-	-	-
	Rattus norvegicus	2	4	-	-	1[1]	1[1]
	Rattus rattus	2	4(1)	-	-	2[2]	1[2]
Cricetidae	*Arvicola sapidus*	7	10(2)	-	-	0[1]	1[1]
	Arvicola scherman	2	2(4)	-	-	-	-
	Chionomys nivalis	3	3(5)	-	-	-	-
	Microtus agrestis P†	-	11	-	-	-	-
	Microtus agrestis S†	-	3	-	-	1[1]	0[1]
	Microtus arvalis	2	5	-	-	3[3]	3[3]
	Microtus cabrerae	14	16(1)	26[26]	3[10]	-	-
	Microtus duodecimcostatus	12	6	-	-	-	-
	Microtus lusitanicus	36	6	11[11]	7[10]	3[3]	2[3]
	Microtus gerbei	1	1	-	-	4[5]	1[5]
	Myodes glareolus	4	5	-	-	4[4]	4[4]
Gliridae	*Eliomys quercinus*	6	5(1)	-	-	-	-
	Glis glis	5*	3(5)	-	-	-	-
Sciuridae	*Sciurus vulgaris*	21	4(2)	1[1]	1[1]	-	-
Unknown				[8]	[7]		
Total	**21**	**154**	**140(28)**	**52[61]**	**16[37]**	**22[25]**	**16[25]**

†*Microtus agrestis* is divided into three taxa by Paupério *et al.* (2012) and only samples from the two taxa present in Iberia (*M. agrestis* P and *M. agrestis* S) were considered in this work. *These samples consistently resulted in numt amplification for the cyt-*b* gene and so our analysis relied entirely on published data for this species.

Because rodents have high mitochondrial mutation rates (Nabholz *et al.* 2008), mitochondrial markers may be valuable for discriminating closely related taxa (Pfunder *et al.* 2004). However, due to the possible existence of mitochondrial DNA nuclear copies (pseudogenes or numts) or mitochondrial introgression, the use of nuclear genes is a valuable additional approach (Zhang & Hewitt 2003; Alves *et al.* 2008). Also, the inclusion of nuclear markers can complement mitochondrial data by allowing the detection of hybridisation, incomplete lineage sorting etc. (Alves *et al.* 2006; Heckman *et al.* 2007). As the interphotoreceptor retinoid-binding protein (IRBP) gene has already successfully been used for the identification of all carnivores in south-western Europe (Oliveira *et al.* 2009), and has recently proved its usefulness in rodent phylogenetics and species identification (Chaval *et al.* 2010; Pagès *et al.* 2010), it can be considered a good candidate for rodent molecular identification.

In developing molecular identification protocols in the present study we not only concern ourselves with the 21 long-recognised rodent species found in Iberia, but we also consider the partition of *Microtus agrestis* (*sensu lato*) into two taxa at the borderline of species separation: the 'Portuguese' lineage (*M. agrestis* P) found in western Iberia and the 'southern' lineage (*M. agrestis* S) found in north-eastern Iberia and elsewhere in southern Europe (Jaarola & Searle 2004; Paupério *et al.* accepted). Using both tissue and non-invasive samples, we utilise the mitochondrial cyt-*b* and nuclear IRBP genes for species identification, through the analysis of single nucleotide polymorphisms (SNPs). We describe our results for near-complete gene sequences of cyt-*b* and IRBP both by examining the phylogenetic relationships of the Iberian rodent taxa and also in the context of the DNA barcoding approach (Hebert *et al.* 2003a). Moreover, we develop a NIGS approach for faeces and bones from owl pellets, by designing shorter fragments of the selected markers. There are difficulties in obtaining amplifiable DNA by NIGS due to low DNA quantity and quality, including DNA degradation, presence of PCR inhibitors and DNA contamination (Rådström *et al.* 2004; Waits & Paetkau 2005). Mitochondrial markers are easier to use on degraded samples than nuclear markers because of the high number of mitochondrial genomes per cell. Nevertheless, as previously mentioned, the use of mtDNA alone can give misleading results, thus the inclusion of nuclear genes is valuable, even when using non-invasive samples (Oliveira *et al.* 2010). Following the validation of species identification methods for all sample types using the mitochondrial and nuclear markers, we consider their value in ecological studies and in the monitoring of Iberian rodents and consequently in biodiversity conservation and management of this important biogeographical region.

3. Material and Methods

Tissue sample collection and DNA extraction

We analysed a set of 261 tissue samples from CIBIO's small mammal repository, as well as from museums, comprising all Iberian rodent species (Table 1). These samples were selected mainly from Iberia, but also from other locations within the ranges of the species in Europe, to maximize the detection of intraspecific variation (Table S1, Supporting Information). Tissue samples included ear, foot, tail and diverse internal organs and DNA was extracted using the EasySpin® Genomic DNA Minipreps Tissue Kit (Citomed, Lisbon, Portugal) following the manufacturer's instructions.

Sample amplification, sequencing and analysis

The whole cyt-*b* gene was amplified in 154 samples using the primers L14727-SP and H15915-SP (Jaarola & Searle 2002). The polymerase chain reaction (PCR) thermal cycling profile consisted of a touch-down protocol ranging from 60 to 56ºC. The IRBP gene was amplified in 140 samples with new universal rodent primers: *IRBP.F2S*: 5'-GCAGGCTATGAAGAGTCRTG -3'; *IRBP.R2S*: 5'- AGCACGGAYACCTGAAACA -3'. The PCR thermal profile for most species consisted of a touch-down protocol from 60 to 54ºC. A different reverse primer was necessary to amplify samples from *Myodes glareolus* (*IRBP.R1S*: 5'-GCAGGTAGCCCACATTGC-3'), with an annealing temperature of 63ºC.

PCR reactions were performed with a total of 5 µl using 2 µl Qiagen© PCR Multiplex Kit Master Mix (Qiagen, Hilden, Germany), 0.2 µM of each primer and 10-20 ng of genomic DNA. The product obtained was purified using ExoSAP-IT® PCR clean-up Kit (GE Healthcare, Piscataway, NJ, USA) and sequences were generated with the amplification primers. Cycle sequencing reactions were carried out using the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Carlsbad, California, USA). Samples were subsequently sequenced for both strands on a 3130*xl* Genetic Analyser Sequencer (Applied Biosystems/HITASHI, Carlsbad, California, USA). Forward and reverse sequences were assembled and edited in Sequencher 4.7 (Gene Codes Corp., Ann Arbor, MI, USA), verified by eye and then aligned using ClustalW (Thompson *et al.* 1994) as implemented in BioEdit 7.0 (Hall 1999). Using the collector's morphological identification and GenBank sequences as reference, each new sequence was assigned to the best-matching species. The assignment was based on the existence of single point

mutations, or combinations of these, that were unique and consistent for each species (Belfiore *et al.* 2003), that we here term 'species discriminating sets of SNPs'.

Sequences obtained for the IRBP gene were phased using the software PHASE as implemented in DNAsp 5.10 (Librado & Rozas 2009). For each species, 1000 burn-in steps and 1000 iterations were performed, and the haplotypes selected had a minimum base call probability of 80%.

In addition to our sequences, we retrieved 85 cyt-*b* and 10 IRBP sequences from GenBank, which we incorporated into the analysis (Table S1, Supporting Information), giving a total of 239 sequences for cyt-*b* and 150 for the IRBP analysis. Since the use of highly divergent sequences of the same species from distant (or isolated) geographic areas can sometimes result in an increased intraspecific divergence due to unrecognized taxonomic differentiation (*e.g.* Paupério *et al.* accepted), we selected only sequences from European localities (up to ten sequences per species for cyt-*b* and all sequences available for the IRBP). This allowed us to minimise the detection of uninformative genetic diversity (from taxonomic subdivisions) while maximising the detection of relevant intraspecific variation.

Diversity indices, namely haplotype number (*H*) and diversity (*Hd*), nucleotide diversity (π), number of variable sites (*V*) and number of parsimony informative sites (*Pi*) were calculated for each species in both genes using DnaSP 5.10 (Librado & Rozas 2009). MEGA 5 (Tamura *et al.* 2011) was used to estimate nucleotide composition (%GC), to check for the presence of stop codons, as well as determine the number of non-synonymous substitutions for each gene and species.

Bayesian inference was used for both cyt-*b* and IRBP to build phylogenetic trees and determine species boundaries. The best-fit model of sequence evolution for each locus alignment was selected based on the Akaike information criterion and using the software jModelTest version 1.0 (Posada 2008). Trees were generated by MrBayes 3.1 (Huelsenbeck & Ronquist 2001) at the Bioportal server (www.bioportal.uio.no), using the rabbit (*Oryctolagus cuniculus*; Lagomorpha) as outgroup. Bayesian posterior probabilities were estimated from two runs with four chains of 10 million generations, with a sampling frequency that provided a total of 10 000 samples for each run, excluding 25% burnin. Tree visualisation was conducted using the software FigTree 1.3.1 (Rambaut 2009).

The species and other phylogenetic groupings were subject to further analysis. Intraspecific and interspecific mean pairwise genetic divergences (both within and among rodent families) were calculated for both loci employing Kimura two-parameter (K2P) distances in MEGA 5 (Tamura *et al.* 2011). Based on these divergence estimates,

histograms were built for each taxon independently to test for a gap between the intra- and interspecific values, called the 'barcoding gap' (Hebert *et al.* 2003b). The detection of very low inter- and very high intraspecific divergence values, often indicative of incomplete lineage sorting, admixture of previously isolated lineages or concealed divergence, can lead to an absence of the barcoding gap (Johnsen *et al.* 2010). Hence, to evaluate its real extent, a second graphical analysis of pairwise divergence was performed, comparing the mean and maximum intraspecific divergence of each taxon with the mean and minimum interspecific divergence of that taxon with the remaining species from the same family (Meier *et al.* 2008).

DNA extraction from non-invasive samples

DNA extractions were attempted on 61 faecal samples and 25 bone samples from owl pellets. Faecal samples were collected from live-traps, within less than 24 h of production, labelled as 'trap' (n=29), or opportunistically collected in the field, labelled as 'field' (n=32). The time of exposure for the 'field' samples could not be assessed accurately but was >24h. DNA was extracted using the QIAamp® DNA Micro Kit (Qiagen, Hilden, Germany), following the 'Isolation of genomic DNA from forensic case work samples' protocol, and half of the reagent volumes. DNA was eluted with 20, 30 or 40 µl AE buffer depending on whether it originated from 1, 2 or 3 faecal pellets respectively.

Bone samples were obtained from owl pellets collected in Iberia and identified to species using dichotomous keys (Blanco 1998; Gosàlbez & Noguera 1987). Mandibles and skulls were extracted using the DNeasy® Blood & Tissue Kit (Qiagen, Hilden, Germany) following the 'Purification of total DNA from compact animal bone' protocol with some adaptations, as follows: Briefly, bone fragments, usually consisting of a half mandible or a piece of the skull, were first cleaned externally with bleach (40%) and left under UV light for a short period of time (~15 min) in order to eliminate any contaminant on the exterior of the bone. The bones were then placed in a 2 ml tube with two 4 mm stainless steel balls and grinded for 3 min at 30 Hz in a Mixer Mill MM 400 (Retsch, Haan, Germany). The tubes were centrifuged in a microfuge for 1 min at 8000 rpm to remove loose bone powder from the lid of the tube and the balls were carefully removed after the addition of the lysis buffer. The remaining extraction procedures were performed as described in the protocol. DNA was eluted in 65 µl AE buffer.

All non-invasive extraction procedures were performed in a physically isolated room maintaining conditions to reduce risk of DNA contamination (Gilbert *et al.* 2005).

Amplification of non-invasive DNA samples

Shorter DNA fragments were selected from each gene to increase the amplification success in non-invasive samples (and also in degraded tissue samples), targeting regions rich in species discriminating sets of SNPs. These regions were identified from our analysis of the long gene fragments. New primers were designed for both cyt-*b* (*CYTB.F2S*: 5'-ATGAGGMCAAATATCATTCTGAGG -3'; *CYTB.R2S*: 5'-CAYGAAACAGGATCYAACAACC -3') and IRBP genes (*IRBP.F5S*: 5'-TAYATCCTSAAGCAGATGCG -3'; *IRBP.R5S*: 5'- CAGRGTGAGGATRGCCA -3'), resulting in two fragments of 220 bp and 238 bp, respectively. The primer design followed the recommendations of Waits & Paetkau (2005) for fragment size in highly degraded DNA samples, not exceeding 300 bp. The PCR thermal profiles consisted of a touch-down protocol from 56 to 51ºC for cyt-*b* and 58 to 52ºC for IRBP, with a final extension of 10 min and a total of 45 cycles. Amplification was performed in a final volume of 10 µl that consisted of: 4 µl of Qiagen© Multiplex PCR Kit Master Mix, 0.4 µM of each primer and 2 µl of genomic DNA. Samples were sequenced and analysed as already described for the tissues, allowing them to be typed for species. This molecular identification could be compared with the morphological species assignment, given that for the 'trap' faecal pellets and skull samples there was already information available on the species of origin (based on morphological identification of either the captured specimen or the skull). In the cases where incongruence between molecular and morphological species identification was observed, the species assignment of each sample was corrected according to the molecular identification. Using these results, misidentification percentages were calculated from the number of samples corrected for species identification out of the total number of samples analysed.

Haplotype maps were constructed for both the cyt-*b* and IRBP short fragments, using all available sequences (all relevant sequences in GenBank and our new sequences) by creating a consensus sequence for each species in BioEdit (Hall 1999), then importing into MEGA 5 (Tamura *et al.* 2011) and exporting as tables. The species-specific states were identified manually.

4. Results

Species identification using mitochondrial and nuclear markers

Cytochrome-b gene: From the 239 cyt-*b* sequences assembled (85 from GenBank) we analysed 1128 bp of the 1143 bp amplified, corresponding to positions

14145–15273 in Bibb *et al.* (1981). We obtained a total of 189 haplotypes. These sequences had a GC content of 41.5%, 46.1% parsimony informative sites, and 23 positions with nonsynonymous substitutions, none of which resulted in stop codons. Haplotype and nucleotide diversities for each species are shown in Table S2, Supporting Information. Some samples from specimens of the genera *Apodemus* and *Glis* generated numts (pseudogenes) which were excluded from the analysis (this particularly reduced the number of haplotypes for *G. glis*). For the *Apodemus* species, pseudogene sequences have already been described in the literature (see Dubey *et al.* 2009). *Glis glis* pseudogenes were identified by the presence of numerous positions with base changes, mostly transitions, together with a four base deletion, which is highly indicative of the presence of a numt (Gojobori *et al.* 1982; Li *et al.* 1984; Triant & DeWoody 2007).

The Bayesian tree inference for cyt-*b* was performed using the GTR+I+G model. The tree recovered the four rodent families present in Iberia with high posterior probability (≥0.95) (Figure 1). Within each family, all species formed monophyletic groups (posterior probability ≥0.85), with the exception of *Microtus lusitanicus*, which forms a paraphyletic group with *M. duodecimcostatus*. Samples of *M. lusitanicus* from the northwest quadrant of Iberia form an independent group with posterior probability of 1, however *M. lusitanicus* individuals from southern localities clustered closer to *M. duodecimcostatus*. Given this unresolved separation, these two taxa were combined and will be referred to as *Microtus* DL from now onwards. Other intraspecific lineages were recovered, such as those already reported in *Apodemus sylvaticus* (Michaux *et al.* 2003).

There is a small overlap between intraspecific (orange) and interspecific (within families, blue) K2P divergences for cyt-*b* (Figure 2A), which require a more detailed analysis (Figure 2B). Despite the high K2P intraspecific divergence observed in some species (*Chionomys nivalis*, 4.3%, *Apodemus sylvaticus*, 4.5%) or species complex (*Microtus* DL, 5.5%), and the low interspecific divergence obtained between *Microtus agrestis* P and *M. agrestis* S (3.3%), a barcoding gap is observed in all taxa.

IRBP gene: For the IRBP gene, 1040 base pairs were evaluated in 150 sequences (ten recovered from GenBank), corresponding to positions 286–1325 in Danciger *et al.* (1990). This fragment, with a GC content of 59.9%, displayed 38.8% parsimony informative sites and a total of 72 nonsynonymous substitution positions, none resulting in stop codons. These sequences resulted in 126 different haplotypes (Table S2, Supporting Information).

Bayesian tree inference was performed using the GTR+I+G model in MrBayes 3.1, the closest to the IRBP best fit model (TrNef+I+G) selected by jModelTest. The IRBP

phylogeny recovered the four rodent families present in Iberia with high posterior probability (Figure 3), however the phylogenetic relations among these families were different from those obtained with cyt-*b* (Figure 1). For the IRBP gene, all four families were recovered with posterior probability of 1 and most species formed monophyletic groups (posterior probabilities ≥71%). Once more, the pair *Microtus duodecimcostatus* and *M. lusitanicus* had shared haplotypes, representing a paraphyletic group.

Finally, possible hybridisation events were detected when comparing samples amplified with both the cyt-*b* and IRBP genes. Four *Mus* individuals from Cádiz (Spain) were identified as *M. spretus* by the cyt-*b* gene and as *M. musculus* by the IRBP gene.

Levels of K2P divergence for the IRBP gene were generally lower than those observed for cyt-*b* (Figure 4A). Despite that, a well-defined gap is observed between the interspecific within family (blue) and the interspecific between family (black) divergences and, in this last category, a gap can also be observed between the comparisons within the mouse- or squirrel-related clades (left peak) and comparisons between these clades (right peak) (mouse-related clade: murids and cricetids; squirrel-related clade: sciurids and glirids: Blanga-Kanfi *et al.* 2009; Churakov *et al.* 2010). However, the overall interspecific divergence within families overlaps with intraspecific divergence, which translates into the absence of a barcoding gap. In particular, *Mus spretus*, both *Apodemus* species and three *Microtus* taxa showed either high intra- or low interspecific divergence values, which resulted in the absence of a barcoding gap for those particular taxa (Figure 4B). The overall mean value of intraspecific divergence is 0.31%, where *Apodemus sylvaticus* shows the highest maximum intraspecific divergence value (1.4%). The mean value of interspecific divergences for species from the same family was 5.1%, though values ranged from 0.2% (*Microtus* DL/*M. gerbei*) to 10.1% (*Rattus norvegicus*/*Micromys minutus*).

Amplification success for non-invasive samples

Of the 61 faecal samples analysed, 85% (52/61) showed amplifiable DNA for cyt-*b*, whereas only 43% (16/37) provided positive results for IRBP. Faeces less than 24 h old (‘trap’) had greater amplification success (85%) than those that were older (‘field’, 44%). We obtained a positive relationship between amplification success and number of faecal pellets in ‘field’ samples (Figure 5). For the 25 samples of bones from owl pellets, 88% (22/25) provided positive amplifications for cyt-*b* and 64% (16/25) for IRBP. Species misidentification for faecal samples was estimated at 10% (6/61), and 12% (3/25) for bone samples. Amplification success did not vary detectably among species

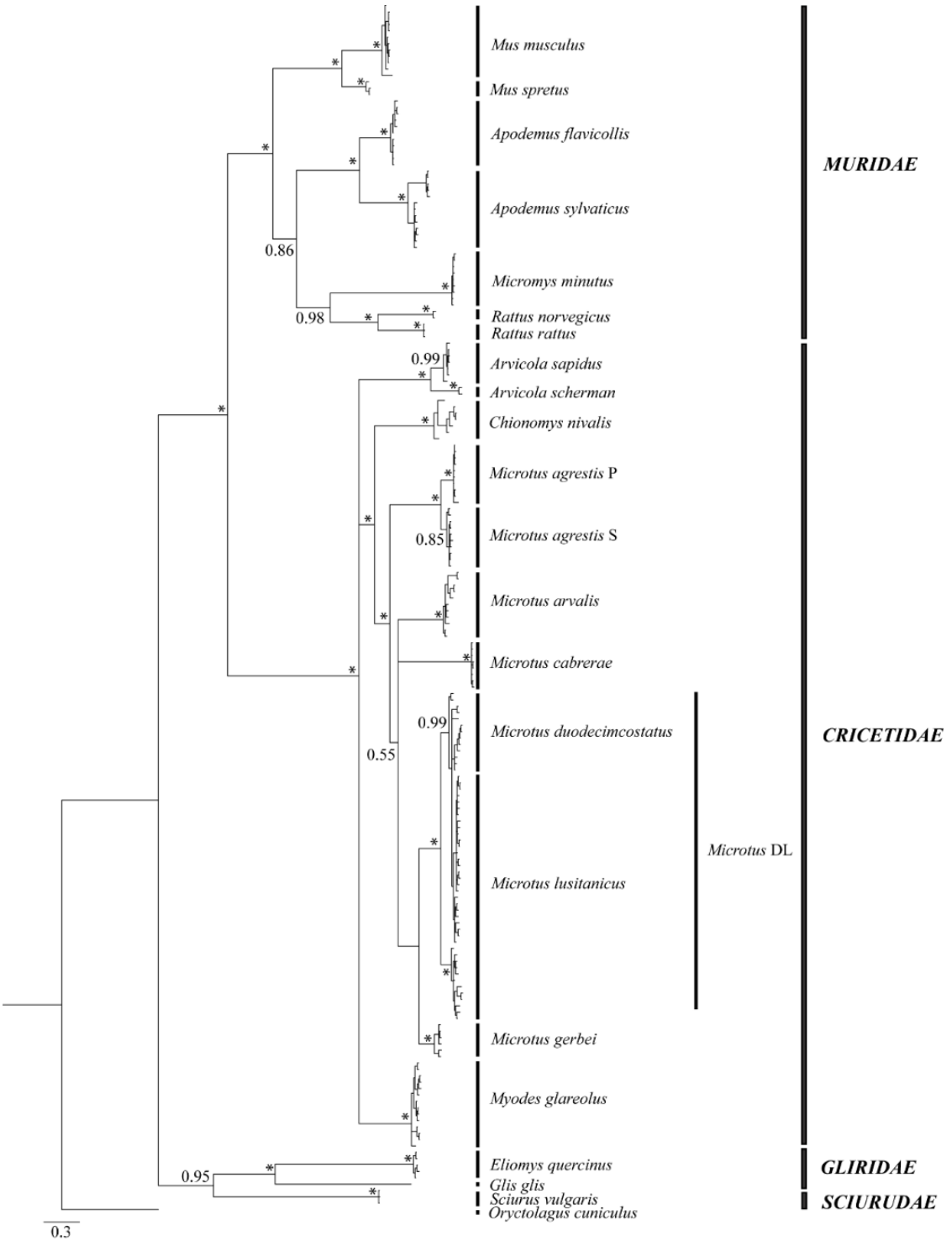


Figure 1. Bayesian inference tree for the cyt-*b* long fragment showing the phylogenetic relationship of all 21 Iberian rodent taxa. Posterior probabilities of major nodes are indicated. Asterisks represent posterior probabilities of 1. Outgroup: Oryctolagus cuniculus.

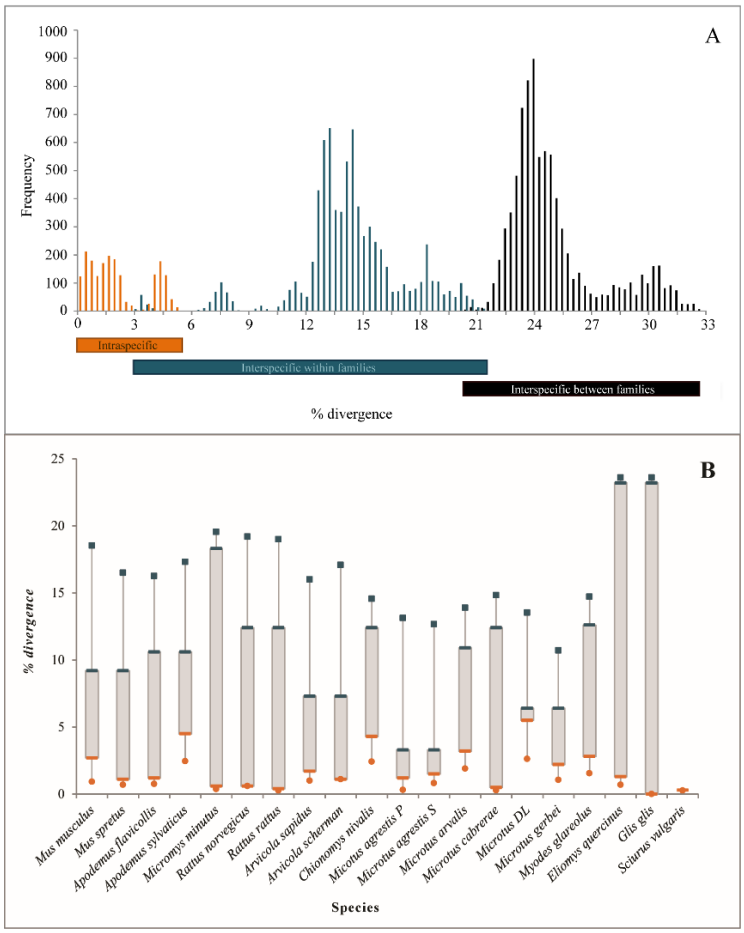


Figure 2. A) Histogram of *K2P* cyt-*b* divergence values of the 21 Iberian rodent taxa. (*Microtus duodecimcostatus* and *M. lusitanicus* were analysed as a single taxon, *Microtus* DL, due to lack of differentiation in the phylogenetic analysis.) Intraspecific divergence appears in orange, interspecific within families in blue and interspecific among families in black. B) Summary of pairwise divergences involving sequences of each species showing mean (orange dot) and maximum (orange dash) intraspecific divergences and mean (blue dot) and minimum (blue dash) interspecific divergences (comparing sequence from the named species with other species in the same family). Grey bars characterize the extent of the barcoding gap.

Variability of short fragments

For the cyt-*b* gene short fragment, 369 sequences were analysed, comprising 239 long fragment sequences and 74 non-invasive samples, along with 56 additional GenBank sequences from European individuals (Table 1; also see Table S1.A, Supporting Information). From these, 166 haplotypes were generated, with a total of 75 variable nucleotide positions along the entire fragment (Tables S2 and S3, Supporting Information). Despite the short size of the selected fragment, the observed high variability always allowed the identification of all Iberian rodent species, with the exception of the species complex *Microtus* DL.

For the IRBP gene, 210 sequences were analysed (150 long fragment sequences and 60 non-invasive and degraded tissue samples; Table 1; also see Table S1.B, Supporting Information), resulting in a total of 58 haplotypes with 66 variable positions (Tables S2 and S4, Supporting Information). This fragment was able to positively distinguish most rodent species with the exception of *Apodemus flavicollis*/*A. sylvaticus*, *Microtus agrestis* P/*M. agrestis* S and *Microtus* DL/*M. gerbei*. The haplotype diversity for each species, as well as inter- and intraspecific divergence, were generally lower than the values obtained for cyt-*b* (Table S2, Supporting Information).

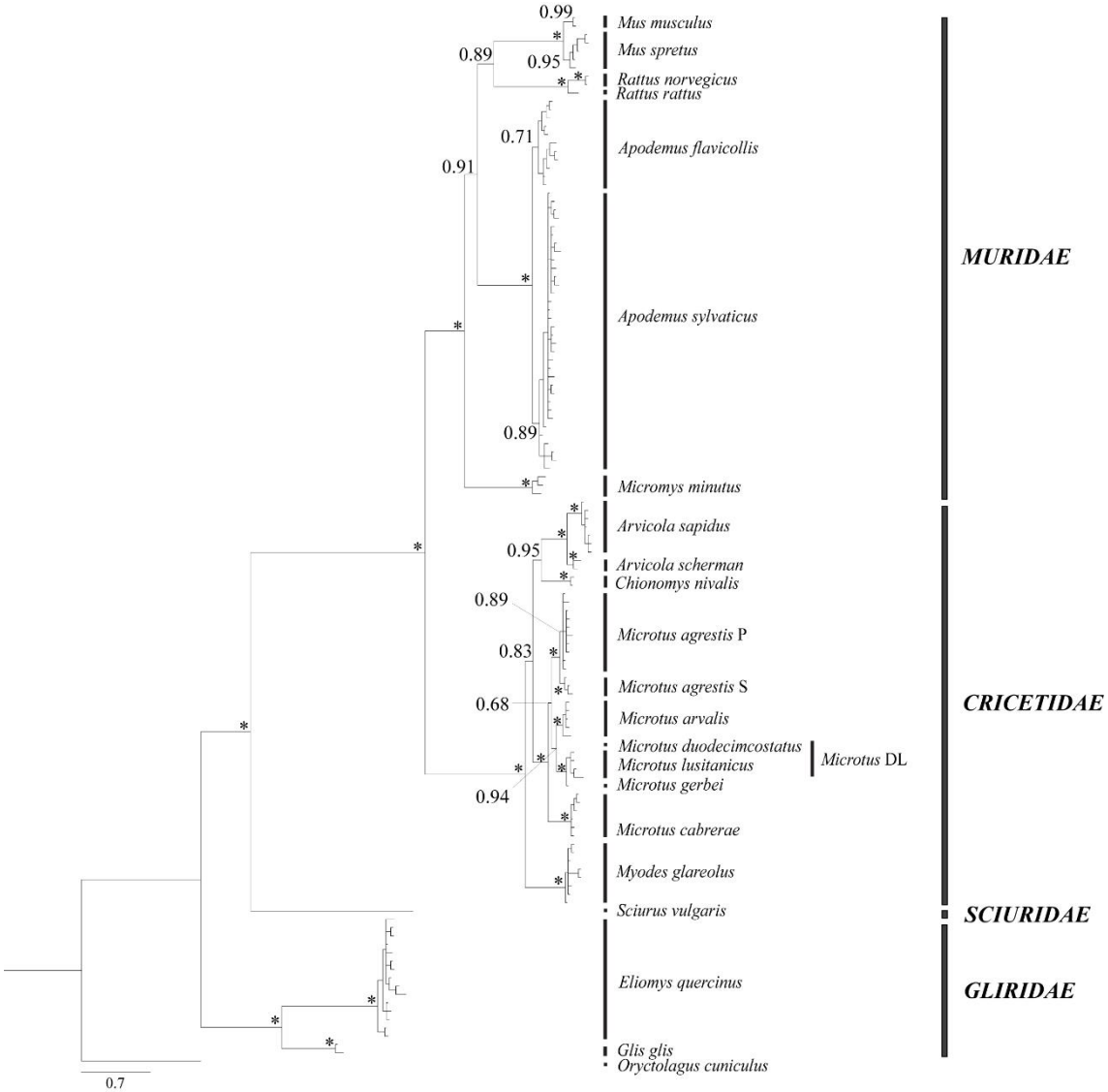


Figure 3. Bayesian inference tree for the IRBP long fragment illustrating the phylogenetic relationship of the 21 Iberian rodent taxa. Posterior probabilities of major nodes are indicated. Asterisks represent posterior probabilities of 1. Outgroup: *Oryctolagus cuniculus*.

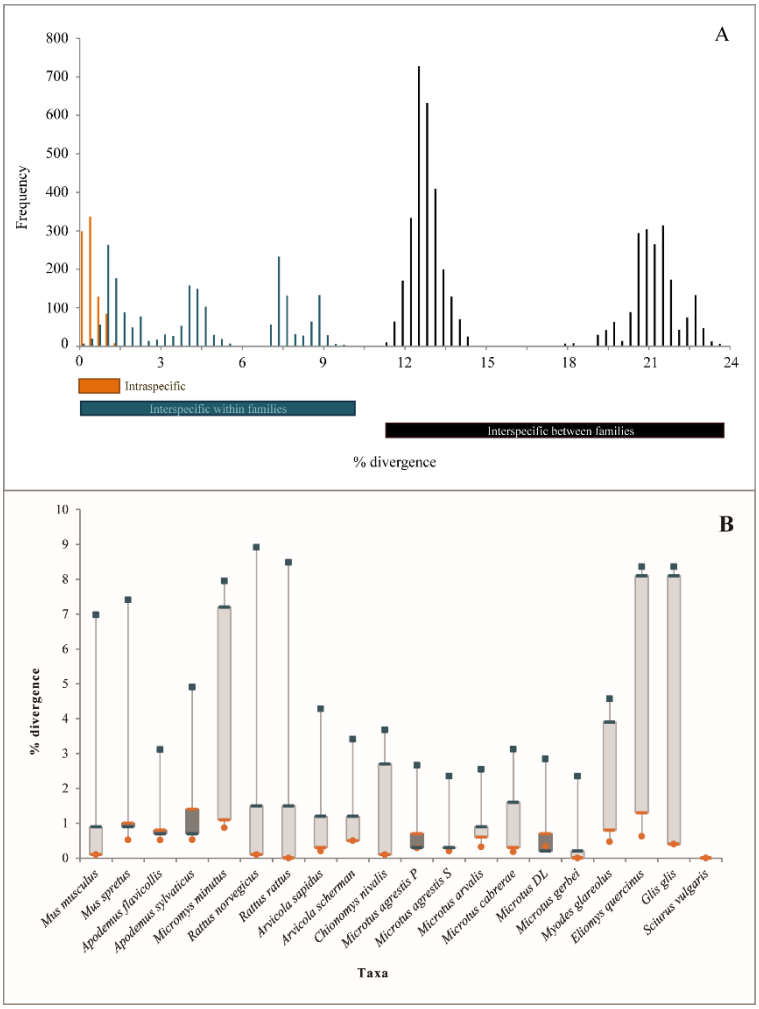
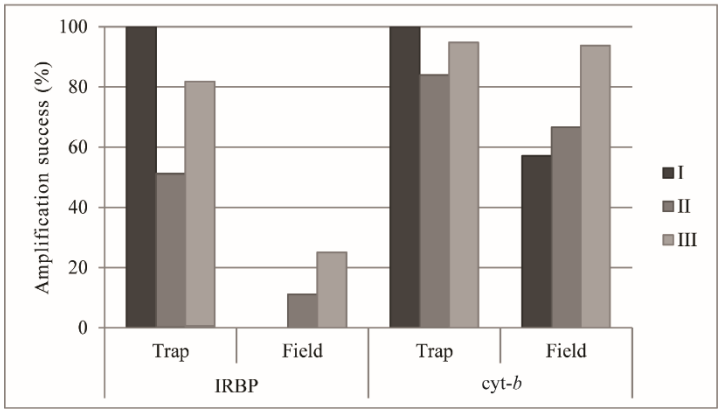


Figure 4. A) Histogram of *K2P* IRBP divergence values of the 21 Iberian rodent taxa. (*Microtus duodecimcostatus* and *M. lusitanicus* were analysed as a single taxon, *Microtus* DL, due to lack of differentiation in the phylogenetic analysis.) Intraspecific divergence appears in orange, interspecific within families in blue and interspecific among families in black. B) Summary of pairwise divergences involving sequences of each species showing mean (orange dot) and maximum (orange dash) intraspecific divergences and mean (blue dot) and minimum (blue dash) interspecific divergences (comparing sequence from the named species with other species in the same family). Grey bars characterize the extent of the barcoding gap; dark grey bars represent the cases where there was an overlap of intraspecific and interspecific divergences and show the extent of that overlap.

Figure 5. Amplification success for the IRBP and the cyt-*b* gene using DNA from faecal material. For each gene, the data is divided into 'Trap' and 'Field' depending on whether the faecal samples were collected from traps or opportunistically in the field. The data are further subdivided to show the amplification success for extractions performed with one (I), two (II) or three (III) faecal pellets



5. Discussion

Phylogenetic and barcoding gap approaches

The mitochondrial and nuclear genes that we analysed proved to be able to differentiate the large majority of the Iberian rodent species studied. As expected, the mitochondrial gene phylogeny showed higher discriminative power, especially for the more recent nodes, due to the higher mutation rate associated to mitochondrial DNA. Even though mutational saturation is expected for the deeper nodes (Hassanin *et al.* 1998), such events did not influence the cyt-*b* gene phylogeny unduly, which was supported with high posterior probability values. The average mammalian cyt-*b* neutral substitution rate is nearly 60 times higher than that of the IRBP gene (Nabholz *et al.* 2008) with particularly high values recorded in the family Arvicolidae (Galewski *et al.* 2006; Triant & DeWoody 2006). For this family we found examples of high intraspecific phylogenetic divergence between cyt-*b* lineages (as in *Chionomys nivalis*) but this tendency also extends to some non-arvicolid rodents, such as *Apodemus sylvaticus*, as already described by Michaux *et al.* (2003). Both *C. nivalis* and *A. sylvaticus* have large geographical ranges and the high intraspecific divergence levels mirror those found in other widely distributed rodent species (Michaux *et al.* 2003; Grill *et al.* 2009; Wójcik *et al.* 2010). *Eliomys quercinus* is also a widely distributed species which, according to Nombela *et al.* (1982), shows karyotypic differentiation. This might explain the occurrence of highly distinct lineages for the IRBP gene, however such differentiation was not recorded for cyt-*b*. Discordance between mitochondrial and nuclear phylogenies is quite common among mammals, often attributed to mitochondrial introgression, e.g, in *Myodes* voles (Boratyński *et al.* 2011) and in hares (Alves *et al.* 2008). These results emphasize the importance in using nuclear genes in addition to mtDNA to reveal major genetic differentiation.

All species formed monophyletic groups for both genes, with the exception of the pair *Microtus duodecimcostatus* and *M. lusitanicus* (Figures. 1 and 3), and in the IRBP gene the species *M. gerbei* could only be distinguished from this species complex on the basis of a single SNP. These three vole species form a monophyletic group in mtDNA with higher support for the closer relatedness of *M. lusitanicus* and *M. duodecimcostatus* (Jaarola *et al.* 2004; Tougard *et al.* 2008a), further supported by comparative dental morphometry and cytogenetics (Chaline *et al.* 1999). The lack of differentiation between *M. lusitanicus* and *M. duodecimcostatus* presumably reflects incomplete lineage sorting or hybridisation (Bastos-Silveira *et al.* accepted, and our study). In addition, due to the low sample size of *M. gerbei*, we cannot be sure that this species can reliably be

distinguished from the species complex *M. duodecimcostatus/M. lusitanicus* with the IRBP gene.

Three levels of comparison were considered in the barcoding analysis: intraspecific, interspecific within families and interspecific between families. At the interspecific divergence between families level, both genes showed a bimodal distribution of values (Figures 2A and 4A) which represent the comparisons within and between the mouse-related (murid and cricetid) and the squirrel-related (glirid and sciurid) rodent clades identified in several previous phylogenetic studies (Blanga-Kanfi *et al.* 2009; Churakov *et al.* 2010).

Considering the intra- and interspecific (within family) variation a barcoding gap was detected for most species in both genes, which was more substantial for cyt-*b* (Figures 2B and 4B). The threshold for species delimitation using the cyt-*b* was around 3%, which is consistent with the maximum limit of intraspecific variation detected for the COI gene in mammals, the common mitochondrial gene used for barcoding (Luo *et al.* 2011). However, for the IRBP gene it is not possible to define a threshold for species delimitation because, when all species are analysed together, an overlap of intra- and interspecific variation is observed due to the low divergences between closely related taxa (Figure 4A), a common finding for nuclear exons (Zhang & Hewitt 2003). In the specific analysis of the IRBP gene, we observe that the two *Apodemus* species, *M. agrestis* P, *M. agrestis* S, *Microtus* DL and *Mus spretus* showed overlap of intra- and interspecific divergences (dark grey in Figure 4B). These represent cases where there are closely related species, which reduces the probability of a barcoding gap for a sequence type that is not exceptionally variable.

Even though nuclear genes like IRBP may show limitations for discrimination between closely related species, they have particular value in the detection of hybridisation events, as revealed between *Mus musculus* and *Mus spretus*. We found that some specimens identified as *M. musculus* morphologically and from IRBP sequences were classified as *M. spretus* from the mitochondrial D-loop (S. Gabriel, personal communication) and cyt-*b* (this study). Hybridisation of these two species has long been recognised (Orth 2002) and recent studies suggest adaptive introgression of rodenticide resistance from *M. spretus* into *M. musculus* (Song *et al.* 2011).

Using nuclear genes also avoids the problem of confusion between mitochondrial genes and numts. In this study we detected numts of the cyt-*b* gene in *Apodemus* and *Glis*, which showed co- or preferential amplification (Dubey *et al.* 2009). A recent study from den Tex *et al.* (2010) confirms that the use of universal primers tends to increase the likelihood of numt amplification. Numt co-amplification has been described for many

rodent species (Groenenberg & Dekker 2011; Mirol *et al.* 2000; Phillips & Dudı 2008; Triant & DeWoody 2008) and their occurrence could severely compromise efforts to make species identifications based purely on mitochondrial markers (Dubey *et al.* 2009).

Thus, we believe the simultaneous use of nuclear and mitochondrial genes increases substantially the likelihood of accurate species identification and provides the opportunity for further perspectives (such as subdivision within species, hybridisation, etc.).

Applicability to non-invasive sampling

We obtained successful cyt-*b* amplifications for the majority of faecal samples (85%, 52/61). The results when using fresh samples ('trap': 93%, 27/29) were similar to those reported previously for arvicolids (95%, Alasaad *et al.* 2011). Regarding the use of exposed rodent faecal samples from the field, our amplification success (78%, 25/32) was also comparable to that obtained in other studies with small mammals (80%, Moran *et al.* 2008).

As expected, the amplification success for the IRBP gene was lower than for cyt-*b*, with an average of 44% (16/37) positive amplifications for faeces. For samples labelled as 'trap' only, there were 78% (14/18) positive amplifications, and using 'field' faecal pellets, there was 11% (2/19) amplification success. The DNA amplification success that we obtained for nuclear genes from fresh rodent faeces was comparable to that obtained for previous studies of faeces, specifically from *Apodemus* (83.3%, Moran *et al.* 2008).

There have been particular efforts to obtain DNA from carnivore scats. Considering nuclear gene amplification, generally there has been similar DNA amplification success for fresh carnivore scats (71.7%, Murphy *et al.* 2007; 79.2%, Oliveira *et al.* 2010) and fresh rodent faeces (78%, our study; 90%, Alasaad *et al.* 2011; 83.3%, Moran *et al.* 2008). However, it is difficult to confirm such a pattern for exposed faecal samples, as, to our knowledge, there is no previous literature making use of rodent faeces collected in the field to amplify nuclear loci. For exposed carnivore scats, nuclear DNA amplification success varies over a wide range of values (20-49%, Murphy *et al.* 2007; 26%, Michalski *et al.* 2011; 48%, Kohn *et al.* 1999; 54%, Mukherjee *et al.* 2010; 84.1%, Oliveira *et al.* 2010). Our amplification success for nuclear genes regarding 'field' samples (11%) is lower than any value recorded for carnivore scats. Nevertheless, the size of rodent faecal pellets must be considered, as these are smaller and presumably contain less amplifiable DNA than larger carnivore faecal samples.

From our results, it is desirable to collect rodent faeces for DNA extraction as soon as possible after they have been produced. Rodents have a tendency to leave faecal pellets on clean surfaces such as cardboard or wooden boards (Emlen *et al.* 1957). Therefore, dropping stations may provide a good solution for collecting fresh faeces for genetic analysis.

Regarding the extractions from bones collected from owl pellets, we obtained positive amplifications for 88% (22/25) of samples using the cyt-*b* gene, which is marginally lower than some equivalent studies (97%, Taberlet & Fumagalli 1996; 92%, Poulakakis *et al.* 2005), but higher than the 62% obtained by Centeno-Cuadros *et al.* (2009). We obtained 64% (16/25) positive amplifications for IRBP, which can only be compared to the 38% amplification success in Taberlet & Fumagalli (1996), although these authors present some possible explanations for their low amplification success. Skulls of small mammals from museum collections can provide greater amplification success for nuclear genes, judging by the 100% obtained by Asher *et al.* (2010) and Asher & Hofreiter (2006).

Overall, we found 10% (6/61) and 12% (3/25) morphological misidentifications of faecal and bone samples, respectively. In 'trap' faecal samples, most identification errors were between *Apodemus sylvaticus* and *Mus spretus*. This might indicate that the trapped species were also misidentified, but cross contamination due to multiple captures cannot be ruled out. For bone samples, misidentification occurred between similar species of the same genus (*Microtus gerbei* / *M. lusitanicus*; *Microtus arvalis* / *M. agrestis*; *Rattus rattus* / *R. norvegicus*), which reinforces the value of molecular techniques for accurate identification of morphologically similar species.

As expected, nucleotide and haplotype variation generally decreased when comparing the long to the short fragments (Table S2, Supporting Information), but successful species identification was nevertheless possible with the short fragments, with the exception of *Apodemus* sp., *Microtus gerbei*, *M. agrestis* P and *M. agrestis* S for the IRBP gene and *M. lusitanicus* and *M. duodecimcostatus* for both genes. SNP haplotype maps of both markers (Tables S3 and S4, Supporting Information) can be used as a reference, either for allocating samples collected in Iberia to the correct species or for providing the basis for developing a species-specific diagnostic test.

6. Conclusions and conservation implications

In this study we were able to distinguish all Iberian rodent species using a mitochondrial and a nuclear gene, with the exception of the species complex *Microtus duodecimcostatus* and *M. lusitanicus*. Examining both the cyt-*b* and IRBP genes provided an effective methodology for species identification, whether applied to living or well preserved specimens (tissue samples), museum samples, or to non-invasive samples (faeces and/or bones from owl pellets). In general, we had a high DNA extraction success for the non-invasive samples and the short amplified fragments allowed identification of all the taxa studied with a few exceptions for the IRBP gene. Basing the analysis on the cyt-*b* and IRBP genes showed the value of using highly variable genetic markers to distinguish very similar taxa.

The methodology we have developed provides a major advance for the conservation genetics of Iberian rodents, since NIGS avoids specimen handling and disturbance. There are opportunities for fine-scale population studies of rodents by faecal analysis. Considerable increases in population sample sizes and geographic coverage may also be achieved by using owl pellets. Additionally, owl pellets of course provide NIGS data on owl diet, which in itself may provide insights relevant to the conservation of both owls and rodents. The sophistication that can be achieved by these various approaches can be substantial given the success that we have demonstrated in amplifying short fragments of nuclear DNA. This allows the use of microsatellite and other markers that can be applied to a wide range of population genetic analyses.

Thus, through our studies, there is the opportunity to carry out detailed non-invasive studies of those rodents that are viewed as under conservation threat in Iberia, either locally or throughout the region.

Acknowledgements

We thank A. Arrizabalaga and the Museum of Natural Science of Granollers, A. Mira and the PERSIST project team, the Portuguese National Tissue Bank/National Institute for Nature and Biodiversity Conservation (BTVS/ICNB), University of Oslo Bioportal for providing a platform for running the statistical software MrBayes v3.1, S. Afonso, C. Bastos-Silveira, Z. Boratynski, E. Castiéns, A. de Castro, A. Centeno-Cuadros, G. Dias (Fran), S. Ferreira, S. Gabriel, J.A. Godoy, V. Gomes, H. Gonçalves, J.M. Grosso-Silva, J. Herman, M. Jaarola, V. Mata, P. Monterroso, R. Pita, R. Ribeiro, R. Rosa-García, P. Sarmento, B. Silva, P. Sousa and R. Vega for providing some of the samples used in this study. This work was partially supported by FEDER through

COMPETE program and Portuguese national funds through the FCT – Fundação para a Ciência e a Tecnologia (PTDC/BIA-BEC/105110/2008 research project) and by Programa Operacional Potencial Humano – Quadro de Referência Estratégico Nacional funds from the European Social funds and Portuguese Ministério da Educac¸a˜o e Cieˆncia (FCT, SFRD/BD/27369/2006 PhD grant to J. Paupério). We thank the reviewers for valuable comments.

7. References

Alasaad S, Soriguer RC, Jowers MJ *et al.* (2011) Applicability of mitochondrial DNA for the identification of arvicolid species from faecal samples: a case study from the threatened Cabrera's vole. *Molecular Ecology Resources*, **11**, 409-414.

Álvarez-Castañeda ST, Ortega-Rubio A (2003) Current status of rodents on islands in the Gulf of California. *Biological Conservation*, **109**, 157-163.

Alves PC, Harris DJ, Melo-Ferreira J *et al.* (2006) Hares on thin ice: introgression of mitochondrial DNA in hares and its implications for recent phylogenetic analyses. *Molecular Phylogenetics and Evolution*, **40,** 640-641.

Alves PC, Melo-Ferreira J, Freitas H, Boursot P (2008) The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **363**, 2831-2839.

Asher RJ, Hofreiter M (2006) Tenrec phylogeny and the non-invasive extraction of nuclear DNA. *Systematic Biology*, **55**, 181–194.

Asher RJ, Maree S, Bronner G, *et al.* (2010) A phylogenetic estimate for golden moles (Mammalia, Afrotheria, Chrysochloridae). *BMC Evolutionary Biology*, **10**, 69–82.

Avenant NL (2005) Barn owl pellets: a useful tool for monitoring small mammal communities? *Belgian Journal of Zoology*, **135**, 39-43.

Bastos-Silveira C, Santos S, Monarca R, *et al.* (accepted) Deep mitochondrial introgression and hybridization among ecologically divergent vole species. *Molecular Ecology*.

Beja-Pereira A, Oliveira R, Alves PC *et al.* (2009) Advancing ecological understandings through technological transformations in non-invasive genetics. *Molecular Ecology Resources*, **9**, 1279-1301.

Belfiore NM, Hoffman FG, Baker RJ, DeWoody JA (2003) The use of nuclear and mitochondrial single nucleotide polymorphisms to identify cryptic species. *Molecular Ecology*, **12**, 2011–2017.

Bertolino S, Amori G, Henttonen H *et al.* (2008) *Eliomys quercinus*. In: IUCN 2011. IUCN Red List of Threatened Species, version 2011.2. www.iucnredlist.org.

Bertolino S, Wauters L, Pizzul A *et al.* (2009) A general approach of using hair-tubes to monitor the European red squirrel: A method applicable at regional and national scales. *Mammalian Biology*, **74**, 210-219.

Bibb MJ, Van Etten RA, Wright CT *et al.* (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell*, **26**, 167-180.

Blanco JC (1998) Mamíferos de España, Volumen I . Editorial Planeta, S.A., Barcelona.

Blanga-Kanfi S, Miranda H, Penn O *et al.* (2009) Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evolutionary Biology*, **9**, 71.

Bonhomme F, Orth A, Cucchi T *et al.* (2011) Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization. *Proceedings of the Royal Society B*, **278**, 1034-1043.

Boratyński Z, Alves PC, Berto S *et al.* (2011) Introgression of mitochondrial DNA among *Myodes* voles: consequences for energetics? *BMC Evolutionary Biology*, **11**, 355-367.

Borkowska A, Ratkiewicz M (2008) Sex-related spatial structure and effective population size in the common vole, *Microtus arvalis*, as revealed by mtDNA analysis. *Annales Zoologici Fennici*, **45**, 255-262.

Castiglia R, Annesi F, Kryštufek B *et al.* (2009) The evolutionary history of a mammal species with a highly fragmented range: the phylogeography of the European snow vole. *Journal of Zoology*, **279**, 243-250.

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.

Centeno-Cuadros A, Godoy JA (2010) Structure, organization and nucleotide diversity of the mitochondrial control region and cytochrome *b* of southern water vole (*Arvicola sapidus*). *Mitochondrial DNA*, **21**, 48-53.

Chaline J, Brunet-Lecomte P, Montuire S, Viriot L (1999) Anatomy of the arvicoline radiation (Rodentia): palaeogeographical, palaeoecological history and evolutionary data. *Annales Zoologici Fennici*, **36**, 239-267.

Chaval Y, Dobigny G, Michaux JR *et al.* (2010) A multi-approach survey as the most reliable tool to accurately assess biodiversity: an example of Thai murine rodents. *Kasetsart Journal (Natural Sciences)*, **44**, 590-603.

Chevret P, Veyrunes F, Britton-Davidian J (2005) Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. *Biological Journal of the Linnean Society*, **84**, 417-427.

Churakov G, Sadasivuni MK, Rosenbloom KR *et al.* (2010) Rodent evolution: back to the root. *Molecular Biology and Evolution*, **27**, 1315-1326.

Colangelo P, Aloise G, Franchini P *et al.* (2012) Mitochondrial DNA reveals hidden diversity and an ancestral lineage of the bank vole in the Italian peninsula. *Journal of Zoology*, **287**, 41-52.

Conroy CJ & Cook JA (1999) MtDNA evidence for repeated pulses of speciation within arvicoline and murid rodents. *Journal of Mammalian Evolution*, **6**, 221-245.

Cook JA, Runck AM, Conroy CJ (2004) Historical biogeography at the crossroads of the northern continents: molecular phylogenetics of red-backed voles (Rodentia: Arvicolinae). *Molecular Phylogenetics and Evolution*, **30**, 767-777.

Danciger F, Kozak C, Nickerson J *et al.* (1990) Localization of the gene for interphotoreceptor retinoid-binding protein to mouse chromosome 14 near Np-1. *Genomics*, **8**, 727-731.

Dubey S, Michaux JR, Brünner H *et al.* (2009) False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene. *Molecular Phylogenetics and Evolution*, **50**, 633-641.

Emlen JT, Hine RL, Fuller WA, Alfonso P (1957) Dropping boards for population studies of small mammals. *Journal of Wildlife Management*, **21**, 300-314.

Fernandes M, Pita R, Mira A (2008) *Microtus cabrerae*. In: IUCN 2011. IUCN Red List of Threatened Species, version 2011.2. www.iucnredlist.org.

Fink S, Excoffier L, Heckel G (2004) Mitochondrial gene diversity in the common vole *Microtus arvalis* shaped by historical divergence and local adaptations. *Molecular Ecology*, **13**, 3501-3514.

Fink S, Fischer MC, Excoffier L, Heckel G (2010) Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia: *Microtus*): the utility of AFLPs versus mitochondrial and nuclear sequence markers. *Systematic Biology*, **59**, 548-572.

Galewski T, Tilak M, Sanchez S *et al.* (2006) The evolutionary radiation of Arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. *BMC Evolutionary Biology*, **6**, 80-97.

Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. *Trends in Ecology & Evolution*, **20**, 541-544.

Gojobori T, Li W-H, Graur D (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. *Journal of Molecular Evolution*, **18**, 360-369.

Gosàlbez I, Noguera J (1987) *Insectívors i rosegadors de Catalunya: metodologia d'estudi i catàleg faunístic*. Editorial Ketres S.A., Barcelona.

Grill A, Amori G, Aloise G *et al.* (2009) Molecular phylogeography of European *Sciurus vulgaris*: refuge within refugia? *Molecular Ecology*, **18**, 2687-2699.

Groenenberg DSJ, Dekker RWRJ (2011) A mouse's tail: how to settle an insurance dispute. *Forensic Science International*, **207**, 24-27.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **42**, 95-98.

Harris RL, Nicol SC (2010) The effectiveness of hair traps for surveying mammals: results of a study in sandstone caves in the Tasmanian Southern Midlands. *Australian Mammalogy*, **32**, 62-66.

Hassanin A, Lecointre G, Tillier S (1998) The "evolutionary signal" of homoplasy in protein-coding gene sequences and its consequences for a priori weighting in phylogeny. *Comptes Rendus de l'Académie des Sciences*, **321**, 611-620.

Haynes S, Jaarola M, Searle JB (2003) Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. *Molecular Ecology*, **12**, 951-956.

Hebert PDN, Ratnasingham S, DeWaard JR (2003a) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B*, **270**, 96-99.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003b) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B*, **270**, 313-321.

Heckman KL, Mariani CL, Rasoloarison R, Yoder AD (2007) Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (*Microcebus*). *Molecular Phylogenetics and Evolution*, **43**, 353-367.

Henry P, Henry A, Russello MA (2011) A non-invasive hair sampling technique to obtain high quality DNA from elusive small mammals. *Journal of Visualized Experiments*, **49**, 1-5.

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

Hurner H, Kryštufek B, Sara M *et al.* (2010) Mitochondrial phylogeography of the edible dormouse (*Glis glis*) in the western Palearctic region. *Journal of Mammalogy*, **91**, 233-242.

IUCN (2011) IUCN Red List of Threatened Species, Version 2011.1

Jaarola M, Martínková N, Gündüz I *et al.* (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **33**, 647-663.

Jaarola M, Searle JB (2002) Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology*, **11**, 2613-2621.

Jaarola M, Searle JB (2004) A highly divergent mitochondrial DNA lineage of *Microtus agrestis* in southern Europe. *Heredity*, **92**, 228-234.

Johnsen A, Rindal E, Ericson PGP *et al.* (2010) DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. *Journal of Ornithology*, **151**, 565-578.

Kohn MH, York EC, Kamradt DA, *et al.* (1999) Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London B*, **266**, 657-663.

Kryštufek B, Lužnik M, Buzan EV (2011) Mitochondrial cytochrome *b* sequences resolve the taxonomy of field mice (*Apodemus*) in the western Balkan refugium. *Acta Theriologica*, **57**, 1-7.

Li W-H, Wu C-I, Luo C-C (1984) Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and evolutionary implications. *Journal of Molecular Evolution*, **25**, 58-71.

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

Lundrigan BL, Jansa SA, Tucker PK (2002) Phylogenetic relationships in the genus *Mus*, based on paternally, maternally, and biparentally inherited characters. *Systematic Biology*, **51**, 410-431.

Luo A, Zhang A, Ho SY *et al.* (2011) Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. *BMC Genomics*, **12**, 84-97.

Malé P-JG, Martin J-F, Galan M *et al.* (2012) Discongruence of Mhc and cytochrome *b* phylogeographical patterns in *Myodes glareolus* (Rodentia: Cricetidae). *Biological Journal of the Linnean Society*, **105**, 881-899.

Martin Y, Gerlach G, Schlötterer C, Meyer A (2000) Molecular phylogeny of European muroid rodents based on complete cytochrome *b* sequences. *Molecular Phylogenetics and Evolution*, **16**, 37-47.

Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the “barcoding gap” and leads to misidentification. *Systematic Biology*, **57**, 809–813.

Mercer JM, Roth VL (2003) The effects of Cenozoic global change on squirrel phylogeny. *Science*, **299**, 1568-1572.

Michalski F, Valdez FP, Norris D, *et al.* (2011) Successful carnivore identification with faecal DNA across a fragmented Amazonian landscape. *Molecular Ecology Resources*, **11**, 862-871.

Michaux JR, Chevret P, Filippucci M-G, Macholan M (2002) Phylogeny of the genus *Apodemus* with a special emphasis on the subgenus *Sylvaemus* using the nuclear IRBP gene and two mitochondrial markers: cytochrome *b* and 12S rRNA. *Molecular Phylogenetics and Evolution*, **23**, 123-136.

Michaux JR, Libois R, Paradis E, Filippucci M-G (2004) Phylogeographic history of the yellow-necked fieldmouse (*Apodemus flavicollis*) in Europe and in the Near and Middle East. *Molecular Phylogenetics and Evolution*, **32**, 788-798.

Michaux JR, Magnanou E, Paradis E *et al.* (2003) Mitochondrial phylogeography of the woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Molecular Ecology*, **12**, 685-697.

Mirol PM, Mascheretti S, Searle JB (2000) Multiple nuclear pseudogenes of mitochondrial cytochrome *b* in *Ctenomys* (Caviomorpha, Rodentia) with either great similarity to or high divergence from the true mitochondrial sequence. *Heredity*, **84**, 538-547.

Moncrief ND, Va RA, Bussche D *et al.* (2008) Diagnostic genetic marker that differentiates eastern fox squirrels from eastern gray squirrels. *Journal of Wildlife Management*, **72**, 320-323.

Moran S, Turner PD, O'Reilly C (2008) Non-invasive genetic identification of small mammal species using real-time polymerase chain reaction. *Molecular Ecology Resources*, **8**, 1267-1269.

Mukherjee S, Cn A, Home C, Ramakrishnan U (2010) An evaluation of the PCR-RFLP technique to aid molecular-based monitoring of felids and canids in India. *BMC Research Notes*, **3**, 159.

Murphy MA, Kendall KC, Robinson A, Waits LP (2007) The impact of time and field conditions on brown bear (*Ursus arctos*) faecal DNA amplification. *Conservation Genetics,* **8**, 1219-1224.

Nabholz B, Glémin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals - the longevity hypothesis. *Molecular Biology and Evolution*, **25**, 120-130

Nilsson MA, Gullberg A, Spotorno AE *et al.* (2003) Radiation of extant marsupials after the K/T boundary: evidence from complete mitochondrial genomes. *Journal of Molecular Evolution*, **57**, 3-12.

Nilsson MA, Härlid A, Kullberg M, Janke A (2010) The impact of fossil calibrations, codon positions and relaxed clocks on the divergence time estimates of the native Australian rodents (*Conilurini*). *Gene*, **455**, 22-31.

Nombela JJA, Murcia CR, Delibes M, Hiraldo F (1982) Comparative karyotype studies between Spanish and French populations of *Eliomys quercinus* L. *Genetica*, **59**, 161-166.

Nunome M, Yasuda SP, Sato JJ *et al.* (2007) Phylogenetic relationships and divergence times among dormice (Rodentia, Gliridae) based on three nuclear genes. *Zoologica Scripta*, **36**, 537-546.

Oliveira R, Castro D, Godinho R *et al.* (2010) Species identification using a small nuclear gene fragment: application to sympatric wild carnivores from south-western Europe. *Conservation Genetics*, **11**, 1023-1032.

Orth A (2002) Natural hybridisation between two sympatric species of mice *Mus musculus domesticus* L. and *Mus spretus* Lataste. *Comptes Rendus Biologies*, **325**, 89-97.

Pagès M, Chaval Y, Herbreteau V *et al.* (2010) Revisiting the taxonomy of the Rattini tribe: a phylogeny-based delimitation of species boundaries. *BMC Evolutionary Biology*, **10**, 184-211.

Paupério J, Herman JS, Melo-Ferreira J *et al.* (2012) Cryptic speciation in the field vole: a multilocus approach confirms highly divergent lineages in Europe. *Molecular Ecology*, **21**, 6015-6032.

Pfunder M, Holzgang O, Frey JE (2004) Development of microarray-based diagnostics of voles and shrews for use in biodiversity monitoring studies, and evaluation of mitochondrial cytochrome oxidase I vs. cytochrome *b* as genetic markers. *Molecular Ecology*, **13**, 1277-1286.

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161-175.

Piggott MP (2004) Effect of sample age and season of collection on the reliability of microsatellite genotyping of faecal DNA. *Wildlife Research*, **31**, 485-493.

Pocock MJO, Jennings N (2006) Use of hair tubes to survey for shrews: new methods for identification and quantification of abundance. *Mammal Review*, **36**, 299-308.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253-1256.

Poulakakis N, Lymberakis P, Paragamian K, Mylonas M (2005) Isolation and amplification of shrew DNA from barn owl pellets. *Biological Journal of the Linnean Society*, **85**, 331-340.

Rambaut A (2009) FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh.

Reyes A, Gissi C, Pesole G *et al.* (2000) Where do rodents fit? Evidence from the complete mitochondrial genome of *Sciurus vulgaris*. *Molecular Biology and Evolution*, **17**, 979-983.

Reyes A, Pesole G, Saccone C (1996) Complete mitochondrial DNA sequence of the fat dormouse, *Glis glis*: Further evidence of rodent paraphyly. *Molecular Biology*, **15**, 499-505.

Rigaux P, Vaslin M, Noblet JF *et al.* (2008) *Arvicola sapidus*. In IUCN 2011. IUCN Red List of Threatened Species, version 2011.2. www.iucnredlist.org.

Rådström P, Knutsson R, Wolffs P *et al.* (2004) Pre-PCR processing: strategies to generate PCR-compatible samples. *Molecular Biotechnology*, **26**, 133-146.

Serizawa K, Suzuki H, Tsuchiya K (2000) A phylogenetic view on species radiation in *Apodemus* inferred from variation of nuclear and mitochondrial genes. *Biochemical Genetics*, **38**, 27-40.

Song Y, Endepols S, Klemann N *et al.* (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between Old World mice. *Current Biology*, **21**, 1296-1301.

Suzuki H, Tsuchiya K, Takezaki N (2000) A molecular phylogenetic framework for the Ryukyu endemic rodents *Tokudaia osimensis* and *Diplothrix legata*. *Molecular Phylogenetics and Evolution*, **15**, 15-24.

Taberlet P, Fumagalli L (1996) Owl pellets as a source of DNA for genetic studies of small mammals. *Molecular Ecology*, **5**, 301-305.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731-2739.

Teerink BJ (1991) *Hair of West European Mammals: Atlas and Identification Key*. Cambridge University Press. Cambridge.

den Tex R-J, Maldonado JE, Thorington R, Leonard JA (2010) Nuclear copies of mitochondrial genes: another problem for ancient DNA. *Genetica*, **138**, 979-984.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673-4680.

Tougard C, Brunet-Lecomte P, Fabre M, Montuire S (2008a) Evolutionary history of two allopatric *Terricola* species (Arvicolinae, Rodentia) from molecular, morphological, and palaeontological data. *Biological Journal of the Linnean Society*, **93**, 309-323.

Tougard C, Renvoisé E, Petitjean A, Quéré J-P (2008b) New insight into the colonization processes of common voles: Inferences from molecular and fossil evidence. *PLoS ONE*, **3**, 1-10.

Triant DA, Dewoody JA (2006) Accelerated molecular evolution in *Microtus* (Rodentia) as assessed via complete mitochondrial genome sequences. *Genetica*, **128**, 95-108.

Triant DA, DeWoody JA (2007) The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematics and phylogeography. *Journal of Mammalogy*, **88**, 908–920.

Triant DA, DeWoody JA (2008) Molecular analyses of mitochondrial pseudogenes within the nuclear genome of arvicoline rodents. *Genetica*, **132**, 21-33.

Waits LP, Paetkau D (2005) Non-invasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, **69**, 1419-1433.

Wójcik JM, Kawałko A, Marková S *et al.* (2010) Phylogeographic signatures of northward post-glacial colonization from high-latitude refugia: a case study of bank voles using museum specimens. *Journal of Zoology*, **281**, 249-262.

Yasuda SP, Vogel P, Tsuchiya K *et al.* (2005) Phylogeographic patterning of mtDNA in the widely distributed harvest mouse (*Micromys minutus*) suggests dramatic cycles of range contraction and expansion during the mid- to late Pleistocene. *Canadian Journal of Zoology*, **83**, 1411-1420.

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

Data Accessibility

DNA sequences: GenBank accessions JX457608–JX457818; Detailed information on each sampled specimen (locality, haplotypes and accession numbers) and DNA haplotype alignments: DRYAD entry doi:10.5061/dryad.fv21h.

8. Supporting information

Table S1. Sequence information obtained in this study and from GenBank for both the 'long' and 'short' (selected for non-invasive samples) fragments of: (A) cyt-*b*; and (B) IRBP (see text for further details). Haplotype names are provided for the new sequences obtained in this study; those including an extra 'S' are short fragment haplotypes. The number of individuals that generated long fragment sequences is listed per locality (short fragments in parentheses). The location of GenBank sequences is represented by an asterisk.

A. Cyt-*b* gene

Species	Haplotype	GenBank Accession Number	Author	Locality														
				PT	ES	AD	FR	GB	CH	IT	BE	NL	DNB	PL	CZ	BLK	SKD	FI
Mus musculus	(Lab strain)	AY172335	Bibb *et al.* 1981															
		AY057804	Lundrigan *et al.* 2002													*		
		HQ270440	Bonhomme *et al.* 2011	(1)						(*)								
		AY332705	Pfunder *et al.* 2004						(*)									
	Mmus1	JX457716	This study	1														
	Mmus2	JX457719		2														
	Mmus3	JX457717		1														
	Mmus4	JX457715		1														
	Mmus5	JX457718		1														
	Mmus6	JX457722			1													
	Mmus7	JX457721			1													
	Mmus8	JX457724			1													
	Mmus9	JX457723			1													
	Mmus10	JX457720			1													
Mus spretus		AF159398	Martin *et al.* 2000			*												
	Mspr1	JX457726	This study	1(9)	3													
	Mspr2	JX457725		1														
Apodemus flavicollis		AB032853	Serizawa *et al.* 2000						*									
		JF819964	Kryštufek *et al.* 2011													*		
		JF819962														*		
		JF819970														*		
		JF819968														*		
		AJ605663	Michaux *et al.* 2004														(*)	
		AJ605608													(*)			
		AJ605600											(*)					
		AJ605613					(*)											
		AJ605659			(*)													
		AJ605648											(*)					
		FJ389660	Dubey *et al.* 2009										(*)					
	Afla1	JX457730	This study		1			1						1				1
	Afla2	JX457731			1									1				
	Afla3	JX457729					1							1				
	Afla4	JX457732												1				
	Afla5	JX457727						1										
	Afla6	JX457728					1											

Species	Haplotype	GenBank Accession Number	Author	Locality													
				PT	ES	AD	FR	GB	CH	IT	BE	NL	DNB	PL	CZ	BLK	SKD	FI
Apodemus sylvaticus		JF819971	Kryštufek *et al.* 2011													*		
		JF819973														*		
		JF819980														*		
		JF819978														*		
		JF819976														*		
		AB033695	Suzuki *et al.* 2000									*						
		FJ389638	Dubey *et al.* 2009														(*)	
		FJ389653					(*)											
		AJ311148	Michaux *et al.* 2002							(*)								
		AY332700	Pfunder *et al.* 2004	(1)					(*)									
		AJ511902	Michaux et al. 2003										(*)					
		AJ511906									(*)							
		AJ511947														(*)		
		AJ511911			(*)													
		AJ511881			(*)													
	Asyl1	JX457739	This study		1													
	Asyl2	JX457733			1													
	Asyl3	JX457734			1													
	Asyl4	JX457735		1(1)														
	Asyl5	JX457736		1														
	Asyl6	JX457737		1(1)														
	Asyl7	JX457738		1(1)														
	AsylS1	§		(1)														
	AsylS2	§		(1)														
	AsylS3	§		(1)														
	AsylS4	§		(1)														
Micromys minutus		AB201981	Yasuda *et al.* 2005										*					
		AB125100											*					
		AB201982											*					
		AB125099											*					
		AB201983						*										
		AB125094							*									
		AB125097							*									
		AB125098					*											
	Mmin1	JX457740	This study															4
Rattus norvegicus		AJ428514	Nilsson *et al.* 2003														*	
	Rnor1	JX457741	This study	1	1													
	RnorS1	§		(1)														
Rattus rattus		FJ355927	Nilsson *et al.* 2010	(2)													*	
	Rrat1	JX457743	This study		1													
	Rrat2	JX457742		1														
Arvicola sapidus		FJ539342	Centeno-Cuadros & Godoy 2010		*													
	Asap1	JX457748	This study	1														
	Asap2	JX457744		1														
	Asap3	JX457745		1														
	Asap4	JX457746		1														
	Asap5	JX457747		2														
	Asap6	JX457749		1														
Arvicola scherman	Asch1	JX457750	This study		1													
	Asch2	JX457751			1													

Species	Haplotype	GenBank Accession Number	Author	Locality														
				PT	ES	AD	FR	GB	CH	IT	BE	NL	DNB	PL	CZ	BLK	SKD	FI
Chionomys nivalis		AY513847	Jaarola *et al.* 2004										*					
		AY513846								*								
		AY513848			*													
		AM392367	Galewski *et al.* 2006		*													
		GQ150791	Castiglia *et al.* 2009													(*)		
		GQ150792														(*)		
		GQ150795								(*)								
		GQ150797								(*)								
		GQ150799								(*)								
		GQ150802								(*)								
		AY332716	Pfunder *et al.* 2004						(*)									
		GU954316	Fink *et al.* 2010						(*)									
	Cniv1	JX457753	This study		1													
	Cniv2	JX457754			1													
	Cniv3	JX457752			1													
Microtus agrestis P		AY167186	Jaarola & Searle 2002	*														
		JX284253	Paupério *et al.* 2012	*														
		JX284254		*														
		JX284267		*														
		JX284268		*														
		JX284262		*														
		JX284259		*														
		JX284270		*														
		JX284257		*														
		JX284265		*														
		AY303160	Jaarola &Searle 2004	(*)														
Microtus agrestis S		AY167187	Jaarola & Searle 2002		*													
		AY167160			*													
		AY167161			*													
		AY167162			*													
		AY167163			*													
		AY167164			*													
		JX284274	Paupério *et al.* 2012		*													
		JX284273			*													
		JX284282				*												
		JX284283				*												
		AY303159	Jaarola & Searle 2004		(*)													
		AY303155			(*)													
		AY303156		(*)														
	MagSS1	§	This study		(1)													
Microtus arvalis		EU439459	Borkowska & Ratkiewicz 2008											*				
		EU439457												*				
		AY220783	Haynes *et al.* 2003					*										
		AY220788						*										
		AY220780						*										
		AY220785						*										
		AY220787					*											
		AY220789			*(3)													
		AM991095	Tougard *et al.* 2008b				*											
		AM991094					*											
		AY708505	Fink *et al.* 2004												(*)			
		AY708461											(*)					
		AY708507											(*)					

Species	Haplotype	GenBank Accession Number	Author	Locality														
				PT	ES	AD	FR	GB	CH	IT	BE	NL	DNB	PL	CZ	BLK	SKD	FI
Microtus arvalis		AY708495	Fink *et al.* 2004										(*)					
		AY708481							(*)									
		AY708513							(*)									
		AY708503			(*)													
		AY708502			(*)													
	Marv1	JX457755	This study		1													
Microtus cabrerae		AY513789	Jaarola *et al.* 2004	*(19)														
		AY513788		*														
		JX284284	Paupério *et al.* 2012	*														
	Mcab1	JX457756	This study	5														
	Mcab2	JX457759		2(7)														
	Mcab3	JX457757		3														
	Mcab4	JX457758		3														
	Mcab5	JX457760		1														
Microtus duodecimcostatus		AJ717744	Tougard *et al.* 2008a				*											
		AY513797	Jaarola *et al.* 2004	*														
		AY513796		*														
	Mduo1	JX457799	This study		1													
	Mduo2	JX457797		1														
	Mduo3	JX457798		1														
	Mduo4	JX457800		1														
	Mduo5	JX457801		1														
	Mduo6	JX457802		1														
	Mduo7	JX457803		2														
	Mduo8	JX457804		2														
	Mduo9	JX457805		1														
	Mduo10	JX457806		1														
Microtus lusitanicus		AJ717746	Tougard *et al.* 2008a		(2)		*											
		AY513813	Jaarola *et al.* 2004		*													
		AY513812			*													
	Mlus1	JX457769	This study		1													
	Mlus2	JX457770			1													
	Mlus3	JX457771			1													
	Mlus4	JX457773			1													
	Mlus5	JX457774			1													
	Mlus6	JX457775			1													
	Mlus7	JX457776			1													
	Mlus8	JX457763		1														
	Mlus9	JX457764		1														
	Mlus10	JX457762		1														
	Mlus11	JX457761		1(1)														
	Mlus12	JX457772		1														
	Mlus13	JX457780		1														
	Mlus14	JX457781		1														
	Mlus15	JX457782		1														
	Mlus16	JX457777		1														
	Mlus17	JX457786		1														
	Mlus18	JX457787		1														
	Mlus19	JX457778		1														
	Mlus20	JX457779		1														
	Mlus21	JX457788		1														
	Mlus22	JX457789		1														

Species	Haplotype	GenBank Accession Number	Author	Locality														
				PT	ES	AD	FR	GB	CH	IT	BE	NL	DNB	PL	CZ	BLK	SKD	FI
Microtus lusianicus	Mlus23	JX457783	This study	1														
	Mlus24	JX457796		1(2)														
	Mlus25	JX457790		1(1)	(1)													
	Mlus26	JX457791		1														
	Mlus27	JX457792		1														
	Mlus28	JX457793		1														
	Mlus29	JX457794		1														
	Mlus30	JX457795		1														
	Mlus31	JX457765		1														
	Mlus32	JX457766		1														
	Mlus33	JX457767		1														
	Mlus34	JX457784		1														
	Mlus35	JX457768		1														
	Mlus36	JX457785		1														
	MlusS1	§		(2)														
	MlusS2	§		(4)														
	MlusS3	§		(1)														
Microtus gerbei		AY513799	Jaarola *et al.* 2004		*(1)													
		AY513800			*(3)													
		AY513801			*													
		AY513802			*													
		AJ717748	Tougard *et al.* 2008a		*													
	Mger1	JX457807	This study		1													
Myodes glareolus		AF119272	Conroy & Cook 1999															*
		AY309420	Cook *et al.* 2004					*										
		AY309421						*										
		FJ881473	Wójcik *et al.* 2010											*				
		FJ881474												*				
		FJ881475												*				
		FJ881476												*				
		FJ881477												*				
		FJ881478												*				
		FJ881479												*				
		HQ288403	Malé *et al.* 2012														(*)	
		HQ288339														(*)		
		HQ288349													(*)			
		HQ288331											(*)					
		HQ288365															(*)	
		HQ288385					(*)											
		DQ768151	Direct submission										(*)					
		HE604965	Colangelo *et al.* 2012							(*)								
	Mgla1	JX457811	This study			1												
	Mgla2	JX457808			1													
	Mgla3	JX457810			1													
	Mgla4	JX457809			1													
	MglaS1	§			(1)													
	MglaS2	§			(3)													
Eliomys quercinus		HE613994	Direct submission													(*)		
		HE614002											(*)					
		HE614004									(*)							
		HE614000								(*)								
		HE614006					(*)											
		HE613991		(*)														

Species	Haplotype	GenBank Accession Number	Author	Locality														
				PT	ES	AD	FR	GB	CH	IT	BE	NL	DNB	PL	CZ	BLK	SKD	FI
Eliomys quercinus	Eque1	JX457812	This study		1													
	Eque2	JX457813			1													
	Eque3	JX457816			1													
	Eque4	JX457814		1														
	Eque5	JX457815		2														
Glis glis		AJ001562	Reyes *et al.* 1996				*											
		FM160664	Hurner *et al.* 2010													(*)		
		FM160651											(*)					
Sciurus vulgaris		AJ238588 [1]	Reyes *et al.* 2000	(1)														
	Svul1	JX457817	This study	20														
	Svul2	JX457818		1														

Country codes: PT – Portugal; ES – Spain; AD – Andorra; FR – France; GB – Great Britain; CH – Switzerland; IT – Italy; BE – Belgium; NL – Netherlands; DNB – Danubian River Countries (Austria, Hungary, Slovakia and Germany); PL – Poland; CZ – Czech Republic; BLK – Balkans (Greece, Bulgaria, Macedonia, Montenegro, Bosnia and Herzegovina, Croatia and Slovenia); SKD – Scandinavia (Sweden, Norway and Denmark); FI – Finland. [1] Location of origin of the haplotype not available. The haplotypes not listed in GenBank (§) are available in Dryad (doi: 10.5061/dryad.fv21h).

B. IRBP gene

Species	Haplotype	GenBank Accession Number	Author	Locality									
				PT	ES	AD	FR	GB	CH	IT	NL	PL	FI
Mus musculus	(Lab strain)	NM_015745	Danciger *et al.* 1990	3(2)	3								
	Mmus1	JX457616	This study	3	9								
Mus spretus	(Lab strain)	AJ698883-1	Chevret *et al.* 2005	(12)	*(2)								
		AJ698883-2			*								
	Mspr1	JX457618	This study	4									
	Mspr2	JX457619-1			1								
	Mspr3	JX457619-2			1								
Apodemus flavicollis		AB032860	Serizawa *et al.* 2000						*				
	Afla1	JX457620	This study					3					
	Afla2	JX457621-2						1					
	Afla3	JX457622-1			1(2)								
	Afla4	JX457622-2			1								
	Afla5	JX457623-1											2
	Afla6	JX457623-2										1	2
	Afla7	JX457624-1										2	
	Afla8	JX457625										5	
	Afla9	JX457626-2										1	
	Afla10	JX457627-2										1	
Apodemus sylvaticus		AB032863	Serizawa *et al.* 2000	(3)	2(4)						*		
	Asyl1	JX457629	This study	1(2)	1			4					
	Asyl2	JX457630			1			2					
	Asyl3	JX457631-1		1(1)	1								
	Asyl4	JX457633		2	3								
	Asyl5	JX457632-2			1								
	Asyl6	JX457634-1			1								
	Asyl7	JX457635-1			1								
	Asyl8	JX457635-2		1	1								
	Asyl9	JX457636-1		(1)	1								
	Asyl10	JX457636-2			1								
	Asyl11	JX457637-1			1								
	Asyl12	JX457637-2			1								
	Asyl13	JX457639-1		1	1								
	Asyl14	JX457640			2								
	Asyl15	JX457641			2								
	Asyl16	JX457642-2			1								
	Asyl17	JX457643-1		2									
	Asyl18	JX457643-2		1									
	Asyl19	JX457644-1		1									
	Asyl20	JX457644-2		1									
	Asyl21	JX457645-1		1									
	Asyl22	JX457645-2		1									
	Asyl23	JX457646-1		1									
	Asyl24	JX457646-2		1									
	Asyl25	JX457647-2		2									
	Asyl26	JX457648-2		1(1)									
	Asyl27	JX457649-1		1									
	Asyl28	JX457649-2		1									
	Asyl29	JX457650-2		1									
	Asyl30	JX457651-2		1									

Species	Haplotype	GenBank Accession Number	Author	Locality									
				PT	ES	AD	FR	GB	CH	IT	NL	PL	FI
Apodemus sylvaticus	Asyl31	JX457653	This study	2									
	Asyl32	JX457654-1		1									
	Asyl33	JX457654-2		1									
Micromys minutus	Mmin1	JX457656	This study										4
	Mmin2	JX457655-2											1
	Mmin3	JX457657-2											1
Rattus norvegicus	Rnor1	JX457658-2	This study	1(2)	1								
	Rnor2	JX457659		3	3								
Rattus rattus	Rrat1	JX457660	This study	2(2)	6(2)								
Arvicola sapidus	Asap1	JX457661-1	This study	4(2)	(4)								
	Asap2	JX457661-2		1									
	Asap3	JX457662		6									
	Asap4	JX457668		6									
	Asap5	JX457663-2		1									
	Asap6	JX457664-2		1									
	Asap7	JX457666-2		1									
Arvicola scherman	Asch1	JX457669	This study		2(8)								
	Asch2	JX457670			2								
Chionomys nivalis	Cniv1	JX457673	This study		3(9)								
	Cniv2	JX457672			3								
	CnivS1	JX457608-2			(1)								
Microtus agrestis P	MagP1	JX457674-1	This study	7									
	MagP2	JX457674-2		1									
	MagP3	JX457675-2		3									
	MagP4	JX457676-2		1									
	MagP5	JX457677-2		1									
	MagP6	JX457678-2		1									
	MagP7	JX457679-1		1									
	MagP8	JX457679-2		4									
	MagP9	JX457680-2		2									
	MagP10	JX457682-2		1									
Microtus agrestis S	MagS1	JX457683	This study		1	2							
	MagS2	JX457685			2								
	MagS3	JX457684-2			1								
Microtus arvalis	Marv1	JX457686	This study		2(2)	3							
	Marv2	JX457687-2				1							
	Marv3	JX457688-2			2								
	Marv4	JX457689-1			1								
	Marv5	JX457689-2			1								
	MarvS1	JX457609			(4)								
Microtus cabrerae	Mcab1	JX457690	This study	19(8)									
	Mcab2	JX457691-2		2									
	Mcab3	JX457692		4									
	Mcab4	JX457694-2		3									

Species	Haplotype	GenBank Accession Number	Author	Locality									
				PT	ES	AD	FR	GB	CH	IT	NL	PL	FI
Microtus cabrerae	Mcab5	JX457695-2	This study	1									
	Mcab6	JX457696-2		3									
M. duodec	Mduo1	JX457699	This study	12									
Microtus lusitanicus	Mlus1	JX457701	This study	(4)	2								
	Mlus2	JX457702		4(6)	(2)								
	Mlus3	JX457703-1		2									
	Mlus4	JX457703-2		3									
	Mduo1	JX457699		1									
	MlusS1	JX457611-1		(2)									
	MlusS2	JX457612		(2)									
	MlusS3	JX457613			(2)								
Microtus gerbei	Mger1	JX457700	This study		2(1)								
	Mlus1	JX457701			(1)								
Myodes glareolus	Mgla1	JX457705-1	This study		1								
	Mgla2	JX457705-2			2								
	Mgla3	JX457707-1			1(6)								
	Mgla4	JX457707-2			2								
	Mgla5	JX457706-2			1								
	Mgla6	JX457708-2			1								
	Mgla7	JX457709-1			1								
	Mgla8	JX457709-2			1								
	MglaS1	JX457614			(2)								
Eliomys quercinus		AB253958	Nunome et al. 2007							*			
		AB253957-1			(1)		*						
		AB253957-2					*						
		FM162056-1	Blanga-Kanfi et al. 2009						*				
		FM162056-2							*				
	Eque1	JX457710-1	This study		1								
	Eque2	JX457710-2			1								
	Eque3	JX457711-1			1								
	Eque4	JX457711-2			1								
	Eque5	JX457712-1			1								
	Eque6	JX457712–2			1								
	Eque7	JX457713-1		1									
	Eque8	JX457713–2		1									
	Eque9	JX457714-1		1									
	Eque10	JX457714–2		1									
	EqueS1	JX457615-1		(1)									
Glis glis		AB253961	Nunome et al. 2007						*				
		AB253962			6(6)	(4)				*			
Sciurus vulgaris		AY227620	Mercer and Roth 2003	8(6)					*				

Country codes: PT – Portugal; ES – Spain; AD – Andorra; FR – France; GB – Great Britain; CH – Switzerland; IT – Italy; NL – Netherlands; PL – Poland; FI – Finland. The haplotypes are available in Dryad (doi: 10.5061/dryad.fv21h).

Table S2. Number of samples analysed, and haplotype and nucleotide diversity and variability obtained for the cyt-*b* and IRBP 'long' and 'short' fragments (see text).

Species	cyt-*b*										IRBP									
	Long fragment					Short fragment					Long fragment					Short fragment				
	N	*H*	*Hd*	π	*V(Pi)*	*N*	*H*	*Hd*	π	*V(Pi)*	*N*	*H*	*Hd*	π	*V(Pi)*	*N*	*H*	*Hd*	π	*V(Pi)*
Mus musculus	13(2)	12	0.987 (0.001)	0.009	40(13)	3(2)	5	0.533 (0.020)	0.007	7(2)	10(1)	2	0.505 (0.003)	0.000	1(1)	1	1	0.000 (0.000)	0.000	0
Mus spretus	6(1)	3	0.700 (0.050)	0.004	11(0)	9	2	0.133 (0.013)	0.002	2(0)	4(1)	5	0.786 (0.023)	0.003	10(6)	7	2	0.091 (0.007)	0.000	1(0)
Apodemus flavicollis	16(5)	11	0.933 (0.002)	0.006	25(14)	(7)	10	0.727 (0.010)	0.010	12(2)	11(1)	11	0.934 (0.001)	0.029	14(13)	1	7	0.826 (0.002)	0.007	5(4)
Apodemus sylvaticus	13(6)	13	1.000 (0.000)	0.020	71(49)	17(9)	22	0.977 (0.000)	0.022	20(10)	28(1)	34	0.973 (0.000)	0.004	33(20)	6	10†	0.711 (0.002)	0.006	8(6)
Micromys minutus	12(8)	9	0.901 (0.004)	0.003	16(5)	-	4	0.626 (0.012)	0.004	3(1)	3	3	0.600 (0.046)	0.005	13(4)	-	2	0.333 (0.046)	0.003	2(0)
Rattus norvegicus	3(1)	2	0.667 (0.100)	0.004	7(0)	1	3	0.833 (0.049)	0.007	2(1)	4	2	0.429 (0.028)	0.000	1(1)	1	1	0.000 (0.000)	0.000	0
Rattus rattus	3(1)	3	1.000 (0.070)	0.002	4(0)	2	1	0.000 (0.000)	0.000	0	4	1	0.000 (0.000)	0.000	0	2	1	0.000 (0.000)	0.000	0
Arvicola sapidus	8(1)	7	0.994 (0.005)	0.007	26(10)	-	3	0.464 (0.040)	0.007	4(1)	10	7	0.811 (0.003)	0.001	6(3)	3	3	0.151 (0.009)	0.001	2(0)
Arvicola scherman	2	2	1.000 (0.250)	0.010	12(0)	-	1	0.000 (0.000)	0.000	0	2	2	0.667 (0.042)	0.003	5(5)	4	1	0.000 (0.000)	0.000	0
Chionomys nivalis	7(4)	7	0.964 (0.006)	0.020	62(44)	(8)	13	0.967 (0.001)	0.026	14(7)	3	2	0.600 (0.017)	0.001	1(1)	5	2	0.125 (0.011)	0.001	1(0)
Microtus agrestis P	(10)	10	1.000 (0.002)	0.004	22(2)	(1)	5	0.618 (0.027)	0.006	5(1)	11	10	0.879 (0.003)	0.003	10(3)	-	1	0.000 (0.000)	0.000	0
Microtus agrestis S	(10)	10	1.000 (0.002)	0.007	31(10)	4(3)	11	0.956 (0.002)	0.015	11(5)	3	3	0.857 (0.012)	0.002	3(1)	-	2†	0.250 (0.032)	0.002	2(0)
Microtus arvalis	11(10)	11	0.985 (0.002)	0.018	65(41)	11(8)	16	0.961 (0.001)	0.029	17(14)	5	5	0.756 (0.017)	0.002	19(13)	3	2	0.400 (0.013)	0.004	2(2)
Microtus cabrerae	17(3)	8	0.875 (0.003)	0.002	9(6)	26	2	0.365 (0.005)	0.002	1(1)	16	6	0.629 (0.008)	0.001	4(3)	4	1	0.000 (0.000)	0.000	0
Microtus duodecimcostatus	15(3)	13	0.981 (0.001)	0.015	54(40)	14	32	0.950 (0.000)	0.036	25(19)	6	1	0.000 (0.000)	0.000	0(0)	9	6	0.568 (0.007)	0.003	5(5)
Microtus lusitanicus	39(3)	39	1.000 (0.000)	0.025	131(84)						6	5†	0.833 (0.005)	0.003	8(8)					
Microtus gerbei	6(5)	6	1.000 (0.010)	0.013	33(16)	4	4	0.644 (0.023)	0.010	6(3)	1	1	-	-	-	1	2†	0.500 (0.070)	0.002	1(0)
Myodes glareolus	14(10)	14	1.000 (0.001)	0.020	59(36)	12(8)	18	0.996 (0.000)	0.027	29(10)	5	8	0.956 (0.004)	0.004	9(9)	4	4	0.714 (0.010)	0.005	3(0)
Eliomys quercinus	6	5	0.933 (0.010)	0.006	15(9)	(8)	10	0.923 (0.004)	0.030	17(10)	8(3)	15	0.992 (0.000)	0.006	26(19)	1	7	0.693 (0.013)	0.004	6(3)
Glis glis	6*(1)	1	0.000 (0.000)	0.000	0	(2)	3	1.000 (0.074)	0.012	3(0)	5(2)	2	0.303 (0.022)	0.001	4(4)	5	2	0.189 (0.012)	0.001	1(1)
Sciurus vulgaris	22(1)	3	0.177 (0.010)	0.000	4(2)	1	1	0.000 (0.000)	0.000	0	5(1)	1	0.000 (0.000)	0.000	0	3	1	0.000 (0.000)	0.000	0
Total	239(85)	189				130(56)	166				150(10)	126				60	58			

N – total number of sequences of which those from GenBank are shown in parentheses; H – number of haplotypes; Hd – haplotype diversity (standard deviation in parentheses); π – nucleotide diversity; V(Pi) – number of variable sites of which the number of parsimony informative sites are shown in parentheses; For the 'Short Fragment', number of haplotypes (H) derived from the data on long fragments plus new non-invasive haplotypes. For the IRBP gene, N refers to the number of unphased sequences. Hd and π for the short fragment were calculated using all long and short fragment sequences. *All tissue samples used in this work (5) resulted in numt sequences. †H considering one shared haplotype with each respective

Table S3. Variable nucleotide positions for the cyt-*b* 'short' fragment (see text) with numbering following the *Mus musculus* mitochondrial genome sequence, accession number AY172335.1 (Bibb *et al.* 1981).

Species	Nucleotide Position
	14582 14585 14588 14589 14591 14594 14597 14598 14600 14603 14606 14609 14610 14612 14615 14616 14618 14621 14622 14624 14627 14630 14633 14636 14639 14642 14645 14648 14651 14654 14657 14658 14660 14663 14666 14667 14669 14670 14672 14675 14678 14681 14684 14687 14690 14693 14696 14699 14702 14705 14708 14709 14711 14712 14714 14717 14718 14720 14721 14722 14723 14724 14725 14726 14729 14732 14733 14735 14736 14738 14741 14742 14744
Mus musculus Ref.Seq.	T A C C C C A A G C C A T A T A A A C C A C A A T A G C C A A C A A C C T G A C A C C T C C C C T A A T T A C G G C C A G C A A T C T C C C C C C C C
Mus musculus	. . . . . R . . R . . . . . M . . . . . Y . . . . . Y . . Y . Y . . .
Mus spretus	. . . . . T . . . . . M . . G . . . . . T . C A . . . . . C . . . . . G . C C . . . A . . . . . R . . . . . T . .
Apodemus flavicollis	. . T . A T . . . A Y . C . C . R T . . . A R . C . A A . . . T . . Y A C A . A T Y Y C . . T . . . . C C . T R Y . . . . T . . . H C Y . Y . A T . H
Apodemus sylvaticus	. . T Y A . . R . A T . C . C . . T . . . A R . C . R A . . . . . T A . R . A T T . Y T . T . C T . . Y . T . T . . . . . T . . . . T G Y . .
Micromys minutus	C C . . . . . T T . . G A . . C . . . T . . C . A A . C . T . . A . C A . A T . . A . T . T C . C C C . . . R . . . . . R . T . T . T . T T . W
Rattus norvegicus	. . . . A T . W . T . T C . . G . T . . . . . C . A . . . . . A . C A . A C . . A . . Y . C C . C . . . . C . . T . . . . . T A T . T . T . . .
Rattus rattus	C . . . A T . . . . T C C . . C . C T . . . . . C . A . . . . . A . C A . A T T . A . . . . C C . C C . . . C . . T . . . . . T A T . . . . .
Arvicola sapidus	. . Y . A . . R . . . C C . C C . . T . . T . . C . A A . . T . . R . T C T . A . . . A . T T . C C . . Y . T A C T . C . T C T . A . . . . . A . . T
Arvicola terrestris	. . T . . . . C C . . C . . . . . T . . C . . A . . C . . . . T C T . A . . T A . T T T C T . . C . . . C T . T . T C T . A . . T A T A . . A
Chionomys nivalis	Y . . . . Y . . . . . C . Y H . V A . . A . . C . R R . . T . . . Y T C Y . A . . Y C T . . T . R T Y C . . A C . . T . T T T . A A T . R . A Y . A
Microtus agrestis P	. . . . W . . . . . T C C G A C . C A . . A R . C . R T T . . . . . M C C . A . . . C . . . . C . C . C . Y A C T . . . T T C . A A T T A . A T . A
Microtus agrestis S	Y . . Y A T . . . . . C C G A C . C M . . A . . Y . R R T . . . . C C . A . T . C . . . . C W M . C . T A C Y . . . T T C . A A T Y A . A Y . A
Microtus arvalis	. . T . . Y . . R A . C C . Y T . W A Y . A R R . . R Y . . . Y . . . . . C C . A . . Y C . . . . C . C . Y . T A C Y . H . T . T . A Y Y . T T A Y . T
Microtus cabrerae	. . . . A . . . . A . . . . . C . . . A . T A . . C G T . . . . T . . T A C C . A . T . C . . . . C R . . C . . A T . . C . T . T . A C . . . . A . . T
Microtus DL	. . . . . C C . C R . B A . R R R R C . V . . . . . R H Y C T . R . T . Y . . . . C R C . C . Y A C . . T . T . T . A Y . . D . R Y T A
Microtus gerbei	. . . . Y . . . . . Y C . C . . C R . . A . . C . Y G . . . T . . . . C Y R A . . . C . . . . C . C . C . T A A . . T . T . C . A C T . A . A . T A
Myodes glareolus	Y . Y . A Y . . . Y Y Y Y . Y M . . . . . A . . C R V R . . . B . . Y Y C Y . A . . . A . . . . . C G T C Y . Y A A Y . Y . T Y Y . H C Y Y R Y R Y . A
Eliomys quercinus	. . T . . . . R Y . Y . C C . C . . . A Y R A R . C . R S . C . . R . W . C A . A . . . A . . Y T C C . C C . Y A T . T . . T . R Y A V . . . T A Y Y A
Glis glis	. T T . T T . T . . . Y C . C T . T . T . A . . . . C . . C . . . . T T C A . . . T Y A T . T T C . C C . . T . A . . . . T . . Y A C . . T . A . . T
Sciurus vulgaris	C . . . . T G T . T T T . . . . . C . . . T . . C . C A . . . T . . A T C A . A . . . C . . T . C T T C C G . A . . . . . T T . . A A . T A . A . . T

Table S4. Variable nucleotide positions for the IRBP 'short' fragment (see text) with numbering following the *Mus musculus* Rbp3, accession number NM_015745.2 (Danciger *et al.* 1990).

Nucleotide Position

Species	825	826	828	830	834	835	840	843	848	852	858	860	861	867	868	873	879	885	891	892	897	898	900	903	909	912	913	915	918	921	931	933	936	939	940
Mus musculus Ref.Seq.	G	C	C	C	C	G	G	C	G	G	T	G	G	G	G	C	G	A	C	C	C	A	C	C	C	A	G	G	T	G	C	G	A	T	C
Mus musculus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mus spretus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Apodemus flavicollis	C	A	G	.	.	.	.	T	.	.	.	.	A	.	.	.	R	.	.	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Apodemus sylvaticus	C	A	G	.	Y	.	R	T	.	R	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Micromys minutus	C	A	G	.	.	.	.	T	.	A	.	.	T	R	R	.	.	.	.	.	.	.	.	.	.	.	.	A	.	T	.	.	.	C	.
Rattus norvegicus	C	A	G	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	T	.	T	G	.	.	T	.	.	.	.	.	.	.	G	C	.
Rattus rattus	C	A	G	.	.	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	G	C	.
Arvicola sapidus	.	A	G	G	.	.	.	T	.	.	.	.	C	.	.	G	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	G	C	T
Arvicola terrestris	.	A	G	G	.	.	.	.	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	G	C	T
Chionomys nivalis	C	A	G	G	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	A	.	.	.	G	C	T
Microtus agrestis P	.	A	G	G	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	.	C	T
Microtus agrestis S	.	A	G	G	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	Y	.	.	.	C	.	C	.	.	.	.	C	T
Microtus arvalis	.	A	G	G	.	.	.	.	.	R	.	.	T	.	.	.	.	.	.	.	.	.	.	Y	.	.	C	.	C	.	.	.	.	C	T
Microtus cabrerae	.	A	G	G	.	.	.	.	.	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	.	C	T
Microtus DL	.	A	G	G	Y	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	C	.	.	C	.	C	T
Microtus gerbei	.	A	G	G	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	C	.	.	C	.	C	T
Myodes glareolus	T	A	G	G	T	.	.	.	R	A	.	.	T	A	.	.	.	.	Y	G	.	.	.	.	.	.	.	.	C	.	.	.	T	C	T
Eliomys quercinus	.	A	G	.	Y	K	.	Y	.	C	G	S	C	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	C	.	.	.	G	G	.
Glis glis	A	A	G	.	.	.	.	.	.	C	G	.	C	.	.	.	.	G	.	.	.	.	.	.	.	C	.	.	C	.	.	.	G	.	.
Sciurus vulgaris	C	A	G	.	.	.	.	T	.	T	A	.	T	.	.	.	.	G	T	.	.	.	T	.	.	T	.	A	.	.	T	.	.	C	.

Species	942	945	951	952	954	957	960	966	969	972	975	979	981	984	987	991	993	995	996	999	1002	1005	1006	1008	1012	1014	1017	1020	1023	1024	1029
Mus musculus Ref.Seq.	A	T	T	G	C	G	A	A	C	T	C	C	G	T	T	G	G	C	C	T	A	G	C	A	C	A	A	G	C	C	C
Mus musculus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mus spretus	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Apodemus flavicollis	G	.	Y	.	.	.	R	.	.	.	A	.	.	C	.	A	.	.	A	.	.	.	.	.	.	.	G	.	.	.	.
Apodemus sylvaticus	G	.	C	.	.	.	G	.	.	.	A	Y	.	C	.	A	.	.	A	.	.	R	.	.	.	.	G	.	.	.	.
Micromys minutus	.	.	.	.	.	.	G	.	.	.	A	.	.	C	.	.	.	T	A	.	.	.	.	.	.	.	.	.	G	.	T
Rattus norvegicus	.	.	C	.	.	A	.	.	.	C	G	.	.	C	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.
Rattus rattus	.	.	A	.	.	A	G	.	.	C	G	.	.	C	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.
Arvicola sapidus	G	A	.	.	.	.	G	.	.	.	A	.	.	C	C	.	.	T	A	A	.	.	A	.	.	.	.	.	.	T	.
Arvicola terrestris	G	A	.	.	.	.	G	.	.	.	A	.	.	C	C	.	.	T	A	A	.	.	A	.	.	.	.	.	.	T	.
Chionomys nivalis	G	A	.	.	T	.	.	.	.	.	A	.	.	C	C	.	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Microtus agrestis P	G	A	.	.	T	.	G	.	.	.	A	.	.	C	C	.	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Microtus agrestis S	G	A	.	.	T	.	R	.	.	.	A	.	.	C	C	.	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Microtus arvalis	G	A	.	.	T	.	G	.	.	.	A	.	.	C	Y	.	.	T	A	A	R	.	.	.	.	.	.	.	.	T	.
Microtus cabrerae	G	A	.	.	.	.	G	.	.	.	A	.	.	C	C	.	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Microtus DL	G	A	.	.	T	.	G	.	.	.	A	.	.	C	C	R	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Microtus gerbei	G	A	.	.	T	.	G	.	.	.	A	.	.	C	C	R	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Myodes glareolus	G	A	.	.	Y	.	G	.	.	.	A	.	.	C	C	.	.	T	A	A	G	.	A	.	.	.	.	.	.	T	.
Eliomys quercinus	G	C	C	A	Y	.	G	.	.	C	.	.	C	C	.	.	R	.	A	A	T	A	.	G	.	G	G	A	.	.	.
Glis glis	G	C	C	A	.	.	.	G	T	.	T	.	C	C	.	.	.	.	G	A	Y	.	.	G	.	G	G	A	A	.	.
Sciurus vulgaris	G	C	C	A	.	.	G	.	.	.	A	T	.	C	.	.	A	.	A	C	T	A	.	.	T	G	G	A	.	.	.

CHAPTER III.

THE EVOLUTIONARY HISTORY OF THE CABRERA VOLE

Paper II. Barbosa S, Paupério J, Pavlova SV, Alves PC and Searle JB (*in prep*) **The *Microtus* voles: resolving the phylogeny of one of the most speciose mammalian genera with nuclear- and mito-genomes**.

Paper III. Barbosa S, Paupério J, Herman JS, Ferreira CM, Pita R, Vale-Gonçalves HM, Cabral JA, Garrido-García JA, Soriguer RC, Beja P, Mira A, Alves PC and Searle JB (2017) **Endemic species may have complex histories: within-refugium phylogeography of an endangered Iberian vole.** *Molecular Ecology*, **26**, 951-967.

The *Microtus* voles: resolving the phylogeny of one of the most speciose mammalian genera with nuclear- and mito-genomes

Barbosa S, Paupério J, Pavlova SV, Alves PC and Searle JB

1. Abstract

Rapid radiations pose some of the greatest challenges in phylogenetics, especially when analysing only a small number of genetic markers. Given that most of these events occurred in a very short period of time, at various points in time, particular challenges have to be addressed to determine the phylogenetic relationships, namely branching order and the time since divergence. With the development of high throughput sequencing, thousands of markers can now readily be obtained to revisit and resolve some of these questions. *Microtus* is a speciose genus currently composed of 65 species that evolved in a very short period of time in the last 2 million years. Although it is a very well-studied group, there is still phylogenetic uncertainty. Building upon previous studies that used mostly a few mitochondrial and nuclear loci, in this work we used partial mitogenomes and 3 426 SNPs obtained through genotyping-by-sequencing (GBS) to clarify the taxonomic position of the described subgenera and the phylogenetic relationships of some species of particular interest. Both types of genome (mitochondrial and nuclear) generated similar tree topologies, with a basal split of the Nearctic (*M. ochrogaster*) and Holarctic (*M. oeconomus*) species, and then a subdivision of the five Old World species into two subgroups. These data support the occurrence of two radiations to Europe (or vicariance within Europe) and a single radiation event to North America, with a later expansion of *M. oeconomus* from Asia to the west and the east. We further resolved the positioning of *M. cabrerae* as sister group of *M. agrestis*, and refute the claim that this species should be elevated to its own genus (*Iberomys*). Finally, we found evidence of ongoing speciation events, especially within *M. agrestis*, with high level of genetic divergence supporting the claim of three different species within this taxon. Similar high levels of divergence were also found within *M. oeconomus*.

2. Introduction

It appears that diversification often occurs through rapid radiations, where an ancestral form takes advantage of a geographical or ecological opportunity that leads to subdivision into many new taxa with distinct ecological niches (Simões *et al.* 2016; Stroud & Losos 2016). The driving factors of rapid radiations are extremely interesting and their elucidation would benefit from detailed phylogenies, but the synchrony of the initial radiation creates challenges in deciphering the branching order and the time since divergence (Whitfield & Lockhart 2007; Giarla & Esselstyn 2015). As well as multiple contemporary divergences over space being a serious challenge to phylogenetics, the analysis of multiple divergences over different time scales can be difficult, because such events are likely affected by different types of evolutionary forces (Jeffroy *et al.* 2006; Degnan & Rosenberg 2009). The ability of molecular phylogenetics to deal with these issues has been hampered in studies using limited data, such as one or two mitochondrial loci, and sometimes a small number of nuclear loci (Simon *et al.* 2006). There are several issues relating to the sole use of mitochondrial DNA (mtDNA) and there has been a recent tendency for a combined approach using different types of markers reflective of different evolutionary trajectories (Morin *et al.* 2004; Simon *et al.* 2006; Melo-Ferreira *et al.* 2012). Fast evolving mitochondrial genes, provide enough resolution to distinguish between recently diverged taxa, but also are more prone to effects of genetic drift and haplotype fixation given their smaller effective population size, and long branch attraction (LBA), given the high chance of mutational saturation and difference in evolutionary rates among lineages (Bergsten 2005; Nabholz *et al.* 2008; Yang & Rannala 2012; Su & Townsend 2015). Thus, mitochondrial DNA often becomes unsuitable for the study of older rapid radiations, but can prove to be very helpful for recent divergences (Hurst & Jiggins 2005) including applying contemporary calibrations (Herman & Searle 2011). Slower evolving markers, namely nuclear exons and introns, have the advantage of avoiding many of the problems of using mitochondrial markers alone (such as LBA), but they often do not display enough genetic variation to resolve recent divergences, being more prone to effects of incomplete lineage sorting (ILS) (Giarla & Esselstyn 2015). With the arrival of high throughput sequencing, genome-wide single nucleotide polymorphism (SNP) variation and whole genome sequences are revolutionising evolutionary studies (Morin *et al.* 2004), including solving difficult issues in phylogenetics. Some examples of the applications of these recent high throughput sequencing methods in phylogenetics ('phylogenomics', Jeffroy *et al.* 2006) range from deep evolutionary events like the eutherian radiation (Murphy *et al.* 2001; Song *et al.*

2012), to more recent diversifications, like the African cichlid fish (Seehausen 2006; Brawand *et al.* 2014). Phylogenomics is increasingly going to replace studies with single or low numbers of markers but it is still desirable to include both nuclear and mitochondrial data in phylogenetics, even at a genomic scale (Moore 1995; Duchêne *et al.* 2011; Filipi *et al.* 2015; Leaché *et al.* 2015).

The genus *Microtus* is part of the rodent family Cricetidae and more specifically the subfamily Arvicolinae, and represents one of the most rapid documented radiations in extant mammals. The *Microtus* radiation was proposed to have originated in central Asia with expansion both westwards – to Europe [*c.* 1 million years (Mya)] – and eastwards – to North America (*c.* 450 kya), followed by multiple speciation events in both regions at various points in time (Repenning 1993; Conroy & Cook 1999, 2000; Jaarola *et al.* 2004; Weksler *et al.* 2010). The later expansion of *M. oeconomus* to Europe and North America is thought to have been occurred *c.* 150 kya, given the estimates of the second opportunity to cross over Beringia – leading to a Holarctic distribution for this species, which is supported by both fossil and molecular data (Lance & Cook 1998; Brunhoff *et al.* 2003; Galbreath & Cook 2004). The biogeographic history of *Microtus* can thus help us better understand the historical biogeography of the northern continents given that this group has species across the Palearctic, Nearctic and Holarctic regions (Conroy & Cook 2000). There are 65 *Microtus* species currently recognised by the IUCN, which are thought to have diversified during the last 2 million years (IUCN 2016): there are 44 species belonging to 5 subgenera that are found in the Old World; there are 20 species belonging to 7 subgenera representing the first wave of subgenera that colonised the New World; and there is one species, *Microtus oeconomus*, with a Holarctic distribution that represents a more recent colonisation of North America (Conroy & Cook 2000).

Many molecular phylogenetic studies on *Microtus* have used mitochondrial cyt-*b* and found that most subgenera represent true monophyletic clades (Conroy & Cook 2000; Jaarola *et al.* 2004; Bannikova *et al.* 2010). Other studies have used both mitochondrial and nuclear markers (single/low numbers of gene sequences and AFLPs) which have added support to the idea of a rapid radiation and two to three independent expansions from central Asia to Europe and two to North America, with associated bursts of speciation (Galewski *et al.* 2006; Robovský *et al.* 2008; Fink *et al.* 2010; Martínková & Moravec 2012). These studies set the first steps towards a greater understanding of the *Microtus* radiation, but the high mitochondrial mutation rate observed in *Microtus* is likely a limiting factor when aiming to resolve the initial splits of this rapid adaptive radiation (due to LBA), and the slower mutation rate of nuclear loci does not allow the effects of

incomplete lineage sorting to be parsed out (Whitfield & Lockhart 2007; Giarla & Esselstyn 2015). As a result, these studies are impaired by phylogenetic inconsistencies across loci, conspecifics splitting into different groups, and low support for the more basal nodes, which is expected when using a small number of loci in groups that have undergone rapid radiations (Fink *et al.* 2010; Martínková & Moravec 2012; Giarla & Esselstyn 2015).

As well as the interest in the overall *Microtus* phylogeny and its biogeographic history, there is an interest in the phylogenetic position of certain species. In particular, the Cabrera vole (*Microtus cabrerae*) has one of the most unresolved positions within the genus which, although an European species, it has been found to often cluster with the North American species (Jaarola *et al.* 2004; Robovský *et al.* 2008; Fink *et al.* 2010). The Cabrera vole is classified in the subgenus *Agricola* together with the field vole (*Microtus agrestis*) based on its karyotype (Zagorodnyuk 1990), however many still prefer its first classification by Chaline (1974) as the sole member of a separate subgenus *Iberomys* (López-García & Cuenca-Bescós 2012; Pita *et al.* 2014). Given the combination of numerous differentiating morphological and biological features – e.g. molar morphology, body size, reproductive strategy, chromosomes (Pita *et al.* 2014) – Cuenca-Bescós *et al.* (2014) have recently proposed the elevation of the Cabrera vole to its own genus, *Iberomys*. However, the elevation of this species to its own genus needs to be considered carefully and be justified on phylogenetic grounds.

The aim of this study was to help clarify the phylogeny of the genus *Microtus*, using mitochondrial and nuclear genomes in a biogeographical context. Given that subgenera are typically monophyletic, we compare taxa from five subgenera defined morphologically following Wilson & Reeder (2005): *Agricola* (*M. agrestis* and *M. cabrerae*), *Alexandromys* (*M. oeconomus*), *Microtus* (*M. arvalis* and *M. levis*), *Pedomys* (*M. ochrogaster*) and *Terricola* (*M. subterraneus*). Our phylogenomic approach using a partial mitogenome and genotyping-by-sequencing (GBS) SNP data demonstrates that increasing the number of loci and their representability of the genome can improve phylogenetic inferences, especially at more recent time scales. However, it is important to be aware that greater detail can be obtained from haplotype rather than SNP data, and even then, rapid radiations might still be very difficult to disentangle.

3. Material and Methods

Sampling

We sampled seven *Microtus* species representative of five out of the 12 recognised subgenera, distributed across the entire northern hemisphere (Figure S1, Supporting information). *M. (Alexandromys) oeconomus*, has a Holarctic distribution, *M. (Pedomys) ochrogaster* is only found in the New World, and five species only occur in the Old World: *M. (Agricola) agrestis*, *M. (Agricola) cabrerae*, *M. (Microtus) arvalis*, *M. (Microtus) levis* and *M. (Terricola) subterraneus*, representing all the subgenera found in Europe. Up to three individuals per taxon were analysed to account for intra-specific variation (Table 1). Whenever possible, conspecific individuals were chosen from geographically distant localities (Figure S1, Supporting information). We also analysed two *Arvicola sapidus* individuals (outgroup), given its close relationship to *Microtus* (Buzan *et al.* 2008; Martínková & Moravec 2012), encompassing a total of 18 specimens analysed separately for their mitochondrial and nuclear genomes (see Table 1 for details). One *M. ochrogaster* and one additional *M. levis* mitochondrial genomes were obtained from GenBank, which were not analysed for nuclear variation. For the nuclear analysis we obtained the full genome of *M. ochrogaster* from GenBank to use as a reference for the SNP calling pipeline, but this genome was not incorporated as a representative of *M. ochrogaster* in our analyses.

Mitochondrial data

Mitochondrial capture: All samples were barcoded for the complete cytochrome-*b* (cyt-*b*) gene following Barbosa *et al.* (2013) to verify species ID and these sequences were later included in the mitogenomic analyses. To obtain mitochondrial sequences we performed a mitochondrial capture technique on two *Arvicola* and 14 *Microtus* samples, following Fu *et al.* (2013), that consisted of two main steps: 1) bait preparation and 2) hybridisation capture. For the bait preparation we performed long-range PCR using one sample of *Microtus cabrerae*, targeting two areas of the mitochondrial genome, one approximately 3 700 base pairs (bp) and the other approximately 4 000 bp (details in Figure S2, Supporting information). We then sheared the DNA into smaller fragments to an average size of 400 bp and prepared the bait for hybridisation with each of our 16 samples (Table 1), which were then individually labelled by double indexing. Library preparation followed Meyer & Kircher (2010) with the modifications described in Kircher (2012) using the standard Illumina multiplex adaptors and indexing PCR followed

Dabney & Meyer (2012). The hybridisation products were sequenced in a paired-end run (500 bp) on a MiSeq desktop sequencer (Illumina) using the MiSeq v2 Reagent Kit. After the run, reads were demultiplexed using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit) and mitogenomes for each sample were assembled with GENEIOUS v8.1.8 (Kearse *et al.* 2012) following two steps: 1) all reads from the same taxon were mapped to the closest related mitochondrial genome available on GenBank [following the phylogeny of Jaarola *et al.* (2004)] to produce a taxon specific reference; 2) reads from each sample were mapped to the taxon specific reference. Only genomic regions with at least 3x coverage were maintained and due to the small size of these genomes, the alignments could be verified by eye. Mitochondrial divergence (*Dxy*) was estimated using *p*-distances between different groups with MEGA v6 (Tamura *et al.* 2013), and the degree of differentiation found both within and between closely related species was compared.

Phylogenetic analyses: For the mitochondrial phylogenies we used only genomic regions with coverage for all 16 samples (plus the two sequences from GenBank) (Figure S2, Supporting information). Sequences were aligned using the ClustalW algorithm (Thompson *et al.* 1994) implemented in Geneious v8.1.8 (Kearse *et al.* 2012). We then built Maximum Likelihood (ML) and Bayesian (BA) trees using RAxML v8.1 (Stamatakis 2014), and ExaBayes v1.4 (Aberer *et al.* 2014), respectively. For all mitochondrial analyses we partitioned the dataset into 5 fully and 2 partially sequenced mitochondrial coding genes and the 10 fully sequenced tRNAs (see 'Results' and Figure S2, Supporting information), for independent analysis. For the ML analysis we applied the GTRGAMMA model with 1000 bootstraps and a rapid ML search. For the BA trees we used ExaBayes v1.4 (Aberer *et al.* 2014) to run four replicates until the average and maximum standard deviation of split frequencies ('asdsf') was lower than 5%, running at least for 1 million generations, each with four heated chains, and sampling every 500 generations using the default model (GTR). Default values were used for tuning and branch swap parameters and branch lengths among partitions were linked. Convergence and proper sampling of the posterior distribution of parameter values were assessed by checking that the effective sample sizes of all estimated parameters and branch lengths were greater than 200 in the Tracer v1.5 software (Rambaut & Drummond 2015), and by ensuring that the 'asdsf' was below 1% (assumed to represent "excellent convergence") and that the potential scale reduction factor across runs was close to 1% (values below 1.1% are considered good convergence). Finally, to check for convergence in topology and clade posterior probabilities, we summarised a consensus

tree from 1 800 post burn-in trees using the ExaBayes 'consense' tool for the four runs, after excluding 10% of the run as burnin.

Table 1. Details of samples used in this study (including geographical provenance) and the sequences generated from them (including coverage for the mitochondrial genome), and taxonomic assignments of the vole species from which they were derived.

Genus	Subgenus	Distribution	Species	Samples				Mitochondrial		Nuclear
				Code	Latitude	Longitude	Country	Mean bp coverage	Accession #	Accession #
Microtus	*Agricola*	Palearctic	*agrestis*	SM0082	41.760	-8.637	Portugal	83.2	XXXX	XXXX
				SM0372	42.974	-1.516	Spain	29.7	XXXX	-
				SM0356	42.309	-3.265	Spain	-	-	XXXX
				SM2434	64.381	47.521	Russia	328.4	XXXX	XXXX
			cabrerae	SM0020	37.610	-8.797	Portugal	131.8	XXXX	XXXX
				SM1994	42.423	-0.437	Spain	349.7	XXXX	XXXX
				SM2049	38.734	-1.174	Spain	29.1	XXXX	XXXX
	Alexandromys	Holarctic	*oeconomus*	SM2444	55.575	41.434	Russia	107.3	XXXX	XXXX
				SM2446	53.453	109.018	Russia	376.9	XXXX	XXXX
	Pedomys	Nearctic	*ochrogaster*	GenBank†			USA	-	KT166982	-
				SM2514	40.144	-88.165	USA	-	-	XXXX
	Microtus	Palearctic	*arvalis*	SM0316	42.474	-5.531	Spain	184.9	XXXX	XXXX
				SM2439	54.501	38.335	Russia	410.8	XXXX	XXXX
			levis	GenBank*			Ukraine	-	DQ015676	-
				SM2450	50.425	46.455	Russia	86.2	XXXX	XXXX
				SM2453	57.154	32.416	Russia	93.9	XXXX	XXXX
			subterraneus	SM2454	56.435	31.311	Russia	442.6	XXXX	XXXX
				SM2456	57.154	32.416	Russia	114.5	XXXX	-
Arvicola	-	Palearctic	*sapidus*	SM0005	37.968	-8.801	Portugal	40.6	XXXX	XXXX
				SM1635	41.913	-8.496	Portugal	108.6	XXXX	XXXX

Nuclear data

Genotyping-by-sequencing and SNP calling: We performed genotyping-by-sequencing (GBS; Elshire *et al.* 2011) at the Cornell Genomic Diversity Facility on 14 *Microtus* and two *Arvicola* samples, following the same protocol as described in White *et al.* (2013) (Table 1). Filtering and SNP calling were performed using the 'TASSEL' pipeline (Bradbury *et al.* 2007), and *Microtus ochrogaster* as reference genome. A minimum base call of five reads and a maximum locus missing data of 20% was defined. For post processing we excluded all loci with missing data and indels, and kept only those loci that had both alleles present in a homozygous state in different individuals (i.e. loci that only had AA and AB individuals were excluded; loci with AA and BB or AA, AB and BB were kept). These filters were applied due to the ML analysis requiring SNP data to fit the ASC_GTRCAT [Lewis correction (Lewis 2001)] model. We additionally calculated Nei's genetic distance (Nei 1972) between all pairs of samples using the function *nei.dist* from the *poppr* R package (Kamvar *et al.* 2014).

Phylogenetic analyses: An initial phylogeny was estimated with the concatenated SNP dataset using maximum likelihood (ML) in the program RAxML v.8.1.16 (Stamatakis, 2014), and the GTR model of nucleotide evolution. Likelihood calculations were corrected for ascertainment bias (-m ASC_GTRCAT). This option is useful for SNP datasets that contain no invariable sites, which will cause the algorithm to exit with errors. Support was assessed by 1 000 nonparametric bootstrap replicates, followed by a search for the best-scoring maximum likelihood tree, using *Arvicola sapidus* as outgroup. For the Bayesian analysis, we ran ExaBayes in the same way as for the mitochondrial data, except that we did not partition the data and ran 2 chains of 1 million generation sampling every 500 generations. Evaluation of convergence followed the same protocol as for the mitochondrial data.

4. Results

Mitochondrial data

For all 16 voles analysed, we confirmed the species ID by matching the cyt-*b* sequences to references in the GenBank nucleotide database and were able to capture mitochondrial DNA using the *Microtus cabrerae* mitochondrion as bait. Capture was not influenced by taxonomic distance, but rather by the initial DNA concentration of the samples concerned (data not shown). Mean base pair (bp) length of each read was around 129 bp and mean coverage was 182 reads (29 - 443) (Table 1). We obtained full

coverage for the target regions and a portion of the flanking regions, resulting in 5 complete coding regions (ATP6, COI, COII, COIII, ND5), two partial coding regions (cyt-*b* and ND4) and 10 tRNAs, including for the two sequences from GenBank (Figure S2, Supporting information). With these data we generated consensus sequences for all samples with varying lengths and mitochondrial regions covered (Figure S2, Supporting information). For the sample for the *Microtus agrestis* Southern lineage (SM0372, Table 1), we found two mitochondrial sequences spanning a large portion of the targeted regions, one similar to the remaining *M. agrestis* mitochondria, and the other differing in a large number of positions. To recover the putative true mitochondrial sequences, we used GENEIOUS to align these reads to the consensus sequence of the remaining two *M. agrestis* samples and kept the least divergent reads. The discarded reads resulted in a sequence with stop codons in coding regions and several insertions and deletions (data not shown), presumably representing a large numt (a nuclear copy of a mitochondrial sequence).

For the mitochondrial phylogenetic analyses, both the ML and BA runs resulted in phylogenies with the same topology and similar levels of support for specific nodes (Figure 1): *Microtus ochrogaster* is basal to the mitochondrial radiation, followed by *M. oeconomus*; both of these species are ancestral to two well supported groups represented by *M. agrestis*/*M. cabrerae* and *M. arvalis*/*M. levis*/*M. subterraneus*; the divergence between *M. agrestis* and *M. cabrerae* is of similar magnitude as *M. subterraneus* from *M. arvalis*/*M. levis*. Within *M. agrestis*, the northern lineage 'N' showed equal level of divergence from the remaining *M. agrestis* lineages as *M. arvalis* from *M. levis* (*Dxy* = 5.8%) (Figure 1). The remaining *Dxy* values show a variety of intraspecific mitochondrial divergences that vary greatly across species (Figure 1), from 0.4% in *M. subterraneus* to 3.6% in *M. oeconomus*.

Nuclear data

Using the TASSEL GBS pipeline and the *M. ochrogaster* genome as a reference, we obtained over 246 400 nuclear SNPs, which were reduced to 5 151 SNPs after excluding all missing data. Additionally, we only maintained loci that showed both alleles present in the homozygous state, resulting in a final supermatrix of 3 426 SNPs with no missing data. The within-species Nei's distance values are very similar to those obtained for the mitochondrial genome, varying from 0.5% in *M. levis* to 3.8% in *M. arvalis* (Figure 1).

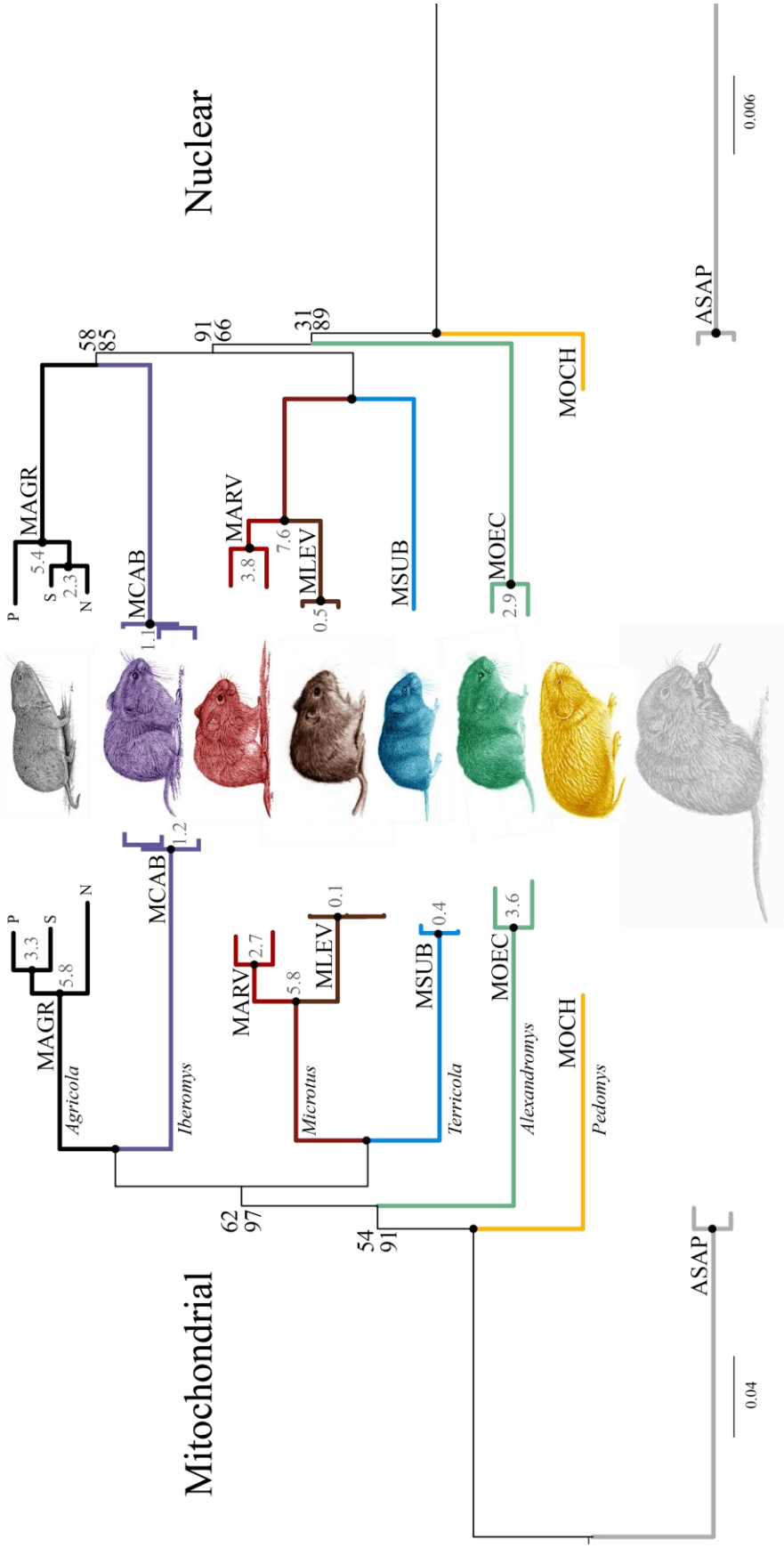


Figure 1. Mitochondrial and nuclear phylogenetic trees for the *Microtus* genus, using *Arvicola sapidus* (ASAP) as outgroup. For each genome type we performed Maximum likelihood (ML) and Bayesian coalescence (BA) analyses and given the high similarity of topologies, and bootstrap (BS) and posterior probabilities (PP), and we chose to represent the maximum clade credibility of the ML runs; branches in bold represent subgenera, identified by different colours and with the respective subgenus below. Node support is represented by black (BS/PP > 0.95) circles, except where noted otherwise (BS above and PP below the node). Grey numbers in front of nodes represent divergence values for the two branches (see text for details). Species are represented by codes: *M. agrestis* (MAGR), *Microtus cabrerae* (MCAB), *M. arvalis* (MARV), *M. levis* (MLEV), *M. subterraneus* (MSUB), *M. oeconomus* (MOEC), *M. ochrogaster* (MOCH) and *Arvicola sapidus* (ASAP). For MAGR, we further identified the three lineages identified in Paupério *et al.* (2012): Northern (N), Southern (S) and Portuguese (P); Copyright: MAGR, MCAB, MARV and ASAP images were obtained from Palomo & Gisbert (2002), © Jordi Mateos, with permission; MLEV image was obtained from "Latvijas daba" online encyclopaedia, © Askolds Kļaviņš, Ltd. "Gandrs" (https://www.latvijasdaba.lv), with permission; MSUB and MOEC illustrations were adapted from images obtained from www.ecosystema.ru, with permission; MOCH image is a painting by Todd Zalewski obtained from Kays and Wilson's *Mammals of North America*, © Princeton University Press (2002), with permission.

Similarly to the mitochondrial data, the ML and BA phylogenetic analyses for the nuclear dataset also show very similar results with equal topology and equivalent support for the different nodes (Figure 1): *M. ochrogaster* and *M. oeconomus* are basal to the *Microtus* phylogeny, with very low support to determine confidently which of these species is the most basal; the *M. arvalis*/*M. levis*/*M. subterraneus* complex is well supported; *M. agrestis* and *M. cabrerae* are clustered together, although with low support, and show a higher divergence than that observed between *M. subterraneus* and *M. arvalis*/*M. levis*; there is deep divergence between lineages within *M. agrestis*, with the Portuguese lineage 'P' as basal.

5. Discussion

Microtus *evolutionary history*

The radiation of *Microtus* has long been studied with mitochondrial and nuclear single markers, including AFLPs, however the diversification order and some clustering patterns are inconclusive across studies, as observed in other rapid radiations (Fink et al., 2010; Galewski et al., 2006; Giarla and Esselstyn, 2015; Jaarola et al., 2004; Martínková and Moravec, 2012; Robovský et al., 2008). In this study we used both mitochondrial and nuclear data to study the rapid radiation within the genus *Microtus* in a genomic context, using partial mitogenomes and genotyping-by-sequencing (GBS) on seven species representing five *Microtus* subgenera and the main biogeographic ranges where this group occurs. One of the questions we address is the branching order of the different subgenera. Both types of genome (nuclear and mitochondrial) indicate that *M. ochrogaster* and *M. oeconomus* are basal to the remaining *Microtus* species analysed, which represent all subgenera currently found in Europe (Figure 1). This pattern has already been observed in previous studies, but with little statistical support (Jaarola et al., 2004; Martínková and Moravec, 2012). However, the genomic dataset used in our study was not able to discriminate accurately which of these species (or the subgenera they represent) is the most basal, maybe reflecting a true polytomy or something close to it, as expected with rapid radiations (Giarla and Esselstyn, 2015). Although GBS has shown to resolve the coffee interspecific phylogeny which is thought to have radiated in a similar time frame (*c.* 2 - 4 Mya) (Hamon et al., 2017; McCormack et al., 2013), it was not able to provide high support values for the deeper nodes of the *Microtus* phylogeny. This may imply the need of applying other high throughput sequencing methods that target specific genomic regions and may provide more power for understanding the deeper branches of phylogenies (i.e. Zeng et al., 2014). The European subgenera are

separated in two well-supported groups, *Terricola* + *Microtus* and *Agricola* + *Iberomys*, suggesting two independent events of diversification in the western Palearctic, possibly representing two colonisation waves to Europe. The close relation of *Terricola* and *Microtus* subgenera, supported both at the mitochondrial and nuclear level has been consistently observed across phylogenetic studies [e.g. Jaarola *et al.* (2004), Martínková & Moravec (2012)]. The difference in our study is that we observed a consistent clustering of *M. agrestis* and *M. cabrerae* for both types of genome, although with low support at the nuclear level. One particular question we were interested in resolving with this study was the phylogenetic positioning of the Cabrera vole. Previous studies have shown a dubious phylogenetic positioning of *M. cabrerae* among the Nearctic species, to which *M. ochrogaster* belongs (e.g. Jaarola *et al.* 2004; Fink *et al.* 2010; Martínková & Moravec 2012). However, in our study we were able to place *M.* (*Iberomys*) *cabrerae* within the European clade, as a sister group to *M.* (*Agricola*) *agrestis*, though with low support at the nuclear level (Figure 1). Our results suggest that previous studies may reflect the effect of long branch attraction (LBA), possibly due to the long time since the divergence of the *Iberomys* and *Agricola* subgenera from other *Microtus* subgenera and among themselves (Bergsten 2005). The bias in previous studies could have been the result of heterogeneity in molecular evolutionary rates due to high selective pressures (leading to very distinct evolutionary trajectories thereon), high levels of homoplasy and thus, LBA with very distinct lineages (Philippe *et al.* 2011; Ellegren 2014; Ho 2014). This lineage specific high molecular evolutionary rate could be associated with the distinct Cabrera vole morphology, life history, and karyotypic characteristics, possibly reflecting strong selective pressures due to its ecological specialisation and multiple bottlenecks since the Pleistocene (Ellegren 2014; Pita *et al.* 2014). Based on this distinctiveness, Cuenca-Bescós *et al.* (2014) have recently proposed to elevate the subgenus *Iberomys* to the rank of genus with the Cabrera vole as its sole representative, *Iberomys cabrerae*. In this work we found support for the subgenus *Iberomys* given that both mitochondrial and nuclear data show deep branching in its divergence from *M. agrestis*, even older than the *Terricola* – *Microtus* subgenera split. However, the elevation of *Iberomys* to the rank of genus is not supported by our data as it is internal to the phylogeny of the genus *Microtus*. Thus, considering *Iberomys* as a new genus would make the genus *Microtus* paraphyletic (Figure 1), and therefore inappropriate.

Microtus *biogeographic history*

The subgenera analysed in this study are particularly important for the biogeographic history of *Microtus*. Previous studies using mitochondrial and nuclear markers (traditional nuclear loci and AFLPs) support the hypotheses of two to three independent expansions from central Asia to Europe, and two to North America, with associated bursts of speciation (Galewski *et al.* 2006; Fink *et al.* 2010; Martínková & Moravec 2012). Our study detected only one radiation in North America, represented by *M. ochrogaster*, and a later expansion by *M. oeconomus*, but the number of species analysed is limited. However, our sampling among the European species includes all five recognised subgenera (*Agricola*, *Iberomys*, *Terricola*, *Microtus* and *Alexandromys*), and our results appear to support the hypothesis that Europe was colonised in two waves of colonisation and diversification from the east: one resulting in the radiation of the ancestor of *M. agrestis*/*M. cabrerae*, with early divergence of these two species; and the other resulting in the expansion of the ancestor of the *Microtus* and *Terricola* subgenera, and their radiations; with a third later expansion by *M. oeconomus* (Figure 1). *M. oeconomus* is assumed to have expanded both east and west from its central Asian origin, resulting in three main intraspecific lineages with a branching consistent with a recent rapid expansion (Brunhoff *et al.* 2003; Haring *et al.* 2011). This indicates that this species possibly colonised Europe and North America at the same time, which has been estimated to have occurred around the time of the Zyryanka glaciation *c.* 70 - 55 kya, thereby being the most recent colonisation out of the three expansions to Europe and the last to North America (Lance & Cook 1998; Galbreath & Cook 2004). Tougard (2017) focussing on the *Terricola* subgenus suggested that there were various time periods of diversification of this group starting in its origin around 4 Mya in the Near East, with its divergence from the *Microtus* subgenus around 5 Mya. This contrasts with the finding that most extant *Terricola* species coalesce to *c.* 0.5 Mya and that the putative common ancestor of all *Microtus* species, *Allophaiomys*, is only found in the European fossil record from the Early Pleistocene onwards (2.4-2.0 Mya) (Chaline *et al.* 1999). The inconsistences may reflect the uncertainty associated to fossil calibrations (dos Reis 2016). In any case, there was an initial divergence of the *Terricola* subgenus in the vicinity of the Near East (here represented by *M. subterraneus*), while *M. arvalis* and *M. levis* diverged from each other at a later time (Tougard 2017). This author further proposes that a combination of dispersal and vicariance, associated to a muroid like speciation rate, have resulted in the high number of species of this group, following a geographic radiation (Simões *et al.* 2016).

Little information is available regarding the colonisation of Europe by the ancestors of *M. agrestis* and *M. cabrerae*, although the first representative of the *Iberomys*

subgenus appeared in the Iberian Peninsula around 0.78 Mya, thus predating the expected time of arrival of *M. oeconomus* to Europe, around 0.25 – 0.15 Mya (Brunhoff *et al.* 2003; Galbreath & Cook 2004; Laplana & Sevilla 2013; Cuenca-Bescós *et al.* 2014; Pita *et al.* 2014).

Continued radiations in relation to geography and climate

The use of SNP variation from restriction enzyme-based techniques such as RAD-seq and GBS has previously proven to be useful for species delimitation (Pante *et al.* 2015). In our study, all species were recovered with maximum support, but we found additional support for the subdivision of various vole lineages, specifically for the field vole (*M. agrestis*), the common vole (*M. arvalis*) and the root vole (*M. oeconomus*) (Figure 1). Although very recent, some of these lineages might already represent different species or are in the process of becoming so. Mitchell-Jones *et al.* (1999) proposed the subdivision of *M. arvalis* into two species, based on their karyotypes: an eastern '*obscurus*' and a western '*arvalis*' form, which is supported by our results, given the observed divergence levels (*Dxy* = 2.7%) (Figure 1). Previous mitochondrial studies found little genetic divergence between the two forms, although more recent studies support the existence of parapatric distributions and existence of a contact zone (Haynes *et al.* 2003b; Bulatova *et al.* 2010). The *M. oeconomus* samples used in this study belong to the 'Northern European' and 'Central Asian' clades and, as in previous studies, have a divergence of about 3.6%, which is similar to that observed between the *M. agrestis* 'Southern' and 'Portuguese' lineages (Brunhoff *et al.* 2003; Paupério *et al.* 2012). Building on the initial proposal of Jaarola & Searle (2002, 2004) and later Paupério *et al.* (2012), our findings support the subdivision of *M. agrestis* into different species: *M. rozianus* (Portuguese form), *M. levernedii* (Southern form) and *M. agrestis* (Northern form). These three taxa are thought to have diverged within the time span of the last glacial period (up to 100 kya), which in our phylogenetic estimates is equivalent to the observed divergence between *M. arvalis* from *M. levis* for both marker types [Figure 1, Paupério *et al.* (2012)]. However, there is a cyto-nuclear discordance that it is confirmed with our genomic data. While the 'Northern' form is the first to diverge on the basis of the mitochondrial data, the nuclear data suggest that the 'Portuguese' form is the more basal. As seen in previous studies, this might reflect differential lineage sorting of the mitochondrial and nuclear DNA, associated to the different evolutionary rates (Paupério *et al.* 2012). The sample from the *M. agrestis* Southern lineage (SM0372, Table 1), showed two mitochondrial sequences spanning a large proportion of the targeted region. The discarded group of mitochondrial sequences likely represent nuclear copies (numts)

given the presence of indels and stop codons in the coding regions of the mitochondria, which has been observed frequently in other arvicolines including many *Microtus*, even though we only found this pattern in this particular sample (Triant & DeWoody 2008). The presence of numts has already been observed in samples from a specific population, in Eugui, Navarre, Spain (J Paupério, *pers. comm.*), and might be indicative of further substructure within the field vole lineages.

6. Conclusions

Our results show that the use of mitogenomes and SNPs from genotyping-by-sequencing improve the accuracy and confidence of the phylogeny of rapid radiating groups, like *Microtus*. We found that *M. oeconomus* and *M. ochrogaster* are the most basal species, representing the Holarctic and Nearctic groups, respectively. Our data suggest at least two movements into North America, one radiation by the group of Nearctic species and the other an expansion by *M. oeconomus*. Europe appears to have been colonised by three groups, one consisting of *M. (Agricola) agrestis* and *M. (Iberomys) cabrerae*, another consisting of *M. (Terricola) subterraneus*, *M. (Microtus) arvalis* and *M. (Microtus) levis*, and a third one represented by *M. (Alexandromys) oeconomus*. Speciation has been pervasive throughout the history of the genus *Microtus*, and ongoing diversification can be found within many recognised species, possibly masking the true number of species present in this group. In this study we confirmed the cryptic divergence previously observed within *M. agrestis*, *M. arvalis* and *M. oeconomus*. At a different scale, we show here that phylogenomic studies are also very important to clarify specific questions that are associated with deeper divergences, such as the positioning of *M. cabrerae*. Finally, with this study we were able to confirm the phylogenetic position of *M. cabrerae* as a member of the *Microtus* genus and sister species to *M. agrestis*, although belonging to its own subgenus, *Iberomys*.

7. References

Aberer AJ, Kobert K, Stamatakis A (2014) ExaBayes: massively parallel bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, **31**, 2553–2556.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews*

Genetics, **17**, 81–92.

Bannikova AA, Lebedev V, Lissovsky A *et al.* (2010) Molecular phylogeny and evolution of the Asian lineage of vole genus *Microtus* (Rodentia: Arvicolinae) inferred from mitochondrial cytochrome *b* sequence. *Biological Journal of the Linnean Society*, **99**, 595–613.

Bergsten J (2005) A review of long-branch attraction. *Cladistics*, **21**, 163–193.

Bonetti MF, Wiens JJ (2014) Evolution of climatic niche specialization: a phylogenetic analysis in amphibians. *Proceedings of the Royal Society B*, **281**, 20133229.

Bradbury PJ, Zhang Z, Kroon DE *et al.* (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Brawand D, Wagner CE, Li YI *et al.* (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375–381.

Brunhoff C, Galbreath KE, Fedorov VB, Cook JA, Jaarola M (2003) Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Molecular Ecology*, **12**, 957–968.

Bulatova NS, Golenoshchev FN, Koval'skaya YM *et al.* (2010) Cytogenetic study of the parapatric contact zone between two 46-chromosomal forms of the common vole in European Russia. *Russian Journal of Genetics*, **46**, 443-448.

Buzan E V, Krystufek B, Hänfling B, Hutchinson WF (2008) Mitochondrial phylogeny of Arvicolinae using comprehensive taxonomic sampling yields new insights. *Biological Journal of the Linnean Society*, **94**, 825–835.

Cao W, Xia Y, Dang X, Xu Q (2016) The first complete mitochondrial genome of the *Microtus ochrogaster*. *Mitochondrial DNA Part A*, **27**, 4292–4293.

Chaline J (1974) Les rongeurs, l'âge et l'environnement de la trés haute terrasse de Grâce à Montières. *Bulletin de l'Association Française pour l'Étude du Quaternaire*, **11**, 151–157.

Chaline J, Brunet-lecomte P, Montuire S, Viriot L, Courant F (1999) Anatomy of the arvicoline radiation (Rodentia): palaeogeographical, palaeoecological history and evolutionary data. *Annales Zoologici Fennici*, **36**, 239–267.

Collins RA, Hrbek T (2015) An in silico comparison of reduced-representation and sequence-capture protocols for phylogenomics. *bioRxiv*, 32565.

Conroy CJ, Cook JA (1999) MtDNA evidence for repeated pulses of speciation within

Arvicoline and Murid rodents. *Journal of Mammalian Evolution*, **6**, 221–245.

Conroy CJ, Cook JA (2000) Molecular systematics of a Holarctic rodent (*Microtus*: Muridae). *Journal of Mammalogy*, **81**, 344–359.

Cuenca-Bescós G, López-García JM, Galindo-Pellicena MA *et al.* (2014) The Pleistocene history of *Iberomys*, an endangered endemic rodent from South Western Europe. *Integrative Zoology*, **9**, 481–497.

Dabney J, Meyer M (2012) Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques*, **52**, 87–94.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution*, **24**, 332–340.

Duchêne S, Archer FI, Vilstrup J, Caballero S, Morin PA (2011) Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS One*, **6**, e27138.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution*, **29**, 51–63.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, e19379.

Felsenstein J (2004) *Inferring Phylogenies*. Sinauer Associates, Sunderland, MA, USA.

Fernandes M, Pita R, Mira A (2008) *Microtus cabrerae*. In: IUCN 2011. IUCN Red List of Threatened Species, version 2011.2. www.iucnredlist.org.

Filipi K, Marková S, Searle JB, Kotlík P (2015) Mitogenomic phylogenetics of the bank vole *Clethrionomys glareolus*, a model system for studying end-glacial colonization of Europe. *Molecular Phylogenetics and Evolution*, **82**, 245–257.

Fink S, Fischer MC, Excoffier L, Heckel G (2010) Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia: *Microtus*): the utility of AFLPs versus mitochondrial and nuclear sequence markers. *Systematic Biology*, **59**, 548–572.

Fu Q, Meyer M, Gao X *et al.* (2013) DNA analysis of an early modern human from

Tianyuan Cave, China. *Proceedings of the National Academy of Sciences*, **110**, 2223–2227.

Galbreath KE, Cook JA (2004) Genetic consequences of Pleistocene glaciations for the tundra vole (*Microtus oeconomus*) in Beringia. *Molecular Ecology*, **13**, 135–148.

Galewski T, Tilak M, Sanchez S *et al.* (2006) The evolutionary radiation of Arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. *BMC Evolutionary Biology*, **6**, 80.

Giarla TC, Esselstyn JA (2015) The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology*, **64**, 727–740.

Haring E, Sheremetyeva IN, Kryukov AP (2011) Phylogeny of Palearctic vole species (genus *Microtus*, Rodentia) based on mitochondrial sequences. *Mammalian Biology*, **76**, 258–267.

Haynes S, Jaarola M, Searle JB (2003) Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. *Molecular Ecology*, **12**, 951–956.

Herman JS, Searle JB (2011) Post-glacial partitioning of mitochondrial genetic variation in the field vole. *Proceedings of the Royal Society B*, **278**, 3601–3607.

Ho SYW (2014) The changing face of the molecular evolutionary clock. *Trends in Ecology and Evolution*, **29**, 496–503.

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

Isaac NJB, Turvey ST, Collen B, Waterman C, Baillie JEM (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One*, **2**, e296.

IUCN (2016) The IUCN Red List of Threatened Species. <*http://www.iucnredlist.org*>, Downloaded on 07 Dec 2016.

Jaarola M, Martínková N, Gündüz I *et al.* (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **33**, 647–663.

Jaarola M, Searle JB (2002) Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology*, **11**, 2613–2621.

Jaarola M, Searle JB (2004) A highly divergent mitochondrial DNA lineage of *Microtus agrestis* in southern Europe. *Heredity*, **92**, 228–234.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H (2006) Phylogenomics: the beginning of incongruence? *Trends in Genetics*, **22**, 225–231.

Kearse M, Moir R, Wilson A *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Kircher M (2012) Analysis of high-throughput Ancient DNA sequencing data. In: *Ancient DNA: Methods and Protocols* (eds Shapiro B, Hofreiter M), pp. 3–16. Springer Science & Business Media.

Lance EW, Cook JA (1998) Biogeography of tundra voles (*Microtus oeconomus*) of Beringia and the southern coast of Alaska. *Journal of Mammalogy*, **79**, 53–65.

Laplana C, Sevilla P (2013) Documenting the biogeographic history of *Microtus cabrerae* through its fossil record. *Mammal Review*, **43**, 309–322.

Leaché AD, Banbury BL, Felsenstein J, Nieto-Montes de Oca A, Stamatakis A (2015) Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, **64**, 1032-1047.

Lewis PO (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, **50**, 913–925.

López-García JM, Cuenca-Bescós G (2012) Changes in the geographical distribution of *Microtus (Iberomys) cabrerae* (Thomas, 1906) from the Late Pleistocene to the Holocene. *Spanish Journal of Palaentology*, **27**, 117–124.

Martínková N, Moravec J (2012) Multilocus phylogeny of arvicoline voles (Arvicolini, Rodentia) shows small tree terrace size. *Folia Zoologica*, **61**, 254–267.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **5**, pdb-prot5448.

Mitchell-Jones AJ, Bogdanowicz W, Kryštufek B *et al.* (1999) *The Atlas of European Mammals*. T&AD Poyser, London, UK.

Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, **49**, 718–726.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, **19**, 208–216.

Murphy WJ, Eizirik E, O'Brien SJ *et al.* (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, **294**, 2348–2351.

Nabholz B, Glémin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals - the longevity hypothesis. *Molecular Biology and Evolution*, **25**, 120–130.

Palomo LJ, Gisbert J (2002) *Atlas de los Mamiferos Terrestres de España*. Dirección General de Conservación de la Naturaleza, SECEM-SECEMU, Madrid.

Pante E, Abdelkrim J, Viricel A *et al.* (2015) Use of RAD sequencing for delimiting species. *Heredity*, **114**, 450–459.

Paupério J, Herman JS, Melo-Ferreira J *et al.* (2012) Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia. *Molecular Ecology*, **21**, 6015–6032.

Philippe H, Brinkmann H, Lavrov D V *et al.* (2011) Perspective resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*, **9**, e1000602.

Pita R, Mira APP, Beja P (2014) *Microtus cabrerae* (Rodentia: Cricetidae). *Mammalian Species*, **912**, 48–70.

Rambaut A, Drummond AJ (2015) Tracer v1.5. Edinburgh, UK

dos Reis M (2016) Notes on the birth-death prior with fossil calibrations for Bayesian estimation of species divergence times. *Philosophical Transactions of the Royal Society B*, **371**, 20150128.

Repenning CA (1993) *Allophaiomys* and the age of the Olyor Suite, Krestovka Sections, Yakutia. *US Geological Survey Bulletin*, **2037**, 1-98.

Robovský J, Řičánková V, Zrzavý J (2008) Phylogeny of Arvicolinae (Mammalia, Cricetidae): utility of morphological and molecular data sets in a recently radiating clade. *Zoologica Scripta*, **37**, 571–590.

Rubin BER, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. *PLoS One*, **7**, e33394.

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society of London B*, **273**, 1987–1998.

Simões M, Breitkreuz L, Alvarado M *et al.* (2016) The evolving theory of evolutionary radiations. *Trends in Ecology and Evolution*, **31**, 27–34.

Simon C, Buckley TRT, Frati F, Stewart JJB, Beckenbach ATA (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics*, **37**, 545–579.

Song S, Liu L, Edwards S V, Wu S (2012) Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences*, **109**, 14942–14947.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Steppan SJ, Storz BL, Hoffmann RS (2004) Nuclear DNA phylogeny of the squirrels (Mammalia: Rodentia) and the evolution of arboreality from c-myc and RAG1. *Molecular Phylogenetics and Evolution*, **30**, 703–719.

Stroud JT, Losos JB (2016) Ecological opportunity and adaptive radiation. *Annual Review of Ecology, Evolution, and Systematics*, **47**, 507–532.

Su Z, Townsend JP (2015) Utility of characters evolving at diverse rates of evolution to resolve quartet trees with unequal branch lengths: analytical predictions of long-branch effects. *BMC Evolutionary Biology*, **15**, 86.

Suh A, Smeds L, Ellegren H (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biology*, **13**, e1002224.

Taylor RW, Turnbull DM (2005) Mitochondrial DNA mutations in human disease. *Nature Reviews Genetics*, **6**, 389–402.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Tougard C (2017) Did the Quaternary climatic fluctuations really influence the tempo and mode of diversification in European rodents? *Journal of Zoological Systematics and Evolutionary Research*, **55**, 46–56.

Triant DA, DeWoody JA (2006) Accelerated molecular evolution in *Microtus* (Rodentia) as assessed via complete mitochondrial genome sequences. *Genetica*, **128**, 95–

108.

Triant DA, DeWoody JA (2008) Molecular analyses of mitochondrial pseudogenes within the nuclear genome of arvicoline rodents. *Genetica*, **132**, 21–33.

Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Weksler M, Lanier HC, Olson LE (2010) Eastern Beringian biogeography: historical and spatial genetic structure of singing voles in Alaska. *Journal of Biogeography*, **37**, 1414–1431.

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

Whitfield JB, Lockhart PJ (2007) Deciphering ancient rapid radiations. *Trends in Ecology and Evolution*, **22**, 258–265.

Wiens JJ, Kozak KH, Silva N (2013) Diversity and niche evolution along aridity gradients in north american lizards (Phrynosomatidae). *Evolution*, **67**, 1715–1728.

Wilson DE, Reeder DM (Eds.) (2005) *Mammal Species of the World: A Taxonomic and Geographic Reference*. Johns Hopkins University Press, Baltimore, MD, USA.

Yang Z, Rannala B (2012) Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, **13**, 303–314.

Zagorodnyuk I V. (1990) Karyotypic variability and systematics of the gray voles (Rodentia, Arvicolini). Communication 1. Species composition and chromosomal numbers. *Vestnik Zoologii*, **2**, 26–37.

8. Supporting information

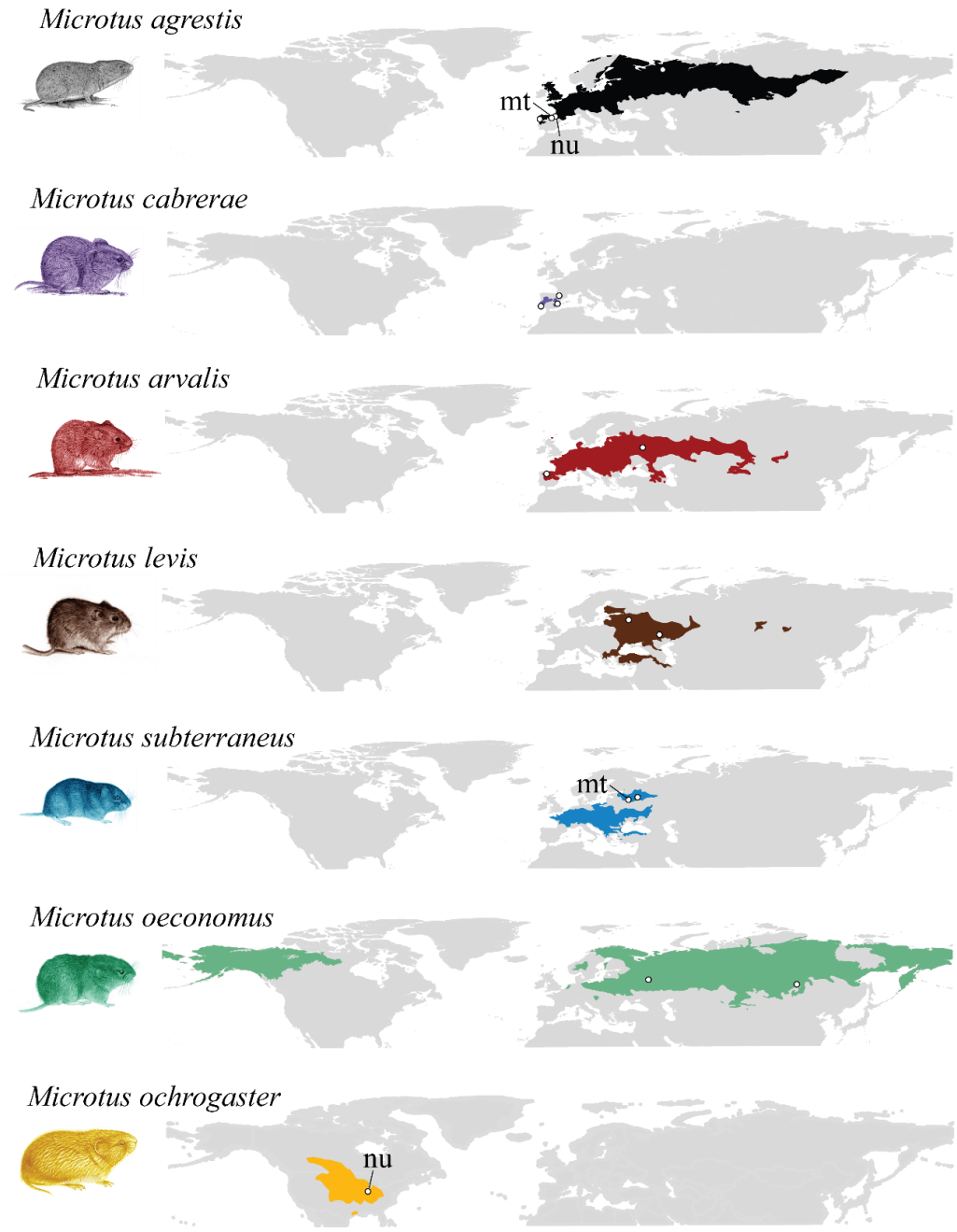


Figure S1. Distribution maps of the *Microtus* species analysed in this study (colours) and location of the samples (white circles), excluding those from GenBank. Some samples were used for the mitochondrial (mt) or nuclear (nu) analyses only and are thus labelled accordingly. Copyright: *Microtus agrestis*, *M. cabrerae* and *M. arvalis* images were obtained from Palomo & Gisbert (2002), © Jordi Mateos, with permission; *M. levis* image was obtained from "Latvijas daba" online encyclopaedia, © Askolds Kļaviņš, Ltd. "Gandrs" (https://www.latvijasdaba.lv), with permission; *M. subterraneus* and *M. oeconomus* illustrations were adapted with permission from images obtained from www.ecosystema.ru; *M. ochrogaster* image is a painting from Kays and Wilson's *Mammals of North America*, Princeton University Press (2002), by © Todd Zalewski and used with permission. Species distributions were downloaded from the IUCN Red List of Threatened Species website (IUCN 2016).

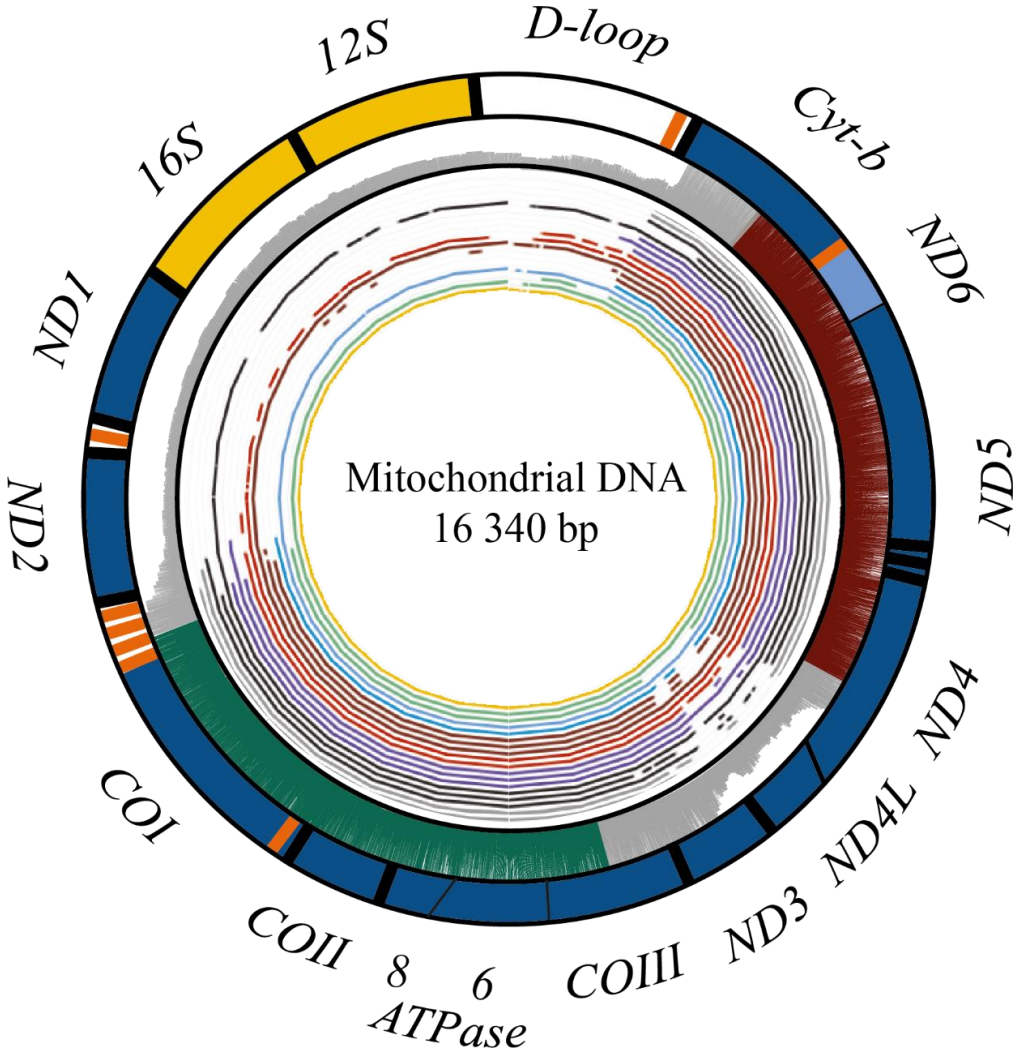


Figure S2. Mitochondrial DNA structure (outer layer) coloured by protein coding genes (blue and light blue for transcription in the heavy and light strains, respectively), ribosomal RNA genes (yellow), and tRNA genes (black and orange thick lines for transcriptions in the heavy and light strains, respectively), all labelled by standardised abbreviations. Thin lines represent gene boundaries and oblique strokes represent areas of transcription overlap between neighbouring genes [based on Taylor & Turnbull (2005), with permission]. Sequencing effort is represented on the intermediate layer, where bars around the mitochondrion represent the average coverage for a given portion of the mitochondrial genome obtained with the samples analysed in this study (ranging from 0 to 100%). Grey represents non-target areas, while green and red represent target areas with the following primer pairs: Green forward – 5'-TCTACACCTTCGAATTTGCAATT-3', Green reverse – 5'-TGTCARTAYCATGCTGCTGCTTC-3' resulting in a fragment of approximately 4030 base pairs (bp); Red forward – 5'-TCTGCCTBCGACAAACAGAC-3', Red reverse 5'-GAAGGCRAARAATCGTGTGAG-3' resulting in a fragment of approximately 3685 bp. For the long range PCR reactions we used TaKaRa LA Taq® Hot Start (Takara, Kyoto, Japan) with the following thermocycling profile: 94ºC for 1 min, followed by 35 cycles of 98ºC for 10 s and 68ºC for 7 min, and ended with an extension at 72ºC for 10 min. Inside circumferences represent the portion of the mitochondrial DNA covered for each individual analysed in this study coloured by species: *Arvicola sapidus* (grey), *Microtus agrestis* (black), *M. cabrerae* (purple), *M. arvalis* (red), *M. levis* (brown), *M. subterraneus* (blue), *M. oeconomus* (green) and *M. ochrogaster* (yellow).

Endemic species may have complex histories: within-refugium phylogeography of an endangered Iberian vole

Barbosa S, Paupério J, Herman JS, Ferreira CM, Pita R, Vale-Gonçalves HM, Cabral JÁ, Garrido-García JÁ, Soriguer RC, Beja P, Mira A, Alves PC and Searle JB

1. Abstract

Glacial refugia protected and promoted biodiversity during the Pleistocene, not only at a broader scale, but also for many endemics that contracted and expanded their ranges within refugial areas. Understanding the evolutionary history of refugial endemics is especially important in the case of endangered species to recognise the origins of their genetic structure and thus produce better informed conservation practices. The Iberian Peninsula is an important European glacial refugium, rich in endemics of conservation concern, including small mammals, such as the Cabrera vole (*Microtus cabrerae*). This near-threatened rodent is characterized by an unusual suite of genetic, life history and ecological traits, being restricted to isolated geographic nuclei in fast-disappearing Mediterranean subhumid herbaceous habitats. To reconstruct the evolutionary history of the Cabrera vole, we studied sequence variation at mitochondrial, autosomal and sex-linked loci, using invasive and non-invasive samples. Despite low overall mitochondrial and nuclear nucleotide diversities, we observed two main well-supported mitochondrial lineages, *west* and *east*. Phylogeographic modelling in the context of the Cabrera vole's detailed fossil record supports a demographic scenario of isolation of two populations during the Last Glacial Maximum from a single focus in the southern part of the Iberian Peninsula. In addition, our data suggest subsequent divergence within the east, and secondary contact and introgression of the expanding western population, during the late Holocene. This work emphasizes that refugial endemics may have a phylogeographic history as rich as that of more widespread species, and conservation of such endemics includes the preservation of that genetic legacy.

2. Introduction

Much of phylogeographic reconstruction has involved tracking the colonisation of widespread temperate species from 'glacial refugia' (Hewitt 2000). However, geographic and climatic heterogeneity within glacial refugia may often have subdivided populations into several discrete areas during the last glaciation, referred to as '*refugia within refugia*' (Gómez & Lunt 2007; Abellán & Svenning 2014). The occurrence of subdivision within refugia is not only important for interpreting the genetic structure of widespread species that have expanded their ranges far beyond the refugia; it is also critical for species that are currently restricted to these refugial areas, that is refugial endemics (Bilton *et al.* 1998; Kryštufek *et al.* 2007). Even considering the whole refugium, such as one of the classical Mediterranean refugia of Iberia, Italy and the Balkans (Hewitt 1999), for some refugial endemics there may be difficulty in maintaining a viable population under current conditions of habitat degradation/destruction and climate change. Therefore, refugial endemics must, by definition, be of conservation concern, and an appreciation of their genetic structure and origin is critical for effective conservation (Avise 2000; D'Amen *et al.* 2013; Malaney & Cook 2013; Moritz & Potter 2013). While such refugial endemics have been identified, and genetic structure determined in some cases, there has been little attempt to use the recent advances in demographic inference and phylogeographic modelling to infer their detailed colonisation histories. We do this for a small mammal endemic to the Iberian refugium: the Cabrera vole (*Microtus cabrerae*). Understanding the phylogeography of the Cabrera vole is important for the conservation of this species, and it can provide clues about the colonisation history of many other Iberian endemics. Finally, in a context of climate change and habitat loss, such analyses can help to identify the main obstacles to range change that refugial endemic species might face today and in the future.

The Cabrera vole is highly distinctive compared with other extant members of *Microtus*, including in the important taxonomic character of dental morphology, causing some to argue that it should be placed in a separate genus, *Iberomys* (Cuenca-Bescós *et al.* 2014). It has an unresolved phylogenetic position within *Microtus* based on mitochondrial and nuclear molecular markers (Jaarola *et al*. 2004; Fink *et al.* 2010). It exhibits giant sex chromosomes that are unique in mammals in having multiple copies of the *SRY* gene in both sexes (Jiménez *et al.* 1991; Bullejos *et al.* 1997). In terms of life history, the species is unusual for *Microtus* in having a *K* reproductive strategy associated with monogamy, small litters and large body size (Ventura *et al.* 1998; Fernández-Salvador *et al.* 2001). Ecologically, the Cabrera vole is narrowly specialized

to the *thermo-* and *meso-mediterranean* bioclimates, which are characterized by subhumid conditions (Rivas-Martínez 1981). Within these climatic ranges, the species is largely confined to very particular habitats, consisting primarily of meadows and perennial herb communities, often near small water courses and temporary ponds (Pita *et al.* 2006; Luque-Larena & López 2007). The *Vulnerable* conservation status of the Cabrera vole in Portugal and Spain, as well as the global *Near Threatened* status attributed by the IUCN Red List of Threatened Species (Palomo & Gisbert 2002; Cabral *et al.* 2005; Fernandes *et al.* 2008), is certainly a consequence of these narrow ecological requirements. Major threats to the species are related to habitat loss through agricultural intensification, including wetland drainage, overgrazing and poor management of possible dispersal corridors (Pita *et al.* 2014), and it may also be severely affected by climate change (Mestre *et al.* 2015).

A recent study by Garrido-García *et al.* (2013) describes the Cabrera vole in four main geographic nuclei (Figure 1): *Lusocarpetan* (*Lc*), extending from the southwest to the northeast of Portugal and into Spain throughout the Central mountain system; *Montiberic* (*Mb*), along the Iberian mountain system extending to the province of Valencia and north of the Murcia region; *Betic* (*Bt*), comprising the eastern Betic mountain chain and the provinces of Albacete, Jaén, Granada and Murcia; and *pre-Pyrenean* (*pP*), where it is restricted to very few locations in the pre-Pyrenean massifs. This distribution suggests a trend of contraction of a previously widespread species into increasingly isolated populations, possibly driven by climatic factors (Mira *et al.* 2008; Mestre *et al.* 2015). Likewise, the excellent fossil record of the Cabrera vole indicates that its past distribution has fluctuated throughout the climatic oscillations of the Late Pleistocene, with colonisation–extinction–recolonisation dynamics (Garrido-García & Soriguer-Escofet 2012; Laplana & Sevilla 2013). In any case, it is postulated that climate induced aridification would have been enhanced by anthropogenic activities that gradually decreased forest area, impacting the Iberian Peninsula more intensely from the south to the north (Garrido-García *et al.* 2013; Laplana & Sevilla 2013). This led to a sharper contraction of the species distribution in the southernmost regions, where the species is now absent, and created more open areas in central Iberia, associated with extensive agro–silvo–pastoral systems, where the species is known to thrive (Pita *et al.* 2014).

So far, only two studies have analysed genetic differentiation of the Cabrera vole in a geographic context: one using allozymes (Cabrera-Millet *et al*. 1982) and the other using RAPD-PCR (Alasaad *et al*. 2013). The first compared the two previously recognized morphological subspecies (Ellerman & Morrison-Scott 1951) and failed to find any genetic divergence between those, while the latter did find a possible subdivision

into two main geographic forms (one in the southern and the other in the northern half of Iberia, not coincident with the morphological subspecies), albeit with high population connectivity. These studies were based on small numbers of populations and on markers with low variability and did not characterize the entire range of the species. Thus, having in mind the study of the evolutionary history and the conservation of refugial endemics, we used mitochondrial and nuclear genetic markers to determine genetic structure of the Cabrera vole over the current species range (restricted to Iberia) and to infer the demographic processes leading to the observed pattern of variation. Furthermore, we combined all the available genetic data to distinguish between various phylogeographic scenarios inferred from both the genetic results and the fossil record. Our study represents a detailed analysis of the colonisation history of an endemic species in one of the main glacial refugial areas in Europe and emphasizes that refugial endemics with limited distributions may have as complex phylogeographic history as more widespread species.

3. Methods

Sample collection and DNA extraction

Being a species of conservation concern, we combined non-invasive sampling with minimal invasive sampling (ear biopsy) from live-caught (and released) Cabrera voles. The non-invasive samples included bones from museum specimens (six samples, collected 1976 - 1984), and from barn owl (*Tyto alba*) pellets (two samples collected in 1905 and 74 samples collected 1980 - 2012), which substantially increased the sampling coverage in many parts of the species distribution (Figure 1). Altogether, we analysed a set of 202 tissue samples and 82 bone samples, totalling 284 voles (Figure 1). DNA from tissue samples was extracted using the EasySpin Genomic DNA Minipreps Tissue Kit (Citomed, Lisbon, Portugal) following the manufacturer's instructions. Bone samples included mandibles and femurs, and DNA was extracted following the protocol described in Barbosa *et al.* (2013), in a laboratory dedicated to the extraction of samples with low DNA quality and quantity to avoid contamination by good quality DNA. Samples within <10 km of each other were combined into a common geographic coordinate using the Integrate tool in ArcGIS v10.0 (http://www.esri.com/arcgis) - resulting in 35 localities [*Live Trapping and Non-invasive* (*LTNI*) *localities,* Table S1, Supporting information], 22 of which comprised tissue samples [*Live Trapping* (*LT*) *localities,* Table S1, Supporting information]. By clustering individuals within 10 km clusters, we limited the impact of site-by-site variation for spatial analyses and accounted for uncertainty in the location of the

vole samples represented by the bones in the barn owl pellets, given that the radius of an owl's hunting territory may be up to 5 km (Taylor 1994).

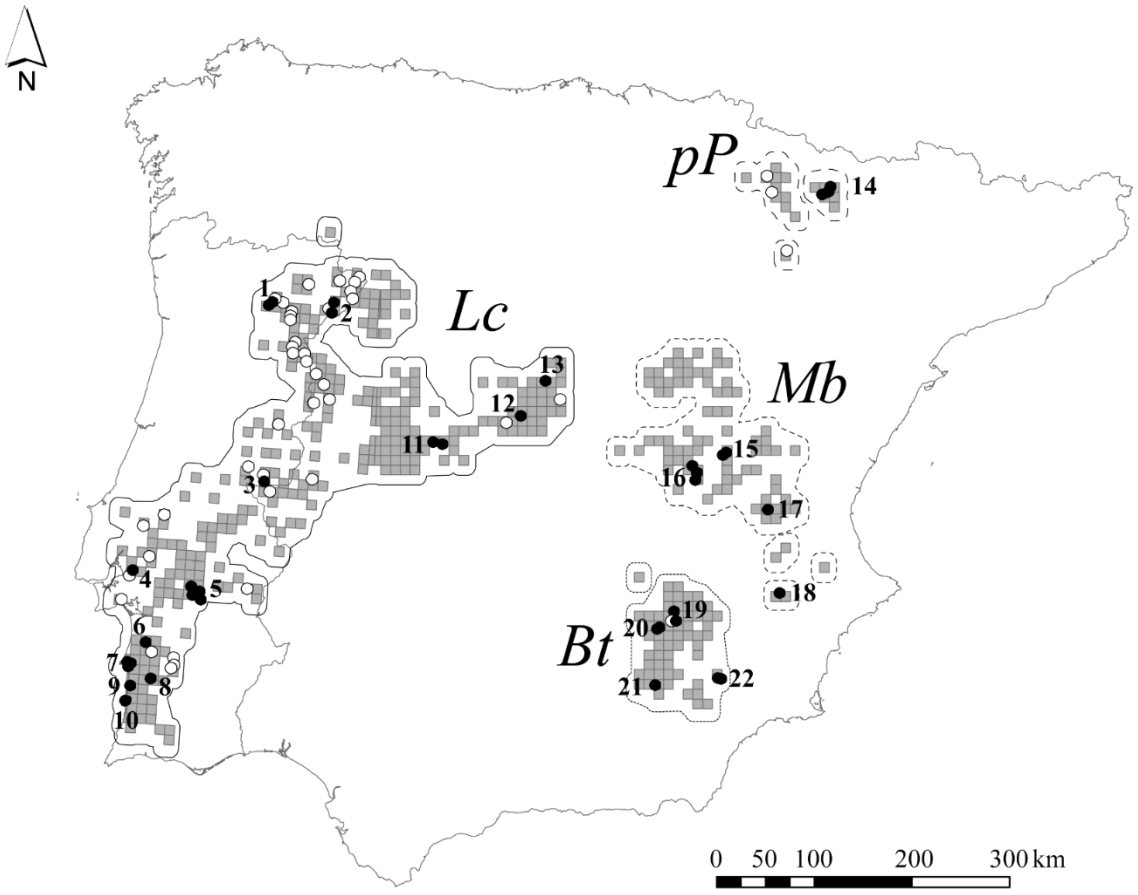


Figure 1. Sampling localities on a map of the Iberian Peninsula with the current distribution of *Microtus cabrerae* in grey 10 x 10 UMT squares (Mira *et al.* 2008; Fernandez-Salvador *et al.* 2002; Garrido-García *et al.* 2013; this study). The lines enclosing different parts of the species distribution specify the four geographic nuclei: *Lusocarpetan* (*Lc* - full line), *pre-Pyrenean* (*pP* - wide dashed line), *Montiberic* (*Mb* - narrow dashed line) and *Betic* (*Bt* - dotted line). Black and white circles indicate tissue and bone samples, respectively. Numbers specify locations where tissue samples were collected.

DNA amplification and sequencing

We analysed one mitochondrial gene for cytochrome *b* (cyt-*b*) in all 284 samples. For tissue samples we used the primers developed by Jaarola & Searle (2002) (Appendix S1, Supporting information). For bone samples, we first confirmed the species ID by amplifying a small fragment of the cyt-*b* gene (Barbosa *et al.* 2013), and the complete cyt-*b* fragment was obtained using four pairs of primers that produced overlapping fragments (Appendix S1, Supporting information). Fragments of 10 nuclear loci were amplified in a subset of 79 tissue samples from 22 populations representative of the species range [Figure 1 and *Live Trapping* (*LT*) *localities,* Table S1, Supporting information]. Due to DNA degradation, non-invasive samples could not be used for nuclear analysis, thus reducing substantially the number of available samples per

population (Barbosa *et al.* 2013). For each population, we selected available samples (between 1 and 9 voles), maximising the geographic distribution of the sampled individuals within each locality, while minimising kinship associated bias. These samples were used to amplify the following autosomal loci: exon 11 of the *BRCA1* gene, exon 1 of the *IRBP* gene and exon 1 of the *RAG1* gene; intron 4 of the *COPS7A* gene, intron 5 of the *OSTA* gene, intron 3 of the *PNPO* gene, and intron 8 of the *SLC38A7* gene. Additionally, we amplified three sex chromosome loci: introns 5 and 7 of the *DBX* gene from the X-chromosome and intron 7 from the *SMCY* gene from the Y-chromosome, also used for gender assignment. Further details of markers used, samples and laboratory procedures are provided in Appendix S1 and Table S1 (Supporting information).

Diversity analysis and summary statistics

We used DnaSP v5 (Librado & Rozas 2009) to calculate the number of sequences (N_{seq}), number of segregating sites (N_{seg}), number of haplotypes (N_{hap}), haplotype diversity (h), nucleotide diversity (π) and Watterson´s Theta (θ_W, computed from the number of segregating sites). Nuclear sequences were phased using the software PHASE as implemented in DnaSP, with 1000 burn-in steps and 1000 iterations. Indels were manually phased given their rarity and the existence of homozygous individuals to help infer the indel phase. We further used MEGA v6 (Tamura *et al*. 2013) to calculate divergence between lineages (*Dxy*) using *p-distances* (with 1000 bootstrap replicates to calculate standard deviation). Significant differences between π values of different groups were assessed variously with different marker types: for the cyt-*b* gene we compared π_{mt} between mitochondrial lineages and the four geographic nuclei; for the nuclear loci, we calculated the average nucleotide diversity ($\bar{\pi}_i$) for each of the four geographic nuclei (i) as follows:

$$\bar{\pi}_i = \frac{1}{n} \times \sum_{j=1}^{n} \left(\frac{\pi_{ij}}{max_i(\pi_{ij})} \right)$$

where j represents each of the nine nuclear loci (seven autosomal and two sex-linked loci, excluding *SMCY7*), π_{ij} represents the π_{nuc} observed in the i^{th} geographic nucleus for the j^{th} locus, and $max_i(\pi_{ij})$ represents the maximum π_{nuc} among all i for the j^{th} locus. We finally normalized between π_{mt} and all average π_{nuc} to have proportions between 0 and 1. The overall tests comparing all groups were performed using one-way ANOVA and the Tukey HSD post-hoc test (Abdi & Williams 2010).

To provide a better visualization of the geographic distribution of mitochondrial variation, we depicted the haplotypes and π_{mt} per locality on maps of the Iberian

Peninsula using data from all 284 voles sampled [analysis adapted from Igea *et al.* (2013)]. To evaluate geographic nucleus specific genetic variation, we grouped samples in *100 km* groups (minimum distance between geographic nuclei) and for each recorded the number of individuals (N_{ind}), segregating sites (N_{seg}) and haplotypes (N_{hap}), and estimated h_{mt} and π_{mt} (and respective error, calculated as the percentage of standard deviation) (Table S3, Supporting information). For localities comprising more than one mitochondrial lineage within the respective buffer zone, the calculation of π_{mt} was performed in two alternative ways: 1) the total π_{mt} considering all sequences in that respective buffer zone; and 2) the total π_{mt} per lineage in the respective buffer zone. We finally interpolated π_{mt} values over the distribution of the Cabrera vole using an Inverse Distance Weighting with the *Interpolation* tool implemented in the software QGIS v 2.4.0 (QGIS Development Team 2008).

Gene genealogy analyses

Gene genealogy networks for all loci were computed with a median-joining algorithm implemented in the software Network v4.6.0.0 (Bandelt *et al.* 1999; http://www.fluxus-engineering.com). For the cyt-*b* gene, we used the entire data set (284 samples) to estimate the genealogy using maximum likelihood (ML), as implemented in MEGA v6, and from a Bayesian posterior distribution of trees that was generated using a coalescent model in BEAST v2.3.2 (Bouckaert *et al.* 2014). For the latter analysis, we made use of the CIPRES Science Gateway online platform (Miller *et al.* 2010). The best-fit model of sequence evolution for the ML analysis was the Tamura–Nei model (Tamura & Nei 1993) with a proportion of invariable sites (TrN+I) based on the corrected Akaike and Bayesian information criteria implemented in the software jModelTest 2.0 (Darriba *et al.* 2012). For the Bayesian analysis, we used *Path* and *Stepping Stone Sampling* to estimate marginal likelihoods of each model that were then used to determine the Bayes factor for preference of one model over the other (Baele & Lemey 2013). Bayes factors for comparisons between models were interpreted according to the widely-used criteria of Kass & Raftery (1995). The selected *site substitution model*, *molecular clock model* and *tree prior* were the Hasegawa-Kishino-Yano model [HKY; Hasegawa *et al.* (1985)], uncorrelated lognormal relaxed clock (Drummond *et al.* 2006) and the coalescent Bayesian skyline (Drummond *et al.* 2005), respectively. The genealogies were calibrated using a substitution rate estimated for *Microtus arvalis* [3.2×10^{-7} substitutions/site/year; Martínková *et al.* (2013)], which was obtained using a combined analysis of modern and ancient DNA. This substitution rate is of similar magnitude to cytochrome *b* substitution

rates ($3.887x10^{-7}$ and $4.572x10^{-7}$ substitutions/site/year) estimated for *M. agrestis* using land-bridge calibration (Herman & Searle 2011; Herman *et al.* 2014) and has previously been used for other *Microtus* species over the timescale applicable here (Beysard & Heckel 2014; Hope *et al.* 2014). We performed four replicate runs of 20 million generations for each model, sampling trees and parameter estimators every 2000 generations, with the exception of the Bayesian skyline *tree prior* model, where we included four replicate runs of 50 million generations. In each case, 10% burn-in was removed using LogCombiner v2.3.0, part of the BEAST package. Tracer v1.5 (Rambaut & Drummond 2007) was used to check for convergence and sufficient sampling (effective size above 200 for each parameter) for the combined runs. The maximum clade credibility tree was obtained from the posterior sample of trees using TreeAnnotator v2.3.1 and visualized with FigTree v1.3.1 (Rambaut 2009). We used Tracer v1.5 to obtain Bayesian skyline plots (BSPs) of effective female population size over time, for the complete cyt-*b* gene and each main lineage.

For the nuclear loci, although different genes might have different evolutionary histories, here we wished to test for a general common pattern. Mean pairwise genetic divergences were calculated between all 79 individuals selected and respective haploid copies in MEGA v6, using *p-distances* for distance calculation and using 1000 bootstraps to determine standard deviation. The *SMCY7* locus was excluded from these analyses given that it did not show nucleotide variation beyond the embedded microsatellite and an adjacent indel. The pairwise distance matrices were used to calculate multilocus interindividual distances using the software POFAD v1.03 (Joly & Bruneau 2006). The distances were averaged and rescaled so that equal weights were given to each locus. The standardized multilocus distances were used to construct a distance network using the NeighborNet algorithm (Bryant & Moulton 2004) in SplitsTree v4.11.3 (Huson & Bryant 2006). Given that the loci used were selected randomly with no bias towards particular parts of the genome, we assume that there is recombination between all of them. Additionally, we tested for recombination within each locus using the software RDP v3.44 (Martin & Rybicki 2000) and we found no evidence for such recombination.

Spatial structure analyses

For comparative purposes, in these analyses, we only used the individuals with information for both mitochondrial and nuclear loci (79 individuals represented in the *LT localities*, Table S1, Supporting information). We used the software BAPS v6 (Corander *et al.* 2007) to test for spatial structure of the mitochondrial and nuclear data using the

Spatial clustering of individuals option. With the nuclear data, we used the Clustering with linked loci option to generate the preprocessed data. For both analyses, we systematically tested for the best number of populations (*K*) from *K* = 20 downwards, as suggested by the software developers. The Voronoi tessellation pictures with the optimal *K* were plotted on the map of the study area for each data set, and used as reference to assign the underlying UTM squares to the respective population. To determine the amount of genetic variation explained by population structure, we performed an analysis of molecular variance (AMOVA) and calculated the average values of population differentiation (F_{ST}) in Arlequin v3.5 (Excoffier & Lischer 2010) separating the loci in three data sets corresponding to maternally (cyt-*b*), paternally (*SMCY7*) and biparentally (remaining nine nuclear loci concatenated) inherited markers.

Demographic analyses

To investigate the possible mito-nuclear discordance (see Results – 'Spatial structure analysis'), we considered whether mitochondrial and nuclear genetic variation could be explained by sex-biased dispersal/isolation-by-distance (IBD), demographic expansions/contractions, and/or selective pressures over one or more loci. For comparative purposes, we used only the 79 individuals with information for both mitochondrial and nuclear loci (*LT localities*, Table S1, Supporting information).

To test for sex-biased dispersal and IBD, we performed a distance-based redundancy analyses (dbRDA) (Legendre & Anderson 1999) with the *capscale* function implemented in the R-package VEGAN (Oksanen *et al.* 2013), following Pavlova *et al*. (2013). As observations we used mitochondrial and mean pairwise nuclear *p-distance* genetic distance matrices. As predictors we used two geographic variables – latitude and longitude, analysed together for ease of interpretation (*LatLong*), and the first significant ($p < 0.05$) principal coordinate of the neighbourhood matrix (*PCNM*) axis, calculated as detailed in Pavlova *et al*. (2013). The geographic pairwise distance matrix used in the above analyses was calculated using the software Geographic Distance Matrix Generator v1.2.3 (Ersts 2015). DbRDAs were then used to test the influence of the predictors (geographic location and distance) in explaining mitochondrial/nuclear genetic distances between samples alone and when controlling for one another. For this, we performed two analyses: fitting each predictor independently to the observations (marginal tests); and fitting each variable to the observations after excluding the weight of the other variable (conditional tests). Statistical significance was assessed using 999

permutations, and then, each predictor was accepted as valid if explaining a significant amount of variance in both the marginal and conditional tests.

To evaluate a scenario of population expansion/decline and/or selection explaining the mito-nuclear discordance, we used DnaSP to test for departures from a neutral demographic model. For this, we calculated three neutrality test statistics, Tajima's (1989) *D*, Fu's (1997) F_S and Ramos-Onsins & Rozas's (2002) R_2, and evaluated their significance using 10 000 coalescent simulations of the genealogy. These tests were employed over each gene and respective groups (mitochondrial lineages for the cyt-*b* gene, and the geographic nuclei for all loci). For each of these groups, we also tested for selection by calculation of deviations from neutral expectations of the *dN/dS* ratio, with the Z-test of selection implemented in MEGA v6. In a first round, we tested for deviations to neutral expectations ($dN \neq dS$). If the ratio was significantly smaller or greater than one, we then tested for either positive ($dN > dS$) or purifying ($dN < dS$) selection, respectively.

Modelling alternative phylogeographic scenarios

The development of the phylogeographic scenarios was informed by the fossil record and aimed to provide realistic alternative explanations for the genetic results obtained. We modelled the probability of each specified scenario using Approximate Bayesian Computation (ABC), as implemented in DIYABC v.2.0.4 (Cornuet *et al.* 2014). We combined all information from the 11 loci (mitochondrial and nuclear), amplified for the 79 individuals of the *LT localities* (Table S1, Supporting information), and applying the parameters described in Table S4 (Supporting information). The samples were grouped into the four geographic nuclei *Lc*: 37 individuals, *Mb*: 14, *Bt*: 20, *pP*: 8. We employed a two-phase study, where in *phase I* we selected five different general scenarios, representing a comprehensive range of alternative phylogeographic hypotheses, for each of which we performed an independent DIYABC analysis keeping the basic topological structure and permuting the four geographic nuclei at the tips, to determine the best supported tree for each topology: SC_1, a model of a complete polytomy of all four populations arising from a common ancestral population; SC_2, a model of divergence of the most distinct population and then a polytomy of the remaining three; SC_3, simultaneous divergence of two lineages in populations pairs; SC_4, splitting of one population every time point from a continuous main population; and SC_5, divergence of two main lineages with recent secondary contact of two of the four diverged populations. The tree with the highest posterior probability for each general

scenario was selected for *phase II* (as reported here) where we compared the best trees for each general scenario against each other.

Further details of the model specifications and DIYABC run parameters are presented in Table S4 (details in footnotes) and Figure S1 (Supporting information).

4. Results

Multilocus diversity analysis

A total of 44 cyt-*b* haplotypes were identified, which grouped into two well-supported mitochondrial lineages, here designated *west* and *east*, with the exception of a distinct haplotype (*H38*) that was not closely associated with either (Figures 2 and 3). From the 284 Cabrera voles analysed, 228 belong to the *west* lineage (orange), comprising 34 different haplotypes; the *east* lineage (blue) included 51 voles, representing nine haplotypes; and *H38* (purple) was found in five voles (Table S2, Supporting information, and Figure 3). The median *tMRCA* (time to most recent common ancestor; and respective 95% highest posterior density) obtained from the Bayesian coalescent analysis of the complete data set was estimated to be 23.7 kya (10.5–46.7), while for the *west* and *east* lineages, it was 10.7 (5.1–18.5) and 12.2 (5.0–21.5) kya, respectively (Figure 2). The *west* lineage is the most widespread, found in three geographic nuclei (*Lc*, *Mb* and *Bt*), with most haplotypes being geographically restricted (Figure 4). The *east* lineage is also composed of geographically restricted haplotypes that have similar levels of differentiation between localities as the *west* lineage haplotypes, however having major gaps in distribution between nuclei (Figure 4). *H38* is restricted to one locality in the southernmost sampling site of the nucleus *Mb*. The haplotype that has by far the widest geographic range is *H1,* which is central within the *west* lineage median-joining network and found in most geographic nuclei, with exception of *pP* (Figures 3 and 4).

The overall mean cyt-*b* gene haplotype (h_{mt}) and nucleotide (π_{mt}) diversities, and respective standard deviations, are 0.896±0.014 and 0.500±0.031%, respectively. There is a significantly lower π_{mt} for the *west* (0.218±0.014%) than for the *east* (0.563±0.022%) lineage ($p < 0.05$; Table S2, Supporting information). The average nucleotide divergence (*Dxy*) between the two main mitochondrial lineages is 1.078±0.024%. Overall, π_{mt} is highest in the southeast of the Iberian Peninsula (Figure S2A, Supporting information).

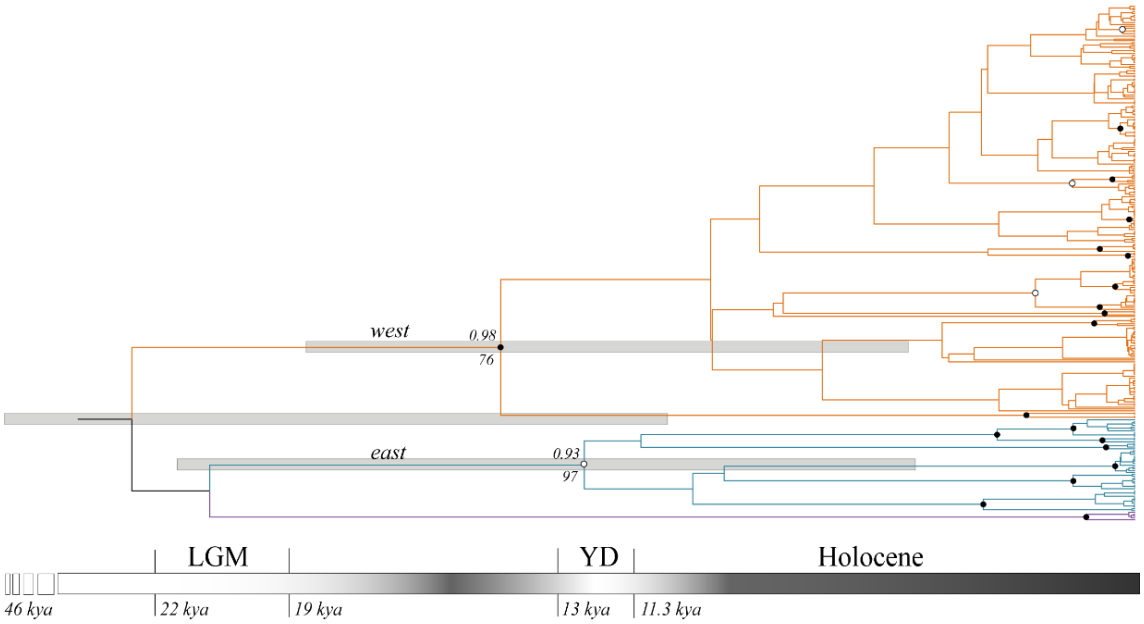


Figure 2. Cyt-*b* MCC tree from coalescent genealogical modelling. The two cyt-*b* lineages are highlighted by their respective colour: *west* (orange) and *east* (blue). The *H38* haplotype is coloured in purple. Open and closed circles represent nodes with posterior probabilities > 0.90 and > 0.95, respectively, and for the two lineages there is information for the Bayesian posterior probability above, and maximum likelihood bootstrap below the nodes; grey bars represent node height 95% HPD for the root and the points of coalescence for the sequences in the *west* and *east* lineages. The bar below the tree gives timescale using the common vole (*Microtus arvalis*) substitution rate (μ = 3.2 x 10^{-7} substitutions/site/year) according to Martínková *et al.* (2013); white represents glacial periods and black, warm periods; timings given of Last Glacial Maximum (LGM) and Younger Dryas (YD) in Iberia (López-García *et al.* 2012; Bañuls-Cardona *et al.* 2014); kya – kiloyears ago.

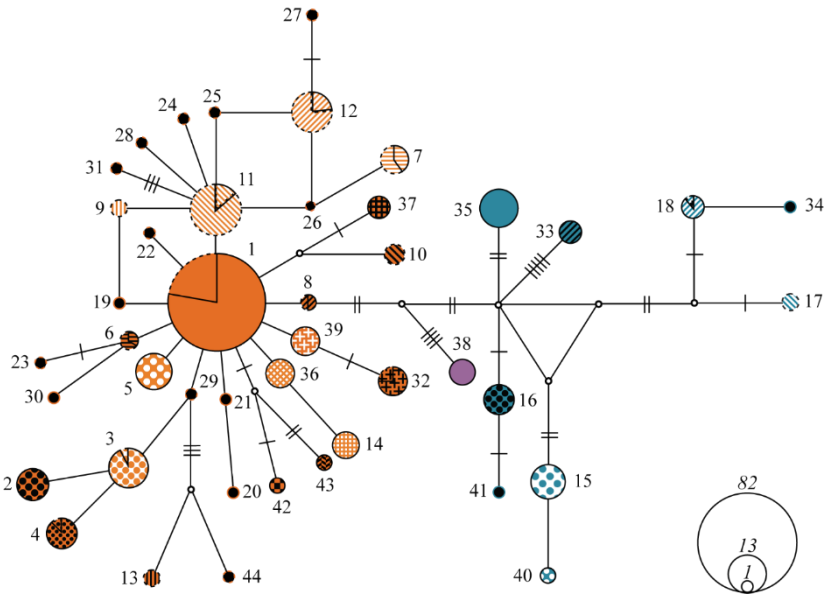


Figure 3. Median joining cyt-*b* haplotype network. Numbered circles represent haplotypes belonging to one of the two major cyt-*b* lineages: *west* (orange) and *east* (blue), and the *H38* haplotype (purple), with individual patterning for each haplotype. The circle area is proportional to the number of sequences of a given haplotype and the dashed outer line indicates the proportion of non-invasive samples. Haplotypes represented by a single sequence are shown as closed circles with an outer line colour according to the respective lineage, inferred haplotypes are represented by open circles; connecting lines between circles correspond to one mutational step, with extra mutations shown by perpendicular strikes.

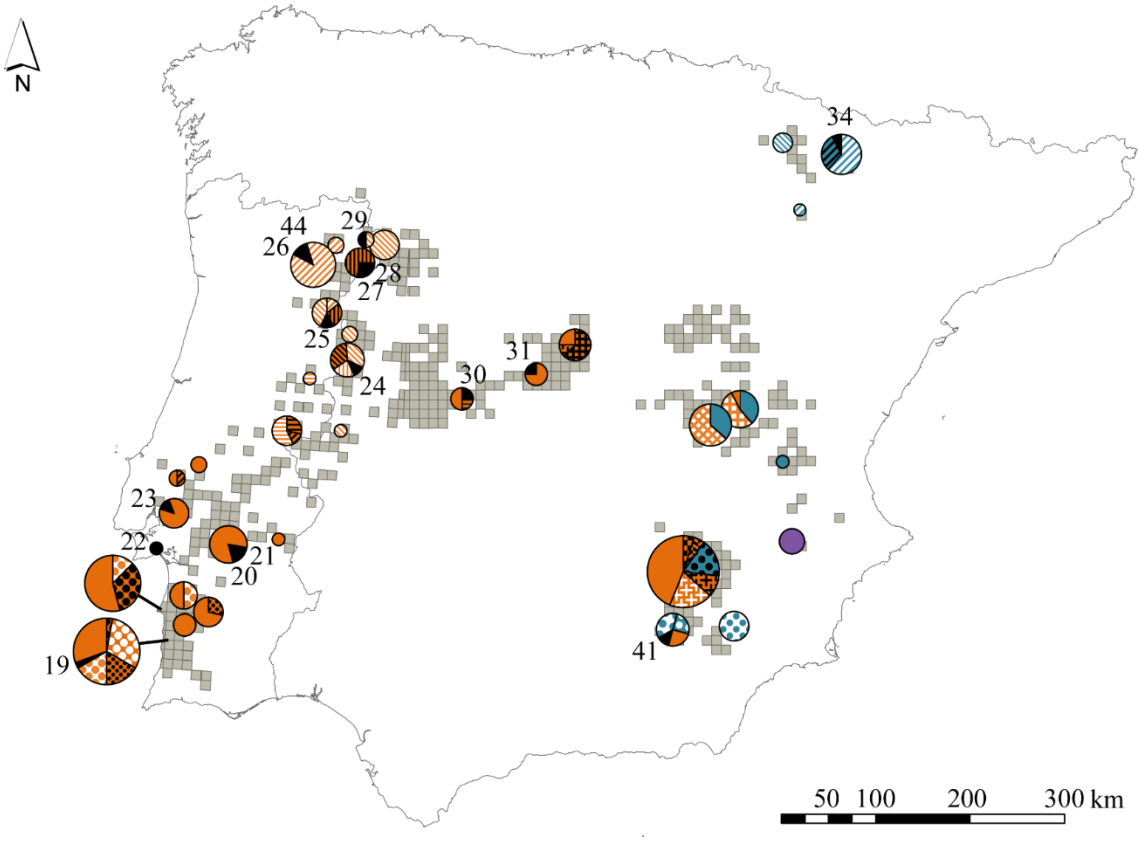


Figure 4. Cyt-*b* haplotype frequencies on map of the Iberian Peninsula with the current distribution of *Microtus cabrerae* in grey (Mira *et al*. 2008; Fernandez-Salvador *et al*. 2002; Garrido-García *et al*. 2013; this study). Samples within 10 km are grouped into pies; the size of the circles is proportional by area (min. 1 sample, max. 25 samples); colours represent the two mitochondrial lineages: *west* (orange) and *east* (blue), and the *H38* haplotype (purple); following Figure 2 each pattern represents a particular haplotype and black indicates singleton haplotypes which are distinguished by their identification numbers.

However, considering the variation within each lineage, the *east* has the highest π_{mt} in *Bt* and *pP* (Figure S2B, Supporting information), while the *west* has the highest π_{mt} in the north of *Lc* (Figure S2C, Supporting information).Considering the geographic nuclei independently, we found that *Lc* has exclusively *west* lineage haplotypes, having significantly lower π_{mt} when compared to the remaining three geographic nuclei (0.216±0.015%) (Figures 4 and S2A, and Table S2, Supporting information). *pP* has exclusively *east* haplotypes, and it is the eastern geographic nucleus with the lowest π_{mt} (0.355±0.065%), however significantly ($p < 0.05$) higher than *Lc* (purely *west*). *Mb* and *Bt* include haplotypes from both lineages and as a consequence have the highest π_{mt} (*Mb*: 0.555±0.042%; *Bt*: 0.582±0.040%). Nucleotide diversities are all significantly different among the four geographic nuclei ($p < 0.05$), with the highest π_{mt} recorded in *Bt*, followed by *Mb*, *pP* and finally *Lc*.

For the 10 nuclear loci, we screened 79 voles: 53 and 24 assigned to the *west* and *east* mitochondrial lineages, respectively, and 2 with the *H38* mitochondrial haplotype (Table S2, Supporting information). Within the 39 males, the *SMCY7* variation was mostly limited to the embedded microsatellite ($TTTTC_n$ repeat motif). Only one male (locality 17, Figure 1) showed variation at the sequence level for this locus – a 5 bp deletion adjacent to the *5'* end of the microsatellite. Considering only the microsatellite, we obtained a total of 11 alleles, ranging from 150 to 200 bp. Eight of the alleles were present in *Lc*, four in *Mb*, three in *Bt* and two in *pP* (3, 0, 1 and 2 were private, respectively) (Figure S3, Supporting information). Regarding the remaining nine loci, *PNPO* showed the highest *h* (0.916±0.011%) and π (0.414±0.017%), while *DBX7* had the lowest *h* (0.290±0.054) and *BRCA1* the lowest π (0.053±0.010%) (Table S2, Supporting information). Some individual nuclear loci networks showed clearer geographic partitioning than others, but specific haplotypes were never limited to particular geographic areas (Figure S3, Supporting Information). *pP* had a significantly lower ($p < 0.05$) proportional π_{nuc} (0.303±0.069) than the other geographic nuclei, which did not differ significantly from each other (*Lc*: 0.685±0.097; *Mb*: 0.692±0.099; *Bt*: 0.735±0.135).

Spatial structure analysis

The BAPS test for possible geographic division of mitochondrial and nuclear genetic diversity showed the best supported number of clusters to be $K = 3$ and $K = 2$ ($p = 1$), respectively (Figure S4A, S4B, Supporting information). The two nuclear clusters show no concordance to the distribution observed with the cyt-*b* lineages, for which the software additionally identified a third population (*pP*; Figure S4A, Supporting information).

The AMOVA results show contrasting patterns for the different types of markers (mitochondrial and nuclear), considering the allocation of genetic diversity within and among the four geographic nuclei (Figure S5, Supporting information). For the cyt-*b* gene, 42.8% of the genetic variation is attributed to differences among nuclei (Figure S5A, Supporting information). For *SMCY7* and autosomal data, the equivalent values are 10.1%and 21.3%, respectively (Figure S5B and S5C, Supporting information). Concordantly, average F_{ST} between samples was the lowest for *SMCY7* (0.27, $p < 0.05$), intermediate for the remaining nuclear loci (0.47, $p < 0.05$), and the highest for cyt-*b* (0.68, $p < 0.05$). The concatenated autosomal NeighborNet showed shallow

substructure, with only some degree of geographic isolation for *pP* (Figure S5C, Supporting information).

Demographic analyses

We found a significant proportion of mitochondrial genetic variance explained by both geographic location (*LatLong*, 37.1%) and geographic distance (*PCNM*, 12.6%), which increased after controlling for the other variable (40.3% and 15.8%, respectively) (Table 1). Regarding the nuclear loci, we also found a significant proportion of variance explained by both predictors (17.9% for *LatLong* and 9.6% for *PCNM*), but only *LatLong* was significant after controlling for the other variable (10.0% for *LatLong* and 1.7% for *PCNM*).

A hypothesis of genetic neutrality was rejected for cyt-*b* considering either the whole data set, or the *west* mitochondrial lineage and *Lc* geographic nucleus separately, based on *D*, *R2* and *Fs* tests, and a significantly higher proportion of nonsynonymous to synonymous substitutions results for the Z-tests for all groups (Table S2, Supporting information). Given that neutrality test results may indicate population expansion, the BSP is relevant. The BSP suggests stable population sizes over time both for the complete data and the *west* and *east* lineages separately, but there are some signs of recent slight population growth in the *west* lineage (from about 5000 years ago) and possible decline in the *east* lineage (from about 2000 years ago) (Figure S6, Supporting information). For the nuclear loci, the various tests indicate deviations from neutrality ($p < 0.05$) for six of the nine nuclear genes (*BRCA1*, *IRBP*, *RAG1*, *OSTA*, *PNPO* and *DBX7*) including a higher proportion of nonsynonymous to synonymous substitutions for *RAG1* (Table S2, Supporting information). Considering each geographic nucleus separately, we found deviations from neutral expectations at 6 (*Lc*), 2 (*Mb*), 2 (*Bt*) and 0 (*pP)* of the 9 loci (Table S2, Supporting information).

Phylogeographic scenarios

Given that the *tMRCA* for cyt-*b* dates on average to the LGM (Figure 2), and the highest number of fossils from this time derive from southern Iberia, all scenarios were hypothesized to begin with a single southern population from which different groups expanded or diverged after expansion. After testing all possible combinations within each of five plausible phylogeographic scenarios in *phase I* and selecting the best supported version (results not shown), these selected versions of the five scenarios were

themselves compared in *phase II* (Figure S1, Supporting information, and the "Methods" section for further details).

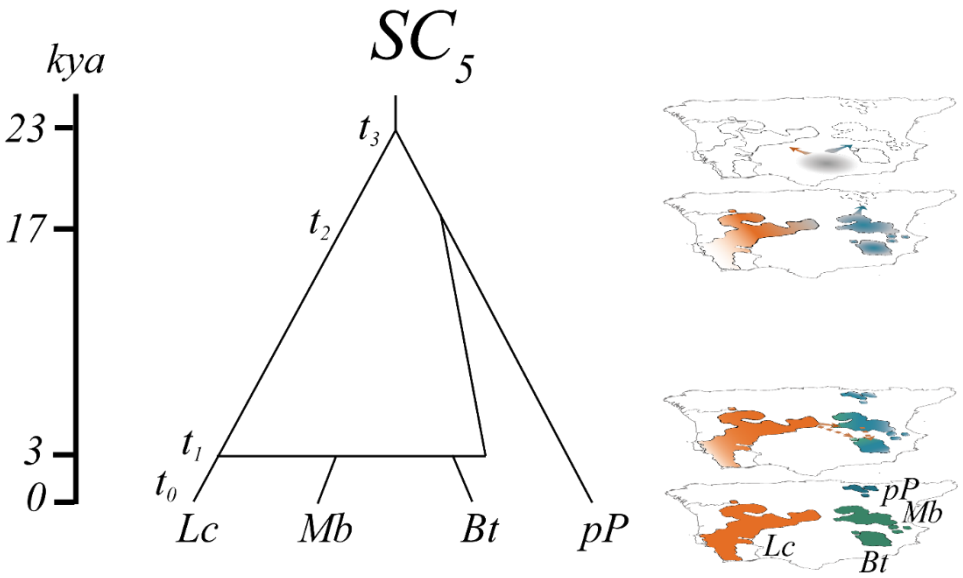


Figure 5. Phylogeographic scenario with the highest posterior probability in the *DIYABC* analysis (see also text, Table S4 and Figure S1, Supporting information). On the left there is a graphical representation of the phylogenetic tree with a time scale in kya (kiloyears ago) for the estimated time points of each split. On the right there is a schematic representation of the supported population splits and the expected cyt-*b* characteristics (colours) of the voles in each geographic area at each given time point; ancestral lineage (grey), *west* lineage (orange), *east* lineage (blue) and both lineages (green); arrows indicate population expansion into new areas: *Lusocarpetan* (*Lc*), *pre-Pyrenean* (*pP*), *Montiberic* (*Mb*) and *Betic* (*Bt*).

In *phase II*, the scenario representing allopatric formation of the *east* and *west* lineages and secondary contact (SC_5) was far better supported (73.6%) than the SC_1 (14.7%), SC_2 (6.1%), SC_3 (1.7%) and SC_4 (3.9%) scenarios (parameter values in Table S4, Supporting information). We performed *bias and precision* and *model checking* estimation within the DIYABC package for the best scenario and obtained congruence between observed and simulated parameters, respectively (Tables S4 and S5, Supporting information, respectively). We generated new pseudo-data sets sampling from SC_5 to test for *type I error*, and sampling from SC_1 (second best supported scenario) for *type II error*, comparing one with the other on both analyses using a logistic approach. Error results were 14.4% and 7.8%, respectively, indicating that the wrong model would be chosen under 15% of the time. The parameter estimates for SC_5 are thus very robust, with low bias values, but they have broad confidence intervals and match to a large extent that of the priors (Table S4, Supporting information). The separation of the *west* and *east* groups is estimated at 22.9 kya (mean), at time point t_3. A subsequent

separation of *pP* from the rest of the *east* group (t_2) is estimated at *c.* 17 kya (Figure 5 and Table S4, Supporting information). The secondary contact between the *west* and *east* lineages (t_1) is estimated to be very recent at *c.* 3 kya.

5. Discussion

Spatial genetic structure

The Cabrera vole is subdivided into two well-supported mitochondrial lineages, here designated *west* and *east*, which partially overlap with the genetic units defined in Alasaad *et al.* (2013) based on RAPD-PCR. The mitochondrial *west* lineage has a central haplotype (*H1*) around which appears to have been a star-like radiation (Figure 3). This contrasts with the *east* lineage, which shows a much more dispersed structure with significantly higher nucleotide diversity (0.563% vs. 0.218%), but fewer haplotypes (9 vs. 34) despite a proportional sampling effort. These results suggest that the *west* lineage may have undergone a more recent radiation than the *east* lineage (see Bamshad & Wooding 2003; Brace *et al*. 2012). We also identified one 'intermediate' haplotype with uncertain phylogenetic positioning (*H38*) in animals captured in a peripheral and likely declining population (Figures 2–4).

Most haplotypes have a restricted geographic distribution (Figure 4), especially in the east of Iberia, where a more complex geology might have promoted greater isolation of populations (Vera 2004). The *west* lineage central haplotype, *H1*, however, has a widespread distribution, being found from southwest Portugal, through central to southeast Spain, and is thus present in *Lc*, *Mb* and *Bt*, but not in *pP* (Figure 4). In a scenario of secondary contact between the two lineages, the western more homogeneous landscape could have facilitated the rapid expansion of this haplotype into the east, leading to the high mitochondrial nucleotide diversity in the eastern nuclei of *Bt* and *Mb*. Concordantly, if we consider diversity on a lineage-by-lineage basis, the regions with the highest π_{mt} differ from the overall picture: for the *west* lineage it becomes the northern portion of *Lc*; and for the *east* lineage, it is especially *pP* and, to some extent, *Bt* (Figure S2, Supporting information). These various perspectives need to be considered in relation to possible phylogeographic scenarios (see next section).

Because mitochondrial DNA is maternally inherited, it is also important to determine how nuclear genetic variation is structured for a complete picture of the population history (Edwards & Bensch 2009). Among the 10 nuclear loci analysed, none showed a clear phylogeographic pattern coinciding with the described geographic nuclei or the identified mitochondrial lineages. The highest total average nuclear nucleotide

diversities ($\bar{\pi}_i$) were observed in *Bt*, *Mb* and *Lc*, while *pP* displayed a significantly lower value and was also the most divergent geographic nucleus for all categories of markers (Figures S3 and S5C, Supporting information), possibly reflecting its small effective population size and geographic isolation. This nucleus is further identified as distinctive in the BAPS mitochondrial analysis, but with the nuclear data *pP* is part of a very shallow west–east trend resulting from low genetic divergence (Figure S4, Supporting information).

There is much that differs between the nuclear and mitochondrial DNA results. The AMOVA and dbRDA results for cyt-*b* show that mitochondrial genetic diversity is mostly explained by differences among geographic nuclei, with a pattern of IBD; while at the nuclear level, more genetic variation is found within than among populations (Table 1 and Figure S5, Supporting information). These results are consistent with sexbiased dispersal (Toews & Brelsford 2012). Male-biased dispersal is well documented in mammals and, together with female philopatry, can lead to patterns similar to those observed with our data, as previously reported for another monogamous vole, *Microtus townsendii* (Lambin 1994; Lawson Handley & Perrin 2007). However, with our data alone we cannot rule out that the greater genetic structure for mitochondrial DNA might merely reflect a higher mitochondrial mutation rate and thus faster lineage sorting (Avise 2000). Another difference between nuclear and mitochondrial data is that the secondary contact of both cyt-*b* lineages occurs within the eastern part of the species' range, while the nuclear variation is difficult to delimit, spanning from west to east Figure S4, Supporting information). This could reflect introgression of cyt-*b* haplotype *H1* and its derivatives, fitting with the three fold increase in female N_{ef} in Figure S6B (Supporting information). The *Lc* nucleus is where the mitochondrial haplotype *H1* presumably originated and where it and its derivatives predominate. *Lc* has a similar proportion of $\bar{\pi}_{nuc}$ to *Mb* and *Bt*, but was significantly less variable in mitochondrial diversity than any other group, including *pP*, which is the least variable group at the nuclear level (Table S2 and Figure S7, Supporting information). This discrepancy between *Lc* and the other geographic nuclei presumably reflects expansion of haplotype *H1* and its derivatives, which likely indicates a demographic event, given that in *Lc* there is departure from neutrality at both mitochondrial and nuclear loci. Also the Z-tests of selection do not indicate a specific response for *Lc* but rather indicate that mitochondrial variation is under purifying selection in all populations (Table S2, Supporting information).

Glacial refugia and post-glacial expansion

The Cabrera vole fossil record shows a continuous presence in Iberia as its divergence from *Microtus (Iberomys) brecciensis* during the Middle-Late Pleistocene (*c.* 130 kya, kiloyears ago), and in France intermittently until the end of the Last Glacial Maximum (LGM, *c.* 24 – 20 kya), from when it has been continuously present until the Middle Ages (*c.* 1.5 – 0.5 kya) (Garrido-García & Soriguer-Escofet 2012; Laplana & Sevilla 2013). Using the mitochondrial data alone, with the mitochondrial mutation rate reported by Martínková *et al.* (2013), we estimated the *tMRCA* for the Cabrera vole to date back to *c.* 24 kya, around the LGM (see also Bañuls-Cardona *et al.* 2014). Given the confidence interval of the *tMRCA*, and uncertainties associated with using a mutation rate from another species (albeit from the same genus), the coalescence point of all current cyt-*b* sequences may have occurred at a later time. Before the LGM, the fossil record suggests that the Cabrera vole was widely distributed in the Iberian Peninsula and southern France (Garrido-García & Soriguer-Escofet 2012; Laplana & Sevilla 2013; Bañuls-Cardona *et al.* 2014). Hence, we propose that, during the LGM, harsh climatic conditions and perhaps competition with other species resulted in the extinction of most Cabrera vole populations. The LGM was associated to a steppe-like environment in central Iberia, judging from the replacement of the Cabrera vole by cold tolerant species like *Microtus oeconomus* in the well-studied archaeological assemblages of that period. At that time, Cabrera vole fossils are only found in the southern half of the Iberian Peninsula and the warmer areas of the Mediterranean coast, possibly following the first episode of range contraction for the species (Pita *et al.* 2014). Accordingly, all the scenarios tested in our phylogeographic modelling assume an expansion from a single southern Iberian focus during or after the LGM, including the favoured scenario, which shows expansion of two main groups presumably to the northwest and northeast *c.* 22.9 kya (SC_5 – t_3, Figure 5 and Table S4, Supporting information).

After the LGM, there was a period of increase in temperature and humidity, the Bølling-Allerød period (from *c.* 15 to 13 kya), that likely created favourable habitat for the species to quickly recolonize the north and periphery of the Iberian Peninsula, and reach southern France (López-García *et al.* 2012; Laplana & Sevilla 2013), which agrees with the separation of the *pre-Pyrenean* geographic nucleus, around 17 kya (Figure 5 and Table S4, Supporting information). The time of the population subdivision and the time of emergence of the mitochondrial lineages have very wide confidence intervals which may result from the low genetic diversity observed in the Cabrera vole, and we cannot exclude the possibility of these events being older than the averages presented here, as

discussed later on. There was, however, a subsequent cold period - the Younger Dryas (YD) (*c.* 13–11.4 kya) - that marked the end of the Pleistocene. At this point, the Cabrera vole's favoured habitat would have become scarcer once again for a short period of time as fossil abundance from this period is low (Robinson *et al.* 2006; Garrido-García & Soriguer-Escofet 2012; Laplana & Sevilla 2013; Bañuls-Cardona *et al.* 2014). Such conditions might have favoured further subdivision within each geographic nucleus, as reflected by the high mitochondrial haplotype diversity (Figure 2). Phylogeographic subdivision driven by the YD has previously been suggested for *Microtus agrestis* (Herman & Searle 2011).

With the beginning of the Holocene (*c.* 11.4 kya), temperatures increased, maximising at the Holocene Climatic Optimum (around 9 - 5 kya), leading to wide availability of habitats and expansion of the Cabrera vole (Fletcher *et al.* 2010; Laplana & Sevilla 2013). Few of the available fossils date from the Holocene (11.7 kya onwards), but this is more likely due to lack of studies of microfauna on archaeological sites, than the absence of the species itself. In the favoured phylogeographic model, the extant populations spread in the Holocene from their YD refugia, with increasing effective population sizes and signals of expansion, as well as a later introgression of the *west* group into the *east* estimated at around 3 kya (Figure 2 and Tables S2 and S4, Supporting information). It is unlikely that selective pressures drove the expansion of the *Lusocarpetan* population into the eastern populations, given that all geographic nuclei show signals of being under purifying selection for the mitochondrial gene. It seems more parsimonious that the expansion of *Lc* was mostly a demographic event possibly driven by favourable anthropogenic modifications of the landscape (Laplana & Sevilla 2013). Further support for the demographic hypothesis is given by the deviations to neutrality in other loci, affecting only *Lc*, indicating that this might be a generalized pattern over the genome.

Although we found evidence of population expansion for the *west* lineage, overall the Cabrera vole may have had a declining effective population size in recent years based on the BSP (Figure S6B, Supporting information). This decline could be related to climate-driven habitat loss, specifically associated with the increasing aridity of the late Holocene, most recently associated with agricultural intensification (Laplana & Sevilla 2013). However, human influence on the species is complex and may have an opposite consequence, that is population expansion due for instance to the creation of more open habitat (Garrido-García & Soriguer-Escofet 2012). These results need to be considered carefully: the eastern geographic nuclei might have been isolated from each other for some time, and thus might violate the assumptions of panmixia, leading the BSP to

falsely indicate population decline (Heller *et al.* 2013). The influence of humans in altering habitat and climate and thereby impacting the Cabrera vole needs to be understood through the Holocene.

Finally, we need to stress that the *DIYABC* phylogeographic modelling results need to be viewed with caution. The models tested here do not reflect an exhaustive comparison of all possible scenarios, but are instead based on a selection of hypothesis that we believe are most likely to represent our data (Csilléry *et al.* 2010). More specifically, our analyses show that posterior probabilities of the model parameters are often as wide as their respective priors [mainly for population size estimates (*N*) and the transition/transversion ratio (*k1*)], even after increasing most of the prior parameters range over 200%. This indicates that there is little resolution in our data to estimate values accurately. Our study likely reflects a case where approaches such as ABC may not have enough power to accurately determine population parameters, either because of low genetic diversity or not enough geographic structure, as has already been seen in other studies (Chen *et al.* 2012; Gaubert *et al.* 2015).

The Cabrera vole as an example of a refugial endemic

It is very unusual for a refugial endemic to have a fossil history as detailed as that of the Cabrera vole, and we have been able to exploit that together with a thorough analysis of contemporary genetic structure, using mitochondrial and nuclear markers, and historic distribution information. This work clearly indicates the existence of at least two mitochondrial lineages of Cabrera vole, and the best supported phylogeographic model suggests that they occupied different YD refugia within the Iberian peninsula, in a pattern consistent with the concept of '*refugia within refugia*' (Gómez & Lunt 2007).

The Cabrera vole exhibits interesting similarities with other species of Iberian small mammals also characterized by multiple lineages, both in terms of their population history within the Iberian Peninsula and their conservation needs. One example is the field vole (*Microtus agrestis*), which is divided into three cryptic taxonomic units over its wide Eurasian distribution, two of which - the *Portuguese* and *Southern* lineages - existed in separate LGM refugial areas in western and eastern Iberia, respectively (Paupério *et al.* 2012). The Pyrenean desman (*Galemys pyrenaicus*) is another threatened small mammal that is restricted to northern Iberia. Although the existing estimates of *tMRCA* for the desman are *c.* 0.35 my, they are based on fossil calibrations with very distant taxonomic groups (Dornburg *et al.* 2011; Igea *et al.* 2013), and the true divergence time between the two desman lineages is likely to be much more recent and more comparable

to our estimates for the Cabrera vole lineages. Based on the comparison of genetic variation with the Cabrera vole, the Pyrenean desman likely contracted to northwest Iberia during the LGM, from where the two lineages diverged, also supported by recent genomic analyses (Querejeta *et al.* 2016).

The complexity of the Cabrera vole colonisation history at a local scale is very informative, and the subdivision of other species, such as the field vole and the Pyrenean desman into multiple lineages in Iberia, suggest similar yet individual histories for other species as well. The multiple refugia, range shifts and genetic introgression seen in the Cabrera vole mean that there is a rich population history, albeit recent, with associated genetic variation in this refugial endemic, and, as far as possible, this genetic legacy should be conserved. There is particular anxiety because the species as a whole is already endangered due to the ongoing land use intensification and climate change and, therefore, special efforts are needed to ensure the conservation of the multiple parts of its remaining distribution. Refugial endemics such as the Cabrera vole are, by definition, of conservation concern and these species illustrate the considerable care that is needed for their conservation from a genetic perspective.

Our study on the Cabrera vole is unusually detailed, and employs contemporary molecular clock and fossil data. It provides a phylogeographic perspective on a refugial endemic species comparable to that achieved for more widespread species. As we have shown, refugial endemics should not just be considered forms that reside, immobile, in a region that happens to have been protected from the extremes of the Late Pleistocene. Instead, they represent dynamic systems that have changed their ranges just as dramatically as more widespread species, but at a more local scale. Other groups of Iberian small vertebrates seem to share similar patterns of genetic variation, and combined phylogeographic analyses with some of the many other species which occupied '*refugia within refugia*' (Gómez & Lunt 2007; Centeno-Cuadros *et al.* 2009; Vega *et al.* 2010; Fitó *et al.* 2011; Miraldo *et al.* 2011; van de Vliet *et al.* 2014), would permit generalizations of the colonisation history of refugial endemics comparable to the generalizations for species that spread outside the refugia (Hewitt 1999).

Acknowledgments

We thank A. Arrizabalaga from the Museum of Natural Science of Granollers, A. Sillero, A. Silva, C. Bastos-Silveira from the Portuguese National Natural History Museum, C. Díaz, D. Galícia from the Museum of the University of Navarra, I. Roque and the *Tyto Tagus* team, J.V. Turón, R. García-González from the Spanish Pyrenean

Ecology Institute (IPE-CSIC; samples donated by M. Cabrera), R. Jato, S. Alasaad and S. Peris (including samples donated from M.C. Escala, J.C. Irurzun, A. Rueda and F.M. Santos) for providing samples for this study. We are especially grateful to M.C. Moreira for her help during field work and also C. Viedma, D. Ortega, F. González, F. Pereira, J.C. Albero, J.P. Palacios, J. Romero, M. González, N. Gómez, Q. Ortiz, S. Alcobia, V. Corcoles and V. Díaz. We also appreciate the contribution of A. Ortuño, A. Tortola, A. Vela, E. Escoriza, F. Almansa, J. Escudero, J. Gisbert, J. Idañez, J. Jiménez, J.A. Lozano, J. Pozo, J.M. Vadillo, M. Alcántara and N. González in providing information on sampling locations and helping facilitate field work. The necessary trapping permits for this work were provided by the Portuguese Instituto da Conservação da Natureza e das Florestas (319/2012/CAPT), Comunidad de Madrid (10/076408), Generalitat Valenciana (60/2014-VS-FAU14-011), Gobierno de Aragon (550201.24.2013.1031; INAGA/500201/24/2014/337), Gobierno de Castilla-la-Mancha (DGMEN/SEN/avp_13_013_aut; DGMEN/SEN/avp_14_022_aut), Junta de Castilla y León (EP/AV/407/2012; EP/SA/71/2014) and Región de Murcia (AUF/2013/0038; AUF/2014/0010).

This work was partially supported by Programa Operacional Potencial Humano – Quadro de Referência Estratégico Nacional funds from the European Social funds and Portuguese Ministério da Educação e Ciência through Fundação para a Ciência e Tecnologia (FCT SFRH/BD/77726/2011 PhD grant to S. Barbosa), by the European Regional Development Fund (ERDF) through COMPETE program and Portuguese national funds through FCT (PTDC/BIA-BEC/105110/2008, PTDC/AAG-MAA/3227/2012 and PTDC/BIA-BIC/6582/2014 research projects). J. Paupério had a post-doctoral grant funded by the project “Genomics and Evolutionary Biology” co-financed by North Portugal Regional Operational Programme 2007/2013 (ON.2 – O Novo Norte), under the National Strategic Reference Framework, through the ERDF.

6. References

Abdi H, Williams LJ (2010) Tukey’s Honestly Significant Difference (HSD) Test. In: *Encyclopedia of Research Design* (eds Salkind NJ, Dougherty DM, Frey B), pp. 583–585. Sage, Thousand Oaks, CA, USA.

Abellán P, Svenning J-C (2014) Refugia within refugia – patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. *Biological Journal of the Linnean Society*, **113**, 13–28.

Alasaad S, Rozhkovan K V, Atopkin DM *et al.* (2013) RAPD-PCR molecular analysis of the threatened Cabrera's vole populations in the Iberian Peninsula. *African Journal of Biotechnology*, **12**, 4065–4073.

Avise JC (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA, USA.

Baele G, Lemey P (2013) Bayesian evolutionary model testing in the phylogenomics era: matching model complexity with computational efficiency. *Bioinformatics*, **29**, 1970–1979.

Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. *Nature Reviews Genetics*, **4**, 99–111.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bañuls-Cardona S, López-García JM, Blain H-A, Lozano-Fernández I, Cuenca-Bescós G (2014) The end of the Last Glacial Maximum in the Iberian Peninsula characterized by the small-mammal assemblages. *Journal of Iberian Geology*, **40**, 19–27.

Barbosa S, Paupério J, Searle JB, Alves PC (2013) Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to non-invasive sampling. *Molecular Ecology Resources*, **13**, 43–56.

Beysard M, Heckel G (2014) Structure and dynamics of hybrid zones at different stages of speciation in the common vole (*Microtus arvalis*). *Molecular Ecology*, **23**, 673–687.

Bilton DT, Mirol PM, Mascheretti S *et al.* (1998) Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London B*, **265**, 1219–1226.

Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Brace S, Palkopoulou E, Dalén L *et al.* (2012) Serial population extinctions in a small mammal indicate Late Pleistocene ecosystem instability. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 20532–20536.

Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255–265.

Bullejos M, Sánchez A, Burgos M *et al.* (1997) Multiple, polymorphic copies of *SRY* in both males and females of the vole *Microtus cabrerae*. *Cytogenetics and Cell Genetics*, **79**, 167–171.

Cabral MJ, Raposo de Almeida P, Almeida J *et al.* (2005) *Livro Vermelho dos Vertebrados de Portugal*. Instituto da Conservação da Natureza, Lisboa, Portugal.

Cabrera-Millet M, Britton-Davidian J, Orsini P (1982) Génétique biochimique comparée de *Microtus cabrerae* Thomas, 1906 et de trois autres espèces d'Arvicolidae méditerranéens. *Mammalia*, **46**, 381–388.

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.

Chen SY, Zhang YJ, Wang XL *et al.* (2012) Extremely low genetic diversity indicating the endangered status of *Ranodon sibiricus* (Amphibia: Caudata) and implications for phylogeography. *PLoS One*, **7**, e33378.

Corander J, Sirén J, Arjas E (2007) Bayesian spatial modeling of genetic population structure. *Computational Statistics*, **23**, 111–129.

Cornuet J-M, Pudlo P, Veyssier J *et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187–1189.

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution*, **25**, 410–8.

Cuenca-Bescós G, López-García JM, Galindo-Pellicena MA *et al.* (2014) The Pleistocene history of *Iberomys*, an endangered endemic rodent from South Western Europe. *Integrative Zoology*, **9**, 481–497.

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

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

Dornburg A, Beaulieu JM, Oliver JC, Near TJ (2011) Integrating fossil preservation biases in the selection of calibrations for molecular divergence time estimation. *Systematic Biology*, **60**, 519–527.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**, e88.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*, **18**, 2930–2933.

Ellerman JR, Morrison-Scott TCS (1951) *Checklist of Palaearctic and Indian Mammals, 1758-1946*. Trustees of the British Museum.

Ersts PJ (2015) Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation.

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

Fernandes M, Pita R, Mira APP (2008) *Microtus cabrerae*. *The IUCN Red List of Threatened Species*.

Fernández-Salvador R, García-Perea R, Ventura J (2001) Reproduction and postnatal growth of the Cabrera vole, *Microtus cabrerae*, in captivity. *Canadian Journal of Zoology*, **79**, 2080–2085.

Fink S, Fischer MC, Excoffier L, Heckel G (2010) Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia: *Microtus*): the utility of AFLPs versus mitochondrial and nuclear sequence markers. *Systematic Biology*, **59**, 548–572.

Fitó F, Rivera X, Roca J *et al.* (2011) Extremely low level of genetic variability in Iberian *Pelobates cultripes* (Cuvier, 1829) (Amphibia; Anura; Pelobatidae). *Butlettí de la Societat Catalana d'Herpetologia*, **19**, 21–28.

Fletcher WJ, Sánchez Goñi MF, Allen JRM *et al.* (2010) Millennial-scale variability during the last glacial in vegetation records from Europe. *Quaternary Science Reviews*, **29**, 2839–2864.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Garrido-García JA, Rosário IT, Gisbert J *et al.* (2013) Revisión a nivel ibérico de la distribución del topillo de Cabrera o iberón, *Iberomys cabrerae* (Thomas, 1906). *Galemys*, **25**, 35–49.

Garrido-García JA, Soriguer-Escofet RC (2012) Cabrera's Vole *Microtus cabrerae* Thomas, 1906 and the subgenus *Iberomys* during the Quaternary: evolutionary implications and conservation. *Geobios*, **45**, 437–444.

Gaubert P, Del Cerro I, Centeno-Cuadros A *et al.* (2015) Tracing historical introductions in the Mediterranean Basin: the success story of the common genet (*Genetta genetta*) in Europe. *Biological Invasions*, **17**, 1897–1913.

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

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Heller R, Chikhi L, Siegismund HR (2013) The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS One*, **8**, e62992.

Herman JS, McDevitt AD, Kawałko A *et al.* (2014) Land-bridge calibration of molecular clocks and the post-glacial colonization of Scandinavia by the Eurasian field vole *Microtus agrestis*. *PLoS One*, **9**, e103949.

Herman JS, Searle JB (2011) Post-glacial partitioning of mitochondrial genetic variation in the field vole. *Proceedings of the Royal Society B*, **278**, 3601–3607.

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

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

Hope AG, Ho SYW, Malaney JL, Cook JA, Talbot SL (2014) Accounting for rate variation among lineages in comparative demographic analyses. *Evolution*, **68**, 2689–2700.

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

Igea J, Aymerich P, Fernández-González A *et al.* (2013) Phylogeography and postglacial expansion of the endangered semi-aquatic mammal *Galemys pyrenaicus*. *BMC Evolutionary Biology*, **13**, 115.

Jaarola M, Martínková N, Gündüz I *et al.* (2004) Molecular phylogeny of the speciose

vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **33**, 647–663.

Jaarola M, Searle JB (2002) Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology*, **11**, 2613–2621.

Jiménez R, Carnero A, Burgos M, Sánchez A, Díaz de la Guardia R (1991) Achiasmatic gant sex chromosomes in the vole *Microtus cabrerae* (Rodentia, Microtidae). *Cytogenetics and Cell Genetics*, **57**, 56–58.

Joly S, Bruneau A (2006) Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. *Systematic Biology*, **55**, 623–636.

Kass RE, Raftery AE (1995) Bayes Factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kryštufek B, Buzan E V, Hutchinson WF, Hänfling B (2007) Phylogeography of the rare Balkan endemic Martino's vole, *Dinaromys bogdanovi*, reveals strong differentiation within the western Balkan Peninsula. *Molecular Ecology*, **16**, 1221–1232.

Lambin X (1994) Sex ratio variation in relation to female philopatry in Townsend's voles. *Journal of Animal Ecology*, **63**, 945–953.

Laplana C, Sevilla P (2013) Documenting the biogeographic history of *Microtus cabrerae* through its fossil record. *Mammal Review*, **43**, 309–322.

Lawson Handley LJ, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology*, **16**, 1559–1578.

Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1–24.

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

López-García JM, Blain H-A, Morales JI *et al.* (2012) Small-mammal diversity in Spain during the late Pleistocene to early Holocene: climate, landscape, and human impact. *Geology*, **41**, 267–270.

Luque-Larena JJ, López P (2007) Microhabitat use by wild-ranging Cabrera voles *Microtus cabrerae* as revealed by live trapping. *European Journal of Wildlife Research*, **53**, 221–225.

Malaney JL, Cook JA (2013) Using biogeographical history to inform conservation: the case of Preble's meadow jumping mouse. *Molecular Ecology*, **22**, 6000–6017.

Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, **16**, 562–563.

Martínková N, Barnett R, Cucchi T *et al.* (2013) Divergent evolutionary processes associated with colonization of offshore islands. *Molecular Ecology*, **22**, 5205–5220.

Mestre F, Pita R, Paupério J *et al.* (2015) Combining distribution modelling and non-invasive genetics to improve range shift forecasting. *Ecological Modelling*, **297**, 171–179.

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, 1–8.

Mira APP, Marques CC, Santos SM, Rosário IT, Mathias ML (2008) Environmental determinants of the distribution of the Cabrera vole (*Microtus cabrerae*) in Portugal: implications for conservation. *Mammalian Biology*, **73**, 102–110.

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

Moritz CC, Potter S (2013) The importance of an evolutionary perspective in conservation policy planning. *Molecular Ecology*, **22**, 5969–5971.

Oksanen J, Blanchet FG, Kindt R *et al.* (2013) Package "vegan." *Community Ecology Package*, **2**.

Palomo LJ, Gisbert J (2002) *Atlas de los Mamiferos Terrestres de España*. Dirección General de Conservación de la Naturaleza, SECEM-SECEMU, Madrid.

Paupério J, Herman JS, Melo-Ferreira J *et al.* (2012) Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia. *Molecular Ecology*, **21**, 6015–6032.

Pavlova A, Amos JN, Joseph L *et al.* (2013) Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution*, **67**, 3412–3428.

Pita R, Mira APP, Beja P (2006) Conserving the Cabrera vole, *Microtus cabrerae*, in intensively used Mediterranean landscapes. *Agriculture, Ecosystems and*

Environment, **115**, 1–5.

Pita R, Mira APP, Beja P (2014) *Microtus cabrerae* (Rodentia: Cricetidae). *Mammalian Species*, **912**, 48–70.

QGIS Development Team (2008) Quantum GIS Geographic Information System. Open source Geospatial Foundation project.

Querejeta M, González-Esteban J, Gómez A *et al.* (2016) Genomic diversity and geographical structure of the Pyrenean desman. *Conservation Genetics*, **17**, 1333–1344.

Rambaut A (2009) FigTree v1.3.1: tree figure drawing tool. Available from *FigTree* website.

Rambaut A, Drummond AJ (2007) Tracer v1.5.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rivas-Martínez S (1981) Les étages bioclimatiques de la végétation de la Péninsule Ibérique. *Anales del Jardín Botánico de Madrid*, **37**, 251–268.

Robinson SA, Black S, Sellwood BW, Valdes PJ (2006) A review of palaeoclimates and palaeoenvironments in the Levant and Eastern Mediterranean from 25,000 to 5000 years BP: setting the environmental background for the evolution of human civilisation. *Quaternary Science Reviews*, **25**, 1517–1541.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpazees. *Molecular Biology and Evolution*, **10**, 512–526.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Taylor I (1994) *Barn Owls: Predator-Prey Relationships and Conservation*. Cambridge University Press, Cambridge, UK.

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

Vega R, Amori G, Aloise G *et al.* (2010) Genetic and morphological variation in a Mediterranean glacial refugium: evidence from Italian pygmy shrews, *Sorex minutus* (Mammalia: Soricomorpha). *Biological Journal of the Linnean Society*, **100**, 774–787.

Ventura J, López-Fuster MJ, Cabrera-Millet M (1998) The Cabrera vole, *Microtus cabrerae*, in Spain: a biological and a morphometric approach. *Netherlands Journal of Zoology*, **48**, 83–100.

Vera JA (2004) *Geología de España*. Instituto Geológico y Minero de España, Madrid.

van de Vliet MS, Diekmann OE, Machado M *et al.* (2014) Genetic divergence for the amphibian *Pleurodeles waltl* in southwest Portugal: dispersal barriers shaping geographic patterns. *Journal of Herpetology*, **48**, 38–44.

Author contributions

This study is a part of S.B.'s PhD thesis work on the Cabrera vole, *Microtus cabrerae*, which she is developing under the supervision of P.C.A. and J.B.S. Genetic laboratory work and analyses were conducted at Cornell University and CIBIO, with contributions from J.P., R.P., C.F., P.B. and A.M., that are part of the Cabrera vole project team and contributed with new samples, to the data analysis and writing of the manuscript. J.S.H. contributed to the data analysis and writing of the manuscript. H.M.V-G., J.A.C., J.A.G-G. and R.C.S. contributed with new samples and to the writing of the manuscript.

Data Accessibility

DNA sequences and information on each sampled specimen (locality, haplotypes and accession numbers) are deposited at GenBank with the accession numbers KY380111 to KY381117; DNA alignments are deposited in DRYAD (doi:10.5061/dryad.9fc86).

7. Supporting information

Appendix S1. List of primers and thermal cycling conditions for the genes used in this study.

Gene	Size (bp)	Primer (Forward)	Sequence (5' - 3')	Primer (Reverse)	Sequence (5' - 3')	Tn (ºC)	Reference	Assession numbers
mitochondrial								
cyt-*b*	1143	*L14727-SP*	GACAGGAAAAATCATCGTTG	*H15915-SP*	TTCATTACTGGTTTACAAGAC	TD 60-56	Jaarola & Searle (2002)	KY380571 - KY380693 KY380776 - KY380854
mitochondrial non-invasive								
cyt-*b*.1S	395	*L14727-SP*	GACAGGAAAAATCATCGTTG	*cyt-b.R1S*	ACATAGCCCATAAATGCTGTTG	TD 61-53	this study	KY380694 - KY380775
cyt-*b*.2S	410	*cyt-b.F2S*	TCTTCATCTGCCTCTTCCTC	*cyt-b.R2S*	TGTATAGTAGGGGTGGAACGG	TD 61-53	this study	
cyt-*b*.3S	379	*cyt-b.F3S*	CACTGCCCTCGTATTAGTCC	*cyt-b.R3S*	GAGTGTTCGTTGCTTTGAGG	TD 61-53	this study	
cyt-*b*.4S	308	*cyt-b.F4S*	CCATCTTACGCTCTATCCCTAAC	*H15915-SP*	TTCATTACTGGTTTACAAGAC	TD 61-53	this study	
autosomal exons								
BRCA1	1153	*BRCA1MF*	ACAAGGAGCCAACGGAACAGAT	*BRCA1MR*	ATGGACCGGTTCCTTGAGGTTT	TD 65-55	Paupério *et al.* (2012)	KY380111 - KY380186
IRBP	948	*IRBP.F2S*	GCAGGCTATGAAGAGTCRTG	*IRBP.R2S*	AGCACGGAYACCTGAAACA	TD 58-52	Barbosa *et al.* (2013)	KY380416 - KY380493
RAG1	1250	*S70*	TCCGAGTGGAAATTTAAGMTGTT	*S73*	GAGGAAGGTRTTGACACGGATG	TD 65-55	Steppan *et al.* (2004)	KY380347 - KY380415
autosomal introns								
COPS7A	779	*COPS7A-4F_Rd*	TACAGCATYGGVCGRGACATCCA	*COPS7A-4R_Rd*	TSACYTGCTCYTCRATGCCYGACA	TD 63-55	adapted from Igea *et al.* (2010)	KY380494 - KY380570
OSTA	874	*OSTA5F*	TGMWGGYCATGGTGGAAGGCTTTG	*OSTA5R*	AGATGCCRTCRGGGAYGAGRAACA		adapted from Igea *et al.* (2010)	KY380855 - KY380928
PNPO	764	*PNPO-3F_Rd*	GAYGGCTTCCGCTTCTTCACTAACTA	*PNPO-3R_Rd*	GGYTCCCARTAGAAGACAAGRGA		adapted from Igea *et al.* (2010)	KY380929 - KY381005
SLC38A7	617	*SLC38A7-8F_Rd*	AGGCCTGGCMGCCTGCTTCATCTT	*SLC38A7-8R_Rd*	TCYGABAGYTTGGCTTGRATGAGGCA		adapted from Igea *et al.* (2010)	KY381006 - KY381078
X-linked introns								
DBX5	225	*DBX5F*	CAACAACTGTCCTCCACA	*DBX5R*	CATGATAATTTCTCCCATCTC	TD 65-55	Hellborg & Ellegren (2004)	KY380187 - KY380265
DBX7	319	*DBX7F*	GCCCATCTTGAGTCAGAT	*DBX7R*	CAACTCTCTCGTTGGTGC		Hellborg & Ellegren (2004)	KY380266 - KY380346
Y-linked intron								
SMCY7	200	*SMCY7F*	TGGAGGTGCCCRAARTGTA	*SMCY7R*	AACTCTGCAAASTRTACTCCT	TD 63-55	Hellborg & Ellegren (2003)	KY381079 - KY381117

Annealing temperature are in degrees Celsius [Tn(ºC)], touch-down temperature range (TD). PCR reactions were performed with a total of 5 µl using 2 µl Qiagen© PCR Multiplex Kit Master Mix (Qiagen, Hilden, Germany), 0.2 µM of each primer and 10-20 ng of genomic DNA. The product obtained was purified using ExoSAP-IT® PCR clean-up Kit (GE Healthcare, Piscataway, NJ, USA) and forward and reverse sequences were generated with both amplification primers. For the mitochondrial data, sequencing reactions were carried out using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Carlsbad, CA, USA) on automated sequencers. Forward and reverse sequences were assembled and edited in Sequencher 4.7 (Gene Codes Corp., Ann Arbor, MI, USA), verified by eye and aligned using ClustalW (Thompson *et al.* 1994) as implemented in BioEdit 7.0 (Hall 1999).

Table S1. Sample origin, number and type for the different datasets used in the mitochondrial analysis [*Live Trapping and Non-Invasive* (*LTNI*) localities] and nuclear analysis [*Live Trapping* (*LT*) localities]. Individuals within a 10 km radius are grouped into localities (letters).

LTNI localities						*LT localities*												Location		
								N_{ind}												
Code	Lat	Long	N_{ind}		Map code	Lat	Long	*BRCA1*	*IRBP*	*RAG1*	*COPS7A*	*OSTA*	*PNPO*	*SLC38A7*	*DBX5*	*DBX7*	*SMCY7*	Locality	District/Province	Country
A	41.518	-6.891		(2)															Bragança	PT
B	41.581	-6.518		(2)															Bragança	PT
C	41.533	-6.330		(7)															Bragança	PT
D	41.304	-7.183	4	(12)	***1***	41.353	-7.300	4	4	3	4	3	4	1	4	4	2	Mirandela	Bragança	PT
E	41.343	-6.577	3	(4)	***2***	41.335	-6.578	2	2	2	2	2	2	2	2	2	-	Mogadouro	Bragança	PT
F	40.891	-6.959		(7)															Guarda	PT
G	40.671	-6.699		(2)															Guarda	PT
H	40.468	-6.669		(9)															Guarda	PT
I	40.226	-7.178		(1)															Castelo Branco	PT
J	39.763	-7.399	3	(4)	***3***	39.717	-7.289	3	3	3	3	3	3	3	3	3	3	Monforte da Beira	Castelo Branco	PT
K	39.736	-6.743		(1)															Santarém	PT
L	39.340	-8.468		(2)															Santarém	PT
M	39.228	-8.726		(2)															Santarém	PT
N	38.870	-8.725	2	(5)	***4***	38.822	-8.806	2	2	2	2	2	2	2	2	2	1	Benavente	Santarém	PT
O	38.530	-8.923		(1)															Setúbal	PT
P	38.632	-8.045	11		***5***	38.657	-8.064	6	6	6	6	6	6	6	6	6	3	Évora	Évora	PT
Q	38.098	-8.561	2	(4)	***6***	38.160	-8.599	1	1	0	1	1	0	0	1	1	1	Grândola	Setúbal	PT
R	37.959	-8.771	25		***7***	37.966	-8.783	3	3	2	3	3	3	3	3	3	2	Sines	Setúbal	PT
S	37.953	-8.284		(7)															Beja	PT
T	37.828	-8.523	4		***8***	37.843	-8.521	2	2	1	2	1	1	2	2	2	1	Bicos	Beja	PT
U	37.687	-8.773	25		***9***	37.755	-8.753	5	5	5	5	5	5	5	5	5	3	Vila Nova de Milfontes	Beja	PT
			10		***10***	37.610	-8.786	2	2	2	2	2	2	2	2	2	1	Cabo Sardão	Beja	PT
V	38.703	-7.456		(1)															Cáceres, Estremadura	SP

N_{ind} - Number of tissue samples and in parentheses the number of non-invasive samples per 10 km grouped locality (letters); Map code – identifier of the populations in Figure 1 for which tissue samples are available; Lat – Latitude, Long – Longitude; PT – Portugal, SP - Spain.

Table S1 (cont.). Sample origin, number and type for the different datasets used in the mitochondrial analysis [*Live Trapping and Non-Invasive (LTNI) localities*] and nuclear analysis [*Live Trapping (LT) localities*]. Individuals within a 10 km radius are grouped into localities (letters).

LTNI localities						*LT localities*												Location		
								N_{ind}												
Code	Lat	Long	N_{ind}		Map code	Lat	Long	*BRCA1*	*IRBP*	*RAG1*	*COPS7A*	*OSTA*	*PNPO*	*SLC38A7*	*DBX5*	*DBX7*	*SMCY7*	Locality	District/Province	Country
W	40.114	-5.256	4		***11***	40.116	-5.251	4	4	4	4	4	4	4	4	4	1	Candeleda	Ávila, Castilla y Leon	SP
X	40.349	-4.350	3	(1)	***12***	40.388	-4.247	1	2	2	1	2	2	2	2	2	1	Navas del Rey	Madrid	SP
Y	40.622	-3.877	5	(3)	***13***	40.718	-3.971	1	1	1	1	1	1	1	1	1	1	Becerril de la Sierra	Madrid	SP
Z	42.548	-1.211		(3)															Navarra/ Huesca, Aragon	SP
AA	42.438	-0.466	13		***14***	42.440	-0.482	8	8	7	8	8	8	8	8	8	6	Sabiñanigo	Huesca, Aragon	SP
AB	41.894	-1.002		(1)															Huesca, Aragon	SP
AC	40.026	-1.817	13		***15***	40.025	-1.806	6	6	5	6	6	7	7	7	7	2	Cañada del Hoyo	Cuenca, Castilla-la-Mancha	SP
AD	39.861	-2.158	14		***16***	39.873	-2.172	3	3	3	3	2	3	3	3	3	1	Las Valeras	Cuenca, Castilla-la-Mancha	SP
AE	39.508	-1.291	1		***17***	39.518	-1.292	1	1	1	1	1	1	1	1	1	1	Utiel	Valencia	SP
AF	38.730	-1.184	5		***18***	38.735	-1.181	2	2	2	2	1	2	2	2	2	-	Yecla	Murcia	SP
AG	38.475	-2.509	24	(1)	***19***	38.534	-2.414	9	9	8	9	9	9	9	9	9	3	Riopar	Albacete, Castilla-la-Mancha	SP
			16		***20***	38.416	-2.609	5	5	3	5	3	5	4	5	5	3	Siles	Jaén, Andalucia	SP
AH	37.895	-2.649	8		***21***	37.897	-2.632	1	2	2	2	1	2	1	2	2	1	Canal de San Clemente	Granada, Andalucia	SP
AI	37.950	-1.888	7		***22***	37.957	-1.878	5	5	5	5	5	5	5	5	5	2	Caravaca de la Cruz	Murcia	SP
Total			202	(82)				76	78	69	77	71	77	73	79	79	39			

N_{ind} - Number of tissue samples and in parentheses the number of non-invasive samples per 10 km grouped locality (letters); Map code – identifier of the populations in Figure 1 for which tissue samples are available; Lat – Latitude, Long – Longitude; PT – Portugal, SP - Spain.

Table S2. Number of samples, genetic diversity indices and neutrality test statistics calculated for the complete dataset for each locus, the three identified cyt-*b* lineages *west* (*W*), *east* (*E*) and haplotype *H38* for the cyt-*b* gene only, and for each geographic nucleus [*Lusocarpetan* (*Lc*), *Montiberic* (*Mb*), *Betic* (*Bt*) and *pre-Pyrenean* (*pP*)] for all loci*.

Gene	N_{ind}	N_{seq}	N_{seg}	N_{hap}	*h (SD)*	*π [%] (SD)*	*θw [%] (SD)*	*D*	F_S	R_2	*Ratio*	*dN/dS*
cyt-*b*	284	284	71	44	0.896 (0.014)	0.500 (0.031)	0.998 (0.232)	**-1.52**	**-12.73**	**0.040**	**4.458**	**<**
W	228	228	44	34	0.847 (0.020)	0.218 (0.014)	0.641 (0.164)	**-1.95**	**-21.19**	0.040	**3.478**	**<**
E	51	51	23	9	0.845 (0.022)	0.563 (0.022)	0.447 (0.151)	0.67	4.27	0.138	**3.858**	**<**
H38	5	5	0	1	-	-	-	-	-	-	-	-
Lc	179	179	38	29	0.841 (0.021)	0.216 (0.015)	0.577 (0.155)	**-1.83**	**-15.92**	**0.032**	**3.077**	**<**
Mb	34	34	16	5	0.758 (0.036)	0.555 (0.042)	0.340 (0.129)	2.08	7.93	0.198	**3.694**	**<**
Bt	54	54	25	9	0.821 (0.031)	0.582 (0.040)	0.480 (0.159)	0.68	4.75	0.132	**3.693**	**<**
pP	17	17	11	4	0.669 (0.091)	0.355 (0.065)	0.285 (0.128)	0.92	3.99	0.180	**3.314**	**<**
BRCA1	76	152	15	12	0.466 (0.049)	0.053 (0.010)	0.232 (0.078)	**-2.02**	**-9.11**	0.033	-0.400	≠
Lc	36	72	6	7	0.495 (0.065)	0.050 (0.008)	0.107 (0.050)	-1.27	**-3.58**	0.053	-1.107	≠
Mb	13	26	5	5	0.409 (0.117)	0.046 (0.016)	0.114 (0.060)	**1.70**	**-2.52**	0.090	-0.694	≠
Bt	19	38	10	6	0.555 (0.077)	0.078 (0.026)	0.206 (0.086)	**-1.88**	-1.69	0.100	1.708	≠
pP	8	16	0	1	-	-	-	-	-	-	-	-
IRBP	78	156	22	29	0.785 (0.032)	0.170 (0.013)	0.413 (0.125)	**-1.64**	**-25.39**	0.036	-0.337	≠
Lc	37	74	12	18	0.804 (0.037)	0.163 (0.016)	0.260 (0.098)	-1.03	**-12.02**	0.065	-0.578	≠
Mb	13	26	6	7	0.723 (0.080)	0.111 (0.021)	0.166 (0.083)	-0.98	**-2.82**	0.087	-0.206	≠
Bt	20	40	13	12	0.868 (0.034)	0.254 (0.025)	0.322 (0.126)	-0.66	-3.36	0.088	-0.138	≠
pP	8	16	1	2	0.125 (0.106)	0.013 (0.011)	0.032 (0.032)	-1.16	-0.70	0.242	1.011	≠
RAG1	69	138	35	39	0.949 (0.007)	0.323 (0.012)	0.509 (0.143)	-1.09	**-22.37**	0.057	**2.695**	**<**
Lc	33	66	17	19	0.909 (0.017)	0.227 (0.020)	0.286 (0.100)	-0.61	**-7.45**	0.082	**2.332**	**<**
Mb	12	24	19	13	0.917 (0.036)	0.344 (0.039)	0.407 (0.158)	-0.56	-3.41	0.056	**2.214**	**<**
Bt	17	34	21	20	0.934 (0.027)	0.325 (0.027)	0.411 (0.149)	-0.72	**-10.28**	0.090	**3.019**	**<**
pP	7	14	4	4	0.692 (0.094)	0.153 (0.025)	0.101 (0.060)	1.71	1.08	0.239	1.111	≠
COPS7A	77	154	10	10	0.622 (0.039)	0.110 (0.011)	0.229 (0.087)	-1.25	-4.17	0.043	-	-
Lc	36	72	8	9	0.696 (0.054)	0.123 (0.015)	0.212 (0.090)	-1.07	**-3.67**	0.060	-	-
Mb	13	26	4	5	0.557 (0.104)	0.083 (0.019)	0.135 (0.077)	-1.03	-1.96	0.086	-	-
Bt	20	40	2	3	0.304 (0.084)	0.040 (0.012)	0.060 (0.044)	-0.64	-0.67	0.102	-	-
pP	8	16	3	3	0.508 (0.126)	0.113 (0.033)	0.116 (0.075)	-0.07	0.73	0.147	-	-
OSTA	71	142	13	12	0.406 (0.052)	0.064 (0.010)	0.270 (0.095)	**-1.96**	**-10.10**	**0.027**	-	-
Lc	35	70	9	8	0.265 (0.070)	0.051 (0.016)	0.215 (0.088)	**-2.01**	**-6.27**	0.044	-	-
Mb	11	22	3	4	0.463 (0.120)	0.058 (0.017)	0.094 (0.060)	-0.98	-1.43	0.099	-	-
Bt	17	34	5	6	0.693 (0.065)	0.112 (0.019)	0.140 (0.073)	-0.53	-1.55	0.098	-	-
pP	8	16	0	1	-	-	-	-	-	-	-	-
PNPO	77	154	17	28	0.916 (0.011)	0.414 (0.017)	0.397 (0.129)	0.12	**-11.50**	0.093	-	-
Lc	35	70	12	15	0.845 (0.029)	0.350 (0.026)	0.326 (0.123)	0.20	-3.69	0.110	-	-
Mb	14	28	12	11	0.873 (0.037)	0.490 (0.034)	0.404 (0.167)	0.71	-1.58	0.149	-	-
Bt	20	40	11	14	0.906 (0.024)	0.454 (0.023)	0.338 (0.138)	1.04	-3.18	0.156	-	-
pP	8	16	4	5	0.700 (0.080)	0.171 (0.021)	0.158 (0.079)	0.27	-0.75	0.156	-	-

Table S2 (cont.). Number of samples, genetic diversity indices and neutrality test statistics calculated for the complete dataset for each locus, the three identified cyt-*b* lineages *west* (*W*), *east* (*E*) and haplotype *H38* for the cyt-*b* gene only, and for each geographic nucleus [*Lusocarpetan* (*Lc*), *Montiberic* (*Mb*), *Betic* (*Bt*) and *pre-Pyrenean* (*pP*)] for all loci*.

Gene	N_{ind}	N_{seq}	N_{seg}	N_{hap}	*h (SD)*	*π [%] (SD)*	*θw [%] (SD)*	*D*	F_S	R_2	*Ratio*	*dN/dS*
SLC38A7	73	146	8	11	0.831 (0.014)	0.330 (0.012)	0.233 (0.096)	0.95	-0.83	0.127	-	-
Lc	33	66	8	10	0.854 (0.023)	0.324 (0.019)	0.272 (0.117)	0.49	-1.44	0.124	-	-
Mb	14	28	7	7	0.807 (0.040)	0.348 (0.020)	0.292 (0.138)	0.58	-0.34	0.151	-	-
Bt	18	36	4	5	0.651 (0.049)	0.187 (0.018)	0.156 (0.088)	0.47	-0.06	0.143	-	-
pP	8	16	6	4	0.642 (0.103)	0.325 (0.057)	0.293 (0.153)	0.38	1.43	0.162	-	-
DBX5	79	119	3	4	0.292 (0.050)	0.134 (0.024)	0.249 (0.151)	-0.80	-1.49	0.050	-	-
Lc	37	54	1	2	0.073 (0.048)	0.032 (0.021)	0.098 (0.098)	-0.88	-0.95	**0.036**	-	-
Mb	14	24	1	2	0.522 (0.030)	0.232 (0.013)	0.119 (0.119)	1.60	1.59	0.261	-	-
Bt	20	31	2	3	0.340 (0.100)	0.159 (0.050)	0.223 (0.164)	-0.58	-0.56	0.089	-	-
pP	8	10	0	1	-	-	-	-	-	-	-	-
DBX7	79	119	5	6	0.290 (0.054)	0.106 (0.022)	0.303 (0.150)	**-1.35**	**-3.94**	0.036	-	-
Lc	37	54	5	6	0.512 (0.076)	0.195 (0.037)	0.346 (0.176)	-1.04	**-2.46**	0.064	-	-
Mb	14	24	0	1	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.00	0.00	0.00	-	-
Bt	20	31	1	2	0.125 (0.077)	0.039 (0.024)	0.078 (0.078)	-0.77	-0.47	0.062	-	-
pP	8	10	0	1	-	-	-	-	-	-	-	-

*except for *SMCY7*, for which there was no sequence variation.

Parameters are abbreviated as follows: number of individuals (N_{ind}), number of sequences (N_{seq}), number of segregating sites (N_{seg}), number of haplotypes (N_{hap}), haplotype diversity and standard deviation [*h(SD)*], percentage of nucleotide diversity and standard deviation [*π [%] (SD)*], percentage of Watterson's theta mutation rate and standard deviation [*θw [%] (SD)*], Tajima's *D*, Fu's F_S, Ramos-Onsins & Rozas's R_2; Z-test of selection using the ratio of non-synonymous to synonymous substitutions in coding regions (*Ratio*), with the respective alternative hypotheses tested (*dN/dS*: if there were significant differences in the *Ratio* (*dN ≠ dS*), we performed a second test for purifying selection (<), given that all significant *Ratio* values were positive); numbers and symbols in bold are significant (p **<** 0.05).

Table S3. Mitochondrial haplotype and nucleotide diversities for groups of individuals within a geographic nuclei and each 100 km buffer zone.

Groups	Centroid		Genetic diversity						
	Latitude	Longitude	N_{ind}	N_{seg}	N_{hap}	*h (SD)*	E_{hap} *(%)*	*π [%] (SD)*	E_{nuc} *(%)*
Lc	39.621	-7.039	179	38	29	0.841(0.021)	2.5	0.216(0.015)	6.9
I	38.445	-8.357	97	13	11	0.686(0.044)	6.4	0.136(0.012)	8.8
II	40.580	-6.930	61	20	15	0.791(0.037)	4.7	0.213(0.031)	14.6
III	40.313	-4.428	15	13	6	0.769(0.081)	10.5	0.252(0.054)	21.4
Mb	39.479	-1.517	34	16	5	0.756(0.038)	5.0	0.559(0.041)	7.3
W	39.479	-1.517	16	2	3	0.592(0.067)	11.3	0.057(0.010)	17.5
E	39.479	-1.517	13	0	1	-	-	-	-
Bt	38.178	-2.301	56	25	9	0.810(0.034)	4.2	0.569(0.042)	7.4
W	38.178	-2.301	35	10	5	0.632(0.075)	11.9	0.160(0.037)	23.1
E	38.178	-2.301	23	16	5	0.704(0.063)	8.9	0.401(0.062)	15.5
pP	42.307	-0.891	17	11	4	0.669(0.091)	13.6	0.355(0.065)	18.3

Parameters abbreviated as follows: total number of individuals (N_{ind}), number of segregating sites (N_{seg}), number of haplotypes (N_{hap}), haplotype diversity and respective standard deviation [*h (SD)*], percentage of error of haplotype diversity estimation [E_{hap} *(%)*], nucleotide diversity and respective standard deviation [*π [%] (SD)*], percentage of error of nucleotide diversity estimation [E_{nuc} (%)]. Groups of *100 km* buffer zones are listed in Figure S2: *Lusocarpetan* (*Lc*) *I*, *II* and *III*, *pre-Pyrenean* (*pP*), *Montiberic* (*Mb*) and *Betic* (*Bt*). All parameters were also calculated independently for each lineage where there is lineage sympatry within the *100 km* buffer zones [*west* (*W*) and *east* (*E*) lineage].

Table S4. Parameters used in the *DIYABC* analysis, respective priors, estimated results and bias computed for scenario SC_5.

Priors			Results					Bias and precision	
Parameter	Conditions	Distribution [min-max]	mean	median	mode	$Q_{2.5}$	$Q_{97.5}$	true	mean
N_1	*Lc*	Uniform [10 - 300 000]	163 000	161 000	165 000	57 800	282 000	145 200	151 900
N_2	*Mb*	Uniform [10 - 300 000]	61 000	44 900	30 100	9 100	226 000	153 400	162 600
N_3	*Bt*	Uniform [10 - 300 000]	62 600	48 900	44 700	11 500	218 000	155 300	156 900
N_4	*pP*	Uniform [10 - 300 000]	17 300	9 620	6 910	2 170	93 800	156 600	165 600
$N_{5.5}$		Uniform [10 - 300 000]	170 000	176 000	281 000	19 000	295 000	136 800	150 600
$N_{5.6}$		Uniform [10 - 300 000]	68 400	50 700	30 300	8 270	240 000	154 900	150 200
$N_{5.7}$		Uniform [10 - 300 000]	193 000	207 000	297 000	33 900	296 000	153 000	147 400
$N_{5.8}$		Uniform [10 - 300 000]	19 100	16 300	12 700	3 380	49 700	142 100	127 600
$t_{5.1}$		Uniform [10 - 60 000]	3 070	2 640	1 910	696	7 980	15 090	13 730
$t_{5.2}$	$t_{5.2}>t_{5.1}$	Uniform [10 - 60 000]	17 000	15 400	11 100	4 810	38 500	30 770	28 090
$t_{5.3}$	$t_{5.3}>t_{5.2}$	Uniform [10 - 60 000]	22 900	20 200	15 500	5 460	53 700	45 640	42 220
$r_{1(M)}$		Uniform [0.001 - 0.999]	0.553	0.561	0.554	0.100	0.950	0.522	0.536
$r_{2(B)}$		Uniform [0.001 - 0.999]	0.470	0.459	0.463	0.068	0.921	0.522	0.518
μ_{mic}		Uniform [10^{-4} - 10^{-3}]	5.37×10^{-4}	5.25×10^{-4}	4.57×10^{-4}	1.83×10^{-4}	9.45×10^{-4}	5.57×10^{-4}	5.70×10^{-4}
P_{mic}		Uniform [10^{-1} - 3×10^{-1}]	0.18	0.17	0.10	0.10	0.28	0.20	0.20
SNI_{mic}		Log-u [10^{-8} - 10^{-5}]	1.52×10^{-7}	2.59×10^{-8}	1.00×10^{-8}	1.00×10^{-8}	1.25×10^{-6}	1.36×10^{-6}	1.67×10^{-6}
u_{exo}	HKY	Uniform [10^{-9} - 10^{-7}]	2.47×10^{-8}	2.04×10^{-8}	1.50×10^{-8}	5.63×10^{-9}	6.79×10^{-8}	4.88×10^{-8}	4.69×10^{-8}
$k1_{exo}$	HKY	Uniform [0.05 - 20]	11.40	12.00	20.00	0.87	19.90	9.84	9.66
u_{int}	HKY	Uniform [10^{-9} - 10^{-7}]	1.78×10^{-8}	1.46×10^{-8}	9.81×10^{-9}	4.42×10^{-9}	5.17×10^{-8}	5.06×10^{-8}	4.81×10^{-8}
$k1_{int}$	HKY	Uniform [0.05 - 20]	11.10	11.60	20.00	0.79	19.90	9.90	9.66
u_{sex}	HKY	Uniform [10^{-9} - 10^{-7}]	3.01×10^{-8}	2.47×10^{-8}	1.38×10^{-8}	5.32×10^{-9}	8.11×10^{-8}	4.85×10^{-8}	4.41×10^{-8}
$k1_{sex}$	HKY	Uniform [0.05 - 20]	11.10	11.60	20.00	0.76	19.80	9.93	9.90
$u_{cyt\text{-}b}$	HKY	Uniform [8.0×10^{-8} – 3.2×10^{-7}]	1.07×10^{-7}	9.97×10^{-8}	9.09×10^{-8}	8.03×10^{-8}	1.83×10^{-7}	1.97×10^{-7}	2.00×10^{-7}
$k1_{cyt\text{-}b}$	HKY	Uniform [0.05 - 20]	10.60	10.70	20.00	0.70	19.70	9.77	8.93

Parameters: Labels regarding effective population sizes (*N*), time of events in generations (*t*) and admixture rates (*r*) are represented in the models depicted in Figures 5 and S1, Supporting information; Microsatellite ($_{mic}$) parameters: mean mutation rate (μ), a parameter determining the shape of the gamma distribution of the individual locus mutation rate (*P*), and the Single Nucleotide Insertion rate (*SNI*). Exonic ($_{exo}$), intronic ($_{int}$), sex-linked ($_{sex}$) and mitochondrial ($_{cyt\text{-}b}$) locus parameters: per site and generation mutation rate (*u*) and transition/transversion ratio (*k1*). **Conditions:** Twin letters indicate the respective geographic nucleus: *Lusocarpetan* (*Lc*), *Montiberic* (*Mb*), *Betic* (*Bt*) and *pre-Pyrenean* (*pP*); time of events (*t*) are controlled relative to each other; admixture rates (*r*) for each admixed population have information on the contribution of $N_{5.6}$ to *M* (r_1) and *B* (r_2) respectively, and the contribution of $N_{5.5}$ is $1\text{-}r_n$; the sequence data were simulated under a Hasegawa-Kishino-Yano (HKY) mutation model. **Distribution:** For all models, parameter distributions were left as default with the exception of maximum *N* values, which were set to four times the estimated mitochondrial effective population size (N_e); maximum *t* values, set to 60 kya (kiloyears years ago) above the upper 95% CI of the *tMRCA* using the cyt-*b* gene; minimum and maximum mutation rate for the cyt-*b* gene ($u_{cyt\text{-}b}$) were set for the slowest [estimated for the *Microtus* voles, Brunhoff *et al.* (2003)] and fastest [estimated for the *Microtus arvalis*, Martínková *et al.* (2013)] rates available. The same priors were used for the modelling of all scenarios, only varying the number of populations (*N*) and time points (*t*) accordingly. **Results:** $Q_{2.5}$ – quintile 2.5%; $Q_{97.5}$ – quintile 97.5%. **Bias and precision:** true – true simulated values; mean – mean average values estimated for the preferred model.

Table S5. Model checking of summary statistics used on the *DIYABC* analysis with values corresponding to the simulations of scenario SC_5 using the posterior distribution of parameters. *SumStat*: summary statistics; *Obs*: observed values; *Sim<Obs*: *goodness-of-fit* of the observed value of each statistic with its simulated distribution by providing the proportion of simulated values lower than observed values. In bold are *p-values* ≤ 0.05 and ≥ 0.95, representing the probability that the simulated data could be more extreme than the observed data.

SumStat	*Obs*	*Sim<Obs*	*SumStat*	*Obs*	*Sim<Obs*
NAL_M_1	8.0	0.229	LIK_M_3&4	0.8	0.641
NAL_M_2	3.0	0.339	LIK_M_4&1	1.8	0.790
HET_M_1	0.9	0.232	LIK_M_4&2	1.3	0.508
HET_M_2	0.8	0.339	LIK_M_4&3	1.5	0.687
HET_M_3	0.8	0.083	DAS_M_1&2	0.2	**0.967**
VAR_M_1	27.6	0.832	DAS_M_1&3	0.1	0.319
VAR_M_2	18.8	0.668	DAS_M_1&4	0.0	0.111
VAR_M_3	10.9	0.391	DAS_M_2&3	0.0	0.090
MGW_M_1	0.4	**0.016**	DAS_M_2&4	0.0	0.259
MGW_M_2	0.3	0.217	DAS_M_3&4	0.0	0.168
MGW_M_3	0.4	0.157	DM2_M_1&2	1.6	0.322
MGW_M_4	0.6	0.313	DM2_M_1&3	34.0	0.894
N2P_M_1&2	8.0	0.113	DM2_M_1&4	108.5	0.905
N2P_M_1&3	9.0	0.116	DM2_M_2&3	21.0	0.788
N2P_M_1&4	10.0	0.308	DM2_M_2&4	136.1	0.933
N2P_M_2&3	6.0	0.109	DM2_M_3&4	264.1	**0.989**
N2P_M_2&4	5.0	0.235	NHA_E_1	14.7	0.600
N2P_M_3&4	5.0	**0.029**	NHA_E_2	8.3	0.525
H2P_M_1&2	0.8	0.113	NHA_E_3	12.7	0.684
H2P_M_1&3	0.9	0.292	NSS_E_1	11.7	0.402
H2P_M_1&4	0.9	0.312	NSS_E_2	10.0	0.477
H2P_M_2&3	0.9	0.231	NSS_E_3	14.7	0.555
H2P_M_3&4	0.8	0.115	MPD_E_1	1.7	0.463
V2P_M_1&2	25.5	0.765	MPD_E_2	2.0	0.501
V2P_M_1&3	29.4	0.799	MPD_E_3	2.5	0.603
V2P_M_1&4	41.2	0.883	VPD_E_1	1.8	0.436
V2P_M_2&3	16.8	0.491	VPD_E_2	2.5	0.506
V2P_M_2&4	43.1	0.890	VPD_E_3	2.9	0.539
V2P_M_3&4	74.5	**0.964**	DTA_E_1	-1.0	0.556
FST_M_1&3	0.1	0.945	DTA_E_2	-1.1	0.346
FST_M_2&3	0.2	0.934	DTA_E_3	-1.1	0.438
LIK_M_1&2	0.6	**0.035**	DTA_E_4	0.2	0.634
LIK_M_1&3	1.0	0.724	PSS_E_1	6.0	0.395
LIK_M_1&4	1.0	0.651	PSS_E_2	2.0	0.242
LIK_M_2&3	0.6	0.827	PSS_E_3	5.7	0.531
LIK_M_2&4	0.5	0.532	PSS_E_4	1.0	0.271
LIK_M_3&1	0.7	0.490	MNS_E_1	5.5	0.506
LIK_M_3&2	0.8	0.710	MNS_E_2	2.4	0.345

SumStat	*Obs*	*Sim<Obs*	*SumStat*	*Obs*	*Sim<Obs*
MNS_E_3	3.5	0.514	MPD_I_1	1.5	0.726
MNS_E_4	2.0	0.330	MPD_I_2	1.8	0.753
VNS_E_1	32.7	0.347	MPD_I_3	1.5	0.655
VNS_E_2	3.4	0.274	MPD_I_4	1.1	0.753
VNS_E_3	11.2	0.494	VPD_I_1	1.5	0.637
NH2_E_1&2	19.7	0.594	VPD_I_2	1.8	0.674
NH2_E_1&3	23.0	0.629	VPD_I_3	1.3	0.565
NH2_E_1&4	15.7	0.521	VPD_I_4	1.2	0.670
NH2_E_2&3	16.0	0.563	DTA_I_1	-0.7	0.651
NH2_E_2&4	9.7	0.468	DTA_I_2	-0.2	0.829
NH2_E_3&4	13.7	0.585	DTA_I_3	-0.1	0.930
NS2_E_1&2	17.0	0.419	DTA_I_4	0.1	0.581
NS2_E_1&3	21.0	0.464	PSS_I_1	4.0	0.484
NS2_E_1&4	13.0	0.367	PSS_I_3	1.3	0.152
NS2_E_2&3	17.0	0.460	MNS_I_1	6.4	0.731
NS2_E_2&4	11.3	0.440	MNS_I_2	4.0	0.851
NS2_E_3&4	15.7	0.509	MNS_I_3	5.7	0.922
MP2_E_1&2	1.7	0.474	MNS_I_4	2.5	0.613
MP2_E_1&3	1.8	0.504	VNS_I_1	37.8	0.528
MP2_E_1&4	1.6	0.464	VNS_I_2	8.0	0.779
MP2_E_2&3	2.3	0.587	VNS_I_3	23.3	0.931
MP2_E_2&4	1.6	0.465	VNS_I_4	3.7	**0.952**
MP2_E_3&4	2.2	0.583	NH2_I_1&2	12.8	0.647
MPB_E_1&2	2.2	0.553	NH2_I_1&3	14.0	0.651
MPB_E_1&3	2.4	0.592	NH2_I_1&4	11.3	0.649
MPB_E_1&4	2.1	0.491	NH2_I_2&3	10.5	0.594
MPB_E_2&3	2.3	0.568	NH2_I_2&4	8.0	0.645
MPB_E_3&4	2.0	0.457	NH2_I_3&4	8.8	0.582
HST_E_1&2	0.2	0.901	NS2_I_1&2	11.0	0.497
HST_E_1&3	0.2	0.918	NS2_I_1&3	11.8	0.488
HST_E_1&4	0.2	0.685	NS2_I_1&4	10.3	0.551
HST_E_2&3	0.0	0.313	MP2_I_1&2	1.6	0.733
HST_E_2&4	0.0	0.147	MP2_I_1&3	1.5	0.703
HST_E_3&4	-0.1	0.083	MP2_I_1&4	1.5	0.725
NHA_I_1	10.8	0.728	MP2_I_2&3	1.6	0.701
NHA_I_2	6.8	0.677	MP2_I_2&4	1.6	0.760
NHA_I_3	7.3	0.569	MP2_I_3&4	1.4	0.672
NHA_I_4	3.5	0.794	MPB_I_1&2	1.8	0.754
NSS_I_1	9.8	0.622	MPB_I_1&3	1.6	0.687
NSS_I_2	6.8	0.594	MPB_I_1&4	1.7	0.661
NSS_I_3	5.8	0.399	MPB_I_2&3	1.8	0.733
NSS_I_4	3.5	0.745	MPB_I_2&4	2.0	0.755

SumStat	*Obs*	*Sim<Obs*	*SumStat*	*Obs*	*Sim<Obs*
MPB_I_3&4	1.8	0.716	NS2_S_2&4	5.0	0.710
HST_I_1&2	0.1	0.737	MP2_S_1&2	0.5	0.607
HST_I_1&3	0.1	0.489	MP2_S_1&3	0.4	0.526
HST_I_1&4	0.1	0.410	MP2_S_1&4	0.5	0.586
HST_I_2&3	0.1	0.889	MP2_S_2&3	0.4	0.443
HST_I_2&4	0.2	0.569	MP2_S_2&4	0.5	0.600
HST_I_3&4	0.2	0.754	MP2_S_3&4	0.2	0.273
NHA_S_1	4.5	0.600	MPB_S_1&2	0.7	0.648
NHA_S_2	2.0	0.251	MPB_S_1&3	0.4	0.446
NHA_S_3	2.5	0.296	MPB_S_1&4	0.3	0.326
NSS_S_1	4.0	0.617	MPB_S_2&3	0.5	0.517
NSS_S_2	5.0	0.786	MPB_S_2&4	0.4	0.467
NSS_S_3	1.5	0.285	HST_S_1&2	0.3	0.804
MPD_S_1	0.5	0.575	HST_S_1&4	-0.8	0.117
MPD_S_2	0.6	0.616	HST_S_2&3	0.2	0.946
MPD_S_3	0.2	0.268	HST_S_3&4	-0.7	0.116
VPD_S_1	0.7	0.732	NHA_C_1	15.0	0.314
VPD_S_2	3.2	**0.950**	NHA_C_2	5.0	0.078
VPD_S_3	0.2	0.279	NHA_C_3	6.0	0.065
DTA_S_1	-0.9	0.302	NHA_C_4	3.0	0.500
DTA_S_2	-0.3	0.607	NSS_C_1	18.0	0.184
DTA_S_3	-0.7	0.420	NSS_C_2	16.0	0.319
DTA_S_4	0.0	0.467	NSS_C_3	20.0	0.326
PSS_S_1	3.5	0.788	NSS_C_4	9.0	0.832
PSS_S_2	4.5	**0.963**	MPD_C_1	2.3	0.134
PSS_S_3	0.5	0.261	MPD_C_2	6.8	0.683
PSS_S_4	0.0	0.244	MPD_C_3	7.2	0.700
MNS_S_1	2.8	0.396	MPD_C_4	2.3	0.711
MNS_S_2	6.0	**0.974**	VPD_C_1	2.7	0.154
MNS_S_3	2.5	0.506	VPD_C_2	22.3	0.813
VNS_S_1	2.1	0.388	VPD_C_3	23.3	0.829
VNS_S_3	0.5	0.334	VPD_C_4	12.2	0.885
VNS_S_4	0.0	0.375	DTA_C_1	-1.6	0.195
NH2_S_1&2	5.5	0.533	DTA_C_2	1.4	**0.980**
NH2_S_1&3	5.5	0.486	DTA_C_3	1.0	**0.971**
NH2_S_1&4	4.5	0.514	PSS_C_1	14.0	0.383
NH2_S_2&3	3.0	0.259	PSS_C_2	8.0	0.516
NH2_S_3&4	2.5	0.249	PSS_C_3	11.0	0.448
NS2_S_1&2	9.0	0.782	PSS_C_4	7.0	0.835
NS2_S_1&3	5.0	0.510	MNS_C_1	2.6	0.147
NS2_S_1&4	4.0	0.542	MNS_C_2	4.3	0.974*
NS2_S_2&3	6.0	0.653	MNS_C_3	5.4	0.979*

SumStat	*Obs*	*Sim<Obs*	*SumStat*	*Obs*	*Sim<Obs*
MNS_C_4	1.0	0.207	MP2_C_1&3	3.4	0.217
VNS_C_1	5.0	0.118	MP2_C_1&4	2.3	0.137
VNS_C_2	3.3	0.777	MP2_C_2&3	7.0	0.716
VNS_C_3	10.8	**0.969**	MP2_C_2&4	5.7	0.693
VNS_C_4	0.0	0.220	MP2_C_3&4	6.5	0.704
NH2_C_1&2	19.0	0.183	MPB_C_1&2	6.6	0.588
NH2_C_1&3	20.0	0.152	MPB_C_1&3	7.3	0.659
NH2_C_1&4	18.0	0.316	MPB_C_2&3	8.4	0.758
NH2_C_2&3	10.0	0.029*	MPB_C_2&4	12.8	0.909
NH2_C_2&4	8.0	0.111	MPB_C_3&4	11.1	0.842
NH2_C_3&4	9.0	0.089	HST_C_1&2	0.6	0.946
NS2_C_1&2	32.0	0.250	HST_C_1&3	0.5	0.943
NS2_C_1&3	36.0	0.272	HST_C_1&4	0.8	**0.971**
NS2_C_1&4	34.0	0.516	HST_C_2&3	0.2	0.713
NS2_C_2&3	29.0	0.269	HST_C_2&4	0.6	0.829
NS2_C_2&4	27.0	0.548	HST_C_3&4	0.4	0.748
NS2_C_3&4	29.0	0.465			
MP2_C_1&2	2.8	0.175			

Summary statistics description contains the analysis code, followed by the locus type and finally, for which population(s) the analysis is being performed for. Microsatellite (M) 'One sample summary statistics': NAL – mean number of alleles, HET – mean genic diversity, VAR – mean size variance, MGW – mean Garza-Williamson's *M*; 'Two sample summary statistics': N2P – mean number of alleles, H2P – mean genic diversity, V2P – mean size variance, FST - F_{ST}, LIK – classification index, DAS – shared allele distance, DM2 – $(d\mu)^2$ distance. Sequence [exon (E), intron (I), sex-linked (S) and mitochondrial (C)], 'One sample summary statistics': NHA – number of haplotypes, NSS – number of segregating sites, MPD – mean pairwise differences, VPD – variance of pairwise differences, DTA – Tajima's *D*, PSS – private segregating sites, MNS – mean of numbers of the rarest nucleotide at segregating sites, VNS – variance of numbers of the rarest nucleotide at segregating sites; 'Two sample summary statistics': NH2 – number of haplotypes, NS2 – number of segregating sites, MP2 – mean pairwise differences (W), MPB – mean pairwise differences (B), HST – F_{ST} (Hudson *et al.* 1992). Population codes are as follow: 1 – *Lusocarpetan*, 2 – *Montiberic*, 3 – *Betic*, 4 – *pre-Pyrenean*.

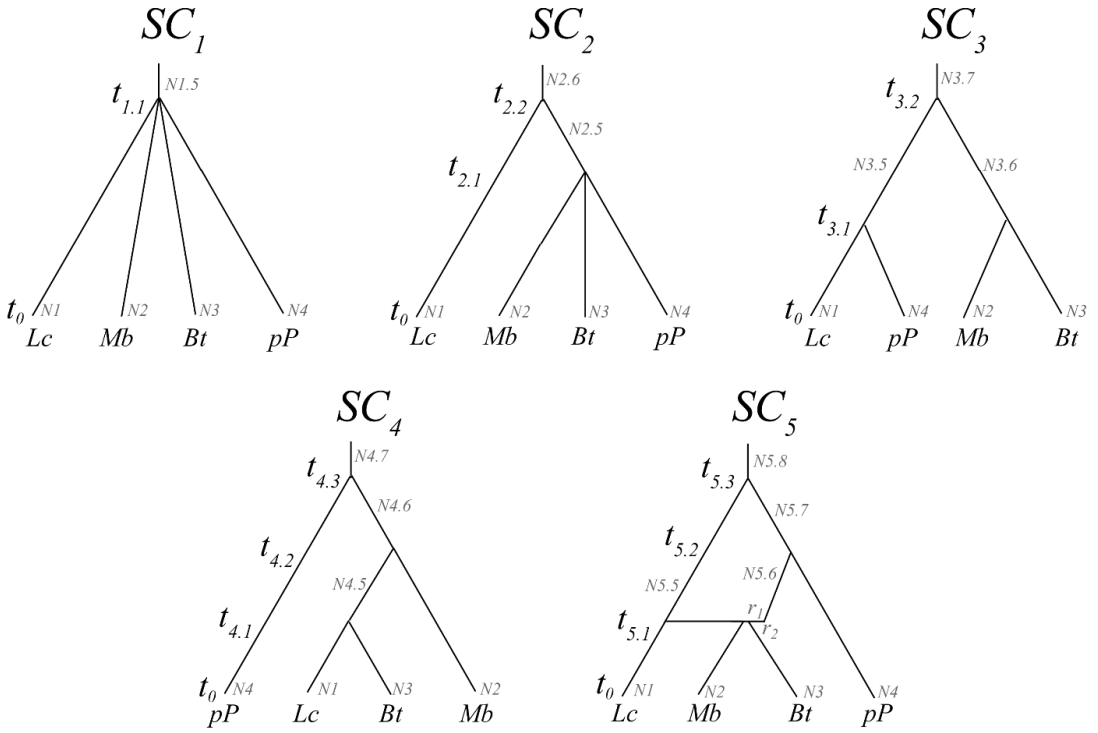


Figure S1. Phylogeographic models tested for the 79 *Microtus cabrerae* tissue samples amplified for the complete loci dataset. Parameters for effective population size (N_x), time splits (t_x) and recombination rates (r), as well as the mutation models specifics for each different marker types, are the same as for SC_5 (as listed in Table S4, Supporting information).

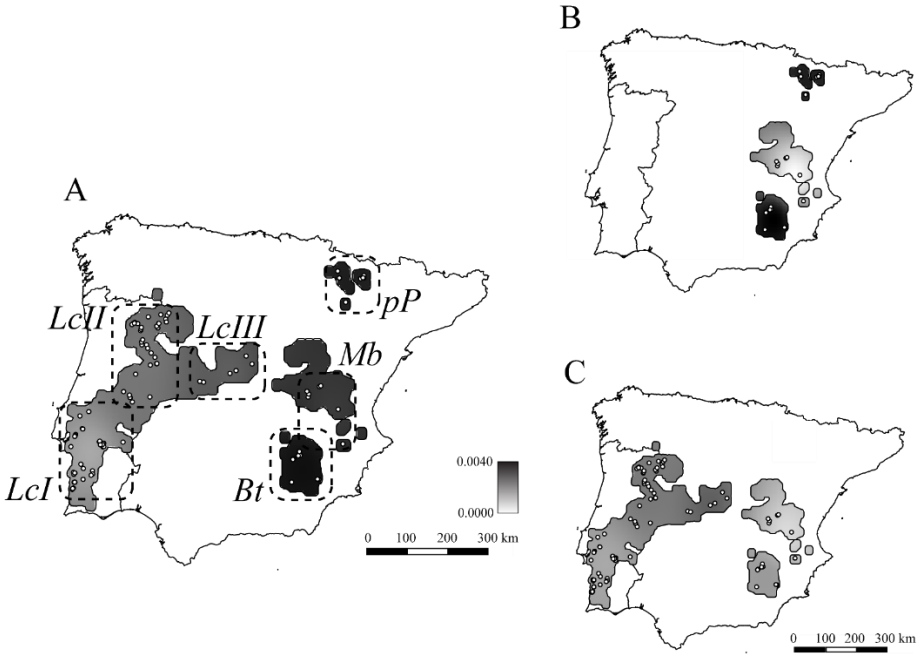


Figure S2. Maps of the Iberian Peninsula with interpolated nucleotide diversities using the cyt-*b* dataset. Cyt-*b* sequences from tissue and non-invasive samples (circles) within 10 km were grouped. Nucleotide diversities were calculated for the centroid of samples grouped within *100 km* buffer zones, and the values were then interpolated for the rest of the species distribution. The four geographic nuclei are the *Lusocarpetan* (*LcI*, *LcII* and *LcIII*), *pre-Pyrenean* (*pP*), *Montiberic* (*Mb*) and *Betic* (*Bt*). **A)** Total nucleotide diversities per group; **B)** Nucleotide diversities per group considering only samples with *east* lineage haplotypes; **C)** Nucleotide diversities per group considering only samples with *west* lineage haplotypes.

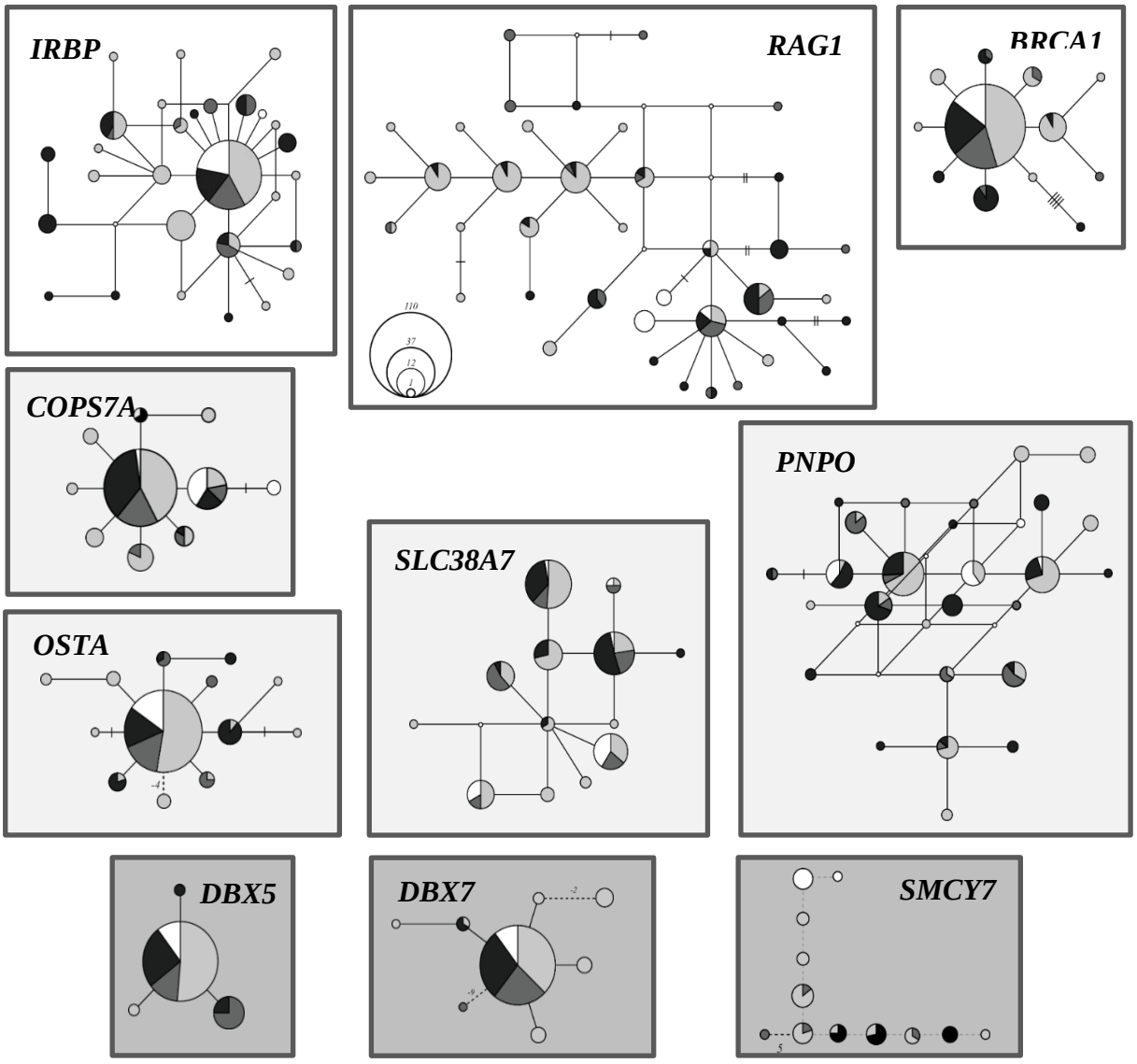


Figure S3. Median-joining haplotype networks for the three nuclear exons (white background), the four nuclear introns (grey background) and the three sex-linked loci (dark grey background). Circles represent haplotypes; shades of grey within circles identify the four geographic nuclei: *Lusocarpetan* (grey), *pre-Pyrenean* (white), *Montiberic* (dark grey) and *Betic* (black); the area of the circle is proportional to the number of sequences; white dots represent inferred haplotypes and each connecting line between any circle corresponds to one mutational step, with extra mutations shown by perpendicular strikes; dashed black represent indels (with the respective number of bp difference); dashed grey lines represent single microsatellite repeats.

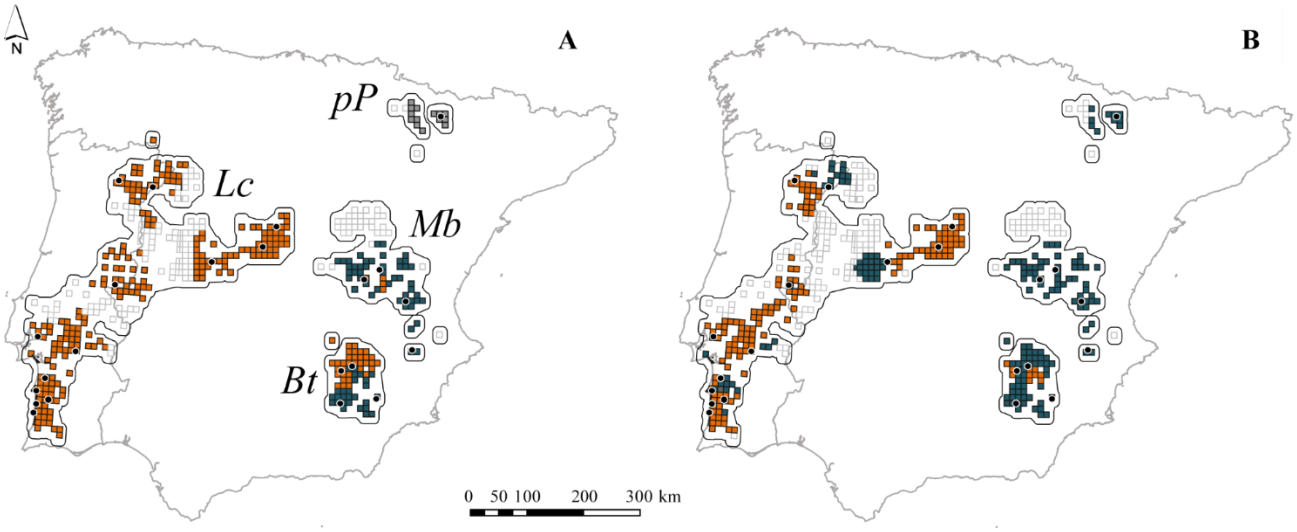


Figure S4. Spatial distribution of the genetic groups supported by BAPS, typed for the 79 *Live Trapping* (*LT*) population samples (black circles) for the mitochondrial (A) and nuclear (B) loci. The four geographic nuclei are the *Lusocarpetan* (*Lc*), *pre-Pyrenean* (*pP*), *Montiberic* (*Mb*) and *Betic* (*Bt*). A) Spatial clustering using cyt-*b* resulting in *K* = 3 populations inferred by BAPS (*p* = 1; orange, blue and grey); B) Spatial clustering using nine nuclear loci (exons: *BRCA1*, *IRBP* and *RAG1*; introns: *COPS7A*, *OSTA*, *PNPO* and *SLC38A7*; X-linked: *DBX5*, *DBX7*) resulting in *K* = 2 populations (*p* = 1, orange and blue); The Voronoi tessellation picture was overlaid onto the *M. cabrerae* distribution and UTM squares were coloured according to the inferred *K* colour for that location.

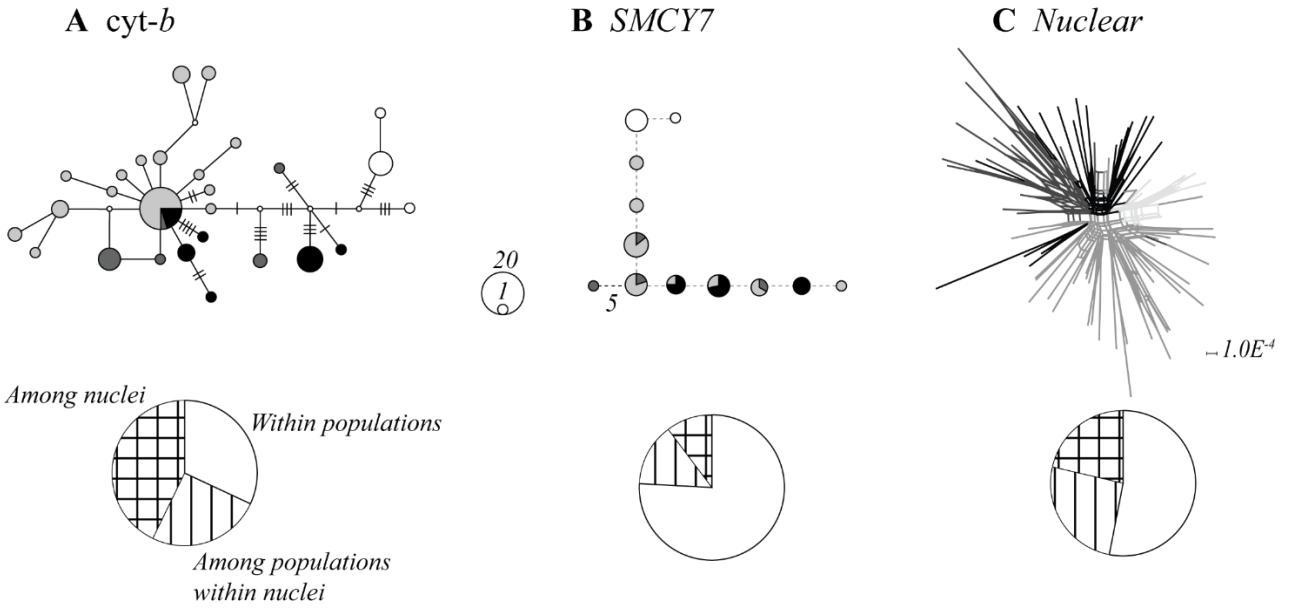


Figure S5. Median-joining haplotype networks for the cyt-*b* gene [**A**, *Live trapping* (*LT*), 79 samples] and *SMCY7* locus (**B**, 39 males), and NeighborNet for the remaining concatenated nuclear loci (**C**, 79 samples, exons: *BRCA1*, *IRBP* and *RAG1*; introns: *COPS7A*, *OSTA*, *PNPO* and *SLC38A7*; X-linked: *DBX5*, *DBX7*), and their respective AMOVA results (as pie-charts). For the cyt-*b* and *SMCY7* results (**A** and **B**), circles represent haplotypes and the area of the circle is proportional to the number of sequences; for cyt-*b*, white dots represent inferred haplotypes and each connecting line corresponds to one mutational step, with extra mutations shown by perpendicular strikes; for *SMCY7* the dashed black line represents a 5 bp indel and the dashed grey lines represent single microsatellite repeats. For the NeighborNet (**C**), each branch tip represents one sample and the distances between them is a proxy for genetic divergence. Shades of grey identify the four geographic nuclei: *Lusocarpetan* (grey), *pre-Pyrenean* (white in the networks and light grey in the NeighborNet network), *Montiberic* (dark grey) and *Betic* (black).

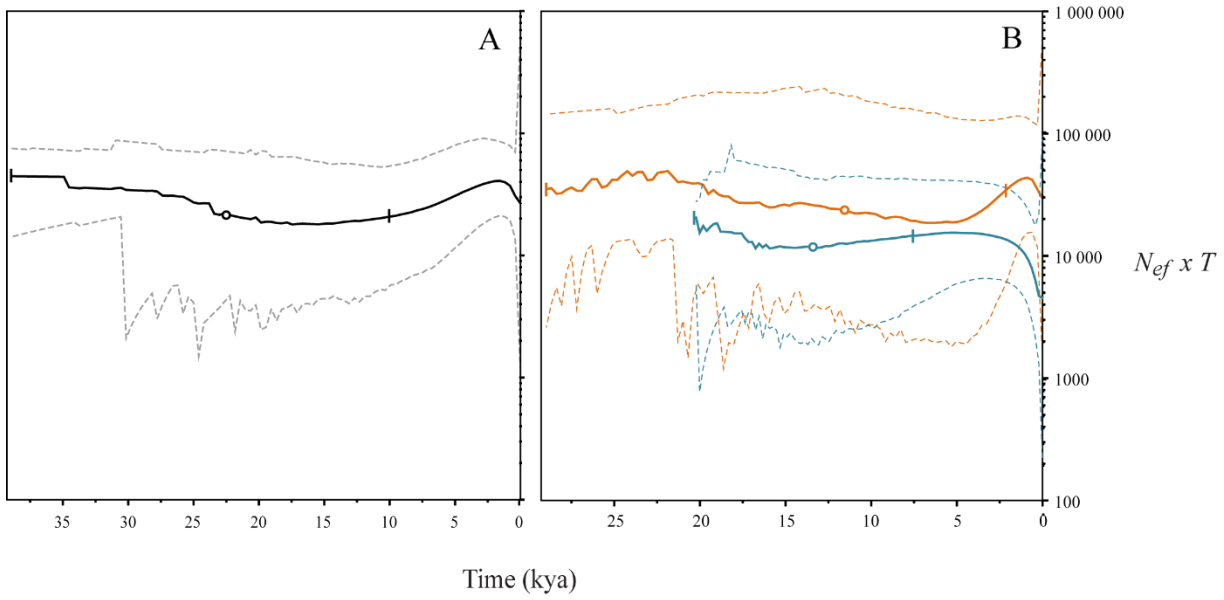


Figure S6. Bayesian skyline plots for *M. cabrerae* based on cyt-*b* data, for A) the whole species assuming a single panmictic population, and B) for the *west* (orange) and *east* (blue) lineages separately, assuming two divergent lineages relevant to a model of secondary contact (see text). The full lines represents the estimated median effective female population size (multiplied by mean generation time *T*) over time for each lineage, with 95% confidence interval (CI) in dashed lines; along the full lines, circles represent the average time to the most recent common ancestor and vertical dashes, the 95% HPD for that estimate; the X axis indicates time in kiloyears ago (kya); the Y axis indicates effective female population size (N_{ef} x T) on a logarithmic scale.

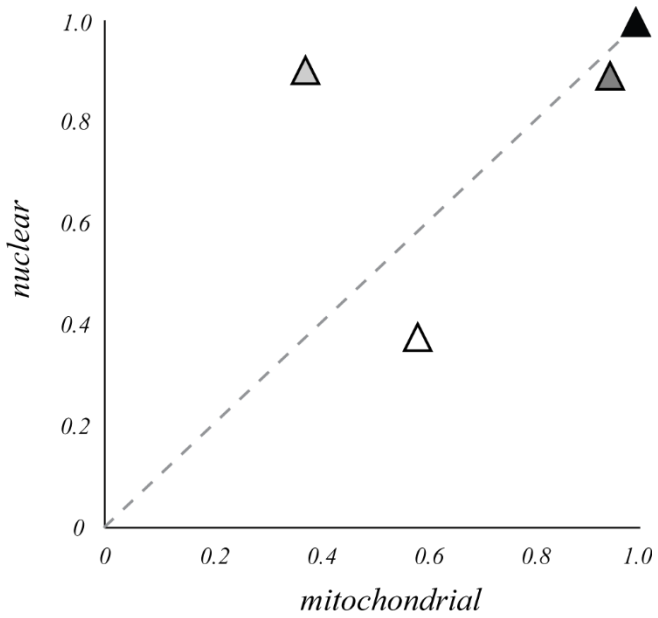


Figure S7. Locus-by-locus nucleotide diversity as a proportion of the highest population value, averaged over all loci, comparing nuclear and mitochondrial loci. Shades of grey within triangles represent the four geographic nuclei: *Lusocarpetan* (grey), *pre-Pyrenean* (white), *Montiberic* (dark grey) and *Betic* (black). The dashed line defines a 1:1 ratio.

CHAPTER IV.

CONSERVATION GENOMICS PLANNING

Paper IV. Barbosa S, Mestre F, White TA, Paupério J, Alves PC and Searle JB (*submitted*) **Integrative approaches to guide conservation decisions: using genomics to define conservation units and functional corridors**.

Paper V. Barbosa S, Paupério J, Mitchell S, Alves PC and Searle JB (*in prep*) **Non-invasive population genomics: applying genotyping-by-sequencing to small mammal conservation**.

Integrative approaches to guide conservation decisions: using genomics to define conservation units and functional corridors

Barbosa S, Mestre F, White TA, Paupério J, Alves PC and Searle JB

1. Abstract

Climate change and increasing habitat loss greatly impact species survival, requiring range shifts, phenotypic plasticity and/or evolutionary change for long term persistence, which may not readily occur unaided in threatened species. Therefore, remediation requires a detailed assessment of evolutionary factors, following the principles of 'evolutionary conservation'. Existing genetic diversity needs to be thoroughly evaluated and spatially mapped to define conservation units (*CUs*) in an evolutionary context, and the levels of gene flow between them assessed to determine functional connectivity. In this study, we propose a new approach to determine functional connectivity between *CUs* by including genetic diversity in the modelling while controlling for isolation-by-distance and phylogeographic history. We evaluate our approach on a near-threatened Iberian endemic rodent by analysing genotyping-by-sequencing (GBS) genomic data from 107 Cabrera voles (*Microtus cabrerae*), screening the entire species distribution to define subtypes of *CUs* and their connectivity: we defined five management units (*MUs*) which can be grouped into three evolutionarily significant units (*ESUs*) or two adaptive units (*AUs*). Thus, we demonstrate that three different subtypes of *CU* can be objectively defined using genomic data, and their characteristics and connectivity can inform conservation decision-making. With our newly proposed approach we established that connectivity is very limited in the south-east of the Cabrera vole distribution, which is an area of high habitat loss and where genetically very distinct populations are found. We argue that a multidisciplinary framework for *CU* definition is essential, and that this framework needs a strong evolutionary basis.

2. Introduction

Among the most pressing factors leading to biodiversity decline are habitat loss and climate change. In this context, species' survival can occur through a combination of three main processes: tracking their optimal habitat (range shifts), tolerating less optimal conditions within their physiological tolerance (phenotypic plasticity), and/or adapting to the new conditions by rapid evolutionary (genetic) change over generations (adaptation) (Mills 2013; Harrisson *et al.* 2014). Climate change induced range shifts will depend on a combination of factors relating to the speed of the climate change, the permeability of the landscape, and the species-specific dispersal abilities (Schloss *et al.* 2012). Species with lower dispersal abilities will be less able to track their preferred habitats, and will often be constrained to increasingly small habitat patches and exposed to new climatic conditions *in situ* (Allendorf *et al.* 2012). Adjusting to new stressors *in situ* can result in a first instance from phenotypic plasticity, but for long term persistence, evolutionary adaptations from standing genetic variation over generations are very important - responding to the environmental change and leading to environment-dependent phenotype expression (Ghalambor *et al.* 2007, 2015; Barrett & Schluter 2008; Harrisson *et al.* 2014). Given the importance of evolutionary factors there is a need for 'evolutionary conservation' – an approach to conservation that maximises genetic variability of species in order to potentiate beneficial associations with the environment, i.e. maximising the 'evolutionary potential' (Eizaguirre & Baltazar-Soares 2014). An evolutionary conservation approach is particularly important for threatened and patchily distributed species with limited dispersal ability, since decreasing effective population sizes (N_e) leads to loss of genetic diversity, increased genetic drift and inbreeding, and therefore less fit populations – a sequence known as the 'extinction vortex' (Frankham *et al.* 2010).

Given the importance of genetic diversity for natural populations, one of the main challenges currently faced in conservation planning is to maximise the species evolutionary potential by defining and being guided by appropriate conservation units (*CUs*) in an evolutionary context (Reed & Frankham 2003; Balkenhol *et al.* 2016). Following this approach the best established *CUs* are 'evolutionarily significant units' (*ESUs*) and although various definitions have been proposed over the years, there is common acceptance that *ESUs* should reflect major genetic diversity within species (Ryder 1986; Moritz 1994; Funk *et al.* 2012).

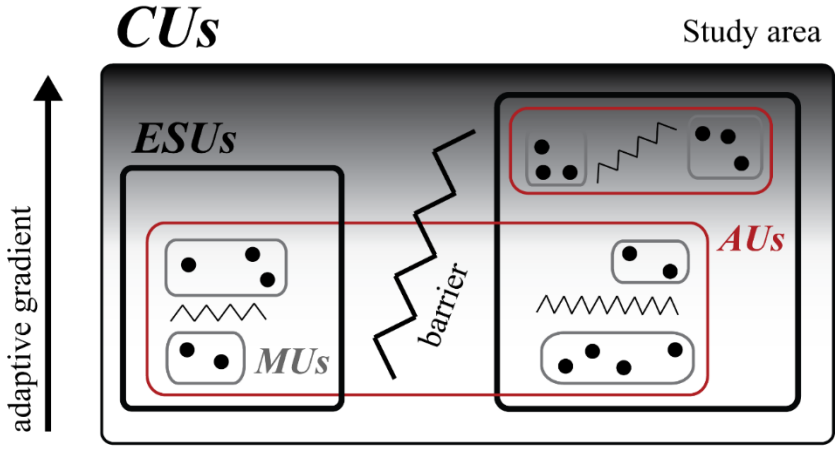


Figure 1. Scheme of the categorisation of conservation units (*CUs*) into different subtypes, as adapted from Funk *et al*. (2012). Black dots represent populations, zig-zag lines of different thickness represent barriers to dispersal of different strengths (e.g. rivers or mountain ranges of different magnitude), and the shading gradient represents an ecological gradient associated with adaptive variation. Rectangles of different colour represent the different subtypes of *CU*: evolutionarily significant units (*ESUs*, thick black line) – defined using all identified loci and representing groups of populations with different evolutionary histories; management units (*MUs*, grey line) – identified using neutral loci and representing groups of populations that maintain current gene flow; adaptive units (*AUs*, red line) – identified using outlier (putatively adaptive) loci and representing groups of *MUs* that have similar allele frequencies. In summary, on this scheme, *ESUs*, *AUs* and *MUs* are all *CUs*; an *MU* is always within one particular *ESU* and one particular *AU*; *ESUs* and *AUs* are two different ways of characterising groups of *MUs*; *ESUs* and *AUs* may or may not be equivalent (in this hypothetical example they are not).

With the arrival of genomic tools, the definition of *CUs* can now encompass greater detail of the species neutral and adaptive variation to provide a more accurate classification of two further subtypes of *CU*: management units (*MUs*) and adaptive units (*AUs*), respectively (Figure 1, Moritz 1994; Funk *et al.* 2012). The combination of such comprehensive genomic datasets with species ecology and the landscape, allows better predictions on the effects of habitat and climate change on species distribution and genetic variation (Steiner *et al.* 2013; Balkenhol *et al.* 2016). In a genomic context, Funk *et al.* (2012) proposed that *ESUs* should be identified using all available loci (e.g. SNPs), reflecting the overall evolutionary processes that have shaped genetic variation up to the present (i.e. biogeographic, demographic, adaptive, etc.); then, by identifying neutral and adaptive loci, we can respectively determine *MU*s, that reflect groups of populations within *ESUs* that are connected by current gene flow, and *AUs*, that reflect groups of *MUs* sharing adaptive variation (Figure 1). While *AUs* are important to promote different adaptive groupings of *MUs* and to maintain the current evolutionary potential of *ESUs* (Funk *et al.* 2012), *MUs* are the basis for short-term management of *ESUs*, more indicative of population monitoring and demographic studies (Moritz 1994). However,

there are some caveats in terms of the influence of species history on the currently observed genetic structure, important in the definition of *MUs*, given that such species history may represent events such as demographic responses to the most recent glacial cycles (phylogeographic history), rather than reflecting current connectivity – a type of 'time lag' (Landguth *et al.* 2010; Wang 2010). Especially at broader geographic scales, effects of phylogeography or isolation-by-distance (IBD) might become more difficult to parse from current gene-flow (Lowe & Allendorf 2010). Given the effects that the 'time lag' might have on current genetic variation, in this study we build upon the proposal of Funk *et al.* (2012), using an approach that controls for the species phylogeographic history and IBD in defining current gene flow between *MU*s, and hence better assessing their functional connectivity. Furthermore, while *AUs* can be identified from loci under divergent selection between populations (Funk *et al.* 2012), in cases where there are no functional studies relating genotypes to phenotypes or the environment, we argue that overall measures of 'adaptive' genetic diversity (based, for example, on F_{ST} outlier loci detection) should be used as a proxy for species 'adaptive potential' (Eizaguirre & Baltazar-Soares 2014), without targeting particular loci assumed to have associations to the environment. Thus, prioritisation of *CUs* should aim to detect and preserve distinctive genetic diversity, especially in small and isolated populations that are susceptible to extinction (Hampe & Jump 2011). Isolated *CUs* require high logistic and financial investment for their preservation, which highlights the importance of testing for connectivity/isolation between *CUs* for conservation planning. Usually, connectivity is evaluated by determining the species distribution model with the best fit to the observed genetic variation and to assume that areas with high species presence suitability between populations are the most likely dispersal corridors (Johnson & Omland 2004; Balkenhol *et al.* 2016). These connecting areas are thought of as areas of functional connectivity and are often targets for conservation efforts in order to avoid genetic isolation. Here we propose that genetic data can be included in the connectivity estimation model, to integrate the areas of climatic suitability (inferred from the species presence probability) with gene flow (a genetic connectivity matrix), providing realistic functional connectivity between *MUs*.

We evaluated our approach using data for an Iberian endemic rodent, that is 'near-threatened' according to the IUCN: the Cabrera vole, *Microtus cabrerae* (Fernandes *et al.* 2008a). This species is highly specialised to wet meadows and perennial herb communities near small streams and temporary ponds across the *thermo-* and *meso-mediterranean* bioclimates, and is typically distributed as metapopulations, often

subjected to high landscape dynamics (Rivas-Martínez 1981; Pita *et al.* 2006, 2007; Luque-Larena & López 2007). The main threats for the Cabrera vole relate to habitat reduction by husbandry activities that lead to smaller, more isolated and more susceptible populations restricted to the increasingly smaller habitat patches available (Pita *et al.* 2014). In this context, the protection of these sparsely distributed habitat patches that appear and disappear through both space and time is likely to reduce the susceptibility of more isolated metapopulations to extinction, preventing a generalised loss of genetic diversity (Frankham *et al.* 2010; Mills 2013). Previous phylogeographic studies identified two main mitochondrial lineages that likely diverged during and after the Last Glacial Maximum and that are now in secondary contact (Barbosa *et al.* 2017). However, the observed nuclear differentiation (based on the analysis of a few autosomal and sex-linked loci) was very low overall and inadequate to define *CUs*. In the present study, we use genomic tools to elucidate *CUs* and we use a multidisciplinary approach (based on genetic, ecological and climate suitability data) to evaluate population connectivity and determine potential corridors, helping to establish conservation priorities for the *MUs* of the Cabrera vole. We finally discuss these results in the light of currently defined conservation practices for the species.

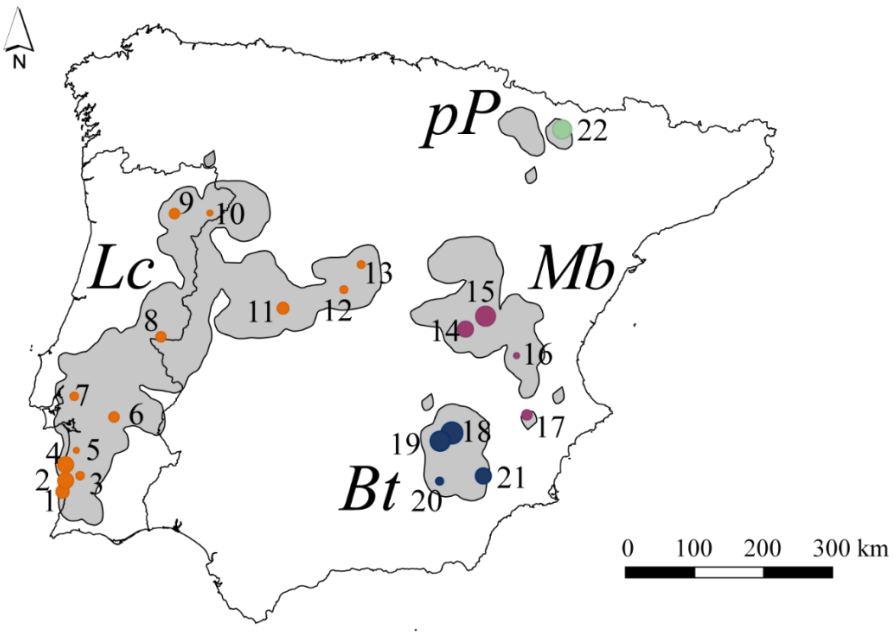


Figure 2. Map of the Iberian Peninsula with the distribution of the Cabrera vole (*Microtus cabrerae*) in grey *(Barbosa et al.* 2017); dots represent the sampled populations used for the GBS analyses and their size is proportional to the number of samples (Table S1, Supporting Information); colours of the dots represent their assignment to the four geographic nuclei as defined by Garrido-García *et al.* (2013) and Barbosa *et al.* (2017): *Lusocarpetan* (*Lc*, orange), *Montiberic* (*Mb*, purple), *Betic* (*Bt*, blue) and *pre-Pyrenean* (*pP*, green).

3. Material and methods

Overall population structure

In this study we performed genotyping-by-sequencing (GBS; Elshire *et al.* 2011) at the Cornell Genomic Diversity Facility on 107 Cabrera vole samples following the same protocol as in White *et al.* (2013). All samples were previously used in Barbosa *et al.* (2017), and are distributed in 22 localities comprising individuals within 10 km of each other (Figure 2 and Table S1). Filtering and SNP calling were performed using the UNEAK pipeline (Lu *et al.* 2013), where we defined a minimum base call of five reads, a maximum locus and individuals missing data of 20% and 30%, respectively. Post processing was performed to combine forward and reverse reads of the same loci and exclude potential paralogs following White *et al.* (2013). We then performed a principal coordinate analysis (PCoA), using the *dudi.pco* function of the *ade4* package in R (Dray & Dufour 2007), to visualise the distribution of genetic variance between the Cabrera vole individuals, in a model-free approach. To evaluate the species evolutionary history using genomic data we generated a maximum clade credibility tree following Lischer *et al.* (2014) using the entire dataset ('all loci'), performing a random sampling of the genotypes with the software RRHS (1000 replicates). Genetic distances were calculated with the Dnadist program of the PHYLIP package (Felsenstein 1993) using *p*-distances, and phylogenomic trees were inferred using a neighbor-joining approach implemented in the program Neighbor of PHYLIP. We summarised the trees by calculating a majority rule consensus tree with mean branch length in SplitsTree v4.13 (Huson & Bryant 2006). We finally performed a simple Mantel test (Mantel 1967) to test for isolation-by-distance (IBD) comparing all pairs of individuals and using the function *mantel* from the R package *ecodist* (Goslee & Urban 2007). We plotted Nei's genetic distances (Nei 1972, 1978), calculated using the function *dist.genet* from the *ade4* R package, against Euclidean geographic distances to visualise the trend. For this analysis we also evaluated the local density of points, measured using a two-dimensional kernel density estimation with the function *kde2d* from the *MASS* R package (Venables & Ripley 2002).

Delimitation of conservation units (CUs)

Evolutionarily significant units (ESUs): To detect *ESU*s we used the software STRUCTURE (Pritchard *et al.* 2000) to test for genetic subdivision using all loci. For that we calculated five replicates of each *K* (1 – 23) with 20 000 iterations each (10% burnin). From this analysis we determined the best-supported number of populations (K_{ESU}) as

determined by the Evanno method (Evanno *et al.* 2005). Gene diversity [$\hat{\pi}$, Nei (1987)] and pairwise population differentiation [F_{ST}, Weir & Cockerham (1984)] were calculated respectively within and between the defined *ESUs* in Arlequin v3.5 (Excoffier & Lischer 2010). The significance of the $\hat{\pi}$ comparisons was assessed using one-way ANOVA and the Tukey HSD post-hoc test (Abdi & Williams 2010). The significance of the F_{ST} values comparison was assessed through 100 bootstraps over loci using the function *diffCalc* implemented in the *diveRsity* R package (Keenan *et al.* 2013).

Management units (MUs): To identify *MUs* we first distinguished between neutral loci and outlier loci (potentially under selection). For that we performed an F_{ST} outlier analysis using BayeScan v2.01 (Foll 2012) to define 'neutral loci' and 'outlier loci' datasets. We set the prior odds for the neutral model to 1000 (recommended when there are thousands of markers), and Log_{10} values of the posterior odds (PO) larger than 1.5 were taken as 'very strong' evidence for selection (Foll 2012). We then analysed the 'neutral loci' in STRUCTURE (Pritchard *et al.* 2000) to test for genetic subdivision within each *ESU*. We assigned each individual to their respective K_{ESU} and then ran five replicates of each K (1 – 23) with 20 000 iterations each (10% burnin), initialising the run at the K_{ESU} (STARTPOPINFO option) to determine the best number of K_{MU} populations given the *ESU* structure previously obtained. Gene diversity ($\hat{\pi}$) and pairwise population differentiation (F_{ST}) were calculated respectively within and between the defined *MUs* in Arlequin v3.5 (Excoffier & Lischer 2010). For each of the populations (1-22) we additionally calculated the inbreeding coefficient (F_{IS}).

Prioritising MUs for conservation: For each *MU* we calculated a measure of population distinctiveness, the *Shapley metric* (SH), which is similar to the *evolutionary distinctiveness metric* used by the Zoological Society of London (www.edgeofexistence.org), with the difference that it can be estimated from an unrooted tree (Volkmann *et al.* 2014). For this we used the pairwise F_{ST} values between all pairs of *MU*s to build a NeighborNet network using SplitsTree v4.13 in order to obtain split weights for the calculations of SH. These measures, together with the expected degree of isolation obtained from the functional connectivity analysis (see below), were used to prioritise populations based on uniqueness and isolation.

Adaptive units (AUs): To test for differentiation of *AUs*, we evaluated adaptive variation by calculating an index of adaptive divergence, based on Bonin *et al.* (2007), where the allele frequencies of the 'outlier loci' are compared between *MUs*. For that we

calculated the average *MU* allele frequency of the alleles at each of the three 'outlier loci'. We then evaluated the correlation between different *MUs* for 'outlier loci' allele frequencies using the *rcorr* function, applying the Spearman test for data not normally distributed, from the *Hmisc* R package (Harrell Jr 2008). *MUs* with allele frequency correlations above 95% (and $p < 0.05$) were considered as the same *AU*. The tests comparing all groups were performed using one-way ANOVA and the Tukey HSD post-hoc test (Abdi & Williams 2010). Finally, we averaged across all three loci and depicted the assignment of each population in the study area to a gradient from fixation of allele A to fixation of allele B for a more intuitive visualisation of adaptive divergence.

Determining functional connectivity between CUs

Figure 3 represents a workflow of our analysis: first, we obtained an ecological niche model (ENM) for the Cabrera vole using the model described in Mestre *et al.* (2015), except that we estimated it at a 1x1 km scale (Figure 3A). We then used Circuitscape (Shah & McRae 2008) to generate a connectivity map between all samples based on that ENM (Figure 3B). For a broad-scale approach, circuit theory is relatively insensitive to choice of cell size and it has been shown to outperform conventional gene flow models by assuming that organisms are capable of using the entire landscape (McRae & Beier 2007). Also, circuit theory has been shown to outperform standard models of genetic differentiation because it incorporates effects of environmental/landscape resistance and range shape, and does not assume that gene flow is mediated by single, optimal pathways, but instead takes into account how alleles move over multiple pathways through intervening populations over many generations (McRae 2006; McRae & Beier 2007). In our approach, the ENM was not fitted to the genetic data but was only based on ecological data instead, to avoid redundancy when later incorporating the genetic connectivity matrix (Figures 3C and 3D). Secondly, we created a genetic connectivity matrix in R which converted the 'neutral' SNP genepop file into a distance matrix of percentage of allelic dissimilarity using the function *diss.dist*, part of the *poppr* R package (Kamvar *et al.* 2014, 2015). With this matrix we performed a distance based redundancy analysis (dbRDA) to control for the effect of IBD and phylogeographic history on the observed genetic divergence (Figure 3C). For this we used the *capscale* function from the *vegan* package in R (Oksanen *et al.* 2013), correlating the individuals' pairwise percent of allelic dissimilarity to a) their Euclidean geographic distance as a proxy for isolation-by-distance (*GEO*, Figure 3C), and b) to a binary distance matrix representing historical fragmentation among sampled

populations, as a proxy for phylogeographic history (*PHY*, Figure 3C). For this binary distance matrix we assigned a value of 1 to pairs of individuals separated during specific historical events and 0 to pairs that were not historically separated. In our case we divided the samples in the three *ESUs* identified in this work, which correspond to the three groups identified in previous work on the phylogeography of the Cabrera vole during the Younger Dryas (Barbosa *et al.* 2017).

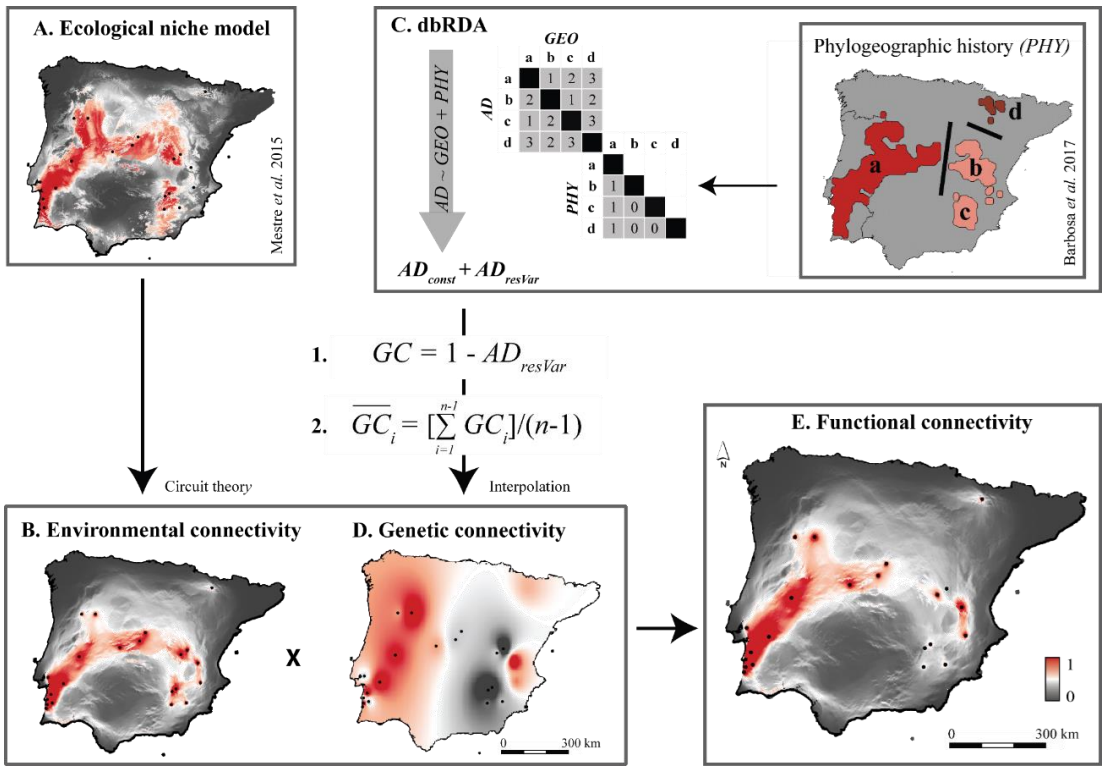


Figure 3. Incorporation of genetic differentiation between populations into ecological niche modelling and connectivity inference between all 22 populations of Cabrera vole used for GBS analysis (see Figure 2). **A. Ecological niche model:** BIOMOD probability of presence of the Cabrera vole in the Iberian Peninsula (Mestre *et al.* 2015). **B. Environmental connectivity:** Circuitscape connectivity analysis between the 22 populations studied using the ecological niche model as a conductance matrix. **C. dbRDA:** Distance based redundancy analysis testing the effect of predictor variables – geographic distance (*GEO*) and phylogeographic history (*PHY*), here shown with hypothetical values – in explaining the percent of allelic dissimilarity (*AD*) using the function *capscale* (*AD ~ GEO + PHY*) implemented in the *vegan* R package (Oksanen *et al.* 2013): AD_{const} – *AD* constrained by the predictor variables; AD_{resVar} – residual variation, i.e. not constrained by the predictor variables; phylogeographic history *(PHY)* – binary pairwise matrix with 0 and 1 for populations respectively connected and separated during given past time event, in our case the Younger Dryas (Barbosa *et al.* 2017), where lines represent barriers and population colours represent populations in the same group. **Equation 1.** Genetic connectivity (*GC*): inverse of the *AD* residual variation (1 – AD_{resVar}); **Equation 2.** Pairwise average connectivity ($\overline{GC_i}$) of each population to every other. **D. Genetic connectivity:** Interpolation of $\overline{GC_i}$ to the study area. **E. Functional connectivity:** Product of environmental and genetic connectivity maps. For maps A, B, D and E, grey and red represent low and high suitability/connectivity, respectively. For purposes of comparison, all maps present normalised values varying from 0 to 1.

The dbRDA provides the percentage of variation explained by each explanatory variable, the percentage of variation explained by the conditional variance (in the case of the conditional tests) and the unconstrained variation (i.e. not explained by the tests). We further calculated pseudo-*F* values, which is a measure of the significance of the overall analysis, along with its significance, using the *anova.cca* function from the *vegan* R package (Oksanen *et al.* 2013). Residual variation from this analysis was inverted to represent similarities (equation 1, Figure 3) and then assumed to represent contemporary gene flow. We then used this matrix to calculate the average pairwise percent of genetic connectivity between all pairs of populations (equation 2, Figure 3). This variation was then interpolated into the study area using standardised inverse distance weighting (IDW) in QGIS v2.4.0 (QGIS Development Team 2008) (Figure 3D). To create the functional connectivity map, we multiplied the environmental and genetic connectivity maps using a raster calculator in QGIS, where realised connectivity (genetic connectivity) then shapes the predicted connectivity (environmental connectivity), highlighting the connectivity of suitable environmental/landscape corridors that are maintaining gene flow between populations (Figure 3E).

Finally, for each population we calculated an average cost for individuals to reach other populations over time by first inverting the values of the matrix (1 – *Functional Connectivity*) and then calculating the least cost paths (LCP) between all pairs of populations using the function *costDistance* implemented in the *gdistance* R package (van Etten 2012). We averaged and normalised between 0 and 1 all pairwise LCP for each sampled locality to represent their genetic, geographical and ecological isolation.

4. Results

We performed GBS on 107 Cabrera voles from 22 populations, obtaining 3 341 unique single nucleotide polymorphisms (SNP) with less than 20% missing data for loci and 30% missing data for samples. The two principal coordinates of the PCoA analysis explained 13.9% or the total genetic variation and reflected very closely latitude/longitude, identifying three main clusters: *Lusocarpetan* (*Lc*) – 42 voles, *Montiberic* with *Betic* (*Mb*+*Bt*) – 55 voles and *pre-Pyrenean* (*pP*) – 10 voles (Figure S1A, Supplementary information). The main clusters of the neighbor-joining majority rule consensus tree correspond to geographic nuclei with high bootstrap support for most splits, including all the basal nodes, and also separate *Mb* from *Bt* and population 17 within *Mb* (Figure S1B, Supplementary information). The IBD analysis indicated a strong

correlation between genetic and geographic distance (r^2 = 0.40, p = 0.001) (Figure S2, Supplementary information).

Conservation units

The STRUCTURE analysis of 'all loci' (3 341 SNPs) identified three main populations (K_{ESU}) or *ESUs* (Figures 4A and S3A, Supplementary information): *ESU1* (42 voles, populations 1 – 13), *ESU2* (55 voles, populations 14 – 21); and *ESU3* (10 voles, population 22). Levels of gene diversity ($\hat{\pi}$) of *ESU2* were significantly higher than *ESU3* but not *ESU1* (Table 1). Levels of F_{ST} show *ESU1* and *ESU2* as the most similar and *ESU3* to be the most distinctive (Table 1). The STRUCTURE analysis also indicated substantial admixture between *ESU1* and *ESU2* (Figures 4A and S3A, Supplementary information).

Table 1. Genetic diversity and divergence values using 'all loci' dataset (3 341 SNPs) between the three *ESUs* identified for the Cabrera vole: *ESU1* (populations 1 – 13), *ESU2* (populations 14 – 21) and *ESU3* (population 22) (see Figure 4A). Diagonal shows average gene diversity ($\hat{\pi}$) over loci [and respective 95% confidence interval (CI)], and below diagonal shows average pairwise F_{ST} values (and respective 95% CI). All values differ significantly from zero (p < 0.05).

	ESU1	***ESU2***	***ESU3***
ESU1	0.209 (0.180-0.240)		
ESU2	0.080 (0.076-0.085)	0.259 (0.230-0.290)	
ESU3	0.245 (0.237-0.256)	0.196 (0.194-0.202)	0.148 (0.100-0.190)

Table 2. Genetic diversity and divergence values using the 'neutral loci' dataset (3 229 SNPs) between the five *MUs* identified for the Cabrera vole: *MU1* (populations 1 – 13), *MU2* (populations 14 – 16), *MU3* (populations 17 – 19), *MU4* (populations 20 – 21) and *MU5* (population 22) (see Figure 4B). Diagonal shows average gene diversity ($\hat{\pi}$) over loci [and respective 95% confidence interval (CI)], and below diagonal shows average pairwise F_{ST} values (and respective 95% CI). All values differ significantly from zero (p < 0.05).

	MU1	***MU2***	***MU3***	***MU4***	***MU5***
MU1	0.205 (0.180-0.230)				
MU2	0.121 (0.114-0.127)	0.240 (0.19-0.290)			
MU3	0.088 (0.084-0.093)	0.081 (0.077-0.087)	0.257 (0.210-0.300)		
MU4	0.163 (0.154-0.174)	0.156 (0.147-0.165)	0.113 (0.105-0.120)	0.185 (0.120-0.250)	
MU5	0.246 (0.244-0.248)	0.247 (0.244-0.249)	0.220 (0.219-0.221)	0.337 (0.329-0.345)	0.144 (0.100-0.190)

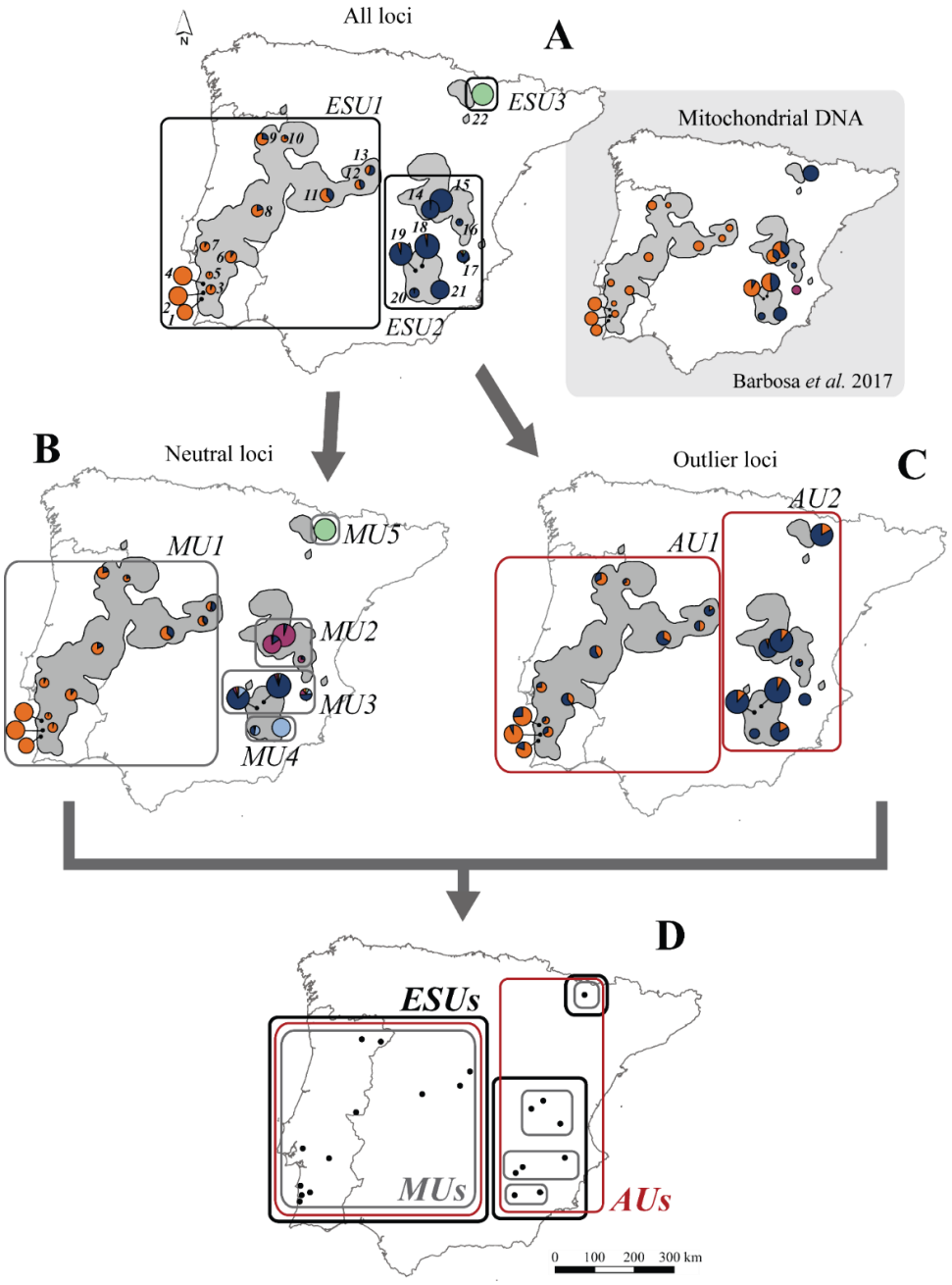


Figure 4. Definition of conservation units (*CUs*) for the Cabrera vole: **A**) grouping of populations into evolutionarily significant units (*ESUs*, where numbers refer to populations in Table S1) using 'all loci' (3 341 SNPs) and showing proportion of assignment to *ESU1* (orange), *ESU2* (dark blue) and *ESU3* (green); for reference, the distribution of the two main mitochondrial lineages identified in Barbosa *et al.* (2017) is shown, where *west* and *east* lineages are represented in orange and blue, respectively, and the purple circle represents a lineage that is not closely related to either; **B**) grouping of populations into management units (*MUs*) using the 'neutral loci' dataset (3 229 SNPs) showing proportion of assignment to *MU1* (orange), *MU2* (purple), *MU3* (dark blue), *MU4* (light blue) and *MU5* (green); **C**) grouping of populations into adaptive units (*AUs*) based on 'outlier loci' dataset (3 loci) showing average assignment to *AU1* (orange) and *AU2* (dark blue) alleles; and **D**) combined view of the Cabrera vole *CUs* following the same scheme as in Figure 1. For parts **A**, **B** and **C**, the pie area represents number of individuals analysed from 1 – 13 (e.g. population 10 and 18, respectively).

For the definition of *MUs* and *AUs* we distinguished between neutral and adaptive variation by identifying F_{ST} outliers, of which there were 14, three with high F_{ST} values (it is this more restrictive set of loci that we refer to as the 'outlier loci' from here on - Figure S4, Supplementary information). Using the 'neutral loci' (3 229 SNPs), the STRUCTURE analysis recovered five populations (K_{MU}) or *MUs* (Figure 4B): *MU1* (42 voles, populations 1 – 13), *MU2* (19 voles, populations 14 – 16), *MU3* (27 voles, populations 17 – 19), *MU4* (9 voles, populations 20 – 21) and *MU5* (10 voles, population 22). Considering $\hat{\pi}$, the observed values did not differ significantly ($p > 0.05$) between most *MUs*, except *MU2* and *MU3* from *MU5* (Table 2). F_{ST} was lowest between *MU2* and *MU3* and highest between *MU4* and *MU5*, although values between *MU5* and all other *MUs* are generally high (Table 2). F_{IS} values per population indicate that there are very different values of inbreeding within populations of each *MU*, no *MU* showing a particularly higher susceptibility to inbreeding than any other (Table S2, Supplementary information).

In order to prioritise *MUs*, we calculated the *Shapley metric* (SH) using F_{ST} distances between *MUs* to generate a NeighborNet network using the software SplitsTree (Figure S5, Supplementary information). SH values in increasing order from lower to greater contribution of *MUs* to the Cabrera vole genetic variation were obtained for *MU3* (0.03), *MU2* (0.06), *MU1* (0.08), *MU4* (0.10) and *MU5* (0.18).

Considering the five identified *MUs* we used the 'outlier loci' for the definition of the *AUs*. This resulted in a separation of *MU1* from the remainder *MUs*, which show a correlation above 95% in their 'outlier' allele frequencies, with $p < 0.05$ (Table S3, Supplementary information). This variation is depicted in Figure 4C where *AU1* has mostly *A* alleles and *AU2* has mostly *B* alleles, but there appears to be more *AU2* alleles in *AU1* than *vice versa*, similar to that observed with 'all loci' from *ESU2* into *ESU1*.

Functional connectivity between populations

The environmental connectivity map obtained from the ecological niche model (Figure 3A, Mestre *et al.* 2015) is shown in Figure 3B. For the genetic connectivity analysis, we performed a dbRDA controlling for geographic distance (*GEO*) and phylogeography (*PHY*) on the observed genetic variation. Tests for each predictor alone (marginal tests) showed significant effects of both *GEO* and *PHY* in explaining genetic variation (74.8% and 18.1%, respectively; Table S4, Supporting information). When we controlled for the other variable (conditional tests), *GEO* still significantly explained 56.7% of neutral genetic variation, but *PHY* explained none. Residual variation after

controlling for both variables (*GEO+PHY*) represented 25.2% of the total genetic diversity, which was used to calculate the percent of allelic dissimilarity (*AD*) per pair of populations (Figure 3C and Table S4, Supplementary information). Applying Equations 1 and 2 (Figure 3), we interpolated these values in the study area to obtain a genetic connectivity map (Figure 3D). The functional connectivity display that we ultimately obtained (Figure 3E) shows high connectivity in the southwest and low connectivity in the southeast, and the *pre-Pyrenees* (i.e. *ESU3*/*MU5*) also highly isolated (population 22, Table S1, Supplementary information). Connectivity between *MU2* and *MU3* appears to be maintained through populations 16 and 17, rather than through more central possibly unsampled populations between both *MUs*.

5. Discussion

Population structure and conservation units

The Cabrera vole genomic diversity appears to fit the geographic division into population nuclei described by Garrido-García *et al.* (2013) together with a strong pattern of isolation-by-distance (IBD) (Figures 2, 4A, and S1 and S2, Supplementary information). This may be indicative of the low dispersal ability of the species, which in terms of conservation may lead to increasingly isolated populations given increasing habitat degradation and climate change. In the absence of dispersal corridors, such isolated populations are susceptible to genetic drift and inbreeding (Allendorf *et al.* 2012; Balkenhol *et al.* 2016).

Evolutionarily significant units (*ESUs*): We identified three Evolutionarily Significant Units (*ESUs*): *ESU1* overlaps with the *Lusocarpetan* (*Lc*) geographic nucleus, *ESU2* with the *Montiberic* (*Mb*) and *Betic* (*Bt*), and *ESU3* with the *pre-Pyrenean* (*pP*) (Figures 2 and 4A). Previous phylogeographic studies also found an *east/west* split at the mitochondrial level that was further supported using phylogeographic modelling with autosomal and sex-linked loci, where the *west* was represented by *Lc* – *ESU1*, and the *east* was represented by *Mb*, *Bt* and *pP* – *ESU2* and *ESU3* (Barbosa *et al.* 2017). In the present study this split corresponds to the first population division (*K*=2) of the STRUCTURE analysis using both 'all loci' (Figure S6, Supporting information) and 'neutral loci' datasets (data not shown - equivalent results as for 'all loci'). However, in other analyses we show *ESU3* to be very divergent from the remaining groups, including *ESU2*: it has significantly lower $\hat{\pi}$, likely reflecting its degree of geographic isolation and possibly a declining trend in population size; it is also more divergent from *ESU1* than

ESU2, suggesting a closer common origin with the south-eastern Iberian populations (Table 1 and Figure S6, Supporting information: $K = 2$). In our study we also found evidence of admixture between *ESU1* and *ESU2* (orange and blue, respectively; Figures 4A and S3A, Supplementary information), with a gradient from a higher proportion of *ESU2* alleles in *ESU1* populations closer to *ESU2* (e.g. populations 12 and 13, Figure 2), to their absence in the southwestern-most part of *ESU1*, furthest away from *ESU2* (populations 1, 2 and 4, Figure 2). This is in clear contrast with the mitochondrial results obtained in Barbosa *et al.* (2017), where the authors described a *west* lineage (*ESU1*) presence in southern *east* lineage populations (*ESU2*). Already with the small number of nuclear loci examined in Barbosa *et al.* (2017) there were indications of mito-nuclear discordance, but it is now much clearer in our study given the extensive coverage of the nuclear genome. Thus, on the basis of the comparison of the mitochondrial data in Barbosa *et al.* (2017) and the nuclear data in the present study, it appears that the mito-nuclear discordance originates between *ESU1* and *ESU2* with nuclear *ESU2* alleles in the west and mitochondrial *ESU1* alleles in the east. Various processes can lead to mito-nuclear discordance, most commonly sex-biased dispersal, demographic and/or adaptive processes, and these can arise *in situ* or following isolation (primary vs. secondary contact, Toews & Brelsford 2012). This mito-nuclear discordance highlights the fact that there has been past connectivity between both western and eastern *ESUs*, which is important to take into account when delineating conservation strategies. For further discussion see below.

Management units (*MUs*): When we tested for further substructure using only 'neutral' variation, we found that only *ESU2* was subdivided in three *MUs* (*MU2*, *MU3* and *MU4*, Figures 4B and S3B, Supplementary information). This subdivision may be due to the high degree of isolation of the populations found in *ESU2* that, together with declining population sizes, has led to a divergence of allele frequencies between these populations (Allendorf *et al.* 2012; Mills 2013). Moreover, the admixture observed in *MU1* appears to originate from *MU3* rather than the neighbouring *MU2* (Figures 4B and S3B, Supplementary information). This might indicate a previous connection between *MU1* and *MU3* that could have been lost or interrupted by *MU2* in more recent times. Of all management units, *MU2* and *MU3* are the most diverse, but only *MU5* shows significantly lower gene diversity than other *MUs*. *MU5* is also the most divergent [especially from *MU4* (Table 2)], which means that this *MU*, as well as being the one with the lowest genetic diversity, appears to have very distinct genetic characteristics. This

fact is shown by the *Shapley metric* that holds *MU5* as the one with the highest contribution to the overall genetic diversity of the Cabrera vole (Figure S5, Supplementary information). This is again likely a reflection of the small size and long term isolation of this *CU* that should be considered important to the overall genetic constitution for this species. Additionally, there appears to be further substructure at smaller scales, especially within *MU3* (Figure S6, Supporting information: *K* = 8, populations 17 – 19), mirroring the lower overall functional connectivity between populations in this *MU* (Figure 3E and Table S2, Supporting information). All of this contributes to population isolation in the Cabrera vole, of concern given that its metapopulation structure relies heavily on dispersal to avoid inbreeding depression and getting caught in an 'extinction vortex' (Frankham *et al.* 2010; Allendorf *et al.* 2012; Balkenhol *et al.* 2016). Thus, the effects of a decreasing population trend and loss of habitat may lead to the loss of important genetic variation in many populations of the Cabrera vole (Pita *et al.* 2014).

Adaptive units (*AUs*): When examining adaptive variation – based on the 'outlier loci' - we found two *AUs* that divide the Cabrera vole populations in an east-west manner, in the same way previous mitochondrial results did (Barbosa *et al.* 2017), however with a contact zone more similar to that observed in the present study with all nuclear loci (Figures 4A and 4C). Although we believe our method for detection of *AUs* to be appropriate, caution is needed with our particular results with the Cabrera vole given the small number of outlier loci detected. The similarity of the result with *AUs* and all nuclear loci may be because the detected 'adaptive' variation is not in fact adaptive but instead reflects a random distribution of genetic variation that resulted in outlier loci; alternatively, the analysed 'outlier loci' may truly be revealing an adaptive introgression of *ESU2* into *ESU1*, which also influences the analysed 'neutral loci' (Figure 4B). There may be further substructure of the identified *AUs*, however due to the low number of outlier loci (and therefore alleles) there is not enough power to detect it. In any case, we wish to argue that our generalised approach is appropriate because locus specific associations of adaptive variation with the environment need to be very well substantiated to be used for conservation decision making – there is a high potential for erroneous associations that may lead to incorrect prioritisation of populations (Balkenhol *et al.* 2016). Thus, in the majority of cases a general maximisation of 'adaptive' genetic diversity should be preferred over choosing specific adaptive loci as a means to increase the species overall adaptive potential.

Mito-nuclear discordance

Altogether, the GBS data support the division of the Cabrera voles into three *ESUs*, two of which (*ESU1* and *ESU2*) have had extensive secondary contact after a period of isolation. However there is a decay of nuclear *ESU2* alleles from east to west but a decay of mitochondrial *ESU1* alleles from west to east (Figure 4A). Mito-nuclear discordances arise from various processes: incomplete lineage sorting, sex-biased dispersal, mitochondrial DNA introgression, demographic disparities or contact-zone movement (Toews & Brelsford 2012). Incomplete lineage sorting tends to result in unstructured variation, at odds with the high level of nuclear and mitochondrial structuring observed here (Figure 4); sex-biased dispersal can be associated with monogamous species such as the Cabrera vole, but ecological and genetic studies in this species have shown little to no evidence of such a phenomenon (Lawson Handley & Perrin 2007; Pita *et al.* 2014; Barbosa *et al.* 2017); mitochondrial DNA introgression could be a viable hypothesis given that there is extensive mitochondrial admixture from the *west* into the *east* lineage not found for the nuclear data, however previous studies showed that the whole mitochondrion is under purifying selection, so there is no particular advantage for *ESU1* mitochondria to be expanding (Barbosa *et al.* 2017); finally, it is possible that the secondary contact zone between both lineages was initially located within what is now *ESU1* and as this unit began to expand to the east, it displaced *ESU2* haplotypes from western Iberia. However, at the same time, the mitochondria seem to have moved far beyond the current nuclear contact zone as *west* lineage mitochondrial haplotypes are widely represented in *ESU2* (Figure 4). This leads once more to the possibility of mitochondrial DNA introgression and thus further studies on the mitochondrial variation of the species are needed to establish the actual processes behind this discordance.

Functional connectivity

The dbRDA results indicate a substantial influence of geography (74.8%) on the genetic variation we found in the Cabrera vole, supporting the finding that isolation-by-distance (IBD) is a strong determinant of the neutral genetic variation (Table S4 and Figure S2, Supplementary information). The phylogeographic history of the Cabrera vole appears to be tightly linked with its demographic history, given that there is conditional genetic variation explained by phylogeography within that explained by geography (18.1%). However, the effect of phylogeography when controlling for geography was not significant. The residual genetic variation from the dbRDA analysis (25.2%), when combined with the environmental connectivity model (Figure 3B), shows lower

connectivity between the east and the west than would be expected given the ecological niche model alone (Figure 3A and 3E). It also shows a highly connected western Iberia and a patchier eastern Iberia, the latter especially prominent in both the extreme north and south of the species distribution. Thus, in the east the *pre-Pyrenean CU* (*ESU3*/*MU5*) again appears highly disconnected from the remaining populations, as is also the case for most populations from *MU3* and *MU4* (Figures 3E and 4, and Table S2, Supporting information). Interestingly, the connectivity between *MU2* and *MU3* appears to be via populations 16 and 17, with an inferred corridor between these populations. *MU1*, which is the largest *CU*, has higher connectivity between its populations, possibly reflecting a more homogenous landscape than the east of the Iberian Peninsula (Vera 2004). Nonetheless, there is variation in connectivity within this *CU*, being higher in the southwest than the northeast.

With climate change it is expected that the rear edge populations that are older and often more genetically diverse will either track the new environmental conditions or disappear (Hampe & Petit 2004). Thus there may be lineage extinctions if individuals are not able to track their preferred habitat, or mixing of these ancient lineages if they are able to move from their previous locations, possibly explaining the admixture levels observed in some of the Cabrera vole populations, e.g. *MU2*/*MU3* and *MU3*/*MU4* (Figure 4B). Some *CUs* seem to have been connected in the past, and appear to continue to be connected – such as *MU2* and part of *MU3*; while others seem to have been connected and have recently lost that connectivity or show greatly reduced gene flow – e.g. *MU1*/*MU3* and *MU3*/*MU4* [also supported by Barbosa *et al.* (2017)]; and finally, some have long been isolated from the main populations, like *MU5* from all others.

The new framework we have developed in this study to infer functional connectivity shows the importance of considering both demographic and phylogeographic processes when aiming to determine current gene flow. With this approach we were able to control for the effects of IBD and phylogeographic history on the current patterns of connectivity, and determine which populations have higher conservation priority given not only their contribution to the species genetic diversity (Figure S5, Supporting information), but also given their overall isolation (Figure 3E). Determining the level of genetic/geographic isolation of a given *MU* is very important in terms of conservation because it provides guidelines for effort allocation, which is especially important in a crisis discipline with often limited financial resources. Moreover, with climate change, these evolutionary conservation parameters are essential to estimate in our efforts to allow species to track their optimal habitat or survive the new conditions *in situ*.

Conservation guidelines

The overall results suggest that varying conservation measures are needed to manage the different *CUs* of the Cabrera vole. Firstly, it is evident that the *pre-Pyrenean CU* (*ESU3*/*MU5*, Figure 4) is of very high conservation priority. This *CU* harbours genetic diversity not found anywhere else in the species distribution [Figures 4 and S1, S3 and S5, Supplementary information; see also Barbosa *et al.* (2017)] and it shows a high degree of isolation from the more central populations (Figure 3E and Table S2). At the opposite end of the spectrum is *ESU1*/*MU1* (Figure 4): this *CU* is very widespread with similar levels of genetic diversity throughout. There is, however, some admixture present in the central Iberian mountains, down to the centre/south of Portugal, which might indicate past connectivity that has been lost, given the absence of a defined corridor from any of the *MUs* in *ESU2* to *ESU1* (Figure 3E). As a conservation priority, we suggest that the connectivity between *ESU1* and *ESU2* should be promoted through the maintenance of field verges in the inferred corridors associated to extensive agriculture, given that the species has a natural capacity to colonise these habitat patches in a stepping-stone manner (Mira *et al.* 2008; Pita *et al.* 2013). Within *ESU2* the preservation of rear end populations is important, given that these possibly represent climate relicts associated with different ecological regimes than those found in the rest of the populations, thus having distinct adaptations and high genetic diversity (Hampe & Jump 2011; Barbosa *et al.* 2017). *MU4* is in fact the second largest contributor to Cabrera vole genetic diversity (Figure S5, Supplementary information), and considering its level of functional isolation (Figure 3E), it is highly likely that this variation will be lost due to habitat conversion into farmland and increasing aridification. This is because, especially in this area of the species distribution, there have been many recent extinctions of previously existing populations due to habitat loss associated with overgrazing and agricultural expansion (Fernández-Salvador 2007; Garrido-García *et al.* 2013).

In sum, all of the Cabrera vole *CUs* are unique and represent different and meaningful contributions to the species genetic diversity. In Spain, where most of the species genetic diversity is found, there should be a higher concern in maintaining and increasing the available habitat for the species, especially in the southern *Betic* (*MU4*) and the northern *pre-Pyrenean* (*MU5*) nuclei, which are the most threatened at this time. We further propose that the conservation of the *Lusocarpetan* geographic nucleus (*MU1*) should result from a combined effort of both Portugal and Spain, given that, although this is the most widespread and homogenous *CU*, it still encompasses many regional clusters of genetic diversity, some apparently expanding (in the northeast of Portugal) others

apparently declining (in the southwest of Portugal and central Spain), and maybe of importance for the future of this species in a context of human induced habitat changes.

Acknowledgements

We would like to thank N. Guiomar (ICAAM) for his help setting up the landscape analyses, and R. Pita, A. Mira and P. Beja for their helpful comments to the manuscript.

This work was partially supported by Programa Operacional Potencial Humano – Quadro de Referência Estratégico Nacional funds from the European Social funds and Portuguese Ministério da Educação e Ciência through Fundação para a Ciência e Tecnologia (FCT SFRH/BD/77726/2011 PhD grant to S. Barbosa), by the European Regional Development Fund (ERDF) through COMPETE program and Portuguese national funds through FCT (PTDC/BIA-BEC/105110/2008, PTDC/AAG-MAA/3227/2012 and PTDC/BIA-BIC/6582/2014 research projects).

6. References

Abdi H, Williams LJ (2010) Tukey's Honestly Significant Difference (HSD) Test. In: *Encyclopedia of Research Design* (eds Salkind NJ, Dougherty DM, Frey B), pp. 583–585. Sage, Thousand Oaks, CA, USA.

Allendorf FW, Luikart G, Aitken SN (2012) *Conservation and the Genetics of Populations*. Wiley-Blackwell, Oxford, UK.

Balkenhol N, Cushman SA, Storfer AT, Waits LP (2016) *Landscape Genetics: Concepts, Methods and Application*. Wiley-Blackwell, Oxford, UK.

Barbosa S, Paupério J, Herman JS *et al.* (2017) Endemic species may have complex histories: within-refugium phylogeography of an endangered Iberian vole. *Molecular Ecology*, **26**, 951–967.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**, 38–44.

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.

Dray S, Dufour A-B (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Eizaguirre C, Baltazar-Soares M (2014) Evolutionary conservation - evaluating the adaptive potential of species. *Evolutionary Applications*, **7**, 963–967.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, e19379.

van Etten J (2012) R package gdistance: distances and routes on geographical grids. *Journal of Statistical Software*.

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

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

Felsenstein J (1993) PHYLIP (phylogeny inference package). *Department of Genetics, University of Washington, Seattle*.

Fernandes M, Pita R, Mira APP (2008) *Microtus cabrerae*. *The IUCN Red List of Threatened Species*.

Fernández-Salvador R (2007) *Microtus cabrerae*. In: *Atlas y Libro Rojo de los Mamíferos Terrestres de España* (eds Palomo LJ, Gisbert J, Blanco JC), pp. 429–433. Dirección General para la Biodiversidad-SECEM-SECEMU, Madrid.

Foll M (2012) BayeScan v2.1 user manual. *Ecology*, **20**, 1450–1462.

Frankham R, Ballou JD, Briscoe DA (2010) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, **27**, 489–496.

Garrido-García JA, Rosário IT, Gisbert J *et al.* (2013) Revisión a nivel ibérico de la distribución del topillo de Cabrera o iberón, *Iberomys cabrerae* (Thomas, 1906). *Galemys*, **25**, 35–49.

Ghalambor CK, Hoke KL, Ruell EW *et al.* (2015) Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, **525**, 372–375.

Ghalambor CK, McKay JK, Carroll SP, Reznick DN (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new

environments. *Functional Ecology*, **21**, 394–407.

Goslee SC, Urban DL (2007) The *ecodist* package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

Hampe A, Jump AS (2011) Climate relicts: past, present, future. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 313–333.

Hampe A, Petit RJ (2004) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, **8**, 461–467.

Harrell Jr FE (2008) Hmisc: harrell miscellaneous. *R package version 3.2*.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, **7**, 1008–1025.

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

Johnson JB, Omland KS (2004) Model selection in ecology and evolution. *Trends in Ecology and Evolution*, **19**, 101–108.

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, **6**, 1–10.

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

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, **4**, 782–788.

Landguth EL, Cushman SA, Schwartz MK *et al.* (2010) Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, **19**, 4179–4191.

Lawson Handley LJ, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology*, **16**, 1559–1578.

Lischer HEL, Excoffier L, Heckel G (2014) Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of *Microtus* voles. *Molecular Biology and Evolution*, **31**, 817–831.

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity?

Molecular Ecology, **19**, 3038–3051.

Lu F, Lipka AE, Glaubitz J *et al.* (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, **9**, e1003215.

Luque-Larena JJ, López P (2007) Microhabitat use by wild-ranging Cabrera voles *Microtus cabrerae* as revealed by live trapping. *European Journal of Wildlife Research*, **53**, 221–225.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.

McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 19885–19890.

Mestre F, Pita R, Paupério J *et al.* (2015) Combining distribution modelling and non-invasive genetics to improve range shift forecasting. *Ecological Modelling*, **297**, 171–179.

Mills LS (2013) *Conservation of Wildlife Populations: Demography, Genetics, and Management*. Wiley-Blackwell, Oxford, UK.

Mira APP, Marques CC, Santos SM, Rosário IT, Mathias ML (2008) Environmental determinants of the distribution of the Cabrera vole (*Microtus cabrerae*) in Portugal: implications for conservation. *Mammalian Biology*, **73**, 102–110.

Moritz C (1994) Defining “evolutionarily significant units” for conservation. *Trends in Ecology and Evolution*, **9**, 373–375.

Nei M (1972) Genetic distance between populations. *The American Naturalist*, **106**, 283–292.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY, USA.

Oksanen J, Blanchet FG, Kindt R *et al.* (2013) Package “vegan.” *Community Ecology Package*, **2**.

Pita R, Beja P, Mira APP (2007) Spatial population structure of the Cabrera vole in Mediterranean farmland: the relative role of patch and matrix effects. *Biological Conservation*, **134**, 383–392.

Pita R, Mira A, Beja P (2006) Conserving the Cabrera vole, *Microtus cabrerae*, in intensively used Mediterranean landscapes. *Agriculture, Ecosystems and Environment*, **115**, 1–5.

Pita R, Mira APP, Beja P (2013) Influence of land mosaic composition and structure on patchy populations: the case of the water vole (*Arvicola sapidus*) in Mediterranean farmland. *PLoS One*, **8**, e69976.

Pita R, Mira APP, Beja P (2014) *Microtus cabrerae* (Rodentia: Cricetidae). *Mammalian Species*, **912**, 48–70.

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

QGIS Development Team (2008) Quantum GIS Geographic Information System. Open source Geospatial Foundation project.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology*, **17**, 230–237.

Rivas-Martínez S (1981) Les étages bioclimatiques de la végétation de la Péninsule Ibérique. *Anales del Jardín Botánico de Madrid*, **37**, 251–268.

Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology and Evolution*, **1**, 9–10.

Schloss CA, Nuñez TA, Lawler JJ (2012) Dispersal will limit ability of mammals to track climate change in the Western Hemisphere. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 8606–8611.

Shah VB, McRae BH (2008) Circuitscape: a tool for landscape ecology. *Proceedings of the 7th Python in Science Conference*, 62–66.

Steiner CC, Putnam AS, Hoeck PEA, Ryder OA (2013) Conservation genomics of threatened animal species. *Annual Review of Animal Biosciences*, **1**, 261–281.

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

Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*. Springer, New York.

Vera JA (2004) *Geología de España*. Instituto Geológico y Minero de España, Madrid.

Volkmann L, Martyn I, Moulton V, Spillner A, Mooers AO (2014) Prioritizing populations for conservation using phylogenetic networks. *PLoS One*, **9**, e88945.

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology*, **19**, 2605–2608.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

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

Author contributions

This study is a part of S.B.'s PhD thesis work on the Cabrera vole, *Microtus cabrerae*, which she is developing under the supervision of P.C.A. and J.B.S. S.B. conducted genetic laboratory work and analyses at Cornell University and CIBIO. F.M. contributed to the ecological modelling analyses, T.A.W. with bioinformatic support and J.P. to the data analysis. All contributed to the writing of the manuscript.

Data accessibility

Illumina reads produced in this study are deposited in NCBI Sequence Read Archive (XXX). Assembled GBS alignments, tree files and other data are deposited in Dryad (XXX).

7. Supporting information

Table S1. Details of the localities and number of Cabrera voles sampled per locality. Individuals are grouped into localities when sampled within a 10 km radius and coordinates represent the centroid of those samples.

Map code	Latitude	Longitude	N_{ind}	Locality	District/Province	Country
1	37.609	-8.794	5	Cabo Sardão	Beja	Portugal
2	37.753	-8.755	7	Vila Nova de Milfontes	Beja	Portugal
3	37.828	-8.523	2	Bicos	Beja	Portugal
4	37.959	-8.771	7	Sines	Setúbal	Portugal
5	38.098	-8.561	1	Grândola	Setúbal	Portugal
6	38.632	-8.045	3	Évora	Évora	Portugal
7	38.870	-8.725	2	Benavente	Santarém	Portugal
8	39.763	-7.399	3	Monforte da Beira	Castelo Branco	Portugal
9	41.343	-6.577	3	Mogadouro	Bragança	Portugal
10	41.304	-7.183	1	Mirandela	Bragança	Portugal
11	40.114	-5.256	4	Candeleda	Ávila, Castilla y Leon	Spain
12	40.349	-4.350	2	Navas del Rey	Madrid	Spain
13	40.622	-3.877	2	Becerril de la Sierra	Madrid	Spain
14	39.861	-2.158	7	Las Valeras	Cuenca, Castilla-la-Mancha	Spain
15	40.026	-1.817	11	Cañada del Hoyo	Cuenca, Castilla-la-Mancha	Spain
16	39.508	-1.291	1	Utiel	Valencia	Spain
17	38.730	-1.184	3	Yecla	Murcia	Spain
18	38.512	-2.401	10	Riopar	Albacete, Castilla-la-Mancha	Spain
			3	Paterna de la Madera	Albacete, Castilla-la-Mancha	Spain
19	38.407	-2.597	11	Siles	Jaén, Andalucia	Spain
20	37.895	-2.649	2	Canal de San Clemente	Granada, Andalucia	Spain
21	37.950	-1.888	7	Caravaca de la Cruz	Murcia	Spain
22	42.438	-0.466	10	Sabiñanigo	Huesca, Aragón	Spain
Total			107			

N_{ind} - Number of individuals sampled. Map code indicates the populations in Figure 2.

Table S2. Inbreeding coefficient (F_{IS}) for each population of Cabrera vole sampled and functional cost value representing normalised average genetic, geographical and ecological isolation of each sampled locality calculated using least cost paths (see text). Significant values of F_{IS} ($p < 0.05$) are in bold.

Map code	F_{IS}	Functional cost
1	**0.250**	0.029
2	**0.126**	0.000
3	0.028	0.365
4	**0.161**	0.052
5	0.000	0.123
6	0.237	0.418
7	0.047	0.421
8	0.021	0.529
9	0.388	0.584
10	0.000	0.587
11	**0.548**	0.655
12	0.512	0.666
13	0.320	0.682
14	**0.136**	0.754
15	**0.142**	0.773
16	0.000	0.730
17	0.164	0.835
18	0.076	0.940
19	**0.092**	0.974
20	0.442	1.000
21	**0.241**	0.978
22	**0.146**	0.961

Map code indicates the populations in Figure 2.

Table S3. Correlation of average allele frequency (%) of the three 'outlier' loci between *MUs* of the Cabrera vole identified in Figure 4B. Adaptive units (*AUs*) are represented by squares in thick lines, defined as those sharing more than 95% of allele frequency correlation ($p < 0.05$, bold).

	MU1	***MU2***	***MU3***	***MU4***	***MU5***
MU1	100				
MU2	-38	100			
MU3	-26	**97**	100		
MU4	-38	**100**	**97**	100	
MU5	-24	**98**	**96**	**98**	100

Table S4. Results of the dbRDA analysis for 'neutral loci' (3 229 SNPs), testing the percentage of genetic variation (given by pairwise distance matrices) among individuals explained by two predictor variables, geographic distance (*GEO*) and phylogeographic structure (*PHY*). The first row for each predictor is a marginal test, calculating the percentage of variance (%Var) and respective *F* value explained by that predictor. When significant ($p < 0.05$, bold), we performed conditional tests controlling for the listed variable on the predictor.

Predictor	Conditional	*F*	%Var	% Conditional	% Unconstrained
GEO		**2.7**	74.8	-	25.2
	PHY	**2.1**	56.7	18.1	25.2
PHY		**11.5**	18.1	-	81.9
	GEO	0.0	0.0	74.8	25.2
GEO + PHY		**2.7**	74.8	-	25.2

Predictors: For the *GEO* matrix, distances between individuals are given in km. For the *PHY* matrix, individuals from the same *ESU* have distance 0, and individuals from different *ESUs* have distance 1.

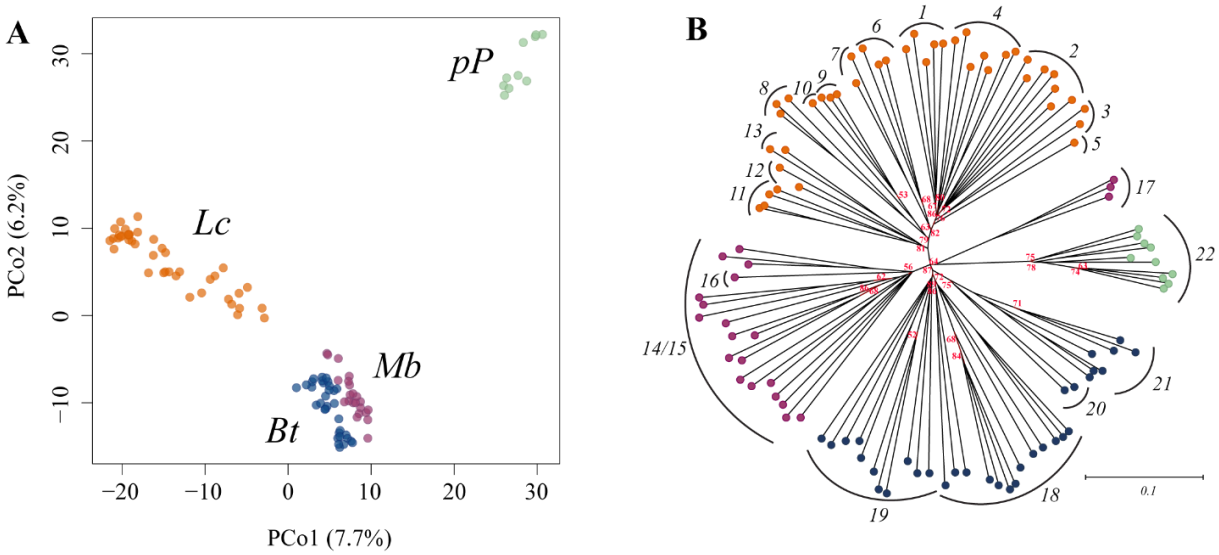


Figure S1. Clustering analysis using 107 Cabrera voles for the total SNP dataset (3 341 loci) (see text for further details). A) Principal coordinate analysis (PCoA) using the *dudi.pco* function implemented in the R package *ade4* 1.7.3 (Dray & Dufour 2007). B) Majority rule consensus tree with random sampling haplotypes obtained using the RRHS software (1000 replicates); branch numbers indicates bootstrap values below 90%, values are indicated on the respective branch in red. Dots in both analyses represent samples coloured by the geographic nucleus of origin: *Lusocarpetan* (*Lc*, orange), *Montiberic* (*Mb*, purple), *Betic* (*Bt*, blue) and *pre-Pyrenean* (*pP*, green).

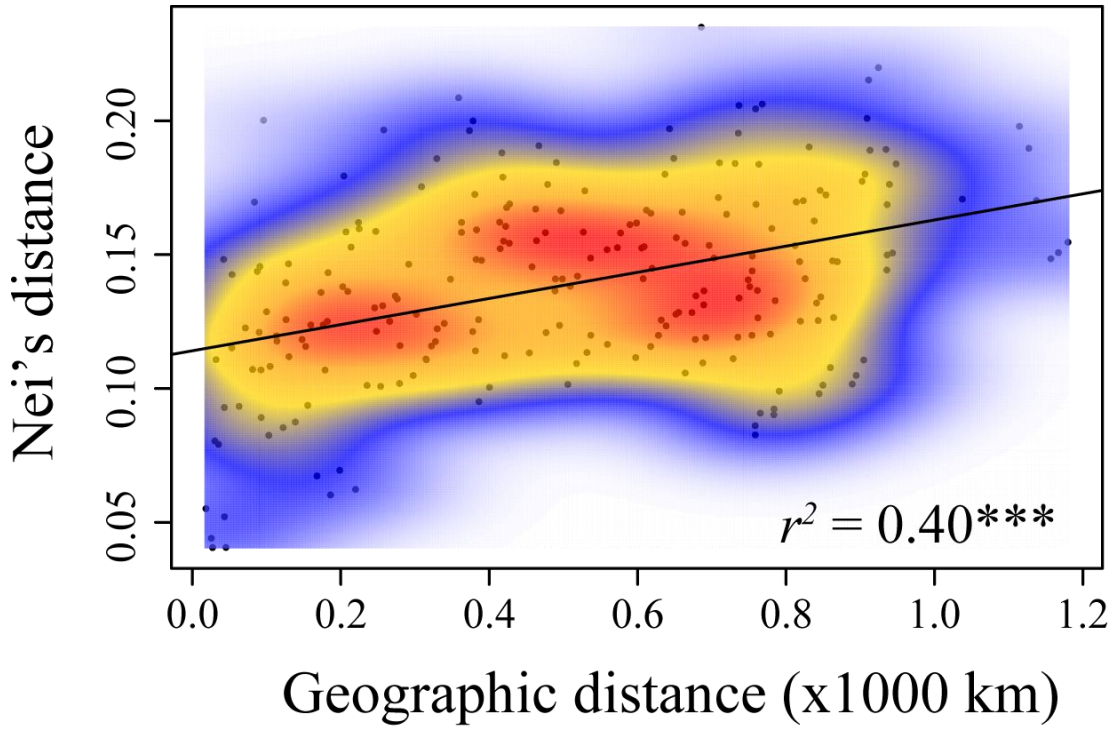


Figure S2. Isolation-by-distance plot between all Cabrera vole individuals. Nei's distances provide the genetic distance measure. Black dots represent pairwise comparisons, with fitted trend line; colours illustrate dot density from low (blue) to high (red). *** $p < 0.001$.

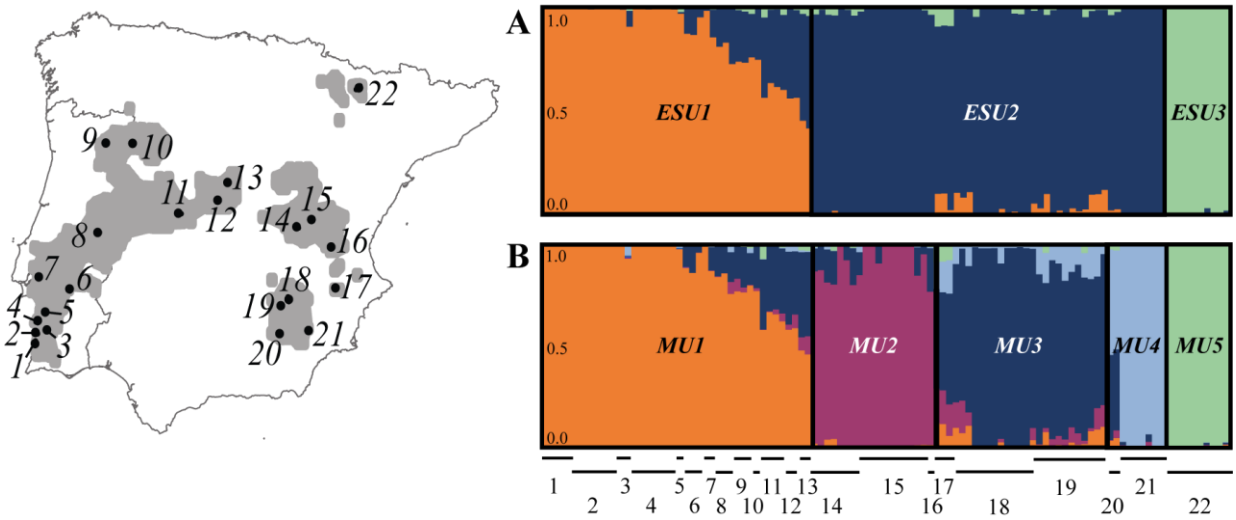


Figure S3. Population assignment cluster analysis for the Cabrera vole using the software STRUCTURE indicating the proportion of sample assignment to each of the inferred best *K* number of populations (colours) using A) 'all loci' (3 341 SNPs) to define the three evolutionary significant units (*ESUs*), and B) the 'neutral loci' dataset (3 229 SNPs) to define the five management units (*MUs*). Horizontal black bars and respective numbers refer to the sampled populations shown in the accompanying map and listed in Table S1.

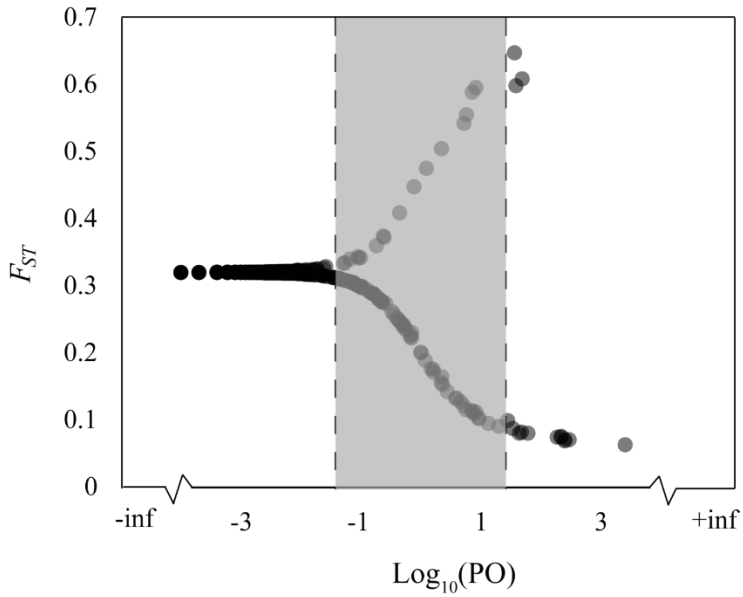


Figure S4. BayeScan F_{ST} outlier analysis using 3 341 SNP loci from 107 Cabrera voles, considering the 22 sampled populations. F_{ST} for each locus (dots) is plotted against the Log_{10} of the posterior odds (PO). The grey area corresponds to PO between -1.5 and 1.5, where there is low confidence in locus classification to either 'neutral loci' or 'outlier loci'. From the identified outliers (14 loci) we selected those with high F_{ST} for the 'outlier loci' dataset as they are assumed to represent markers under positive selection (3 loci); the 3 229 'neutral loci', non-outlier, are shown on the left of the grey area; locus density is depicted with increasing darkness proportional to the overlap of the dots.

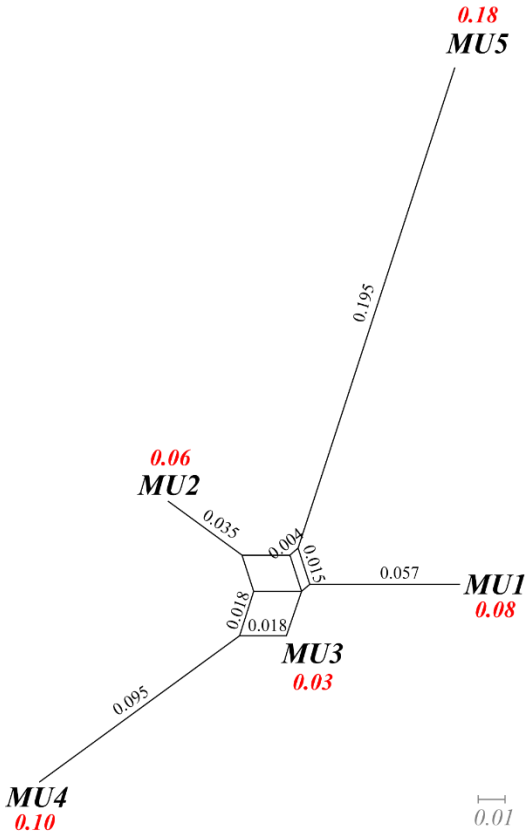


Figure S5. NeighborNet network of the five *MUs* of Cabrera vole identified in Figure 4B, calculated using pairwise F_{ST} values (Table 2). Numbers on branches represent split weights [i.e. phyletic distance (Fitch 1997)] and numbers in red around each *MU* refer to their respective *Shapley metric* value.

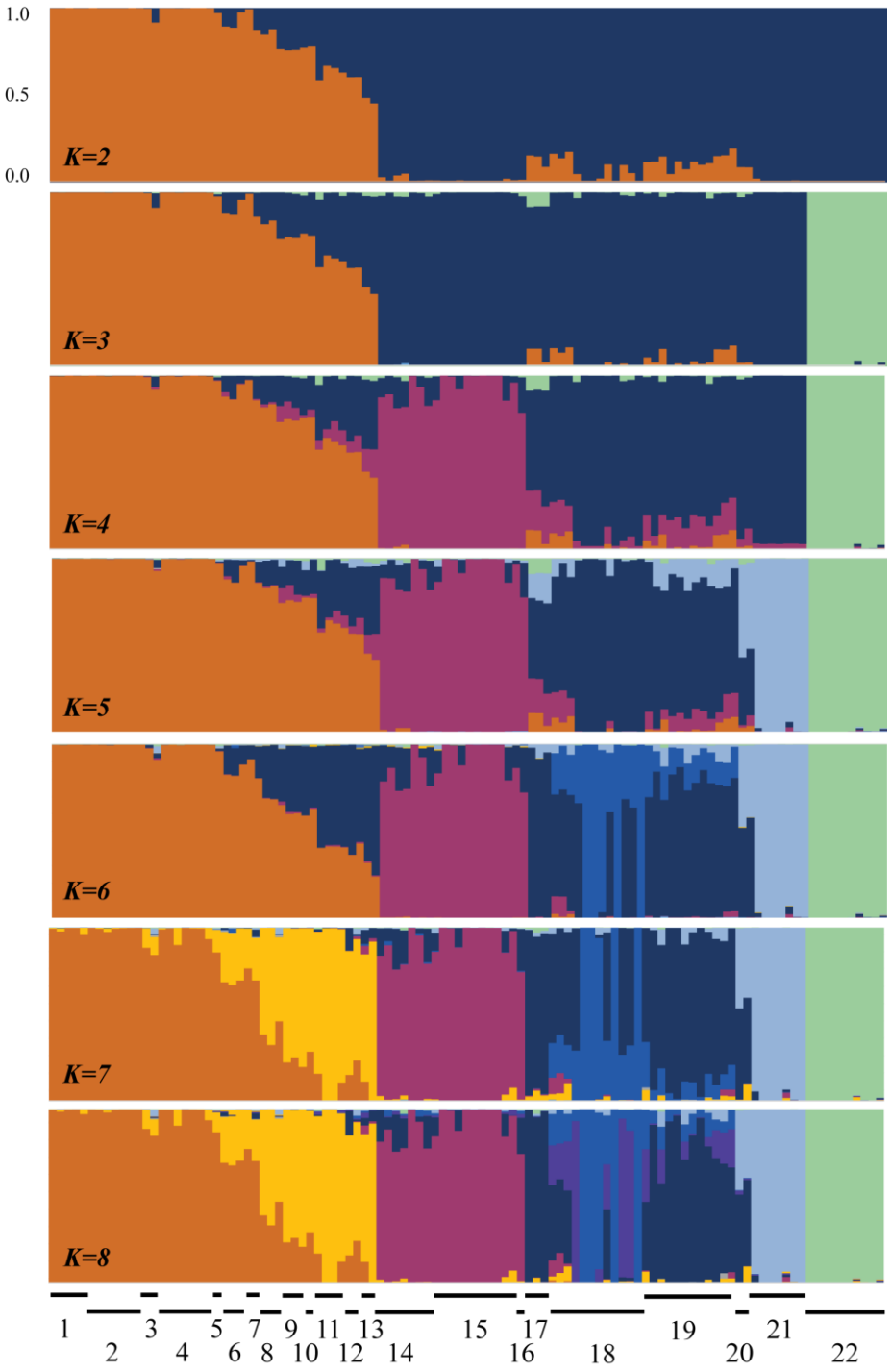


Figure S6. Population assignment clustering analysis for the Cabrera vole using the software STRUCTURE indicating the proportion of sample assignment to each of the inferred *K* number of populations (colours) using 'all loci' (3 341 SNPs) from *K*=2 to *K*=8. Bars and respective numbers refer to the populations in Figure 2.

Non-invasive population genomics: applying genotyping-by-sequencing to small mammal conservation

Barbosa S, Paupério J, Mitchell S, Alves PC and Searle JB

1. Abstract

Collection of genetic data from elusive or endangered species poses a challenge for many conservation studies. Non-invasive genetic sampling (NiGS) does not require animal capturing or handling, but provides genetic data useful to answer a variety of ecological or evolutionary questions relevant to conservation-orientated studies. The main drawback of using NiGS is the low DNA quantity and quality, which demands special precautions in the early stages of sample collection to the final data analyses. Current directions involving NiGS include the optimization of novel high throughput sequencing techniques and the development of analytical tools that account for the inherent problems of non-invasive DNA. In this study we tested a simple, inexpensive protocol using GBS (Genotyping-by-Sequencing) on small mammal faecal samples and bones from owl pellets, as two means of monitoring small mammals without using invasive (trapping) techniques. As anticipated, results show a high proportion of non-target DNA, with a maximum of 1.6% and 3.0% of reads belonging to the target species in faecal and bone DNA pools, respectively. However, we were able to recover the full panel of SNPs identified in previous genomic studies, with a low level of missing data per sample (4-6%). This novel approach allowed us to identify the geographic origin of both NiGS DNA pools (faecal and bones) when using a reference database previously obtained with tissue samples, although the genetic diversity of the NiGS DNA pools was low, for unknown reasons. Thus, we have shown that it is possible to obtain usable genomic data from NiGS DNA pools at low sequencing cost for non-model species, in this case using two viable methods for monitoring small mammals (faeces and bones from owl pellets). Our results open new opportunities in the application of genomic data for the conservation of elusive or endangered species using NiGS.

2. Introduction

Our understanding of the field biology of rare species has benefited greatly from the use of genetic information (DeSalle & Amato 2004). However, collecting genetic data from elusive or threatened species poses a challenge, which may be seen by wildlife managers and conservation practitioners as a barrier to its applicability (Taylor *et al.* 2017). Even so, over the past decade, numerous conservation-orientated studies have used non-invasive samples (e.g. faeces, hairs, feathers, skin, etc.) from endangered species to obtain DNA samples (Reed *et al.* 1997; Henry *et al.* 2011). Non-invasive genetic sampling (NiGS) does not require capturing or handling the animals, but provides genetic data for various applications that can have conservation relevance: species detection/distribution/abundance, phylogeography, population genetics, adaptation, social behaviour, hybridisation, predator diet, and forensics (Ogden 2011; Paupério *et al.* 2012; Barbosa *et al.* 2013, 2017; Bi *et al.* 2013; Mumma *et al.* 2015, 2016; Mestre *et al.* 2015; Bischof *et al.* 2016). The major drawback of using NiGS is the low quantity and quality of DNA obtained [e.g. Beja-Pereira *et al.* (2009)]. To overcome these problems, special precautions are necessary from sample collection to the laboratory bench and with data analyses, imposing severe quality control systems, which significantly increase the workload and costs per sample (Waits & Paetkau 2005; Beja-Pereira *et al.* 2009).

The probability of successful NiGS strategies has increased with current improvements in DNA sequencing, such as the optimization of high throughput sequencing (HTS) for difficult samples and the development of analytical tools that account for the inherent problems of extraction of non-invasive DNA extraction (Perry *et al.* 2010; Bohmann *et al.* 2014; Peñalba *et al.* 2014; Costa *et al.* 2017). Although it is possible to sequence whole genomes using degraded DNA samples, such studies rely mostly on museum specimens that hold higher amounts of DNA than those found in samples like faeces, and on the existence of a reference genome (Rowe *et al.* 2011). Specifically for non-invasive samples, reduction of genome complexity is desirable to increase the chances of amplifying homologous genomic regions in different samples. This may be accomplished by using probes or hybridisation capture protocols (Andrews *et al.* 2016; Lynn *et al.* 2016). However, such techniques are technically challenging and costly as they require *a priori* definition of target sequences using a reference genome from a closely related taxon, thus limiting opportunities for non-model species (Bi *et al.* 2013; Lynn *et al.* 2016). Moreover, the cost of analysing individuals using NiGS can become prohibitive given the low success of amplification of all genomic regions of interest. In these cases, it can be cost-effective to use restriction site associated next-

generation DNA sequencing on pools of DNA, as long as there is enough DNA and its degradation is not too extensive (Schlötterer *et al.* 2014; Graham *et al.* 2015).

A recent study by Beck & Semple (2015) analysed plant museum specimens with genotyping-by-sequencing (GBS), a restriction enzyme-based HTS technique (Elshire *et al.* 2011), to obtain informative single nucleotide polymorphisms (SNPs) from species with no previous genomic data. Additionally, an experimental study by Graham *et al.* (2015) has shown that restriction enzyme-based HTS methods are very robust to DNA degradation. These works gave credence to the application of restriction-based HTS to non-model species using small amounts of low quality DNA. In the present study, we applied GBS to DNA extracted from faecal samples and bones from owl pellets of a near-threatened rodent endemic to the Iberian Peninsula, the Cabrera vole (*Microtus cabrerae*). Faeces and owl pellets represent two of the most readily detected types of non-invasive sampling for small mammals in general, and for this species in particular (Barbosa *et al.* 2013, 2017). Like the plant species analysed previously (Beck & Semple 2015), the Cabrera vole does not have a sequenced reference genome, which is also typical for most small mammals, including endangered species. The goal of this study was to determine the possibility of gathering a substantial number of genome-wide SNPs from pooled DNA samples extracted from *M. cabrerae* faeces and skeletal material of *M. cabrerae* from owl pellets. Additionally, we wanted to test whether the SNPs obtained can properly be used for population assignment and to estimate genetic diversity, by comparing the DNA pools from NiGS to a reference collection of tissue samples representative of the same distribution areas as the NiGS samples. Given that genomic data will soon become the standard for conservation genetic assessments of current populations, we further discuss the application of NiGS for threatened small mammals, obviating the need for techniques which may potentially be more detrimental to individuals and populations, and which can be more labour intensive and costly (e.g. live-trapping).

3. Material and Methods

Preparation of pooled NiGS DNA samples and construction of GBS libraries

We used DNA extracted three years previously from 67 Cabrera vole bone samples obtained from barn owl (*Tyto alba*) regurgitated pellets collected over a wide geographic range in Portugal, and 8 freshly collected and extracted faecal samples collected in the south of Portugal (Figure 1 and Table S1, Supporting information). The

DNA extraction protocol followed Barbosa *et al.* (2013). All DNA samples were individually amplified for a short fragment of the mitochondrial cytochrome *b* (cyt-*b*) for species identification (Barbosa *et al.* 2013). After confirming the samples as *Microtus cabrerae*, we quantified the DNA using a Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA) and pooled individual DNA samples into two groups, hereafter referred to as 'bone' and 'faecal' DNA pools. Pooled DNA samples were quantified again using a Qubit dsDNA BR Assay Kit (Invitrogen, Carlsbad, CA). GBS libraries were then prepared from pooled DNA samples (100 ng) following Elshire *et al.* (2011) with the adaptations from Wallace & Mitchell (2017). After checking the library fragment size profiles on a capillary instrument (Experion, Bio-Rad), 100 bp single-end sequences were collected on the HiSeq2500 (Illumina). GBS data from 107 individual tissue samples were already available from previous work (Barbosa *et al.* submitted).

Bioinformatics

For the tissue samples and the two non-invasive DNA pools we evaluated sequence quality using the software FastQC (Andrews 2011). For all three categories, a random subsample of 500 reads were BLASTed against the NCBI nucleotide database to determine the proportion of reads that belonged to target (mammalian genomes) and non-target DNA (e-value < $1x10^{-4}$). For SNP calling and filtering we applied the UNEAK pipeline (Lu *et al.* 2013), using GBS data previously obtained from 107 tissue samples representative of 22 Cabrera vole populations (see Figure 1; Barbosa *et al.* submitted) combined with the sequences obtained for the bone and faecal samples. For this analysis, we defined a minimum base call of 50 reads in all 109 samples. We also set the minor and major allele frequency threshold to 5% and 50%, respectively. Filtering on a 5% minor allele frequency allows us to avoid errors from nucleotide base deamination that are common in degraded DNA samples and to eliminate Illumina sequencing errors, by excluding alleles only present in five or less samples (Hofreiter *et al.* 2014). Post-processing of the data was performed to combine forward and reverse reads of the same loci and to exclude potential paralogs following White *et al.* (2013): filtering out loci with more than 75% heterozygosity and with over 20% missing data and filtering out samples with over 25% missing data. Finally, we plotted the read counts per locus for the two DNA pools and the average of all tissue samples.

Population genomics analyses

We first performed a principal coordinate analysis (PCoA), using the *dudi.pco* function of the *ade4* package in R (Dray & Dufour 2007), to visualize the distribution of genetic variance between the Cabrera vole individuals and to match the genetic similarity of the pooled DNA samples to existing populations, in a model-free approach. We then used the software STRUCTURE (Pritchard *et al.* 2000) to test for genetic subdivision using all loci, calculating five replicates of each *K* (1 – 23) with 20 000 iterations each (10% burnin). From this analysis we determined the best-supported number of populations (*K*) as determined by the Evanno method (Evanno *et al.* 2005). Given that we did not perform individual ID for the pooled DNA samples, we could not estimate the observed heterozygosity for the non-invasive DNA pools, but we could calculate expected heterozygosity (H_e) from the number of reads for each allele at each locus as a proxy for gene diversity. We calculated expected heterozygosity (H_e) using the software GenAlEx v6.5 (Peakall & Smouse 2012) for each geographic region and for both DNA pools. We additionally selected two groups of tissue samples representing populations with overlapping geographic range with the DNA pools, to test for similarity in genetic parameters: the faecal DNA pool was compared to population 2 (P2) and the bone DNA pool was compared to populations 3-10 (P3-10) (Figure 1 and Table S1, Supporting information). The significance of the H_e comparisons was assessed using one-way ANOVA and the Tukey HSD post-hoc test (Abdi & Williams 2010).

4. Results

GBS of pooled DNA samples

Quantifications showed that the pooled DNA of 8 faecal and 67 bone samples amounted to 360 and 175 ng of DNA, respectively. The FastQC results showed that sequences from the pooled DNA samples were mostly high quality (Figure S1, Supporting information). Many of the sequences from the DNA pools returned no BLAST hits in the NCBI nucleotide database (64.4% and 35.4% for the faecal and bone DNA pools, respectively) (Figure 2). When BLAST results were positive, they showed very low percentages of target DNA (blasting to mammalian genomes), with 1.6% and 3.0% for the faecal and bone DNA pools, respectively, while most of the remaining hits were assigned to bacteria or adapters (Figure 2). The number of reads per locus differed between sample types, with the bone DNA pool resembling the average tissue read depth, but the faecal DNA pool had higher average read depth (Figure S1, Supporting information).

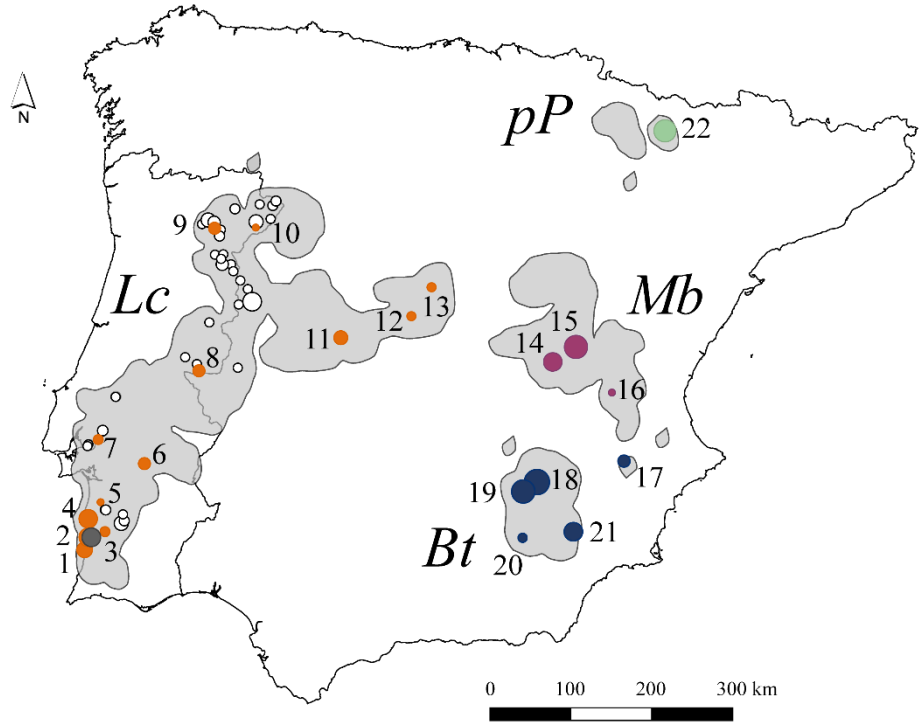


Figure 1. Map of the Iberian Peninsula with the distribution of the Cabrera vole (*Microtus cabrerae*) in grey (Barbosa *et al.* 2017); numbered and coloured dots represent 107 tissue samples from 22 populations genotyped in Barbosa *et al.* (submitted), which are distributed in four geographic nuclei as defined by Garrido-García *et al*. (2013) and Barbosa *et al.* (2017): *Lusocarpetan* (*Lc*, orange), *Montiberic* (*Mb*, purple), *Betic* (*Bt*, blue) and *pre-Pyrenean* (*pP*, green); grey and white dots represent the 8 faecal samples and 67 bones from owl (*Tyto alba*) pellets used this study, respectively (Table S1, Supporting Information); size of the circles is proportional to the number of samples at the same sampling location (1-13 samples).

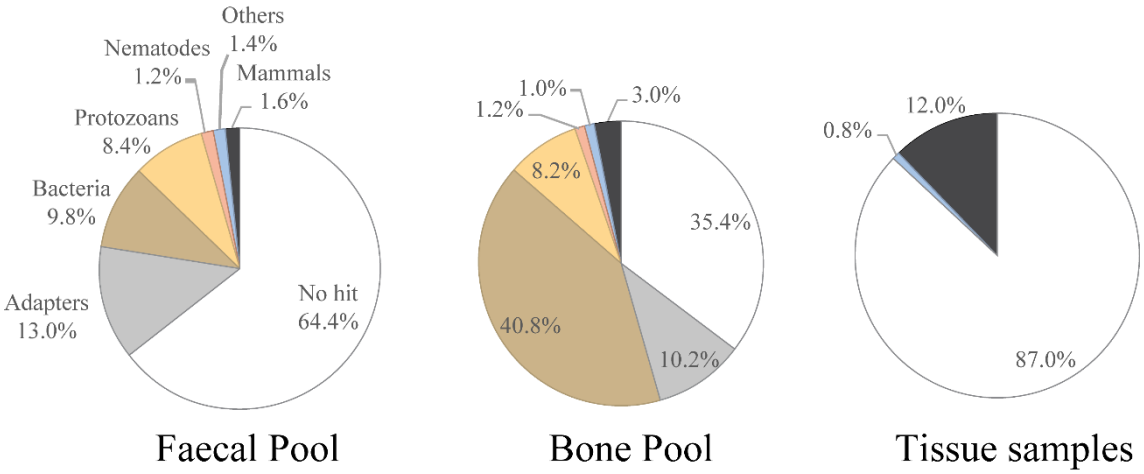


Figure 2. Pie charts of BLASTn results of 500 randomly selected sequencing fragments each for the faecal pools, the bone pools and tissue samples, screened against the NCBI nucleotide database. Colors refer to the same categories across pie charts

Population genomic analyses

We obtained a total of 3 968 unique single nucleotide polymorphisms (SNP). The faecal and bone DNA pools showed very low percentages of missing data, with 5.6% and 4.2%, respectively. The two principal coordinates of the PCoA analysis explained 13.1% of the total genetic variation and reflected well the sampling geography (Figure 3A). Three main clusters were identified corresponding to previously identified *ESUs* (Barbosa *et al.* submitted): *ESU1* [*Lusocarpetan* (*Lc*) region] – 42 voles plus bone and faecal DNA pools, *ESU2* [*Montiberic* with *Betic* (*Mb*+*Bt*)] – 55 voles and *ESU3* [*pre-Pyrenean* (*pP*)] – 10 voles (Figure 3A). The STRUCTURE analysis identified three main populations (*K*), matching the PCoA clusters (Figure 3B); the bone and faecal DNA pools clustered with *ESU1* in both analyses (Figure 3).

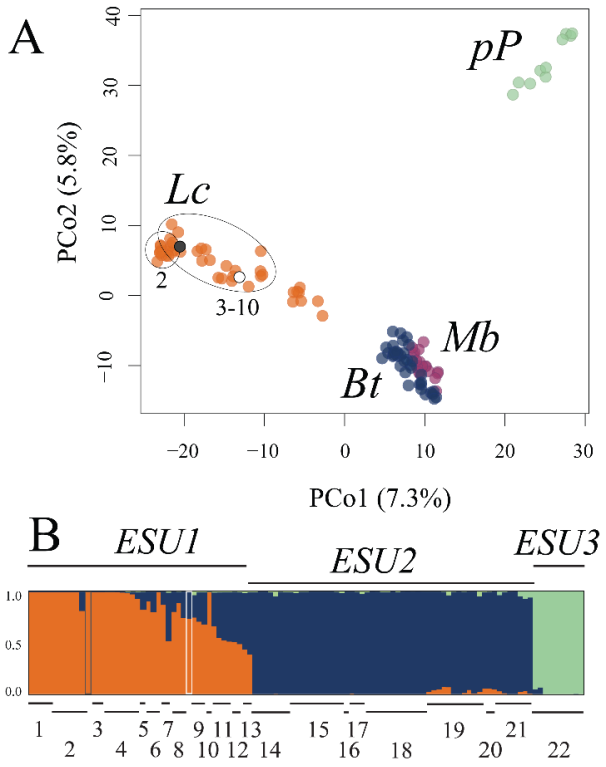


Figure 3. A) Principal coordinate analysis (PCoA) with the 3 968 loci identified in this study using the *dudi.pco* function implemented in the R package *ade4* 1.7.3 (Dray & Dufour 2007); coloured dots represent samples coloured by the geographic nucleus of origin as in A, and the grey and white dots represents the faecal and bone pools, respectively; black ellipses delimit samples belonging to the populations with the respective numbers, which have a similar geographic range as the non-invasive pools. **B)** Population assignment cluster analysis for the Cabrera vole using the software STRUCTURE indicating the proportion of sample assignment to each of the inferred best *K* number of populations (colours) that represent the three evolutionary significant units (*ESUs*) identified in Barbosa *et al.* (2017). Horizontal black bars and respective numbers and codes refer to the sampled populations as in A. Faecal and bone pools are represented by a grey and white outline, respectively.

The values of expected heterozygosities (H_e) are listed in Table 1. H_e values show that *Lc* diversity is not significantly different from P3-10. The non-invasive DNA pools have significantly lower H_e than the populations within the same geographic range (faecal DNA pool vs. P2 and bone DNA pool vs. P3-10). However, we must point out that the temporal time frame of sampling of the non-invasive DNA pools and their respective comparative populations are not the same: whereas P2 and P3-10 tissues were collected between 2011-2012 and 2003-2012, respectively, the respective non-invasive samples were collected in 2014 for the faecal DNA pool, and in 1980, 1996, 2010-2011 for the bone DNA pool.

Table 1. Genetic diversity for the fecal and bone non-invasive pools based on 3 968 SNPs presented as expected heterozygosities (H_e; mean and standard error). For comparison, population 2 (P2) and P3-10 represent sampled Cabrera vole populations of the same geographic range as the DNA pools, and *Lc*, *Mb*, *Bt*, and *pP* are the results based on all voles typed in each of the four geographic nuclei (see Figure 1).

Same geographic range

Population	H_e
Fecal pool	0.192 (0.003)
P2	0.230 (0.002)
Bone pool	0.221 (0.003)
P3-10	0.261 (0.002)
Lc	0.267 (0.002)
Mb	0.256 (0.002)
Bt	0.274 (0.002)
pP	0.163 (0.002)

Not significantly different ($p = 0.01$)

5. Discussion

This study confirms the application of high-throughput sequencing using a restriction enzyme-based method on degraded DNA pool samples, by gathering data appropriate for accurate population assignment. The electrophoretic patterns of the non-invasive DNA pools did not mimic exactly that of the tissue samples, and this was reflected in the low percentage of target sequencing found in both pools, which varied between 1.6% and 3.0% (Figures 2 and S1, Supporting information). Per base quality was found to be high on average for the non-invasive DNA pools, but we also found a higher than expected GC content in these samples, possibly reflecting the high proportion of bacterial sequences that are known to be very GC rich (Lassalle *et al.* 2015). However, with the non-invasive DNA pools we were able to recover the full panel

of SNPs identified in the previous tissue-based studies, with a low level of missing data per sample (4-6%).

Using the software STRUCTURE, we were able to confidently assign the faecal DNA pool to the south-west of *Lc*, where the faecal samples came from, and the bone DNA pool to the middle of *Lc*, as this was a mix of DNA from samples originally from the geographic area of P3-10, from different parts of *Lc* (Figure 3). Looking at the values of expected heterozygosity (H_e) as a proxy for gene diversity, we found that the DNA pools had generally lower diversity than the comparative populations (Table 1). This could be interpreted as a lower power to detect genetic diversity using these samples due to, for example, allele dropout. However, the differences could also reflect the temporal difference in the collection of the tissue vs. the non-invasive DNA pools, given that the different sample types were not collected during the same time period.

This work demonstrates the possibility of conducting genomic studies using NiGS in small mammals. Moreover, we have shown that genomic data can be obtained for non-model species with no genomic resources available other than control data based on the genotyping of available tissue samples. Up to now, most studies based on NiGS that have not required prior genomic resources were restricted to museum samples, due to their generally higher DNA concentration than most field non-invasive samples (Beck & Semple 2015). However, previous studies have shown that restriction enzyme-based HTS can be applied to samples with very low DNA quality with high success (Graham *et al.* 2015). Our study opens new opportunities in conservation genetics, such as identifying the population of origin of migrants and, in more extreme cases: in wildlife traffic; determining population structure; compare levels of genetic diversity between populations; etc. Future work should include the use of individual barcoding on non-invasive samples to test if there is enough coverage of orthologous SNPs for individual based studies in conservation, as is possible for studies based on well-preserved tissues.

Acknowledgments

We would like to thank J.A. Cabral, H.M. Vale-Gonçalves, I. Roque and the *Tyto Tagus* team, and D. Galícia from the Museum of the University of Navarra for providing bone samples for this study, as well as R. Pita, A. Mira and P. Beja for providing both bone and fecal samples in the scope of the Fundação para a Ciência e Tecnologia research project PTDC/BIA-BIC/6582/2014.

6. References

Abdi H, Williams LJ (2010) Tukey's Honestly Significant Difference (HSD) Test. In: *Encyclopedia of Research Design* (eds Salkind NJ, Dougherty DM, Frey B), pp. 583–585. Sage, Thousand Oaks, CA, USA.

Andrews S (2011) FastQC: a quality control tool for high throughput sequence data, Babraham Institute, Cambridge, UK.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.

Barbosa S, Mestre F, White TA *et al.* (*submitted*) Integrative approaches to guide conservation decisions: using genomics to define conservation units and functional corridors.

Barbosa S, Paupério J, Herman JS *et al.* (2017) Endemic species may have complex histories: within-refugium phylogeography of an endangered Iberian vole. *Molecular Ecology*, **26**, 951–967.

Barbosa S, Paupério J, Searle JB, Alves PC (2013) Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling. *Molecular Ecology Resources*, **13**, 43–56.

Beck JB, Semple JC (2015) Next-generation sampling: pairing genomics with herbarium specimens provides species-level signal in *Solidago* (Asteraceae). *Applications in Plant Sciences*, **3**, 1500014.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, **9**, 1279–1301.

Bi K, Linderoth T, Vanderpool D *et al.* (2013) Unlocking the vault: next-generation museum population genomics. *Molecular Ecology*, **22**, 6018–6032.

Bischof R, Gregersen ER, Brøseth H, Ellegren H, Flagstad Ø (2016) Noninvasive genetic sampling reveals intrasex territoriality in wolverines. *Ecology and Evolution*, **6**, 1527–1536.

Bohmann K, Evans A, Gilbert MTP *et al.* (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, **29**, 358–367.

Costa V, Rosenbom S, Monteiro R, O'Rourke SM, Beja-Pereira A (2017) Improving DNA quality extracted from fecal samples - a method to improve DNA yield. *European Journal of Wildlife Research*, **63**, 3.

DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Reviews Genetics*, **5**, 702–712.

Dray S, Dufour A-B (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, e19379.

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

Graham CF, Glenn TC, McArthur AG *et al.* (2015) Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, **15**, 1304–1315.

Henry P, Henry A, Russello MA (2011) A noninvasive hair sampling technique to obtain high quality DNA from elusive small mammals. *Journal of Visualized Experiments*, **49**, e2791.

Hofreiter M, Paijmans JLA, Goodchild H *et al.* (2014) The future of ancient DNA: technical advances and conceptual shifts. *BioEssays*, **37**, 284–293.

Lassalle F, Périan S, Bataillon T *et al.* (2015) GC-Content evolution in bacterial genomes: the biased gene conversion hypothesis expands. *PLoS Genetics*, **11**, e1004941.

Lu F, Lipka AE, Glaubitz J *et al.* (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, **9**, e1003215.

Lynn MS, Sechi P, Chikhi L, Goossens B (2016) Primate conservation genetics at the dawn of conservation genomics. In: *An Introduction to Primate Conservation* (eds Wich SA, Marshall AJ), pp. 53–77. Oxford University Press, Oxford, UK.

Mestre F, Pita R, Paupério J *et al.* (2015) Combining distribution modelling and non-invasive genetics to improve range shift forecasting. *Ecological Modelling*, **297**, 171–179.

Mumma MA, Adams JR, Zieminski C *et al.* (2016) A comparison of morphological and

molecular diet analyses of predator scats. *Journal of Mammalogy*, **97**, 112–120.

Mumma MA, Zieminski C, Fuller TK, Mahoney SP, Waits LP (2015) Evaluating noninvasive genetic sampling techniques to estimate large carnivore abundance. *Molecular Ecology Resources*, **15**, 1133–1144.

Ogden R (2011) Unlocking the potential of genomic technologies for wildlife forensics. *Molecular Ecology Resources*, **11**, 109–116.

Paupério J, Herman JS, Melo-Ferreira J *et al.* (2012) Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia. *Molecular Ecology*, **21**, 6015–6032.

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

Peñalba J V, Smith LL, Tonione MA *et al.* (2014) Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Molecular Ecology Resources*, **14**, 1000–1010.

Perry GH, Marioni JC, Melsted P, Gilad Y (2010) Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology*, **19**, 5332–5344.

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

Reed JZ, Tollit DJ, Thompson PM, Amos W (1997) Molecular scatology: the use of molecular genetic analysis to assign species, sex and individual identity to seal faeces. *Molecular Ecology*, **6**, 225–234.

Rowe KC, Singhal S, Macmanes MD *et al.* (2011) Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources*, **11**, 1082–1092.

Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, **15**, 749–763.

Taylor HR, Dussex N, Van Heezik Y (2017) Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Global Ecology and Conservation*, **10**, 231–242.

Waits LP, Paetkau D (2005) Noninvasive genetic sampling tools for wildlife biologists: a

review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, **69**, 1419–1433.

Wallace JG, Mitchell SE (2017) Genotyping-by-Sequencing. *Current Protocols in Plant Biology*, **2**, 64–77.

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

7. Supporting information

Table S1. Sample types and location of the Cabrera vole specimens used in this study. Map code is only available for tissue samples and it refers to the numbers in Figure 1.

Map Code	Type	Number of samples	Community/District	Country	Latitude	Longitude
1	Tissue	5	Beja	Portugal	37.609	-8.794
2	Tissue	7	Beja	Portugal	37.753	-8.755
3	Tissue	2	Beja	Portugal	37.828	-8.523
4	Tissue	7	Setúbal	Portugal	37.959	-8.771
5	Tissue	1	Setúbal	Portugal	38.098	-8.561
6	Tissue	3	Évora	Portugal	38.632	-8.045
7	Tissue	2	Santarém	Portugal	38.870	-8.725
8	Tissue	3	Castelo Branco	Portugal	39.763	-7.399
9	Tissue	3	Bragança	Portugal	41.343	-6.577
10	Tissue	1	Bragança	Portugal	41.304	-7.183
11	Tissue	4	Ávila	Spain	40.114	-5.256
12	Tissue	2	Madrid	Spain	40.349	-4.350
13	Tissue	2	Madrid	Spain	40.622	-3.877
14	Tissue	7	Cuenca	Spain	39.861	-2.158
15	Tissue	11	Cuenca	Spain	40.026	-1.817
16	Tissue	1	Valencia	Spain	39.508	-1.291
17	Tissue	3	Murcia	Spain	38.730	-1.184
18	Tissue	13	Albacete	Spain	38.512	-2.401
19	Tissue	11	Jaén	Spain	38.407	-2.597
20	Tissue	2	Granada	Spain	37.895	-2.649
21	Tissue	7	Murcia	Spain	37.950	-1.888
22	Tissue	10	Huesca	Spain	42.438	-0.466
	Fecal	1	Odemira	Portugal	37.744	-8.736
	Fecal	1	Odemira	Portugal	37.745	-8.743
	Fecal	1	Odemira	Portugal	37.742	-8.736
	Fecal	1	Odemira	Portugal	37.752	-8.755
	Fecal	1	Odemira	Portugal	37.750	-8.753
	Fecal	1	Odemira	Portugal	37.750	-8.742
	Fecal	1	Odemira	Portugal	37.745	-8.737
	Fecal	1	Odemira	Portugal	37.748	-8.738
	Bone	2	Aljustrel	Portugal	37.960	-8.260
	Bone	4	Aljustrel	Portugal	37.930	-8.300
	Bone	1	Ferreira do Alentejo	Portugal	38.030	-8.270
	Bone	2	Setúbal	Portugal	38.070	-8.530
	Bone	1	Santarém	Portugal	38.954	-8.628
	Bone	1	Santarém	Portugal	38.946	-8.615

	Bone	2	Santarém	Portugal	39.341	-8.473
	Bone	2	Santarém	Portugal	38.790	-8.820
	Bone	1	Santarém	Portugal	38.770	-8.830
	Bone	1	Castelo Branco	Portugal	39.828	-7.503
	Bone	1	Castelo Branco	Portugal	40.231	-7.178
	Bone	2	Castelo Branco	Portugal	39.757	-7.330
	Bone	8	Salamanca	Spain	40.488	-6.570
	Bone	1	Salamanca	Spain	39.737	-6.736
	Bone	1	Salamanca	Spain	40.447	-6.762
	Bone	1	Salamanca	Spain	40.717	-6.753
	Bone	1	Salamanca	Spain	40.624	-6.638
	Bone	1	Bragança	Portugal	41.422	-6.343
	Bone	1	Bragança	Portugal	40.994	-7.142
	Bone	2	Bragança	Portugal	41.518	-6.885
	Bone	3	Bragança	Portugal	40.890	-7.036
	Bone	1	Bragança	Portugal	40.950	-7.050
	Bone	1	Bragança	Portugal	41.002	-7.021
	Bone	2	Bragança	Portugal	41.375	-7.288
	Bone	1	Bragança	Portugal	40.894	-6.901
	Bone	1	Bragança	Portugal	40.818	-6.862
	Bone	1	Bragança	Portugal	41.330	-7.356
	Bone	4	Bragança	Portugal	41.379	-7.276
	Bone	2	Bragança	Portugal	41.205	-7.095
	Bone	1	Bragança	Portugal	41.236	-7.085
	Bone	1	Bragança	Portugal	41.277	-7.075
	Bone	3	Bragança	Portugal	41.357	-7.186
	Bone	4	Bragança	Portugal	41.383	-6.562
	Bone	1	Bragança	Portugal	41.581	-6.514
	Bone	2	Bragança	Portugal	41.573	-6.323
	Bone	2	Bragança	Portugal	41.625	-6.277
	Bone	1	Bragança	Portugal	41.361	-7.289

Samples from localities with Map Code were obtained from Barbosa *et al.* (submitted)

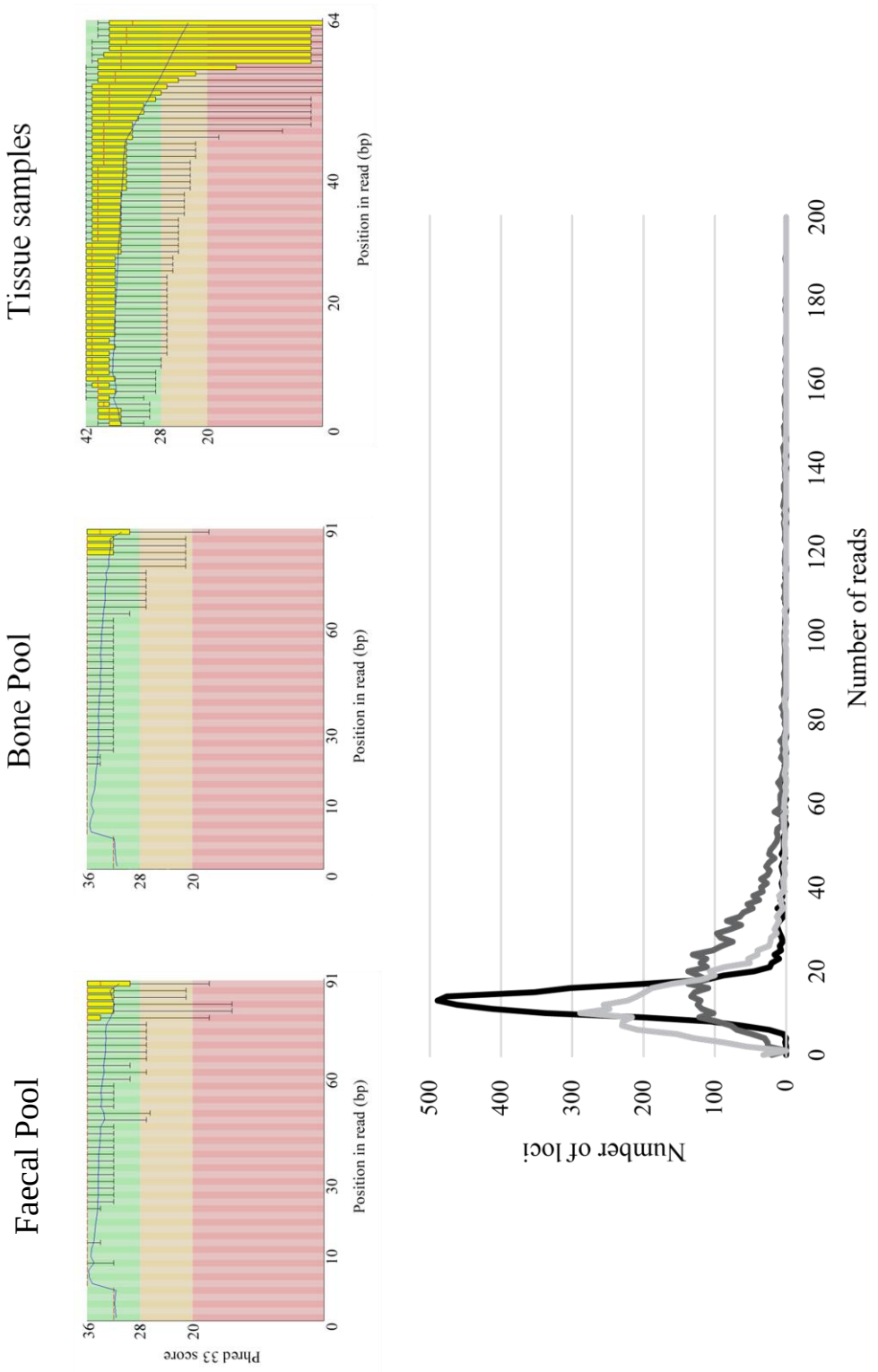


Figure S1. Sequencing quality of both non-invasive pools compared to the remaining tissue samples analysed in the software FastQC: per base sequencing quality. Below there is a plot of the number of reads per locus for each sample type: faecal pool (grey), bone pool (light grey) and tissue samples (average of all samples, black).

CHAPTER V.

GENERAL DISCUSSION

1. Discussion

The main goal of this thesis was to understand the evolutionary history of the Cabrera vole, to provide a broader context for conservation planning in this species. This goal was achieved using molecular tools, given that genetic diversity contributes to decrease a species' extinction risk, being also a predictor of a species evolutionary potential (Moritz & Potter 2013; Willoughby *et al.* 2015).

It is commonly accepted that many species are facing extinction due to human induced climate change and habitat loss, and due to the consequential population reduction, genetic diversity is expected also to be decreasing (Lacy 1987; Ceballos *et al.* 2015). Given that genetic diversity plays a fundamental role in maintaining healthy populations, effective conservation measures for particular taxa need to take into account how much genetic diversity there is, how it is distributed across the landscape, and what were the evolutionary factors driving the observed patterns (Crandall *et al.* 2000; Allendorf *et al.* 2012; Mills 2013).

Molecular conservation tools

One of the limitations in obtaining genetic data from wildlife is the difficulty in collecting samples from rare, threatened and elusive species, for which the disturbance of the individuals in their natural habitat needs to be minimised (Waits & Paetkau 2005; Beja-Pereira *et al.* 2009; Bastos *et al.* 2011; Paupério *et al.* 2012; Latinne *et al.* 2013). As our model species was the Cabrera vole, we first developed a DNA-based protocol using both mitochondrial and nuclear markers to make use of non-invasive samples (faeces, hairs and bones) of all rodents from the Iberian Peninsula (*Paper I*). We were able to accurately identify all species of Iberian rodents with the exception of *Microtus lusitanicus* and *M. duodecimcostatus*, which are currently hybridising (Bastos-Silveira *et al.* 2012). We confidently identified and distinguished Cabrera voles by genetic typing of ear biopsies (the species may be confused with juvenile *Arvicola sapidus* when caught in the field), and also were able to identify the species from faecal samples based on small DNA fragments (< 300 bp). The molecular protocol developed for the Iberian rodent species identification has already been applied in ecological niche modelling of the Cabrera vole and population dynamic studies and using non-invasive samples (Mestre *et al.* 2015; Ferreira *et al. in prep*). Additionally, for the other Iberian rodents we observed a broad range of intra- and interspecific genetic distances, highlighting the importance of wide-scale barcoding initiatives to increase our understanding of taxonomy and

evolutionary processes. We were able to assess the feasibility of using nuclear markers on non-invasive population genetic studies, particularly when working with small samples, a limitation typically shown by small mammals.

The evolutionary history of the Cabrera vole

Phylogenetic positioning

The genus *Microtus* represents one of the fastest mammalian radiations known to date, thought to have diversified during the last 2 Mya into 12 subgenera and 65 species currently recognised by the IUCN (Conroy & Cook 1999, 2000; IUCN 2016). Its subgenera are very well supported at the genetic level, distributed throughout the entire northern hemisphere (Jaarola *et al.* 2004; Galewski *et al.* 2006; Fink *et al.* 2010; Martínková & Moravec 2012). In our study we have sampled five of those subgenera, three found in the Palearctic, one in the Nearctic and one Holarctic (*Paper II*). Our data is consistent with the contention that the Cabrera vole belongs to its own subgenus, *Iberomys*, given the time since divergence from the sister subgenus *Agricola*, and by comparison to the divergence time of the other two well established Palearctic subgenera, *Microtus* and *Terricola* (Chaline 1974). The separation of *Agricola* and *Iberomys* from *Microtus* and *Terricola* suggests that these groups represent two independent radiations from Asia into Europe, while only one radiation into North America has been detected, originating all the endemic species existing in this region. Finally, it appears that *M.* (*Alexandromys*) *oeconomus* shows a Holarctic distribution as a result of more recent expansions from Asia both west and east. To the best of our knowledge, *Agricola* and *Iberomys* are monospecific subgenera, represented by the field (*M. agrestis*) and Cabrera (*M. cabrerae*) voles, respectively. This implies that the Cabrera vole's closest living relative is the field vole (*M. agrestis*), a proximity that is also supported by karyotypic similarities, which previously have led to the placement of the Cabrera vole in the *Agricola* subgenus (Zagorodnyuk 1990). However, the Cabrera vole is a highly distinctive species within *Microtus* due to various life history, morphologic and karyotypic features, and this led some authors to propose the elevation of the Cabrera vole to its own genus, *Iberomys* (Cuenca-Bescós *et al.* 2014; Pita *et al.* 2014). It has been suggested that the Iberian Peninsula was the site of origin of highly specialised and distinct arvicoline lineages such as *Iberomys* during the Pleistocene (Cano *et al.* 2013). Our results based on a phylogenomic approach support the distinctiveness of species such as the Cabrera and field voles but confidently disprove the hypothesis of *Iberomys*

as a separate genus, given that the elevation of the Cabrera vole to its own genus would make the genus *Microtus* paraphyletic.

Phylogeographic history

For the study of the Cabrera vole phylogeographic history we sampled 284 tissue and NiGS (bones from owl pellets) samples from the entire distribution of the species. The use of NiGS (*Paper I*) was necessary to obtain a good representation of some of the geographic nuclei occupied by the Cabrera vole (of which there are four: *Paper III*), especially the *Lusocarpetan* (*Lc*) and the *pre-Pyrenean* (*pP*)

Despite its long divergence within the *Iberomys* subgenus, assumed to have occurred in western Europe around 100 kya, our study on the phylogeography of the Cabrera vole suggests that the mitochondrial tMRCA of current populations dates to the LGM (*Paper III*). Combining mitochondrial and traditional nuclear loci with the Cabrera vole fossil record, we determined that this species expanded from a single refugium around 17 kya, possibly in south-eastern Iberia, as expected from the species fossil record and for a species adapted to the Mediterranean climate (*Paper III*) (Garrido-García & Soriguer-Escofet 2012; Laplana & Sevilla 2013). However, it is important to avoid pinpointing precise refugial locations as, for instance, there is a bias towards a greater likelihood of fossilisation around south-eastern Iberia (Laplana & Sevilla 2013), which can also bias our perceptions.

The modelling that we carried out in *Paper III* suggests that, as a result of increased available suitable habitat after the LGM, there was a population expansion followed by the separation of various populations, including (but likely not limited to) the current western (*Lc*) and eastern [*pP*, *Montiberic* (*Mb*) and *Betic* (*Bt*)] populations (López-García *et al.* 2012). By then, the fossil record suggests that the Cabrera vole had established resilient populations in the pre-Pyrenean massifs and southern France (Garrido-García & Soriguer-Escofet 2012; Laplana & Sevilla 2013). We consider that these range expansions were interrupted by the last cold period that marked the end of the Pleistocene, the Younger Dryas (YD). We argue that during this time period, from *c.* 13 – 11 kya, the Cabrera vole populations got separated in west and east Iberia, leading to the divergence in allopatry of the two currently known well supported mitochondrial lineages, *west* and *east* (*Paper III*). This is further supported in *Paper IV*, where we used a wider genomic approach (3 341 SNPs) and observed that at *K*=2, the populations are genetically divided into western and eastern Iberia, and only at *K*=3 the *pre-Pyrenean* (*pP*) geographic nucleus is considered a different cluster. Interestingly, the higher

divergence of *pP* within the eastern populations is inferred from both mitochondrial and nuclear data, suggesting an older divergence from the other eastern populations, or a greater divergence as a consequence of long term isolation (*Paper III* and *Paper IV*).

A recent study from Varela (2016) analysing ancient DNA of Cabrera voles shows a significant increase in genetic diversity a few thousand years after the YD, during the Neolithic (c. 7.5 - 5 kya), indicative of population expansions, also supported by an increased representation in the fossil record (Laplana & Sevilla 2013). These expansions may be associated to human alterations of the landscape, such as forest reduction and the emergence of the initial vegetation succession stages, in which the Cabrera vole is known to thrive (Fernandes *et al.* 2008a; Laplana & Sevilla 2013). The time of diversification of the two Cabrera vole lineages is consistent with these time periods, right after the YD (*Paper III*). The Neolithic haplotype and nucleotide diversities appear to be maintained until the Bronze and Iron ages (*c.* 2 kya) (Varela 2016). From this point on, a decrease in genetic diversity and population reduction in the fossil records may be associated with extinctions of many eastern Iberian and southern French populations due to increased human pressures on the Cabrera vole habitat (Laplana & Sevilla 2013; Varela 2016).

Mito-nuclear discordance

In the Cabrera vole, it is possible that the mitochondrial *west* lineage started expanding east during the Neolithic population growth, coming into secondary contact with the *east* lineage (*Paper III*). We also found evidence of such secondary contact in *Paper IV*, using a landscape genomic approach based on 3 341 SNPS, where we observed an extensive admixture between eastern and western alleles in populations of *Lc*, i.e. in the opposite direction of the mitochondrial data. This mito-nuclear discordance suggests that: i) there is mitochondrial introgression of the *west* lineage into the eastern populations; and/or ii) the initial contact zone was located further west and there was a replacement of the mitochondrial *east* lineage haplotypes in the west; and/or iii) differential introgression of both mitochondrial and nuclear loci occurred, though this is more frequent in taxa that are more divergent than the Cabrera vole lineages (Carling & Brumfield 2008; Toews & Brelsford 2012; Gralka *et al.* 2016). We further observed that the nuclear admixture originated mostly from *Bt* rather than the closer *Mb*, which suggests that this secondary contact might have happened after the establishment and vicariance of the current eastern geographic nuclei.

Molecular conservation planning

Given our new knowledge of the past and current patterns of genetic diversity in the Cabrera vole, there is now the opportunity for effective evidence-based conservation planning (Crandall *et al.* 2000).

In *Paper IV* we conducted a study that integrated genomics with landscape analysis, aiming to define conservation units (*CUs*) for the Cabrera vole. We found that the species is divided into three well-defined evolutionarily significant units (*ESUs*): *ESU1* is coincident with the *Lc* geographic nucleus; *ESU2* is represented by the *Mb* and *Bt* geographic nuclei; and *ESU3* is represented by the *pP* geographic nucleus. *ESU1* is the most widespread *CU* and it appears to be genetically more homogenous than the remaining *ESUs*. This population currently appears to be expanding especially in the north, possibly as a sign of climate induced habitat range shifts (Vale-Gonçalves & Cabral 2014). In the northern and central regions of *ESU1* we observed some levels of admixture with *ESU2* possibly from past connectivity (*Paper III* and *Paper IV*). *ESU2* is currently the most diverse, and it is subdivided into three *MUs* (*MU2*, *MU3* and *MU4*; *Paper IV*). This is also related to the fact that connectivity is very limited between most of its populations, especially between and within *MU3* and *MU4*, similarly to the increased structure associated to population decline found in the populations of the alpine chipmunk (Holmes *et al.* 2016). The most recent population assessments show that various southern populations of *ESU2* have disappeared due to habitat destruction, and also possibly because they are located at the species environmental tolerance limits (Garrido-García *et al.* 2013); as edge populations, these *MUs* can store genetic diversity of importance for the species survival and evolution, and their extinction could lead to a serious loss for the Cabrera vole gene pool (Hampe & Petit 2004). The increased genetic structure observed within *ESU2* might already be an effect of inbreeding within these *MUs*, which makes allele frequencies very similar within populations due to high levels of identity-by-descent (Allendorf *et al.* 2012), but which leads to divergence between populations. *ESU3* is currently the least diverse and most isolated conservation unit, and also has some of the lowest levels of heterozygosity (*Paper IV*). Thus, it should be considered as a priority for conservation given that it is also the one that contributes the most to the Cabrera vole genetic diversity (*Paper IV*). In terms of adaptive variation, we observed a west-east separation of ‘adaptive’ genetic diversity, consistent with the patterns of nuclear variation (*Paper IV*). However, this was based on very few outlier loci and conclusions are thus very limited.

The small number of populations, their isolation, small number of individuals, and genetic homogeneity are likely to lead to strong genetic and ecological Allee effects which may, in turn, cause the extinction of many of the Cabrera vole populations (Fagan & Holmes 2006; Garrido-García *et al.* 2013; Luque *et al.* 2016). This is especially problematic in the Cabrera vole given that the species is distributed in metapopulations and, with increasing genetic diversity loss, theory predicts that the species will have decreased colonisation ability, even if favourable habitat is available (Amarasekare 1998). On this basis, we must rely on overall levels of variation as a proxy for evolutionary potential and, as such, there is a need to protect as much genetic diversity as possible to ensure that the species has the best chances of persisting under increasingly dynamic environmental conditions (Eckert *et al.* 2008; Moritz & Potter 2013; Eizaguirre & Baltazar-Soares 2014).

Given the importance of population genetic studies in conservation and the power of genomics to detect population structure, a combination of genomics with non-invasive sampling could greatly benefit the study of threatened species and their populations. Microsatellites have been the markers of choice when using non-invasive samples like faeces, as they are usually represented by small fragments (< 300 bp), and have high levels of polymorphism (due to high mutation rates), allowing a variety of ecological and evolutionary questions to be answered (Selkoe & Toonen 2006). However, microsatellites lack standardisation across laboratories and also suffer from high error rates (allele dropout, false alleles) (Taberlet *et al.* 1999). With the arrival of high throughput sequencing, new techniques are rendering microsatellite limitations less pronounced (de Barba *et al.* 2016). Nevertheless, SNPs are rapidly becoming the marker of choice for population genetic studies and monitoring due to their higher genotyping efficiency, data quality, genome wide coverage, and analytical simplicity, while allowing the study of a greater variety of questions on multiple evolutionary scales, including a focus on adaptation (Morin *et al.* 2004; Selkoe & Toonen 2006; Schwartz *et al.* 2007; Kraus *et al.* 2015). So far, the biggest limitation of SNP application to non-invasive studies is the need of targeted methodologies to maximise the probability of amplification of homologous DNA sequences. However, targeted methodologies (e.g. exon capture, SNPchips) are very costly to develop as they require a reference genome/transcriptome and the design of hundreds to thousands of probes, and they may additionally suffer from ascertainment bias (Andrews *et al.* 2016). We thus aimed at developing a non-invasive genotyping methodology based on a genomic approach that could be useful for conservation studies. This required a methodology based on anonymous sequencing, such as RAD-seq (and GBS), allowing for studies in non-model species that usually have

few genomic resources available (Morin *et al.* 2004). However, these non-targeted methodologies can lead to high levels of missing data, especially in samples with degraded DNA. We tested the applicability of GBS in DNA pools extracted from Cabrera vole faecal and bone (from owl pellets) samples, in order to perform population assignment and assessment of levels of genetic diversity (*Paper V*). We obtained a very low percentage of missing data (around 5%), even when compared to the tissue sample dataset, and were able to achieve a reliable indication of the population of origin of the two DNA pools. This information was obtained from less than 3% of all reads, which suggest that increasing the endogenous DNA sequencing with techniques such as double-digest RAD-seq could greatly improve the amount of information obtained from an individual sample (Graham *et al.* 2015). Moreover, these methodologies can be reused with additional samples, either from the same or different species, for different questions that may arise in the future. Further discussion can be found in *Future perspectives*.

2. Conservation recommendations

As previously indicated, the Cabrera vole populations are divided into three *ESUs* and five *MUs*, and each of which have different priorities for conservation.

The most distinctive and also the most vulnerable population is *ESU3* (or *MU5*) that represents the *pP* geographic nucleus. This population has the lowest estimates of genetic diversity and is the population with the lowest connectivity to the remaining Cabrera vole distribution (*Paper IV*). Even so, levels of inbreeding appear to be low. This is further supported by recent census of the population in the province of Huesca, Aragón, that have detected various UTM squares with newly described Cabrera vole colonies (Jato R, *pers. comm.*), suggesting that there are ample opportunities for outbreeding. In any case, it is important not to be complacent about *ESU3*. For instance, a recent study from Escribano *et al.* (2016) shows that the Cabrera vole is currently extinct in Navarra, within the *pP* geographic nucleus, given that most records are more than 10 years old and a recent census confirmed the absence of the species in most areas where it was previously described. As population censuses are conducted in an increasing number of areas of the Cabrera vole distribution, it is common to conclude that many of the textbook locations are no longer occupied. Garrido-García *et al.* (2008) found that most populations in Andalucía (southern areas of *ESU2*) have recently become extinct, which suggests that the population is also receding in the southernmost

regions of the Cabrera vole distribution. As previously mentioned, this area encapsulates a high proportion of the overall species genetic diversity, and the connectivity between and within *MUs* appears to be highly limited, possibly due to conversion of suitable habitat into farmland, associated to increase in temperature and decrease in precipitation (Mira *et al.* 2008; Laplana & Sevilla 2013).

The studies on the current distribution of the Cabrera vole suggest that its populations, which have extreme colonisation-extinction-recolonisation dynamics, need to be carefully managed. There is a need for a reassessment of suitable available habitat and the protection of corridors that facilitate dispersal between habitat patches, especially within *MUs* (do Rosário & Mathias 2007; Albert *et al.* 2017). A framework for the preservation of Cabrera vole populations should encompass the following (Crandall *et al.* 2000; Fraser & Bernatchez 2001; Albert *et al.* 2017):

1) the generation of habitat quality and resistance maps for the species and its *ESUs*;
2) the analysis of the connectivity of *ESUs* habitat networks, quantifying the contribution of each habitat pixel to short- and long-range connectivity to define corridors through time;
3) the modelling of changes of climate suitability at various time points in the future;
4) the identification of spatial conservation priorities based on a set of criteria (e.g. contribution to the species overall diversity, most diverse and/or isolated populations, etc);
5) the establishment of different prioritisation schemes into the future with spatially explicit dynamic land-use simulations.

At a broader scale, Cabrera vole conservation should be integrated in a wider conservation approach for the Iberian Peninsula, considering various vertebrate species that represent the range of bioclimates, dispersal abilities, and responses to habitat fragmentation and climate change observed in Iberian species, so that protected areas and dispersal corridors can provide effective conservation outcomes (Albert *et al.* 2017).

The Cabrera vole is mostly found in agro-silvo-pastoral ecosystems and farmlands, particularly agricultural fields and road verges, that maintain tall grass for the species to find shelter, feed and disperse with a low predation risk (Pita *et al.* 2006; Santos *et al.* 2006). Thus corridors need to be incorporated in the intensively managed Iberian farmland, which for the short term might just be a matter of avoiding vegetation succession, especially on road verges (Pita *et al.* 2014). It is also important to maintain

small ‘stepping-stone’ habitat patches around 1000 m apart, particularly in areas with low vole densities. This would allow efficient colonisation of dispersing individuals and of moving colonies, given that the few estimates of Cabrera vole dispersal abilities were one individual dispersing *c.* 500 m in one night and an entire colony dispersing 1300 m in eight months (do Rosário & Mathias 2007; Pita *et al.* 2007).

ESU1 appears to be the most stable *CU*, with high connectivity between many of its populations, especially in southwestern Iberia. However, connectivity with the central Iberian populations and the northwest are apparently decreasing (*Paper IV*). In addition, with increasing pressures in southern areas of the species distribution, it is likely that those populations will undergo dramatic range contractions as is happening in *ESU2*, and thus maintaining connectivity to northern suitable areas will be very important. The difference in functional connectivity currently observed within the west and the east of the species distribution might be related to the topography of the terrain, given that western Iberia is more homogenous than central and eastern Iberia (Vera 2004). Thus effective conservation management will require the protection of dispersal corridors from less suitable to more suitable areas. Finally, we would like to stress the importance of developing international conservation strategies for the Cabrera vole given that most of the genetic variation is currently found in Spain, but most of the suitable habitat for the species and the most stable and suitable areas in the future appear to be present in Portugal (Mestre *et al.* 2015).

3. Conclusions

We found that non-invasive methodologies can greatly benefit the study of species under conservation concern, either by providing more extensive sampling without capturing animals, or by granting access to more detailed genetic variation representative of the genome wide population diversity. We were able to identify all species of rodents present in the Iberian Peninsula, using mitochondrial and nuclear markers, with exception of *Microtus lusitanicus* and *M. duodecimcostatus*. We additionally optimised the barcoding method for non-invasive samples, including bones from barn owl (*Tyto alba*) pellets, hair and faecal samples. We used this protocol in order to obtain a comprehensive sampling of the Cabrera vole’s distribution using bones from barn owl pellets.

Our data suggest that the Cabrera vole represents a distinct subgenus (*Iberomys*) within *Microtus*, and one of the first *Microtus* radiations to Europe, together with *M.* (*Agricola*) *agrestis*. This resolves previous uncertainties about the species phylogenetic

positioning, often placed together with Nearctic species. We also found that the three lineages of *M. agrestis* are highly distinctive (consistent with them being different species), and that other species such as *M. arvalis* and *M. oeconomus* have also very distinct genetic lineages throughout their ranges. After the spread of *Microtus* to Europe from a likely source area in Asia, the *Iberomys* subgenus started diverging, leading to the appearance of the Cabrera vole in Western Europe around 100 kya. The distribution of the species would have fluctuated with the Quaternary glaciations, with the last major contraction during the LGM. From an Iberian LGM refugium, it appears that the species expanded and got separated once again during the YD, around the time of divergence of the two western and eastern Iberian groups. Expansions after this last cold period apparently led to secondary contact of these lineages, likely associated to the emergence of agriculture.

The Cabrera vole is currently distributed in four main geographic nuclei, which we deemed to be represented in three *ESUs*. The most threatened of these *ESUs* is the *pre-Pyrenean*, which shows the lowest diversity and highest divergence, being also the one that contributes the most to the species overall genetic diversity. The *ESU* representing eastern and south-eastern populations has the highest diversity, but its southernmost populations have already lost functional connectivity and may soon become extinct. The *ESU* representing western and central Iberian populations has the larger range and likely will occupy the majority of suitable habitat areas for the Cabrera vole in the future. The conclusions of this thesis need to be carefully considered for the planning for the conservation planning of the Cabrera vole.

Finally, we found that it is possible to perform genomic-based population monitoring on non-model species of rodent even with non-invasive procedures and in the absence of pre-existing genomic resources.

4. Future perspectives

The research conducted in this thesis allowed us to develop new tools for the study of wild populations using NiGS, but also better understand the evolutionary history of the Cabrera vole, as well as its conservation needs. However, various questions have arisen that remain unanswered or could benefit from further research:

Resolving and timing the Microtus *rapid radiation*

Previous phylogenetic studies on the *Microtus* genus could not determine the branching order of this rapid radiation. We proposed to do so by analysing a partial mitogenome and genome wide SNP data, however the velocity of this radiation might not ever allow certain answers to such questions. We gained further clarity on the phylogenetic relations within this genus, however it would be interesting to include more species, especially from the Asian and Nearctic species, to see if the branch supports would be stronger, given a more complete picture of the entire radiation. Future studies should use sequences rather than SNPs given that mutation models could then easily be incorporated, possibly improving the support of the deeper nodes.

Another interesting aspect would be to incorporate fossils into the estimation times of these vicariant events, especially given that the genus *Microtus* has an abundant fossil record.

This is an especially interesting group given that the number of species in a given habitat is thought to be positively correlated with the time the lineage has occupied that habitat for (Wiens *et al.* 2013). In such conditions, the fact that the climatic niche is stable over time, gives species the opportunity to adapt to different regimes – a phenomenon called ‘climatic niche conservationism’ (Bonetti & Wiens 2014). However, Lv et al. (2016) suggest that the Arvicoline rapid radiation is mostly a result of ecological opportunity, where the expansion and adaptation to warmer areas of the Palearctic led to an explosion in the number of lineages, rather than the adaptation to very similar climatic niches. Looking at adaptive variation within this group might bring new clues to the process of its radiation and associating it with environmental changes through time might be the way to solve some of the major basal splits.

Cabrera vole improved phylogeography and mito-nuclear discordance

The Cabrera vole phylogeographic history can now be studied in increased depth given the existence of genomic data. One of the future steps would naturally be the

detection of refugial areas using range expansion theory, to determine the likely origin of the current genetic diversity for various points in time. Coupled with ecological niche modelling and the fossil record, it would be interesting to evaluate if the Cabrera vole's niche was maintained or if it changed, when comparing genetic estimates to the fossil record. This could also help determine the demographic processes behind the detected mito-nuclear discordance, which at this point appears to be driven mostly by demographic factors rather than associated to selective pressures. Further research is needed to explain the observed patterns, especially on the role of selection in the Cabrera vole mitochondria and its historical demography using SNP variation. This could inform on whether the observed patterns are a result of allele surfing or selection.

Design of protected areas accounting for genetic diversity

It would be important to use the knowledge on the current genetic variation of the Cabrera vole and its dispersal ability to design conservation areas that account for future areas of suitability for the species. This would imply determining areas of future climatic suitability and designing corridors that would be managed through time and as the species tracks its niche, so that we guarantee that the most genetic diversity is maintained, but also avoiding connecting populations that would undergo outbreeding depression. Outbreeding could be tested in experimental populations by performing crossings of individuals from different *ESUs* and determine the outcome of hybridisation at those *ex-situ* locations.

Another important aspect would be to perform comprehensive genome wide association studies on different *ESUs* evaluating their adaptive capacity to increasing temperatures and aridity, to make sure that those populations with the highest fitness are maintained in case dramatic management practices are necessary.

The promise of non-invasive genomics

With the non-invasive GBS study we tested the feasibility of using HTS methods in non-invasive DNA pools. Ideally, this protocol should be tested using individual barcoding, i.e. each individual can be identified, so that we can move from population to individual-based non-invasive studies. This would greatly benefit wildlife conservation of non-model organisms as it could be the standard for monitoring natural populations using NiGS.

One by-product of the non-invasive genomics study was the presence of various other non-target reads. As expected for these sample types, bacteria were the non-target

group with the highest representation in both the faecal and the bone DNA pools. In the case of the faecal material, that opens up the opportunity for microbiome work, with opportunities for future analyses relating to comparative phylogeography, population genetics and individual health.

5. References

Albert CH, Rayfield B, Dumitru M, Gonzalez A (2017) Applying network theory to prioritize multi-species habitat networks that are robust to climate and land-use change. *Conservation Biology*, doi: 10.1111/cobi.12943.

Allendorf FW, Luikart G, Aitken SN (2012) *Conservation and the Genetics of Populations*. Wiley-Blackwell, Oxford, UK.

Amarasekare P (1998) Allee effects in metapopulation dynamics. *The American Naturalist*, **152**, 298–302.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.

de Barba M, Miquel C, Lobréaux S *et al.* (2016) High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, **17**, 492–507.

Bastos-Silveira C, Santos SM, Monarca R, Mathias ML, Heckel G (2012) Deep mitochondrial introgression and hybridization among ecologically divergent vole species. *Molecular Ecology*, **21**, 5309–5323.

Bastos AD, Nair D, Taylor PJ *et al.* (2011) Genetic monitoring detects an overlooked cryptic species and reveals the diversity and distribution of three invasive *Rattus* congeners in South Africa. *BMC Genetics*, **12**, 26.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, **9**, 1279–1301.

Cano ARG, Cantalapiedra JL, Mesa A, Moreno Bofarull A, Hernández Fernández M (2013) Global climate changes drive ecological specialization of mammal faunas: trends in rodent assemblages from the Iberian Plio-Pleistocene. *BMC Evolutionary*

Biology, **13**, 94.

Carling MD, Brumfield RT (2008) Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the *Passerina* bunting hybrid zone. *Evolution*, **62**, 2600–2615.

Ceballos G, Ehrlich PR, Barnosky AD *et al.* (2015) Accelerated modern human-induced species losses: entering the sixth mass extinction. *Science Advances*, **1**, e1400253.

Chaline J (1974) Les rongeurs, l'âge et l'environnement de la trés haute terrasse de Grâce à Montières. *Bulletin de l'Association Française pour l'Étude du Quaternaire*, **11**, 151–157.

Conroy CJ, Cook JA (1999) MtDNA evidence for repeated pulses of speciation within arvicoline and murid rodents. *Journal of Mammalian Evolution*, **6**, 221–245.

Conroy CJ, Cook JA (2000) Molecular systematics of a Holarctic rodent (*Microtus*: Muridae). *Journal of Mammalogy*, **81**, 344–359.

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

Cuenca-Bescós G, López-García JM, Galindo-Pellicena MA *et al.* (2014) The Pleistocene history of *Iberomys*, an endangered endemic rodent from south western Europe. *Integrative Zoology*, **9**, 481–497.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–1188.

Eizaguirre C, Baltazar-Soares M (2014) Evolutionary conservation - evaluating the adaptive potential of species. *Evolutionary Applications*, **7**, 963–967.

Escribano N, Ariño AH, Galicia D (2016) Biodiversity data obsolescence and land uses changes. *PeerJ*, **4**, e2743.

Fagan WF, Holmes EE (2006) Quantifying the extinction vortex. *Ecology Letters*, **9**, 51–60.

Fernandes M, Pita R, Mira APP (2008) *Microtus cabrerae*. *The IUCN Red List of Threatened Species 2008: e.T13418A3915236.* Downloaded on 06 May 2017.

http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T13418A3915236.en.

Ferreira CM, Sabino-Marques H, Paupério J *et al.* The applicability of non-invasive genetic sampling for monitoring small mammals' metapopulations. *(in prep)*.

Fink S, Fischer MC, Excoffier L, Heckel G (2010) Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia: *Microtus*): the utility of AFLPs versus mitochondrial and nuclear sequence markers. *Systematic Biology*, **59**, 548–572.

Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, **10**, 2741–2752.

Galewski T, Tilak M, Sanchez S *et al.* (2006) The evolutionary radiation of Arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. *BMC Evolutionary Biology*, **6**, 80.

Garrido-García JA, Rosário IT, Gisbert J *et al.* (2013) Revisión a nivel ibérico de la distribución del topillo de Cabrera o iberón, *Iberomys cabrerae* (Thomas, 1906). *Galemys*, **25**, 35–49.

Garrido-García JA, Soriguer-Escofet RC (2012) Cabrera's vole *Microtus cabrerae* Thomas, 1906 and the subgenus *Iberomys* during the Quaternary: evolutionary implications and conservation. *Geobios*, **45**, 437–444.

Garrido-García JA, Soriguer RC, Serrano DP-A, Pardávila X (2008) A revision of the distribution of Cabrera's vole (*Microtus cabrerae* Thomas 1906) in Andalusia (southern Spain). *Hystrix, Italian Journal of Mammalogy*, **19**, 121–131.

Graham CF, Glenn TC, McArthur AG *et al.* (2015) Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, **15**, 1304–1315.

Gralka M, Stiewe F, Farrell F *et al.* (2016) Allele surfing promotes microbial adaptation from standing variation. *Ecology Letters*, **19**, 889–898.

Hampe A, Petit RJ (2004) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, **8**, 461–467.

Holmes MW, Hammond TT, Wogan GOU *et al.* (2016) Natural history collections as windows on evolutionary processes. *Molecular Ecology*, **25**, 864–881.

IUCN (2016) The IUCN Red List of Threatened Species. Version 2016-3. *Downloaded*

on 07 Dec 2016. <http://www.iucnredlist.org>.

Jaarola M, Martínková N, Gündüz I *et al.* (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **33**, 647–663.

Kraus RHS, Vonholdt B, Cocchiararo B *et al.* (2015) A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Molecular Ecology Resources*, **15**, 295–305.

Lacy RC (1987) Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology*, **1**, 143–158.

Laplana C, Sevilla P (2013) Documenting the biogeographic history of *Microtus cabrerae* through its fossil record. *Mammal Review*, **43**, 309–322.

Latinne A, Waengsothorn S, Rojanadilok P *et al.* (2013) Systematics and biodiversity diversity and endemism of Murinae rodents in Thai limestone karsts. *Systematics and Biodiversity*, **11**, 323–344.

López-García JM, Blain H-A, Morales JI *et al.* (2012) Small-mammal diversity in Spain during the late Pleistocene to early Holocene: climate, landscape, and human impact. *Geology*, **41**, 267–270.

Luque GM, Vayssade C, Facon B *et al.* (2016) The genetic Allee effect: a unified framework for the genetics and demography of small populations. *Ecosphere*, **7**, e01413.

Martínková N, Moravec J (2012) Multilocus phylogeny of arvicoline voles (Arvicolini, Rodentia) shows small tree terrace size. *Folia Zoologica*, **61**, 254–267.

Mestre F, Pita R, Paupério J *et al.* (2015) Combining distribution modelling and non-invasive genetics to improve range shift forecasting. *Ecological Modelling*, **297**, 171–179.

Mills LS (2013) *Conservation of Wildlife Populations: Demography, Genetics, and Management*. Wiley-Blackwell, Oxford, UK.

Mira APP, Marques CC, Santos SM, Rosário IT, Mathias ML (2008) Environmental determinants of the distribution of the Cabrera vole (*Microtus cabrerae*) in Portugal:

implications for conservation. *Mammalian Biology*, **73**, 102–110.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, **19**, 208–216.

Moritz C, Potter S (2013) The importance of an evolutionary perspective in conservation policy planning. *Molecular Ecology*, **22**, 5969–5971.

Paupério J, Herman JS, Melo-Ferreira J *et al.* (2012) Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia. *Molecular Ecology*, **21**, 6015–6032.

Pita R, Beja P, Mira APP (2007) Spatial population structure of the Cabrera vole in Mediterranean farmland: the relative role of patch and matrix effects. *Biological Conservation*, **134**, 383–392.

Pita R, Mira A, Beja P (2006) Conserving the Cabrera vole, *Microtus cabrerae*, in intensively used Mediterranean landscapes. *Agriculture, Ecosystems and Environment*, **115**, 1–5.

Pita R, Mira APP, Beja P (2014) *Microtus cabrerae* (Rodentia: Cricetidae). *Mammalian Species*, **912**, 48–70.

do Rosário IT, Mathias ML (2007) Post-fire recolonisation of a montado area by the endangered Cabrera vole (*Microtus cabrerae*). *International Journal of Wildland Fire*, **16**, 450–457.

Santos SM, Simões MP, Mathias ML, Mira APP (2006) Vegetation analysis in colonies of an endangered rodent, the Cabrera vole (*Microtus cabrerae*), in southern Portugal. *Ecological Research*, **21**, 197–207.

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

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

Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution*, **14**, 323–327.

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

Vale-Gonçalves HM, Cabral JA (2014) New records on the distribution of three rodent species in NE Portugal from barn owl (*Tyto alba*) diet analysis. *Galemys*, **26**, 1–5.

Varela RR (2016) Variaciones en la distribución y la estructura genética de los linces paleárticos y el topillo de Cabrera desde el Pleistoceno Superior. Universidad Complutense de Madrid, Madrid, Spain.

Vera JA (2004) *Geología de España*. Instituto Geológico y Minero de España, Madrid.

Waits LP, Paetkau D (2005) Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, **69**, 1419–1433.

Willoughby JR, Sundaram M, Wijayawardena BK *et al.* (2015) The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biological Conservation*, **191**, 495–503.

Zagorodnyuk I V. (1990) Karyotypic variability and systematics of the gray voles (Rodentia, Arvicolini). Communication 1. Species composition and chromosomal numbers. *Vestnik Zoologii*, **2**, 26–37.