U.PORTO

FC FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

CRYPTIC SPECIATION IN THE FIELD VOLE (*Microtus agrestis*):

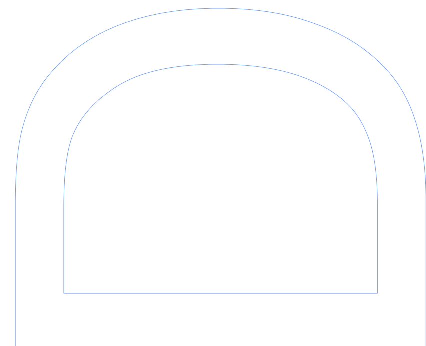
INSIGHTS FROM GENETIC, KARYOTYPIC AND MORPHOMETRIC DATA

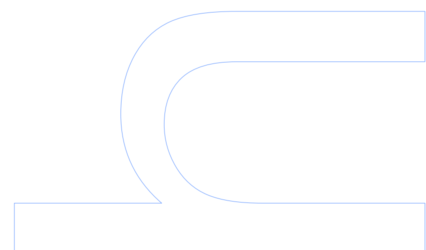
Joana Castro Paupério
2012

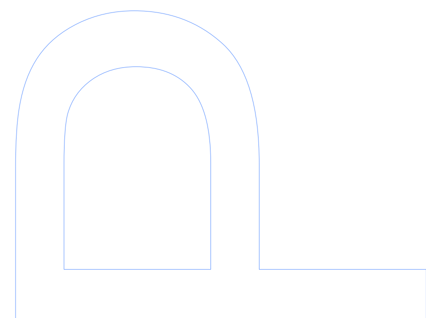
PhD

CRYPTIC SPECIATION IN THE FIELD VOLE (*Microtus agrestis*):

INSIGHTS FROM GENETIC, KARYOTYPIC AND MORPHOMETRIC DATA







Joana Maria Jorge Pereira de Castro Paupério
Doutoramento em Biologia
Departamento de Biologia
2012

Orientador
Paulo Célio Alves, Professor Auxiliar, Faculdade de Ciências da Universidade do Porto

Coorientador
Jeremy Byron Searle, Professor, University of York

Ao Adriano e às minhas três princesas,

Nota prévia:

Na elaboração desta dissertação, e nos termos do número 2 do Artigo 4º do Regulamento Geral dos Terceiros Ciclos de Estudos da Universidade do Porto e do Artigo 31º do D.L. 74/2006, de 24 de Março, com a nova redação introduzida pelo D.L. 230/2009, de 14 de Setembro, foi efetuado o aproveitamento total de um conjunto coerente de trabalhos de investigação já publicados ou submetidos para publicação em revistas internacionais indexadas e com arbitragem científica, os quais integram alguns dos capítulos da presente tese. Tendo em conta que os referidos trabalhos foram realizados com a colaboração de outros autores, o candidato esclarece que, em todos eles, participou ativamente na sua conceção, na obtenção, análise e discussão de resultados, bem como na elaboração da sua forma publicada.

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/27369/2006).

This work was developed in

This thesis should be cited as:
Paupério J (2012) Cryptic speciation in the field vole (*Microtus agrestis*): insights from genetic, karyotypic and morphometric data. PhD Thesis, University of Porto, Porto, Portugal.

ACKNOWLEDGEMENTS

Lots of people contributed to the content of this thesis. Either by giving samples, suggestions or support, or by making me think about things or just by being present... To all I would like to say: Thank you!

I would like to thank my two supervisors, Paulo and Jeremy, without them this thesis would not have been accomplished.

Paulo, já me aturas há bastante tempo… há mais de uma década!!! E, na verdade és totalmente responsável pelo tema desta tese. Foste tu que me puseste o 'bichinho do agrestis atrás da orelha'... mostraste-me os primeiros trabalhos realizados pela Maarit com esta espécie… Obrigada por isso! Obrigada também por me teres acompanhado todos estes anos, por todas as tuas sugestões, ensinamentos, pelo teu companheirismo, apoio e amizade e também pelo entusiasmo que sempre demonstraste com este trabalho e que sempre me deu grande ânimo.

Jeremy, it has been a pleasure to know you and work with you! Thanks a lot for being so welcoming and for sharing with me all your knowledge and experience not only in the lab but also in our field trips. You were my good luck charm for finding voles…The voles seemed to throw themselves into the pool at CIBIO whenever you were in town! I enjoyed a lot my time in York, specially our nice suppers with Sofia's chocolate mousse and our memorable conversations about pudding... ups, sorry, I meant to say the continuous hard work! Thank you for being such a good supervisor!

I would also like to thank the institutions that made this work possible. To FCT, for the attribution of my PhD fellowship. To CIBIO and Professor Nuno Ferrand for the opportunity of developing this study in this excellent research centre. And to the University of York, for welcoming me during my stays.

I also want to express my appreciation to the institutions that gave me the appropriate permits for trapping and collecting tissue samples from field voles both in Portugal and northern Spain: Instituto da Conservação da Natureza e da Biodiversidade, Consejería de Medio Ambiente do Gobierno del Principado de Asturias, Consellería de Medio Ambiente y Desenvolvemento Sostible de Galicia, Servició Territorial de Medio Ambiente de la Comunidad Autónoma de Castilla Léon, Parque Nacional Picos da Europa, Dirección General de Biodiversidad de la Comunidad Autónoma de Cantabria, Departamento de Agricultura de la Deputación Foral de Bizkaia, Departamento de Medio Ambiente de la Diputación Foral de Álava, Dirección General de Medio Natural da Comunidad Autónoma de Cataluña.

The results presented in this thesis wouldn't have been accomplished without the contribution of a great number of people and institutions.

First of all to Maarit Jaarola, thank you for your availability in answering my doubts and sending me some of your samples. It was a pleasure to have you as a co-author of the first paper of the thesis!

To Alejandro Centeno-Cuadros, Alfonso Ceña, Antoni Arrizabalaga and the Museum of Natural Science of Granollers, António Mira and the PERSIST project team, Augusto de Castro, Beatriz e Vasco Alves, Carlos Fonseca, Carlos Nores, Cristianne Bastos-Silveira, Enrique Castiéns, Fernando Jubete, Guilherme Dias, Hélia Gonçalves, Jacinto Róman, José António Godoy, José Manuel Grosso-Silva, Maria José Lopez-Fuster and the University of Barcelona,

the Portuguese National Tissue Bank/National Institute for Nature and Biodiversity Conservation (BTVS/ICNB), Rocio Rosa-García, Sara Afonso, Vanessa Mata, Verónica Gomes and Zbyszek Boratynski, thanks a lot for providing samples for this study! I am also grateful to Maria da Luz Mathias and the University of Lisbon, for lending some field materials. And, as not only people may contribute to a PhD thesis, I would also like to express my gratitude in a special way to Rafi (Paulo's dog), Tomás and Manuchau (Pedro's cats) and Fran (Guilherme's cat).

To the York gang, Mabel, Sofia, Ellie, Rodrigo, Sophie, Natalia, Jerry, Dan, Frida, Bruce, Mike, Nancy and Ollie thank you guys for being so welcoming! Despite the rainy weather, I really enjoyed the time I spend in York!

Jerry, thank you for teaching me the ways around the lab and for getting me started in the molecular analysis. I was very happy that we could collaborate in part of the work developed in the thesis and hopefully we will continue to work together in the next projects. Natalia, the other mum in the lab, with you I shared my family homesickness and I learned a lot, especially how to be efficient in the lab! Thank you for that! Rodrigo, I really enjoyed our field trip in northern Spain. We've seen very nice places and trapped lots of voles, and some shrews… Thank you for all our conservations and discussions, I learned a lot with that. Mabelita, cariño… how nice was your company… I couldn't have done it without you! Thank you for all the moments we spent at the lab and in the city… thanks a lot for keeping me company. Thanks for teaching me all about the karyotypes and working with me on the paper, it turned out to be a very promising work! I really enjoyed your visit to Porto and hope you can come and visit us soon! Sofia, a minha companheira portuguesa nas terras britânicas! Que boas as nossas conversas, os nossos jantares e passeios na cidade… Obrigado por me ajudares a superar as saudades de casa e da minha família! Com a tua companhia tudo foi mais fácil! Obrigado por tudo!!!!

Aos meus colegas do CIBIO, em especial ao gabinete do GEPE mas não só, um muito obrigado. Sónia, Peter spider man, Xico, Francisco, Raquel, Cândida, Ana, Fátima, João, Verónica, Tiago meus colegas de gabinete ou no lab, em um ou outro momento muito obrigado pelas conversas, dicas e pela companhia. Tarroso, obrigada pelas eternas ajudas no arcgis! À Ana Silva, Alexandra, Joana, Rómulo, agradeço a ajuda no trabalho de campo e na análise das egagrópilas. Ao Zef e ao Miguel Carneiro, muito obrigado pelas conversas e discussões sobre species trees e evolução dos *M*icrotus. Zef, a tua participação nessa fase do trabalho foi essencial. À Sandra e à Sara muito obrigado por estarem sempre disponíveis e fazerem sempre todos os possíveis para tornarem o nosso trabalho mais fácil!

Aos colegas do lab da FCUP (Tiago, Sandra, Pedro e Zé Manel) obrigadinha por me terem acompanhado nestes últimos momentos de escrita. Tiago, obrigada pelo apoio, por me responderes às minhas dúvidas, mesmo as mais idiotas, pelos teus treinos de relaxamento…

Às meninas do CTM (e meninos) muito obrigado por me terem acolhido quando precisei de ir analisar os meus crânios. Diana, obrigado por tudo o que me ensinaste, a tua experiência foi fundamental para o sucesso deste trabalho!

À Hélia, por toda a ajuda que me deste no campo, pela amizade e por todas as amostras que me conseguiste, um muito obrigado. És a minha colaboradora nº 1.

Ao Ricardo Pita e ao António Mira (Universidade de Évora) por todo o apoio que sempre me deram. António, obrigada pela disponibilidade e pela maneira como sempre me apoiaste neste trabalho. Ricardo, obrigada por me ensinares a capturar os *Microtus*…as dicas que me deste foram essenciais para o sucesso das minhas saídas de campo.

À Soraia, tem sido um prazer trabalhar contigo! O teu companheirismo, amizade e ajuda foram fundamentais para este trabalho! Obrigado!

Catas, companheira! Demos os nossos primeiros passos na ciência juntas e apesar de termos seguido caminhos um pouco diferentes estiveste sempre presente. Muito obrigado pela tua amizade, pelos momentos de hiperventilação conjunta e por estares sempre disponível mesmo nos momentos menos bons.

A todos os meus amigos, que não vou nomear (eles sabem quem são) muito obrigado por serem quem são e por estarem presentes!

Obrigado, Mariana Sampaio, pelo desenho maravilhoso que fizeste para a minha capa.

E finalmente, aos meus pais, por todo o apoio que sempre me deram! Sem eles não estaria aqui hoje. A toda a minha família pela amizade e ajuda que sempre nos deram.

Ao meu eterno companheiro e às minhas três princesinhas, vocês são a luz que ilumina os meus dias! Muito obrigado por estarem sempre presentes…

RESUMO

A delimitação de 'espécies crípticas', espécies próximas que apresentam uma elevada divergência genética sem diferenciação morfológica aparente, contribui de forma significativa para o aumento do conhecimento relativo à especiação, bem como para a conservação da biodiversidade. Este trabalho incide sobre o rato-de-campo-de-cauda-curta, *Microtus agrestis*, uma espécie de pequeno mamífero com distribuição euroasiática e que apresenta pouca variabilidade ao nível morfológico e cariotípico. No entanto, estudos recentes baseados na análise de DNA mitocondrial e de intrões localizados nos cromossomas sexuais, demonstraram que esta espécie se encontra diferenciada em duas linhagens, com distribuições a norte e a sul da Eurásia e que possuem uma divergência genética suficientemente elevada para que possam ser consideradas duas espécies crípticas. Estes estudos demonstraram ainda a existência de duas linhagens mitocondriais na Península Ibérica, embora baseados num número limitado de amostras.

O objetivo geral deste trabalho consistiu no estudo da história evolutiva do rato-de-campo-de-cauda-curta através da análise de marcadores moleculares de diferentes regiões genómicas. Os padrões geográficos da variabilidade genética, cariotípica e morfométrica foram avaliados, dando-se maior enfâse à Península Ibérica, local onde foi previamente detetada uma elevada divergência ao nível do DNA mitocondrial.

Os resultados obtidos demonstraram que o rato-de-campo-de-cauda-curta consiste de facto num complexo de três unidades evolutivas distintas: Norte, Sul e Português. As análises de coalescência sugerem que a divergência destas três linhagens terá ocorrido durante o último período glaciar, o que indica um processo de especiação rápido. A linhagem Portuguesa terá sido a primeira a diferenciar-se (há cerca de 70 000 anos) e as linhagens do Norte e Sul ter-se-ão separado durante o último máximo glaciar (há cerca de 18 500 anos).

A análise cariotípica de dois indivíduos pertencentes à linhagem Portuguesa revelou a existência de alguma variabilidade na morfologia dos cromossomas sexuais, nomeadamente na estrutura do cromossoma Y e no tamanho dos cromossomas X e Y. Estes resultados sugerem ainda uma possível diferenciação cariotípica entre as diferentes linhagens e realçam o interesse de se realizar uma análise cariotípica mais extensiva das linhagens Portuguesa e do Sul.

O estudo detalhado das populações ibéricas de *M. agrestis* através de métodos de genética não-invasivos (análise de crânios provenientes de egagrópilas e de amostras de museus) permitiu conhecer com maior rigor a distribuição espacial das linhagens do Sul e Portuguesa na Península Ibérica. Este trabalho revelou também a existência de uma extensa área de introgressão de marcadores nucleares e mitocondriais entre as duas linhagens no nordeste de Espanha. Contudo, a introgressão observada é assimétrica, indiciando um possível movimento da zona hibrida e, por outro lado, foi detetada a existência de mecanismos de isolamento pós-zigótico, evidenciada sobretudo pela não ocorrência de introgressão nos machos.

A forma das mandíbulas utilizadas no estudo molecular acima descrito, foi analisada através de métodos de morfometria geométrica, tendo revelado a existência de diferenças significativas entre as linhagens Portuguesa e do Sul. Este facto sugere que, apesar destas linhagens serem crípticas, evidenciam já uma diferenciação morfométrica subtil em estruturas como a mandibula.

A análise extensiva realizada na Península Ibérica foi possível devido à incorporação de amostras não-invasivas. De forma a otimizar a incorporação deste tipo de amostras desenvolveu-se um método de identificação molecular dos roedores da Península Ibérica. Este método permite a identificação inequívoca de todas as espécies de roedores ibéricos, com a exceção das espécies próximas, *Microtus lusitanicus* e *M. duodecimcostatus*. A utilização deste método irá potenciar os estudos de distribuição, dinâmica populacional e de conservação da biodiversidade nesta importante região biogeográfica.

Em resumo, todos os dados analisados suportam a existência de três linhagens de *Microtus agrestis* caraterizadas por uma elevada diferenciação genética (em todas as regiões genómicas), pela existência de alguma variabilidade morfológica e cariotípica, bem como pela quantidade limitada de fluxo genético nas áreas de contacto identificadas. Face a estas evidências e de acordo com alguns conceitos de espécies, estas linhagens podem mesmo ser consideradas como espécies distintas. De qualquer forma, e apesar de ser desejável a recolha de informação adicional tanto ao nível cariotípico e morfológico como genómico, estas três linhagens devem ser reconhecidas como Unidades Evolutivas Significativas para efeitos de conservação.

Palavras-chave: Pequenos mamíferos, roedores, espécies crípticas, história evolutiva, análise multilocus, evolução cromossómica, isolamento reprodutivo, morfometria geométrica, métodos não-invasivos.

SUMMARY

The delimitation of ‘cryptic species’, sister forms that show substantial genetic differentiation without apparent morphological variation, is important for understanding speciation and is also relevant for biodiversity conservation. This thesis focuses on the field vole, *Microtus agrestis*, a Eurasian mammal with little morphological and karyotypic differentiation but which, on the basis of previous molecular studies of mitochondrial DNA and X and Y chromosome introns, was subdivided into a Northern and a Southern lineage. In the earlier work these lineages were considered sufficiently divergent to possibly represent two cryptic species and there were indications of further subdivision of the Southern lineage within Iberia, although based in few samples.

The general goal of this thesis was to study the evolutionary history of the field vole focusing on the major subdivisions previously detected and using a multilocus approach. The geographic patterns of genetic, karyotypic and morphometric variation were thoroughly investigated, with a stronger emphasis than previously on the Iberian Peninsula where high genetic variability had been detected and available information was limited.

The main results obtained here reveal that the field vole is in fact a complex of three distinct evolutionary units with neighboring geographic distributions in Eurasia, which may represent three cryptic species: Northern, Southern and Portuguese. Divergence among these lineages was inferred to have occurred during the last glacial period indicating a rapid pace of speciation. The Portuguese lineage split occurred first (at ca. 70 000 years BP) and the separation of the Northern and Southern lineages is dated to the Last Glacial Maximum (at ca. 18 500 years BP).

The analysis of the Portuguese lineage karyotypes showed some variation in sex chromosome morphology relative to previously described Northern lineage karyotypes, namely in the structure of the Y chromosome and in the size of both X and Y chromosomes. These results point to a possible karyotypic differentiation between the lineages and stress the interest in performing an extensive analysis of karyotypic variability within the Portuguese and Southern lineages.

A landscape-scale analysis of Iberian field vole populations through noninvasive genetic sampling (bones from barn owl pellets and from museum collections) revealed the existence of a wide area of introgression of nuclear and mitochondrial loci between the Southern and Portuguese lineages, suggesting possible movement of the hybrid zone formed on initial contact of these lineages. This introgression was found to be

asymmetric, with a higher introgression of Southern lineage alleles into the Portuguese lineage than vice versa, suggesting a particularly large movement of the zone in favour of the Portuguese lineage. In addition, evidence of post-mating isolation mechanisms was also found; the complete absence of hybrid males suggests strong selection against them.

Analysis of the variability of mandible shape through geometric morphometrics revealed the existence of significant differences between the Portuguese and Southern lineages. These results suggest that these forms, although essentially ‘cryptic’, show some subtle morphometric differentiation in mandible shape.

The extensive and detailed landscape-scale analysis of field voles in the Iberian Peninsula was only possible due to the use of noninvasive sampling. To optimize the analysis of this type of samples, we first developed a genetic method for identifying the Iberian rodent species, which could be applied to noninvasive samples. This method allowed the unambiguous identification of all Iberian rodents with the exception of the sibling species *Microtus lusitanicus* and *M. duodecimcostatus*. The genetic method can be used in studies of distribution, population dynamics and for biodiversity conservation in this important biogeographical region.

Overall, the data in this study show consistent results, identifying three lineages of the field vole, genetically distinct (in all genomic regions), with some karyotypic and morphological variability and a limited amount of gene flow within the areas of contact of the lineages. Although collecting additional information both on karyotypic and morphological variability as well as genomic data would nonetheless be desirable, these three lineages should already be recognized as Evolutionarily Significant Units for conservation purposes.

Keywords: Small mammals, rodents, cryptic species, evolutionary history, multilocus dataset, chromosomal evolution, reproductive isolation, geometric morphometrics, noninvasive genetic sampling.

TABLE OF CONTENTS:

RESUMO i

SUMMARY iii

TABLE OF CONTENTS v

LIST OF FIGURES vii

LIST OF TABLES xi

CHAPTER 1 – General Introduction 1

1.1. Speciation and species delimitation 3
- 1.1.1. 'Cryptic' speciation 3
- 1.1.2. Chromosomes and speciation 4
- 1.1.3. Gene flow and speciation 6
- 1.1.4. Species delimitation 9

1.2. Molecular tools 10
- 1.2.1. Noninvasive sampling 12

1.3. Geometric morphometrics 13
1.4. The Eurasian field vole 14
1.5. Objectives and structure of the thesis 17
1.6. List of papers 19
1.7. References 19

CHAPTER 2 – Evolutionary history of the field vole in Eurasia 29

Paper I - Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia 31

Paper II - Giant sex chromosomes retained within the Portuguese lineage of the field vole (*Microtus agrestis*) 67

CHAPTER 3 – Differentiation of the two lineages of the field vole in Iberia: insights from molecular and morphometric data 77

Paper III - Asymmetric introgression between the two lineages of the field vole in the Iberian Peninsula revealed by noninvasive genetic sampling 79

Paper IV - Morphometric differentiation of the two field vole lineages in the Iberian Peninsula: a geometric analysis of mandible shape 117

CHAPTER 4 – A method for genetic identification of Iberian rodents 141

Paper V - Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling 143

CHAPTER 5 – General Discussion 187

5.1. Three cryptic lineages of the field vole in Eurasia 189
5.2. The evolutionary history of *Microtus agrestis sensu lato* 191
5.3. Multiple evidence suggest the revision of species status 193
5.4. The value of noninvasive sampling 195
5.5. The field vole as a model system 196
5.6. Future research prospects 198
5.7. References 199

CHAPTER 6 – Conclusions 205

LIST OF FIGURES:

CHAPTER 1

Fig. 1 Distribution range (green) of the field vole (*Microtus agrestis*). 15

CHAPTER 2

Paper I

Fig. 1 Map showing in yellow shading the distribution range of the field vole (*Microtus agrestis*) according to IUCN (2010), as well as the geographical location of the sampling localities (filled symbols) and of the sequences retrieved from GenBank (empty symbols). The locality numbers refer to those in Table 1. Symbols indicate the lineages identified in the gene trees (squares – Northern; circles – Southern; triangles – Portuguese). 34

Fig. 2 Phylogenetic tree based on Bayesian inference showing the relationships among field voles for the cytochrome b gene (n = 412). Values on branches indicate posterior probability support and bootstrap values of the maximum likelihood and neighbour joining analysis. Lineages are coloured according to their geographic distribution, as in Fig. 1 41

Fig. 3 Median-joining networks of the X-chromosome introns (DBX5 and DBX6), Y-chromosome introns (DBY7 and SCMY7) and autosomal genes (BRCA1 and RAG1) in *M. agrestis* (number of sequences for each locus in Table 2). *M. arvalis* included as outgroup (n = 1). Each circle represents one haplotype and circle area is proportional to the frequency of each haplotype. Total frequency is indicated for more common haplotypes. Branches are proportional to the number of nucleotide differences between haplotypes and dots on branches indicate mutational steps. Colours indicate geographic distribution of the lineages and correspond to those used in Fig. 1. 42

Fig. 4 Phylogenetic trees based on Bayesian inference showing the relationships among field voles for the nuclear genes. Values on branches indicate posterior probability support for clades and bootstrap support for the maximum likelihood and neighbour joining analyses respectively. The lineages are coloured according to their geographic distributions, as in Fig. 1. 44

Fig. 5 Consensus species trees: (a) *BEAST species tree inference output for all 7 loci analysed. The posterior probability of each clade is presented on each node and grey bars show 95% highest posterior density intervals for times to most recent common ancestor (tMRCA in substitutions per site; values presented below bar) of each lineage. (b) Consensus species tree inferred by ML in STEM for all loci for a θa value of 0.0118 (see text). Grey bars show 95% highest posterior density intervals of branch lengths in coalescent units (number of 2Ne generations). Bootstrap support (above) and branch lengths (mean and 95%HPD, below) are shown for each node. 46

Fig. 6 Extended Bayesian Skyline plots of the effective population size through time estimated from 1450 million post-burn-in generations from three MCMC simulations. The accumulation of thin lines gives a view of the full posterior distribution. Dashed black line is the median effective population size Ne, in thousands, multiplied by one (mean generation time in years). Solid black lines are the 95% highest posterior density 49

limits. The y-axis is presented on a log scale for clarity. (a) – Northern lineage; (b) – Southern lineage; (c) – Portuguese lineage.

Paper II

Fig. 1 Localities where field voles have been karyotyped in Europe (adapted from Fredga and Jaarola 1997). Yellow shading shows distribution of the field vole. Colouring of symbols indicates inferred molecular lineage, based on maps in Jaarola and Searle (2002, 2004). Blue: Northern lineage, standard karyotype. White: Northern lineage, Lund Y karyotype. Red: Southern lineage, standard karyotype. Star: our new Portuguese samples. Note that the actual density of sampling sites in southern Scandinavia is about double that shown here (Fredga and Jaarola 1997). 70

Fig. 2 Karyotypes of the female (a) and male (b) field vole from Portugal. 71

Fig. 3 Individual Y chromosomes from different metaphases from the Portuguese male field vole. Arrow indicates an example of the distinct short arm on the Y chromosome. 72

CHAPTER 3

Paper III

Fig. 1 Median-joining networks of the molecular markers analysed in the Portuguese and Southern lineages of *M. agrestis* (number of sequences for each locus in Table 1): a) Cytb gene; b) BRCA1 gene; c) DBX5 intron on the X-chromosome; d) DBY7 intron on the Y-chromosome. Each circle represents one haplotype and circle area is proportional to the frequency of each haplotype. Total frequency is indicated for the most common haplotypes. Branches are proportional to the number of nucleotide differences between haplotypes and dots on branches indicate mutational steps. Black is for Southern lineage haplotypes and white for Portuguese lineage haplotypes. 88

Fig. 2 Population structure inferred by Geneland, under the UFM, for the field vole in the Iberian Peninsula (K=2). Dots represent sampling localities coloured according to the modal membership to each cluster (blue: Portuguese; red: Southern). The size of the dots is proportional to the number of samples analysed. X represents localities where we were able to amplify only one locus, which were not used in the analysis. The colour gradient refers to the posterior probability of assignment to each of the two clusters. 90

Fig. 3 Sampling localities in the Iberian Peninsula. Each pie chart represents one locality and the area is proportional to the number of samples analysed. Colours indicate the frequencies of each lineage, considering the lineage assignment at all molecular markers analysed: black: Southern lineage; white: Portuguese lineage; grey: hybrid individuals. X represents localities where we were able to amplify only one locus. The dashed line corresponds to the one dimensional transect designed considering the field vole distribution in Iberia. The green shading corresponds to the distribution range of the field vole (*Microtus agrestis*) in the Iberian Peninsula according to Kryštufek et al. (2008). 91

Fig. 4 Detail of the frequencies of the lineage typical genetic markers per locality in the hybrid zone: a) Cytb gene; b) BRCA1 gene; c) DBX5 intron on the X-chromosome; d) DBY7 intron on the Y-chromosome. Each pie chart represents one locality and the area is proportional to the number of samples analysed. Colours indicate the assignment of the individuals to each lineage. Black: Southern lineage; white: Portuguese lineage. 91

Fig. 5 Southern lineage allele frequencies plotted against geographic position on the transect (Km) for the four markers analysed (only localities between km 200 and 800 of 92

the transect were plotted for ease of visualisation). Dots correspond to the observed frequencies and lines to the cline fitted; c is the cline centre and w the cline width.

Paper IV

Fig. 1 Sampling localities in the Iberian Peninsula. Each circle represents one locality and the circle area is proportional to the number of samples analysed. Colours of the circle indicate the lineage assignment according to the existing molecular data: white - localities where only individuals from the Portuguese lineage were identified; black - localities where only individuals from the Southern lineage were identified; grey - localities where hybrids were identified. The four geographic regions used in the analysis are identified by coloured contours: Blue - Portugal; Purple - northwestern Spain; Brown - Burgos, Basque Country and Navarra; Green - northeastern Spain (see Methods for details). The green shading corresponds to the distribution range of the field vole (*Microtus agrestis*) in the Iberian Peninsula according to Kryštufek et al. (2008). 121

Fig. 2 Landmark configuration on the external view of the field vole right side mandible 122

Fig. 3 Box-plot of mandible CS variation (median and percentiles, 25-75%, as well as maximum and minimum values) in the two lineages of field vole (Portuguese, n= 35; Southern, n=56) and hybrid individuals (n=4) by sex (Females, n=56; Males, n=39). 124

Fig. 4 Ordination of the samples in the space of the first two shape variables (RW1 and RW2) based on the matrix of shape variables for the full data-set (n=226). a) Samples coloured according to geographic regions. PT - blue; NW - purple; BBN - brown; NE - green. b) Samples grouped according to the lineage assignment. Circles correspond to individuals for which genetic information exists and diamonds to samples that were not sequenced. Symbols are coloured according to the lineage assignment (circles) or populations of origin (diamonds). White - Portuguese lineage; Black - Southern lineage; Grey - hybrids. Each symbol represents one individual. Two samples from Burgos and other two samples from central Portugal are not shown as they represent outliers to the range of values of RW1 and RW2 shown in this scatter plot. 126

Fig. 5 Ordination of the samples according to the two discriminant functions (DF1 and DF2) obtained in the analysis between geographic regions. Circles are coloured by region. PT: blue; NW: purple; BBN: brown; NE: green. 127

Fig. 6 Frequency histogram of discriminant function scores (DF) obtained in the analysis between lineages. a) Considering the dataset of genotypic individuals (n=111). Bars are coloured according to the lineage assignment. White - Portuguese lineage; Black - Southern lineage. b) and c) TPS deformation grids showing the shape changes at DF. b) represents the extreme configuration for the Portuguese lineage and c) for the Southern lineage. The deformations were magnified three times for clarity. The mandibles in the deformation grids are stylised representations obtained by linking the pairs of landmarks (see Fig. 2). d) Considering all samples (n=226). Bars are coloured according to the lineage assignment of the population of origin. White - Portuguese lineage; Black - Southern lineage; Grey - hybrid populations. 128

Fig. 7 Population structure inferred by Geneland for the morphometric data (n=226). Each circle represents one locality and the circle area is proportional to the number of samples analysed. Colours of the circle indicate the modal assignment of the individuals to each cluster. The four clusters are identified by different colours. The green shading corresponds to the distribution range of the field vole (*Microtus agrestis*) in the Iberian Peninsula according to Kryštufek et al. (2008). 130

Fig. 8 Population structure inferred by Geneland for the dataset including the genetic 131

and morphometric data (n=327). Each circle represents one locality and the circle area is proportional to the number of samples analysed. Colours of the circle indicate the modal assignment of the individuals to each cluster. The black and white colour refers to the two clusters identified in the analysis. The green shading corresponds to the distribution range of the field vole (*Microtus agrestis*) in the Iberian Peninsula according to Kryštufek et al. (2008).

CHAPTER 4

Paper V

Fig. 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.* 152

Fig. 2 (a) Histogram of K2P cyt-b divergence values (intraspecific, interspecific within families, interspecific among families), for 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. (b) Summary of pairwise divergences involving sequences of each species showing mean (pale dot) and maximum (associated dash) intraspecific divergences and mean (dark square) and minimum (associated 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. 153

Fig. 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.* 155

Fig. 4 a) Histogram of K2P IRBP divergence values (intraspecific, interspecific within families, interspecific among families), for 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. (b) Summary of pairwise divergences involving sequences of each species showing mean (pale dot) and maximum (associated dash) intraspecific divergences and mean (dark square) and minimum (associated 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. 156

Fig. 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. 157

LIST OF TABLES:

CHAPTER 2

Paper I

Table 1 Details of localities sampled for the field vole (*Microtus agrestis s.l.*) in this study and number of samples sequenced for each locus in each locality. 35

Table 2 Species delimitation based on clocklike loci and for all loci, considering an average theta value for all lineages (θa) of 0.0118 (see text and Table S2, Supporting information). 43

Table 3 Divergence between the three field vole lineages (standard errors based on 10 000 replicates are given in parentheses; all estimates are presented as percentages). 47

Table 4 Levels of polymorphism at each locus in the three lineages of field vole and for *Microtus agrestis s.l.* 48

CHAPTER 3

Paper III

Table 1 Levels of polymorphism at each locus in both lineages of field vole. 89

Paper IV

Table 1 Effect of age, sex and of the interaction sex x age on mandible centroid size considering the dataset of genotyped individuals (n=95). 125

Table 2 Effect of sex, lineage and of the interaction sex x lineage on the 10 shape variables considering the dataset of genotyped individuals (n=95). 125

CHAPTER 4

Paper V

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. 147

CHAPTER 1

GENERAL INTRODUCTION

1.1. SPECIATION AND SPECIES DELIMITATION

Speciation, the process by which species arise, is the force that drives biodiversity. More than 150 years has passed after the publication of Darwin's book, *On the Origin of Species*, and, although there is a large amount of literature concerning species and speciation, understanding speciation processes still remains a major challenge for evolutionary biologists. Indeed, as Hendry (2009) has stated "In speciation, as in other fields of study, it seems that the more we know, the more we know we don't".

1.1.1. 'CRYPTIC' SPECIATION

Speciation can be considered the process of development of reproductive isolation as well as phenotypic, behavioural and molecular differences between populations through time. The differentiation process however, does not occur at all levels at the same time. Sometimes, reproductive isolation and molecular differences are observed without any perceivable morphological changes, in a process that can be referred to as 'cryptic speciation'. 'Cryptic species' are forms that would previously have been described as single taxa due to the lack of visible morphological differentiation.

Probably due to the fact that most sensory information processed by the human brain is visual, morphological characteristics have been traditionally the basis of the identification and description of species. However, over the last two decades other methods for delimiting species have been developed and integrative approaches that combine morphological with molecular, ecological and developmental methods are being used (Dayrat 2005). In fact, the advances of molecular tools, has led to an exponential growth in the discovery of 'cryptic species', even in well studied taxonomic groups (birds: *e.g.* Olsson *et al.* 2005; bats: *e.g.* Ibañez *et al.* 2006; rodents: *e.g.* Gündüz *et al.* 2007), which has important consequences in the assessment of biodiversity and in conservation planning (Bickford *et al.* 2007, Pfenninger and Schwenk 2007).

Despite this boost in the detection of 'cryptic species' the mechanisms behind their formation are still mostly unknown. Cryptic species are mostly thought to be of recent origin, such that morphological and other traits have not yet evolved. However, there are examples of cryptic species that show apparently ancient divergences

(Šlapeta *et al.* 2006; Bickford *et al.* 2007). Morphological stasis upon speciation may be also explained by directional selection on ecological, behavioural or reproductive traits that have no known morphological correlates (Bickford *et al.* 2007). For instance, Schröngge *et al.* (2002) in their study of myrmecophilous hoverflies state that specialist parasite species that exhibit behavioural or physiological adaptations to the host may be expected not to show morphological variation among species. Incongruence between genetic and morphological divergences is often also explained by convergence in morphological characters, like the ones related, for example, with feeding habitats. In a morphometric study of three taxa of bats (*Myotis myotis*, *M. blythi*, *M. punicus*), Evin *et al.* (2008) found that although *M. myotis* and *M. blythi* were considered sister taxa, *M. myotis* had a high similarity with *M. punicus* in skull morphology, which they interpret as morphological convergence due to similarity in feeding habits. Also, taxa that use reproductive signalling through other ways than visual may be more prone to contain cryptic species, since changes in reproductive traits do not imply morphological variation. Overall, there is still plenty to learn about 'cryptic speciation'. It is even possible that characters previously considered of minor importance in the formation of species, will be highlighted as relevant due to the study of 'cryptic speciation' (Bickford *et al.* 2007).

Finally, although initial data suggested otherwise (Bickford *et al.* 2007), cryptic speciation seems to be occurring relatively homogeneously across taxa and biogeographical regions, which suggest that morphological stasis upon speciation may represent an evolutionary constant independent of phylogenetic relations or ecological circumstances (Pfenninger and Schwenk 2007).

1.1.2. CHROMOSOMES AND SPECIATION

Most animal and plant species vary greatly in their karyotypes both in the number and structure of the chromosomes. In fact, looking at the recent Atlas of Mammalian Chromosomes (O'Brien *et al.* 2006) we can observe a large variation in the number of chromosomes in mammals, which ranges from 6 to 102. This variation remains substantial even if we narrow it to rodent karyotypes (from 15 to 102 chromosomes; O'Brien *et al.* 2006). In addition to this interspecific variability in karyotypes, there are also differences in chromosome numbers and structure within species. Chromosomal races are documented in several species, from plants (*e.g. Carex pachystachya*, Whitkus 1988) to invertebrates (*e.g. Podisma pedestris*, Barton and Hewitt 1981), fish (*e.g. Fundulus notatus*, Black and Howell 1978), reptiles (*e.g. Sceloporus grammicus*,

Arèvalo *et al.* 1994) and mammals (*e.g. Mus musculus*, Nachman and Searle 1995; *Sorex araneus*, Zima *et al.* 1996).

This diversification in karyotypes is mostly due to chromosomal rearrangements, but differences in chromosome numbers may also originate through duplication of whole chromosome sets, that is, polyploidy. Polyploidy has been considered to play an important role in plant evolution (Tate *et al.* 2005). However, it has been traditionally regarded as relatively unimportant in animal evolution, since it is a less frequent situation (see Otto and Whitton 2000; Mable 2004; Gregory and Mable 2005 for a review). Nevertheless it has already been reported for several vertebrates (*e.g.* anuran amphibians, Brunes *et al.* 2010; rodents, Gallardo *et al.* 2004). In fact, Otto and Whitton (2000) suggest that the occurrence of polyploidy may have extensively influenced the tempo and mode of evolution.

Chromosomal rearrangements, like translocations, fusions, fissions, inversions or heterochromatin additions or deletions, lead to a diversification in karyotypes and are considered likely to be important in speciation as they can contribute to reproductive isolation between populations. These chromosomal rearrangements can lead to hybrid sterility if hybrid meiosis is highly disturbed. Moreover, if the influence on hybrids is less extreme, these differences in karyotypes may still have negative effects on gene flow and therefore promote genetic divergence between populations (Searle 1998). Rieseberg (2001) considers that this reduction of gene flow may be caused by suppression of recombination in heterozygotes for the rearrangements around the position of the rearrangement.

As mentioned above, there is much variability in the karyotypes of mammals, and more specifically of rodents, both between and within species. One of the species that presents a high variability in chromosome number (due to centric fusions) is the house mouse (*Mus musculus*) where it has been demonstrated that the accumulation of chromosomal rearrangements may lead to reproductive isolation between populations (Piálek *et al.* 2005). Moreover, recent work by Lemskaya *et al.* (2010) that investigated the karyotypic variability of eight *Microtus* species and compared it with available phylogenetic information indicates that chromosomal rearrangements have played an important role in the speciation of voles.

Another striking example of karyotypic variability is the increased amount of repetitive DNA (heterochromatin) that has been detected in the chromosomes of several mammals. This variation in presence/absence and size and positioning of the heterochromatic blocks in the chromosomes has been detected both between and within species. One particularly well-studied example of this variability within species is the western hedgehog (*Erinaceus europaeus*) where different populations in Western

Europe have large blocks of heterochromatin on different autosomes (Mandahl 1978; Searle and Erskine 1985). One other group with remarkable examples of between and within species chromosomal variability is the genus *Microtus*. Within this genus, several species have enlarged sex chromosomes (known as 'giant sex chromosomes'). Other species, although not having 'giant sex chromosomes' show different amounts of heterochromatin on the X and Y chromosomes (Modi 1987a, Borodin *et al.* 1995). The most notable examples within this genus are *M. cabrerae* and *M. agrestis*, as the latter possesses the largest heterochromatin blocks on the sex chromosomes of any vole (Nanda *et al.* 1988; Marchal *et al.* 2004).

This variability and often large quantity of sex chromosome heterochromatin in *Microtus* voles is associated with asynapsis of the sex chromosomes during male meiosis, which has been confirmed in many species (Borodin *et al.* 1995, 2012). However, not all voles with sex chromosome asynapsis have giant sex chromosomes which indicates that this meiotic behaviour is not caused by the accumulation of heterochromatin, instead it seems that sex chromosome asynapsis may 'allow' for an accumulation of heterochromatin which would otherwise interfere with the meiotic process (Ashley *et al.* 1989; Borodin *et al.* 2012). Recent work suggests that X-Y synapsis at meiosis should be the ancestral condition in microtine voles (Borodin *et al.* 2012). However, these authors also found that X-Y synapsis has been lost independently at least three times during the evolution of microtine rodents which seems to indicate that this lack of ability of pairing by sex chromosomes does not follow any phylogenetic trend (Borodin *et al.* 2012).

1.1.3. GENE FLOW AND SPECIATION

As mentioned above, speciation is usually considered as a process of evolution of reproductive isolation between taxa that diverge at genetic, phenotypic and/or behavioural levels. Traditionally it was thought that this process of divergence between populations would mostly occur in allopatry (between geographically isolated populations), eventually causing the emergence of genetic incompatibilities that would prevent gene flow between the divergent taxa. However, it has been demonstrated that speciation is also possible in sympatry (without geographic separation between taxa), as Darwin had already proposed with his principle of divergence, and some putative examples are now widely accepted (Bolnick and Fitzpatrick 2007). In fact, considerable attention has recently been given in the literature to the modes and factors that facilitate the occurrence of divergence with gene flow (Bolnick and Fitzpatrick 2007;

Pinho and Hey 2010; Smadja and Butlin 2011). Even so, most biologists continue to argue that sympatric speciation is relatively rare (Hendry 2009).

When two populations become geographically isolated, genetic drift coupled with local selective pressures may lead to their differentiation. There will then be a time in this divergence process when the two populations will have differentiated to a point that, if the geographic barrier that isolates them is removed, they would not merge. Though some degree of gene flow would be possible at this point, the process of speciation would continue until full reproductive isolation would be achieved, in a process that can be considered as parapatric speciation. Nonetheless, the removal of the isolating barrier could happen at various points of the divergence process leading to an area of contact that could go from a situation of no interbreeding to one of complete admixture, which could result in two merging populations. Therefore, the study of areas where divergent taxa make contact and exchange genes - secondary contact zones - is fundamental for the understanding of speciation processes as well as for the assessment of the degree of reproductive isolation between populations.

Where there is hybridisation in secondary contact zones these are called hybrid zones (Barton and Hewitt 1985). Hybrid zones can reflect a wide variety of stages of speciation (Jiggins and Mallet 2000). Their structure depends on the degree of genetic and ecological differentiation of the hybridising populations, as well as on rates of dispersal and on the fitness of the hybrids (Harrison 1993). For instance, if the two divergent taxa interbreed in a narrow area, with a width that results from the balance between dispersal abilities of each population and of selection acting on hybrid individuals, this area of contact will represent a tension zone (Barton and Hewitt 1985). In this case, the genetic integrity of the two populations is maintained over most of the distribution area.

There are various parameters that can be estimated in a hybrid zone that are important in assessing the degree of reproductive isolation between the taxa/populations involved. For instance, the distribution of genotypic classes within a hybrid zone can be used for its classification as 'unimodal' or 'bimodal' (Jiggins and Mallet 2000). In 'unimodal' hybrid zones intermediate hybrid genotypes predominate, while in 'bimodal' hybrid zones populations consist mainly of individuals similar to the parental populations, with relatively few hybrids. However, bimodal hybrid zones also display considerable variation in the overall abundance of hybrid genotypes as well as in the relative frequency of F1 and backcross types of hybrids. For example, Arias *et al.* (2008) detected around 25% of hybrid individuals in mixed populations of two subspecies of the butterflies *Heliconius erato*, while between hybridizing populations of *Triturus* newts, only 1.3% of genotypes were hybrid (Arntzen and Wallis 1991). When

comparing the characteristics of unimodal and bimodal hybrid zones, Jiggins and Mallet (2000) argue that many bimodal hybrid zones show strong prezygotic isolation due to assortative mating, though postzygotic barriers to gene flow seem to be equally found in both unimodal and bimodal hybrid zones. Anyway, the detection of bimodality in a hybrid zone usually indicates that the speciation process is advanced (Jiggins and Mallet 2000).

Another feature of hybrid zones that is worth mentioning is that differential selection on different traits may lead to discrepancies in the introgression of genes. In fact, differential introgression is often related with the mode of inheritance of the genes. For example, in birds, autosomal markers usually show more introgression than maternally inherited loci in hybrid zones (*e.g.* Secondi *et al.* 2006; Martinez-Cruz and Godoy 2007). However, in mammals, mitochondrial DNA (maternally inherited) is frequently more introgressed than autosomal markers in hybrid zones, which in turn show more introgression than the Y chromosome (paternally inherited; *e.g.* Dod *et al.* 1993, Balloux *et al.* 2000). These differential patterns of introgression of the sex chromosomes are often related to Haldane's (1992) rule that states that "when in the F1 offspring of the two animal races one sex is absent, rare, or sterile, that sex is the heterogametic sex". Therefore, the loci inherited only from the heterogametic sex, which is the female in birds and the male in mammals, are expected to be less introgressed. In addition, asymmetries in introgression are often documented between populations, with the genetic traits of one population introgressing more into the other. For example, mtDNA introgresses often only in one direction, into one of the populations (*e.g.* Berthier *et al.* 2006; Alves *et al.* 2008; Boratyński *et al.* 2011). This asymmetric introgression of different genes may be caused by different factors, namely differences in population structure, fitness, cline movement or selection (Barton and Hewitt 1985). Currat *et al.* (2008) showed by spatially explicit simulations that massive introgression of neutral genes occurs during the invasion of an occupied territory if gene flow is not severely impeded. Actually, demographic fluctuations have been demonstrated to create patterns of substantial asymmetric introgression in three species of hares (Melo-Ferreira *et al.* 2005). However, the hypothesis that selection could have also favoured the massive introgression observed in hares should also not be neglected (Melo-Ferreira *et al.* 2007). Furthermore, it has also been demonstrated that sex biased dispersal can account for asymmetries in introgression (Petit and Excoffier 2009).

Hence, the analysis of gene flow between two divergent populations within a hybrid zone may not only help clarify the stage of the evolutionary process but also

contribute to the understanding of the demographic and selective processes that shaped speciation of those taxa.

1.1.4. SPECIES DELIMITATION

The literature on species definitions is quite extensive reflecting the long lasting debate on this subject. For some authors, species are real entities in nature, not subjective to human divisions, and have an unambiguous definition, unlike lower or higher taxonomic ranks (*e.g.* Coyne 1994; Coyne and Orr 2004). Others, however, agree with Charles Darwin's view expressed in *On the Origin of Species*, that although species are natural groups recognised by consistent gaps (in morphology), there is a continuum with varieties within species, which do not show such gaps. In this sense, although the entities that are called species, biological groups of distinct lineages with potentially independent futures, are real, the term 'species' as a level of biological classification is ambiguous and amorphous, being considered a man-made grouping category developed for the ease of communication among biologists (Mallet 2008; Hendry 2009).

Apart from these different views, more than two dozen species definitions have been published and although none is unanimously accepted, one of the most frequently applied definitions is Mayr's (1963) Biological Species Concept (BSC). According to the BSC, a species consists of a network of populations of potentially interbreeding organisms that are reproductively isolated from other such groups of populations. Reproductive isolation is clearly important in the formation and maintenance of sexually distinct taxa; however, often taxa that can and do hybridise are recognizable as different from morphological, genetic, behaviour and ecological data (Mallet 2008). Actually, some estimates state that about 25% of all plant species and 10% of all animal species hybridise successfully with at least one other species (Hendry 2009). This contributes to the fact that, at present, most evolutionary biologists rely on a more relaxed BSC, considering that different groups are considered different species if they can maintain their genetic integrity in nature, that is if they can remain distinct, which goes along with the ideas expressed by Darwin (Mallet 2008; Hendry 2009).

Regardless of the species definition considered, delimiting species boundaries is an important task in the context of understanding many evolutionary mechanisms and processes and in biodiversity assessments. In fact, the recognition of nascent

divergent lineages and the designation of Evolutionarily Significant Units (ESU, Moritz, 1994), which might even be later considered as different species, are fundamental in the definition of priorities for biodiversity conservation and management.

In the last decade, species delimitation has emerged as a major topic in modern systematics. Several methods have been proposed, which handle different types of information (morphological, genetic, geographical, or a combination) and are applicable to different situations (*e.g.* bisexual taxa vs bisexual and asexual taxa; Sites and Marshall 2003; Wiens 2007). However, as these authors mention, most of the methods developed are strongly based on the use of genetic markers and although some highlight the need of corroboration with other types of information, only few methods explicitly include morphological data. In fact, molecular data has become a very important tool for the purpose of species delimitation. Among the different methods described, some focus on the inference of the presence/absence of gene flow, while others rely on the congruence of phylogenetic tree topologies (*e.g.* monophyly, concordance with geography, concordance with morphology; Sites and Marshall 2003). Indeed, multilocus sequence data in a coalescent framework, involving newly developed theoretical models can be a particularly powerful way to identify recently diverged species (Yang and Rannala 2010). Moreover, novel approaches that model the membership of individuals to evolutionary lineages have also been described (Yang and Rannala 2010; Ence and Carstens 2011; O'Meara 2010). More recently, a new method that analyses georeferenced phenotypic and genetic data under a unified model and inference framework has been developed, opening the way to robust comparisons between different types of data and possibly combined analyses (Guillot *et al.* 2012).

1.2. MOLECULAR TOOLS

As described above, molecular data has become a fundamental tool in species delimitation and in the study of speciation. In fact, the analysis of multiple genetic markers is nowadays considered to be of primary importance in the study of the evolutionary history of species.

The first phylogeographic studies were mostly based in the analysis of the mitochondrial DNA (mtDNA). This is a highly abundant molecule in the cell, which is easily extracted and analysed. MtDNA has a set of characteristics that makes it very popular in evolutionary and population genetic studies, namely its high mutation rate,

the absence of recombination in most animal species (though there are studies suggesting otherwise, Tsaousis *et al.* 2005), and the fact that it is maternally inherited and haploid in transmission (Avise 1989).

However, there are also some problems associated with the analysis of mtDNA. There are several taxa where non-functional nuclear copies of mitochondrial genes have been found (pseudogenes or numts; reviewed in Zhang and Hewitt, 1996; Benasson *et al.* 2001) and although these numt sequences are valuable by themselves in evolutionary studies (*e.g.* Triant and DeWoody 2008), if undetected, they can also confound phylogenetic and population genetic studies based on mtDNA (Zhang and Hewitt 1996; Dubey *et al.* 2009). Moreover, the effective population size of mtDNA is only a quarter that of nuclear autosomal loci, which implies that mtDNA has a faster lineage sorting and a higher allele extinction rate, and can lead to oversimplification of evolutionary relationships and underestimation of genetic diversity (Zhang and Hewitt 2003). Furthermore, the sole analysis of mtDNA corresponds effectively to the study of a single locus, and as such, has some widely recognised limitations. It is well documented that individual gene trees are often different from the underlying species tree. In fact, discordance between mitochondrial and nuclear phylogenies is quite common and it can be attributed to mitochondrial introgression (*e.g.* Boratyński *et al.* 2011; Melo-Ferreira *et al.* 2012) as well as the retention and incomplete sorting of ancestral polymorphisms, especially in the cases of recently diverged species or populations (Avise *et al.* 1983; Tajima 1983; Pamilo and Nei 1988; Maddison 1997).

Thus, as an organism comprises many genes, it is important to analyse several different genealogies in order to adequately describe the organism phylogeny. Indeed, complex evolutionary processes that shape the patterns of genetic variation in animal populations can only properly be addressed through the use of several independent markers, as has been demonstrated in several studies (*e.g.* Godinho *et al.* 2008; Melo-Ferreira *et al.* 2012).

Genetic data is nowadays easily and rapidly attainable, allowing for the assembly of large datasets that provide high resolution. Furthermore, several sophisticated statistical methods that analyse complex datasets, and integrate data from multiple loci have been developed recently (*e.g.* Heled and Drummond 2010; Hey *et al.* 2010). It is therefore expected that all these new methods and technological advances will greatly contribute to our understanding of the field of evolutionary biology and population and conservation genetics.

1.2.1. NONINVASIVE GENETICS

The first studies using noninvasive genetic sampling were two decades ago (Höss *et al.* 1992; Taberlet and Bouvet 1992) and since then major technological advances have been made in this area (Beja-Pereira *et al.* 2009).

Noninvasive approaches allow genetic studies of wild populations without the need to capture or even observe the animals, providing the opportunity to answer questions that cannot be addressed using traditional molecular methods (Taberlet *et al.* 1999). Noninvasive methods are particularly helpful in the study of rare and elusive species under conservation threat. Most of the studies published using these techniques refer to mammals (Beja-Pereira *et al.* 2009). In fact, many mammalian species have elusive behaviour, being difficult to observe and therefore to study directly. Putman (1984), some years before the first noninvasive genetic studies, had already highlighted the value of dung as 'the most readily-available and easily collected source of information' for field mammalogists looking for observational data on their elusive study species. Furthermore, given that many mammalian populations are declining (thus increasing the difficulty of direct observations) and that direct methods of capture and handling can be expensive and even harmful to the individuals, the use of these noninvasive techniques has great value, allowing for the collection of data on the distribution, demography and life-history information of those species (Kohn and Wayne 1997). These techniques are also considered valuable even on abundant species, since they facilitate the implementation of large scale studies and allow for the collection of data without altering the dynamic of populations in ways that are difficult to predict (Kohn and Wayne 1997).

Several types of noninvasive sample have been used for molecular analysis (faeces, hairs, urine, feathers, bones, etc.) with faeces being the most commonly used type of sample (see Beja-Pereira *et al.* 2009 for a review). In molecular studies of small mammals, however, in addition to faecal samples, bones from barn owl pellets are also widely used (*e.g.* Taberlet and Fumagalli 1996; Poulakakis *et al.* 2005). The majority of the studies using noninvasive genetic sampling focus on individual identification, wildlife forensics, assessment of population genetic parameters and on studies of mating systems and behavioural ecology.

Though some concern has been expressed about the shortcomings of noninvasive genetic methods, namely related with low success rates, contamination problems and genotyping errors rates (Taberlet *et al.* 1999; Waits and Paetkau 2005), recent advances have led to the development of techniques that provide good enough

DNA and low enough genotyping errors to dispel these doubts (Beja-Pereira *et al.* 2009). Therefore, noninvasive genetic approaches are an exciting and promising field of study, which are expected to continue to have a growing contribution in molecular ecology and conservation genetic studies.

1.3. GEOMETRIC MORPHOMETRICS

Historically, species delimitation and description has mostly been based on the morphological characteristics of the taxa being studied. Reflecting this importance, morphological studies transitioned from a descriptive field to a quantitative science during the twentieth century, with the quantitative analysis of form becoming known as 'multivariate morphometrics' (Bookstein 1998). This 'multivariate morphometrics', now typically called 'traditional morphometrics', consists of the quantitative description of morphological shape (based on measurements of length, depth and width or shape features such as angles and ratios) and its multivariate analysis in order to describe the patterns of variation within and among groups (Bookstein 1998; Adams *et al.* 2004). This type of morphometrics is the 'study of covariances of the biological form' (Bookstein 1991).

In the late 1980's morphometrics underwent a 'major revolution', through the development of coordinate-based methods (that capture the geometry of the morphological structure of interest), of the statistical theory of shape and of the computation of deformation grids (see Bookstein 1998; Rohlf and Marcus 1993; Adams *et al.* 2004 for an historical review). These methods were named by Rohlf and Marcus (1993) as 'geometric morphometrics'.

Geometric morphometrics consists therefore of the analysis of the shape of a structure defined by a set of topographically corresponding points (Bookstein 1991; Rohlf and Marcus 1993; Corti *et al.* 2000; O'Higgins 2000). By taking into account the Cartesian geometric coordinates of the morphological structures it becomes a much more powerful tool than 'traditional morphometrics' that does not consider the geometric relationship among measurements (*e.g.* that pairs of measurements were made from a common landmark; Rolhf and Marcus 1993). In fact, geometric morphometrics are nowadays considered as one of the most powerful techniques for the identification of patterns of morphological variation even below the species level (Loy 1996; Zelditch *et al.* 2004). Indeed, these methods are becoming an increasingly important complement of molecular phylogenetics, as they provide a powerful tool for

examining congruence between morphology and genetics (Cardini 2003). Geometric morphometrics is now being widely applied, for instance, in studies of ecogeographical and clinal variation (*e.g.* Cardini *et al.* 2007), in analysis of the ontogeny of organism shape and its genetic basis (Atchley *et al.* 1992; Cheverud *et al.* 1997; Klingenberg *et al.* 2001, 2003), in phylogenetic relationships (*e.g.* Cardini 2003; Macholán 2006; Gündüz *et al. 2007*) and even in the study of hybrid zones (*e.g.* Gaubert *et al.* 2005; see Lawing and Polly 2009 for a review). In fact, statistical approaches that integrate phenotypic measurements, molecular markers and geographic or environmental information have been developed (*e.g.* Davis and Nixon 1997; Guillot *et al.* 2012). The integration of these different types of data is expected to ease the interpretation of complex phylogeographic systems as well as the identification and delimitation of populations and/or species (Guillot *et al.* 2012).

1.4. THE EURASIAN FIELD VOLE

Microtus is a highly speciose rodent genus comprising nearly half of all arvicolines, yet with a conservative external morphology (Musser and Carleton 2005). About 65 species are currently recognised in the Palearctic and Nearctic; these were described mainly from morphological characters and karyotypes. The variation in the number of recognised species of *Microtus* in different accounts (Nowak 1999; Musser and Carleton 2005), together with the fact that several new species have been proposed over the last 15 years (Golenishchev *et al.* 2003; Kefelioğlu and Kryštufek 1999; Kryštufek and Kefelioğlu 2001; Yiğit and Colak 2002), indicates that the taxonomic characters traditionally used have not been sufficient for solving all systematic issues in the genus (Jaarola *et al.* 2004). In fact, *Microtus* species demonstrate a high intraspecific variability in terms of skull and dental morphology (Chaline *et al.* 1999). Moreover, karyotypic variation is extensive in *Microtus*, with chromosome number varying between 17 and 64 (O'Brien *et al.* 2006) and with considerable variation in the size of the sex chromosome; yet this variability seems to be mostly unrelated with any phylogenetic trend (Modi 1987b; Macholán *et al.* 2001; Borodin *et al.* 2012). Also, high intraspecific variation in sex chromosome heterochromatin has been recorded (Macholán *et al.* 2001; Zima and Macholán 1995; Mitsainas *et al.* 2009).

Data from fossils and molecular analysis suggest that *Microtus* species have been generated during continuing radiation events that began 2 million years ago

(Chaline *et al.* 1999; Jaarola *et al.* 2004; Fink *et al.* 2010). Molecular evidence suggests that the evolutionary history of the genus is characterised by a scenario of repeated colonization events of America and Europe from Asia by similar colonization routes, promoting diversification (Jaarola *et al.* 2004; Fink *et al.* 2010). Furthermore, mitochondrial DNA analyses suggest that some of the currently recognised species are clearly still differentiating, as revealed by the high intraspecific diversity observed in some taxa (Jaarola *et al.* 2004). These high levels of diversity may even indicate that these taxa contain cryptic entities that might, in some cases, be revealed as being different species.

The Eurasian field vole (*Microtus agrestis*) is one of the species where high intraspecific variability has been detected, both in mitochondrial and sex-linked nuclear DNA (Jaarola and Searle 2002, 2004; Hellborg *et al.* 2005; Herman and Searle 2011). This is one of the most widely distributed species of the genus *Microtus*, ranging from the Iberian Peninsula northwards and eastwards through Russia to eastern Siberia (Fig. 1). In Europe, it occupies Great Britain and the central and northern parts of the continent with the exception of Iceland and Ireland (Kryštufek *et al.* 2008).

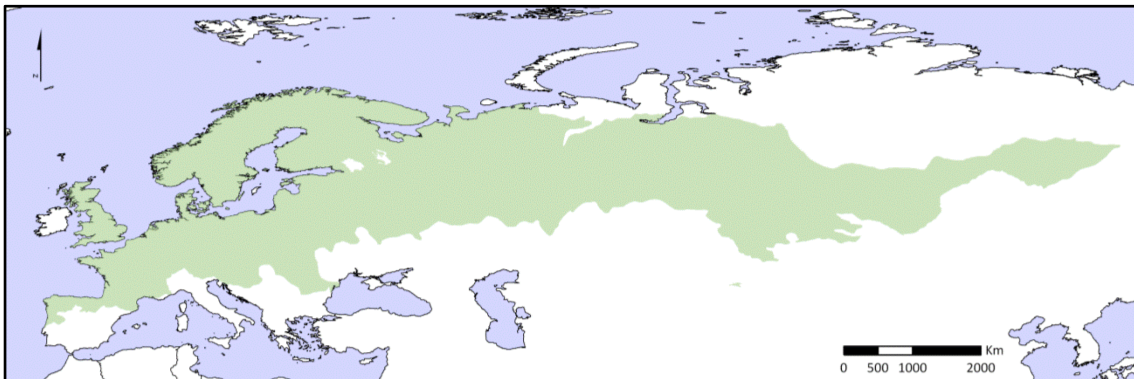


Fig. 1 Distribution range (green) of the field vole (*Microtus agrestis*) according to Kryštufek *et al.* (2008).

Microtus agrestis is considered to be basal in the phylogeny of the genus, not forming significant associations with the other European species (Jaarola *et al.* 2004; Fink *et al.* 2010). It seems to have derived from very old diversifications that occurred early in the evolutionary history of the genus, possible representing a trace of a first radiation in Europe, while other current European species appear to have originated from independent and evolutionary younger radiations (Fink *et al.* 2010). Recent mitochondrial DNA studies, involving the Asian lineage of *Microtus* (Bannikova *et al.* 2009, 2010), place *M. agrestis* as closely related with *Microtus* species of central Asia, subgenus *Blanfordimys* (*M. afghanus*, *M. bucharensis*, *M. juldaschi*). In fact, the earlier studies by Jaarola *et al.* (2004) already indicated *M. agrestis* as possibly a sister

species of *M. juldaschi*. Moreover, all of these Asian species show asynapsis of the sex chromosomes at male meiosis like *M. agrestis*, but have 'normal sized' sex chromosomes (Gileva *et al.* 1982, O'Brien *et al.* 2006, Borodin *et al.* 2012).

The field vole has one of the best-known and remarkable karyotypes of all mammals, possessing giant sex chromosomes (Matthey 1949). Actually, this species has the largest heterochromatin blocks on the sex chromosomes of any vole (and probably of any mammal). Typically the X chromosome of eutherians represents about 5% of the haploid genome. In the field vole this value is about 20% (Nanda *et al.* 1988), and the Y chromosome is also very large. Although there are *Microtus* species with a high intraspecific variability in the amount of sex chromosome heterochromatin (Macholán *et al.* 2001, Mitsainas *et al.* 2009), including one species also with giant sex chromosomes (*M. kirgisorum*: Zima and Macholán 1995), in the field vole only a small degree of variation has been detected until now. Despite the extensive studies performed on field voles over central and northern Europe, variation was only found in Southern Sweden, where the Y chromosome was found to have a substantial short arm ('the Lund Y'), while elsewhere the short arm on the Y chromosome is barely detectable (Fredga and Jaarola 1997).

High intraspecific molecular variability within *M. agrestis* was first detected by mitochondrial DNA (cytochrome *b*) analysis (Jaarola and Searle 2002, 2004) and then subsequently confirmed by the analysis of X- and Y-chromosome introns (Hellborg *et al.* 2005). These studies identified two main lineages with northern and southern distributions in Europe, with genetic divergences of 5.2 to 6.3% for mtDNA, 0.68% for the X-chromosome and 0.70% for the Y-chromosome. The Northern lineage occurs over a very large range through northern and central Europe and into Eurasia, extending from Britain in the west to eastern Siberia, from Fennoscandia in the north to the Alps in the south. The Southern lineage is found in southern Europe, including the Iberian Peninsula, much of France and in a narrow belt in the south of the Alps extending as far as Hungary. Recent studies in the areas where the two lineages come into contact have shown indications of interbreeding, but introgression appears to be limited to a narrow area (Hellborg *et al.* 2005; Beysard *et al.* 2012). All this evidence indicates that these two main lineages may comprise distinct evolutionary units that can, depending on what species definition is used, be considered separate species or close to being so. Furthermore, additional differentiation may be anticipated within the Southern field vole lineage, with a subdivision into two clades, one restricted to central Portugal (Jaarola and Searle 2004). However, relatively few markers have been examined to establish the subdivision into Northern and Southern lineages and the differentiation within the Southern lineage was only analysed with a single marker

(cytochrome *b*). Moreover, the geographic coverage of the previous studies was not very detailed, being very limited within the Iberian Peninsula, where the third mtDNA lineage was detected.

All the molecular data collected, together with the lack of detection of any morphological and karyotypic variability between these lineages, highlights the field vole as a striking example of 'cryptic speciation' that merits further investigation.

1.5. OBJECTIVES AND STRUCTURE OF THE THESIS

The general goal of this thesis is to study the evolutionary history of the field vole, *Microtus agrestis*, focusing on the major subdivisions previously detected and using a multilocus approach. In this thesis, the geographic patterns of genetic, karyotypic and morphometric variation will be thoroughly investigated, particularly in the Iberian Peninsula where previous limited studies showed a higher variability in mtDNA than elsewhere. To accomplish the general objective, several specific goals were set:

i) To study the geographic patterns of genetic variation in the entire range of the species in Eurasia, more specifically investigating the phylogenetic relationship of the lineages identified and gaining insights into their timing of divergence and demography, through the analysis of loci with different modes of inheritance;

ii) To investigate if there is any karyotypic differentiation in the mitochondrial lineage previously detected in Portugal;

iii) To determine the distribution limits of the field vole lineages in the Iberian Peninsula and to assess the existence of possible secondary contact, through noninvasive genetic sampling and using a multilocus approach;

iv) To determine if there is any morphological differentiation between the lineages identified in the Iberian Peninsula congruent with the molecular variability.

Considering the need to use noninvasive genetic sampling techniques, an additional goal was set to develop a method for Iberian rodent genetic identification using both mitochondrial and nuclear loci that could be applied to noninvasive samples like faeces or bones recovered from owl pellets.

The topics developed in this thesis are presented in five scientific papers already published in international journals or in a final phase of preparation. These are organized in chapters and answer the main questions placed. Therefore, the thesis

contains five chapters. The present Chapter 1 consists of a general introduction to the themes discussed, with a summary of the literature already published and the objectives of this work.

Chapter 2 contains two papers focusing on the evidence relating to cryptic speciation collected from both genetic and karyotype data. In the first paper (paper I, in press in *Molecular Ecology*) a multilocus dataset is explored with recently developed coalescent analysis confirming three distinct evolutionary units, which can be considered new cryptic species. Interesting issues such as the nature of species boundaries and tempo of speciation are addressed. Paper II (published in *Acta Theriologica*) describes the karyotype of field voles from Portugal revealing for the first time that these voles also have giant sex chromosomes and highlighting some minor differences in their structure when compared to field voles elsewhere. Aspects of chromosomal evolution and speciation in *Microtus* are discussed.

Chapter 3 refers to a detailed study of the field vole in the Iberian Peninsula through the analysis of multilocus sequence data and geometric morphometrics. It contains two papers, both in the final phase of preparation, one relating to the molecular and the other to the morphometric analysis. Paper III uses a noninvasive genetic sampling approach and loci with different modes of inheritance to clarify the distribution of the Iberian lineages of field vole and assess the extent of reproductive isolation. The process of cryptic speciation in the field vole species complex is discussed in light of the evidence found that post-mating isolation mechanisms already exist. The geometric morphometric analysis of field vole mandibles from the Iberian Peninsula is presented in Paper IV. This paper analyses the variability of mandible shape within the Iberian Peninsula, coupling the results with the molecular data obtained in the previous paper. It highlights the existence of subtle differentiation in mandible shape congruent with the molecular data examined.

Chapter 4 consists of Paper V (in press in *Molecular Ecology Resources*) and relates to the additional goal set of defining a molecular method for Iberian rodent identification. General considerations on rodent phylogeny and species delimitation are also made, and the performance of the method using noninvasive genetic samples is evaluated. The value of this method in ecological studies and in biodiversity conservation and management of rodents is discussed.

In Chapter 5 a general discussion is presented, integrating the main achievements of this study in the context of the field vole evolutionary history and considering its repercussions for the understanding of cryptic speciation processes in the presence of gene flow. The role of the *Microtus agrestis* species complex as a model system for the study of speciation and the nature of species boundaries are

highlighted. Main lines of future research are also proposed. Finally, the key conclusions of this thesis are laid out in Chapter 6.

1.6. LIST OF PAPERS

There is a great variability in the format and graphical presentation of the papers published in scientific international journals. For that reason, the text, tables and figures of the papers that compose this thesis were formatted in a uniform way, without any change in content. Below, a list of the papers that are published or in a final phase of preparation is presented:

Paper I - Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB (2012) Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia. *Molecular Ecology*. DOI: 10.1111/mec.12024.

Paper II – Giménez M[†], Paupério J[†], Alves PC, Searle JB (2012) Giant sex chromosomes retained within the Portuguese lineage of the field vole (*Microtus agrestis*). *Acta Theriologica*, **57**: 377-382.

Paper III - Paupério J, Barbosa S, Searle JB, Alves PC (manuscript) Asymmetric introgression between the two lineages of the field vole in the Iberian Peninsula revealed by noninvasive genetic sampling

Paper IV – Paupério J, Alves PC, Searle JB (manuscript) Morphometric differentiation of the two field vole lineages in the Iberian Peninsula: a geometric analysis of mandible shape

Paper V - Barbosa S[†], Paupério J[†], Searle JB, Alves PC (2012) Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling. *Molecular Ecology Resources.* DOI: 10.1111/1755-0998.12024.

1.7. REFERENCES

Adams DC, Rohlf FJ and Slice DE (2004) Geometric morphometrics: ten years of progress following the “revolution”. *Italian Journal of Zoology*, **71**, 5–16.

† Joint first authors

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 B*, **363**, 2831-2839.

Arèvalo E, Davis SK, Sites Jnr JW (1994) Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. *Systematic Biology*, **43**, 387-418.

Arias CF, Muñoz AD, Jiggins CD, Mavárez J, Bermingham E, Linares M (2008) A hybrid zone provides evidence for incipient ecological speciation in *Heliconius butterflies. Molecular Ecology*, **17**, 4699-4712.

Arntzen JW, Wallis GP (1991) Restricted gene flow in a moving hybrid zone of the newts *Triturus cristatus* and *T. marmoratus* in western France. *Evolution*, **45**, 805-826.

Ashley T, Jaarola M, Fredga K (1989) The behaviour during pachynema of a normal and inverted Y chromosome in *Microtus agrestis*. *Hereditas*, **111**, 281-294.

Atchley ER, Cowley DE, Vogl C, McLellan T (1992) Evolutionary divergence, shape change, and genetic correlation structure in the rodent mandible. *Systematic Biology*, **41**, 196-221.

Avise JC, Shapira JF, Daniel SW, Aquadro CF, Lansman RA (1983) Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. *Molecular Biology and Evolution*, **1**, 38-56.

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

Balloux F, Brünner H, Lugon-Moulin N, Hausser J, Goudet J (2000) Microsatelites can be misleading: an empirical and simulation study. *Evolution*, **54**, 1414-1422.

Bannikova AA, Lebedev VS, Golenishchev FN (2009) Taxonomic position of Afghan vole (subgenus *Blanfordimys*) by the sequence of the mitochondrial cytb gene. *Russian Journal of Genetics*, **45**, 91-97.

Bannikova AA, Lebedev VS, Lissovky AA, Matrosova V, Abramson NI, Obolenskaya EV, Tesakov AS (2010) Molecular phylogeny and evolution of the Asian lineage of the vole genus *Microtus* (Rodentia: Arvicolinae) inferred from mitochondrial cytochrome *b* sequence. *Biological Journal of the Linnean Society*, **99**, 595-613.

Barton NH, Hewitt GM (1981) A chromosomal cline in the grasshopper *Podisma pedestris*. *Evolution*, **35**, 1008-1018.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113-148.

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

Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution*, **16**, 314-321.

Berthier P, Excoffier L, Ruedi M (2006) Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society B*, **273**, 3101-3109.

Beysard M, Perrin N, Jaarola M, Heckel G, Vogel P (2012) Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (*Microtus agrestis*). *Journal of Evolutionary Biology*, **25**, 400-408.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**, 148-155.

Black A, Howell WM (1978) A distinctive chromosomal race of the cyprinodontid fish, *Fundulus notatus*, from the upper Tombigbee river system of Alabama and Mississippi. *Copeia*, **2**, 280-288.

Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 459-487.

Bookstein FL (1991) *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge.

Bookstein FL (1998) A hundred years of morphometrics. *Acta Zoologica Academiae Scientiarum Hungarica*e, **44**, 7-59.

Boratyński Z, Alves PC, Berto S, Koskela E, Mappes T, Melo-Ferreira J (2011) Introgression of mitochondrial DNA among *Myodes* voles: consequences for energetics? *BMC Evolutionary Biology*, **11**, 355-367.

Borodin PM, Sablina OV, Rodionova MI (1995) Pattern of X-Y chromosome pairing in microtine rodents. *Hereditas*, **123**, 17-23.

Borodin PM, Basheva EA, Torgasheva AA, Dashkevich OA, Golenishchek FN, Kartavtseva IV, Mekada K, Dumont BL (2012) Multiple independent evolutionary losses of XY pairing at meiosis in the grey voles. *Chromosome Research*, **20**, 259-268.

Brunes TO, Sequeira F, Haddad CFB, Alexandrino J (2010) Gene and species trees of a Neotropical group of treefrogs: genetic diversification in the Brasilian Atlantic Forest and the origin of a polypoid species. *Molecular Phylogenetics and Evolution*, **57**, 1120-1133.

Cardini A (2003) The geometry of the marmot (Rodentia: Sciuridae) mandible: phylogeny and patterns of morphological evolution. *Systematic Biology*, **52**, 186-205.

Cardini A, Jansson AU, Elton S (2007) A geometric morphometric approach to the study of ecogeographical and clinal variation in vervet monkeys. *Journal of Biogeography*, **34**, 1663–1678.

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.

Cheverud JM, Routman EJ, Irschick DJ (1997) Pleiotropic effects of individual gene loci on mandibular morphology. *Evolution*, **51**, 2006-2016.

Corti M, Marcus LF, Hingst-Zaher E (2000) Introduction to symposium: geometric morphometrics in mammalogy. *Hystrix*, **11**, 3-7.

Coyne JA (1994) Ernst Mayr and the Origin of Species. *Evolution*, **48**, 19-30.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**¸1908-1920.

Dayrat B (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society*, **85**, 407-415.

Davis JI, Nixon KC (1992) Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology*, **41**, 421–435.

Dod B, Jermiin LS, Boursot P, Chapman VH, Nielsen JT, Bonhomme F (1993) Counterselection in sex chromosomes in the *Mus musculus* European hybrid zone. *Journal of Evolutionary Biology*, **6**, 529-546.

Dubey S, Michaux J, Brünner H, Hutterer R, Vogel P (2009) False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene. *Molecular Phylogenetics and Evolution*, **50**, 633-641.

Ence DD, Carstens BC (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**, 473-480.

Evin A, Baylac M, Ruedi M, Mucedda M, Pons JM (2008) Taxonomy, skull diversity and evolution in a species complex of *Myotis* (Chiroptera: Vespertilionidae): a geometric morphometric appraisal. *Biological Journal of the Linnean Society*, **95**, 529-538.

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.

Fredga K, Jaarola M (1997) The origin and distribution of the Lund Y chromosome in *Microtus agrestis* (Rodentia, Mammalia). *Hereditas*, **126**, 25-34.

Gallardo MH, Kausel G, Jiménez A, Bacquete C, González C, Figueroa J, Köhler N, Ojeda R (2004) Whole genome duplications in South American desert rodents (Octdontidae). *Biological Journal of the Linnean Society*, **82**, 443-451.

Gaubert P, Taylor PE, Fernandes CA, Bruford MW, Veron G (2005) Patterns of cryptic hybridization revealed using an integrative approach: a case study on genets (Carnivora, Viverridae, *Genetta* spp.) from the southern African subregion. *Biological Journal of the Linnean Society,* **86**, 11–33.

Gileva EA, Bolshakov VN, Chernousova NF, Mamina VP (1982) Cytogenetical differentiation of forms in the group of Pamir (*Microtus juldaschi*) and Carruther's (*M. carruthersi*) voles (Mammalia, Microtinae). *Zoolicheskii Zhurnal*, **61,** 912-922 (in Russian).

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

Golenishchev FN, Malikov VG, Nazari F, Vaziri AS, Sablina OV, Polyakov AV (2003) New species of vole of "*guentheri*" group (Rodentia, Arvicolinae, *Microtus*) from Iran. *Russian Journal of Theriology*, **1**, 117-123.

Gregory TR, Mable BK (2005) Polyploidy in animals. In *The Evolution of the Genome* (ed. Gregory TR). Elsevier, San Diego, 427–517.

Guillot G, Renaud S, Ledevin R, Michaux J, Claude J (2012) A unifying model for the analysis of phenotypic, genetic and geographic data. *Systematic Biology*. DOI: 10.1093/sysbio/sys038.

Gündüz I, Jaarola M, Tex C, Yeniyurt C, Polly PD, Searle JB (2007) Multigenic and morphometric differentiation of ground squirrels (*Spermophilus*, Sciuridae, Rodentia) in Turkey with the description of a new species. *Molecular Phylogenetics and Evolution*, **43**, 916-935.

Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, **12**, 101-109.

Harrison RG (1993) *Hybrid Zones and the Evolutionary Process*. Oxford University Press, New York.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.

Hellborg L, Gündüz I, Jaarola M (2005) Analysis of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology*, **14**, 2025–2031.

Hendry AP (2009) Speciation. *Nature*, **458**, 162-164.

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.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905-920.

Höss M, Kohn M, Pääbo S, Knauer F, Schroder W (1992) Excrement analysis by PCR. *Nature*, **359**, 199.

Ibañez C, García-Mudarra JL, Ruedi M, Stadelmann B, Juste J (2006) The Iberian contribution to cryptic diversity in European bats. *Acta Chiropterologica*, **8**, 277-297.

Jaarola M, Martínková N, Gündüz I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedaski-Tsolis S, González-Esteban J, López-Fuster MJ, Kandaurov AS, Kefelioğlu H, Mathias ML, Villate I, Searle JB (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.

Jiggins CD, Mallet J (2007) Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution*, **15**, 250-255.

Kefelioğlu H, Kryštufek B (1999) The taxonomy of *Microtus socialis* group (Rodentia: Microtinae) in Turkey, with the description of a new species. *Journal of Natural History*, **33**, 289-303.

Klingenberg CP, Leamy LJ, Routman EJ, Cheveraud JM (2001) Genetic architecture of mandible shape in mice: Effects of quantitative trait loci analyzed by geometric morphometrics. *Genetics*, **157**, 785-802.

Kohn MH, Wayne RK (1997) Facts from faeces revisited. *Trends in Ecology and Evolution*, **6**, 223–227.

Kryštufek B, Kefelioğlu H (2001) Redescription of *Microtus irani*, the species limits and a new social vole from Turkey. *Bonner Zoologishe Beiträge*, **50**, 1-14.

Kryštufek B, Vohralík V, Zima J, Zagorodnyuk I (2008) *Microtus agrestis*. In IUCN 2012. *IUCN Red list of threatened species*. Version 2012.1 <www.iucnredlist.org>. Downloaded on 20 June 2012.

Lawing AM, Polly PD (2009) Geometric morphometrics: recent applications to the study of evolution and development. *Journal of Zoology*, **280**, 1-7.

Lemskaya NA, Romanenko SA, Golenishchev FN, Rubtsova NV, Sablina OV, Serdukova NA, O'Brien PCM, Fu B, Yiğit N, Ferguson-Smith MA, Yang F, Graphodatsky AS (2010) Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). III. Karyotype relationships of 10 *Microtus* species. *Chromosome Research*, **18**, 459-471.

Loy A (1996) An introduction to geometric morphometrics and intraspecific variation: a fascinating adventure. In *Advances in Morphometrics* (eds Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE). Plenum, New York, 271-273.

Mable BK (2004) 'Why polyploidy is rarer in animals than plants': myths and mechanisms. *Biological Journal of Linnean Society*, **82**, 453–466.

Macholán M, Filippucci MG, Zima J (2001) Genetic variation and zoogeography of pine voles of the *Microtus subterraneus/majori* group in Europe and Asia Minor. *Journal of Zoology*, **255**, 31-42.

Macholán M (2006) A geometric morphometric analysis of the shape of the first upper molar in mice of the genus *Mus* (Muridae, Rodentia). *Journal of Zoology*, **270**, 672-681.

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

Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B,* **363**, 2971-2986.

Mandahl N (1978) Variation in C-stained chromosome regions in European hedgehogs (Insectivora, Mammalia). *Hereditas*, **89,** 107-128.

Marchal JA, Acosta MJ, Nietzel H, Sperling K, Bullejos M, Díaz de la Guardia R, Sánchez A (2004) X chromosome painting in *Microtus*: origin and evolution of the giant sex chromosomes. *Chromosome Research*, **12**, 767-776.

Martinez-Cruz B, Godoy JA (2007) Genetic evidence for a recent divergent and subsequent gene flow between Spanish and Eastern imperial eagles. *BMC Evolutionary Biology*, **7**, 170-177.

Matthey R (1949) Chromosomes sexuels géant chez un campagnol, *Microtus agrestis* L. *Experientia*, **5,** 72.

Mayr E (1963) *Animal Species and Evolution.* Harvard University Press, Cambridge, Massachusetts.

Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC (2005) Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology*, **14**, 2459-2464.

Melo-Ferreira J, Boursot P, Randi E, Kryukov A, Suchentrunk F, Ferrand N, Alves PC (2007) The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. *Molecular Ecology*, **16**, 605-618.

Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farelo L, Alves PC (2012) Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulation. *Systematic Biology*, **61**, 367–381.

Mitsainas GP, Rovatsos MT, Rizou EI, Giagia-Athanasopoulou E (2009) Sex chromosome variability outlines the pathway to the chromosomal evolution in *Microtus thomasi* (Rodentia, Arvicolinae). *Biological Journal of the Linnean Society*, **96**, 685-695.

Modi WS (1987a) C-banding analysis and the evolution of heterochromatin among arvicolid rodents. *Journal of Mammalogy*, **68**, 704-714.

Modi WS (1987b) Phylogenetic analysis of chromosomal banding patterns among the Nearctic Arvicolidae (Mammalia: Rodentia). *Systematic Biology*, **36**, 109-136.

Moritz C (1994) Defining "Evolutionarily Significant Units" for conservation. *Trends in Ecology and Evolution*, **9**, 373-375.

Musser GG, Carleton MD (2005) Superfamily Muroidea. In *Mammal Species of the World. A Taxonomic and Geographic Reference* 3rd ed. (eds Wilson DE, Reeder DM). Johns Hopkins University Press, Baltimore, Maryland, 834-1531.

Nachman MW, Searle JB (1995). Why is the house mouse karyotype so variable? *Trends in Ecology and Evolution*, **10**, 397-402.

Nanda I, Neitzel H, Sperling K, Studer R, Epplen JT (1988) Simple GAT/CA repeats characterize the X chromosomal heterochromatin of *Microtus agrestis*, European field vole (Rodentia, Cricetidae). *Chromosoma*, **96**, 213-219.

Nowak RM (1999*) Walker's Mammals of the World* 6th ed. John Hopkins University Press, Baltimore, Maryland.

O'Brien SJ, Menninger JC, Nash WG (2006) *Atlas of Mammalian Chromosomes*. Wiley-Liss, New York.

O'Higgins P (2000) The study of morphological variation in the hominid fossil record: biology, landmarks and geometry. *Journal of Anatomy*, **197**, 103-120.

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, **59**, 59-73.

Olsson U, Alström P, Ericson PGR, Sundberg P (2005) Non-monophyletic taxa and cryptic species – evidence from a molecular phylogeny of leaf-warblers (*Phylloscopus*, Aves). *Molecular Phylogenetics and Evolution*, **36**, 261-276.

Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annual Review of Genetics*, **34**, 401–437.

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

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386-393.

Piálek J, Hauffe HC, Searle JB (2005) Chromosomal variation in the house mouse. *Biological Journal of the Linnean Society*, **84**, 535-563.

Pfenninger M, Schwenk K (2007) Cryptic animal species are homogeneously distributed among taxa and geographical regions. *BMC Evolutionary Biology*, **7**, 121-127.

Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 215-230.

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.

Putman RJ (1984) Facts from faeces. *Mammal Review*, **14**, 79-97.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution*, **16**, 351-358.

Rohlf FJ, Marcus LF (1993) A revolution in morphometrics. *Trends in Ecology and Evolution,* **8**: 129-132.

Schönrogge K, Barr B, Wardlaw JC, Napper E, Gardner MG, Breen J, Elmes GW, Thomas JA (2002) When rare species become endangered: cryptic speciation in myrmecophilous hoverflies. *Biological Journal of the Linnean Society*, **75**, 291-300.

Searle JB (1998) Speciation, chromosomes and genomes. *Genome Research*, **8**, 1-3.

Searle JB, Erskine I (1985). Evidence for a widespread karyotypic race of hedgehog (*Erinaceus europaeus*) in Britain. *Journal of Zoology,* **206**, 276-278.

Secondi J, Faivre B, Bensch S (2006) Spreading introgression in the wake of a moving contact zone. *Molecular Ecology*, **15**, 2463-2475.

Sites Jnr JW, Marshall JC (2003) Delimiting species: a renaissance issue in systematic biology. *Trends in Ecology and Evolution*, **18**, 462-470.

Šlapeta J, López-García P, Moreira D (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Molecular Biology and Evolution*, **23**, 23-29.

Smadja CM, Butlin RK (2011) A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, **20**, 5123-5140.

Taberlet P, Bouvet J (1992) Bear conservation genetics. *Nature*, **358**, 197.

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

Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics*, **105**, 437-460.

Tate JA, Soltis DE, Soltis PS (2005) Polyploidy in plants. In *The Evolution of the Genome* (eds Gregory TR). Elsevier, San Diego, 371-426.

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

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E (2005) Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution*, **22**, 925–933.

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.

Whitkus R (1988) Experimental hybridizations among chromosome races of *Carex pachystachya* and the related species *C. macloviana* and *C. preslii* (Cyperaceae). *Systematic Biology*, **13**, 146-153.

Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, **56**, 875-878.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences, USA*, **107**, 9264-9269.

Yiğit N, Colak E (2002) On the distribution and taxonomic status of *Microtus guentheri* (Danford and Alston, 1880) and *Microtus lidius* Blackler 1916 (Mammalia: Rodentia) in Turkey. *Turkish Journal of Zoology*, **26**, 197-204.

Zhang DX, Hewitt GM (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution*, **11**, 247-251.

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

Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) *Geometric Morphometrics for Biologists: a Primer*. Elsevier Academic Press, London.

Zima J, Fedyk S, Fredga K, Hausser J, Mishta A, Searle JB, Volobuev VT, Wójcik JM (1996) The list of chromosomal races of the common shrew (*Sorex araneus*). *Hereditas*, **125**, 97-107.

Zima J, Macholán M (1995) Karyotypes of common voles from Kyrghyzstan and heterochromatin variation in the sex chromosomes of *Microtus kirgisorum*. *Caryologia*, **48**, 65-74.

CHAPTER 2

EVOLUTIONARY HISTORY OF THE FIELD VOLE IN EURASIA

Paper I - Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB (2012). Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia. *Molecular Ecology*. DOI: 10.1111/mec.12024.

Paper II – Giménez M†, Paupério J†, Alves PC, Searle JB (2012). Giant sex chromosomes retained within the Portuguese lineage of the field vole (*Microtus agrestis*). *Acta Theriologica*, **57**: 377-382.

† Joint first authors

PAPER I

Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia

Joana Paupério, Jeremy S Herman, Maarit Jaarola, José Melo-Ferreira, Paulo C. Alves and Jeremy B. Searle

Molecular Ecology (in press). DOI: 10.1111/mec.12024.

CRYPTIC SPECIATION IN THE FIELD VOLE: A MULTILOCUS APPROACH CONFIRMS THREE HIGHLY DIVERGENT LINEAGES IN EURASIA

PAUPÉRIO, J.[1,2,3], HERMAN, J.S. [2,4], MELO-FERREIRA J. [1], JAAROLA, M. [5], ALVES, P.C.[1,3,6] SEARLE, J.B.[2,7]

ABSTRACT

Species are generally described from morphological features, but there is growing recognition of sister forms that show substantial genetic differentiation without obvious morphological variation and may therefore be considered ‘cryptic species’. Here, we investigate the field vole (*Microtus agrestis*), a Eurasian mammal with little apparent morphological differentiation but which, on the basis of previous sex-linked nuclear and mitochondrial DNA (mtDNA) analyses, is subdivided into a northern and a southern lineage, sufficiently divergent that they may represent two cryptic species. These earlier studies also provided limited evidence for two major mtDNA lineages within Iberia. In our present study we extend these findings through a multilocus approach. We sampled 160 individuals from 46 localities, mainly in Iberia, and sequenced 7 loci, maternally, paternally and bi-parentally inherited. Our results show that the mtDNA lineage identified in Portugal is indeed a distinct third lineage on the basis of other markers as well. In fact, multilocus coalescent based methods clearly support three separate evolutionary units which may represent cryptic species: Northern, Southern and Portuguese. Divergence among these units was inferred to have occurred during the last glacial period; the Portuguese lineage split occurred first (estimated at ca. 70 000 years BP) and the Northern and Southern lineages separated at around the Last Glacial Maximum (estimated at ca. 18 500 years BP). Such recent formation of evolutionary units that might be considered species has repercussions in terms of understanding evolutionary processes and the diversity of small mammals in a European context.

Keywords: *Microtus agrestis*, phylogeography, cryptic species, species tree, demography.

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, InBIO Laboratório Associado, Campus Agrário de Vairão, 4485-661, Vairão, Portugal;
[2] Department of Biology, University of York, Wentworth Way, York YO10 5DD, United Kingdom
[3]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal;
[4]Department of Natural Sciences, National Museums Scotland, Chambers St, Edinburgh EH1 1JF, United Kingdom;
[5]Department of Clinical and Experimental Medicine, Linköping University, 581 85 Linköping, Sweden;
[6]Wildlife Biology Program, College of Forestry and Conservation, University of Montana, 32 Campus Drive, Missoula MT 59801, MT, USA.
[7]Department of Ecology and Evolutionary Biology, E139 Corson Hall, Cornell University, Ithaca, NY 14853-2701, USA

INTRODUCTION

The numerous phylogeographic studies conducted over recent years have led to the perception that species with wide distribution ranges are frequently subdivided into genetically distinct forms, often not discriminated by morphology and that are therefore 'cryptic'. The detection and delimitation of such forms contribute to the understanding of speciation and its underlying processes, one of the major goals of evolutionary biologists (Sites and Marshall 2003) also relevant to the conservation and management of biodiversity (Moritz 1994). Indeed, in some cases it may be appropriate to classify these major genetic forms as 'cryptic species' and the process generating such forms as 'cryptic speciation'.

The molecular and analytical tools to investigate cryptic speciation have much improved in recent years. Multilocus sequence data in a coalescent framework, involving newly developed theoretical models, can be particularly helpful to generate 'species trees' in the sense of Yang and Rannala (2010), i.e. the branching pattern of closely related taxa. Several methods have been developed to deal with incongruence between gene trees and species trees due to lineage sorting (Edwards *et al.* 2007; Heled and Drummond 2010; Kubatko *et al.* 2009). Novel approaches that model the membership of individuals to evolutionary lineages have also been established (Yang and Rannala 2010; Ence and Carstens 2011; O'Meara 2010). We here use these methodologies to investigate major cryptic genetic forms in voles of the genus *Microtus*.

On the basis of described species, *Microtus* is a highly differentiated rodent genus comprising nearly half of all arvicolines (Musser and Carleton 2005). There are about 65 recognised species in the Palearctic and Nearctic that, from fossil evidence and mitochondrial DNA (mtDNA) analyses (Jaarola *et al.* 2004), have been generated during continuing radiation events that began 2 million years ago. The Eurasian field vole (*M. agrestis*) is one of the most widely distributed described species of the genus *Microtus*, but is geographically subdivided into major genetic forms based on analyses of mitochondrial and sex-linked nuclear DNA (Jaarola and Searle 2002; Jaarola and Searle 2004; Hellborg *et al.* 2005, Herman and Searle 2011). These previous studies identified two main lineages, with northern and southern distributions in Europe and with genetic divergences of 5.2 to 6.3% for mtDNA, 0.68% for X-chromosome and 0.70% for Y-chromosome introns. The high genetic divergence between these two lineages, their clear geographical distribution and the limited introgression between them within a narrow area of contact led Hellborg *et al.* (2005) to consider that these

two lineages could represent cryptic species, which was further supported by more recent work (Beysard *et al.* 2012). In fact, according to the 'genetic species concept' (Baker and Bradley 2006), two genetically distinct sister lineages with high mitochondrial divergence (>5%), concordant nuclear variation, geographically defined distributions and limited hybridization restricted to narrow areas of contact, probably have independent evolutionary fates and should be considered different species. In this paper we are cautious about the definition of species, because we realise that some mammalogists would be content with a definition like Baker and Bradley's, whilst others would prefer a more strict adherence to the Biological Species Concept of Mayr (1963) and only consider forms as separate species if there is a complete absence of gene flow between them. The cryptic taxa that we discuss in this paper are at the cusp of being evolutionarily independent entities, located on the evolutionary scale precisely where species concepts tend to disagree the most.

The field vole system may therefore represent a striking example of 'cryptic speciation' that merits further investigation. However, relatively few markers have been examined to establish the subdivision into Northern and Southern lineages. Furthermore, additional differentiation based on mtDNA may be anticipated within the Southern field vole lineage, with a subdivision into two clades, one restricted to central Portugal (Jaarola and Searle 2004). In this work we thoroughly investigate the geographic patterns of genetic variation in *Microtus agrestis sensu lato* (*s.l.*) focussing on the major subdivisions and using a multilocus approach. We significantly increased the geographic coverage over the previous studies, especially in the Iberian Peninsula where the third mtDNA lineage was identified but based on limited sampling. We compared patterns of differentiation among loci on chromosomes with different modes of inheritance and different effective population sizes: mtDNA, X chromosome, Y chromosome and autosomes. We have applied the recently developed methods of species delimitation and species tree inference to assess the stability and congruence of the detected lineages, and discuss their possible species status. Finally, we performed demographic analyses to consider fluctuations in the size of the populations of each lineage and the timing of their divergence. We expect that these comprehensive data will add substantially to our knowledge of the evolutionary history of the field vole and contribute to a better understanding of speciation of small terrestrial vertebrates within the geographical setting of Eurasia.

MATERIALS AND METHODS

Samples and DNA extraction

A total of 142 field vole tissue samples from 39 localities were collected from live trapped animals or museum collections and analysed for the first time. Furthermore, 21 samples from 7 localities already used in previous papers (Jaarola and Searle 2002; Hellborg *et al.* 2005; Herman and Searle 2011) were analysed with additional molecular markers (Fig.1, Table 1). The newly collected samples were largely from Iberia. A small number of tissue samples of *M. arvalis* and *M. cabrerae* were also collected in Iberia, and included in the analysis as outgroups (Table 1). In species phylogenies with nuclear and mitochondrial sequences the phylogenetic position of *M. arvalis* and *M. cabrerae* relative to *M.agrestis* has been variable (Jaarola *et al.* 2004; Fink *et al*. 2010) and in a preliminary analysis we found that the combination of the two species as outgroups was essentially equivalent to having either species on its own in terms of topology, but increased the statistical support of the phylogenies. All tissue samples were preserved in absolute ethanol at 4ºC. Total genomic DNA was extracted from ears or tail tips with the Qiagen Dneasy™ Tissue kit.

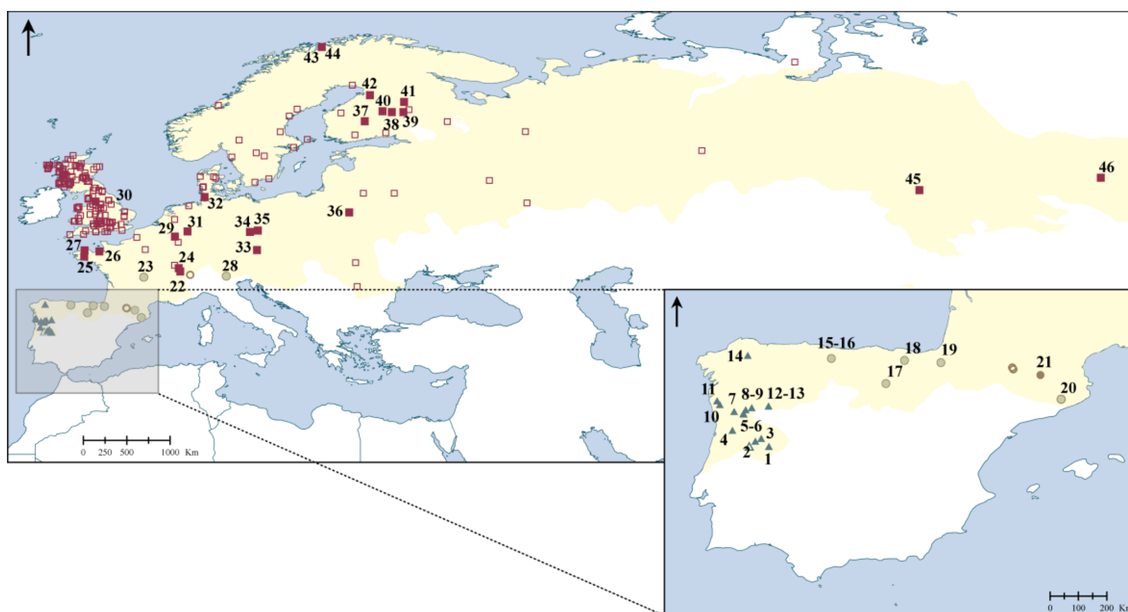


Fig. 1 Map showing in yellow shading the distribution range of the field vole (*Microtus agrestis*) according to IUCN (2010), as well as the geographical location of the sampling localities (filled symbols) and of the sequences retrieved from GenBank (empty symbols). The locality numbers refer to those in Table 1. Symbols indicate the lineages identified in the gene trees (squares – Northern; circles – Southern; triangles – Portuguese).

Table 1 Details of localities sampled for the field vole (*Microtus agrestis s.l.**) in this study and number of samples sequenced for each locus in each locality.

				Sample size						
Map ref.	**Locality**	**Country**	**Lat/Long**	***Cytb***	**DBX5**	**DBX6**	**DBY7**	**SMCY7**	**RAG1**	**BRCA1**
1	Malcata	Portugal	40°18' N 06°59' W	7	7	7	5	0	7	7
2	Serra da Estrela	Portugal	40°29' N 07°25' W	5	5	5	1	1	5	5
3	Guarda	Portugal	40°34' N 07°13' W	8	8	8	6	6	8	8
4	Serra da Arada	Portugal	40°49' N 08°09' W	3	3	3	2	2	3	3
5	Serra do Alvão	Portugal	41°20' N 07°47' W	3	3	3	2	2	3	3
6	Lamas de Olo	Portugal	41°22' N 07°47' W	7	7	7	5	5	7	7
7	Celorico de Bastos	Portugal	41°25' N 08°05' W	6	6	6	4	4	6	6
8	Gouvães	Portugal	41°28' N 07°43' W	1	1	1	0	0	1	1
9	Vila Pouca de Aguiar	Portugal	41°32' N 07°31' W	4	4	4	3	3	4	4
10	Ponte de Lima	Portugal	41°38' N 08°33' W	5	5	5	3	3	5	5
11	Bertiandos	Portugal	41°46' N 08°37' W	5	5	5	3	3	5	5
12	Macedo Cavaleiros1	Portugal	41°36' N 07°00' W	6	6	6	3	3	6	6
13	Macedo Cavaleiros2	Portugal	41°36' N 06°59' W	6	6	6	3	3	6	6
14	Lugo	Spain	43°12' N 07°39' W	1	0	0	0	0	0	0
15	Picos da Europa1	Spain	43°06' N 05°00' W	12	12	12	5	5	12	12
16	Picos da Europa2	Spain	43°06' N 05°00' W	1	1	1	0	0	1	1
17	Burgos	Spain	42°18' N 03°16' W	7	8	8	5	5	7	7
18	Bizcaia	Spain	43°02' N 02°40' W	1	1	1	1	1	1	1
19	Eugui	Spain	42°58' N 01°31' W	0	12	12	6	6	10	10
20	Granollers	Spain	41°48' N 02°17' E	0	2	2	0	0	2	2
21	Andorra	Andorra	45°08' N 17°36' E	3	4	4	1	0	3	4
22	Vallée de Joux	Switzerland	46°35' N 06°21' E	0	0	2	0	0	0	0
23	Dontreix	France	45°49' N 02°32' E	3	0	0	0	0	0	0
24	Combe d'Orgeval	France	46°55' N 06°11' E	0	0	2	2	0	0	0
25	Gourin	France	48°08' N 03°36' W	1‡	1	1	0	0	1	1
26	St Malo	France	48°38' N 02°01' W	3‡	3	3	0	0	3	3
27	Trebeurden	France	48°46' N 03°33' W	1‡	1	1	0	0	1	1
28	Trento	Italy	46°06' N 11°05' E	0	0	1	1	0	0	0
29	Bihain	Belgium	50°12' N 05°47' E	0	0	1	3	0	0	0
30	Billington	England	53°48' N 02°25' W	5‡	5	5	4	4	5	5
31	Bonn	Germany	50°44' N 07°06' E	0	0	1	0	0	0	0
32	Kattinger Watt	Germany	54°16' N 08°51' E	6‡	6	6	5	5	6	6
33	Kladenské Rovne	Czech Rep.	48°48' N 14°17' E	0	0	1	0	0	0	0
34	Flaje	Czech Rep.	50°41' N 13°34' E	0	0	1	1	1§	0	0
35	Filipov	Czech Rep.	50°49' N 14°23' E	0	0	1	0	0	0	0
36	Bialowieza	Poland	52°42' N 23°51' E	0	0	1	0	0	0	0
37	Jyvaskyla	Finland	62°07' N 25°26' E	2	2	2	2	0	2	2
38	Sotkamo	Finland	63°03' N 28°15' E	0	1	1	1	0	1	1
39	Koli	Finland	63°03' N 29°27' E	2	2	2	2	0	2	2
40	Maaninka	Finland	63°08' N 27°18' E	0	0	1	1	0	0	0
41	Kuhmo	Finland	64°07' N 29°31' E	0	0	0	1	0	0	0
42	Muhos	Finland	64°48' N 25°59' E	1	1	2	1	0	1	1
43	Nord Reisa	Norway	69°47' N 21°00' E	0	0	1	1	0	0	0
44	Kagen	Norway	70°01' N 20°48' E	0	0	1	0	0	0	0
45	Novosibirsk	Russia	55°00' N 82°55' E	4†	4	4	6	3	1	3
46	Bratsk	Russia	55°17' N 101°42' E	0	0	1	1	0	0	0
Total				119	132	148	90	65	125	128

* Additionally, one sample of *Microtus arvalis* (from Andorra) and another of *Microtus cabrerae* (from Northern Portugal) were sequenced for all loci analysed and used as outgroups.
† Jaarola and Searle (2002), GenBank accession numbers AY167149, AY167156, AY167157, AY167175
‡ Searle *et al.* (2009), GenBank accession numbers FJ619765, FJ619776; Herman and Searle (2011), GenBank accession numbers GU563252, GU563286, GU563288, GU563290, GU563293- GU563296
§ Hellborg *et al.* (2005), GenBank accession number AY541735

DNA amplification and sequencing

One mitochondrial locus (cytochrome *b*, cyt*b*, 1143 bp) and six nuclear loci (total 4272 bp) were amplified by the polymerase chain reaction (PCR). Amplification of the entire cyt*b* gene was performed in a single PCR using standard protocols and published primers (Jaarola and Searle 2002). Of the nuclear loci, we amplified exon 11 of the

BRCA1 gene on chromosome 11 of the laboratory mouse, exon 1 of RAG1 gene on chromosome 2 of the mouse, introns 5 and 6 from the DBX gene on the X chromosome and introns DBY7 and SMCY7 on the Y chromosome (Table 1; primers, references and PCR conditions in Table S1, Supporting information). All PCR products were purified with a commercial kit (Qiagen) and both strands were sequenced (Macrogen Inc.). An additional 211 cyt*b* gene sequences were retrieved from GenBank and 14 published sequences for the X and Y chromosome introns were also used in the analysis (AY167149-AY167213, Jaarola and Searle 2002; AY541707, AY541712, AY541718/9, AY541730, AY541735, AY541741/2, Hellborg *et al.* 2005; FJ619746–FJ619786, Searle *et al.* 2009; GU563195-GU563299, Herman and Searle 2011).

Sequence alignment and phylogenetic analysis

The new sequences were visually inspected and assembled in Sequencher version 4.7 (Gene Codes Corporation, Ann Arbor, MI) and deposited in GenBank (accession numbers in Table S1, Supporting information). Alignments for each locus were generated with Clustal W (Thompson *et al.* 1994), implemented in BioEdit version 7.0.5.3 (Hall 1999) and manually edited in order to maximize blocks of base pair identity, while clustering insertion/deletions to minimize their number. Polymorphic positions for individual sequences from nuclear loci were carefully inspected to ensure correct and consistent identification of double peaks in heterozygotes. Additionally, heterozygous positions for insertion/deletions were resolved directly from offset chromatogram peaks (Flot *et al.* 2006). For subsequent analysis, haplotypes for each nuclear locus were inferred using Phase version 2.1 (Stephens *et al.* 2001; Stephens and Donnelly 2003) with five runs performed for each locus with 1000 burn-in steps and 1000 iterations. When the dataset included tri-allelic positions we used the parent-independent mutation model (-d1 command) and in the input file these positions were designated as locus type M. Phase results were consistent across runs for all loci. Moreover, all haplotypes for the X linked introns could be assigned unambiguously. The results for the autosomal exons included a small percentage of low probability phase calls (less than 8% with probability lower than 0.95). These correspond to rare alleles (usually found heterozygous with a common allele) and were included in the analysis since omitting them could lead to substantial reductions in the estimated diversity and a bias in the estimates of population genetic parameters (Garrick *et al.* 2010). All insertion/deletion polymorphisms present in the X and Y introns and in BRCA1 were used for network construction but were excluded from other analyses. The microsatellite motif present in SMCY7 was excluded from all analyses. Although DBX5 and DBX6 are located in the same gene, no significant linkage disequilibrium

was found between them (FSTAT 2.9.3.2; Goudet 2002). These loci were therefore treated as independent.

Phylogenetic analysis was first performed separately for all loci analysed. We used the Akaike Information Criterion (AIC) to select the best-fit model of sequence evolution for each locus alignment among the 88 available in the software jModelTest version 1.0 (Posada 2008). The phylogenetic relationships among haplotypes were inferred by the neighbour-joining (NJ) method implemented in PAUP*4.0b10 (Swofford 2002), the maximum likelihood (ML) approach in PHYML v3.0 (Guindon and Gascuel 2003) and the Bayesian phylogenetic inference (BI) of MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001). NJ and ML branch support was obtained through 10 000 and 1 000 bootstrap pseudo-replicates each (100 for cyt*b*). Bayesian posterior probabilities were estimated from two runs with four chains of one million generations for the nuclear genes and two runs with five chains of 10 million generations for cyt*b*, with a sampling frequency that provided a total of 10 000 samples for each run, excluding 25% burn-in.

Phylogenetic networks for nuclear exons and introns were computed with a median-joining algorithm implemented in the software Network 4.6.0.0 (Bandelt *et al.* 1999; http://www.fluxus-engineering.com). In the analysis of closely related sequences, networks are useful tools as they can provide more information than a strict consensus tree and still present a reliable estimate of the true genealogy (Cassens *et al.* 2005; Woolley *et al.* 2008).

Species delimitation and species tree inference

Individual gene trees may differ substantially from the underlying species tree. Recently developed phylogenetic methods that incorporate coalescent models have sought to estimate accurate species trees using data from multiple loci. These methods assume that lineage sorting is the main source of inconsistency between gene trees and species trees, and that there is no recombination within loci, free recombination between loci and no gene flow following speciation (Heled and Drummond 2010; Carstens and Dewey 2010, Ence and Carstens 2011). They require the prior assignment of individuals to particular taxa, therefore lineage membership should first be assessed, which can be done through the application of multilocus species delimitation methods (Hird *et al.* 2010). In this context, we have used the maximum likelihood method that is implemented in SpeDeSTEMv2 (Ence and Carstens 2011). This program estimates species phylogeny using STEM (Kubtako *et al.* 2009) and evaluates the independence of the evolutionary lineages represented by the defined groups of individuals. For that it calculates the likelihood of the species trees given a

set of gene trees for all hierarchical permutations of lineage grouping, ranging from considering all lineages as different species to clustering them all in one species. It then computes a series of metrics based on information theory (corrected AIC, AICc; relative likelihood, L and model probability, ωi) that indicate the probability of each permutation (Ence and Carstens 2011). This process additionally assumes that all loci are evolving in a clocklike manner and that theta (θ) is constant across lineages (Kubtako *et al.* 2009).

To adhere as closely as possible to the assumptions of the method, we first reduced the dataset to the largest non-recombining blocks, using the software IMgc (Woener *et al.* 2007). Due to the size of the dataset and to the fact that there are missing data for a few loci, the estimates of the species tree with SpeDeSTEM were performed with replicated subsampling (Hird *et al.* 2010, Ence and Carstens 2011). Hird *et al.* (2010) showed that replicated subsampling using 3 to 10 alleles per species produces accurate estimates of the species tree. Therefore, the non-recombinant dataset consisting of the individual alignments for all loci was analysed with replicated subsampling (100 replicates; 10 alleles for all lineages except for the outgroups). The analysis was repeated for only clocklike loci and with three different theta values (lowest theta estimated for the lineages identified, θp; average theta for all lineages, θa; and theta for the whole field vole population, θma).

Following the species delimitation process, the species tree was inferred with all lineages considered as different entities. The 100 replicate species trees resulting from the application of SpeDeSTEMv2 were used to construct a majority-rule consensus tree with SumTrees v3.3.1 from the DendroPy v3.9.0 package (Sukumaran and Holder 2010). This was carried out for both datasets (all loci and only clocklike loci) and for all theta values.

We also inferred the species tree using the multispecies coalescent model that is implemented in *BEAST (Heled and Drummond 2010), part of the BEAST v1.6.1 package (Drummond and Rambaut 2007). No outgroup was included in the analysis as *BEAST estimates the roots of the individual gene trees which are combined through the multispecies coalescent to estimate the root of the species tree (Heled and Drummond 2010). The input file for *BEAST was obtained with the application BEAUti v1.6.1, part of the BEAST package. Preliminary analyses were carried out to determine whether each locus evolved according to a strict molecular clock, by examining the posterior distribution of the standard deviation of an uncorrelated lognormal relaxed clock. Based on these trial runs (results not shown) the final analysis was performed with the HKY substitution model for the non-coding genes (X and Y chromosome introns) under a strict molecular clock. For the mitochondrial locus and the nuclear

exons (BRCA1 and RAG1) the analysis was carried out with a 2-partition (first and second codon position linked; third position separate) HKY (+ Γ for cyt*b*) substitution model, as recommended for protein coding sequence data (Shapiro *et al.* 2006). The uncorrelated lognormal relaxed clock was used for cyt*b* and BRCA1 and a strict molecular clock was assumed for RAG1. Cyt*b* was used as the reference gene, with its clock rate set to one. The prior of the relaxed clock standard deviation was set to an exponential distribution with a mean of 0.5, as recommended in BEAST. We used the Yule process tree prior and all remaining priors were set to the defaults. Four replicate runs of 500 million generations were performed, sampling trees and parameter estimators every 50 000 generations for all loci. Tracer v1.5 (Rambaut and Drummond 2007) was used to check for convergence of the runs. *BEAST independent runs appeared to sample all parameters well, except for the effective population sizes (see Discussion). The results from all runs were combined, after removal of a 10% burn-in, with LogCombiner v1.6.1 and summary trees were generated with TreeAnnotater v1.6.1 from the BEAST package.

Population genetic and demographic analysis

Total and net divergences between lineages were calculated using Kimura 2-parameter distances (Kimura 1980) in MEGA v.5.03 (Tamura *et al.* 2011). The standard deviations for these divergences were estimated from 10 000 bootstrap pseudo-replicates.

Nucleotide diversity (π), θw (computed from the number of segregating sites) and haplotype diversity (Hd) were calculated per lineage for all loci. Three neutrality test statistics, Tajima's D (Tajima 1989), Fu's F_S (Fu 1997) and R_2 (Ramos-Onsins and Rozas 2002) were calculated to detect recent population expansion, which results in an excess of singleton mutations in the external branches of the phylogenies. We tested for departures from a neutral model with 10 000 coalescent simulations of the genealogy. These statistics were estimated per locus for each lineage and for the whole field vole population, using DnaSP v5 (Librado and Rozas 2009). Separate calculations were performed for the whole data set and for the non-recombining blocks obtained with IMgc (Woener *et al.* 2007).

We used the Extended Bayesian Skyline Plot (EBSP, Heled and Drummond 2008), as implemented in BEAST v1.6.1, to investigate changes in the effective population size of each lineage over time. This method uses Markov chain Monte Carlo (MCMC) genealogy sampling to estimate a posterior distribution of effective population size through time, based on sequence data from more than one locus (Drummond *et al.* 2005; Heled and Drummond 2008). The EBSP analyses were performed for each

lineage with the linear demographic function. The substitution models for the different loci were the same as described above for the *BEAST analysis. A strict molecular clock was assumed for all loci since preliminary analysis with an uncorrelated lognormal relaxed clock resulted in a posterior distribution of the standard deviation that included zero. As a substitution rate for the reference gene (cyt*b*) we selected the rate recently determined by Herman and Searle (2011), based on the possible time of origin for a mitochondrial lineage of the field vole confined to the northern part of the British mainland and adjacent smaller islands. Due to the restricted distribution of this clade Herman and Searle (2011) considered that it most likely had originated after the rapid warming at the end of the Last Glacial Maximum (LGM) but before the separation of Britain from mainland Europe. Using these dates as a calibration these authors estimated the substitution rate for cyt*b* to be: 3.887×10^{-7} substitutions/site/year. The input file for the analyses was obtained with the application BEAUti v1.6.1. Priors for the substitution rates were the same as those for the *BEAST analysis and the prior for the mean distribution of population sizes was corrected according to the population sizes estimated in trial runs performed with a coalescent prior and a constant population size, as suggested by the EBSP authors (results not shown). All remaining priors were set to the defaults. Three replicate runs for each lineage of 500 million generations were performed, sampling trees and parameter estimators every 50 000 generations, respectively. Tracer v1.5 (Rambaut and Drummond 2007) was used to assess convergence of MCMC; BEAST appeared to sample all parameters well, as the Effective Sample Sizes (ESS) for the independent runs were higher than 200 for the majority of parameters and were always higher than 200 in the combined runs. The results of the independent runs were combined with LogCombiner v1.6.1, part of the BEAST package, after removal of 10% burn-in.

RESULTS

The field vole phylogeographic structure

In the majority of cases, the phylogenetic analysis of the different loci by the different tree inference methods retrieved three well defined and strongly supported clades (Figs. 2 to 4). The mtDNA phylogeny recovered these three main lineages with high posterior probability and bootstrap values (Fig. 2). The three lineages have coherent geographic distributions with limited overlap (Fig. 1) and are here identified as Northern, Southern and Portuguese. Members of the three mtDNA lineages can likewise be consistently distinguished with all of the sex-linked and autosomal

sequences (Figs. 3 and 4), although the support is not always strong. In particular, the analysis of the RAG1 exon, by the different methods, generated trees with varied topologies and with lower support values for the three lineages than recovered for other loci, although even with this locus the Portuguese and Northern lineages were consistent in the analyses. Notwithstanding the recovery of three lineages with all of the markers used here, the pattern of relationship between them varies. All possible permutations were observed in at least one analysis (Figs. 2 to 4). Finer level geographical structure is only evident in cyt*b* from the Northern lineage, but this has been the focus of a previous study (Herman and Searle 2011).

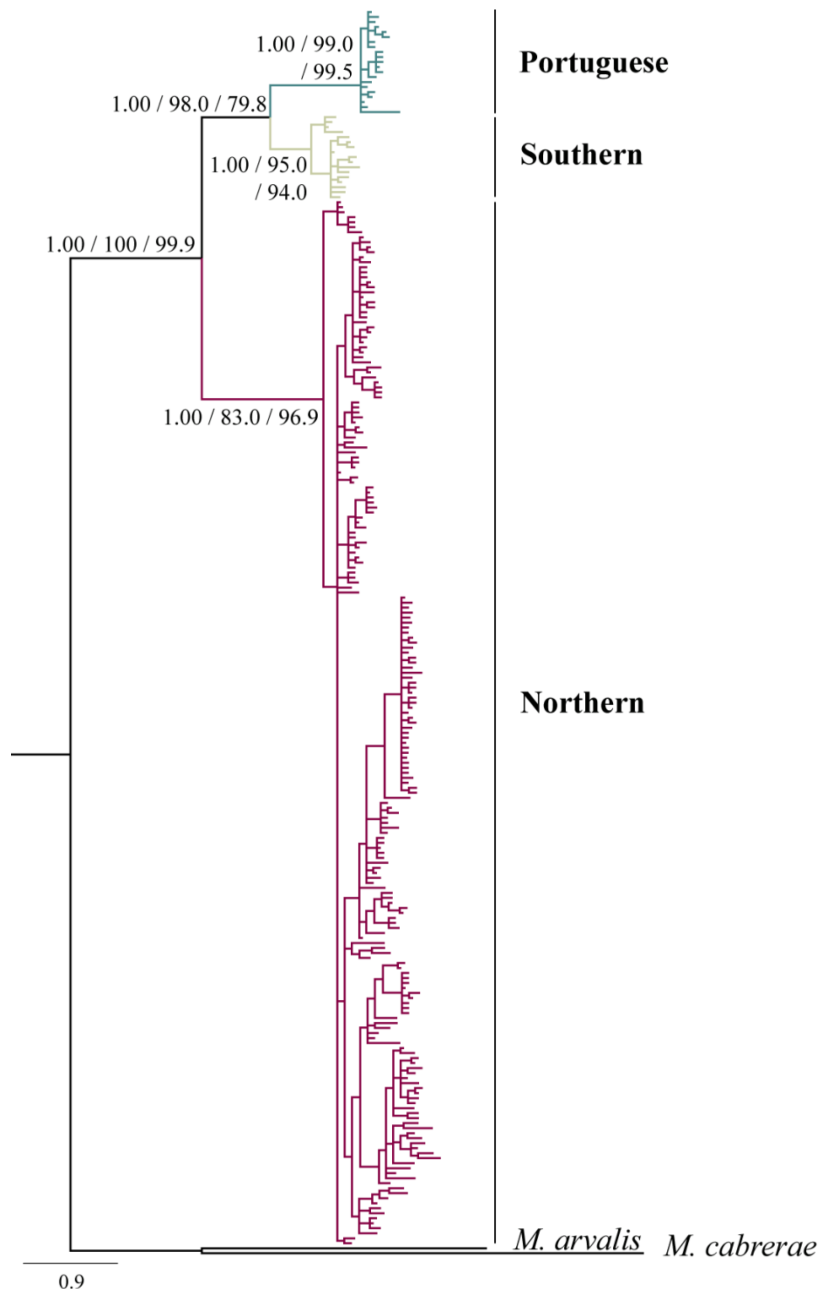


Fig. 2 Phylogenetic tree based on Bayesian inference showing the relationships among field voles for the cytochrome *b* gene (n = 412). Values on branches indicate posterior probability support and bootstrap values of the maximum likelihood and neighbour joining analysis. Lineages are coloured according to their geographic distribution, as in Fig. 1.

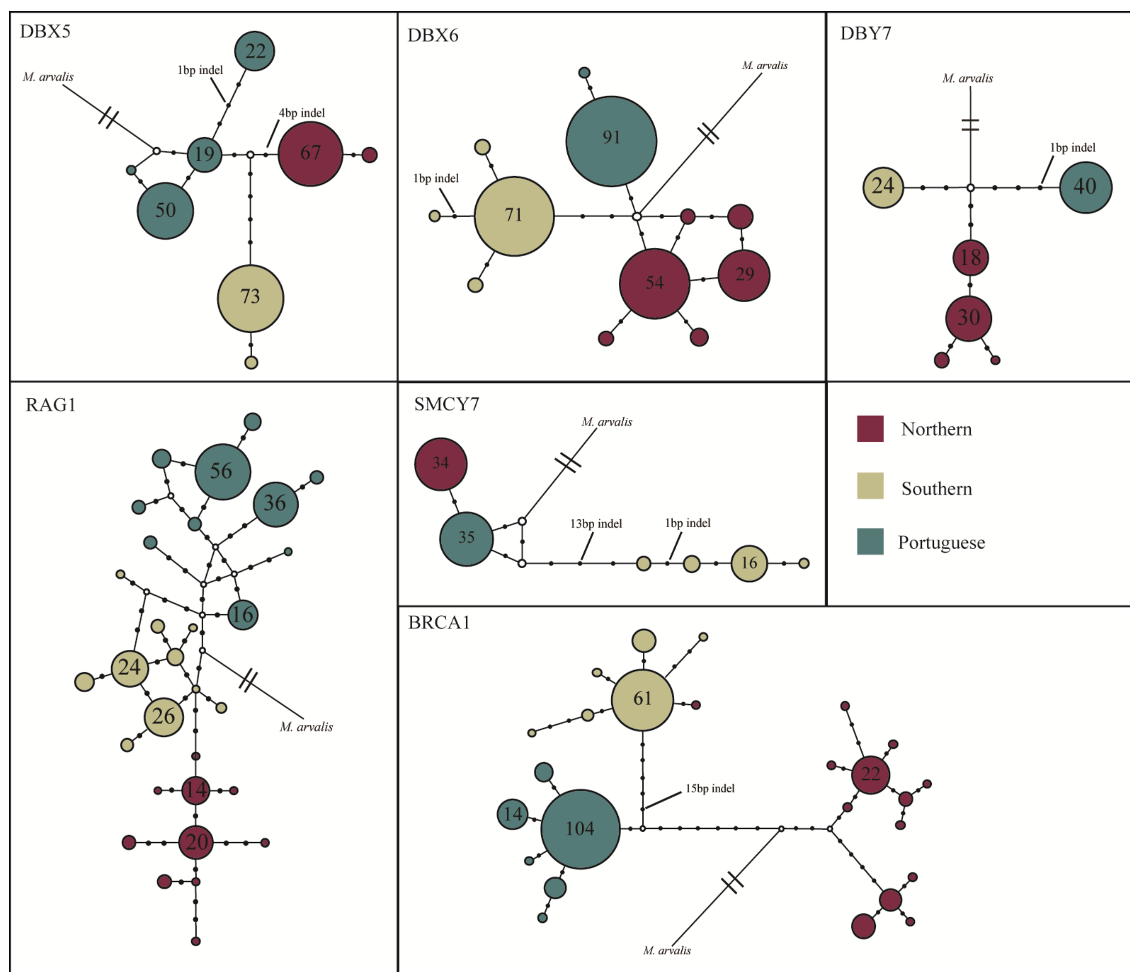


Fig. 3 Median-joining networks of the X-chromosome introns (DBX5 and DBX6), Y-chromosome introns (DBY7 and SCMY7) and autosomal genes (BRCA1 and RAG1) in *M. agrestis* (number of sequences for each locus in Table 2). *M. arvalis* included as outgroup (n = 1). Each circle represents one haplotype and circle area is proportional to the frequency of each haplotype. Total frequency is indicated for more common haplotypes. Branches are proportional to the number of nucleotide differences between haplotypes and dots on branches indicate mutational steps. Colours indicate geographic distribution of the lineages and correspond to those used in Fig. 1.

No clear evidence of gene flow was found between the three lineages. In all loci the three lineages were clearly separated and allele sharing was nearly non-existent - it was only detected for one of the exons, BRCA1. For this locus, one individual from St. Malo, a locality in northwestern France (Northern lineage), had a genotype consisting of a Northern allele and a singleton allele that clustered within the Southern lineage (Figs. 3 and 4). This individual had Northern alleles at all other loci analysed. This instance of shared variation could have resulted from the retention of ancestral polymorphism or from recent gene flow. However, the IMgc software, which detects and removes any traces of recombination within each locus, excluded this singleton, and thus this possible instance of gene flow was not included in the final dataset.

Other possible cases of recombination were identified by IMgc for the RAG1 exon. These included two of the haplotypes detected and a segment of the gene

(22.5% of the 3' end: 278 bp). One of these haplotypes was present in two specimens from St. Malo, including the same individual in which the Northern/Southern heterozygote in BRCA1 was detected. The other haplotype was only present in one individual from Picos da Europa, in northwestern Spain. Removal of these possibly recombinant haplotypes in RAG1 produced a phylogeny similar to the one obtained with the whole dataset, but with higher support for the nodes (data not shown).

Species delimitation and species tree inference

The three alleles identified as possible recombinants with the IMgc software, as well as the 278 bp segment at the 3' end of the RAG1 exon, were excluded from the species delimitation and species tree inference.

The results obtained with SpeDeSTEM showed that the three lineages are as distinct as may be expected for separate species. The fully resolved tree has a probability of 1.00 in every scenario, i.e. for the whole dataset or only the clocklike loci, and for the three values of theta that were used. Results for θa, 0.0118, are shown in Table 2; results for θp, 0.0056, and θma, 0.0268, are in Table S2, Supporting information.

Table 2 Species delimitation based on clocklike loci† and for all loci, considering an average theta value for all lineages (θa) of 0.0118 (see text and Table S2, Supporting information).

Lineage composition	-lnL	k	AICc	Δi	L	ωi
Clocklike loci						
P, S, N, Marv, Mcab	-2487.563	4	4984.607	0	1.000	1.000
P, S_N, Marv, Mcab	-2512.389	3	5031.636	47.029	0.000	0.000
P_N, S, Marv, Mcab	-2527.386	3	5061.630	77.023	0.000	0.000
P_S, N, Marv, Mcab	-2533.556	3	5073.969	83.362	0.000	0.000
P_S_N, Marv, Mcab	-2577.954	2	5160.323	175.716	0.000	0.000
All loci						
P, S, N, Marv, Mcab	-4911.420	4	9832.322	0	1.000	1.000
P, S_N, Marv, Mcab	-4999.162	3	10005.181	172.858	0.000	0.000
P_N, S, Marv, Mcab	-5027.651	3	10062.159	229.837	0.000	0.000
P_S, N, Marv, Mcab	-5063.910	3	10134.677	302.354	0.000	0.000
P_S_N, Marv, Mcab	-5169.367	2	10343.149	510.826	0.000	0.000

†Includes introns 5 and 6 from the DBX gene, intron 7 from the DBY gene, intron 7 from the SMCY gene and the RAG1 exon. P – Portuguese lineage ; S – Southern lineage ; N – Northern lineage ; Marv – *Microtus arvalis*, Mcab – *Microtus cabrerae*, -lnL – log-likelihood of the model given the data; k – number of parameters; AICc – corrected Akaike Information Criteria; Δi – AIC differences; L – relative likelihood of the model given the data; ωi – model probability.

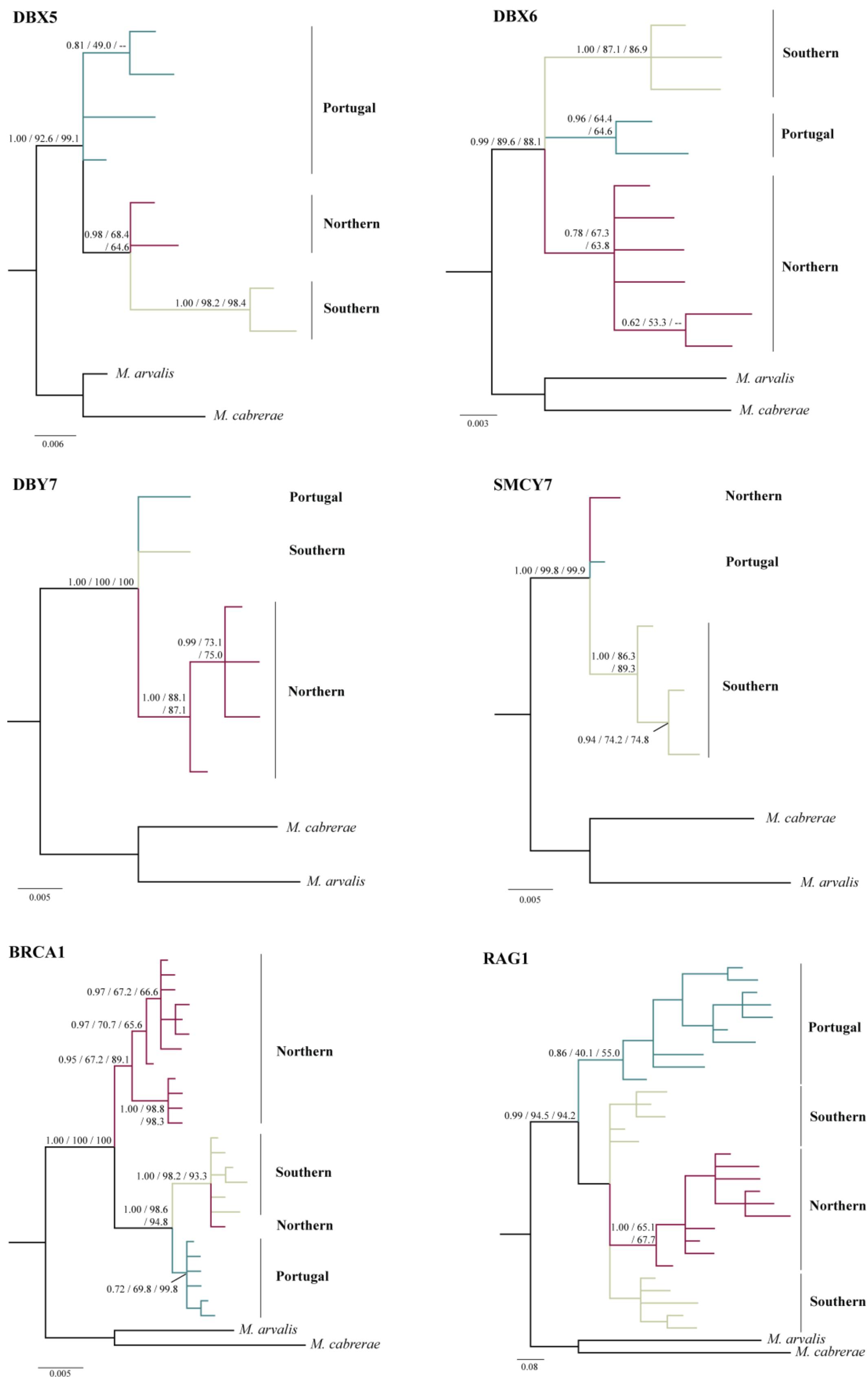


Fig. 4 Phylogenetic trees based on Bayesian inference showing the relationships among field voles for the nuclear genes. Values on branches indicate posterior probability support for clades and bootstrap support for the maximum likelihood and neighbour joining analyses respectively. The lineages are coloured according to their geographic distributions, as in Fig. 1.

The species trees inferred by both *BEAST and STEM suggest that the Northern and Southern lineages are more closely related to each other than to the Portuguese lineage (Fig. 5). The species tree inferred by *BEAST is very well supported, having a consistent posterior probability of 1 for the nodes in all analyses (results for the combined four runs presented in Fig. 5a). Using the substitution rate from Herman and Searle (2011) for cyt*b*, *BEAST estimations show that all the lineage separations took place during the last glacial period, with the first split that separated the Portuguese lineage occurring around 69 200 years ago (95% Highest Posterior Density (HPD) between 40 906 to 101 878 years before present (BP)). The split between the Northern and Southern lineages is estimated to have occurred more recently, close to the Last Glacial Maximum, around 18 500 years BP (95% HPD: 12 092 to 24 955 years BP).

The maximum likelihood approach, STEM, results in the same topology as *BEAST for both clocklike loci and for all loci, but there is better support for the node between the Northern and Southern lineages for clocklike loci only. The use of lower or higher theta values does not affect the topology of the species tree, but it does affect the branch lengths (higher thetas produce shorter branch lengths) and support values for the node between the Northern and Southern lineages (Fig. 5b; Fig. S1, Supporting information).

Population genetics and demographic analysis

The three lineages diverge more substantially for mtDNA than for the nuclear sequences, and for cyt*b* and BRCA1 the lowest divergences are between the Portuguese and Southern lineages while three of the four sex-linked loci show lowest divergences for the Portuguese and Northern lineages (Table 3).

Substantial variation in the levels of polymorphism among lineages and among loci was observed (Table 4). The cyt*b* gene showed the highest diversity values of all the loci examined, and of the lineages at this locus the Northern was by far the most variable with a haplotype diversity of 0.99, a nucleotide diversity of 1.16% and a θw of 3.08%. The Portuguese lineage showed a particularly low nucleotide diversity (0.29%). The diversity values for the X and Y chromosome introns were generally low but with considerable variation among lineages. Relative to the other loci, the autosomal exons showed an intermediate level of polymorphism. Between 39% (BRCA1) to 66% (RAG1) of the individuals analysed were heterozygous, but divergence between haplotypes in the same individual was mostly low, on average 0.30% for BRCA1 and 0.22% for RAG1. The only exception to these low values was the individual from St Malo for BRCA1 (the heterozygous individual with haplotypes from the Northern and Southern

lineage) that showed a divergence between the two haplotypes (1.22%) equivalent to that observed between lineages for this gene (Table 3). Comparing between lineages, for RAG1 the three lineages were similarly variable, while for BRCA1, the Northern lineage was clearly the most diverse. This was found even after the removal of the putative recombinant haplotype that clustered within the Southern lineage (Figs. 3 and 4) and contributed a considerable number of segregating sites (Table 4).

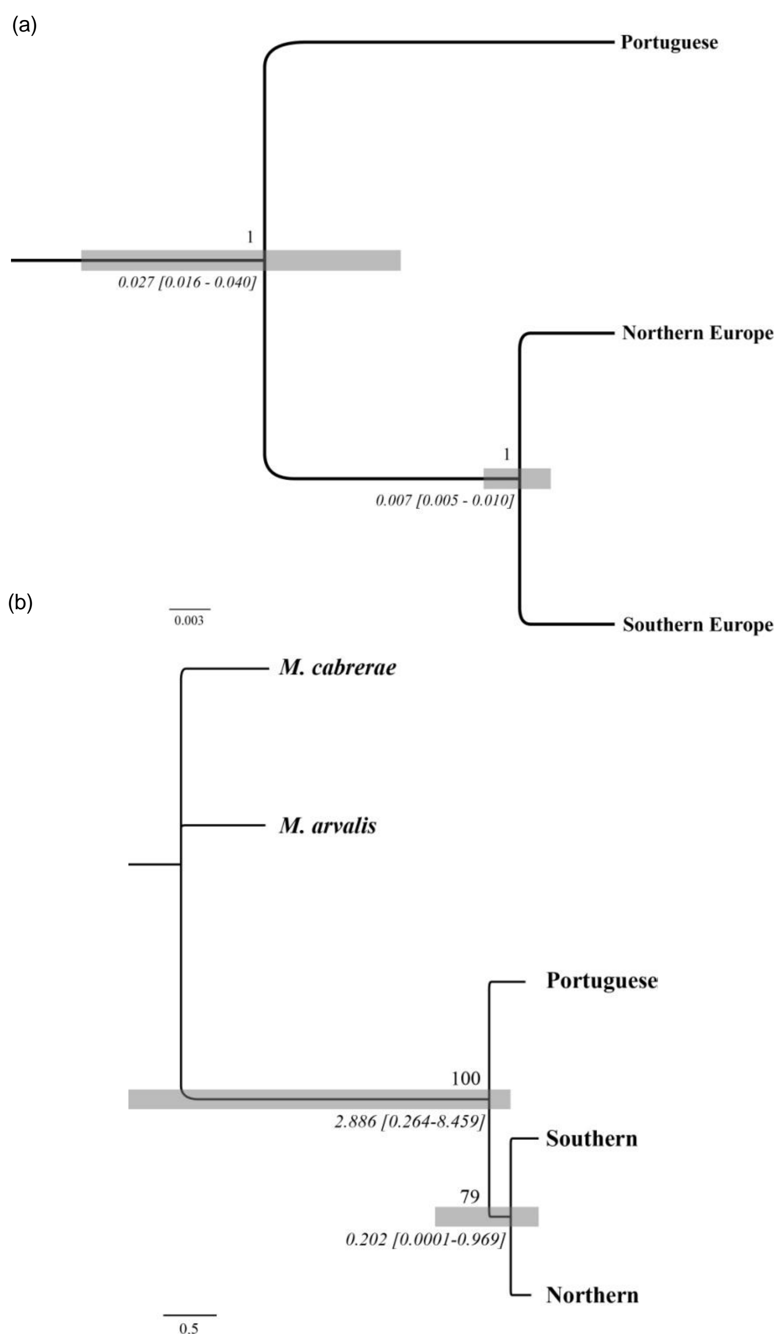


Fig. 5 Consensus species trees: (a) *BEAST species tree inference output for all 7 loci analysed. The posterior probability of each clade is presented on each node and grey bars show 95% highest posterior density intervals for times to most recent common ancestor (tMRCA in substitutions per site; values presented below bar) of each lineage. (b) Consensus species tree inferred by ML in STEM for all loci for a θa value of 0.0118 (see text). Grey bars show 95% highest posterior density intervals of branch lengths in coalescent units (number of $2N_e$ generations). Bootstrap support (above) and branch lengths (mean and 95%HPD, below) are shown for each node.

Table 3 Divergence between the three field vole lineages (standard errors based on 10 000 replicates are given in parentheses; all estimates are presented as percentages).

Locus		Lineage comparisons P - S	P - N	S - N
Cyt*b*	Dxy	3.49 (0.48)	6.59 (0.74)	5.66 (0.68)
	Da	3.07 (0.47)	5.85 (0.72)	4.79 (0.67)
DBX5	Dxy	2.18 (0.81)	0.74 (0.41)	1.46 (0.70)
	Da	1.95 (0.81)	0.50 (0.38)	1.43 (0.71)
DBX6	Dxy	0.84 (0.46)	0.67 (0.39)	0.95 (0.47)
	Da	0.83 (0.46)	0.56 (0.36)	0.84 (0.45)
DBY7	Dxy	0.81 (0.38)	0.95 (0.42)	0.95 (0.41)
	Da	0.81 (0.38)	0.89 (0.41)	0.89 (0.39)
SMCY7	Dxy	0.51 (0.28)	0.18 (0.17)	0.69 (0.33)
	Da	0.46 (0.27)	0.18 (0.17)	0.64 (0.32)
RAG1	Dxy	0.55 (0.18)	0.77 (0.22)	0.45 (0.17)
	Da	0.39 (0.14)	0.61 (0.20)	0.34 (0.15)
BRCA1	Dxy	0.42 (0.17)	0.95 (0.25)	1.17 (0.28)
	Da	0.39 (0.17)	0.78 (0.24)	1.00 (0.26)

P – Portuguese lineage ; S – Southern lineage ; N – Northern lineage; Dxy - average raw DNA divergence ; Da - average net DNA divergence.

In general, the average number of pairwise differences (π) was lower than the proportion of segregating sites (θw), which resulted in negative values for Tajima's D for most combinations of loci and lineages (Table 4). The same was observed for Fu's F_S. There was an excess of rare variants in cyt*b* in the Northern lineage, resulting in statistically significant values for all the neutrality statistics. The most sensitive statistic, Fu's F_S (Ramos-Onsins and Rozas 2002), was also significant for the Portuguese lineage for cyt*b*. The autosomal genes also showed some deviations from neutrality for both the Southern lineage (BRCA1 exon) and the Northern lineage (RAG1 exon) (Table 4).

The Extended Bayesian Skyline plot (EBSP), showing effective population size through time, revealed a dramatic expansion in the Northern lineage starting around 12 000 years BP, with a 10-fold increase in population size in 5 000 years (Fig. 6a). A second rapid expansion event was detected in this analysis, beginning around 3 000 years BP. The Southern lineage also showed signs of a recent expansion event, but with a more subtle increase in population size, that began around 7 000 years BP (Fig. 6b). The EBSP for the Portuguese lineage showed no signs of expansion (Fig. 6c), despite the significant Fu's F_S value for this population with mtDNA. The estimates of contemporary population size for the Southern and Portuguese lineages were about 10- to 100-fold smaller than for the Northern lineage (Fig. 6), reflecting their relative range sizes across Eurasia.

Table 4 Levels of polymorphism at each locus in the three lineages of field vole and for *Microtus agrestis s.l.*.

Locus	Lineage	*L*	*n*	*S*	*H*	*Hd* (s.d)	π ,% (s.d.)	θw,% (s.d.)	D	F_S	R_2
Cyt*b*	*M. agrestis*	1143	412	269	247	0.99 (0.002)	2.98 (0.1)	3.57 (0.7)	-0.78	-176.28	0.06
	Portuguese	1143	68	29	21	0.93 (0.012)	0.29 (0.03)	0.53 (0.16)	-1.43	-7.92**	0.06
	Southern	1143	34	38	17	0.92 (0.026)	0.56 (0.05)	0.81 (0.27)	-1.10	-2.96	0.07
	Northern	1143	310	222	209	0.99 (0.003)	1.16 (0.23)	3.08 (0.64)	-2.03**	-286.54**	0.03*
DBX5	*M. agrestis*	282	237	11	8	0.77 (0.01)	1.04 (0.03)	0.64 (0.23)	1.44	3.81	0.13
	Portuguese	287	92	4	4	0.61 (0.03)	0.44 (0.04)	0.27 (0.15)	1.21	2.17	0.16
	Southern	286	75	1	2	0.05 (0.03)	0.02 (0.01)	0.07 (0.07)	-0.90	-1.24	0.03**
	Northern	282	70	1	2	0.08 (0.04)	0.03 (0.02)	0.07 (0.07)	-0.74	-0.72	0.04*
DBX6	*M. agrestis*	367	264	10	11	0.75 (0.01)	0.56 (0.01)	0.44 (0.16)	0.61	-0.04	0.10
	Portuguese	367	92	1	2	0.02 (0.02)	0.01 (0.01)	0.05 (0.05)	-1.04	-2.15	0.10
	Southern	368	76	2	3	0.08 (0.04)	0.02 (0.01)	0.11 (0.08)	-1.31	-2.89	0.06
	Northern	367	96	4	6	0.59 (0.04)	0.20 (0.02)	0.21 (0.11)	-1.38	-1.40	0.09
DBY7	*M. agrestis*	498	116	9	6	0.75 (0.02)	0.61 (0.01)	0.34 (0.14)	1.98	4.73	0.17
	Portuguese	498	40	-	1	-	-	-	-	-	-
	Southern	499	24	-	1	-	-	-	-	-	-
	Northern	499	52	3	4	0.55 (0.5)	0.12 (0.01)	0.13 (0.08)	-0.16	-0.34	0.11
SMCY7	*M. agrestis*	559	91	5	5	0.69 (0.02)	0.28 (0.02)	0.18 (0.09)	1.23	1.86	0.15
	Portuguese	574	35	-	1	-	-	-	-	-	-
	Southern	559	22	2	3	0.44 (0.10)	0.08 (0.02)	0.10 (0.07)	-0.37	-0.31	0.14
	Northern	574	34	-	1	-	-	-	-	-	-
RAG1	*M. agrestis*	1234	250	34	30	0.90 (0.01)	0.44 (0.01)	0.45 (0.12)	-0.07	-3.70	0.08
	Portuguese	1234	132	14	10	0.73 (0.03)	0.22 (0.01)	0.21 (0.07)	0.17	0.92	0.10
	Southern	1234	72	10	10	0.75 (0.03)	0.10 (0.01)	0.17 (0.07)	-1.04	-3.36	0.07
	Northern	1234	46	12	10	0.72 (0.05)	0.11 (0.02)	0.22 (0.09)	-1.51*	-3.95*	0.06*
RAG1 without recomb	*M. agrestis*	956	246	27	21	0.87 (0.01)	0.46 (0.01)	0.46 (0.13)	-0.13	-1.03	0.08
	Portuguese	956	132	10	8	0.71 (0.03)	0.18 (0.01)	0.19 (0.07)	-0.33	0.34	0.09
	Southern	956	71	4	5	0.58 (0.03)	0.07 (0.01)	0.09 (0.05)	0.41	-0.92	0.08
	Northern	956	43	9	8	0.66 (0.05)	0.10 (0.02)	0.22 (0.09)	-1.56*	-3.49*	0.06
BRCA1	*M. agrestis*	1299	256	42	25	0.77 (0.02)	0.47 (0.03)	0.53 (0.13)	-0.31	-0.43	0.07
	Portuguese	1314	132	5	6	0.37 (0.05)	0.03 (0.005)	0.07 (0.03)	-1.13	-3.02	0.04
	Southern	1299	74	8	6	0.31 (0.07)	0.03 (0.01)	0.13 (0.05)	-1.87**	-3.32*	0.06
	Northern	1314	50	28	13	0.76 (0.05)	0.32 (0.04)	0.48 (0.16)	-1.11	-0.89	0.08
BRCA1 without recomb	*M. agrestis*	1299	255	42	24	0.76 (0.02)	0.47 (0.03)	0.53 (0.13)	-0.31	-0.05	0.07
	Portuguese	1314	132	5	6	0.37 (0.05)	0.03 (0.005)	0.07 (0.03)	-1.13	-3.02	0.04
	Southern	1299	74	8	6	0.31 (0.07)	0.03 (0.01)	0.13 (0.05)	-1.87**	-3.32*	0.06
	Northern	1314	49	16	12	0.75 (0.05)	0.28 (0.02)	0.27 (0.01)	0.08	-0.83	0.11

L – number of sites excluding gaps; *n* – number of chromosomes; *S* – number of segregating sites; *H* – number of haplotypes; *Hd* – haplotype diversity; π – nucleotide diversity per site; θw, computed from the number of segregating sites; D – Tajima's D; F_S – Fu's F_S; R_2 – Ramos-Onsins and Rozas's R_2; *p<0.05; **p<0.01.

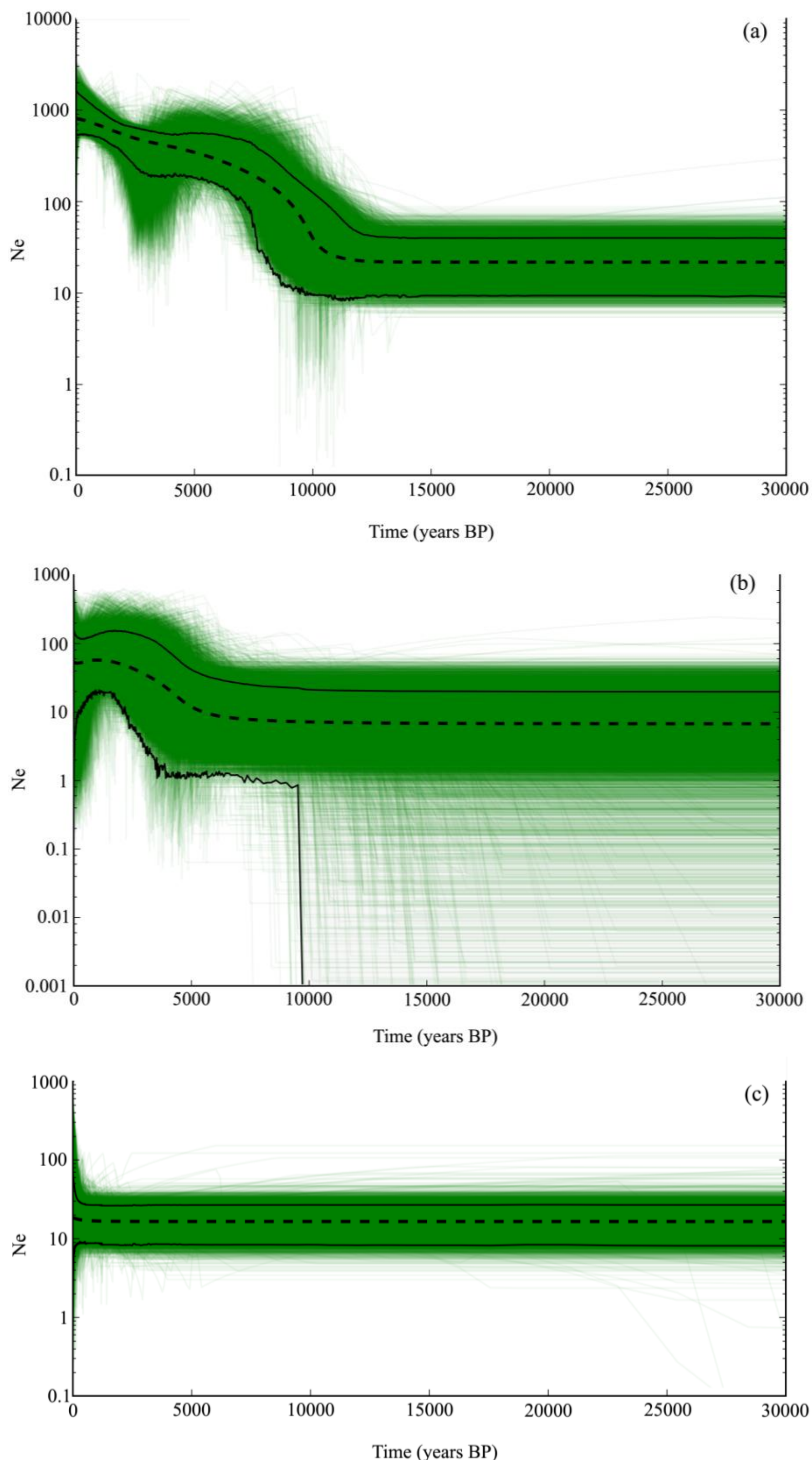


Fig. 6 Extended Bayesian Skyline plots of the effective population size through time estimated from 1450 million post-burn-in generations from three MCMC simulations. The accumulation of thin lines give a view of the full posterior distribution. Dashed black line is the median effective population size N_e, in thousands, multiplied by one (mean generation time in years). Solid black lines are the 95% highest posterior density limits. The y-axis is presented on a log scale for clarity. (a) – Northern lineage; (b) – Southern lineage; (c) – Portuguese lineage.

DISCUSSION

Three field vole lineages in Europe

In this study we were able to clarify the phylogeographic structure of *Microtus agrestis s.l.* We have confirmed the existence of three highly divergent lineages in Europe with geographically congruent distributions for multiple loci. The two lineages (Northern and Southern) that were first identified from mtDNA (Jaarola and Searle 2002 and 2004), and later supported by X- and Y-linked markers (Hellborg *et al.* 2005), were also recovered with the two autosomal markers BRCA1 and RAG1 and were further corroborated with the new samples analysed (Figs. 2 - 4). In addition, the third mtDNA lineage previously detected in central Portugal (Jaarola and Searle 2004) was confirmed in this study as a highly divergent population at every locus analysed. Moreover, our work showed that this lineage has a larger geographical distribution than previously realised, ranging from central Portugal to Galicia in northwestern Spain (Figs. 1 - 4).

Hence, the seven loci that we analysed appear to reflect a genome-wide result as they are generally consistent in their support of the three lineages. However, there are some differences in the topology of the individual gene trees and the support for the lineages within them. It is well recognised that individual gene trees are often different from the species tree, due to the retention and incomplete sorting of ancestral polymorphisms, especially in the case of recently diverged species or populations (Avise *et al.* 1983; Tajima 1983; Pamilo and Nei 1988; Maddison 1997). In addition, branch length heterogeneity is also considered to be a potent source of phylogenetic inconsistency in gene trees (Edwards 2009). Nevertheless, when we applied recently developed coalescent methods that consider incomplete lineage sorting as the main cause of discrepancy between trees (Kubatko *et al.* 2009; Heled and Drummond 2010; Ence and Carstens 2011), they produced concordant results when modelling the membership of individuals to the evolutionary lineages detected and in the estimation of the species tree. Species delimitation analysis showed that the three lineages identified in the gene trees undoubtedly reflect three distinct evolutionary units, as the fully resolved tree had the lowest AICc and an estimated probability of one. This result was achieved regardless of the value of theta imputed and whether or not the analysis was limited to clocklike loci. This consistency of results with variation in conditions has been found previously, for example Carstens and Dewey (2010) likewise found that changing conditions created little variation in the results of their species delimitation analysis of North American *Myotis* bats.

Although we have demonstrated the existence of three different evolutionary entities - comprising well separated and geographically delimited lineages and nearly no sharing of variation found between them - there is the possibility of very limited introgression between the Northern and Southern lineages. This is evident both from our studies and the more detailed analysis of Beysard *et al.* (2012). We found one possible hybrid genotype, for the BRCA1 gene, and possibly a recombinant allele in the RAG1 gene of the same individual, which was sampled in northwestern France, where the Northern and Southern lineages likely make contact. Beysard *et al.* (2012) localised the contact zone between the Northern and Southern lineages in western Switzerland and found only small numbers of hybrids (of which a high proportion were possible F_1s) and greater introgression at mitochondrial than nuclear loci. This contact zone likely extends through France from the Alps to the northwestern coast although further data are needed to map it more accurately than at present (Fig. 1).

There is virtually no indication of genetic exchange between the Portuguese and the Southern lineages. Only in Picos da Europa, the westernmost locality where we found the Southern lineage, a possible recombinant allele was detected in RAG1 exon, which could reflect hybridisation between the lineages. There appears to be a low density of field voles in the expected area of contact between the Portuguese and Southern lineages (based on repeated unsuccessful efforts to live trap voles in this area), making it difficult to study the interaction of these lineages based on field collections. Fortunately owl pellets provide a future opportunity to bridge this gap, through DNA analysis of bone material (Taberlet and Fumagalli 1996; Beja-Pereira et al, 2009; Barbosa *et al.* 2012).

Species tree inference

The species tree inference methods (Bayesian: *BEAST and maximum likelihood: STEM) gave consistent results with the Portuguese lineage as the oldest of the three lineages and with later divergence of the Southern and Northern lineages, but with all events taking place during the last (Weichselian) glacial period, between approximately 10 000 and 100 000 years ago (Fig. 5; Fig. S1, Supporting information). The recent separations reflect the application of the previously estimated field vole cyt*b* nucleotide substitution rate of 3.887 x 10^{-7} substitutions/site/year (Herman and Searle 2011) to the estimates of divergence times that we obtained with *BEAST. This substitution rate itself is based on a calibration dated ca. 10 000 years ago, close to the inferred separation times. However, some caution is needed with the results of the species tree inference. With *BEAST, although almost all parameters produced robust estimates across runs, there was a lack of convergence of the estimates of effective population

sizes. As in the Kubatko *et al.* (2011) study of the phylogeny of *Sistrurus* rattlesnakes, we assume that this single lack of convergence associated with a high consistency of all other parameters does not materially affect our results. For the STEM analysis, there is an assumption that theta should be similar for each lineage, but actually it is not: the three lineages have unequal theta values relating to very different geographical ranges (and therefore population sizes). We used a wide range of theta values in the analysis (following Carstens and Dewey 2010), and recovered consistent species trees with all (which are also identical to those obtained with *BEAST), and thus assume that the variation in theta within an analysis does not likely affect its outcome.

Evolutionary history of Microtus agrestis s.l.

The implications of our multilocus species tree results (Fig. 5) for the evolutionary history of the field vole, and for small mammals in general in a European setting, are quite profound. The single species of field vole has differentiated into three distinct forms, perhaps even species, within the timespan of the last glacial period (up to ca. 100 000 years BP). This is a more rapid pace of speciation than is typically envisaged for mammals, including small mammals in Europe. For instance, four closely related species of the *Sorex araneus* species complex (*S. araneus sensu stricto*, *S. coronatus*, *S. granarius*, *S. antinorii*) show similar levels of mtDNA divergence as found in the field vole, but the species separations were estimated to have occurred earlier, 54 000 to 580 000 years BP (Yannic *et al.* 2008). Other European taxa of small mammals close to the species transition also show earlier estimates of separation, such as *Microtus arvalis arvalis* and *M. a. obscurus* around 237 000 to 587 000 years BP (Fink *et al.* 2004), and *Mus musculus musculus* and *M. m. domesticus*, at ca. 250 000 years BP (Bonhomme and Searle 2012).

It is not only the divergence times within the multilocus species tree that are critical with regards the evolutionary history of the field vole, but also the branching order. The first proposed separation event of the Portuguese lineage from a combined Northern/Southern lineage would most reasonably have occurred in the vicinity of Iberia. This is the only area where the Portuguese lineage is currently found, and descendant populations of the Southern lineage are present there as well. Interestingly, the estimated timing of this separation (approximately 70 000 years BP ± 30 000) matches the maximum glacial extent in northern Iberia during the Weichselian, around 50 000 to 80 000 years BP (Lewis *et al.* 2009; García-Ruíz *et al.* 2010; Moreno *et al.* 2012).

The separation of the Southern and Northern lineages is dated to the Last Glacial Maximum (LGM) in the multilocus species tree, and on this scenario, the

Portuguese lineage may have been located in a refugium in western Iberia, the Southern lineage in a refugium in eastern Iberia or nearby, and the Northern lineage in a refugium further to the east (Jaarola and Searle 2002 suggest the Carpathians). Possible sites for the Southern lineage refugium are indicated by fossils dated to the LGM located in southern France, northern Italy and one location in northeastern Spain (The Stage Three Project, http://www.esc.cam.ac.uk/research/research-groups/oistage3).

The historical demography for the Northern lineage has already been described in detail, based on mtDNA analysis (Herman and Searle, 2011) and the current analyses add nothing to that. The substantial Northern lineage expansion apparently followed the Younger Dryas cold period 12 000 years ago and has carried on through the temperate conditions that have continued after that (Fig. 6a; Herman and Searle 2011). For the Portuguese and Southern lineages, although the neutrality tests generate significant values for some loci, the EBSP analysis shows only some indication of a small expansion event in the Southern lineage and no sign of expansion in the Portuguese lineage. This may partly reflect constraints on the distribution area; in particular, the Portuguese lineage is limited to a small range in western Iberia. Also, very large population sizes of field vole are currently seen in northern Europe, especially in Britain and Scandinavia (Harris *et al.* 1995; Myllymäki *et al.* 1977) suggesting particularly favourable conditions may have fuelled population growth of the Northern lineage.

One of the most striking aspects of our results is the clarity of the genetic differentiation signal for the Northern, Southern and Portuguese lineages, despite their recent separation. For all seven loci, the three lineages are evident (Figs. 2 – 4). The ecology of the field vole may help understand this. The species is less tolerant of dry conditions than the common vole *Microtus arvalis* (Dienske 1979), with which it coexists over much of its range; other *Microtus* species almost certainly provide competition in other respects, including thermal tolerance. During cold, arid, glacial conditions the field vole may have only survived as small populations in places where environmental conditions were permissible and competitors not overwhelming. Thus, population bottlenecking and strong genetic drift in small populations may have been important in generating the three distinctive forms with sorted genetic variation. Consistent with this, EBSP analyses of all lineages reveal relatively low population sizes before the Holocene expansions (Fig. 6).

In the broader context of small mammals in a European setting, it is important to consider whether the field vole is unusual in the speed of its diversification. Maybe not, if, as emphasised by Herman and Searle (2011), the nucleotide substitution rates generally used in previous studies of phylogeography and diversification of small

mammals may have been too low. With the continuous increase in genetic information available from different organisms, time inferences are becoming more robust (Ho *et al.* 2005). In fact, a recent study in a closely related vole species, *Microtus arvalis*, that used several loci, obtained divergence time estimates between lineages that were more recent than those previously published and also related to events in the LGM (Heckel *et al.* 2005). Therefore, the timings of origination of major genetic lineages within widespread species of small mammals, and of sibling species, need to be re-examined, to assess whether the field vole is an outlier in terms of rapid divergence, or whether it is reasonably representative of other European small mammals.

Speciation: from lineages to species

Traditionally species are distinguished on the basis of morphological characteristics. However, species without morphological identity ('cryptic' species) have increasingly been revealed through the use of molecular tools and integrative taxonomy (Bickford *et al.* 2007). In our study there was no *a priori* expectation based on morphology that the field vole *M. agrestis* is more than one species. As with all widespread European small mammals, there are a number of named subspecies of *M. agrestis* based on morphology (Krapp and Niethammer 1982), but none of these have any particularly strong case to be elevated to species status.

Our molecular study revealed the existence of three highly divergent evolutionary units in the field vole, but their distributions do not closely match those of any currently recognised subspecies. Therefore they can be viewed as 'cryptic', although future targeted studies based on the molecular groupings may reveal consistent morphological differences. The Northern, Southern and Portuguese lineages of the field vole are at the borderline between being considered separate populations or separate species, on the basis of cyt*b* divergences (3 – 6%), which are the only comparative data available. In particular, the sibling species pairs *M. duodecimcostatus / M. lusitanicus*, *M. arvalis / M. rossiaemeridionalis* [*epiroticus*] and *M. bavaricus* / *M. liechtensteini* exhibit cyt*b* divergences of 2 - 8% (Jaarola *et al.* 2004; Martínková *et al.* 2007), and in these cases species status is supported by independent data (*e.g.* fertility, karyotypes, morphology) and widely accepted by specialists (*e.g.* Mitchell-Jones *et al.* 1999, Wilson and Reeder 2005). There are also instances within *Microtus* of species (*M. arvalis*, *M. oeconomus*) that show high cyt*b* variability and large divergences between clades (1.0 - 4.3%; Brunhoff *et al.* 2003; Fink *et al.* 2004; Heckel *et al.* 2005).

In the case of the separation of the Northern, Southern and Portuguese lineages of *M. agrestis*, there are not only the high cyt*b* divergences, but also the

differentiation at multiple nuclear markers, to support species status. Moreover, there is evidence that genetic interchange is very limited on the contact of the Northern and Southern lineages (our data in northern France; data of Helborg *et al.* 2005; Beysard *et al.* 2012 in Switzerland), and possibly between the Southern and Portuguese lineages (our data in Picos da Europa). Therefore, within *M. agrestis s.l.* there are three major evolutionary units that may be considered separate species or close to being species. The Northern, Southern and Portuguese lineages are undoubtedly Evolutionarily Significant Units (ESUs) in the sense of Moritz (1994), and should be recognised as such for conservation purposes. The Portuguese lineage, in particular, has a relatively small geographic range in northwestern Iberia, is difficult to collect within this range (being notably elusive in Galicia) and, like other Iberian endemics, is liable to extinction or severe reduction in future distribution under climate change (*e.g.* Carvalho *et al.* 2010).

There are clear taxonomic priorities in terms of species designations for the three lineages. The Northern lineage represents the field vole *Microtus agrestis* first described by Linnaeus in Uppsala in 1761 from a local specimen. The Portuguese lineage relates to the species described as *Arvicola rozianus* (Bocage 1865) based on a specimen from a locality in central Portugal. This species was later on synonymised with *Microtus agrestis* by Lataste (1883) and subsequently considered as the subspecies *M. agrestis rozianus* with a distribution that included Portugal and Galicia (Trouessart 1910; Miller 1912; Cabrera 1914). The type specimen has been lost and over the last 50 years *M. agrestis rozianus* has not been considered a valid subspecies, based on Niethammer's (1964, 1970) assessment that it should be part of the more widespread *M. a. bailloni* (de Sélys-Longchamps 1841), which is found in western continental Europe from the Baltic through to southwestern France, according to Miller (1912). Despite these various changes in nomenclature, the name *rozianus* should have priority in the designation of the Portuguese lineage. In similar fashion, the priority for the Southern lineage should be *levernedii*, which is the oldest available name. It was attributed to a species, *Microtus levernedii* (Crespon 1844), that was later synonymised with *M. agrestis* and was described from a type locality in southern France, within the distribution area of this lineage. The name *M. a. levernedii* has been applied by various authors to field voles with a relatively southern distribution in Europe including southern France and Switzerland (Miller 1912; Ellerman and Morisson-Scott 1951).

Forms that are at the cusp of becoming species provide a challenge in terms of taxonomic nomenclature, but such systems are also the most interesting for understanding speciation. Thus, *M. agrestis s.l.* is an obvious candidate system to

investigate both the speciation process and the nature of species boundaries, including genomic and morphometric approaches, and further studies in these directions are clearly desirable. *M. agrestis s.l.* is not only interesting on its own merits, but as a representative of a genus with many species throughout the Holarctic, providing multiple opportunities to extend the findings of the one system to understand continental species radiations.

Acknowledgements

We thank Antoni Arrizabalaga, the Museum of Natural Science of Granollers, Augusto de Castro, Hélia Gonçalves and Zbyszek Boratyński for providing some of the samples used in this study. We are grateful to Maria da Luz Mathias for lending us field materials and to Alexandra Valongueiro, Ana Silva, Catarina Ferreira, Raquel Ribeiro, Rodrigo Vega, Rómulo Sobral and Sónia Ferreira for help during field work. This work was supported by Programa Operacional Potencial Humano - Quadro de Referência Estratégico Nacional funds from the European Social Fund and Portuguese Ministério de Educação e Ciência (FCT SFRH/BD/27369/2006 PhD grant to JP and SFRH/BPD/43264/2008 Post-Doc grant to JMF). We thank the reviewers for valuable comments.

REFERENCES

Avise JC, Shapira JF, Daniel SW, Aquadro CF, Lansman RA (1983) Mitochondrial DNA differentiation during speciation process in *Peromyscus*. *Molecular Biology and Evolution*, **1**, 38-56.

Baker RJ, Bradley RD (2006) Speciation in mammals and the genetic species concept. *Journal of Mammalogy*, **87**, 643-662.

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

Barbosa S, Paupério J, Searle JB, Alves PC (2012) Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling. *Molecular Ecology Resources*. DOI: 10.1111/1755-0998.12024.

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.

Beysard M, Perrin N, Jaarola M, Heckel G, Vogel P (2012) Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (*Microtus agrestis*). *Journal of Evolutionary Biology*, **25**, 400-408.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**, 148-155.

Bonhomme F, Searle JB (2012) House mouse phylogeography. In *Evolution of the House Mouse* (eds Macholán M, Baird SJE, Munclinger P, Piálek J). Cambridge University Press, Cambridge, 278-296.

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

Cabrera A (1914) *Fauna Iberica*. Mamíferos. Madrid.

Carstens BC, Dewey T (2010) Species delimitation using a combined coalescent and information-theoretic approach: an example from North American *Myotis* bats. *Systematic Biology*, **59**, 400-414.

Carvalho SB, Brito JC, Crespo EJ, Possingham HP (2010) From climate change predictions to actions – conserving vulnerable animal groups in hotspots at a regional scale. *Global Change Biology*, **16**, 3257-3270.

Cassens I, Mardulyn P, Milinkovitch MC (2005) Evaluating intraspecific "network" construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? *Systematic Biology*, **54**, 363–372.

Dienske H (1979) The importance of social interactions and habitat in competition between *Microtus agrestis* and *Microtus arvalis*. *Behaviour*, **71**, 1-126.

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.

Drummond AJ, Rambaut A (2007) Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214-222.

Ence DD, Carstens BC (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**, 473-480.

Edwards SV, Liu L, Pearl DK (2007) High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences, USA*, **104**, 5936-5941.

Edwards SV (2009) Is a new and general theory of molecular systematic emerging? *Evolution*, **63**, 1-19.

Ellerman JR, Morisson-Scott TCS (1951) *Checklist of Palaearctic and Indian Mammals*. British Museum (Natural History), London.

Fink S, Excoffier L, Heckel G (2004) Mitochondrial gene diversity in the common vole *Microtus arvalis* shaped by historical diversity 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.

Flot JF, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes*, **6**, 627-638.

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

García-Ruiz JM, Moreno A, González-Sampériz P, Valero Garcés B, Martí Bono C (2010) La cronología del último ciclo glaciar en las montañas del sur de Europa. Una revisión. *Cuaternario y Geomorfología*, **24**, 35-46.

Garrick RC, Sunnucks P, Rodney JD (2010) Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evolutionary Biology*, **10**, 118-135.

Goudet J (2002): *FSTAT Computer Package for PCs.* Institute of Ecology, UNIL, Lausanne, Switzerland.

Guindon S, Gascuel O (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium Series*, **41**, 95-98.

Harris S, Morris P, Wray S, Yalden D (1995) A review of British mammals: population estimates and conservation status of British mammals other than cetaceans. Peterborough, UK: Joint Nature Conservation Committee.

Heckel G, Burri R, Fink S, Desmet JF, Excoffier L (2005) Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. *Evolution*, **59**, 2231-2242.

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289-304.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.

Hellborg L, Gündüz I, Jaarola M (2005) Analysis of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology*, **14**, 2025–2031.

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.

Hird S, Kubatko L, Carstens B (2010) Rapid and accurate species tree estimation for phylogeographic investigations using replicated subsampling. *Molecular Phylogenetics and Evolution*, **57**, 888-898.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561-1568.

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

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.

Jaarola M, Martínková N, Gündüz I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedaski-Tsolis S, González-Esteban J, López-Fuster MJ, Kandaurov AS, Kefelioğlu H, Mathias ML, Villate I, Searle JB (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia), inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **33,** 647-663.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111-120.

Krapp F, Niethammer J (1982) Microtus agrestis (Linnaeus, 1761) - Erdmaus. In *Handbuch der Säugetiere Europas Band 2/I*, Nagetiere II (eds Niethammer J, Krapp F). Akademische Verlagsanstalt, Wiesbaden, 349-373.

Kubatko LS, Carstens BC, Knowles LL (2009) STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*, **25**: 971-973.

Kubatko LS, Gibbs HL, Bloomquist EW (2011) Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in *Sistrurus* rattlesnakes. *Systematic Biology*, **60**, 393-409.

Lataste F (1883). *Arvicola rozianus* Barboza du Bocage est synonyme de *Mus agrestis* Linné. *Le Naturaliste*, **47**, 373-374.

Lewis CJ, McDonald EV, Sancho C, Peña JL, Rhodes EJ (2009) Climatic implications of correlated Upper Pleistocene glacial and fluvial deposits on the Cinca and Gállego Rivers (NE Spain) based on OSL dating and soil stratigraphy. *Global and Planetary Change*, **67**, 141-152.

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

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

Martínková N, Zima J, Jaarola M, Macholán M, Spitzenberger (2007) The origin and phylogenetic relationships of *Microtus bavaricus* based on karyotype and mitochondrial DNA sequences. *Folia Zoologica*, **56**, 39-49.

Mayr E (1963) *Animal Species and Evolution.* Harvard University Press, Cambridge, Massachusetts.

Miller GS (1912) *Catalogue of Mammals of Western Europe.* British Museum, London.

Mitchell-Jones AJ, Amori G, Bogdanowicz W, Krystufek B Reijnders PJH, Spitzenberger F, Stubbe M, Thissen JBM, Vohralik V, Zima J (1999) *The Atlas of European Mammals*. Poyser, London.

Moreno A, González-Sampériz P, Morellón M, Valero-Garcés BL, Fletcher, WJ (2012) Northern Iberian abrupt climate change dynamics during the last glacial cycle: a view from lacustrine sediments. *Quaternary Science Reviews*, **36**, 139-153.

Moritz C (1994) Defining "Evolutionarily Significant Units" for conservation. *Trends in Ecology and Evolution*, **9**, 373-375.

Musser GG, Carleton MD (2005) Superfamily Muroidea. In *Mammal Species of the World. A Taxonomic and Geographic Reference* 3rd ed. (eds Wilson DE, Reeder DM), Johns Hopkins University Press, Maryland, 834-1531.

Myllymäki A, Christiansen E, Hansson L (1977) Five-year surveillance of small mammal abundance in Scandinavian. *EPPO Bulletin*, **7**, 385-396.

Niethammer J (1964) Ein Beitrag zur Kenntnis der Kleinsauger Nordspaniens. *Zeitschrift für Säugetierkunde*, **29**, 193-220.

Niethammer J (1970) Uber Kleinsauger aus Portugal. *Bonner Zoologische Beiträge*, **21**, 89-118.

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, **59**, 59-73.

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

Posada D (2008) jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Rambaut A, Drummond AJ (2007) Tracer v1.5. Available from http://beast.biod.ed.ac.uk/Tracer.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality test against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

S3P home page (2011) Databases and climate simulations. Available at: http://www.esc.cam.ac.uk/index.php/component/content/article/274 (last accessed 17 November 2011).

Searle JB, Kotlík P, Rambau RV, Marková S, Herman JS, McDevitt AD (2009) The Celtic fringe of Britain: insights from small mammal phylogeography. *Proceedings of the Royal Society B*, **276**, 4287–4294.

Shapiro B, Rambaut A, Drummond AJ (2006) Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution*, **23**, 7–9.

Sites Jnr JW, Marshall JC (2003) Delimiting species: a renaissance issue in systematic biology. *Trends in Ecology and Evolution*, **18**, 462-470.

Stephens M, Smith NJ, Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978-989.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods of haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162-1169.

Sukumaran J, Holder MT (2010) DendroPy: A Python library for phylogenetic computing. *Bioinformatics*, **26**, 1569-1571.

Swofford DL (2002) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates, Sunderland, Massachusetts.

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

Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics*, **105**, 437-460.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731-2739.

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.

Trouessart EL (1910) *Faune des Mammiféres d'Europe*. R. Friedlander und Sohn. Berlin.

Wilson DE, Reeder DM (2005). *Mammal Species of the World. A Taxonomic and Geographic Reference*, 3rd ed. Johns Hopkins University Press, Maryland.

Woener AE, Cox MP, Hammer MF (2007) Recombination filter genomic datasets by information maximization. *Bioinformatics*, **23**, 1851-1853.

Woolley SM, Posada D, Crandall KA (2008) A comparison of phylogenetic network methods using computer simulation. *PLoS ONE*, **3**, e1913.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences, USA*, **107**, 9264-9269.

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.

Data Accessibility

DNA sequences: Genbank accessions JX269135, JX269139, JX269141, JX269142 and JX284248 - JX284397.

Detailed information on each sampled specimen (locality, haplotypes, and accession numbers) and DNA sequence alignments: DRYAD entry DOI: 10.5061/dryad.kp79t

Author contributions box

This study is a part of JP's PhD thesis work on the field vole, *Microtus agrestis*, which she is developing under the supervision of PCA and JBS. Genetic laboratory work and analyses were conducted at the University of York and CIBIO. JSH and MJ were involved in earlier phylogeographic studies in this species and contributed with new samples and to the data analysis and writing of the paper. JMF, who is engaged in various studies relating to speciation and evolutionary biology, also contributed to the data analysis and writing of the paper.

APPENDIX – Supporting Information

Table S1 Details of loci and PCR primers used.

	Locus	Size (bp)	Forward primer		Reverse primer		T_n†	Accession No.‡
mtDNA	Cyt*b*	1143	L14727-SP[a]	GACAGGAAAAATCATCGTTG	H15915-SP[a]	GTCTTGTAAACCAGTAATGAA	55	JX284248 - JX284284
X	DBX5	287	DBX5F[b]	CAACAACTGTCCTCCACA	DBX5R[b]	CATGATAATTTCTCCCATCTC	TD 65-55	JX284366 - JX284378
	DBX6	367	DBX6F[b]	AGACTTGATGGCTTGTGC	DBX6R[b]	ATGCTGCAGTTTTTCCAGA	TD 65-55	JX284379 - JX284389
Y	DBY7	501	DBY7F[c]	GGTCCAGGAGARGCTTTGAA	DBY7R[c]	CAGCCAATTCTCTTGTTGGG	TD 65-55	JX269135, JX269139, JX269141 - JX269142
	SMCY7	637	SMCY7F[c]	CCCCTGCAGACCKRCAT	SMCY7R[c]	AAGGCCAAGTTGATRAARCT	TD 65-55	JX284390 - JX284397
Chr 2*	RAG1	1234	S70[d]	TCCGAGTGGAAATTTAAGMTGTT	S73[d]	GAGGAAGGTRTTGACACGGATG	55	JX284319 - JX284365
Chr11*	BRCA1	1313	BRCA1MF	ACAAGGAGCCAACGGAACAGAT	BRCA1MR	ATGGACCGGTTCCTTGAGGTTT	65	JX284285 - JX284318

*Location based on the position on the mouse genome; †T_n is the annealing temperature used in the PCR (where "TD" represents a touchdown procedure); ‡ GenBank Accession numbers of the sequences produced in this study per locus;
[a] Jaarola M, Searle JB (2002) Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology*, **11**, 2613–2621.
[b] Hellborg L, Ellegren H (2004) Low levels of nucleotide diversity in mammalian Y chromosome. *Molecular Biology and Evolution*, **21**, 158-163.
[c] Hellborg L, Ellegren H (2003) Y chromosome conserved anchored tagged sequences (YCATS) for the analysis of mammalian male specific DNA. *Molecular Ecology*, **12**, 283-291.
[d] 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.

Table S2 Assessment of species delimitation for clocklike loci† and all loci, considering θ*p* (0.0056) and θ*ma* (0.0268) (see text).

Lineage composition	-lnL	k	AICc	Δi	L(Model|data)	ωi
θ = 0.0056						
Clocklike loci						
P, S, N, Marv, Mcab	-4266.975	4	8543.432	0	1.000	1.000
P, S_N, Marv, Mcab	-4326.658	3	8660.173	116.741	0.000	0.000
P_N, S, Marv, Mcab	-4383.610	3	8774.077	230.254	0.000	0.000
P_S, N, Marv, Mcab	-4378.537	3	8763.931	220.499	0.000	0.000
P_S_ N, Marv, Mcab	-4487.254	2	8978.923	435.490	0.000	0.000
All loci						
P, S, N, Marv, Mcab	-8160.782	4	16331.046	0	1.000	1.000
P, S_N, Marv, Mcab	-8278.564	3	16563.984	232.939	0.000	0.000
P_N, S, Marv, Mcab	-8362.429	3	16731.715	400.669	0.000	0.000
P_S, N, Marv, Mcab	-8364.292	3	16735.441	404.395	0.000	0.000
P_S_ N, Marv, Mcab	-8573.301	2	17151.017	819.971	0.000	0.000
θ = 0.0268						
Clocklike loci						
P, S, N, Marv, Mcab	-1349.555	4	2708.591	0	1.000	1.000
P, S_N, Marv, Mcab	-1368.282	3	2743.422	34.831	0.000	0.000
P_N, S, Marv, Mcab	-1381.729	3	2770.314	61.724	0.000	0.000
P_S, N, Marv, Mcab	-1389.382	3	2785.622	77.031	0.000	0.000
P_S_ N, Marv, Mcab	-1411.981	2	2828.376	119.785	0.000	0.000
All loci						
P, S, N, Marv, Mcab	-2447.600	4	4905.121	0	1.000	1.000
P, S_S, Marv, Mcab	-2479.967	3	4966.791	61.669	0.000	0.000
P_N, S, Marv, Mcab	-2494.755	3	4996.366	91.245	0.000	0.000
P_S, N, Marv, Mcab	-2503.074	3	5013.005	107.884	0.000	0.000
P_S_N, Marv, Mcab	-2551.600	2	5107.614	202.493	0.000	0.000

P – Portuguese lineage ; S – Southern lineage ; N – Northern lineage ; Marv – *Microtus arvalis*, Mcab – *Microtus cabrerae*, -lnL, log-likelihood of the model given the data; k, number of parameters; AICc, corrected Akaike Information Criteria; Δ*i*, AICc differences; L, relative likelihood of the model given the data; ω*i*, model probability.
†Includes introns 5 and 6 from the DBX gene, intron 7 from the DBY gene, intron 7 from the SMCY gene and the RAG1 exon.

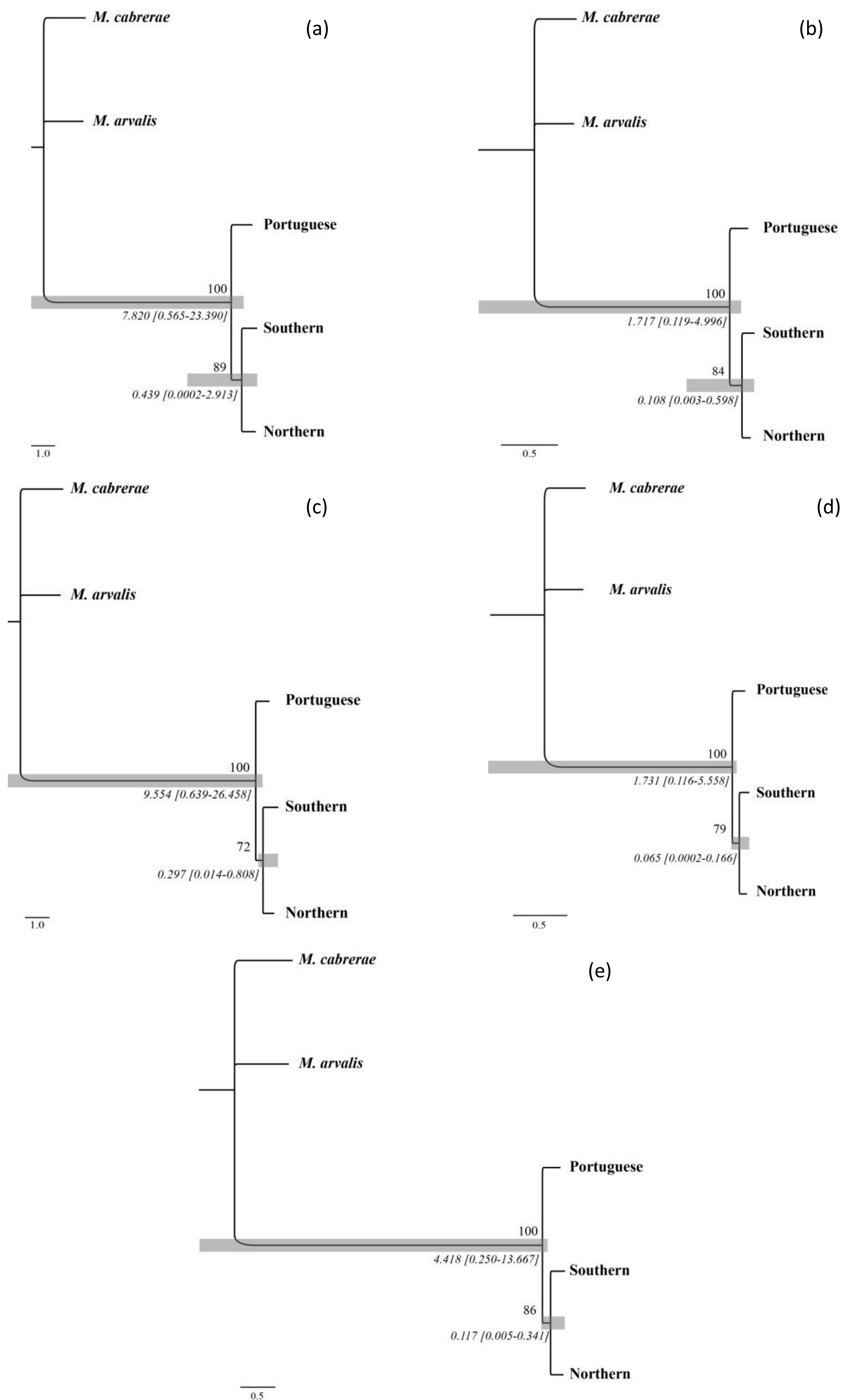


Fig. S1 Consensus species tree inferred with ML in STEM for clocklike loci (a, b, e) and for all loci (c, d), with θp = 0.0056 (a, c), θma = 0.0268 (b, d) and θa = 0.0118 (e). Grey bars show 95% highest posterior density intervals of branch lengths in coalescent units (number of $2N_e$ generations). Bootstrap support (above) and branch lengths (mean and 95%HPD, below) are shown for each node.

PAPER II

Giant sex chromosomes retained within the Portuguese lineage of the field vole (*Microtus agrestis*).

Mabel Giménez†, Joana Paupério†, Paulo C. Alves and Jeremy B. Searle

Acta Theriologica (2012), **57**, 377-382.

† Joint first authors

GIANT SEX CHROMOSOMES RETAINED WITHIN THE PORTUGUESE LINEAGE OF THE FIELD VOLE (*Microtus agrestis*)

GIMÉNEZ, M. [1,2]†, PAUPÉRIO, J.[1,3,4]†, ALVES, P.C.[3,4,5] SEARLE, J.B.[1,6]

ABSTRACT

The field vole (*Microtus agrestis*) is characterised by extremely large blocks of heterochromatin on both the X and Y chromosome. Some other *Microtus* also have blocks of heterochromatin on their sex chromosomes but not as extensive and always of independent origin from the heterochromatic expansion found in *M. agrestis*. Coupled with evidence of geographic variation in large heterochromatic blocks within other species (*e.g.* in the western hedgehog *Erinaceus europaeus*), it might be expected that field voles would show substantial variation in size and disposition of the sex chromosome heterochromatin. In fact, only minor variation has been described up to now. Those studies conducted previously were largely on field voles from central and northern Europe. Here, we describe the karyotype of field voles from Portugal, of interest because recent molecular studies have shown field voles from western Iberia to be a separate evolutionary unit that might be considered a cryptic species, distinct from populations further to the east. The two Portuguese field voles (one female, one male) that we examined also had essentially the same karyotype as seen in other field voles, including the giant sex chromosomes, but with small differences in the structure of the Y chromosome from that described previously. The finding that field voles throughout Europe show relatively little variation in their giant sex chromosomes is consistent with molecular data which suggest a recent origin for this complex of species/near-species.

Keywords: Chromosomal evolution, cryptic species, geographic variation, heterochromatin, karyotype.

---

† joint first authors

[1]Department of Biology (area 2), University of York, PO Box 37, York YO10 5YW, United Kingdom

[2] Facultad de Ciencias Exactas, Químicas y Naturales, Universidad Nacional de Misiones, Félix de Azara 1552, N3300LQH Posadas, Misiones, Argentina

[3]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, InBIO Laboratório Associado, Campus Agrário de Vairão, 4485-661, Vairão, Portugal;

[4]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal;

[5]Wildlife Biology Program, College of Forestry and Conservation, University of Montana, 32 Campus Drive, Missoula, MT 59801, MT, USA.

[6]Department of Ecology and Evolutionary Biology, Cornell University, Corson Hall, Ithaca, NY 14853-2701, USA

INTRODUCTION

The field vole (*Microtus agrestis*) has one of the best-known and distinctive karyotypes of all mammals, being characterised by giant sex chromosomes (Matthey 1949). Typically, the X chromosome of eutherians represents about 5% of the haploid genome. In the field vole, this value is about 20% (Nanda *et al.* 1988), and the Y chromosome is also very large. The increased size of the sex chromosomes represents expansion of repetitive elements (heterochromatin), with similarity in that suite of sequences between the X and Y chromosome (Marchal *et al.* 2004). Having such large sex chromosomes may be expected to cause difficulties at cell division and there is evidence for this at mitosis (Rens *et al.* 2006).

The presence of large blocks of heterochromatin has been reported in other mammals, and one of the features associated with that is variation in size and disposition of the heterochromatin between individuals and between populations within species, mediated by processes such as unequal crossing-over (Mandahl 1978). One particularly well-studied example of this is the western hedgehog (*Erinaceus europaeus*) where different populations in western Europe have large blocks of heterochromatin on different autosomes (Mandahl 1978; Searle and Erskine 1985).

It might be expected therefore within the field vole, and between the field vole and related species, that there would be variation in presence/absence and size and positioning of the heterochromatic blocks. Indeed, this is the case between species of *Microtus*. Although *M. agrestis* has the largest heterochromatin blocks on the sex chromosomes of any vole (and probably of any mammal), there are other *Microtus* with 'giant sex chromosomes', *e.g. Microtus cabrerae* with an X chromosome that comprises 15% of the haploid genome (Marchal *et al.* 2004). There are also *Microtus* with very small amounts of heterochromatin on the X chromosome and Y chromosomes of typical size; and also species with intermediate amounts of heterochromatin (Borodin *et al.* 1995). For those species that show large quantities of heterochromatin, each one appears to be characterised by an independent expansion of a different suite of repetitive elements (Marchal *et al.* 2004). It is striking that this variable and often large quantity of sex chromosome heterochromatin in *Microtus* is associated with asynapsis of the sex chromosomes at male meiosis, confirmed in many species (Borodin *et al.* 1995, 2012), including *M. agrestis* (Wolf *et al.* 1988; Ashley *et al.* 1989a). This unusual meiotic behaviour of the sex chromosomes does not appear to have arisen in response to the accumulation of blocks of heterochromatin (*Microtus* with normal-sized sex chromosomes also show asynapsis: Ashley *et al.*

1989b). Rather, it appears to be the other way around: having arisen, sex chromosome asynapsis may 'allow' an accumulation of heterochromatin which would otherwise interfere with the meiotic process.

As well as the dramatic variation in sex chromosome heterochromatin within the genus *Microtus*, a much smaller degree of variation has also been described within *M. agrestis*. Thus, in the karyotypes of field voles from southern Sweden, the Y chromosome is found to have a substantial short arm ('the Lund Y'), while elsewhere the short arm on the Y chromosome is barely detectable (Fredga and Jaarola 1997). In Fredga and Jaarola's review of sex chromosome variation in field voles over central and northern Europe (Fig. 1), a total of 491 voles from 120 localities in ten countries were examined and the only variation recorded was the Lund Y.

Much greater variation in sex chromosome heterochromatin has been recorded within other *Microtus* species (Macholán *et al.* 2001, Mitsainas *et al.* 2009), including one species also with giant sex chromosomes (*Microtus kirgisorum*: Zima and Macholán 1995).

Recently, it has been demonstrated that *M. agrestis* is subdivided into three major genetic forms (Jaarola and Searle 2002, 2004; Hellborg *et al.* 2005; Herman and Searle 2011; Beysard *et al.* 2012; Paupério *et al.* 2012). The Northern lineage occurs over a very large range through northern and central Europe and into Eurasia, extending from Britain in the west to Lake Baikal in the east, from Fennoscandia in the north to the Alps in the south. The Southern lineage is found in southern Europe in northeastern Iberia, much of France and in a narrow belt to the south of the Alps extending as far as Hungary. The Portuguese lineage is limited to western Iberia. These lineages differ by multiple genetic markers, but are not known to differ in morphology, so are in that sense 'cryptic'. They have a level of cytochrome *b* divergence that suggests that they are separate species or nearly so, and there is consistent nuclear divergence at loci on X-linked, Y-linked and autosomal loci (Jaarola and Searle 2002, 2004; Hellborg *et al.* 2005; Paupério *et al.* 2012). Where the lineages come into contact, there is evidence of interbreeding, but genetic interchange appears to be limited (Beysard *et al.* 2012; Paupério *et al.* 2012). Depending on what species concept is used, the Northern, Southern and Portuguese lineages of field vole are separate species or close to being so.

Previous studies on the sex chromosome variation in field voles have largely been restricted to the range of the Northern lineage (Fig. 1). One of the southern European localities sampled in the past was most likely from within the range of the Southern lineage, but none from the range of the Portuguese lineage (Fig. 1). The characteristics of the sex chromosomes of the Portuguese field voles are of real

interest. At the species level, other *Microtus* differ substantially in the quantity of sex chromosome heterochromatin; yet among the field voles examined so far there has been little variation. The Portuguese field voles have a chance to help us understand the evolution of some of the most extreme chromosomes in mammals.

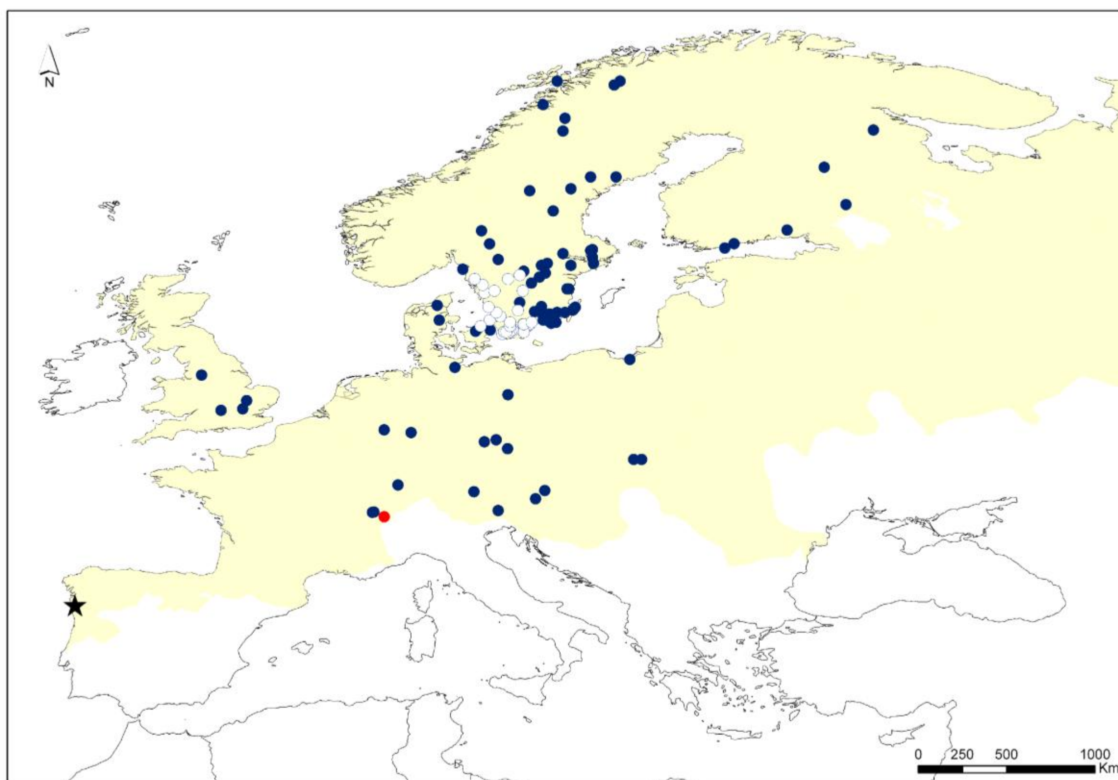


Fig. 1 Localities where field voles have been karyotyped in Europe (adapted from Fredga and Jaarola 1997). Yellow shading shows distribution of the field vole. Colouring of symbols indicates inferred molecular lineage, based on maps in Jaarola and Searle (2002, 2004). Blue: Northern lineage, standard karyotype. White: Northern lineage, Lund Y karyotype. Red: Southern lineage, standard karyotype. Star: our new Portuguese samples. Note that the actual density of sampling sites in southern Scandinavia is about double that shown here (Fredga and Jaarola 1997).

MATERIALS AND METHODS

One male and one female field vole were collected in Lagoas de Bertiandos, northern Portugal (41º55'59''N 8º38'32''W; mapped in Fig. 1). These individuals were genetically typed as belonging to the Portuguese lineage based on sequences of mitochondrial cytochrome *b*, two X-linked DBX introns (DBX5, DBX6), two Y-linked introns (DBY7, SMCY7) and two exons of the autosomal genes BRCA1 and RAG1 (Paupério *et al.* 2012). The relevant haplotypes for these data are DBY7 P1, SMCY7 P1, DBX5 P2 and CYTB PORTUGAL11 (the male); DBX5 PH23 and CYTB PORTUGAL18 (the female) and BRCA1 PH15, RAG1 P1 and DBX6 P1 (both)."

Direct bone marrow preparations were used to obtain mitotic metaphases from the two voles following the procedures of Giménez and Bidau (1994) and the chromosomes were stained with Giemsa. Cells were scored down the microscope (Olympus CX41) and illustrative metaphases were photographed using an in-built digital camera (Olympus SC30) and processed with cell^B software (Olympus). Additionally, chromosome length measurements were made on 22 spreads of the male vole using the cell^B software. The values are given as means, with variation around those means reflecting degree of chromosome contraction in different spreads.

RESULTS

The standard chromosome number of $2n = 50$ was observed in numerous metaphases of both the male and female Portuguese field voles (27 and 26 out of 30 and 32 spreads, respectively; the remainder had reduced chromosome numbers due to cell breakage). The autosome complement of the Portuguese field vole was as described previously (Matthey 1949, Ashley *et al.* 1989a): 23 pairs of acrocentrics and the smallest pair of autosomes were metacentric (Fig. 2).

For the Portuguese field voles, both the X and Y chromosomes were much larger than the autosomes and clearly should be classified as 'giant sex chromosomes' (Fig. 2).

Fig. 2 Karyotypes of the female (a) and male (b) field vole from Portugal.

The X chromosome was a metacentric chromosome with a ratio of 1.5 for the length of the long arm to the length of the short arm. The X chromosome was on

average 3.7 times longer than the length of autosome 1 (the longest autosome). The arm ratio is close to that previously recorded for field voles by Fredga and Jaarola (1997), but with a greater total length than the typical value cited by these authors: 2.5–3 times longer than chromosome 1.

The Y chromosome had a distinct but small short arm that was on average 8.3% of the total length of the Y chromosome. As such, it has an intermediate morphology between the standard Y ("minute short arm") and the Lund Y (substantial short arm, 17.8% of total length of the Y) as described and illustrated in Fredga and Jaarola (1997). Examples of Y chromosomes with different degrees of contraction are shown in Fig. 3.

The long arm of the Y chromosome was on average 13.3% longer than the long arm of the X chromosome. This is closer to typical values for the long arm of the standard Y, ranging between the same size as the X to 10% longer, than to the Lund Y, ranging between the same size as the X to 20% shorter (Fredga and Jaarola 1997).

The Y chromosome was on average 2.7 times longer than the length of autosome 1. This is greater than the typical value for field vole Y chromosomes of 1.5–1.9 times longer than chromosome 1 given by Fredga and Jaarola (1997).

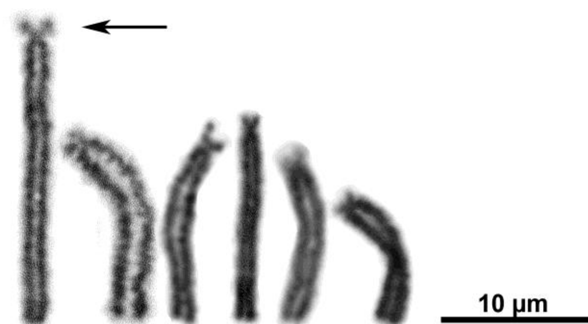


Fig. 3 Individual Y chromosomes from different metaphases from the Portuguese male field vole. Arrow indicates an example of the distinct short arm on the Y chromosome.

DISCUSSION

The giant sex chromosomes of the field vole *M. agrestis* are one of the most extraordinary features of any mammalian karyotype. The recent data that show that the field vole is a complex of three closely-related species or near-species, rather than just a single homogeneous species (Jaarola *et al.* 2004; Hellborg *et al.* 2005; *Paupério et al.* 2012), makes it particularly worthwhile to revisit the possibility of geographic

variation in the form of these sex chromosomes. Fredga and Jaarola (1997) wrote a comprehensive review of the available data on field vole sex chromosomes from wild-caught specimens, but almost all voles that have been karyotyped are from the distribution area of the Northern lineage. Therefore, our addition of karyotypic data from the Portuguese lineage is valuable. On the basis of our studies, coupled with the previous findings of Fredga and Jaarola (1997), it appears that giant sex chromosomes are found in all three lineages (Northern, Southern and Portuguese), with only minor variation in chromosome morphology. Therefore, it is reasonable to suggest that the giant sex chromosomes were already present in the common ancestor to the three lineages, which can be dated to around 70,000 years ago (Paupério *et al.* 2012). Given this recent origin of the three lineages, perhaps it is not so surprising that the size and disposition of the giant sex chromosomes is no more variable than observed. Over a longer evolutionary period, greater variability would undoubtedly be expected.

Recent mitochondrial DNA phylogenies place *M. agrestis* as closely related with *Microtus* species of central Asia, subgenus *Blanfordimys* (Jaarola *et al.* 2004, Bannikova *et al.* 2009, 2010). All of these species (*Microtus afghanus*, *Microtus bucharensis*, *Microtus juldaschi*) show asynapsis of the sex chromosomes at male meiosis but have 'normal sized' sex chromosomes (Gileva *et al.* 1982, O'Brien *et al.* 2006, Borodin *et al.* 2012). During the history of the field vole and its three constituent lineages, it is believed that population bottlenecks were important in evolution and genetic divergence (Paupério *et al.* 2012). Thus, the incredible expansion of repetitive elements that characterise the sex chromosomes in field voles may represent a feature that became fixed under genetic drift, during the early evolution of *M. agrestis*. There is evidence that the giant sex chromosomes may be mildly disadvantageous at cell division (Rens *et al.* 2006) and so their fixation by genetic drift in a bottlenecked population may explain their occurrence in and characterisation of *M. agrestis*. However, this is a great simplification of the evolutionary processes that led to the occurrence of the giant sex chromosomes in present-day populations of field voles. Clearly, there is a mutational tendency for sex chromosomes to accumulate repetitive sequences in those *Microtus* which lack meiotic sex chromosome pairing (Borodin *et al.* 1995). However, it is completely unknown whether the accumulation in *M. agrestis* occurred in multiple small steps over a long period of time or in a short burst. Thus, if genetic drift was involved to fix the new variants, it is unsure whether a single population bottleneck or repeated population bottlenecks were important. Another possibility is that the new variants were somehow favoured by chromosomal drive, and that some combination of drive and selection against variants that are too big has led to the particular size of sex chromosome that is seen in *M. agrestis*.

Whatever the evolutionary basis for the origin and fixation of the giant sex chromosomes of *M. agrestis*, they appear to be relatively stable within the species complex now. There is, however, minor variation in sex chromosome morphology. Studies previous to ours showed that populations in southern Sweden are characterised by a distinct short arm on the Y chromosome (17.8% of the total length of the Y chromosome; the Lund Y) compared with the typical condition of a “minute short arm” (the standard Y; Fredga and Jaarola 1997). Interestingly, the one male from Portugal that we examined also had a Y chromosome with a distinct short arm, but half the length of that on the Lund Y (the short arm on the Portuguese Y chromosome was 8.3% of the Y chromosome length). The total lengths of the X and Y chromosome in this Portuguese male also suggested that its sex chromosomes were somewhat larger than may be normal for the Northern lineage. However, many more specimens need to be examined to determine the ‘typical’ characteristics of the giant sex chromosomes of the Portuguese lineage. Application of G- and C-banding would also be desirable for comparison with data for the other lineages (Fredga and Jaarola 1997), and more intensive molecular cytogenetic comparisons would also be valuable.

The fact that the genus *Microtus* in general, and *M. agrestis* in particular are increasingly used as models incorporating study of molecular markers means that there is real potential for detailed analysis of chromosomal evolution, adding a context of dating and population history. In this way, the existence of the giant sex chromosomes of the field vole may become much better understood in the future. Here, we have added an extra piece to their evolutionary history, but clearly it would be fascinating to establish better the origin of these astonishing chromosomes.

Acknowledgements

This work was supported by the Fundação para a Ciência e a Tecnologia (FCT) SFRH/BD/27369/2006 PhD grant to JP. We thank the reviewer for valuable comments.

Ethical standards

Animals were trapped and ethically sacrificed according to Portuguese law and under the permit no 176/2008/CAPT.

REFERENCES

Ashley T, Jaarola M, Fredga K (1989a) The behaviour during pachynema of a normal and inverted Y chromosome in *Microtus agrestis*. *Hereditas*, **111**, 281-294.

Ashley T, Jaarola M, Fredga K (1989b) Absence of synapsis during pachynema of the normal size sex chromosomes of *Microtus arvalis*. *Hereditas*, **111**, 295-304.

Bannikova AA, Lebedev VS, Golenishchev FN (2009) Taxonomic position of Afghan vole (subgenus *Blanfordimys*) by the sequence of the mitochondrial cytb gene. *Russian Journal of Genetics*, **45**, 91-97.

Bannikova AA, Lebedev VS, Lissovky AA, Matrosova V, Abramson NI, Obolenskaya EV, Tesakov AS (2010) Molecular phylogeny and evolution of the Asian lineage of the vole genus *Microtus* (Rodentia: Arvicolinae) inferred from mitochondrial cytochrome *b* sequence. *Biological Journal of the Linnean Society*, **99**, 595-613.

Beysard M, Perrin N, Jaarola M, Heckel G, Vogel P (2012) Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (*Microtus agrestis*). *Journal of Evolutionary Biology*, **25**, 400-408.

Borodin PM, Sablina OV, Rodionova MI (1995) Pattern of X-Y chromosome pairing in microtine rodents. *Hereditas*, **123**, 17-23.

Borodin PM, Basheva EA, Torgasheva AA, Dashkevich OA, Golenishchek FN, Kartavtseva IV, Mekada K, Dumont BL (2012) Multiple independent evolutionary losses of XY pairing at meiosis in the grey voles. *Chromosome Research*, **20**, 259-268.

Fredga K, Jaarola M (1997) The origin and distribution of the Y chromosome in *Microtus agrestis* (Rodentia, Mammalia). *Hereditas*, **126**, 25-34.

Gileva EA, Bolshakov VN, Chernousova NF, Mamina VP (1982) Cytogenetical differentiation of forms in the group of Pamir (*Microtus juldaschi*) and Carruther's (*M. carruthersi*) voles (Mammalia, Microtinae). *Zoolicheskii Zhurnal*, **61**, 912-922 (in Russian).

Giménez MD, Bidau CJ (1994) A first report of HSRs in chromosome 1 of *Mus musculus domesticus* from South America. *Hereditas*, **121**, 291–294.

Hellborg L, Gündüz I, Jaarola M (2005) Analysis of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology*, **14**, 2025–2031

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.

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.

Jaarola M, Martínková N, Gündüz I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedaski-Tsolis S, González-Esteban J, López-Fuster MJ, Kandaurov AS, Kefelioglu H, Mathias ML, Villate I, Searle JB (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia), inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **33**, 647-663.

Macholán M, Filippucci MG, Zima J (2001) Genetic variation and zoogeography of pine voles of the *Microtus subterraneus/majori* group in Europe and Asia Minor. *Journal of Zoology*, **255**, 31-42.

Mandahl N (1978) Variation in C-stained chromosome regions in European hedgehogs (Insectivora, Mammalia). *Hereditas*, **89**, 107-128.

Marchal JA, Acosta MJ, Nietzel H, Sperling K, Bullejos M, Díaz de la Guardia R, Sánchez A (2004) X chromosome painting in *Microtus*: origin and evolution of the giant sex chromosomes. *Chromosome Research*, **12**, 767-776.

Matthey R (1949) Chromosomes sexuels géant chez un campagnol, *Microtus agrestis* L. *Experientia*, **5**, 72.

Mitsainas GP, Rovatsos MT, Rizou EI, Giagia-Athanasopoulou E (2009) Sex chromosome variability outlines the pathway to the chromosomal evolution in *Microtus thomasi* (Rodentia, Arvicolinae). *Biological Journal of the Linnean Society*, **96**, 685-695.

Nanda I, Neitzel H, Sperling K, Studer R, Epplen JT (1988) Simple GAT/CA repeats characterize the X chromosomal heterochromatin of *Microtus agrestis*, European field vole (Rodentia, Cricetidae). *Chromosoma*, **96**, 213-219.

O'Brien SJ, Menninger JC, Nash WG (2006) *Atlas of Mammalian Chromosomes*. Wiley-Liss, New York.

Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB (2012) Cryptic speciation in the field vole: a multilocus approach confirms highly divergent lineages in Europe. *Molecular Ecology*. DOI: 10.1111/mec.12024.

Rens W, Torosantucci L, Degrassi F, Ferguson-Smith MA (2006) Incomplete sister chromatic separation of long chromosome arms. *Chromosoma*, **115**, 481-490.

Searle JB, Erskine I (1985). Evidence for a widespread karyotypic race of hedgehog (*Erinaceus europaeus*) in Britain. *Journal of Zoology*, **206**, 276-278.

Wolf KW, Baumgart K, Winking H (1988) Meiotic association and segregation of the achiasmatic giant sex chromosomes in the male field vole (*Microtus agrestis*). *Chromosoma*, **97**, 124-133.

Zima J, Macholán M (1995) Karyotypes of common voles from Kyrghyzstan and heterochromatin variation in the sex chromosomes of *Microtus kirgisorum*. *Caryologia*, **48**, 65-74

CHAPTER 3

DIFFERENTIATION OF THE TWO LINEAGES OF THE FIELD VOLE IN IBERIA: INSIGHTS FROM MOLECULAR AND MORPHOMETRIC DATA

Paper III - Paupério J, Barbosa S, Searle JB, Alves PC (manuscript). Asymmetric introgression between the two lineages of the field vole in the Iberian Peninsula revealed by noninvasive genetic sampling

Paper IV – Paupério J, Alves PC, Searle JB (manuscript). Morphometric differentiation of the two field vole lineages in the Iberian Peninsula: a geometric analysis of mandible shape

PAPER III

Asymmetric introgression between the two lineages of the field vole in the Iberian Peninsula revealed by noninvasive genetic sampling

Joana Paupério, Soraia Barbosa, Jeremy B. Searle and Paulo C. Alves

Manuscript

ASYMMETRIC INTROGRESSION BETWEEN THE TWO LINEAGES OF THE FIELD VOLE IN THE IBERIAN PENINSULA REVEALED BY NONINVASIVE GENETIC SAMPLING

PAUPÉRIO, J.[1,2,3], BARBOSA, S.[1,3,4], SEARLE, J.B.[2,4], ALVES, P.C.[1,3,5]

ABSTRACT

A complete understanding of the speciation process requires evaluation of different genomic regions and their role in the reproductive isolation of sister taxa. The *Microtus agrestis* species complex has recently proved to be a good candidate system in the study of speciation. Three highly divergent lineages with neighbouring geographic distributions have been described in Europe: Northern, Southern and Portuguese, the last two found in the Iberian Peninsula. Previous data on the extent of reproductive isolation between these lineages was limited to a restricted region of the contact zone between the Northern and Southern lineages, where evidence of post-mating isolation was detected. Considering the previously documented low abundances of field vole populations in northwestern Iberia we used noninvasive genetic sampling techniques to establish the distribution of the Portuguese and Southern lineages and to assess the degree of introgression in the putative secondary contact zone. Our results show that the Portuguese lineage has a wider distribution than previously reported, ranging from central Portugal to the western part of Asturias and León districts in northwestern Spain. Hybrids were found across a wide area but there was also evidence of the existence of barriers to gene flow. None of the hybrids detected (29) were male, suggesting that hybrid males are strongly selected against. The introgression of Southern lineage alleles into the Portuguese lineage was higher than in the other direction. The introgression most likely reflects a trail of alleles left from movement of the hybrid zone between the two lineages. All the evidence collected suggests an already advanced, although recent, speciation process, supporting the view that most likely *Microtus agrestis* is a complex of three separate species or near-species.

Keywords: *Microtus agrestis*, secondary contact, Haldane's rule, speciation, hybrid zone

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, InBIO Laboratório Associado, Campus Agrário de Vairão, 4485-661, Vairão, Portugal;
[2]Department of Biology, University of York, Wentworth Way, York YO10 5DD, United Kingdom
[3]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal;
[4]Department of Ecology and Evolutionary Biology, E139 Corson Hall, Cornell University, Ithaca, NY 14853-2701, USA
[5] Wildlife Biology Program, College of Forestry and Conservation, University of Montana, 32 Campus Drive, Missoula MT 59801, MT, USA.

INTRODUCTION

How species are formed is a fundamental question for evolutionary biologists. Speciation can occur through the progressive differentiation of geographically isolated gene pools, if these are maintained isolated for enough time. This allopatric isolation in combination with genetic drift and coupled with local selective pressures can result in the establishment of barriers to gene flow that ultimately may lead to the emergence of new species. However, if secondary contact between these isolated populations occurs, before the formation of pre or post-zygotic isolation mechanisms, introgression of some genes between these populations or incipient species may be observed.

Speciation may occur without perceivable changes in morphology, resulting in what may be designated as 'cryptic' species. In fact, the advent of molecular techniques has led to the detection of numerous morphologically 'cryptic' lineages, which in some cases comprise incipient or even fully differentiated species (Bickford *et al.* 2007). In the cases of these cryptic lineages, where high genetic divergence at one or more markers is found in the absence of morphological distinctiveness, analysis of existing secondary areas of contact can elucidate the extent and nature of reproductive isolation (Barton and Hewitt 1985). Although the majority of studies in secondary areas of contact concerns taxa that are morphologically or karyotypically distinct, some detailed work has already been performed in contact zones of highly differentiated cryptic lineages (Barton and Hewitt 1985; Hewitt 2001; Phillips *et al.* 2004).

Hybrid zones established between two genetically distinct populations can exhibit a variety of structures depending on the degree of genetic and ecological differentiation of these groups (Harrison 1993). When the hybrid zone is formed, clines in character state frequencies appear. The shape of these clines is often a result of a balance between dispersal capabilities and the selection acting on hybrid individuals, with the contact zone referred to as a tension zone in this case (Barton and Hewitt 1985). A strong effect from selection coupled with small dispersal distances will result in narrow and steep clines between populations, while relative large dispersal and neutral mixing will generate wide and shallow clines. Differential selection on different traits may lead to cline shapes that are not concordant among loci. In fact, introgression of different genes depends often on their mode of inheritance. For instance, in mammals, mitochondrial DNA (maternally inherited) frequently shows more introgression than autosomal markers, which in turn show more introgression than the Y chromosome (paternally inherited; *e.g.* Dod *et al.* 1993, Balloux *et al.* 2000). Actually,

there is evidence that sex chromosomes carry more genes related with low hybrid fertility and/ or viability than autosomes (Turelli and Orr 1995) and the observation of differential patterns of introgression in sex chromosomes is often explained by Haldane's (1992) rule "when in the F1 offspring of the two animal races one sex is absent, rare, or sterile, that sex is the heterogametic sex". In addition, barriers to gene flow may be asymmetric, allowing for more introgression of traits of one population into the other than in the reverse direction. For example, mtDNA introgression is often asymmetric (*e.g.* Berthier *et al.* 2006; Alves *et al.* 2008; Boratyński *et al.* 2011). This may be caused by differences in population structure, fitness, by cline movement or selection (Barton and Hewitt 1985). Recently, Petit and Excoffier (2009), in a review of studies of introgression between species with sex-biased dispersal, considered that in these situations a simple neutral model, which relies on demographic processes, could account for differences in introgression among both loci and species.

The Eurasian field vole, *Microtus agrestis*, is a species complex where three distinct evolutionary units have been described, Northern, Southern and Portuguese, yet without any apparent morphological differentiation (Jaarola and Searle 2002, 2004, Helborg *et al.* 2005, Paupério *et al.* 2012). Data from different genomic regions was analysed in these studies (mitochondrial and autosomal genes, X and Y chromosome introns), and gave congruent results, showing three highly differentiated lineages. The estimated times of divergence are fairly recent, dating back to the last glacial period; the Portuguese lineage split occurred first (estimated at ca. 70 000 years BP) and the Northern and Southern lineages separated at around the Last Glacial Maximum (estimated at ca. 18 500 years BP; Paupério *et al.* 2012). An area of secondary contact between the Southern and Northern lineages, located in the Swiss Jura was studied recently (Beysard *et al.* 2011) and showed evidence of gene flow. However, only a small number of hybrids were detected and most of the introgression was at mitochondrial loci. In addition, in this study no introgression was detected on the Y chromosome (no mismatches were observed between the autosomal and Y chromosome loci) which suggested that at least some post-mating isolation mechanisms were already in place. Regarding the contact of the Portuguese and Southern lineage, the only possible instance of gene flow was detected by Paupério *et al.* (2012) in Picos da Europa, however the limits of the lineage range in the Iberia Peninsula have not previously been well-described.

Karyotypic variability is also an important aspect that should be considered when studying the degree of reproductive isolation between two populations or species. In fact, chromosomal rearrangements may contribute or lead to reproductive isolation as has been shown for example in the house mouse (Searle 1998, Piálek *et*

al. 2005). In the field vole there is also some indication of sex chromosome variability. The three lineages are characterised by giant sex chromosomes (with extremely large blocks of heterochromatin) and some variation in the structure of the Y chromosome has been observed within the Northern lineage (the Lund Y variant, Fredga and Jaarola 1997). Recently, few Portuguese individuals were karyotyped, which displayed some minor differences in the sex chromosome size and structure in relation to the karyotype described for the Northern lineage (Giménez *et al.* 2012).

Even though previous work on this species complex has been extensive, ranging through the entire area of the field vole distribution and considering different genomic regions (mtDNA, autosomes and X and Y chromosome), questions concerning the distribution limits of the two lineages in Iberia remain unanswered. Also, an assessment of the extent of reproductive isolation between the lineages would be of major importance, since it would greatly contribute to clarifying their possible species status. However, previous analyses showed that the Portuguese and Southern lineages have contemporary population sizes 10- to 100- fold smaller than the Northern lineage, and that they are particularly difficult to collect in northwestern Iberia, mainly in Galicia (Paupério *et al.* 2012). Fortunately, rodent DNA can be obtained from bone material in owl pellets or from museum collections, allowing considerable augmentation of individuals collected by live-trapping (Taberlet and Fumagalli 1996; Beja-Pereira et al, 2009; Barbosa *et al.* 2012).

Therefore, in this study we focussed specifically on patterns of genetic diversity within the Iberian Peninsula by using a noninvasive genetic sampling approach. We performed detailed sampling of bones from owl pellets and museum samples within Iberia to determine the distribution and genetic structure of the two lineages in this area. We compared patterns of genetic variability at loci with different modes of inheritance and different population sizes: mtDNA, X chromosome, Y chromosome and autosomes. We then applied cline shape analysis to assess the location of the secondary contact zone and compare patterns of introgression between the loci studied. We expect that the comprehensive data collected will identify the area of secondary contact between the Portuguese and Southern lineages and clarify the degree of reproductive isolation between the two lineages.

Considering the recentness of separation of these groups and the different patterns of introgression already detailed for various other species, the information collected will greatly contribute to a better understanding of the processes of allopatric speciation and secondary contact acting on small mammals in southern Europe.

MATERIALS AND METHODS

Samples and DNA extraction

Mandibles, skulls and some long bones of *M. agrestis* were obtained from owl pellets retrieved from 109 localities within the Iberian Peninsula. Most of the bones from owl pellets were from long-established collections: material from Spain corresponded mostly to owl pellets collected during the 1970s and 1980s and were analysed for the Spanish mammal atlas (Palomo and Gisbert 2002) and for the Galician atlas of vertebrates (SGHN 1995); material from Portugal was more recent but also used for faunistics (Mira *et al.* 2003; Gonçalves 2006). Material from 4 of the 17 Portuguese localities was collected and screened by us. In total 377 field vole bones, ranging from 1 to 5 samples per locality (Appendix I), were extracted using the DNeasy® Blood and Tissue Kit (Qiagen, Hilden, Germany) following the “Purification of total DNA from compact animal bone” protocol with some adaptations (Barbosa *et al.* 2012). Additionally 23 skins (museum samples), from 11 localities in Northeast Spain, were extracted with the QIAamp® DNA Micro Kit (Qiagen, Hilden, Germany) after soaking for two weeks in Tris-EDTA buffer. All the extraction procedures were performed in a physically isolated room maintaining conditions to reduce risk of DNA contamination (Gilbert *et al.* 2005).

DNA amplification and sequencing

Considering the loci previously studied on this species (Paupério *et al.* 2012), we selected one mitochondrial locus (cytochrome *b* - cyt*b*) and three nuclear loci (exon 11 of BRCA1, intron 5 from the DBX gene on the X chromosome and intron 7 from the DBY gene on the Y chromosome) for amplification in the bone and skin samples. For the cyt*b* gene a central fragment of 385bp was amplified with the primers L15162M2 and H15576M (Jaarola and Searle 2002, 2004). For the nuclear loci, shorter DNA fragments were selected, from alignments produced with published sequences (AY541718 and AY541707, Hellborg *et al.* 2005; JX269135, JX284285-JX284300, JX284366/7, JX284370-JX284376, Paupério *et al.* 2012), targeting regions that discriminated the two lineages of field vole present in Iberia. The primer design for these sequences followed the recommendations of Waits and Paetkau (2005) for fragment size in highly degraded DNA samples, not exceeding 300 bp. Polymerase chain reactions (PCR) were 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. A total of 45 cycles was carried out with annealing temperatures of 57ºC for the cyt*b* and DBX5 loci and 62ºC for the BRCA1 and DBY7 and final extensions of 5 min. The PCR product was purified using ExoSAP-IT® PCR clean-up Kit (GE Healthcare, Piscataway, NJ, USA) and sequences were obtained with the amplification primers. Sequencing reactions were carried out using the BigDye XTerminator cycle sequencing kit (Applied Biosystems, Carlsbad, California, USA). Each DNA fragment was sequenced in one direction (usually the reverse) on a 3130*xl* Genetic Analyser Sequencer (Applied Biosystems/HITASHI).

Genetic analysis

The sequences obtained were visually inspected and ambiguous bases were resolved in Sequencher version 4.7 (Gene Codes Corporation, Ann Arbor, MI). Polymorphic positions for individual sequences from the DBX5 and BRCA1 loci were carefully inspected to ensure correct and consistent identification of double peaks in heterozygotes. Additionally, heterozygous positions for insertion/deletions were resolved directly from offset chromatogram peaks (Flot *et al.* 2006). For the DBX5 and BRCA1 loci, haplotypes were inferred using Phase version 2.1 (Stephens *et al.* 2001; Stephens and Donnelly 2003) with three runs performed for each locus with 1000 burn-in steps and 1000 iterations. For the tri-allelic positions the parent-independent mutation model (-d1 command) was used and in the input file these positions were designated as locus type M. Phase results were consistent across runs but included a small percentage of low probability phase calls (2% and less than 1% with probability lower than 0.95 for BRCA1 and DBX5 respectively). These correspond to rare alleles (usually found heterozygous with a common allele) and were included in the analysis since omitting them could lead to reductions in the estimated diversity and a bias in the estimates of population genetic parameters (Garrick *et al.* 2010). All the new sequences will be deposited in GenBank.

Previously published sequences for the Portuguese and Southern lineage of field vole which included the targeted regions of the studied loci were retrieved from GenBank. Thus, 48 cyt*b* gene haplotypes (AY167186/7 and AY167160-AY167164, Jaarola and Searle 2002; AY303151-AY303160, Jaarola and Searle 2004; JX284253-JX284283, Paupério *et al.* 2012), and all the sequences available for DBX5 and BRCA1 in Paupério *et al.* (2012) were included in the analysis. Alignments with published and new sequences for each locus were generated with Clustal W (Thompson *et al.* 1994), implemented in BioEdit version 7.0.5.3 (Hall 1999) and manually edited.

As networks are useful tools in the analyses of closely related sequences, providing a reliable estimate of the true genealogy (Cassens *et al.* 2005; Woolley *et al.* 2008), we computed phylogenetic networks for each locus with the median-joining algorithm implemented in the software Network 4.6.0.0 (Bandelt *et al.* 1999; http://www.fluxus-engineering.com). In the network construction the insertion/deletion polymorphisms present in the DBX5 locus were considered as biallelic characters (presence/absence of the insertion or deletion), independently of their size.

Diversity statistics (nucleotide diversity, π and haplotype diversity, *Hd*) and neutrality tests (Tajima's D, Tajima 1989; Fu's F_S, Fu 1997 and R_2, Ramos-Onsins and Rozas 2002) were estimated per locus for each lineage using DnaSP v5 (Librado and Rozas, 2009). Departures from a neutral model were tested with 10 000 coalescent simulations of the genealogy.

Cluster analysis

The spatial structure in the data was inferred through a Bayesian cluster analysis with Geneland v.4.0.2 (Guillot *et al.* 2005b; Guillot *et al.* 2012). This program incorporates the spatial distribution of the individuals (geographic coordinates) in the model in order to detect genetic discontinuities between populations (Guillot *et al.* 2005a). For this analysis we have only considered the individuals for which there was information for more than one locus and each unique haplotype was coded as a different allele (Appendix II). For the spatial information we considered some uncertainty of the spatial coordinates (0.01º) in the analysis, to account for the fact that most of the samples were collected from barn owl pellets and thus originated from populations within the hunting area of this species (considered as a 1 km radius from the nesting site, Andries *et al.* 1994). The dataset was analysed using the spatial model under the uncorrelated allele frequency prior (UFM). We performed 10 independent Markov Chain Monte Carlo (MCMC) runs of 1000000 iterations with a thinning of 500 iterations, and discarding the first 20% as burnin. We allowed for the number of populations (*K*) to vary between 1 and 10.

Cline analysis

The median joining networks of the loci clearly separated the two field vole lineages in Iberia and allowed the unambiguous assignment of all haplotypes to either the Portuguese or the Southern lineage. Based on this assignment, for the purpose of cline analysis, individuals were labelled as pure 'Portuguese', when all loci had Portuguese lineage haplotypes, as pure 'Southern', when all loci had Southern lineage haplotypes,

or as 'hybrids', if there were both Portuguese and Southern lineage haplotypes in the same individual.

For the purpose of cline analysis and given the nature of the available data, the sample sites were collapsed into a one dimensional transect. For that we considered the geographic distribution of the two lineages in the Iberian Peninsula and extracted the median line of that area with the tool Polygon to Centerline in ESRI® ArcMap™ 10.0. Each sampling locality was then projected orthogonally onto this transect and geographic distances between sampling points in the transect were measured, going from the site located in the southwestern limit of the range to the north-east of Spain. Considering the uncertainty associated with the geographic coordinates of most of the samples mentioned above, localities that were separated by less than 1km were grouped. For the purpose of cline fitting, all markers were collapsed to a two allele system (Portuguese or Southern). Allele frequencies were determined for each locality in the transect by gene counting and weighted by sample size.

Maximum likelihood clines were fitted independently to each locus using the compound tanh and exponential model developed by Szymura and Barton (1986) implemented in Analyse v.1.3 (Barton and Baird, 1998). Hence, the cline shape parameters, cline centre (c) and cline width (w), were estimated for each locus by fitting a parametric sigmoid cline in allele frequencies (p), according to the equation

$$p = (1 + \tanh [2 (x - c) / w]) / 2$$

where $(x - c)$ is the distance to the centre of the cline (Szymura and Barton, 1986).
The coincidence of cline centre and the concordance of cline width were evaluated by the pairwise comparison of cline shape of the distinct loci as described in Yannic *et al.* (2008). Cline centre and width were estimated for one of the loci. Then the likelihood of the observed data was determined for the second locus by assuming the values of cline centre and width estimated for the first locus (constrained model). This likelihood was compared with the likelihood obtained in the unconstrained model (likelihood of the parameter obtained for the second genomic region). The two likelihoods were then compared assuming that twice the difference in log-likelihood ($G = -2\Delta LL$) between the constrained and unconstrained model follows a chi-squared distribution (Szymura and Barton 1986). Bonferroni corrections were used to account for multiple testing.

In addition, sex linked introgression was investigated by analysing the association between individual lineage assignments on different loci. These were tested using Fisher's exact test.

RESULTS

Amplification success

From the 377 bone samples analysed, we obtained amplifiable DNA in 342 samples (90% of the total) from 108 localities (Appendix I). However, amplification success varied among the amplified markers. The mitochondrial marker (cyt*b* gene) provided the highest rates, since we obtained unambiguous sequences in 334 samples (89%). The nuclear markers showed lower amplification success (75%), with the X-linked marker (DBX5) giving similar results (76%) to the autosomal gene (BRCA1; 73%). The full set of mitochondrial, autosomal and X-linked markers was scored in 248 samples (66%). For the Y-chromosome intron (DBY7), we obtained clear sequences in 101 samples, 90 of which had a complete genotype (i.e, all markers sequenced). All other 190 individuals, for which we were able to successfully amplify the X-linked marker, were considered to be females. However, there is the possibility that a certain proportion of these are males for which the Y-linked marker failed.

Considering the DNA extractions from skins, amplifiable DNA was obtained in only 11 samples (48%) from 4 localities. Of these, eight samples (35%) were amplified for the nuclear markers (DBX5 and BRCA1), but we only obtained unambiguous sequences for the mitochondrial gene, cyt*b*, in six individuals (26%). Five of the eight samples sequenced were males from which we generated clear sequences of the DBY7 intron.

Genetic diversity within Iberia

All molecular markers analysed resulted in networks that show the two highly divergent clades (Portuguese and Southern) separated always by ten mutational steps for cyt*b* and two/three mutational steps for nuclear markers (Fig. 1). In addition, considerable variation in the levels of diversity was found among lineages and among loci. The mitochondrial gene was the marker that showed the highest levels of diversity and divergence between lineages, while the Y chromosome intron revealed the lowest levels of polymorphism, with all males within the Portuguese lineage sharing the same haplotype (Fig. 1, Table 1). Of the two lineages, the Portuguese revealed particularly low values of haplotype (0.68) and nucleotide (0.28) diversity for the mitochondrial gene. These diversities were also lower in the Portuguese lineage for other loci, with the exception of the X chromosome intron where this lineage showed higher diversity values than the Southern lineage. Both lineages gave highly significant values for most neutrality statistics for most loci, reflecting an excess of rare alleles (Table 1).

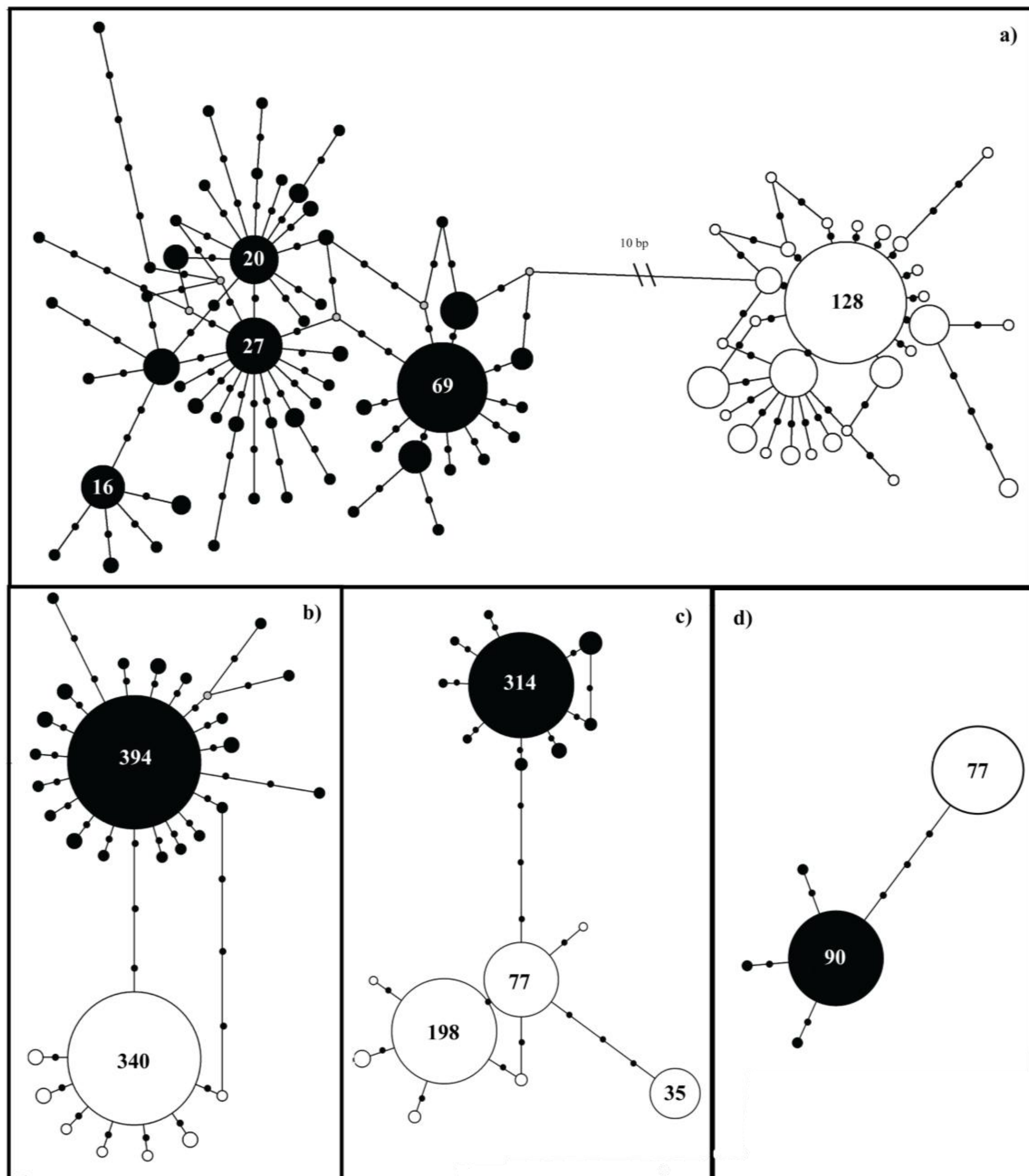


Fig. 1 Median-joining networks of the molecular markers analysed in the Portuguese and Southern lineages of *M. agrestis* (number of sequences for each locus in Table 1): a) Cyt*b* gene; b) BRCA1 gene; c) DBX5 intron on the X-chromosome; d) DBY7 intron on the Y-chromosome. Each circle represents one haplotype and circle area is proportional to the frequency of each haplotype. Total frequency is indicated for the most common haplotypes. Branches are proportional to the number of nucleotide differences between haplotypes and dots on branches indicate mutational steps. Black is for Southern lineage haplotypes and white for Portuguese lineage haplotypes.

Cluster analysis

The cluster analysis performed with Geneland provided congruent results across the ten independent runs. Two populations were always estimated that had a distribution consistent with the two divergent lineages described. The results obtained in the best run (with the highest likelihood) are presented in Fig. 2. One of the populations corresponds to the Portuguese lineage (blue) and has a distribution that includes central Portugal and the northwest of Spain, reaching the western part of Asturias and

León region. The other cluster corresponds to the Southern lineage (red) with a range in Iberia that goes from Asturias and León to the northeast of Spain. These two clusters appear as being highly differentiated (F_{ST} = 0.72). Regarding the assignment of the individuals to each cluster, individuals with hybrid genotypes were always assigned to one of the populations with a probability higher than 0.8, which results in a narrow border between the two lineages crossing Asturias and León regions. We only detected one misclassified individual (in the 420 individuals included in the analysis); this individual had a Southern lineage genotype and was assigned to the Portuguese lineage.

Table 1 Levels of polymorphism at each locus in both lineages of field vole.

Locus	Lineage	*L*	*n*	*S*	*H*	*Hd* (s.d)	*π* % (s.d.)	D	F_S	R_2
Cyt*b*	Portuguese	385	220	26	29	0.65 (0.04)	0.28 (0.02)	-2.07**	-31.34**	0.02*
	Southern	385	232	54	54	0.88 (0.02)	0.86 (0.03)	-1.88**	-46.46**	0.03*
DBX5	Portuguese	123	320	7	8	0.55 (0.02)	0.74 (0.05)	-0.58	-1.31	0.06
	Southern	123	330	6	7	0.05 (0.02)	0.05 (0.02)	-1.80**	-13.77**	0.02**
DBY7	Portuguese	233	77	0	1	-	-	-	-	-
	Southern	233	93	3	4	0.06 (0.03)	0.03 (0.01)	-1.62*	-5.56	0.06
BRCA1	Portuguese	164	350	7	8	0.06 (0.02)	0.03 (0.01)	-1.88**	-16.58**	0.01**
	Southern	164	422	25	22	0.13 (0.02)	0.10 (0.02)	-2.49**	-48.88**	0.01

L – number of sites; *n* – number of chromosomes; *S* – number of segregating sites; *H* – number of haplotypes; *Hd* – haplotype diversity; π – nucleotide diversity per site; D – Tajima's D; F_S – Fu's F_S; *p<0.05; **p<0.01; R_2 – Ramos-Onsins and Rozas's R_2; *p<0.05; **p<0.01.

Distribution of hybrids

The classification of the individuals as pure Portuguese, pure Southern or 'hybrid' based on the assignment of each haplotype to the Portuguese and Southern lineage allowed the detection of a small percentage of 'hybrids'. Considering the individuals for which there was information for at least two loci (n=420), 7% (29/420) were assigned as 'hybrids', while 44% (184/420) were considered as pure Portuguese and the remaining 49% (207/420) as pure Southern. Taking into account only individuals with complete genotypes, 48% (166/344) were assigned as pure Portuguese, 47% (163/344) were pure Southern and only 4% (15/344) were 'hybrids'. The spatial analysis of this individual assignment provides results congruent with the cluster analysis, with all of the pure Portuguese lineage individuals occupying the area that encompasses central Portugal and northwest Spain, while the pure Southern individuals are located in the area that goes from Asturias and León into the northeast of Spain (Fig. 3). All of the hybrids were detected in an area that ranged from Galicia to Burgos region, revealing extensive introgression of Southern alleles in the Portuguese lineage, and a smaller introgression in the other direction (Figs. 3 and 4).

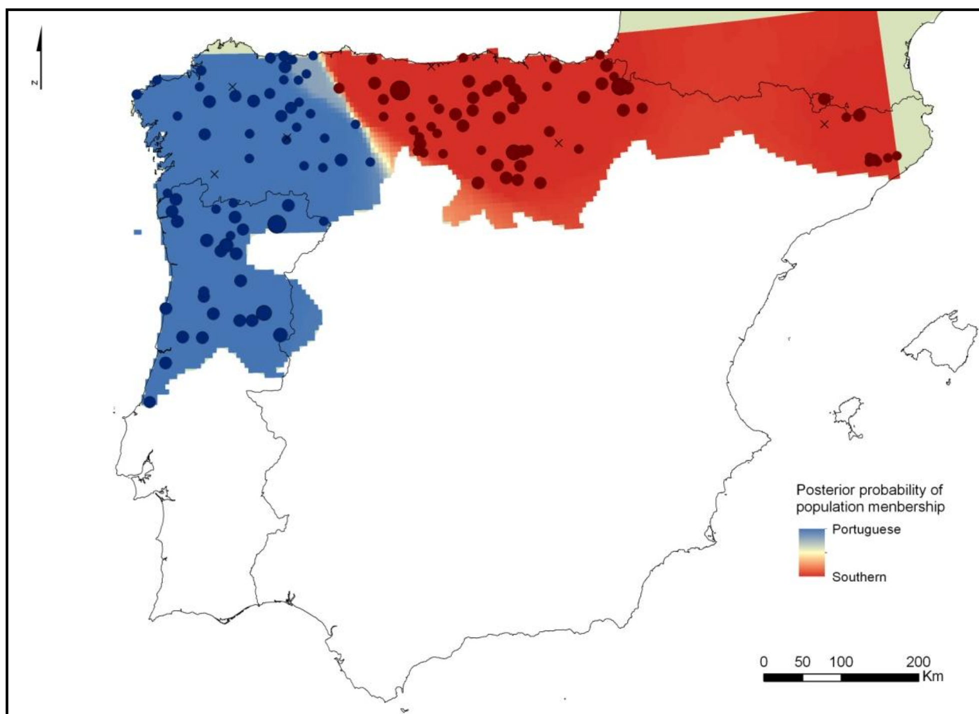


Fig. 2 Population structure inferred by Geneland, under the UFM, for the field vole in the Iberian Peninsula (*K*=2). Dots represent sampling localities coloured according to the modal membership to each cluster (blue: Portuguese; red: Southern). The size of the dots is proportional to the number of samples analysed. X represents localities where we were able to amplify only one locus, which were not used in the analysis. The colour gradient refers to the posterior probability of assignment to each of the two clusters.

Although the area of occurrence of hybrids is wide, only 18% of the individuals (or 12% if considering only individuals with complete genotype) from that area show signs of admixture. Of these, the proportion of the genotype assigned to each lineage is quite variable and no F1 hybrid (considering only the diploid loci, DBX5 and BRCA1, in females) was detected. Moreover, none of the localities sampled included pure individuals from both lineages (Fig. 4).

A more thorough analysis of the haplotype composition at different loci, suggests differential introgression of the studied loci (Fig.4). The Y intron appears as the least introgressed, while BRCA1 reveals the most extensive introgression. In fact, of the 15 individuals, with all loci scored, labelled as 'hybrids' 10 (66%) refer to introgression of the nuclear exon. Moreover, in the area of occurrence of hybrids there is a significantly higher proportion of rare alleles both in BRCA1 and cyt*b* genes (cyt*b*: 27.3% in the hybrid zone vs 16.4% in the rest of the area, Fisher's exact test $p = 0.0079$; BRCA1: 16% in the hybrid zone vs. 2.1% in the rest of the area, Fisher's exact test $p < 0.0001$).

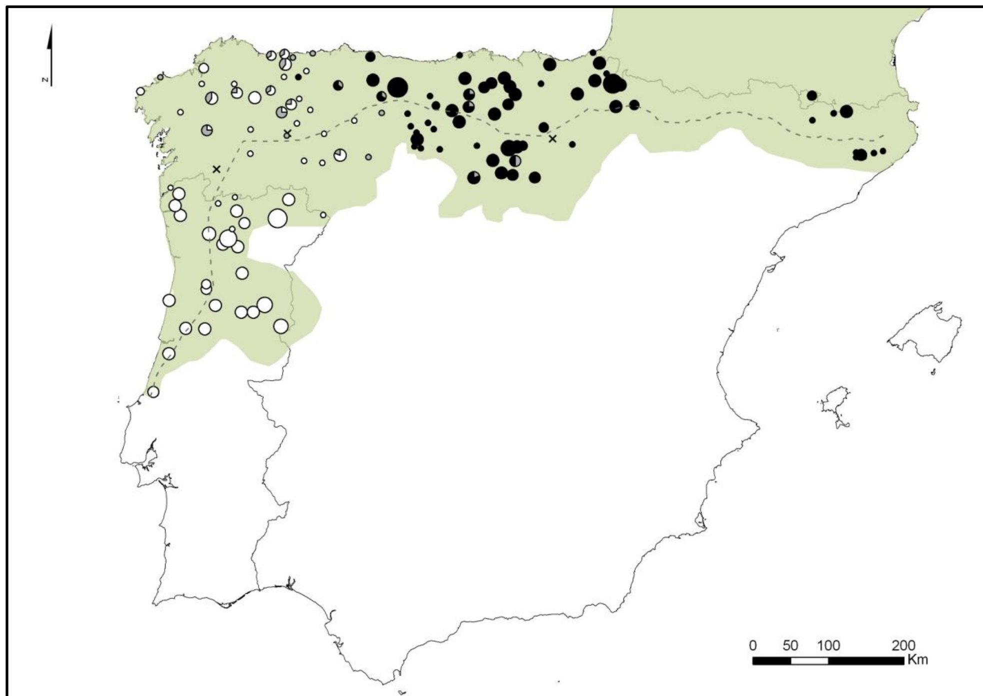


Fig. 3 Sampling localities in the Iberian Peninsula. Each pie chart represents one locality and the area is proportional to the number of samples analysed. Colours indicate the frequencies of each lineage, considering the lineage assignment at all molecular markers analysed: black: Southern lineage; white: Portuguese lineage; grey: hybrid individuals. X represents localities where we were able to amplify only one locus. The dashed line corresponds to the one dimensional transect designed considering the field vole distribution in Iberia. The green shading corresponds to the distribution range of the field vole (*Microtus agrestis*) in the Iberian Peninsula according to Kryštufek *et al.* (2008).

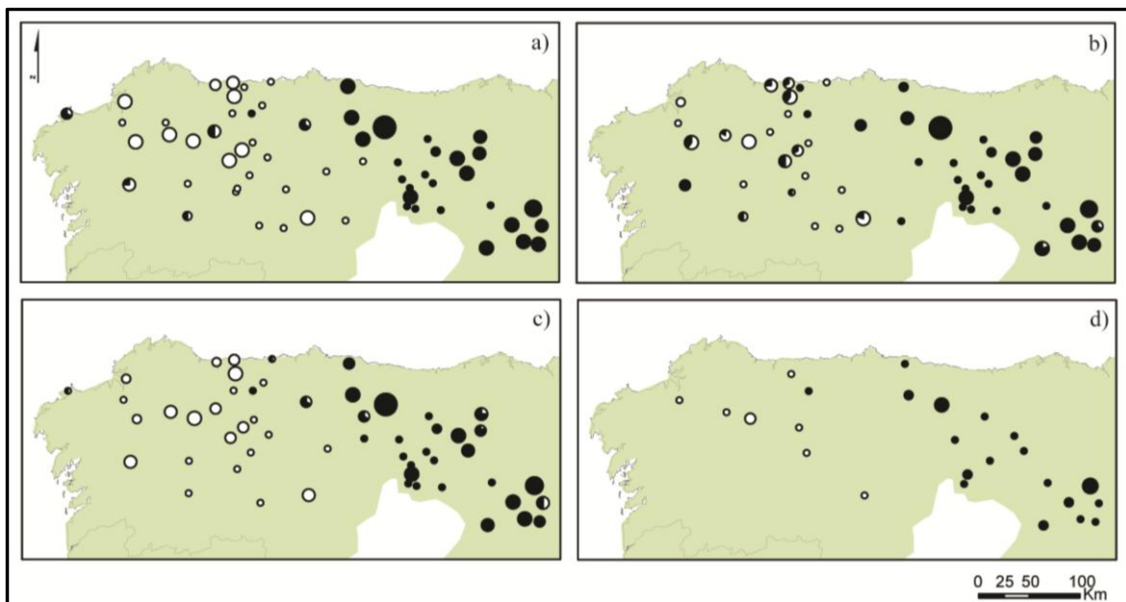


Fig. 4 Detail of the frequencies of the lineage typical genetic markers per locality in the hybrid zone: a) Cyt*b* gene; b) BRCA1 gene; c) DBX5 intron on the X-chromosome; d) DBY7 intron on the Y-chromosome. Each pie chart represents one locality and the area is proportional to the number of samples analysed. Colours indicate the assignment of the individuals to each lineage. Black: Southern lineage; white: Portuguese lineage.

Cline shape analysis

Cline shape estimates for each genomic region are presented in Fig. 5. Cline centre estimates of all loci ranges between around km 470 and 530 of the transect, which corresponds to the area encompassing the western part of the districts of Astúrias and Léon. Estimated cline width varies between 55 and 208 km. Cline shape comparison revealed significant differences between genomic regions both for cline centre and cline width. Specifically, the cline centre of DBX5, located more to the east in the transect (*c* = 530 km, Fig. 5), is significantly different from the cline centres obtained for all other loci ($p<0.001$). Moreover, the cline width estimated for the Y chromosome intron is the narrowest (*w* = 55 Km, Fig. 5), exhibiting highly significant differences from other genomic regions ($p<0.001$). Finally, the BRCA1 cline shape shows significant differences from the other loci in both cline width ($p<0.001$) and cline centre ($p<0.001$), with the exception of the comparison with DBY7 for cline centre ($p>0.05$).

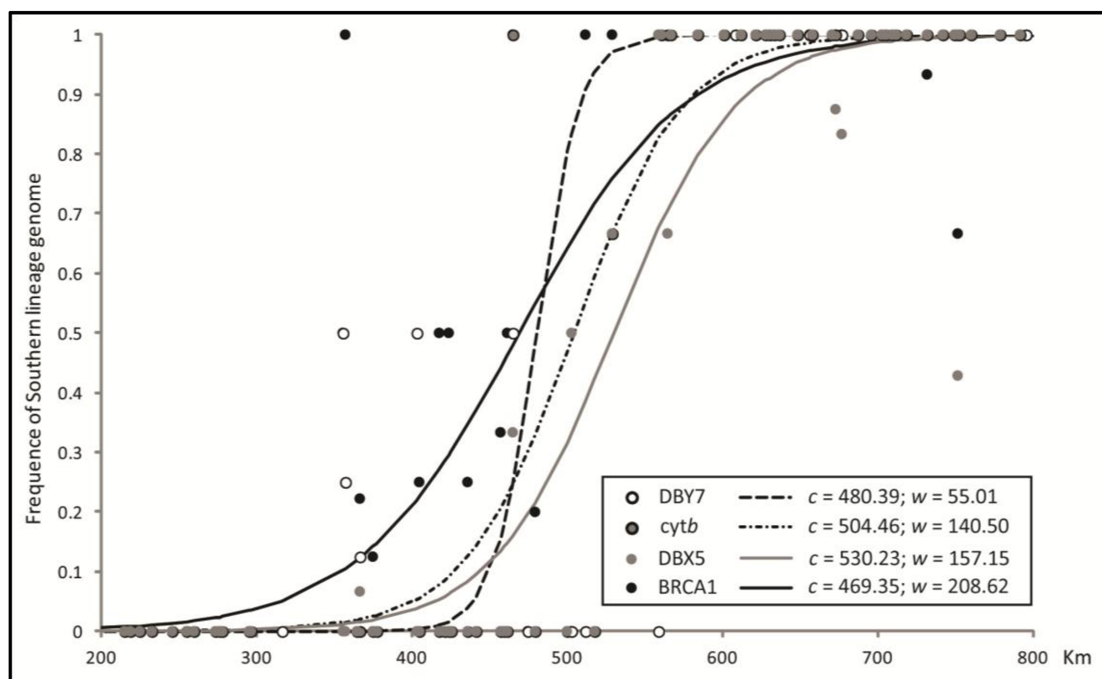


Fig. 5 Southern lineage allele frequencies plotted against geographic position on the transect (Km) for the four markers analysed (only localities between km 200 and 800 of the transect were plotted for ease of visualisation). Dots correspond to the observed frequencies and lines to the cline fitted; *c* is the cline centre and *w* the cline width.

Contrasting the assignment at different loci at the individual level revealed differential introgression of the sex-linked markers. The proportion of 'hybrid' individuals detected is significantly higher in females either considering all individuals or only individuals with a complete genotype (N_{female} = 29/236; N_{male} = 0/161; Fisher's exact test $p < 0.0001$; N_{female} = 15/200; N_{male} = 0/144; Fisher's exact test $p = 0.0003$), since all

males were considered either as pure Portuguese (51.4%; 74/144) or pure Southern (48.6%; 70/144). In addition, the proportion of males within the area of occurrence of hybrids is highly significantly lower than for the rest of the ranges of both lineages (all individuals: $N_{hybridarea}$ = 39/145; $N_{outside}$ = 125/255; Fisher's exact test $p < 0.0001$; complete genotype: $N_{hybridarea}$ = 36/126; $N_{outside}$ = 108/218; Fisher's exact test p = 0.0002). However, most of these males refer to the individuals identified as such through the amplification of the Y intron, since the majority of the samples came from barn owl pellets. Although we have only considered as females individuals from which the X marker had successful amplification, failure of amplification of the Y locus is nevertheless possible in these individuals. Moreover, as the specimens analysed had different collection dates and states of conservation, there was the possibility of a bias in the amplification of the different markers in different areas. Hence, in order to check if the significant lower proportion of males within the area of occurrence of hybrids could be due to a bias in the amplification of sex linked markers we have estimated the success of the X intron both within and outside the hybrid area. The results indicate that the amplification of DBX5 has a significantly lower success in the hybrid area (hybridarea: 123/160 - 77%; outside: 171/193 – 89%; Fisher's exact test p = 0.004). Considering this result, and that the Y marker could have similar bias of amplification, we have applied a correction factor to the number of males both within and outside the hybrid area, which reflect the failure of the X marker. Thus, the raw numbers of males were multiplied by 1.23 and 1.11 within and outside the hybrid area respectively. Nevertheless, and despite the application of this correction factor, the proportion of males is still highly significantly lower within the hybrid area (all individuals: $N_{hybridarea}$ = 46/145; $N_{outside}$ = 134/255; Fisher's exact test $p < 0.0001$; complete genotype: $N_{hybridarea}$ = 44/126; $N_{outside}$ = 120/218; Fisher's exact test p = 0.0005) which indicates that this result is not biased.

DISCUSSION

Portuguese and Southern lineage distribution in Iberia

The extensive analyses of the genetic variability of field voles in the Iberian Peninsula presented here give a clear picture of the distributional range of the Portuguese and Southern lineage. Previous work already identified these two lineages in Iberia as distinct evolutionary units (Jaarola and Searle 2004; Paupério *et al.* 2012). These previous studies also characterised the genetic variability of the lineages and gave some insights into the evolutionary history of the *M. agrestis* species complex, but were

not able to define accurately their range limits in Iberia due to difficulties in sampling. The large availability of bones from owl pellets throughout almost the whole range of the species in Iberia, allowed us to bridge that gap. A good sample size and geographic coverage was obtained in a species that is uncommon in the area studied. Moreover, despite the constraints associated with non-invasive sampling (low DNA quantity and quality, PCR inhibitors and DNA contamination; Rådström *et al.* 2004; Waits and Paetkau 2005) we were able to amplify DNA fragments from different genomic regions with a success comparable to the obtained by Barbosa *et al.* (2012) in their study on Iberian rodents (cyt*b* - 88%; nuclear gene - 64%).

The range of the Portuguese lineage is wider than previously shown (Jaarola and Searle 2004; Paupério *et al.* 2012), occupying all of Galicia and a considerable area of the provinces of Asturias and León in northern Spain. Interestingly, the distribution range of the two field vole lineages in the Iberian Peninsula resembles that of another small mammal species complex, the *Sorex araneus* complex, in which the constituent species *Sorex granarius* and *Sorex coronatus* have distributions comparable to those of the Portuguese and Southern lineage, respectively (Aulagnier *et al.* 2008, Palomo *et al.* 2008). Remarkably, these two species have similar levels of mtDNA divergence to those estimated in the field vole, but are thought to have diverged earlier (Yannic *et al.* 2008). There are also some species of amphibians and reptiles with similar distribution ranges to that of the Portuguese lineage, being restricted to northwestern Iberia, although to different extents, such as *Lacerta schreiberi* (Sá-Sousa *et al.* 2009a) and *Podarcis bocagei* (Sá-Sousa *et al.* 2009b). Other examples exist of intraspecific differentiation within amphibians and reptiles in which subspecies or mtDNA lineages have been described for that area (*e.g. Salamandra salamandra*, Garcia-Paris *et al.* 2003; *Lissotriton boscai*, Martinez-Solano *et al.* 2006; *Lacerta lepida*, Miraldo *et al.* 2011).

Cluster analysis

The spatial cluster analysis, performed with Geneland provided new insights into the partitioning of the genetic diversity within the Iberian populations of the field vole. The use of the uncorrelated allele frequencies prior, which has the ability to detect highly differentiated populations, consistently detected the two lineages and identified as an area of contact a line that crossed the western part of the regions of Asturias and León from northwest to southeast (see Fig. 2). This area of contact coincides with the ranges of cline centre estimates obtained for each marker with the software Analyse. Thus the contact zone between the two lineages seems to be located in an area that encompasses the westernmost part of the Cantabrian Mountains and its surroundings.

There are other mammal species whose limits meet in this region. The three hare species present in the Iberian Peninsula (*Lepus granatensis*, *Lepus castroviejoi* and *Lepus europaeus*) have their range limits in this area and are known to hybridise (Melo-Ferreira *et al.* 2012). The area of contact between the two Iberian *Sorex* species distribution ranges has been documented further to the west (within Galicia), but there is no information available on the patterns of genetic variability of either species (*S. granarius* and *S. coronatus*) within this area. Then again, in a study of the Iberian populations of fire salamander (*Salamandra salamandra*), contact zones between the subspecies present in northern Iberia were described in that area (García-París *et al.* 2003).

Although the Geneland analysis of sequence data was based only on coded alleles, incurring a considerable loss of information, it still provided meaningful results, as already noted it should by Guillot *et al.* (2012). In this analysis of *Microtus agrestis* in Iberia it allowed for the visualisation of the range of the two highly divergent lineages and of the possible area where they currently make contact.

Asymmetric introgression

A hybrid zone between the Portuguese and Southern lineages was identified in northwestern Spain. Although the estimates of the hybrid zone width are high (approximately 50 – 200 km), this is likely to be inflated by the area of introgression, which largely involves Southern lineage alleles within the Portuguese lineage area in northwestern Spain, but does occur in a more limited way further to the east with Portuguese alleles in the Southern lineage area. Most individuals within the area of introgression are classified as belonging to the Portuguese lineage (to the west) or Southern lineage (to the east), based on all loci scored. The number of individuals within this area that showed mismatches for the assignment between the different loci (considered as ‘hybrid’ individuals) was only 12 to 18% (individuals with complete genotype or all individuals) and among those, considering only the diploid loci, no F1s were detected. When considering these data, and as most of the samples came from barn owl pellets, we must take into consideration barn owl movements. Though on average the hunting area of this bird of prey is about 1 Km radius from the nesting site, in areas of low productivity of small mammals it might range several kilometres. This could lead to a situation where we would be observing a bias towards higher levels of admixture. However, both the low amount of hybrids and the fact that no localities were found with ‘pure’ individuals of both lineages indicates that barn owl movement is not influencing the observed admixture results.

The likely explanation for the wide area of introgression is movement of the hybrid zone. Probably the small amount of introgression of Portuguese lineage alleles in the Southern lineage relates to an earlier movement in favour of the Southern lineage while the more substantial trail of Southern lineage alleles within the Portuguese lineage reflects more recent movement in favour of the Portuguese lineage. Even though there is very limited evidence for possible expansion events of the Portuguese lineage, from the previous demographic analysis performed by Paupério *et al.* (2012), that work only included one sample from the area of introgression, which indicates that more information is needed to confirm this possible demographic scenario. Similar trails of alleles attributed to moving hybrid zones have been obtained, for example, with the Australian grasshopper *Caledia* (Shaw *et al.* 1979; Arnold *et al.* 1987; Buggs 2007). In the *Caledia* case the introgression is believed to be due to climatic shift, and the two taxa hybridising show low hybrid fitness. The same is likely to be true in the field vole: that the introgression relates to range changes associated with climate changes at the end of the last glaciation and it is likely that the Southern and Portuguese lineages are very largely reproductively isolated from each other. If our supposition that the hybrid zone moved because of climate change is correct, then that indicates that the two lineages are differentiated enough to have different climatic optima. Further indications of strong differentiation and hybrid unfitness come from consideration of the different loci studied.

Both the cline and the individual based analysis displayed clear signs of differential introgression between the different loci. While the Y chromosome marker showed no signs of introgression (no males with mismatch information for the different loci were detected), the X chromosome and mitochondrial DNA showed a small degree of gene flow and BRCA1 was the most introgressed gene, with the widest cline and a higher proportion of mismatches with other loci.

The lack of male mediated gene flow fits Haldane's rule of sterility or inviability of the heterogametic sex in animals with sex-determined chromosomes (Haldane 1922). That there may be other issues related to male hybrids is suggested by the significantly lower proportion of males within the area of occurrence of hybrids (~30%) than on the rest of the distribution area of both lineages (~50%), but the basis of this is unclear and it requires further investigation.

Chromosomal rearrangements may be important in speciation, since they can contribute to reproductive isolation between populations (Searle 1998). These rearrangements can in some cases disrupt hybrid meiosis so much that the hybrids are sterile. In fact, there are cases described in mammals where chromosomal rearrangements may lead to hybrid male sterility (*e.g.* Forejt 1996; Yannic *et al.* 2008).

The field vole has an extraordinary karyotype with giant sex chromosomes but only minor variation has been described. Yet this variation comprises small differences in the structure of the Y chromosome within the northern lineage (Fredga and Jaarola 1997) but also in the Portuguese lineage (Giménez *et al.* 2012). Although these differences are small, they should not be neglected as possibly influencing hybrid fitness. In fact, an earlier study on the hybrid zone in the field vole between the two Y chromosome types in the Northern lineage also detected differential gene introgression between the mtDNA and the Y markers, which were related to possible incompatibilities between maternal and paternal genomes (Jaarola *et al.* 1997). Although further investigations should be performed, including experimental crosses and a detailed study within the contact zone that includes karyotyping of the individuals, it is possible that these small differences in sex chromosome structure could be impeding Y introgression. Either by chromosome variability or by genetic incompatibilities, the Y chromosome has proven to show reduced introgression in hybrid zone studies of several mammals (house mice, Tucker *et al.* 1992; shrews, Balloux *et al.* 2000; European rabbits, Geraldes *et al.* 2008), suggesting that the Y chromosome may have an important role in speciation (Geraldes *et al.* 2008).

Along with the lack of gene flow in the Y chromosome we also detected lower introgression in the X chromosome and mitochondrial DNA when compared with the introgression at the autosomal loci. Several studies have already demonstrated that loci located on different sections of the X chromosome show differential patterns of introgression between divergent lineages (house mice, Payseur *et al* 2004, Payseur *et al.* 2005; European rabbit, Geraldes *et al.* 2006). In the present study we have only examined one intron, DBX5 that shows reduced introgression. Thus, in order to clearly assess the degree of introgression on the level of the X chromosome, further data would be required through the analysis of more loci located on this chromosome.

The unusual characteristics of the contact between the Southern and Portuguese lineages of field vole may also be seen in other taxa in northwestern Spain, if climate is driving those characteristics. Indeed, a combination of range shifts due to climatic fluctuations together with biased genetic admixture across moving contact zones has also been proposed as potential mechanisms responsible for the characteristics of a hybrid zone between subspecies of *Salamandra salamandra,* located in the same area (northwestern Iberia, García-Paris *et al.* 2003).

In the area of hybrids between the Portuguese and Southern lineages of field vole we also detected a high frequency of rare alleles for the BRCA1 and cyt*b* gene. This is a common situation in hybrid zones, and according to Barton and Hewitt (1985) it can be caused by a higher mutation rate, by intragenic recombination or by relaxed

selection. When BRCA1 is analysed with IMgc (Woener *et al.* 2007), a software that reduces the alignments to the largest nonrecombining blocks, one haplotype is removed. This unique haplotype (found only in a heterozygote) is located within the area of introgression and contains one mutation that produces a stop codon. There is one other haplotype within the hybrid area that also contains a stop codon, but this one is not identified as recombinant. This excess of rare alleles is an added support for the genomic differentiation between the Southern and Portuguese lineages.

The *Microtus agrestis* species complex has already proven to be a good candidate system to investigate both the speciation process and the nature of species boundaries. The high divergences detected across the different genomic regions that place the three lineages at the borderline of becoming species, and the recentness of the diversification makes it extremely important to assess the degree of introgression between them. Both this study and Beysard *et al.* (2011) found evidence for the existence of barriers to gene flow in the studied hybrid zones. Also, hybridization detected was low in both studies and there seems to be a strong selection against hybrid males. One major difference between the Southern/Northern hybrid zone and the Portuguese/Southern hybrid zone is the differential introgression of the loci. While in northwestern Iberia we observe a scenario of asymmetric and wider introgression of the nuclear locus, Beysard *et al.* (2011) describes a situation with higher introgression of the mitochondrial marker. However these differences are possibly explained by the different demographic fluctuations that shaped the contact between these three lineages. Regardless of these differences, all the evidence collected in both studies suggest that the speciation process between the three lineages is already advanced, which supports the view that *Microtus agrestis* is best viewed as a complex of three separate species or near-species. Obviously, more work is desirable, including genomic, morphometric and also karyotypic approaches to better characterise these secondary contact areas detected and their temporal dynamics and to determine more clearly the extent of the reproductive isolation between these distinct evolutionary lineages.

ACKNOWLEDGEMENTS

We thank Alfonso Ceña, Antoni Arrizabalaga and the Museum of Natural Science of Granollers, António Mira, Augusto de Castro, Enrique Castiéns, Fernando Jubete, Hélia Gonçalves, Jacinto Róman, Maria José Lopez-Fuster and the University of Barcelona, Carlos Fonseca, Carlos Nores, for providing some of the samples used in this study. We are grateful to Alexandra Valongueiro and Ana Silva for help in owl

pellet material identification. This work was supported by Programa Operacional Potencial Humano – Quadro de Referência Estratégico Nacional funds from the European Social funds and Portuguese Ministério da Educação e Ciência (FCT, SFRH/BD/27369/2006 PhD grant to J. Paupério).

REFERENCES

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 B,* **363**, 2831-2839.

Andries AM, Gulink H, Herremans M (1994) Spatial model of the barn owl *Tyto alba* habitat using landscape characteristics derived from SPOT data. *Ecography*, **17**: 278-287.

Arnold ML, Shaw DD, Contreras N (1987) Ribosomal RNA-encoding DNA introgression across a narrow hybrid zone between two subspecies of grasshopper. *Proceedings of the National Academy of Sciences, USA*, **84**, 3946-3950.

Aulagnier S, Hutterer R, Amori G, Kryštufek B (2008). *Sorex coronatus*. In IUCN 2012. *IUCN Red List of Threatened Species*. Version 2012.1. <www.iucnredlist.org>.

Balloux F, Brünner H, Lugon-Moulin N, Hausser J, Godet J (2000) Microsatelites can be misleading: an empirical and simulation study. *Evolution*, **54**, 1414-1422.

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

Barbosa S, Paupério J, Searle JB, Alves PC (2012) Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to non-invasive sampling. *Molecular Ecology Resources*. DOI: 10.1111/1755-0998.12024.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113-148.

Barton NH, Baird SJE (1998) *Analyse – an application for analysing hybrid zones*. Edinburgh.

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

Berthier P, Excoffier L, Ruedi M (2006) Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society B,* **273**, 3101-3109.

Beysard M, Perrin N, Jaarola M, Heckel G, Vogel P (2012) Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (*Microtus agrestis*). *Journal of Evolutionary Biology*, **25**, 400-408.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**, 148-155.

Boratyński Z, Alves PC, Berto S, Koskela E, Mappes T, Melo-Ferreira J (2011) Introgression of mitochondrial DNA among *Myodes* voles: consequences for energetics? *BMC Evolutionary Biology*, **11**, 355-367.

Bugs RJA (2007) Empirical study of hybrid zone movement. *Heredity*, **99**, 301-312.

Cassens I, Mardulyn P, Milinkovitch MC (2005) Evaluating intraspecific "network" construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? *Systematic Biology*, **54**, 363–372.

Dod B, Jermiin LS, Boursot P, Chapman VH, Nielsen JT, Bonhomme F (1993) Counterselection in sex chromosomes in the *Mus musculus* European hybrid zone. *Journal of Evolutionary Biology*, **6**, 529-546.

Flot JF, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes*, **6**, 627-638.

Forejt J (1996) Hybrid sterility in the mouse. *Trends in Genetics*, **12**, 412-417.

Fredga K, Jaarola M (1997) The origin and distribution of the Lund Y chromosome in *Microtus agrestis* (Rodentia, Mammalia). *Hereditas*, **126**, 25-34.

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

García-París M, Alcobendas M, Buckley D, Wake DB (2003) Dispersal of viviparity across contact zones in Iberian populations of fire salamanders (*Salamandra salamandra*) inferred from discordance of genetic and morphological traits. *Evolution*, **57**, 129-143.

Garrick RC, Sunnucks P, Rodney JD (2010) Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evolutionary Biology*, **10**, 118-135.

Geraldes A, Ferrand N, Nachman MW (2006) Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, **173**, 919-933.

Geraldes A, Carneiro M, Delibes-Mateos M, Villafuerte R, Nachman W, Ferrand N (2008) Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. *Molecular Ecology*, **17**, 4489-4499.

Giménez M, Paupério J, Alves PC, Searle JB (2012) Giant sex chromosomes retained within the Portuguese lineage of the field vole *Microtus agrestis*. *Acta Theriologica*, **57**: 377-382.

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

Gonçalves H (2006) *Inventariação e caracterização dos micromamíferos no Parque Natural do Alvão.* Degree in Applied Ecology. University of Trás-os-Montes e Alto Douro, Vila Real.

Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261-1280.

Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712-715.

Guillot G, Renaud S, Ledevin R, Michaux J, Claude J (2012) A unifying model for the analysis of phenotypic, genetic and geographic data. *Systematic Biology*. DOI: 10.1093/sysbio/sys038.

Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, **12**, 101-109.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium Series*, **41**, 95-98.

Harrison RG (1993) *Hybrid Zones and the Evolutionary Process*. Oxford University Press, New York.

Hellborg L, Gündüz I, Jaarola M (2005) Analysis of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology*, **14**, 2025–2031.

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

Kryštufek B, Vohralík V, Zima J, Zagorodnyuk I (2008) *Microtus agrestis*. In IUCN 2012. *IUCN Red list of threatened species*. Version 2012.1 <www.iucnredlist.org>.

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

Jaarola M, Telgeström H, Fredga K (1997) A contact zone with noncoincident clines for sex-specific markers in the field vole (*Microtus agrestis*). *Evolution*, **51**, 241-249.

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.

Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farelo L, Alves PC (2012) Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulation. *Systematic Biology*, **61**, 367–381.

Mira A, Ascenção F, Alcobia S (2003). *Distribuição das espécies de roedores e insectívoros*. Final Report. Department of Biologia, Évora University, Évora.

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**, 170-189.

Palomo LJ, Amori G, Hutterer R (2008) *Sorex granarius*. In IUCN 2012*. IUCN Red List of Threatened Species.* Version 2012.1. <www.iucnredlist.org>.

Palomo J, Gisbert J (2002) *Atlas de los Mamíferos Terrestres de España.* Sociedad Espanhola para la Conservación y Estudio de los Mamíferos. Madrid.

Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB (2012) Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Europe. *Molecular Ecology*. DOI: 10.1111/mec.12024.

Payseur BA, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, **58**, 2064-2078.

Payseur BA, Nachman MW (2005) The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. *Biological Journal of the Linnean Society*, **84**, 523-534.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386-393.

Piálek J, Hauffe HC, Searle JB (2005) Chromosomal variation in the house mouse. *Biological Journal of the Linnean Society*, **84**, 535-563.

Phillips BL, Baird SJE, Moritz C (2004) When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution*, **58**, 1536-1548.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

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

Sá-Sousa P, Marquez R, Pérez-Mellado V, Martínez-Solano I (2009). *Lacerta schreiberi*. In IUCN 2012. *IUCN Red List of Threatened Species*. Version 2012.1. <www.iucnredlist.org>.

Sá-Sousa P, Pérez-Mellado V, Martínez-Solano I (2009). *Podarcis bocagei*. In IUCN 2012. *IUCN Red List of Threatened Species*. Version 2012.1. <www.iucnredlist.org>.

Searle JB (1998) Speciation, chromosomes and genomes. *Genome Research*, **8**, 1-3.

SGHN (1995). *Atlas de Vertebrados de Galicia, tomo I*. Consello da Cultura Gallega. Santiago de Compostela, Galiza, Spain.

Shaw DD, Wilkinson P, Moran C (1979) A comparison of chromosomal and allozymal variation across a narrow hybrid zone in the grasshopper *Caledia captiva*, *Chromosoma*, **75**, 333-351.

Stephens M, Smith NJ, Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978-989.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods of haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162-1169.

Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution*, **40**, 1141-1159.

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

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

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.*

Tucker PK, Sage RD, Warner J, Wilson AC, Eicher EM (1992). Abrupt cline for sex-chromosomes in a hybrid zone between two species of mice. *Evolution, **46**, 1146-1163.*

Turelli M, Orr HA (1995) The dominance theory of Haldane's rule. *Genetics*, **140**, 389-402.

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.

Woener AE, Cox MP, Hammer MF (2007) Recombination filter genomic databases by information maximization. *Bioinformatics*, **23**, 1851-1853.

Woolley SM, Posada D, Crandall KA (2008) A comparison of phylogenetic network methods using computer simulation. *PLoS ONE*, **3**, e1913.

Yannic G, Basset P, Hausser J (2008) A hybrid zone with coincident clines for autosomal and sex-specific markers in the *Sorex araneus* group. *Journal of Evolutionary Biology*, **21**, 658-667.

Appendix I

Details of the localities sampled for the field vole in this study. The numbers of samples in which the extraction was tested, as well as the number of samples amplified for each locus in each locality is also shown.

Locality	Lat*	Long*	Type†	Tested‡	Amp§	Cytb	DBY7	BRCA1	DBX5
Alcobaça	39.55	-8.98	bone	5	5	4	3	4	5
Louriçal	40.00	-8.74	bone	5	5	5	2	5	5
Barcouço	40.30	-8.47	bone	5	5	5	1	5	5
Ílhavo	40.64	-8.73	bone	5	5	5	2	5	5
São Pedro de Alva	40.30	-8.18	bone	5	5	5	1	5	5
Parada de Gonta	40.57	-8.01	bone	5	5	5	2	4	5
Gouveia	40.49	-7.60	bone	5	5	5	2	5	5
Santa Cruz de Trapa	40.77	-8.15	bone	5	5	5	1	1	4
Rua	40.95	-7.58	bone	5	5	5	2	5	5
Lugar de Anta	41.27	-7.64	bone	5	5	5	3	5	5
Aveçãozinho	41.30	-7.88	bone	5	5	5	1	5	5
Granja	41.69	-7.65	bone	5	5	5	3	5	5
Paredes de Coura	41.91	-8.57	bone	5	5	5	1	2	5
Outeiro, Serra do Gêres	41.79	-7.95	bone	1	1	1	0	1	1
Campos	41.98	-8.70	bone	1	1	1	0	1	1
Montalegre	41.86	-7.67	bone	1	1	1	1	1	1
Freixo, Celanova	42.19	-7.96	bone	5	1	1	0	0	0
Castro Caldelas	42.37	-7.42	bone	5	3	2	0	2	1
Lalín	42.66	-8.11	bone	5	4	4	0	3	4
Santiago de Compostela	42.88	-8.54	bone	5	3	3	0	1	1
Camariñas	43.13	-9.19	bone	5	2	2	0	1	2
Torre de Mens, Malpica	43.30	-8.86	bone	5	3	3	0	0	1
Sobrado de Monxes	43.04	-8.03	bone	5	5	5	0	5	2
Oza dos Rios	43.22	-8.19	bone	1	1	1	1	1	1
Pontedeumes	43.40	-8.15	bone	5	5	5	0	2	2
Outeiro de Rei	43.10	-7.62	bone	5	5	5	1	3	4
O Incio	42.66	-7.40	bone	1	1	1	0	1	1
Bolaño, Castroverde	43.04	-7.33	bone	5	5	5	3	5	5
Fontsagrada, San Cristobal	43.12	-7.07	bone	5	4	4	0	1	3
Ribadeon	43.54	-7.04	bone	5	4	3	0	4	2
Donis	42.86	-6.90	bone	5	5	5	0	4	3
Pelorde	43.28	-6.84	bone	1	1	1	0	1	1
S. Vicente de Nimbra			bone	5	0	0	0	0	0
Corullón	42.58	-6.82	bone	1	1	1	0	1	1
Grandais	41.82	-6.82	bone	5	5	5	3	5	5
Villafranca del Bierzo	42.61	-6.81	bone	1	1	1	0	0	0
Tormaleo	42.95	-6.74	bone	5	5	5	1	3	3
Vega de Espinareda	42.72	-6.66	bone	1	1	1	1	1	1
Ambasaguas	42.27	-6.56	bone	1	1	1	0	1	1
Constantim	41.62	-6.28	bone	1	1	1	0	1	1
Boal	43.43	-6.82	bone	5	5	5	1	5	5
Gedrez	43.02	-6.61	bone	1	1	1	0	1	1
Morla	42.24	-6.26	bone	1	1	1	0	1	0
La Caridad	43.55	-6.83	bone	5	4	4	0	3	3
Anleo	43.51	-6.69	bone	1	1	1	0	1	0
Pola de Allande	43.27	-6.61	bone	1	1	1	1	1	1
Palacios del Sil	42.88	-6.43	bone	1	1	1	0	0	1
Manzanal del Puerto	42.59	-6.22	bone	1	1	1	0	1	0
Fresno de la Valduerna	42.33	-5.97	bone	5	5	5	1	5	4
Obona	43.34	-6.48	bone	1	1	1	0	0	1
Ribón	43.55	-6.36	bone	1	1	1	0	1	1

Locality	Lat*	Lon*	Type†	Tested‡	Amp§	Cytb	DBY7	BRCA1	DBX5
Valencia de D. Juan	-5.52	42.29	bone	1	1	1	0	1	0
Camposagrado	-5.73	42.73	bone	1	1	1	0	0	1
Vallín	-5.96	43.16	bone	5	4	3	0	3	3
Barrillos de las Arrimadas	-5.28	42.81	bone	1	1	1	0	1	1
Tanes	-5.39	43.21	bone	5	5	5	2	4	5
Puebla de Lillo	-5.27	43.01	bone	5	5	5	0	0	3
Villaviciosa	-5.42	43.49	bone	5	5	5	1	2	3
Guardo	-4.86	42.79	bone	1	1	1	1	1	1
Muñorrodero	-4.49	43.37	bone	5	1	1	0	0	0
Mogrovejo			bone	1	0	0	0	0	0
Pino del Río	-4.82	42.64	bone	1	1	1	0	1	1
Los Llazos	-4.48	42.98	bone	1	1	1	1	1	1
Villorquite del Páramo	-4.73	42.56	bone	1	1	1	0	1	1
Gañinas	-4.73	42.48	bone	5	5	5	2	5	5
San Llorente del Páramo	-4.77	42.39	bone	1	1	1	1	1	1
Villaturde	-4.68	42.37	bone	1	1	1	0	1	1
Báscones de Ojeda	-4.54	42.67	bone	1	1	1	0	1	1
Sotobañado y Priorato	-4.45	42.59	bone	1	1	1	1	1	1
Mudá	-4.40	42.87	bone	5	2	2	0	2	2
Las Cabañas de Castilla	-4.37	42.35	bone	1	1	1	0	1	1
Liencres	-3.96	43.45	bone	1	1	1	0	1	0
Valderredible	-4.14	42.80	bone	5	5	5	1	5	5
Entrambasmestas	-3.90	43.17	bone	5	5	5	1	4	5
Cilleruelo de Bezana	-3.85	42.97	bone	5	4	4	0	3	4
San Martín de Humada	-4.03	42.66	bone	5	5	5	1	5	4
San Martín de Elines	-3.87	42.83	bone	5	4	4	0	4	3
Hornillalatorre	-3.60	43.05	bone	5	5	5	0	4	2
Noceco	-3.48	43.09	bone	5	5	5	2	3	3
Ungo	-3.26	43.14	bone	5	5	5	0	5	4
Encima Angulo	-3.18	43.04	bone	5	5	5	3	5	5
Bentretea	-3.46	42.73	bone	5	5	5	0	5	4
Villalonquéjar	-3.77	42.36	bone	1	1	1	1	1	1
Herrán	-3.23	42.83	bone	5	4	4	2	4	4
Aostri de Losa	-3.10	42.94	bone	5	5	5	2	5	5
Torrecitores	-3.85	41.99	bone	5	5	5	2	5	4
Quintanalara	-3.53	42.18	bone	5	5	5	2	5	5
Contreras	-3.41	42.02	bone	5	5	5	1	5	5
Castrillo de la Reina	-3.23	41.99	bone	5	5	5	1	4	3
Barbadillo de Herreros	-3.18	42.15	bone	5	5	4	1	3	4
Fresneda de la Sierra	-3.13	42.32	bone	5	5	5	3	5	5
Zorraquín	-3.04	42.33	bone	5	4	4	1	3	3
Covaleda	-2.88	41.94	bone	5	5	4	1	3	3
Baños de Ebro	-2.68	42.52	bone	3	3	3	3	3	3
Medrano	-2.55	42.38	bone	1	1	1	0	0	0
La Villa de Ócon	-2.25	42.30	bone	2	2	2	1	1	1
Markina-Xemein	-2.51	43.26	bone	5	5	5	5	5	5
Iturmendi	-2.10	42.89	bone	5	5	5	4	5	5
Fuenterrabia	-1.79	43.36	skin	1	1	1	1	1	1
Vilafranca de Ordivia			skin	1	0	0	0	0	0
Zaldivia			skin	2	0	0	0	0	0
Arbide			skin	1	0	0	0	0	0
Astigarraga			skin	2	0	0	0	0	0
Arrarás	-1.80	43.03	bone	5	5	4	4	3	5
Arraratz			skin	1	0	0	0	0	0
Lesaca	-1.70	43.23	bone	5	5	4	3	2	5
Bertiz	-1.60	43.10	skin	2	1	0	0	1	1

Locality	Lat*	Long*	Type†	Tested‡	Amp§	Cytb	DBY7	BRCA1	DBX5
Equisoain, Monreal	-1.50	42.70	bone	5	5	5	3	4	5
Ureta	-1.40	42.95	bone	5	5	5	1	4	5
Usón	-1.20	42.70	bone	3	3	3	0	0	3
Quinto Real			skin	1	0	0	0	0	0
Lerida			skin	2	0	0	0	0	0
Sierra del Cadí	1.61	42.29	tissue	1	1	1	0	0	0
La Molina	1.96	42.34	skin	2	1	0	1	1	1
Caralps	2.16	42.34	skin	8	8	5	3	5	5
Tona	2.23	41.85	bone	1	1	1	0	1	1
Centelles	2.22	41.79	bone	1	1	1	1	1	1
El Brull	2.30	41.81	bone	4	4	4	3	3	3
La castanyera	2.34	41.78	bone	1	1	1	0	1	1
Arbúcies	2.51	41.81	bone	1	1	1	1	1	1
L'Esparra	2.65	41.82	bone	1	1	1	0	1	0

* Latitude and longitude presented in decimal degrees; † - Type of sampled used: bone, bones collected from barn owl pellets; skin and tissue, degraded tissue sample; ‡ - number of samples tested for each locality; § - number of samples which had successful amplification for at least one locus.

Appendix II

Details of the samples used in the Geneland analysis. Information on the codes attributed to of each haplotype (for each locus sequenced) is given. Details on the locality of each samples are also shown.

Sample	Locality	Lat*	Long*	Cyt*b*	DBY7	DBX5a	DBX5b	BRCA1a	BRCA1b
AB1P	Alcobaça	39.55	-8.98	021	021	022	NA	021	021
AB2P	Alcobaça	39.55	-8.98	021	NA	022	022	021	021
AB3P	Alcobaça	39.55	-8.98	021	021	022	NA	021	021
AB4P	Alcobaça	39.55	-8.98	021	021	021	NA	021	021
LO1P	Louriçal	40.00	-8.74	021	021	021	NA	021	021
LO2P	Louriçal	40.00	-8.74	021	021	021	NA	021	021
LO3P	Louriçal	40.00	-8.74	021	NA	021	022	021	021
LO4P	Louriçal	40.00	-8.74	021	NA	021	022	021	021
LO5P	Louriçal	40.00	-8.74	021	NA	022	022	021	021
B1P	Barcouço	40.30	-8.47	021	NA	022	022	021	021
B2P	Barcouço	40.30	-8.47	214	NA	022	022	021	021
B3P	Barcouço	40.30	-8.47	021	NA	022	022	021	021
B4P	Barcouço	40.30	-8.47	021	NA	022	022	021	021
B5P	Barcouço	40.30	-8.47	021	021	022	NA	021	021
I1P	Ílhavo	40.64	-8.73	021	NA	022	022	021	021
I2P	Ílhavo	40.64	-8.73	021	NA	022	022	021	021
I4P	Ílhavo	40.64	-8.73	021	NA	022	022	021	021
I5P	Ílhavo	40.64	-8.73	021	021	022	NA	021	021
I6P	Ílhavo	40.64	-8.73	021	021	022	NA	021	021
SPA10	São Pedro de Alva	40.30	-8.18	021	NA	022	022	021	021
SPA11	São Pedro de Alva	40.30	-8.18	021	NA	021	022	021	021
SPA2	São Pedro de Alva	40.30	-8.18	021	NA	022	022	021	021
SPA7	São Pedro de Alva	40.30	-8.18	021	021	021	NA	021	021
SPA9	São Pedro de Alva	40.30	-8.18	214	NA	021	022	021	021
PG11P	Parada de Gonta	40.57	-8.01	021	NA	022	022	021	021
PG17P	Parada de Gonta	40.57	-8.01	026	021	022	NA	021	021
PG1P	Parada de Gonta	40.57	-8.01	027	021	021	NA	021	021
PG2P	Parada de Gonta	40.57	-8.01	210	NA	022	022	021	021
PG9P	Parada de Gonta	40.57	-8.01	021	NA	022	022	NA	NA
GO18	Gouveia	40.49	-7.60	021	NA	022	023	021	021
GO26	Gouveia	40.49	-7.60	021	NA	021	021	021	021
GO33	Gouveia	40.49	-7.60	021	021	022	NA	021	021
GO5	Gouveia	40.49	-7.60	021	021	022	NA	021	021
GO9	Gouveia	40.49	-7.60	021	NA	022	022	021	021
SCT2	Santa Cruz da Trapa	40.77	-8.15	021	NA	022	022	NA	NA
SCT3	Santa Cruz da Trapa	40.77	-8.15	021	NA	022	022	NA	NA
SCT4	Santa Cruz da Trapa	40.77	-8.15	021	NA	021	021	021	021
SCT5	Santa Cruz da Trapa	40.77	-8.15	021	021	021	NA	NA	NA
QT1	Videmonte	40.49	-7.42	024	NA	022	022	021	021
QT2	Videmonte	40.49	-7.42	021	021	021	NA	021	021
QT3	Videmonte	40.49	-7.42	021	NA	021	021	021	021
QT4	Videmonte	40.49	-7.42	024	NA	022	023	021	021
QT5	Videmonte	40.49	-7.42	024	NA	022	022	021	021
SM.865	Serra da Malcata	40.31	-6.99	026	NA	021	021	021	021
SM.866	Serra da Malcata	40.31	-6.99	021	021	021	NA	021	021
SM.867	Serra da Malcata	40.31	-6.99	026	NA	021	022	021	021
SM.868	Serra da Malcata	40.31	-6.99	026	021	022	NA	021	021
SM.869	Serra da Malcata	40.31	-6.99	025	021	021	NA	021	021
SM.870	Serra da Malcata	40.31	-6.99	026	021	021	NA	021	021
SM.871	Serra da Malcata	40.31	-6.99	021	021	022	NA	021	021
QM1	Guarda	40.57	-7.23	021	021	021	NA	021	021
QM2	Guarda	40.57	-7.23	024	021	022	NA	021	021

Sample	Locality	Lat*	Long*	Cyt*b*	DBY7	DBX5a	DBX5b	BRCA1a	BRCA1b
QM3	Guarda	40.57	-7.23	021	021	022	NA	021	021
QM4	Guarda	40.57	-7.23	024	NA	022	022	021	021
QM6	Guarda	40.57	-7.23	021	NA	022	022	021	021
QM7	Guarda	40.57	-7.23	025	021	022	NA	021	021
QM9	Guarda	40.57	-7.23	026	021	022	NA	021	021
QM11	Guarda	40.57	-7.23	021	021	021	NA	021	021
CE1	Serra da Arada	40.83	-8.15	021	NA	022	022	021	021
CE2	Serra da Arada	40.83	-8.15	021	021	021	NA	021	021
CE3	Serra da Arada	40.83	-8.15	021	021	021	NA	021	021
R1P	Rua	40.95	-7.58	021	021	021	NA	021	021
R2P	Rua	40.95	-7.58	021	NA	021	022	021	021
R3P	Rua	40.95	-7.58	021	NA	021	022	021	021
R4P	Rua	40.95	-7.58	021	NA	021	022	021	021
R5P	Rua	40.95	-7.58	021	021	022	NA	021	021
LA1P	Lugar de Anta	41.27	-7.64	021	021	021	NA	021	021
LA2P	Lugar de Anta	41.27	-7.64	021	NA	021	021	021	021
LA3P	Lugar de Anta	41.27	-7.64	021	021	023	NA	021	021
LA4P	Lugar de Anta	41.27	-7.64	022	021	022	NA	021	021
LA5P	Lugar de Anta	41.27	-7.64	021	NA	022	022	021	021
A14P	Aveçãozinho	41.30	-7.88	021	NA	022	022	021	021
A17P	Aveçãozinho	41.30	-7.88	021	021	021	NA	021	021
A1P	Aveçãozinho	41.30	-7.88	021	NA	022	022	021	021
A20P	Aveçãozinho	41.30	-7.88	021	NA	022	022	021	021
A23P	Aveçãozinho	41.30	-7.88	021	NA	022	022	021	021
LO1	Serra do Alvão	41.37	-7.79	021	NA	021	023	021	021
LO2	Serra do Alvão	41.37	-7.79	023	021	022	NA	021	021
LO3	Serra do Alvão	41.37	-7.79	021	NA	021	022	021	021
LO4	Serra do Alvão	41.37	-7.79	024	021	021	NA	021	021
LO6	Serra do Alvão	41.37	-7.79	024	021	021	NA	021	021
LO7	Serra do Alvão	41.37	-7.79	021	021	021	NA	021	021
AL3	Serra do Alvão	41.37	-7.79	021	021	022	NA	021	021
AL4	Serra do Alvão	41.37	-7.79	021	021	023	NA	021	021
AL5	Serra do Alvão	41.37	-7.79	021	NA	022	023	021	021
AL6	Serra do Alvão	41.37	-7.79	024	021	021	NA	021	021
GV2	Gouvães	41.48	-7.73	021	NA	022	023	021	021
F1	Celorico de Bastos	41.43	-8.09	028	021	022	NA	021	021
F2	Celorico de Bastos	41.43	-8.09	021	NA	022	022	021	021
F3	Celorico de Bastos	41.43	-8.09	028	021	022	NA	021	021
F4	Celorico de Bastos	41.43	-8.09	021	021	022	NA	021	021
F5	Celorico de Bastos	41.43	-8.09	021	021	022	NA	021	021
F6	Celorico de Bastos	41.43	-8.09	021	NA	022	022	021	021
LA1	Vila Pouca de Aguiar	41.55	-7.53	021	021	021	NA	021	021
LA5	Vila Pouca de Aguiar	41.55	-7.53	021	021	022	NA	021	021
LA6	Vila Pouca de Aguiar	41.55	-7.53	021	NA	022	022	021	021
LA10	Vila Pouca de Aguiar	41.55	-7.53	021	021	022	NA	021	021
V1	Ponte de Lima	41.65	-8.55	027	NA	022	024	021	021
V2	Ponte de Lima	41.65	-8.55	027	021	022	NA	021	021
V3	Ponte de Lima	41.65	-8.55	027	021	022	NA	021	021
V4	Ponte de Lima	41.65	-8.55	027	NA	022	022	021	021
V5	Ponte de Lima	41.65	-8.55	027	021	023	NA	021	021
BE2	Bertiandos	41.77	-8.63	027	NA	021	022	021	021
BE6	Bertiandos	41.77	-8.63	027	NA	022	023	021	021
BE8	Bertiandos	41.77	-8.63	028	021	022	NA	021	021
BE9	Bertiandos	41.77	-8.63	021	021	022	NA	021	021
BE13	Bertiandos	41.77	-8.63	027	021	022	NA	021	021
GJ10P	Granja	41.69	-7.65	021	NA	022	022	021	021
GJ11P	Granja	41.69	-7.65	218	021	022	NA	021	021

Sample	Locality	Lat*	Long*	Cyt*b*	DBY7	DBX5a	DBX5b	BRCA1a	BRCA1b
GJ6P	Granja	41.69	-7.65	021	021	022	NA	021	021
GJ8P	Granja	41.69	-7.65	021	021	022	NA	021	021
GJ9P	Granja	41.69	-7.65	021	NA	022	022	021	021
PC1P	Paredes de Coura	41.91	-8.57	023	021	021	NA	NA	NA
PC2P	Paredes de Coura	41.91	-8.57	021	NA	027	027	021	021
PC3P	Paredes de Coura	41.91	-8.57	021	NA	022	022	NA	NA
PC4P	Paredes de Coura	41.91	-8.57	023	NA	022	022	NA	NA
PC6P	Paredes de Coura	41.91	-8.57	021	NA	022	022	021	021
OU2P	Outeiro	41.79	-7.95	021	NA	027	027	021	021
CA2P	Campos	41.98	-8.70	021	NA	021	021	021	021
MON1	Montalegre	41.86	-7.67	021	021	022	NA	021	021
MC1	Macedo de Cavaleiros	41.60	-7.00	021	NA	022	023	021	021
MC2	Macedo de Cavaleiros	41.60	-7.00	021	021	023	NA	021	021
MC3	Macedo de Cavaleiros	41.60	-7.00	021	NA	022	023	021	021
MC6	Macedo de Cavaleiros	41.60	-7.00	021	NA	023	023	021	021
MC7	Macedo de Cavaleiros	41.60	-7.00	021	NA	023	023	021	021
MC8	Macedo de Cavaleiros	41.60	-7.00	021	021	023	NA	021	021
MC10	Macedo de Cavaleiros	41.60	-7.00	021	021	023	NA	021	021
MC12	Macedo de Cavaleiros	41.60	-7.00	021	NA	021	023	021	021
MC19	Macedo de Cavaleiros	41.60	-7.00	021	021	023	NA	021	021
MC20	Macedo de Cavaleiros	41.60	-7.00	021	NA	023	023	021	021
MC21	Macedo de Cavaleiros	41.60	-7.00	021	021	023	NA	021	021
MC22	Macedo de Cavaleiros	41.60	-7.00	021	021	023	NA	021	021
CC7	Castro Caldelas	42.37	-7.42	216	NA	022	023	021	021
LALC1	Lalín	42.66	-8.11	021	NA	022	022	111	111
LALC2	Lalín	42.66	-8.11	021	NA	022	022	NA	NA
LALC3	Lalín	42.66	-8.11	016	NA	022	022	118	118
LALC6	Lalín	42.66	-8.11	021	NA	022	022	011	011
SA5	Santiago de Compostela	42.88	-8.54	021	NA	022	022	021	021
CM1	Camariñas	43.13	-9.19	216	NA	022	022	NA	NA
CM2	Camariñas	43.13	-9.19	216	NA	022	022	021	021
TM2	Torre de Mens	43.30	-8.86	021	NA	021	011	NA	NA
SM1	Sobrado de Monxes	43.04	-8.03	229	NA	021	021	027	027
SM11	Sobrado de Monxes	43.04	-8.03	229	NA	NA	NA	021	021
SM2	Sobrado de Monxes	43.04	-8.03	229	NA	NA	NA	011	121
SM5	Sobrado de Monxes	43.04	-8.03	023	NA	NA	NA	011	011
SM8	Sobrado de Monxes	43.04	-8.03	023	NA	022	022	021	028
OR4	Oza dos Rios	43.22	-8.19	221	021	022	NA	021	021
PO10	Pontedeumes	43.40	-8.15	226	NA	022	022	021	021
PO2	Pontedeumes	43.40	-8.15	023	NA	022	022	021	021
OU10	Outeiro de Rei	43.10	-7.62	025	NA	021	021	021	021
OU11	Outeiro de Rei	43.10	-7.62	222	NA	021	021	021	011
OU12	Outeiro de Rei	43.10	-7.62	023	021	022	NA	NA	NA
OU7	Outeiro de Rei	43.10	-7.62	023	NA	021	021	026	026
IN2	O Íncio	42.66	-7.40	216	NA	021	022	021	024
BOL2	Bolaño	43.04	-7.33	025	021	022	NA	021	021
BOL3	Bolaño	43.04	-7.33	025	021	022	NA	021	021
BOL5	Bolaño	43.04	-7.33	025	021	022	NA	021	021
BOL6	Bolaño	43.04	-7.33	215	NA	022	022	021	023
BOL7	Bolaño	43.04	-7.33	025	NA	021	022	021	021
FS10	Fonsagrada	43.12	-7.07	216	NA	021	021	021	021
FS8	Fonsagrada	43.12	-7.07	147	NA	022	022	NA	NA
FS9	Fonsagrada	43.12	-7.07	023	NA	021	022	NA	NA
RB1	Ribadeon	43.54	-7.04	023	NA	021	021	021	021
RB2	Ribadeon	43.54	-7.04	221	NA	022	022	021	021
RB4	Ribadeon	43.54	-7.04	227	NA	NA	NA	011	011

Sample	Locality	Lat*	Long*	Cyt*b*	DBY7	DBX5a	DBX5b	BRCA1a	BRCA1b
DA1	Donis	42.86	-6.90	021	NA	022	022	021	116
DA10	Donis	42.86	-6.90	211	NA	022	022	021	021
DA2	Donis	42.86	-6.90	211	NA	NA	NA	011	117
DA6	Donis	42.86	-6.90	021	NA	023	023	021	011
ASPE1	Pelorde	43.28	-6.84	021	NA	021	022	021	021
LECO1	Corullón	42.58	-6.82	021	NA	021	021	021	119
GR1P	Grandais	41.82	-6.82	021	021	022	NA	021	021
GR2P	Grandais	41.82	-6.82	021	NA	023	023	021	021
GR3P	Grandais	41.82	-6.82	021	021	023	NA	021	021
GR4P	Grandais	41.82	-6.82	021	021	022	NA	021	021
GR5P	Grandais	41.82	-6.82	021	NA	022	022	021	021
ASTO2	Tormaleo	42.95	-6.74	021	NA	NA	NA	011	015
ASTO3	Tormaleo	42.95	-6.74	021	021	022	NA	021	021
ASTO4	Tormaleo	42.95	-6.74	021	NA	021	022	NA	NA
ASTO6	Tormaleo	42.95	-6.74	023	NA	022	022	021	021
LEVE2	Vega de Espinareda	42.72	-6.66	211	021	022	NA	021	021
LEAM3	Ambasaguas	42.27	-6.56	219	NA	023	023	021	021
C1P	Constantim	41.62	-6.28	021	NA	022	022	021	021
ASBO1	Boal	43.43	-6.82	210	NA	021	022	011	011
ASBO2	Boal	43.43	-6.82	211	NA	022	022	021	021
ASBO4	Boal	43.43	-6.82	211	NA	021	021	021	021
ASBO5	Boal	43.43	-6.82	211	021	021	NA	021	021
ASBO6	Boal	43.43	-6.82	211	NA	021	021	011	011
ASGE2	Gedrez	43.02	-6.61	021	NA	021	022	022	022
LEMO1	Morla	42.24	-6.26	021	NA	NA	NA	021	021
ASLC1	La Caridad	43.55	-6.83	212	NA	022	022	021	021
ASLC3	La Caridad	43.55	-6.83	211	NA	021	021	021	021
ASLC4	La Caridad	43.55	-6.83	211	NA	022	022	011	011
ASAN2	Anleo	43.51	-6.69	029	NA	NA	NA	011	013
ASPA1	Pola de Allande	43.27	-6.61	011	011	011	NA	011	011
LEPSIL1	Palacios de Sil	42.88	-6.43	211	NA	022	022	NA	NA
LEMP7	Manzanal del Puerto	42.59	-6.22	021	NA	NA	NA	021	021
LEFV3	Fresno de la Valduerna	42.33	-5.97	211	021	023	NA	021	025
LEFV4	Fresno de la Valduerna	42.33	-5.97	021	NA	022	022	021	021
LEFV5	Fresno de la Valduerna	42.33	-5.97	021	NA	021	021	021	021
LEFV7	Fresno de la Valduerna	42.33	-5.97	021	NA	025	025	021	021
LEFV8	Fresno de la Valduerna	42.33	-5.97	021	NA	NA	NA	011	120
ASOB2	Obona	43.34	-6.48	213	NA	022	022	NA	NA
ASRI2	Ribón	43.55	-6.36	211	NA	022	011	021	021
LEVJ1	Valencia de D Juan	42.29	-5.52	021	NA	NA	NA	011	011
LECS3	Camposagrado	42.73	-5.73	021	NA	023	023	NA	NA
ASVA1	Vallín	43.16	-5.96	211	NA	022	022	011	011
ASVA2	Vallín	43.16	-5.96	011	NA	011	011	011	011
ASVA4	Vallín	43.16	-5.96	NA	NA	011	011	011	011
LEBA2	Barrillos de las Arrimadas	42.81	-5.28	021	NA	011	011	011	011
ASTA1	Tanes	43.21	-5.39	011	011	011	NA	014	014
ASTA2	Tanes	43.21	-5.39	121	NA	011	011	NA	NA
ASTA3	Tanes	43.21	-5.39	122	011	011	NA	011	011
ASTA4	Tanes	43.21	-5.39	016	NA	011	011	011	011
ASTA5	Tanes	43.21	-5.39	121	NA	011	011	011	011
LEPLIL1	Puebla de Lillo	43.01	-5.27	013	NA	011	011	NA	NA
LEPLIL2	Puebla de Lillo	43.01	-5.27	013	NA	026	026	NA	NA
LEPLIL4	Puebla de Lillo	43.01	-5.27	149	NA	011	011	NA	NA
ASVI4	Villaviciosa	43.49	-5.42	125	NA	011	011	011	011
ASVI6	Villaviciosa	43.49	-5.42	011	011	011	NA	011	011
ASVI8	Villaviciosa	43.49	-5.42	126	NA	011	011	NA	NA

Sample	Locality	Lat*	Long*	Cyt*b*	DBY7	DBX5a	DBX5b	BRCA1a	BRCA1b
ES-37	Picos da Europa	43.11	-5.00	011	NA	011	011	011	011
ES-39	Picos da Europa	43.11	-5.00	011	NA	011	011	011	011
ES-40	Picos da Europa	43.11	-5.00	011	NA	011	011	011	011
ES-41	Picos da Europa	43.11	-5.00	011	NA	011	011	011	011
ES-42	Picos da Europa	43.11	-5.00	011	011	011	NA	011	011
ES-43	Picos da Europa	43.11	-5.00	011	011	011	NA	011	011
ES-44	Picos da Europa	43.11	-5.00	011	NA	011	011	011	011
ES-45	Picos da Europa	43.11	-5.00	011	011	011	NA	011	011
ES-50	Picos da Europa	43.11	-5.00	011	011	011	NA	011	011
ES-52	Picos da Europa	43.11	-5.00	011	NA	011	011	011	011
ES-53	Picos da Europa	43.11	-5.00	011	NA	011	011	011	011
ES-58	Picos da Europa	43.11	-5.00	115	NA	011	011	011	011
ES-60	Picos da Europa	43.11	-5.00	011	011	011	NA	011	011
PAL47	Guardo	42.79	-4.86	011	011	011	NA	011	011
PAL29	Pino del Río	42.64	-4.82	155	NA	011	011	011	011
PAL60	Los Llazos	42.98	-4.48	016	011	011	NA	011	018
PAL28	Villorquite del Páramo	42.56	-4.73	016	NA	011	011	011	011
PAL32	Gañinas	42.48	-4.73	155	011	011	NA	011	011
PAL33	Gañinas	42.48	-4.73	155	NA	011	011	011	011
PAL34	Gañinas	42.48	-4.73	011	011	011	NA	011	011
PAL35	Gañinas	42.48	-4.73	155	NA	011	011	011	011
PAL36	Gañinas	42.48	-4.73	016	NA	011	011	011	011
PAL48	San Llorente del Páramo	42.39	-4.77	011	011	011	NA	011	011
PAL49	Villaturde	42.37	-4.68	011	NA	011	011	011	011
PAL43	Báscones de Ojeda	42.67	-4.54	016	NA	011	011	011	011
PAL54	Sotobañado y Priorato	42.59	-4.45	016	011	011	NA	011	011
PAL26	Mudá	42.87	-4.40	155	NA	011	011	011	011
PAL3	Mudá	42.87	-4.40	112	NA	011	011	011	011
PAL45	Las Cabañas de Castilla	42.35	-4.37	016	NA	011	011	011	011
CALI3	Liencres	43.45	-3.96	016	NA	NA	NA	011	011
CAVD1	Valderredible	42.80	-4.14	145	NA	011	011	011	011
CAVD2	Valderredible	42.80	-4.14	016	011	011	NA	011	011
CAVD3	Valderredible	42.80	-4.14	016	NA	011	011	011	011
CAVD5	Valderredible	42.80	-4.14	146	NA	019	019	011	011
CAVD7	Valderredible	42.80	-4.14	112	NA	011	018	011	011
BUR360	Entrambasmestas	43.17	-3.90	016	NA	011	011	011	011
BUR361	Entrambasmestas	43.17	-3.90	016	NA	017	017	011	011
BUR362	Entrambasmestas	43.17	-3.90	016	NA	011	011	011	011
BUR363	Entrambasmestas	43.17	-3.90	016	013	011	NA	NA	NA
BUR364	Entrambasmestas	43.17	-3.90	139	NA	011	011	011	011
BUR479	Cilleruelo de Bezana	42.97	-3.85	142	NA	011	011	011	011
BUR481	Cilleruelo de Bezana	42.97	-3.85	142	NA	011	011	011	011
BUR482	Cilleruelo de Bezana	42.97	-3.85	011	NA	021	021	NA	NA
BUR483	Cilleruelo de Bezana	42.97	-3.85	139	NA	011	011	011	114
BUR474	San Martín de Humada	42.66	-4.03	112	011	NA	NA	011	113
BUR475	San Martín de Humada	42.66	-4.03	112	NA	011	011	011	011
BUR476	San Martín de Humada	42.66	-4.03	112	NA	011	011	011	011
BUR477	San Martín de Humada	42.66	-4.03	112	NA	011	011	011	011
BUR478	San Martín de Humada	42.66	-4.03	112	NA	011	011	011	011
BUR489	San Martín de Elines	42.83	-3.87	112	NA	NA	NA	011	011
BUR490	San Martín de Elines	42.83	-3.87	143	NA	011	011	011	011
BUR492	San Martín de Elines	42.83	-3.87	112	NA	011	011	011	011
BUR493	San Martín de Elines	42.83	-3.87	143	NA	011	028	011	011
BUR272	Hornillalatorre	43.05	-3.60	133	NA	011	011	011	011
BUR273	Hornillalatorre	43.05	-3.60	016	NA	011	011	011	011
BUR280	Hornillalatorre	43.05	-3.60	131	NA	NA	NA	011	011

Sample	Locality	Lat*	Long*	Cyt*b*	DBY7	DBX5a	DBX5b	BRCA1a	BRCA1b
BUR281	Hornillalatorre	43.05	-3.60	134	NA	NA	NA	011	011
BUR459	Noceco	43.09	-3.48	016	011	011	NA	011	011
BUR460	Noceco	43.09	-3.48	131	011	NA	NA	NA	NA
BUR462	Noceco	43.09	-3.48	131	NA	011	011	011	011
BUR463	Noceco	43.09	-3.48	131	NA	011	011	011	011
BUR344	Ungo	43.14	-3.26	131	NA	011	016	011	011
BUR345	Ungo	43.14	-3.26	131	NA	NA	NA	011	017
BUR346	Ungo	43.14	-3.26	131	NA	011	011	011	011
BUR347	Ungo	43.14	-3.26	138	NA	011	016	011	011
BUR348	Ungo	43.14	-3.26	112	NA	011	011	011	011
BUR368	Encima Angulo	43.04	-3.18	016	011	011	NA	011	011
BUR369	Encima Angulo	43.04	-3.18	112	011	011	NA	011	011
BUR370	Encima Angulo	43.04	-3.18	112	NA	016	016	018	019
BUR371	Encima Angulo	43.04	-3.18	011	011	011	NA	011	011
BUR372	Encima Angulo	43.04	-3.18	131	NA	011	014	011	011
BUR264	Bentretea	42.73	-3.46	129	NA	011	011	011	011
BUR265	Bentretea	42.73	-3.46	131	NA	011	011	011	011
BUR266	Bentretea	42.73	-3.46	112	NA	NA	NA	011	011
BUR267	Bentretea	42.73	-3.46	112	NA	011	011	011	011
BUR268	Bentretea	42.73	-3.46	129	NA	011	011	011	011
BUR504	Villalonquéjar	42.36	-3.77	112	011	011	NA	011	011
BUR464	Hérran	42.83	-3.23	112	011	011	NA	011	011
BUR465	Hérran	42.83	-3.23	131	NA	011	011	011	011
BUR466	Hérran	42.83	-3.23	129	011	011	NA	011	011
BUR468	Hérran	42.83	-3.23	110	NA	011	011	011	011
BUR443	Aostri de Losa	42.94	-3.10	112	NA	011	011	011	011
BUR444	Aostri de Losa	42.94	-3.10	131	NA	011	011	011	011
BUR445	Aostri de Losa	42.94	-3.10	016	011	016	NA	011	011
BUR446	Aostri de Losa	42.94	-3.10	131	NA	011	011	011	011
BUR447	Aostri de Losa	42.94	-3.10	112	011	011	NA	011	011
BUR330	Torrecitores	41.99	-3.85	110	011	011	NA	011	011
BUR331	Torrecitores	41.99	-3.85	112	NA	011	011	011	011
BUR332	Torrecitores	41.99	-3.85	112	011	011	NA	011	011
BUR333	Torrecitores	41.99	-3.85	112	NA	011	011	011	011
BUR334	Torrecitores	41.99	-3.85	112	NA	NA	NA	021	021
BUR499	Quintanalara	42.18	-3.53	112	011	011	NA	011	011
BUR500	Quintanalara	42.18	-3.53	112	NA	011	011	011	011
BUR501	Quintanalara	42.18	-3.53	112	NA	011	011	011	011
BUR502	Quintanalara	42.18	-3.53	129	011	011	NA	011	011
BUR503	Quintanalara	42.18	-3.53	112	NA	011	011	011	011
BUR469	Contreras	42.02	-3.41	112	NA	011	011	011	011
BUR470	Contreras	42.02	-3.41	136	NA	011	011	011	012
BUR471	Contreras	42.02	-3.41	139	014	011	NA	011	011
BUR472	Contreras	42.02	-3.41	112	NA	011	011	011	017
BUR473	Contreras	42.02	-3.41	112	NA	011	011	011	011
ES-90	Burgos	42.31	-3.27	112	011	011	NA	011	011
ES-91	Burgos	42.31	-3.27	112	011	011	NA	011	011
ES-103	Burgos	42.31	-3.27	112	NA	011	011	011	011
ES-104	Burgos	42.31	-3.27	112	NA	011	011	011	011
ES-113	Burgos	42.31	-3.27	112	NA	011	011	011	011
ES-114	Burgos	42.31	-3.27	112	011	011	NA	011	011
ES-117	Burgos	42.31	-3.27	NA	011	011	NA	NA	NA
ES-122	Burgos	42.31	-3.27	112	011	011	NA	011	011
BUR336	Castrillo de la Reina	41.99	-3.23	136	NA	011	011	011	011
BUR337	Castrillo de la Reina	41.99	-3.23	129	NA	011	011	011	011
BUR338	Castrillo de la Reina	41.99	-3.23	137	NA	NA	NA	011	011
BUR340	Castrillo de la Reina	41.99	-3.23	112	011	011	NA	011	011

Sample	Locality	Lat*	Long*	Cyt*b*	DBY7	DBX5a	DBX5b	BRCA1a	BRCA1b
BUR449	Barbadillo de Herreros	42.15	-3.18	112	011	011	NA	011	011
BUR450	Barbadillo de Herreros	42.15	-3.18	016	NA	011	011	011	011
BUR451	Barbadillo de Herreros	42.15	-3.18	112	NA	022	022	021	021
BUR452	Barbadillo de Herreros	42.15	-3.18	110	NA	022	022	NA	NA
BUR454	Fresneda de la Sierra	42.32	-3.13	112	NA	011	011	011	011
BUR455	Fresneda de la Sierra	42.32	-3.13	112	NA	011	011	011	011
BUR456	Fresneda de la Sierra	42.32	-3.13	112	011	011	NA	011	011
BUR457	Fresneda de la Sierra	42.32	-3.13	140	011	011	NA	011	011
BUR458	Fresneda de la Sierra	42.32	-3.13	129	011	011	NA	011	011
BUR383	Zorraquín	42.33	-3.04	112	NA	011	016	011	110
BUR384	Zorraquín	42.33	-3.04	129	NA	011	011	011	011
BUR386	Zorraquín	42.33	-3.04	112	011	011	NA	011	011
BUR260	Covaleda	41.94	-2.88	011	NA	011	011	016	016
BUR261	Covaleda	41.94	-2.88	128	NA	011	011	011	011
BUR262	Covaleda	41.94	-2.88	129	NA	011	013	011	011
BUR263	Covaleda	41.94	-2.88	130	012	NA	NA	NA	NA
RIOJ1	Baños de Ebro	42.52	-2.68	112	011	011	NA	011	011
RIOJ2	Baños de Ebro	42.52	-2.68	112	011	011	NA	011	011
RIOJ3	Baños de Ebro	42.52	-2.68	112	011	011	NA	011	011
RIOJ4	La Villa de Ócon	42.30	-2.25	112	011	011	NA	011	011
ES-82	Biscaya	43.05	-2.67	110	011	011	NA	011	011
MX791	Marquina Xemein	43.26	-2.51	110	011	011	NA	011	011
MX851	Marquina Xemein	43.26	-2.51	150	011	011	NA	011	011
MX869	Marquina Xemein	43.26	-2.51	110	011	011	NA	011	011
MX877	Marquina Xemein	43.26	-2.51	150	011	011	NA	011	011
MX886	Marquina Xemein	43.26	-2.51	150	011	011	NA	011	011
NAV57	Iturmendi	42.89	-2.10	112	011	011	NA	011	011
NAV61	Iturmendi	42.89	-2.10	153	011	011	NA	011	011
NAV63	Iturmendi	42.89	-2.10	154	011	011	NA	011	011
NAV65	Iturmendi	42.89	-2.10	112	011	011	NA	011	011
NAV67	Iturmendi	42.89	-2.10	110	NA	011	011	011	011
01200483	Fuenterrabia	43.36	-1.79	110	011	011	NA	011	011
NAV26	Arrarás	43.03	-1.80	152	011	011	NA	011	011
NAV27	Arrarás	43.03	-1.80	112	011	011	NA	NA	NA
NAV30	Arrarás	43.03	-1.80	110	011	011	NA	011	011
NAV33	Arrarás	43.03	-1.80	018	NA	011	011	NA	NA
NAV35	Arrarás	43.03	-1.80	NA	011	011	NA	011	011
NAV1	Lesaca	43.23	-1.70	NA	011	011	NA	NA	NA
NAV2	Lesaca	43.23	-1.70	151	NA	011	011	NA	NA
NAV3	Lesaca	43.23	-1.70	151	NA	011	011	NA	NA
NAV4	Lesaca	43.23	-1.70	110	011	011	NA	011	011
NAV8	Lesaca	43.23	-1.70	110	011	011	NA	011	011
01130680	Bertiz	43.10	-1.60	NA	NA	011	011	011	012
NAV12	Equisoain, Monreal	42.70	-1.50	018	011	011	NA	011	011
NAV13	Equisoain, Monreal	42.70	-1.50	018	NA	011	015	011	011
NAV14	Equisoain, Monreal	42.70	-1.50	018	011	011	NA	011	011
NAV15	Equisoain, Monreal	42.70	-1.50	018	011	011	NA	NA	NA
NAV16	Equisoain, Monreal	42.70	-1.50	112	NA	011	011	011	011
EUG-401	Eugui	42.97	-1.52	NA	011	012	NA	011	011
EUG-551	Eugui	42.97	-1.52	NA	NA	011	012	011	011
EUG-561	Eugui	42.97	-1.52	NA	011	011	NA	011	011
EUG-424	Eugui	42.97	-1.52	NA	011	011	NA	011	011
EUG-409	Eugui	42.97	-1.52	NA	NA	011	011	011	011
EUG-508	Eugui	42.97	-1.52	NA	011	011	NA	011	011
EUG-476	Eugui	42.97	-1.52	NA	011	011	NA	011	011
EUG-544	Eugui	42.97	-1.52	NA	011	011	NA	011	011
EUG-536	Eugui	42.97	-1.52	NA	NA	011	011	011	011

Sample	Locality	Lat*	Long*	Cyt*b*	DBY7	DBX5a	DBX5b	BRCA1a	BRCA1b
EUG-511	Eugui	42.97	-1.52	NA	NA	011	011	011	011
NAV46	Ureta	42.95	-1.40	011	NA	011	012	011	011
NAV47	Ureta	42.95	-1.40	110	NA	011	016	011	011
NAV48	Ureta	42.95	-1.40	018	NA	011	011	011	011
NAV50	Ureta	42.95	-1.40	018	NA	011	011	NA	NA
NAV51	Ureta	42.95	-1.40	110	011	011	NA	011	011
NAV41	Usón	42.70	-1.20	018	NA	011	011	NA	NA
NAV42	Usón	42.70	-1.20	018	NA	011	011	NA	NA
NAV43	Usón	42.70	-1.20	018	NA	011	011	NA	NA
AND9	Andorra	42.58	1.65	112	NA	011	011	011	011
AND10	Andorra	42.58	1.65	112	NA	011	011	011	011
AND11	Andorra	42.58	1.65	112	NA	011	011	011	011
SM.852	Andorra	42.58	1.65	NA	011	011	NA	011	011
74060303	La Molina	42.34	1.96	NA	011	011	NA	011	011
73122919	Caralps	42.34	2.16	112	NA	011	011	011	011
73123110	Caralps	42.34	2.16	112	011	011	NA	011	011
74010111	Caralps	42.34	2.16	112	011	011	NA	011	011
74040808	Caralps	42.34	2.16	112	011	011	NA	011	011
74041015	Caralps	42.34	2.16	112	NA	011	011	011	011
TO1810	Tormaleo	41.85	2.23	112	NA	011	011	011	011
CE1629	Centelles	41.79	2.22	112	011	011	NA	011	011
BR1500	El Brull	41.81	2.30	127	011	011	NA	011	011
BR1521	El Brull	41.81	2.30	127	011	011	NA	011	011
BR1534	El Brull	41.81	2.30	112	011	011	NA	011	011
CA1122	La Castanyera	41.78	2.34	112	NA	011	011	011	011
AR1488	Arbúcies	41.81	2.51	112	011	011	NA	011	011
RI1241	Riudaurenes	41.82	2.65	112	NA	NA	NA	011	011
GRAN1	Granollers	41.8	2.28	NA	NA	011	011	011	011
GRAN2	Granollers	41.8	2.28	NA	NA	011	011	011	011

* Latitude and Longitude are presented in decimal degrees; NA - information is not available.

PAPER IV

Morphometric differentiation of the two field vole lineages in the Iberian Peninsula: a geometric analysis of mandible shape

Joana Paupério, Paulo C. Alves and Jeremy B. Searle

Manuscript

MORPHOMETRIC DIFFERENTIATION OF THE TWO FIELD VOLE LINEAGES IN THE IBERIAN PENINSULA: A GEOMETRIC ANALYSIS OF MANDIBLE SHAPE

PAUPÉRIO, J.[1,2,3], ALVES, P.C.[1,3,4], SEARLE, J.B.[2,5]

ABSTRACT

Sister forms that show substantial genetic differentiation without apparent morphological variation are being increasingly detected as a consequence of a growing number of phylogeographic studies on species with wide distribution ranges. The *Microtus agrestis* species complex is just such a species, with three highly divergent lineages with parapatric geographic distributions in Europe: Portuguese, Southern and Northern, yet without any obvious morphological differentiation. In this study we investigate the level of morphological differentiation in the size and shape of the mandibles between the two lineages present in the Iberian Peninsula. Mandibles from several localities within the field vole range in Iberia were obtained either from owl pellet collections or from skeletal material in university or museum depositories. The external side of the right mandibles was photographed and twelve 2-dimensional landmarks were digitized for geometric morphometric analysis. These samples had been previously genotyped, so information on their genetic lineage was available. No sexual dimorphism or allometric shape variability was detected in the dataset. In addition, no significant variation in size was detected, either considering geographically defined categories or genetic lineages. However, the multivariate analysis of shape variability revealed the effect of both geographically defined populations and lineages to be significant. Yet, discriminant analyses showed that the geographic break between the morphological distinct populations is at the boundary of the two lineages rather than other geographical positions. Furthermore, cluster analyses also provided results congruent with the existence of morphometric differentiation between the Portuguese and Southern lineage. Thus, these two forms, defined as 'cryptic species', appear to show already some subtle morphological differences that could only be detected with a robust fine-scale analysis of morphometric variability.

Keywords: *Microtus agrestis*, geometric morphometrics, mandible shape, speciation.

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, InBIO Laboratório Associado, Campus Agrário de Vairão, 4485-661, Vairão, Portugal;
[2]Department of Biology, University of York, Wentworth Way, York YO10 5DD, United Kingdom
[3]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal;
[4] Wildlife Biology Program, College of Forestry and Conservation, University of Montana, 32 Campus Drive, Missoula MT 59801, MT, USA.
[5]Department of Ecology and Evolutionary Biology, E139 Corson Hall, Cornell University, Ithaca, NY 14853-2701, USA

INTRODUCTION

Speciation involves the development of phenotypic, behavioural and morphological differentiation between populations through time. However, these processes do not always occur simultaneously. Sometimes, genetic and/or behavioural differentiation is detected without perceivable changes in morphology, revealing a process that can be referred to as 'cryptic speciation'. Despite the fact that the mechanisms behind cryptic speciation are still mostly unknown, the detection of 'cryptic species' has increased exponentially, largely fuelled by advances of molecular tools (Bickford *et al.* 2007, Pfenninger and Schwenk 2007). In addition, the use of other alternative methods, such as bioacoustics, or of integrative taxonomy approaches that take into account several fields of study, also contributed to the recognition of morphologically cryptic species (*e.g.* von Helverson *et al.* 2001).

Although morphological changes in 'cryptic speciation' are, by definition, not easily perceivable, subtle morphological divergence may be detected in some situations. In these cases, traditional morphometric methods might not be able to differentiate groups. However, geometric morphometrics is nowadays considered as one of the most powerful techniques for the identification of patterns of variation below the species level (Loy 1996; Zelditch *et al.* 2004). In fact these methodologies are becoming an increasingly important complement of molecular phylogenetics, as they provide a powerful tool for examining congruence between morphology and genetics and for studying the ontogeny of organism shape (Cardini 2003).

Geometric morphometrics (Bookstein 1991; Rohlf and Marcus 1993; Corti *et al.* 2000; O'Higgins 2000) allows the comparisons of the shape of a structure described by a set of topographically corresponding points (landmarks). Also, the analysis of shape variation is independent of the structure size, which is measured by centroid size (Loy 1996; Zelditch *et al.* 2004). In mammals, the mandible is usually considered a suitable structure for geometric morphometric analysis (Cardini 2003). It is a simple, highly informative and nearly flat bone that has been the subject of several morphological and phylogenetic studies in rodents (Corti *et al.* 1996; Duarte *et al.* 2000; Corti and Rohlf 2001; Cardini 2003; Gündüz *et al.* 2007). Moreover, mandible ontogeny and the genetic basis of its development have been investigated (Atchley *et al.* 1992; Cheverud *et al.* 1997; Klingenberg *et al.* 2001, 2003). The size and shape of the mandible are the product of numerous interacting genes (Klingenberg *et* al. 2001) and can provide valuable insights into the differentiation of related taxa. Recent studies on cryptic bat species have focused on the analysis of mandible and skull shape, by geometric

morphometric methods, and revealed the existence of morphological differences concordant with proposed taxonomic assignments (*Myotis*: Evin *et al.* 2008; *Pipistrellus*: Sztencel-Jabłonka *et al.* 2009). These studies provide additional support to the power of these tools for detecting subtle morphological variation in cryptic species. Furthermore, new methods that integrate genetic and phenotypic data are been developed. Guillot *et al.* (2012) propose a cluster method, based on a statistical model that deals with genetic, phenotypic and geographical information. This method provides a framework for integrative analysis of molecular and phenotypic data in a spatial background, which may ease the interpretation of complex phylogeographic patterns and the identification and delimitation of populations and/or species. In addition it allows the comparison of different datasets, which can provide clues about the hierarchical patterns of differentiation (Guillot *et al.* 2012).

In the present study we focus on the *Microtus agrestis* species complex which has a wide geographical distribution range across Eurasia and where three highly divergent lineages with parapatric geographic distributions have been described in Europe: Portuguese, Southern and Northern (Jaarola and Searle 2002, 2004, Helborg *et al.* 2005, Paupério *et al.* 2012). However, the distribution range of these lineages does not correspond to any currently accepted subspecies. Moreover, despite the high molecular divergence in every genomic lineage analysed, no obvious morphological differentiation has yet been detected (Almaça, 1991). Hence, the *M. agrestis* species complex might be considered as a good example of cryptic speciation. The estimated times of divergence of these cryptic lineages are fairly recent, dating back to the last glacial period (estimated from ca. 18500 to 70 000 years BP; Paupério *et al.* 2012), which might explain the lack of observable morphological variability. Evidence of gene flow between the lineages has been detected, but the existence of some post-mating isolation mechanisms has also been found (Beysard *et al.* 2011, this chapter, paper III).

Therefore, in this study we complement the recent genetic studies with a comprehensive morphometric analysis of mandible shape in the Iberian Peninsula, where the Southern and Portuguese lineage distribution ranges meet. With this analysis we aim to determine the level of functional morphometric differentiation between the two lineages and relate it to the observed genetic divergence. Hence, we have made use of mandibles from owl pellets and museum samples from over the whole Iberian species range from specimens already analysed genetically (this chapter, paper III) together with some extra samples. We applied geometric morphometric techniques for the analysis of mandible shape since, as already mentioned, these consist of powerful methods for the analysis of intraspecific variability. We also applied recently developed cluster methods that integrate

phenotypic, genetic and geographic data to assess the correspondence between morphometric variability and genetic divergence detected within the Iberian field vole populations. We expect to be able to detect any possible morphometric differentiation in mandible shape between the two lineages. Moreover, although phenotypic and molecular markers may reflect different evolutionary or demographic history, we expect that the integration of these different types of data will provide additional support for the distinctiveness of these evolutionary units and helps gain further insights into the mechanisms of differentiation.

MATERIALS AND METHODS

Sample collection

Mandibles of *Microtus agrestis* from 59 localities within the Iberian Peninsula were selected from a set of 342 samples previously analysed in a study of the genetic structure of the Portuguese and Southern lineages (this chapter, paper III). Individuals from all of the sampled localities have previously been genotyped for mitochondrial and nuclear loci in that study. However, not all individuals genotyped in the study detailed on paper III (this chapter) had intact mandibles that allowed morphometric analysis, so additional samples from some of the studied localities were added to the analysis.

In total, 226 mandibles, consisting of 1 to 18 samples per locality were analysed (Fig. 1, Appendix I). From these, 118 had previously been genotyped, detaining information for more than one locus. For 95 of the samples it was possible to assign gender based on genetic information (56 females and 39 males scored for all loci; see paper III, this chapter).

Geometric morphometric analysis

The external side of the right mandibles was photographed using a Canon EOS 350D digital camera (Canon Inc., Tokyo, Japan) at a 8 megapixel resolution equipped with a Sigma 105mm f/2.8 EX DG Macro Lens (Sigma Coorporation of America, USA) and a 12mm Kenko extension tube (Kenko Tokina Co., Japan), locked to a copy stand at a fixed distance of 39 cm. The mandibles were placed on a horizontal plane and a spirit level was used to check that the lens and the specimen plane were parallel.

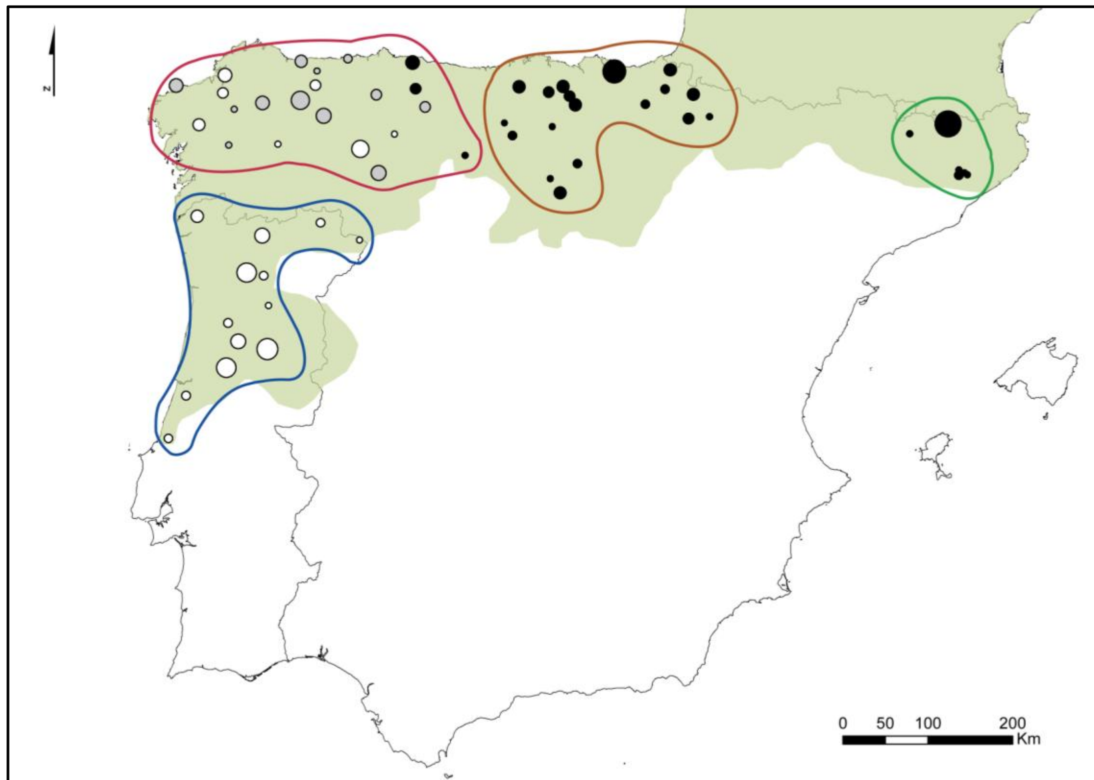


Fig. 1 Sampling localities in the Iberian Peninsula. Each circle represents one locality and the circle area is proportional to the number of samples analysed. Colours of the circle indicate the lineage assignment according to the existing molecular data: white - localities where only individuals from the Portuguese lineage were identified; black - localities where only individuals from the Southern lineage were identified; grey - localities where hybrids were identified. The four geographic regions used in the analysis are identified by coloured contours: Blue - Portugal; Purple - northwestern Spain; Brown - Burgos, Basque Country and Navarra; Green - northeastern Spain (see Methods for details). The green shading corresponds to the distribution range of the field vole (*Microtus agrestis*) in the Iberian Peninsula according to Kryštufek *et al.* (2008).

Twelve landmarks were digitized using tpsDig2 2.16 (Rohlf 2008; Fig. 2). We used TpsSmall 1.20 (Rohlf 2003) to test the approximation of the distances among configurations in the tangent Euclidian space to the Procrustes distance in the shape space. Afterwards, all specimens were superimposed with TpsRelw 1.45 (Rohlf 2007a) by applying a generalized least squares Procrustes superimposition (GPA; Rohlf and Slice, 1990) that allowed for the scaling, translation and rotation of landmark configurations. The resulting shape coordinates of the aligned specimens were then reduced through a principal components analysis (namely, relative warp (RW) analysis), using TpsRelw to obtain shape variables (Rohlf 2007a; n=20).

Centroid size (CS, the square root of the sum of the square distances between each landmark and the centroid, Bookstein 1991) was calculated for estimating the size of each mandible with software TpsRelw (Rohlf 2007a).

Sexual dimorphism and possible differences between the two lineages of field vole were tested by means of two-way (sex x lineage) analysis of variance for both the centroid size (ANOVA) and shape variables (RWs; MANOVA). These analyses were

performed on the dataset for which information on gender (based on genetic data) and on individual genetic assignment was available (n = 95). For the variable 'lineage', three categories were considered: 'Portuguese lineage', individuals that were genetically assigned to the Portuguese lineage; 'Southern lineage', individuals that were genetically assigned to the Southern lineage; 'Hybrid', individuals that had both Southern and Portuguese lineage haplotypes for the studied markers (see this chapter, paper III).

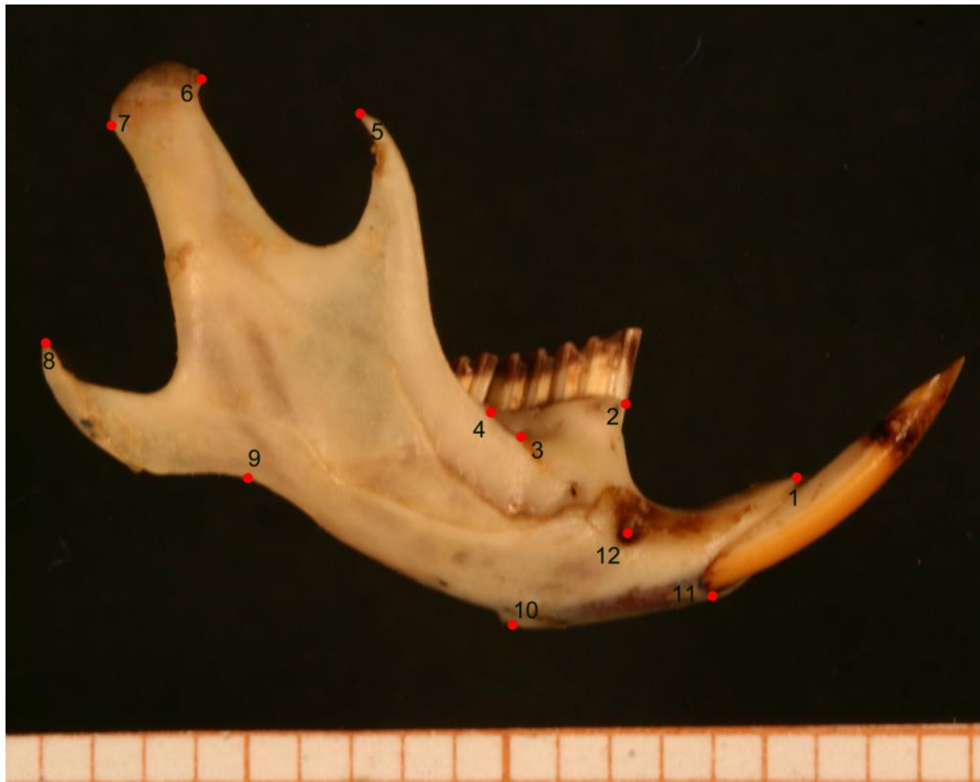


Fig. 2 Landmark configuration on the external view of the field vole right side mandible.

The full dataset (n=226) was then tested separately for differences between the 'lineages' (considering Portuguese localities, Southern localities and localities with hybrids), and 'geographic regions' [Portugal (PT); northwestern Spain, which includes Galicia, Asturias, Cantabria and León (NW); the region that encompasses Burgos, the Basque Country and Navarra (BBN); and northeastern Spain, namely Cataluña (NE); Fig. 1]. One-way analyses of variance for both the centroid size (ANOVA) and shape variables (MANOVA) were carried out.

Discriminant function analyses (DFA) was also performed to test shape differences between 'geographic regions' and 'lineages'. For the test between lineages, discriminant functions were obtained from the analysis of the specimens that were genetically assigned as pure Portuguese and pure Southern (n=111). The other

samples were then classified according to the estimated discriminant functions. To evaluate the performance of the DFA analyses, cross-validation was performed using a jackknife procedure, in which each specimen was classified into a group using the discriminant functions derived from all specimens except the specimen being classified. All the ANOVA, MANOVA and DFA analysis were performed using only the RW's that accounted for the first 90% of shape variance (RW 1 to 10). Both univariate normality and homogeneity of variances of the centroid size and shape variables were tested as a preliminary for this analysis.

The patterns of shape variation related to mandibular size (allometry) were compared for the whole data set with a multivariate regression of the shape variables onto CS. The log-transformed CS was used in this analysis in order to take into account the progressive decrease of the rate of shape change during growth (Zelditch and Fink 1995).

The statistical analyses were performed with the computer programs TpsRegr 1.35 (Rohlf 2007b) and SPSS 20 (IBM, SPSS, 2011).

Cluster analysis

A cluster analysis was employed to assess the congruence between the morphometric variability and the genetic divergence detected within the Iberian field vole populations. For that we have applied the recently developed statistical method implemented in GENELAND v.4.0.2 that combines genetic, phenotypic and geographical data to estimate the number of clusters within a population (Guillot *et al.* 2012). The genetic dataset consisted of 327 samples used in the previous study from the same localities (this chapter, paper III) where more than one locus has been genotyped. Of these, 119 were analysed for mandible shape. These samples were sequenced for 4 loci with different modes of inheritance (maternal, paternal and bi-parental) and each unique haplotype was coded as a different allele (see this chapter, paper III). The phenotypic dataset is composed by the first ten shape variables (RWs) that explain over 90% of the total variance and by CS. For the spatial information we considered some uncertainty of the spatial coordinates (0.01º) in the analysis, to account for the fact that most of the samples were collected from barn owl pellets and thus originated from populations within the hunting area of this species (considered as a 1 km radius from the nesting site, Andries *et al.* 1994).

Finally, two different analyses were performed. The first analysis considered only the morphometric dataset. Then an analysis was performed that incorporated both morphometric and genetic data. These datasets were analysed using the spatial model and under the uncorrelated allele frequency prior (UFM). In each case, we performed 5

independent Markov chain Monte Carlo (MCMC) runs of 10000000 iterations with a thinning of 5000 iterations, and discarded the first 10% as burnin. We allowed the number of populations (*K*) to vary between 1 and 10.

RESULTS

Geometric morphometric analysis

The linear tangent space closely approximates the shape space, as no specimens were found to deviate significantly from the linear regression line (TpsSmall: r = 1.00, slope = 0.99). No significant effect of sex or lineage, or of the interaction sex x lineage was detected for CS (n=95; Table 1; Fig. 3). For the lineages, the median centroid sizes (Portuguese lineage: 20.15; Southern lineage: 19.82, Fig. 3) as well as the means (mean ± s.e.: Portuguese: 19.91 ± 0.15; Southern: 19.91 ± 0.11) are very similar. Furthermore, the analysis of the full data set (n=226) also showed no significant differences in CS either between the lineages (ANOVA: $F_{2,223}$ = 0.541, p = 0.583) or considering the four geographic regions (ANOVA: $F_{3,222}$ = 1.862, p = 0.137).

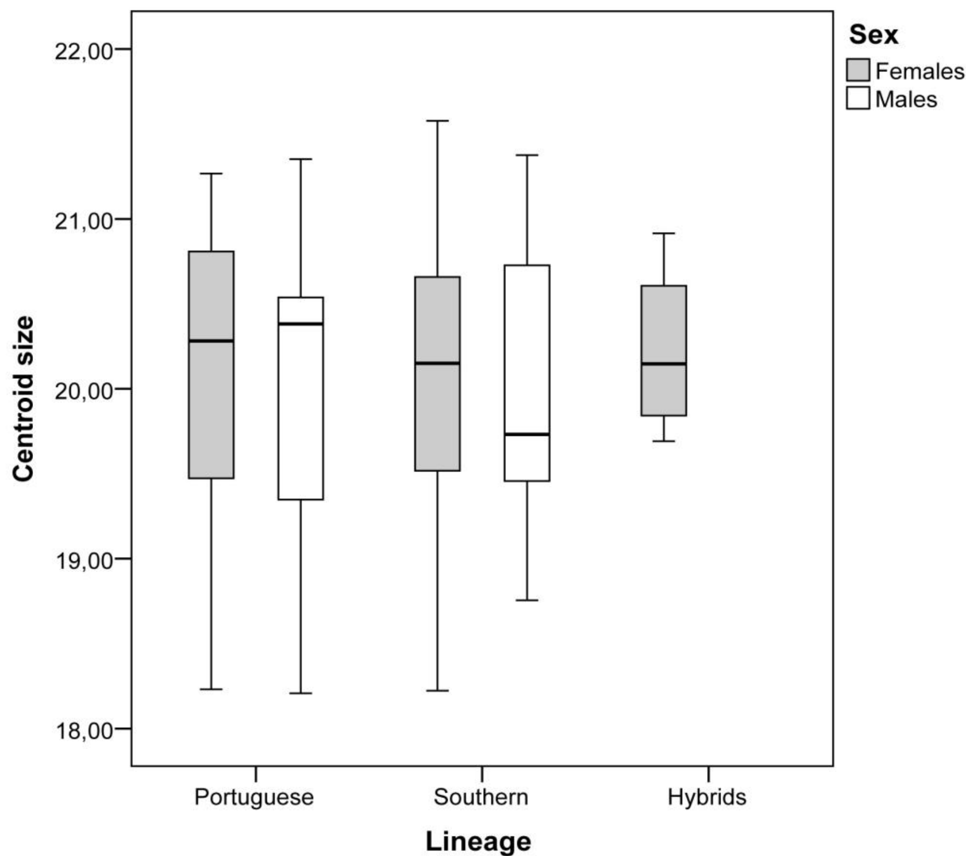


Fig. 3 Box-plot of mandible CS variation (median and percentiles, 25-75%, as well as maximum and minimum values) in the two lineages of field vole (Portuguese, n= 35; Southern, n=56) and hybrid individuals (n=4) by sex (Females, n=56; Males, n=39).

Table 1 Effect of lineage, sex and of the interaction sex x lineage on mandible centroid size considering the dataset of genotyped individuals (n=95).

Effect[a]	**Sum of squares**	**df**	***F***	***p***
Lineage	0.234	2	0.163	0.849
Sex	0.000	1	0.001	0.981
Sex x lineage	0.000	1	0.001	0.981
Error	64.453	90		

[a] The assumptions of normality and homogeneity of group variances were not violated (Kolmogorov-Smirnov test: p=0.095; Levene's test: $F_{(4,90)}$=0.626, p=0.645).

The 10 shape variables (RWs) were first tested for univariate normality and homocedasticity. The Kolmogorov-Smirnov test revealed some deviations from normality. In addition, variance-covariance matrices of shape variables were not homogeneous (χ^2 = 206.96, df = 110, p < 0.001; BoxM = 222.08, $F_{110,77099}$ = 1.878, p < 0.0001). The lack of homogeneity of variance-covariance matrices is a rather common problem in taxonomic comparisons. However, ANOVA's and discriminant analysis are considered rather robust to violations of the assumption of homoscedasticity (Marcus 1990). Nevertheless, the results from these analyses should be considered with caution.

The analysis of mandibular shape of the genotyped individuals shows the existence of significant differences between the lineages, but no significant effect of sex on shape or of the interaction of sex with lineages was detected (Table 2). In addition, testing the full dataset (n = 226) also revealed a significant effect of lineage on the shape of the mandible (λWilks = 0.272, $F_{20,428}$ = 19.622, p < 0.0001). However, when considering the four geographical regions, a significant effect of region was also observed (λWilks = 0.256, $F_{30,625.9}$ = 12.329, p < 0.0001).

Table 2 Effect of sex, lineage and of the interaction sex x lineage on the 10 shape variables considering the dataset of genotyped individuals (n=95).

Effect	**λWilks**	***F***	**Df**	***p***
Lineage	0.257	7.887	20, 162	0[a]
Sex	0.816	1.822	10, 81	0.070
Sex × lineage	0.813	1.856	10, 81	0.064

[a] p values smaller than the minimum value that can be displayed by SPSS.

The distribution of the samples relative to the first two components of mandible shape (which explain 56% of the variability; Fig. 4) shows a wide variability in mandible shape both within each lineage and within geographic regions as well as a broad overlap between every group considered. Nevertheless, the significant differences detected in the MANOVAs indicate that some degree of differentiation between the groups (geographic regions and lineages) can be detected.

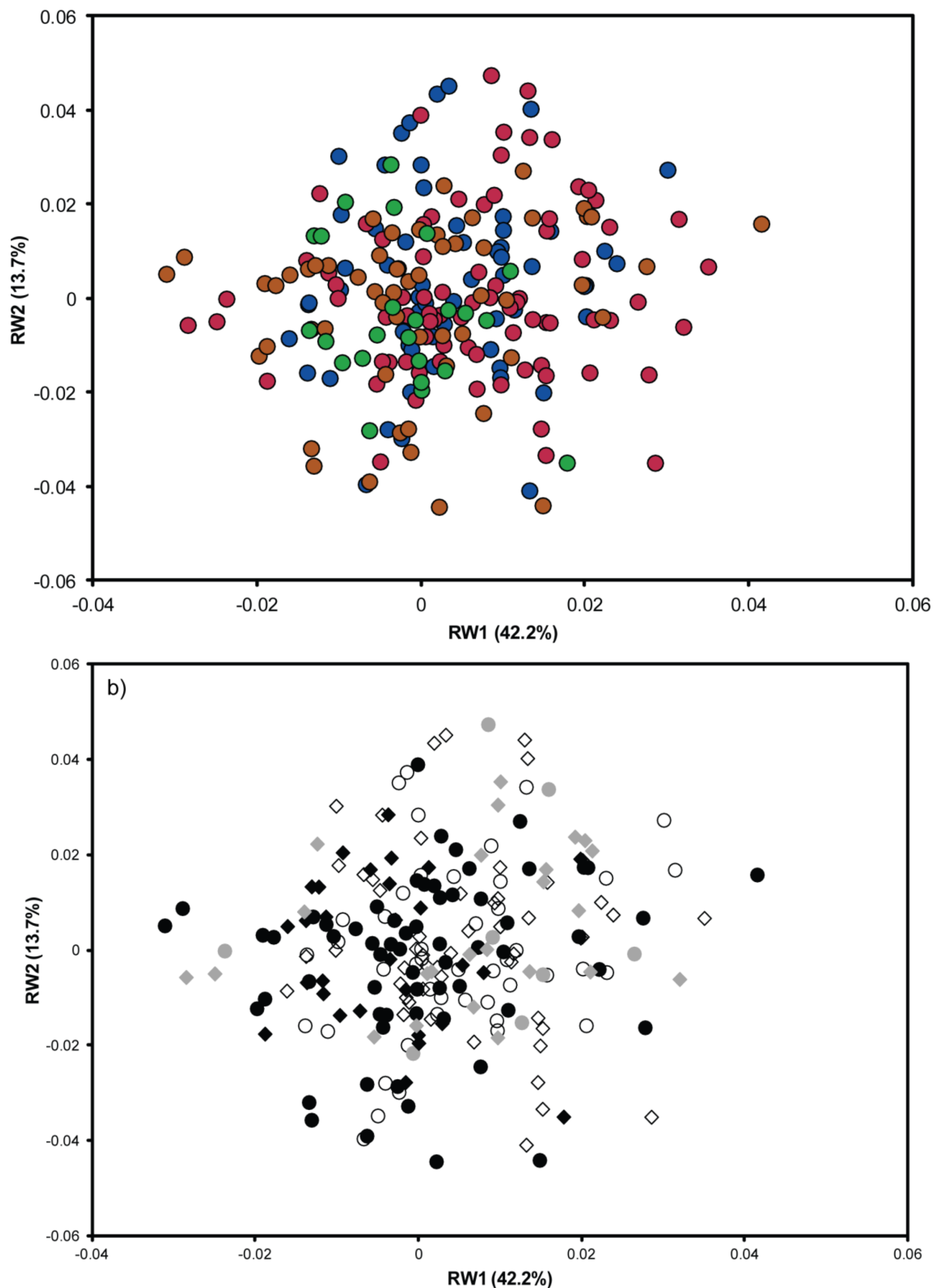


Fig. 4 Ordination of the samples in the space of the first two shape variables (RW1 and RW2) based on the matrix of shape variables for the full data-set (n=226). a) Samples coloured according to geographic regions. PT - blue; NW - purple; BBN - brown; NE - green. b) Samples grouped according to the lineage assignment. Circles correspond to individuals for which genetic information exists and diamonds to samples that were not sequenced. Symbols are coloured according to the lineage assignment (circles) or populations of origin (diamonds). White - Portuguese lineage; Black - Southern lineage; Grey - hybrids. Each symbol represents one individual. Two samples from Burgos and other two samples from central Portugal are not shown as they represent outliers to the range of values of RW1 and RW2 shown in this scatter plot.

DFA, a method that maximizes the differences between groups, was applied to the first ten shape variables considering either the effect of the geographic region and the lineage. The discriminant function obtained for the geographic region showed that the differences between the four groups considered only allowed for the correct classification of 68.1% of the individuals (cross validated 63.7%). In fact, the ordination of the samples according to the two discriminant functions indicates a wide overlap between the PT and NW regions and between the BBN and NE regions (Fig. 5). However, considering the differentiation between the lineages, the DFA indicated that there was enough difference in mandible shape to classify correctly 94.6% of all individuals in the analysis (Fig. 6; n=111; cross validated 91%) despite the morphometric overlap in principal components space (Fig. 4). When the discriminant function estimated was applied to the entire dataset most of the samples was correctly classified (96%). In total, only 8 individuals assigned as Southern lineage were classified as belonging to the Portuguese lineage and one individual genetically assigned to the Portuguese lineage was classified as Southern lineage by the DFA. Of the 32 individuals labelled as hybrids (which includes both animals with hybrid genotype and individuals without genetic information from populations with hybrids) 26 were classified as Portuguese lineage and only 6 were assigned to the Southern lineage.

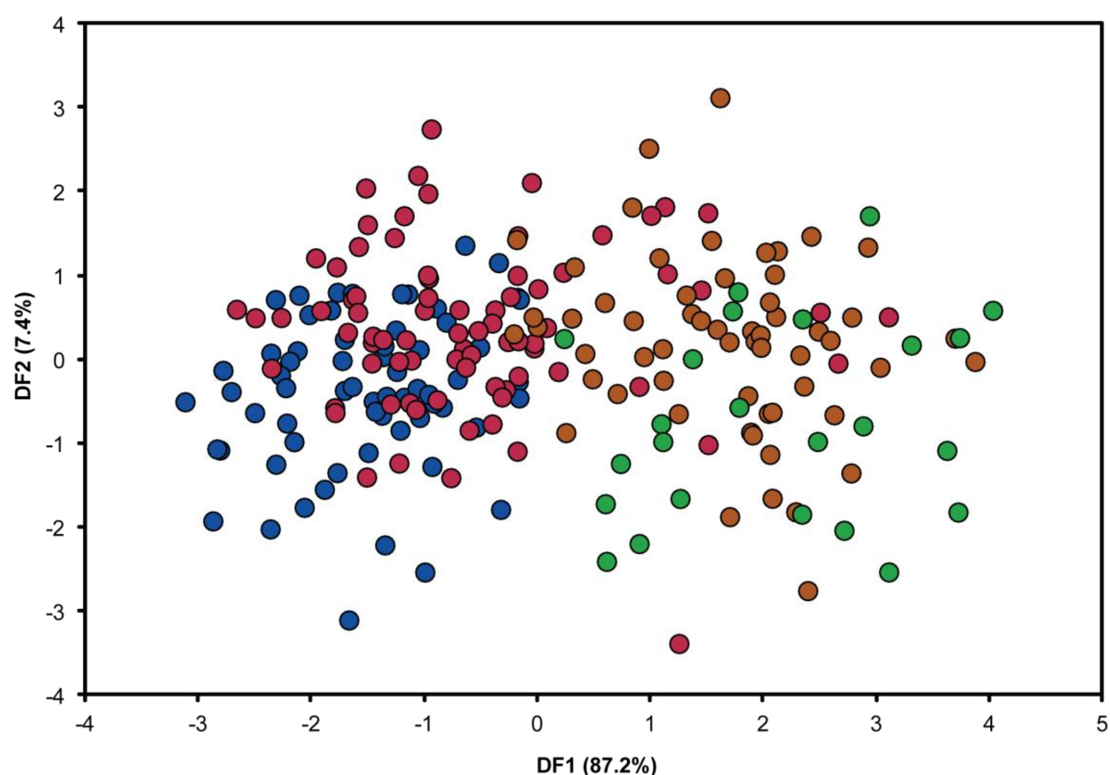


Fig. 5 Ordination of the samples according to the two discriminant functions (DF1 and DF2) obtained in the analysis between geographic regions. Circles are coloured by region. PT: blue; NW: purple; BBN: brown; NE: green.

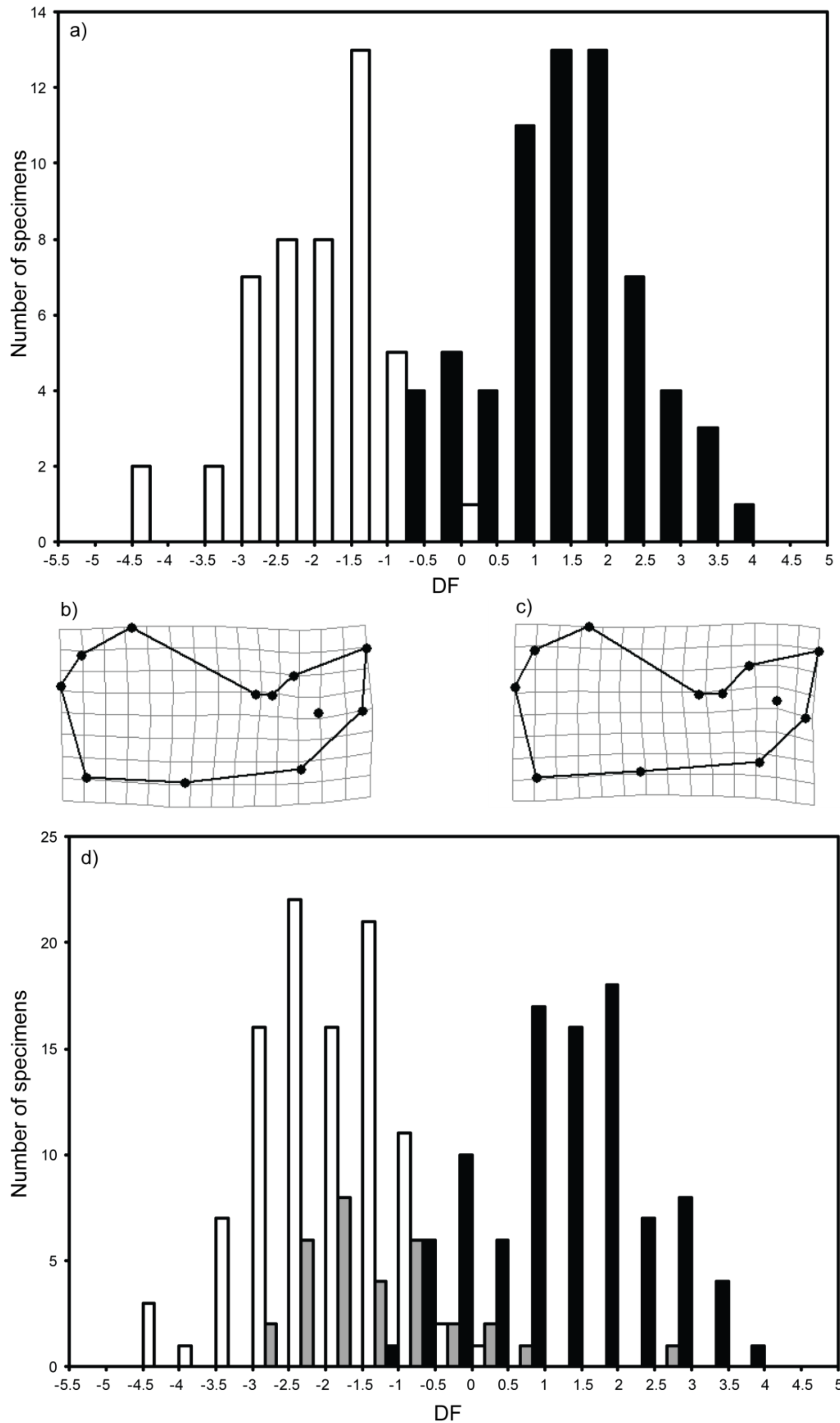


Fig. 6 Frequency histogram of discriminant function scores (DF) obtained in the analysis between lineages. a) Considering the dataset of genotypic individuals (n=111). Bars are coloured according to the lineage assignment. White - Portuguese lineage; Black - Southern lineage. b) and c) TPS deformation grids showing the shape changes at DF. b) represents the extreme configuration for the Portuguese lineage and c) for the Southern lineage. The deformations were magnified three times for clarity. The mandibles in the deformation grids are stylised representations obtained by linking the pairs of landmarks (see Fig. 2). d) Considering all samples (n=226). Bars are coloured according to the lineage assignment of the population of origin. White - Portuguese lineage; Black - Southern lineage; Grey - hybrid populations.

The main shape differences between the two lineages are illustrated by thin-plate spline (TPS) deformation grids obtained through the regression of the shape variables onto the discriminant function (Fig. 6). The deformation grids correspond to the extreme points of the axis and were magnified three times in order to make changes more evident. The subtle shape differences identified between the Portuguese and Southern lineages seem to be mostly located in the anterior part of the mandible. In the Southern lineage individuals the anterior extremity of the maxillary tooth row (landmark 2; Fig. 2) is in an upper position when related to the intersection of the dental ridge with the dorsal portion of the masseteric ridge (landmark 4). In addition there is a slight anterior displacement of the mental foramen (landmark 12) in the Southern lineage mandibles.

The linear regression of shape variables onto log-transformed centroid size, that tests for size related shape changes (allometry), was not significant (λWilks = 0.935, $F_{20, 205}$ = 0.714; p = 0.809) considering the full data set (n = 226).

Cluster analysis

The cluster analysis performed with Geneland provided consistent results across the five independent runs for both datasets (morphometric only and genetic and morphometric together). The morphometric dataset resulted in 5 clusters for all runs, while the joint analysis of the morphometric and genetic data resulted in three clusters in all runs. However, in both analyses one of the clusters was an empty cluster, considered as a 'ghost population', since no individual belonged to it (Guillot *et al* 2005). As in these situations Guillot *et al.* (2005) consider that the runs performed to estimate K give poor results regarding the individual assignment to each of the clusters, we have performed a second set of analyses, with fixed K, where we have disregarded the 'ghost populations' as recommended by these authors. Therefore, in these analyses, we performed 5 independent MCMC runs of 10000000 iterations with a thinning of 5000 iterations, considering a value of K=4 for the morphometric dataset and a value of K=2 for joint analysis of the morphometric and genetic data. Ten percent of the iterations were discarded as burnin in both analyses.

The assignment results obtained in the best run (with the highest likelihood) for the morphometric dataset are presented in Fig. 7. One of the populations appears to correspond to the Portuguese lineage (red) and has a distribution that includes central Portugal and the northwest of Spain. Of the 124 individuals assigned to this cluster, 90% had an assignment probability over 0.80. The remaining 12 individuals, that belonged to two populations located in central Portugal, one in northern Portugal and the other in Astúrias had assignment probabilities that ranged between 0.47 and 0.67.

The other three clusters are mostly located within the Southern lineage range. One of these is composed by only one population (blue) in the Navarra region, with an assignment probability of 0.85. The individual assignment to the green cluster had a relatively high probability for the majority of samples (> 0.72), although some individuals located close to the red cluster showed low probabilities of assignment (~0.56; Fig. 7). The orange cluster, that included localities from the Burgos region to Cataluña, showed high assignment probabilities for most individuals (>0.83). Only 5 individuals from the southernmost localities in Burgos had lower assignment probabilities (~0.73).

The joint analysis of the morphometric and genetic data resulted in the assignment of the individuals to two clusters that have distributions equivalent to the ranges of the two Iberian field vole lineages (Fig. 8). The assignment of the individuals to each cluster was very well supported, since 99% of the individuals had an assignment probability over 0.95. Only 4 individuals from one population in Asturias were assigned to the black cluster (corresponding to the Southern lineage) with a lower probability (0.71). This locality is nevertheless also assigned to the Southern lineage by the genetic data alone (see this chapter, paper III).

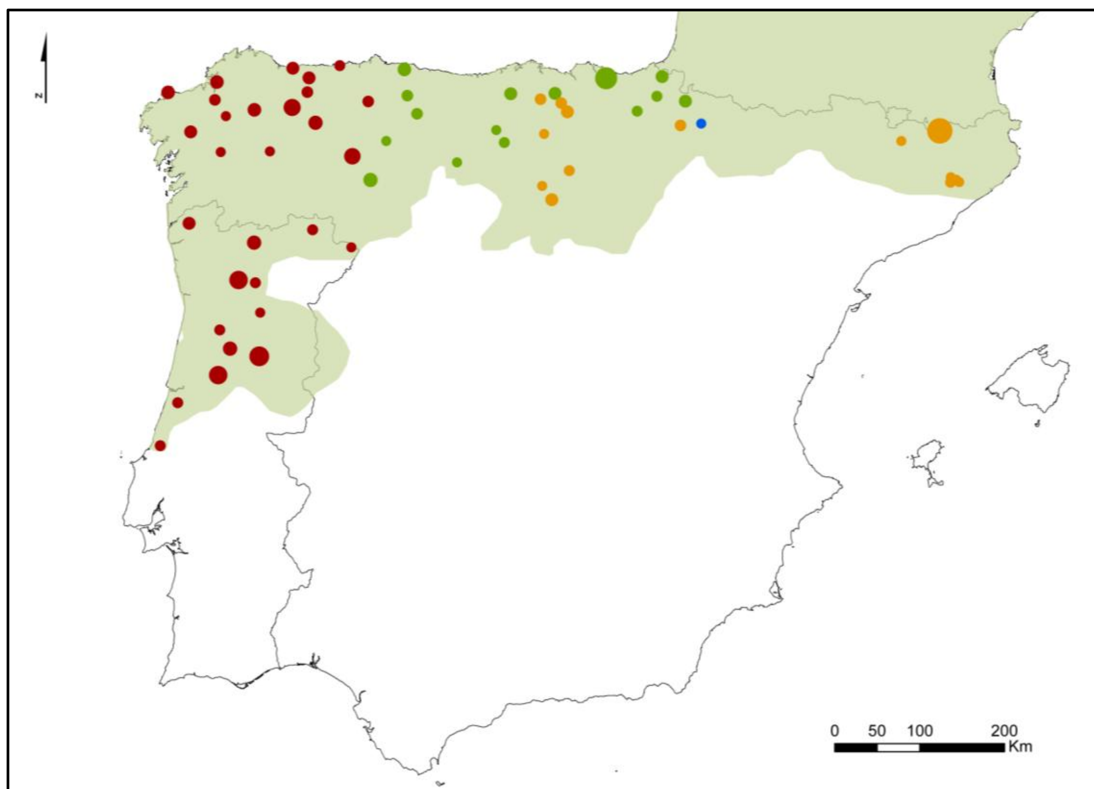


Fig. 7 Population structure inferred by Geneland for the morphometric data (n=226). Each circle represents one locality and the circle area is proportional to the number of samples analysed. Colours of the circle indicate the modal assignment of the individuals to each cluster. The four clusters are identified by different colours. The green shading corresponds to the distribution range of the field vole (*Microtus agrestis*) in the Iberian Peninsula according to Kryštufek *et al.* (2008).

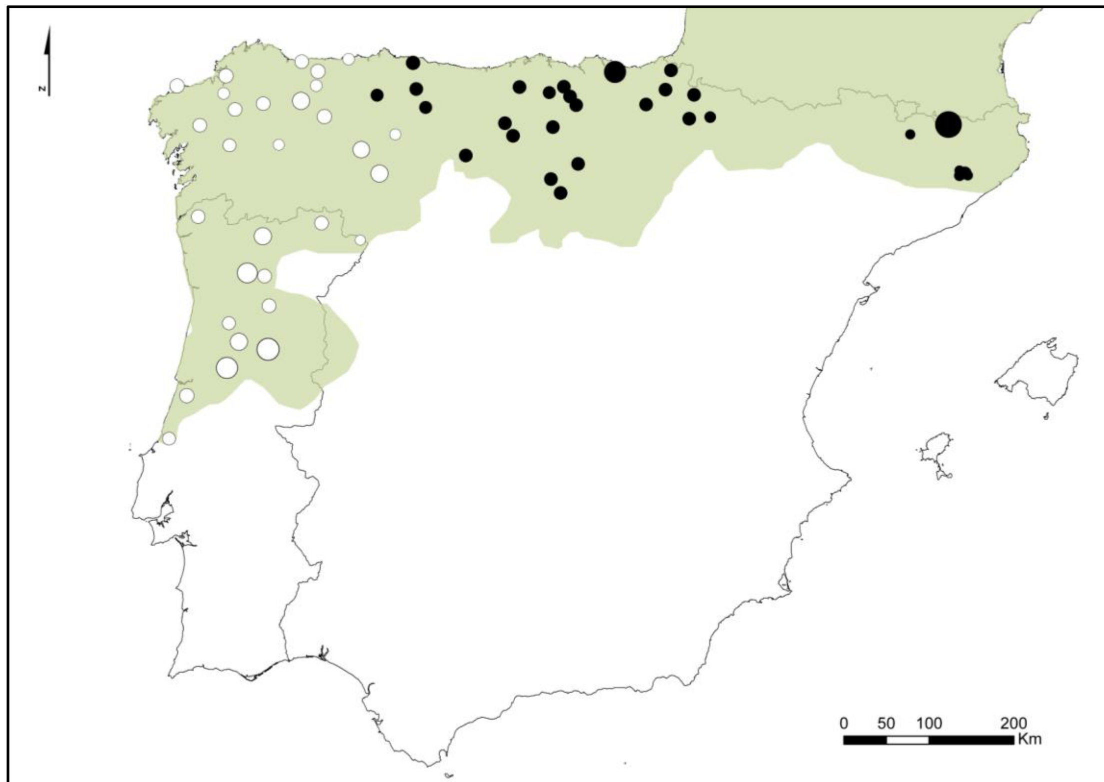


Fig. 8 Population structure inferred by Geneland for the dataset including the genetic and morphometric data (n=327). Each circle represents one locality and the circle area is proportional to the number of samples analysed. Colours of the circle indicate the modal assignment of the individuals to each cluster. The black and white colour refers to the two clusters identified in the analysis. The green shading corresponds to the distribution range of the field vole (*Microtus agrestis*) in the Iberian Peninsula according to Kryštufek *et al.* (2008).

DISCUSSION

No sexual dimorphism in the size and shape of the mandible was observed within the field vole populations, either considering them as a whole or as two different evolutionary units. However, since all of the samples consisted of mandibles found in barn owl pellets, the gender assignment of the specimens was only based on the genetic data available, and thus only conducted in a subset of the samples. The specimens were considered as males if they showed successful amplification of the Y chromosome intron and females if this amplification failed (see this chapter, paper III). Although we have only considered for this assignment individuals that had information for all loci analysed (this chapter, paper III) it is nevertheless possible that the Y chromosome failed to amplify in some males, despite the successful amplification of all other loci. In view of these facts, the results of the tests relative to sexual dimorphism should be considered with caution. Even so, absence of sexual dimorphism has been

previously observed in the analysis of cranial and mandible shape variability in other small mammals (*e.g.* Vega *et al.* 2010) including voles (*e.g.* Kryštufek *et al.* 2012).

Moreover, allometric shape variation was not detected in the analysed dataset. Allometric shape changes of the mandible have been reported during the growth of marmots (Cardini and Tongiorgi 2003) and have also been detected in studies of morphological variability within rodents (Michaux *et al.* 2007) and shrews (Vega *et al.* 2010). However, Cardini (2003) did not detect any influence of allometry on mandible shape variability among phylogenetic clades.

Regarding the size variability of field vole mandibles within the Iberian Peninsula, we have not found any divergence in CS either between geographic regions or between lineages. Despite of this lack of size variability among the groups considered (geographic and genetic), the analysis of mandible shape showed significant differentiation both between regions and lineages. The significant differences found between geographic regions could possibly imply that the observed shape divergence between lineages might be due mostly to geographic variation. However, the discriminant function analysis based on geographic region revealed a low percentage of correct classification while correct classification according to lineage was high. So, it appears that the geographic break between the morphologically distinct populations is at the boundary of the Portuguese and Southern lineages rather than other geographical positions.

Cluster analysis with Geneland identified 5 populations for the morphometric dataset and 3 populations for the joint analysis of morphometric and genetic data. However, both analyses resulted in the inference of one ghost population each (empty cluster which was modal for none of the analysed individuals). The occurrence of ghost populations has been reported several times in the analysis of molecular data after it was first identified by Guillot *et al.* (2005; *e.g.* Coulon *et al.* 2006; Coulon *et al.* 2008). It is thought that this problem is mostly due to departures from model assumptions and weak mixing of the Markov chain Monte Carlo (MCMC) algorithms that lead to spurious classes in the runs (Coulon *et al.* 2006; Guillot 2008). This has been considered a difficult problem to deal with and has also been reported in non-genetic mixture model literature (Jasra *et al.*, 2005; Guillot 2008). We have thus followed Guillot *et al.* (2005) and repeated the analysis with a fixed K value that corresponded only to the non-empty populations, disregarding the empty populations.

Therefore, the morphometric dataset resulted in the estimate of four different clusters within the Iberian Peninsula. One of these populations seems to overlap mostly with the Portuguese lineage distribution, while the other three are mainly located within the Southern lineage range. Individual assignment to each cluster exhibited a

small variability between runs (10% of the sampled specimens), particularly regarding the individuals which showed lower probabilities of assignment to the red and green clusters. Moreover, most of these individuals are from localities near the putative area of contact of the two lineages. The combined analysis of the morphometric and genetic data for the same localities resulted in only two populations. The distribution of these two clusters is congruent with the range of the two field vole lineages in Iberia. Moreover, the assignment of the individuals to each of the two clusters is highly similar to the obtained when the Geneland analysis was performed only with the genetic data (this chapter, paper III). Although there is no mismatch between the assignment of the individuals using only genetic or including also morphometric data, some disparity is observed in the probability of assignment of four individuals from the putative area of contact, which display lower probabilities of assignment when the morphometric variables are included.

The evidence collected from the morphometric analysis of the field vole mandible shape in the Iberian Peninsula in combination with the molecular data for the same individuals (this chapter, paper III) identifies a subtle morphometric differentiation between the two field vole lineages. When analysing the specific differences observed between the extreme shapes of the mandible between lineages, shape variation seems to be mostly located in the anterior part of the mandible. In fact, Monteiro and Reis (2005) in their study of the spiny rats (genus *Trinomys*), highlighted the coronoid process (in the posterior part of the mandible) as the most conserved morphogenetic component of the mandible of these species. Even though these small differences are identified in the analysis, they are not easily detected (mostly perceivable due to the magnification of the deformation grid) and cannot be used as a diagnostic feature between the two lineages.

Although the distribution of these two lineages does not correspond to the range of any currently recognised subspecies, the Portuguese lineage seems to relate to a taxa described by Bocage in 1865, *Microtus rozianus* from a specimen in central Portugal, but that was latter disregarded as a valid subspecies by Niethammer (1964, 1970), who considered it to be a part of the more widespread *M. agrestis bailoni* (de Sélys-Longchamps 1841; see Paupério *et al* 2012). Later, Almaça (1993) who analysed the variation in pelage colour, morphometric traits of the skull and mandible and dental traits of field voles from the Iberian Peninsula and southern France also found little support for the designation of either *M. agrestis rozianus* or *M. a. bailoni* as subspecies, since he attributed the detected morphological variability to clinal variation between highland and lowland populations. The present study provides however overall evidence that there is subtle morphometric variability within the Iberian

Peninsula that is explained by the differentiation of the two lineages. The detection of these differences was only possible due to the application of a robust fine-scale analysis of mandible shape.

Although the present study concerns only the two lineages occurring in the Iberian Peninsula we believe that the analysis of the mandible shape of all three lineages of this species complex may also reveal the existence of morphometric distinction.

Acknowledgements

We thank Alfonso Ceña, Antoni Arrizabalaga and the Museum of Natural Science of Granollers, António Mira, Augusto de Castro, Enrique Castiéns, Fernando Jubete, Hélia Gonçalves, Jacinto Róman, Maria José Lopez-Fuster and the University of Barcelona, Carlos Fonseca, Carlos Nores, for providing some of the samples used in this study. We are grateful to Alexandra Valongueiro and Ana Silva for help in owl pellet material identification. This work was supported by Programa Operacional Potencial Humano – Quadro de Referência Estratégico Nacional funds from the European Social funds and Portuguese Ministério da Educação e Ciência (FCT, SFRH/BD/27369/2006 PhD grant to J. Paupério).

REFERENCES

Almaça C (1993) On the Portuguese population of *Microtus agrestis* (Linnaeus, 1761). *Arquivos do Museu Bocage Nova Série*, **2**, 319-335.

Andries AM, Gulink H, Herremans M (1994) Spatial model of the barn owl *Tyto alba* habitat using landscape characteristics derived from SPOT data. *Ecography*, **17**: 278-287.

Atchley ER, Cowley DE, Vogl C, McLellan T (1992) Evolutionary divergence, shape change, and genetic correlation structure in the rodent mandible. *Systematic Biology*, **41**, 196-221.

Beysard M, Perrin N, Jaarola M, Heckel G, Vogel P (2012) Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (*Microtus agrestis*). *Journal of Evolutionary Biology*, **25**, 400-408.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**, 148-155.

Bookstein FL (1991) *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge.

Cardini A (2003) The geometry of the marmot (Rodentia: Sciuridae) mandible: phylogeny and patterns of morphological evolution. *Systematic Biology*, **52**, 186-205.

Cardini A, Tongiorgi P (2003) Yellow-bellied marmots (*Marmota flaviventris*) 'in the shape space' (Rodentia, Sciuridae): sexual dimorphism, growth and allometry of the mandible. *Zoomorphology*, **122**, 11-23.

Cheverud JM, Routman EJ, Irschick DJ (1997) Pleiotropic effects of individual gene loci on mandibular morphology. *Evolution*, **51**, 2006-2016.

Corti M, Fadda C, Simson S, Nevo E (1996) Size and shape variation in the mandible of the fossorial rodent *Spalax ehrenbergi*. In *Advances in Morphometrics* (eds Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE). Plenum, New York, 303-320.

Corti M, Marcus LF, Hingst-Zaher E (2000) Introduction to symposium: geometric morphometrics in mammalogy. *Hystrix*, **11**, 3-7.

Corti M, Rohlf FJ (2001). Chromosomal speciation and phenotypic evolution in the house mouse. *Biological Journal of the Linnean Society*, **73**, 99-112.

Coulon A, Guillot G, Cosson JF, Angibault JMA, Aulagnier S, Cargnelutti B, Gallan M, Hewison AJM (2006) Genetic structure is influence by landscape features: empirical evidence from a roe deer population. *Molecular Ecology*, **15**, 1669-1679.

Coulon A, Fitzpatrick JW, Bowman R, Stith BM, Makarewich CA, Stenzler LM, Lovette IJ (2008) Congruent population structure inferred from dispersal behavior and intensive genetic surveys of the threatened Florida scrub-jay (*Aphelocoma coerulescens*). *Molecular Ecology*, **17**, 1685-1701.

Duarte LC, Monteiro LR, Von Zuben FJ, dos Reis SF (2000) Variation in the mandible shape in *Thrichomys aperoides* (Mammalia: Rodentia): geometric analysis of a complex morphological structure. *Systematic Biology*, **49**, 563-578.

Evin A, Baylac M, Ruedi M, Mucedda M, Pons JM (2008) Taxonomy, skull diversity and evolution in a species complex of *Myotis* (Chiroptera: Vespertilionidae): a geometric morphometric appraisal. *Biological Journal of the Linnean Society*, **95**, 529-538.

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261-1280.

Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation – the correlated allele frequencies model revisited. *Bioinformatics*, **24**, 2222-2228.

Guillot G, Renaud S, Ledevin R, Michaux J, Claude J (2012) A unifying model for the analysis of phenotypic, genetic and geographic data. *Systematic Biology*. DOI: 10.1093/sysbio/sys038.

Gündüz I, Jaarola M, Tex C, Yeniyurt C, Polly PD, Searle JB (2007) Multigenic and morphometric differentiation of ground squirrels (*Spermophilus*, Sciuridae, Rodentia) in Turkey with the description of a new species. *Molecular Phylogenetics and Evolution*, **43**, 916-935.

Hellborg L, Gündüz I, Jaarola M (2005) Analysis of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology*, **14**, 2025–2031.

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.

Jasra A, Holmes CC, Stephens DA (2005) Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, **20**, 50-67.

Klingenberg CP, Leamy LJ, Routman EJ, Cheveraud JM (2001) Genetic architecture of mandible shape in mice: Effects of quantitative trait loci analyzed by geometric morphometrics. *Genetics*, **157**, 785-802.

Klingenberg CP, Mebus K, Auffray JC (2003) Developmental integration in a complex morphological structure: how distinct are the modules in the house mouse mandible? *Evolution and Development*, **5**¸522-531.

Kryštufek B, Klenovšek T, Bužan EV, Loy A, Janžekovič F (2012) Cranial divergence among evolutionary lineages of Martino's vole, *Dinaromys bogdanovi*, a rare Balkan paleoendemic rodent. *Journal of Mammalogy*, **93**, 818-825.

Loy A (1996) An introduction to geometric morphometrics and intraspecific variation: a fascinating adventure. In *Advances in Morphometrics* (eds Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE). Plenum, New York, 271-273.

Marcus LF (1990) Traditional morphometrics. In *Proceedings of the Michigan morphometrics workshop* (eds Rohlf FJ, Bookstein FL). University of Michigan Museum of Zoology, Ann Arbor.

Michaux J, Chevret P, Renaud S (2007) Morphological diversity of Old World rats and mice (Rodentia: Muridae) mandible in relation with phylogeny and adaptation. J*ournal of Zoological Systematics and Evolutionary Research*, **45**, 263-279.

Monteiro LR, Reis SF (2005) Morphological evolution in the mandible of the spiny rats, genus *Tricomys* (Rodentia, Echimyidae). J*ournal of Zoological Systematics and Evolutionary Research*, **43**, 332-338.

Niethammer J (1964) Ein Beitrag zur Kenntnis der Kleinsauger Nordspaniens. *Zeitschrift für Säugetierkunde*, **29**, 193-220.

Niethammer J (1970) Uber Kleinsauger aus Portugal. *Bonner Zoologische Beiträge*, **21**, 89-118.

O'Higgins P (2000) The study of morphological variation in the hominid fossil record: biology, landmarks and geometry. *Journal of Anatomy*, **197**, 103-120.

Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB (2012) Cryptic speciation in the field vole: a multilocus approach confirms highly divergent lineages in Europe. *Molecular Ecology*. DOI: 10.1111/mec.12024.

Pfenninger M, Schwenk K (2007) Cryptic animal species are homogeneously distributed among taxa and geographical regions. *BMC Evolutionary Biology*, **7**, 121-127.

Rohlf FJ (2003) *tpsSmall*, Version 1.20. Department of Ecology and Evolution. State University of New York. Stony Brook, New York.

Rohlf FJ (2007a) *tpsRelw,* Version 1.45. Department of Ecology and Evolution. State University of New York. Stony Brook, New York.

Rohlf FJ (2007b) *tpsRegr*, Version 1.35. Department of Ecology and Evolution. State University of New York. Stony Brook, New York.

Rohlf FJ (2008) *tpsDig*, Version 2.12. Department of Ecology and Evolution. State University of New York. Stony Brook, New York.

Rohlf FJ, Marcus LF (1993) A revolution in morphometrics. *Trends in Ecology and Evolution*, **8**, 129-132.

Rohlf FJ, Slice D (1990) Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Biology*, **39**, 40-59.

Sztencel-Jabłonka A, Jones G, Bogdanowic ZW (2009) Skull morphology of two cryptic bat species: *Pipistrellus pipistrellus* and *P. pygmaeus* - a 3D geometric morphometrics approach with landmark reconstruction. *Acta Chiropterologica*, **11**, 113-126.

Vega R, Amori G, Aloise C, Cellini S, Loy A, Searle JB (2010) Genetic and morphological variation in a Mediterranean glacial refugium: evidence from Italy pygmy shrews, *Sorex minutus* (Mammalia: Soricomorpha). *Biological Journal of the Linnean Society*, **100**, 774-787.

von Helversen O, Heller KG, Mayer F, Nemeth A, Volleth M, Gombkötö P (2001) Cryptic mammalian species: a new species of whiskered bat (*Myotis alcathoe* n. sp.) in Europe. *Naturwissenschaften*, **88**, 217–223.

Zelditch ML, Fink WL (1995) Allometry and developmental integration of body growth in a piranha, *Pygocentrus nattereri* (Teleostei: Ostariophysi). *Journal of Morphology*, **223**, 341-355.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) *Geometric Morphometrics for Biologists: a Primer*. Elsevier Academic Press, London.

Appendix I

Details of the localities sampled for the field vole in this study. The numbers of samples for each dataset is also shown.

LOCAL	Lat*	Long*	Morpho†	Genotyped‡	Full genotype§	♀	♂	Full dataset‖
Alcobaça	39.55	-8.98	2	2	2	0	2	4
Louriçal	40.00	-8.74	2	1	1	0	1	6
São Pedro de Alva	40.30	-8.18	10	2	2	1	1	13
Parada de Gonta	40.57	-8.01	6	2	2	1	1	9
Gouveia	40.49	-7.60	12	3	3	1	2	14
Santa Cruz da Trapa	40.77	-8.15	2	2	0	0		4
Rua	40.95	-7.58	1	1	1	0	1	5
Lugar de Anta	41.27	-7.64	2	2	2	1	1	5
Aveçãozinho	41.30	-7.88	10	3	3	2	1	12
Granja	41.69	-7.65	6	2	2	2	0	9
Paredes de Coura	41.91	-8.57	4	4	1	1	0	5
Lalin	42.66	-8.11	1	1	1	1	0	4
Santiago de Compostela	42.88	-8.54	4	0	0	0	0	5
Torre de Mens, Malpica	43.30	-8.86	5	0	0	0	0	6
Sobrado de Monxes	43.04	-8.03	1	1	0	0	0	5
Oza dos Rios	43.22	-8.19	3	1	1	0	1	3
Pontedeumes	43.40	-8.15	5	1	1	1	0	6
Outeiro de Rei	43.10	-7.62	5	4	3	3	0	5
O Íncio	42.66	-7.40	1	0	0	0	0	2
Fontsagrada, San Cristobal	43.12	-7.07	9	9	1	1	0	9
Ribadeon	43.54	-7.04	4	2	1	1	0	5
Pelorde	43.28	-6.84	3	1	1	1	0	3
Grandais	41.82	-6.82	2	2	2	1	1	5
Tormaleo	42.95	-6.74	6	4	2	1	1	6
Constantim	41.62	-6.28	1	1	1	1	0	1
Boal	43.43	-6.82	4	3	3	2	1	6
Manzanal del Puerto	42.59	-6.22	8	1	0	0	0	8
Fresno de la Valduerna	42.33	-5.97	6	2	2	2	0	9
Ribón	43.55	-6.36	2	0	0	0	0	3
Camposagrado	42.73	-5.73	1	0	0	0	0	2
Vallin	43.16	-5.96	3	2	1	1	0	4
Tanes	43.21	-5.39	3	3	2	1	1	5
Puebla de Lillo	43.01	-5.27	3	2	0	0	0	4
Villaviciosa	43.49	-5.42	5	2	2	1	1	6
Gañinas	42.48	-4.73	1	1	1	1	0	5
Valderredible	42.80	-4.14	1	1	1	1	0	5
San Martin de Huma	42.66	-4.03	2	2	1	1	0	5
Entrambasmestas	43.17	-3.90	4	4	4	4	0	5
Quintanalara	42.18	-3.53	1	1	1	0	1	5
Noceco	43.09	-3.48	3	3	2	2	0	4
Ungo	43.14	-3.26	4	4	3	3	0	5
Encima Angulo	43.04	-3.18	3	3	3	1	2	5
Bentretea	42.73	-3.46	1	1	1	1	0	5
Aostri de Losa	42.94	-3.10	4	4	4	3	1	5
Contreras	42.02	-3.41	4	4	4	3	1	5
Fresneda de la Sierra	42.32	-3.13	2	2	2	1	1	5
Markina-Xemein	43.26	-2.51	14	5	5	0	5	14
Iturmendi	42.89	-2.10	2	2	2	0	2	5
Arraras	43.03	-1.80	2	2	2	0	2	5
Lesaca	43.23	-1.70	4	4	2	0	2	5
Equisoain	42.70	-1.50	3	3	3	1	2	5
Ureta	42.95	-1.40	4	4	4	3	1	5
Usón	42.70	-1.20	1	1	0	0	0	3
Sierra del Cadí	42.29	1.61	1	0	0	0	0	1
Caralps	42.34	2.16	18	4	4	2	2	19

LOCAL	Lat*	Long*	Morpho†	Genotyped‡	Full genotype§	♀	♂	Full dataset‖
Tona	41.85	2.23	1	1	1	1	0	1
Centelles	41.79	2.22	2	1	1	0	1	2
El Brull	41.81	2.30	1	0	0	0	0	4
La castanya	41.78	2.34	1	1	1	1	0	1

* Latitude and longitude presented in decimal degrees; † - Number of samples used in the geometric morphometric analysis; ‡ - number of samples for which information on genotype was available; § - number of samples for which complete genotype was available. ♀ - number of females with complete genotype. ♂ - number of males with complete genotype; ‖ - number of samples used in the Geneland analysis that considered both genetic and morphometric data.

CHAPTER 4

A METHOD FOR GENETIC IDENTIFICATION OF IBERIAN RODENTS

Paper V - Barbosa S†, Paupério J†, Searle JB, Alves PC (2012). Genetic identification of Iberian Rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling. *Molecular Ecology Resources.* DOI: 10.1111/1755-0998.12024

† Joint first authors

PAPER V

Genetic identification of Iberian Rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling

Soraia Barbosa†, Joana Paupério†, Jeremy B. Searle and Paulo C. Alves

Molecular Ecology Resources (in press). DOI: 10.1111/1755-0998.12024.

† Joint first authors

GENETIC IDENTIFICATION OF IBERIAN RODENT SPECIES USING BOTH MITOCHONDRIAL AND NUCLEAR LOCI: APPLICATION TO NONINVASIVE SAMPLING

BARBOSA, S.[1,3,4]†, PAUPÉRIO, J.[1,2,3]†, SEARLE, J.B.[2,4], ALVES, P.C.[1,3,5]

ABSTRACT

Species identification through noninvasive sampling is increasingly used in animal conservation genetics, given that it obviates the need to handle free-living individuals. Noninvasive 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 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 noninvasive 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 *M. duodecimcostatus*. Overall, cyt-*b* showed higher resolution than IRBP, revealing a clear barcoding gap. To allow these markers to be applied to noninvasive 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 noninvasive samples, with application in studies of distribution, spatial ecology and population dynamics, and for conservation.

Keywords: Conservation genetics, cyt-*b*, DNA barcoding, interphotoreceptor retinoid-binding protein, noninvasive genetic sampling, rodent phylogeny, species ID

---

† joint first authors
[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, InBIO Laboratório Associado, Campus Agrário de Vairão, 4485-661, Vairão, Portugal;
[2]Department of Biology, University of York, Wentworth Way, York YO10 5DD, United Kingdom
[3]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal;
[4]Department of Ecology and Evolutionary Biology, E139 Corson Hall, Cornell University, Ithaca, NY 14853-2701, USA
[5] Wildlife Biology Program, College of Forestry and Conservation, University of Montana, 32 Campus Drive, Missoula MT 59801, MT, USA.

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 or near species (Table 1, see below), 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.* 2008). 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, makes monitoring and status assessment a difficult task (*e.g.* Álvarez-Castañeda and 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 noninvasive genetic sampling (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).

Noninvasive genetic sampling 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 noninvasive source of rodent DNA, found scattered throughout an individual's activity range (Beja-Pereira *et al.* 2009; Centeno-Cuadros and 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 noninvasive 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 noninvasive genetics through various sampling strategies, such as hair tubes or duct tape webs (Harris and Nicol 2010;

Henry *et al.* 2011; Pocock and 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 and Godoy 2010; Oliveira *et al.* 2010; Henry *et al.* 2011). Overall, noninvasive genetic sampling has great potential for rodent monitoring programs, including those focusing on geographic areas considered diversity hotspots, such as the Iberian Peninsula.

As 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 and Hewitt 2003; Melo-Ferreira *et al.* 2009, 2012). 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 southwestern Europe (Oliveira *et al.* 2010), 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 20 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 northeastern Iberia and elsewhere in southern Europe (Jaarola and Searle 2004; Paupério *et al.* 2012). Using both tissue and noninvasive 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 noninvasive genetic approach for faeces and bones from owl pellets, by designing shorter fragments of the selected markers. There are difficulties in obtaining amplifiable DNA by noninvasive genetic sampling due to low DNA quantity and quality, including DNA degradation, presence of PCR inhibitors and DNA contamination (Rådström *et al.* 2004; Waits and 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 noninvasive 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.

MATERIALS 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 and 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, CA,

USA). Samples were subsequently sequenced for both strands on a 3130*xl* Genetic Analyser Sequencer (Applied Biosystems/HITASHI). 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'.

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.

Family	Species	Tissue samples		Noninvasive 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]**

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 noninvasive 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.
†*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.

Sequences obtained for the IRBP gene were phased using the software PHASE as implemented in DNAsp 5.10 (Librado and 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. Because 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.* 2012), 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 and 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 and 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 noninvasive 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 (Gosàlbez and Noguera 1987; Blanco 1998). Mandibles and skulls were extracted using the DNeasy® Blood and Tissue Kit (Qiagen, Hilden, Germany) following the 'Purification of total DNA from compact animal bone' protocol with some adaptations. 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 (c.15 min) 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 6000 g 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 noninvasive extraction procedures were performed in a physically isolated room maintaining conditions to reduce risk of DNA contamination (Gilbert *et al.* 2005).

Amplification of noninvasive DNA samples

Shorter DNA fragments were selected from each gene to increase the amplification success in noninvasive 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 and 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.

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 and 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) (Fig. 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 and interspecific (within families) K2P divergences for cyt-*b* (Fig. 2A), which require a more detailed analysis (Fig. 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.

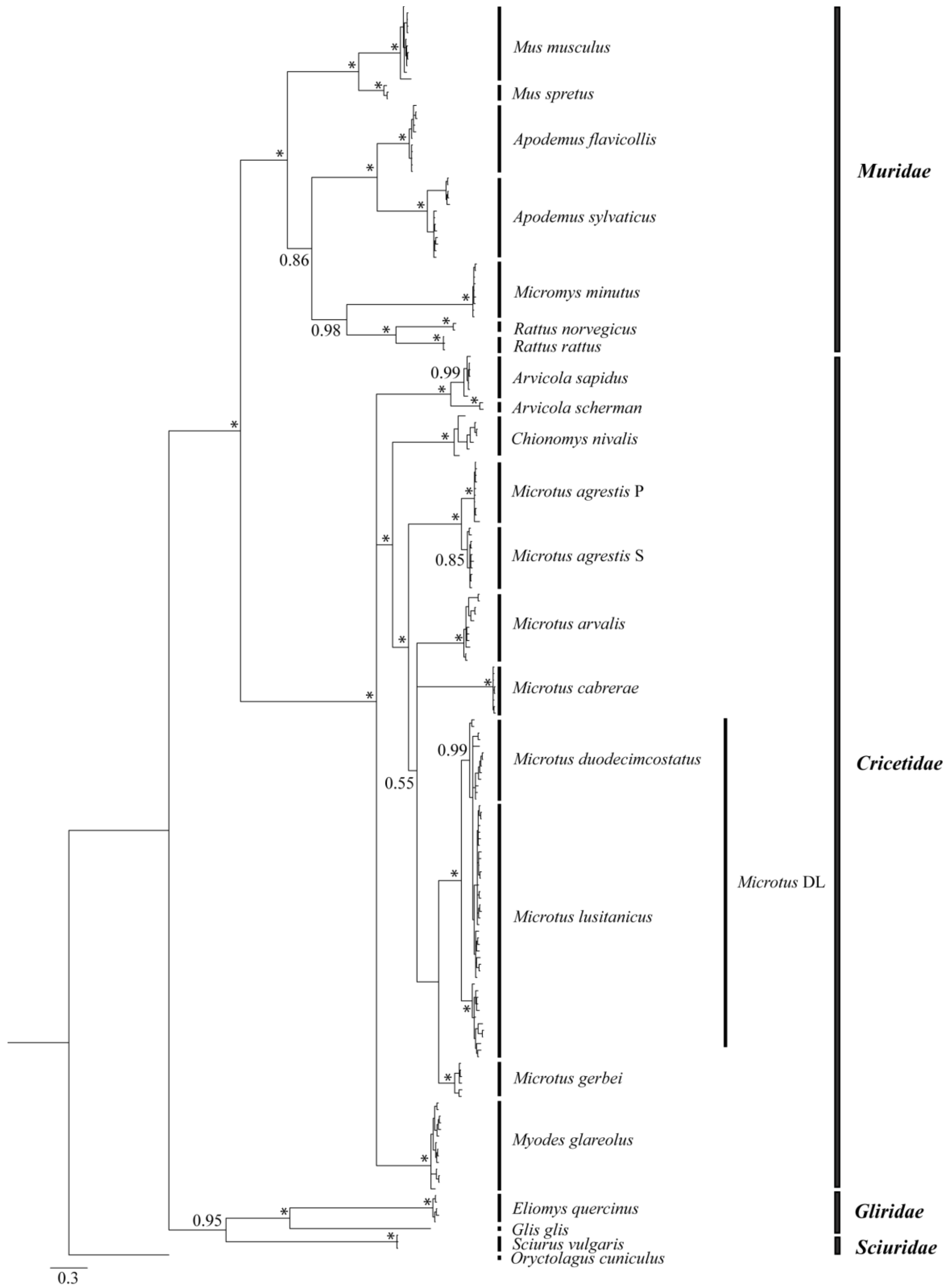


Fig. 1 Bayesian inference tree for the cyt-*b* long fragment showing the phylogenetic relationship of all 21 Iberian rodent taxa. Posterior probabilities of major nodes are indicated. Asterisks represent posterior probabilities of 1. Outgroup: *Oryctolagus cuniculus*.

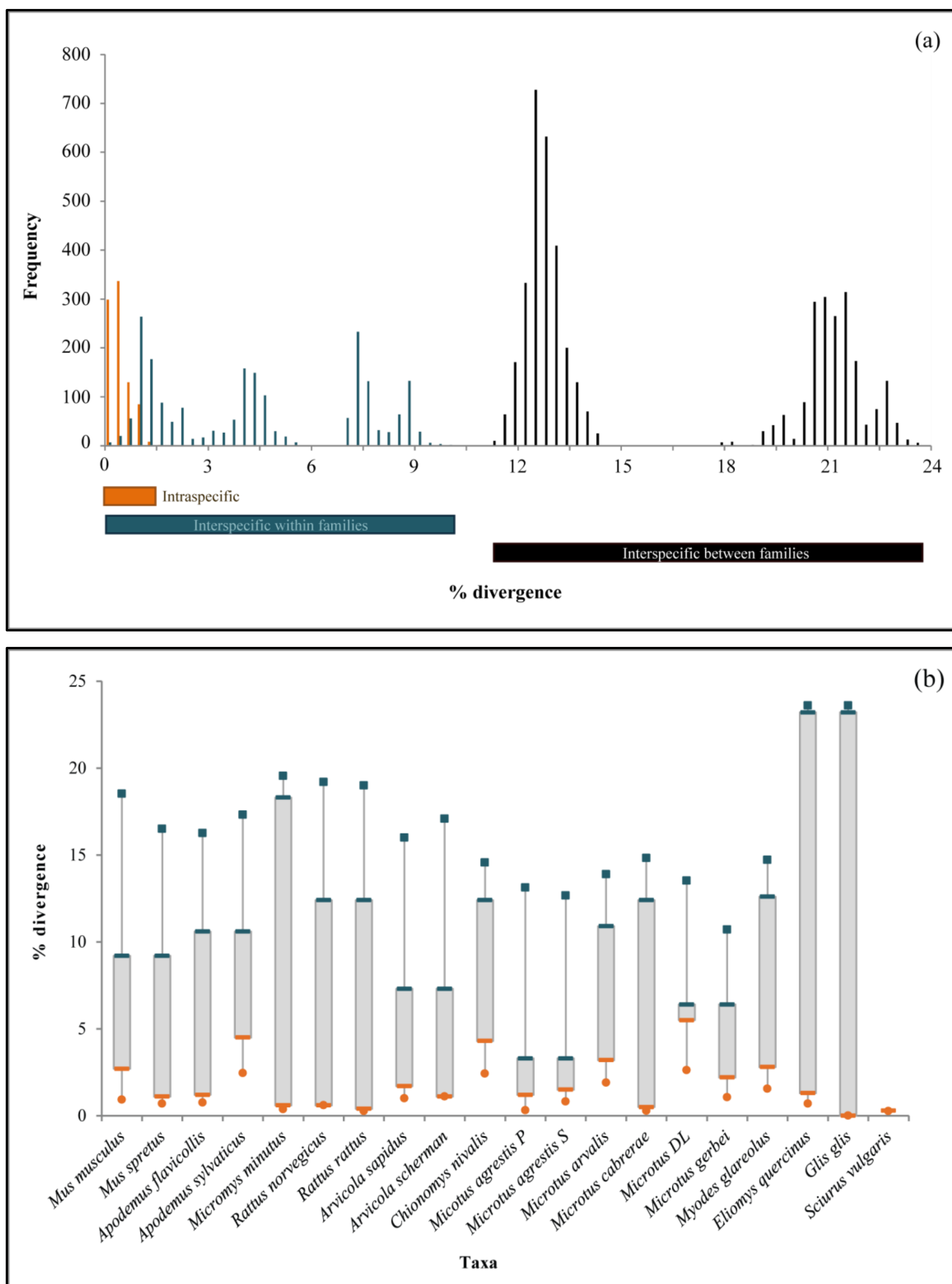


Fig. 2 (a) Histogram of K2P cyt-*b* divergence values (intraspecific, interspecific within families, interspecific among families), for the 21 Iberian rodent taxa. *Microtus duodecimcostatus* and *M. lusitanicus* were analysed as a single taxon, *Microtus* DL, due to lack of differentiation in the phylogenetic analysis. (b) Summary of pairwise divergences involving sequences of each species showing mean (pale dot) and maximum (associated dash) intraspecific divergences and mean (dark square) and minimum (associated 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.

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 (Fig. 3), however the phylogenetic relations among these families were different from those obtained with cyt-*b* (Fig. 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.

Levels of K2P divergence for the IRBP gene were generally lower than those observed for cyt-*b* (Fig. 4A). Despite that, a well-defined gap is observed between the interspecific within family and the interspecific between family 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 (Fig. 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*).

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.

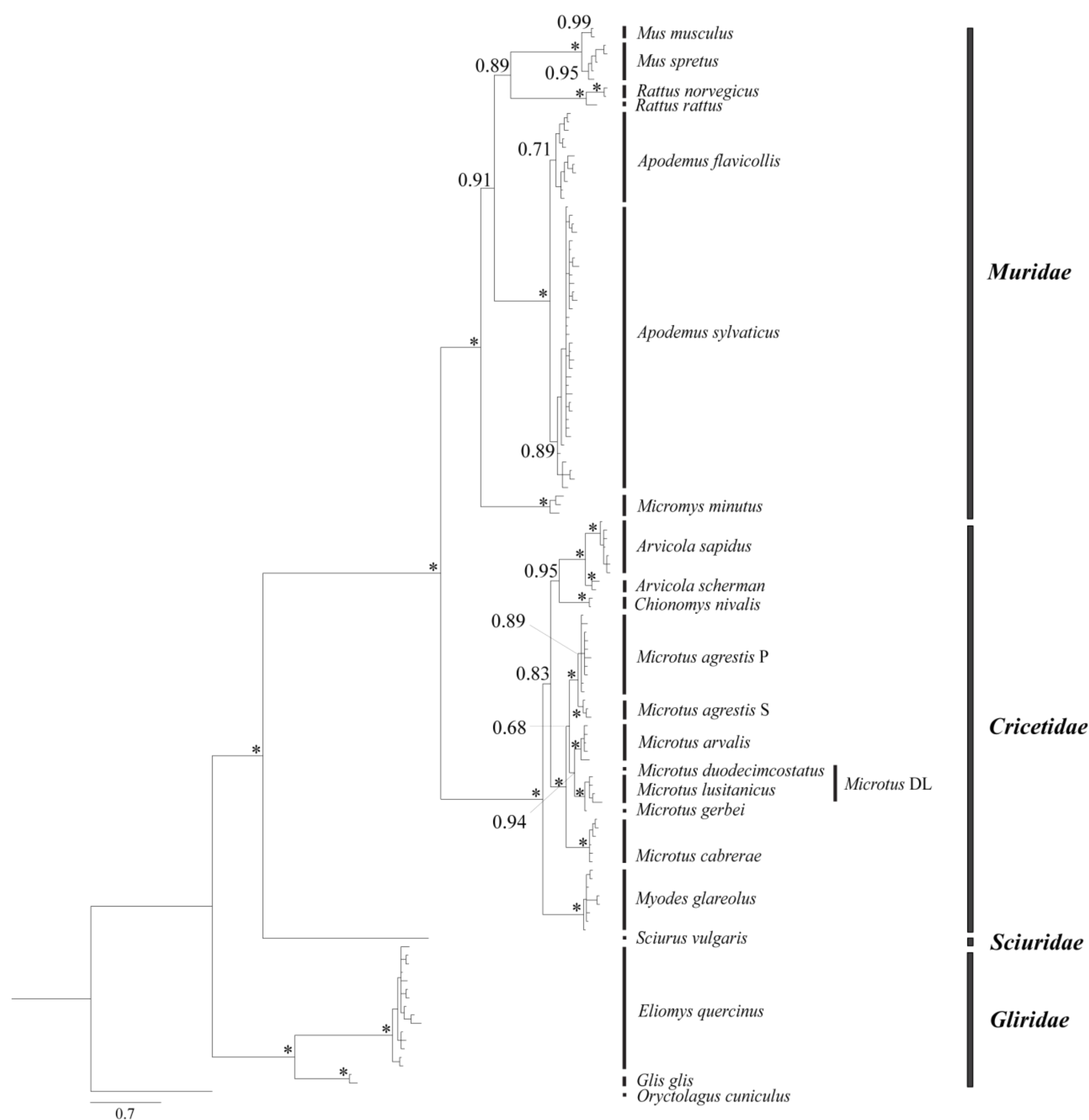


Fig. 3 Bayesian inference tree for the IRBP long fragment illustrating the phylogenetic relationship of the 21 Iberian rodent taxa. Posterior probabilities of major nodes are indicated. Asterisks represent posterior probabilities of 1. Outgroup: *Oryctolagus cuniculus*.

Amplification success for noninvasive 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 (Fig. 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.

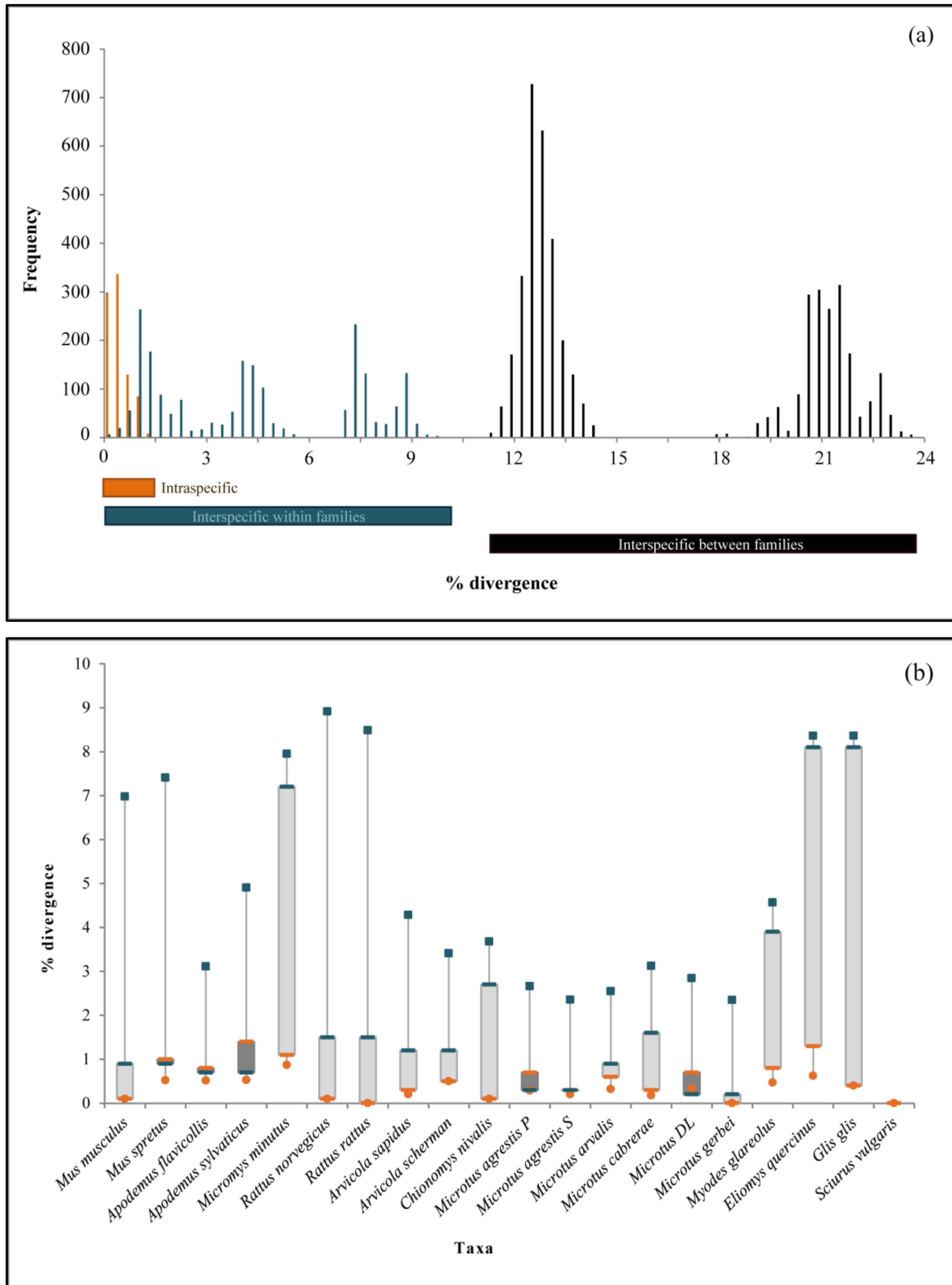


Fig. 4 a) Histogram of K2P IRBP divergence values (intraspecific, interspecific within families, interspecific among families), for the 21 Iberian rodent taxa. *Microtus duodecimcostatus* and *M. lusitanicus* were analysed as a single taxon, *Microtus* DL, due to lack of differentiation in the phylogenetic analysis. (b) Summary of pairwise divergences involving sequences of each species showing mean (pale dot) and maximum (associated dash) intraspecific divergences and mean (dark square) and minimum (associated 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.

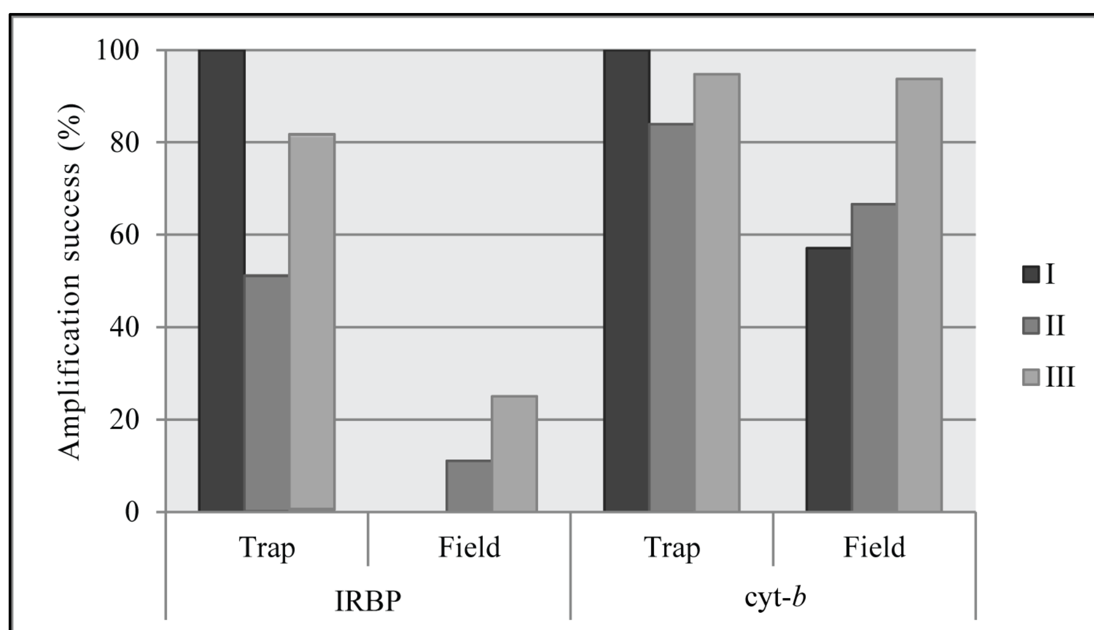


Fig. 5 Amplification success for the IRBP and the cyt-*b* gene using DNA from faecal material. For each gene, the data is divided into 'Trap' and 'Field' depending on whether the faecal samples were collected from traps or opportunistically in the field. The data are further subdivided to show the amplification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

Variability of short fragments

For the cyt-*b* gene short fragment, 369 sequences were analysed, comprising 239 long fragment sequences and 74 noninvasive 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 noninvasive 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).

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 and 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 (Melo-Ferreira *et al.* 2012). 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* (Figs. 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.* 2008), 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 hybridization (Bastos-Silveira *et al.* 2012, 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 (Figs 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* (Figs 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 (Fig. 4A), a common finding for nuclear exons (Zhang and 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 Fig. 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 and Dekker 2011; Mirol *et al.* 2000; Phillips and Dudı 2008; Triant and 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 noninvasive 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 and 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 and 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 and 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.

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 noninvasive samples (faeces and/or bones from owl pellets). In general, we had a high DNA extraction success for the noninvasive 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 noninvasive 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/105110/2011 research project) and by Programa Operacional Potencial Humano – Quadro de Referência Estratégico Nacional funds from the European Social funds and Portuguese Ministério da Educação e Ciência (FCT, SFRH/BD/27369/2006 PhD grant to J. Paupério). We thank the reviewers for valuable comments.

REFERENCES

Alasaad S, Soriguer RC, Jowers MJ, Marchal A, Romero I, Sánchez A (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, Branco M, Ferrand N, Suchentrunk F, Boursot P (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.

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

Asher RJ, Maree S, Bronner G, Bennett NC, Bloomer P, Czechowski P, Meyer M, Hofreiter M (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 SM, Monarca R, Mathias ML, Heckel G (2012) Deep mitochondrial introgression and hybridization among ecologically divergent vole species. *Molecular Ecology*. DOI: 10.1111/mec.12018.

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.

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, Zagarodnyuk I, Zima J, Juškaitis R, Meinig H, Kryštufek B (2008) *Eliomys quercinus*. In IUCN 2011. *IUCN Red List of Threatened Species*, version 2011.2. www.iucnredlist.org.

Bertolino S, Wauters L, Pizzul A, Molinari A, Lurz P, Tosi G (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, Walberg MW, Clayton DA (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, Pupko T, Debry RW, Huchon D (2009) Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evolutionary Biology*, **9**, 71.

Boratyński Z, Alves PC, Berto S, Koskela E, Mapes T, Melo-Ferreira, J (2011) Introgression of mitochondrial DNA among *Myodes* voles: consequences for energetics? *BMC Evolutionary Biology*, **11**, 355-367.

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, Pagès M, Corbisier C, Cosson JF, Herbreteau V (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.

Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, Schmitz J (2010) Rodent evolution: back to the root. *Molecular Biology and Evolution*, **27**, 1315-1326.

Danciger F, Kozak C, Nickerson J, Redmond M, Farber D (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, Hutterer R, Vogel P (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.

Galewski T, Tilak M, Sanchez S, Chevret P, Paradis E, Douzery EJP (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 HJ, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. *Trends in Ecology and Evolution*, **20**, 541-544.

Gojobori T, Li WH, 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, Lisi I, Tosi G, Wauters LA, Randi E (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.

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 B*, **270**, 96-99.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003b) Biological identifications through DNA barcodes. *Proceedings of the Royal Society 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 noninvasive 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.

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

Jaarola M, Martínková N, Gündüz I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedaski-Tsolis S, González-Esteban J, López-Fuster MJ, Kandaurov AS, Kefelioglu H, Mathias ML, Villate I, Searle JB

(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, Zuccon D, Kerr KCR, Stoeckle MY, Lifjeld JT (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, Haught G, Sauvajot RM, Wayne RK (1999) Estimating population size by genotyping faeces. *Proceedings of the Royal Society B*, **266**, 657-663.

Li WH, Wu CI, Luo CC (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.

Luo A, Zhang A, Ho SY, Xu W, Zhang Y, Shi W, Cameron SL, Zhu C (2011) Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. *BMC Genomics*, **12**, 84-97.

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.

Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P (2009) The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. *Molecular Ecology*, **18**, 2643-2658.

Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farelo L, Alves PC (2012) Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulation. *Systematic Biology*, **61**, 367–381.

Michalski F, Valdez FP, Norris D, Zieminski C, Kashivakura CK, Trinka CS, Smith HB, Vynne C, Wasser SK, Metzger JP, Eizirik E (2011) Successful carnivore identification with faecal DNA across a fragmented Amazonian landscape. *Molecular Ecology Resources*, **11**, 862-871.

Michaux JR, Libois R, Paradis E, Filippucci MG (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.

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, Dueser RD, Loftis D, Cockett NE, Culver N (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) Noninvasive 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

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.

Oliveira R, Castro D, Godinho R, Luikart G, Alves PC (2010) Species identification using a small nuclear gene fragment: application to sympatric wild carnivores from southwestern 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, Jaarola M, Alves PC, Searle JB (in press) Cryptic speciation in the field vole: a multilocus approach confirms highly divergent lineages in Europe. *Molecular Ecology*. DOI: 10.1111/mec.12024

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.

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

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

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

Song Y, Endepols S, Klemann N, Richter D, Matuschka FR, Shih CH, Nachman MW, Khon MH (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between Old World mice. *Current Biology*, **21**, 1296-1301.

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, Stecher G, Nei M, Kumar S (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 (2008) Evolutionary history of two allopatric *Terricola* species (Arvicolinae, Rodentia) from molecular, morphological, and palaeontological data. *Biological Journal of the Linnean Society*, **93**, 309-323.

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) Noninvasive 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, Searle JB, Kotlík P (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.

Zhang DX, 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

Author contributions box

This study is a part of SB's Master thesis supervised by PCA, and of JP's PhD thesis work that is being developed under the supervision of PCA and JBS. Genetic laboratory work and analyses were conducted at the University of York and CIBIO.

APPENDIX – 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 and 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 and Searle 2002	*														
		JX284253	Paupério *et al.* 2012	*														
		JX284254		*														
		JX284267		*														
		JX284268		*														
		JX284262		*														
		JX284259		*														
		JX284270		*														
		JX284257		*														
		JX284265		*														
		AY303160	Jaarola and Searle 2004	(*)														
Microtus agrestis S		AY167187	Jaarola and Searle 2002		*													
		AY167160			*													
		AY167161			*													
		AY167162			*													
		AY167163			*													
		AY167164			*													
		JX284274	Paupério *et al.* 2012		*													
		JX284273			*													
		JX284282				*												
		JX284283				*												
		AY303159	Jaarola and Searle 2004		(*)													
		AY303155			(*)													
		AY303156		(*)														
	MagSS1	§	This study		(1)													
Microtus arvalis		EU439459	Borkowska and 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	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 and 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
A. sylvaticus	Asyl31	JX457653		2									
	Asyl32	JX457654-1	This study	1									
	Asyl33	JX457654-2		1									
Micromys minutus	Mmin1	JX457656											4
	Mmin2	JX457655-2	This study										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		4(2)	(4)								
	Asap2	JX457661-2		1									
	Asap3	JX457662		6									
	Asap4	JX457668	This study	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			3(9)								
	Cniv2	JX457672	This study		3								
	CnivS1	JX457608-2			(1)								
Microtus agrestis P	MagP1	JX457674-1		7									
	MagP2	JX457674-2		1									
	MagP3	JX457675-2		3									
	MagP4	JX457676-2		1									
	MagP5	JX457677-2	This study	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			1	2							
	MagS2	JX457685	This study		2								
	MagS3	JX457684-2			1								
Microtus arvalis	Marv1	JX457686			2(2)	3							
	Marv2	JX457687-2				1							
	Marv3	JX457688-2	This study		2								
	Marv4	JX457689-1			1								
	Marv5	JX457689-2			1								
	MarvS1	JX457609			(4)								
Microtus cabrerae	Mcab1	JX457690		19(8)									
	Mcab2	JX457691-2	This study	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 sister species.

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).

Nucleotide Position

Species	14582	14585	14588	14589	14591	14592	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
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
Mus musculus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	R	.	.	.	.	.	.	.	M	.	.	.	.	.
Mus spretus	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	M	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	C	A	.	.
Apodemus flavicollis	.	.	T	.	A	T	.	.	.	A	Y	.	C	.	C	.	R	T	.	.	.	A	R	.	C	.	A	A	.	.	.	T	.	.	Y	A	C	A	.	A
Apodemus sylvaticus	.	.	T	Y	A	.	.	R	.	A	T	.	C	.	C	.	.	T	.	.	.	A	R	.	C	.	R	A	.	.	.	.	.	.	T	A	.	R	.	A
Micromys minutus	C	C	.	.	.	.	.	.	.	T	T	.	.	G	A	.	.	C	.	.	.	T	.	.	C	.	A	A	.	C	.	T	.	.	A	.	C	A	.	A
Rattus norvegicus	.	.	.	.	A	T	.	W	.	T	.	T	C	.	.	G	.	T	.	.	.	.	.	.	C	.	A	.	.	.	.	.	.	.	A	.	C	A	.	A
Rattus rattus	C	.	.	.	A	T	.	.	.	.	T	C	C	.	.	C	.	C	T	.	.	.	.	.	C	.	A	.	.	.	.	.	.	.	A	.	C	A	.	A
Arvicola sapidus	.	.	Y	.	A	.	.	R	.	.	.	C	C	.	C	C	.	.	T	.	.	T	.	.	C	.	A	A	.	.	T	.	.	R	.	T	C	T	.	A
Arvicola terrestris	.	.	T	.	.	.	.	.	.	.	.	C	C	.	.	C	.	.	.	.	.	T	.	.	C	.	.	A	.	.	C	.	.	.	.	T	C	T	.	A
Chionomys nivalis	Y	.	.	.	.	Y	.	.	.	.	.	.	C	.	Y	H	.	V	A	.	.	A	.	.	C	.	R	R	.	.	T	.	.	.	Y	T	C	Y	.	A
Microtus agrestis P	.	.	.	.	W	.	.	.	.	.	T	C	C	G	A	C	.	C	A	.	.	A	R	.	C	.	R	T	T	.	.	.	.	.	.	M	C	C	.	A
Microtus agrestis S	Y	.	.	Y	A	T	.	.	.	.	.	C	C	G	A	C	.	C	M	.	.	A	.	.	Y	.	R	R	T	.	.	.	.	.	.	.	C	C	.	A
Microtus arvalis	.	.	T	.	.	Y	.	.	R	A	.	C	C	.	Y	T	.	W	A	Y	.	A	R	R	.	.	R	Y	.	.	.	Y	.	.	.	.	C	C	.	A
Microtus cabrerae	.	.	.	.	A	.	.	.	.	A	.	.	.	.	C	.	.	.	A	.	T	A	.	.	C	G	T	.	.	.	.	T	.	.	T	A	C	C	.	A
Microtus DL	.	.	.	.	.	.	.	.	.	.	.	C	C	.	C	R	.	B	A	.	R	R	R	R	C	.	V	.	.	.	.	.	.	R	H	Y	C	T	.	R
Microtus gerbei	.	.	.	.	Y	.	.	.	.	.	.	Y	C	.	C	.	.	C	R	.	.	A	.	.	C	.	Y	G	.	.	.	T	.	.	.	.	C	Y	R	A
Myodes glareolus	Y	.	Y	.	A	Y	.	.	.	Y	Y	Y	Y	.	Y	M	.	.	.	.	.	A	.	.	C	R	V	R	.	.	.	B	.	.	Y	Y	C	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
Glis glis	.	T	T	.	T	T	.	T	.	.	.	Y	C	.	C	T	.	T	.	T	.	A	.	.	.	.	C	.	.	C	.	.	.	.	T	T	C	A	.	.
Sciurus vulgaris	C	.	.	.	.	T	G	T	.	T	T	T	.	.	.	.	.	C	.	.	.	T	.	.	C	.	C	A	.	.	.	T	.	.	A	T	C	A	.	A

Species	14675	14678	14681	14684	14687	14690	14693	14696	14697	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.	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	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	Y	.	Y	.	.	.
Mus spretus	.	.	.	C	.	.	.	.	.	G	.	C	C	.	.	.	A	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	T	.	.
Apodemus flavicollis	T	Y	Y	C	.	.	T	.	.	.	.	C	C	.	T	R	Y	.	.	.	.	T	.	.	.	H	C	Y	.	Y	.	A	T	.	H
Apodemus sylvaticus	T	T	.	Y	T	.	T	.	C	T	.	.	Y	.	T	.	T	.	.	.	.	T	.	.	.	T	.	.	.	.	T	G	Y	.	.
Micromys minutus	T	.	.	A	.	T	.	T	C	.	C	C	C	.	.	.	R	.	.	.	.	.	.	R	.	T	.	T	.	T	.	T	T	.	W
Rattus norvegicus	C	.	.	A	.	.	Y	.	C	C	.	C	.	.	.	.	C	.	.	T	.	.	.	.	.	T	A	T	.	T	.	T	.	.	.
Rattus rattus	T	T	.	A	.	.	.	.	C	C	.	C	C	.	.	.	C	.	.	T	.	.	.	.	.	T	A	T	.	.	.	.	.	.	.
Arvicola sapidus	.	.	.	A	.	T	T	.	C	C	.	.	Y	.	T	A	C	T	.	C	.	T	C	T	.	A	.	.	.	.	.	A	.	.	T
Arvicola terrestris	.	.	T	A	.	T	T	T	C	T	.	.	C	.	.	.	C	T	.	T	.	T	C	T	.	A	.	.	T	A	T	A	.	.	A
Chionomys nivalis	.	.	Y	C	T	.	.	T	.	R	T	Y	C	.	.	A	C	.	.	T	.	T	T	T	.	A	A	T	.	R	.	A	Y	.	A
Microtus agrestis P	.	.	.	C	.	.	.	.	C	.	C	.	C	.	Y	A	C	T	.	.	.	T	T	C	.	A	A	T	T	A	.	A	T	.	A
Microtus agrestis S	.	T	.	C	.	.	.	.	C	W	M	.	C	.	T	A	C	Y	.	.	.	T	T	C	.	A	A	T	Y	A	.	A	Y	.	A
Microtus arvalis	.	.	Y	C	.	.	.	.	C	.	C	.	Y	.	T	A	C	Y	.	H	.	T	.	T	.	A	Y	Y	.	T	T	A	Y	.	T
Microtus cabrerae	.	T	.	C	.	.	.	.	C	R	.	.	C	.	.	A	T	.	.	C	.	T	.	T	.	A	C	.	.	.	.	A	.	.	T
Microtus DL	.	T	.	Y	.	.	.	.	C	R	C	.	C	.	Y	A	C	.	.	T	.	T	.	T	.	A	Y	.	.	D	.	R	Y	T	A
Microtus gerbei	.	.	.	C	.	.	.	.	C	.	C	.	C	.	T	A	A	.	.	T	.	T	.	C	.	A	C	T	.	A	.	A	.	T	A
Myodes glareolus	.	.	.	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	.	.	.	A	.	.	Y	T	C	C	.	C	C	.	Y	A	T	.	T	.	.	T	.	R	Y	A	V	.	.	.	T	A	Y	Y	A
Glis glis	.	T	Y	A	T	.	T	T	C	.	C	C	.	.	T	.	A	.	.	.	.	T	.	.	Y	A	C	.	.	T	.	A	.	.	T
Sciurus vulgaris	.	.	.	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 (positions 825–898)

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

Nucleotide Position (positions 900–969)

Species	900	903	909	912	913	915	918	921	931	933	936	939	940	942	945	951	952	954	957	960	966	969
Mus musculus Ref.Seq.	C	C	C	A	G	G	T	G	C	G	A	T	C	A	T	T	G	C	G	A	A	C
Mus musculus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mus spretus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Apodemus flavicollis	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	Y	.	.	.	R	.	.
Apodemus sylvaticus	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	C	.	.	.	G	.	.
Micromys minutus	.	.	.	.	.	A	.	T	.	.	.	C	.	.	.	.	.	.	.	G	.	.
Rattus norvegicus	.	.	T	.	.	.	.	.	.	.	G	C	.	.	.	C	.	.	A	.	.	.
Rattus rattus	.	.	T	.	.	.	.	.	.	.	G	C	.	.	.	A	.	.	A	G	.	.
Arvicola sapidus	.	.	.	.	C	.	.	.	.	.	G	C	T	G	A	.	.	.	.	G	.	.
Arvicola terrestris	.	.	.	.	C	.	.	.	.	.	G	C	T	G	A	.	.	.	.	G	.	.
Chionomys nivalis	.	.	.	.	S	.	A	.	.	.	G	C	T	G	A	.	.	T	.	.	.	.
Microtus agrestis P	.	.	.	.	C	.	C	.	.	.	.	C	T	G	A	.	.	T	.	G	.	.
Microtus agrestis S	Y	.	.	.	C	.	C	.	.	.	.	C	T	G	A	.	.	T	.	R	.	.
Microtus arvalis	.	Y	.	.	C	.	C	.	.	.	.	C	T	G	A	.	.	T	.	G	.	.
Microtus cabrerae	.	.	.	.	C	.	C	.	.	.	.	C	T	G	A	.	.	.	.	G	.	.
Microtus DL	.	T	.	.	C	.	C	.	.	C	.	C	T	G	A	.	.	T	.	G	.	.
Microtus gerbei	.	T	.	.	C	.	C	.	.	C	.	C	T	G	A	.	.	T	.	G	.	.
Myodes glareolus	.	.	.	.	.	.	C	.	.	.	T	C	T	G	A	.	.	Y	.	G	.	.
Eliomys quercinus	T	.	.	C	.	.	C	.	.	.	G	G	.	G	C	C	A	Y	.	G	.	.
Glis glis	.	.	.	C	.	.	C	.	.	.	G	.	.	G	C	C	A	.	.	.	G	T
Sciurus vulgaris	T	.	.	T	.	A	.	.	T	.	.	C	.	G	C	C	A	.	.	G	.	.

Nucleotide Position (positions 972–1029)

Species	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.	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	.	A	.	.	C	.	A	.	.	A	.	.	.	.	.	.	.	G	.	.	.	.
Apodemus sylvaticus	.	A	Y	.	C	.	A	.	.	A	.	.	R	.	.	.	.	G	.	.	.	.
Micromys minutus	.	A	.	.	C	.	.	.	T	A	.	.	.	.	.	.	.	.	.	G	.	T
Rattus norvegicus	C	G	.	.	C	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.
Rattus rattus	C	G	.	.	C	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.
Arvicola sapidus	.	A	.	.	C	C	.	.	T	A	A	.	.	A	.	.	.	.	.	.	T	.
Arvicola terrestris	.	A	.	.	C	C	.	.	T	A	A	.	.	A	.	.	.	.	.	.	T	.
Chionomys nivalis	.	A	.	.	C	C	.	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Microtus agrestis P	.	A	.	.	C	C	.	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Microtus agrestis S	.	A	.	.	C	C	.	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Microtus arvalis	.	A	.	.	C	Y	.	.	T	A	A	R	.	.	.	.	.	.	.	.	T	.
Microtus cabrerae	.	A	.	.	C	C	.	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Microtus DL	.	A	.	.	C	C	R	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Microtus gerbei	.	A	.	.	C	C	R	.	T	A	A	.	.	.	.	.	.	.	.	.	T	.
Myodes glareolus	.	A	.	.	C	C	.	.	T	A	A	G	.	A	.	.	.	.	.	.	T	.
Eliomys quercinus	C	.	.	C	C	.	.	R	.	A	A	T	A	.	G	.	G	G	A	.	.	.
Glis glis	.	T	.	C	C	.	.	.	.	G	A	Y	.	.	G	.	G	G	A	A	.	.
Sciurus vulgaris	.	A	T	.	C	.	.	A	.	A	C	T	A	.	.	T	G	G	A	.	.	.

REFERENCES

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell*, **26**, 167–180.

Blanga-Kanfi S, Miranda H, Penn O, Pupko T, DeBry RW, Huchón D (2009) Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evolutionary Biology*, **9**, 71-82.

Bonhomme F, Orth A, Cucchi T, Rajabi Maham H, Catalan J, Boursot P, Auffray JC, Britton-Davidian J (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.

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, Filippucci MG, Amori G (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, 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.

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.

Colangelo P, Aloise G, Franchini P, Annesi F, Amori G (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, Redmond T, Farber D (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, Hutterer R, Vogel P (2009) False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene. *Molecular Phylogenetics and Evolution*, **50**, 633–641.

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, Chevret P, Paradis E, Douzery EJP (2006) The evolutionary radiation of Arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. *BMC Evolutionary Biology*, **6**, 80–97.

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.

Hurner H, Kryštufek B, Sarà M Ribas A, Ruch T, Sommer R, Ivashkina V, Michaux JR (2010) Mitochondrial phylogeography of the edible dormouse (*Glis glis*) in the western Palearctic region. *Journal of Mammalogy*, **91**, 233–242.

Jaarola M, Martínková N, Gündüz I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedaski-Tsolis S, González-Esteban J, López-Fuster MJ, Kandaurov AS, Kefelioglu H, Mathias ML, Villate I, Searle JB (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.

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*, **91**, 233–242.

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.

Malé PJG, Martin J-F, Galan M, Deffontaine V, Bryja J, Cosson JF, Michaux J, Charbonnel N (2012) Discongruence of *Mhc* and cytochrome *b* phylogeographical patterns in *Myodes glareolus* (Rodentia: Cricetidae). *Biological Journal of the Linnean Society*, 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.

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

Michaux JR, Chevret P, Filippucci MG, 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 MG (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, Nieberding C, Libois R (2003) Mitochondrial phylogeography of the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Molecular Ecology*, **12**, 685–697.

Nilsson MA, Gullberg A, Spotorno AE, Arnason U, Janke A (2003) Radiation of extant marsupials after the K/T boundary: evidence from complete mitochondrial genomes. *Journal of Molecular Evolution*, **57**, S3–S12.

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.

Nunome M, Yasuda SP, Sato JJ, Vogel P, Suzuki H (2007) Phylogenetic relationships and divergence times among dormice (Rodentia, Gliridae) based on three nuclear genes. *Zoologica Scripta*, **36**, 537–546.

Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB (2012) Cryptic speciation in the field vole: a multilocus approach confirms highly divergent lineages in Europe. *Molecular Ecology*, DOI: 10.1111/mec.12024.

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.

Reyes A, Gissi C, Pesole G, Catzeflis FM, Saccone C (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.

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.

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.

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 JP (2008b) New insight into the colonization processes of common voles: inferences from molecular and fossil evidence. *PLoS ONE*, **3**, 1–10.

Wójcik JM, Kawałko A, Marková S, Searle JB, Kotlík P (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, Han SH, Lin LK, Suzuki H (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.

CHAPTER 5

GENERAL DISCUSSION

5.1 THREE CRYPTIC LINEAGES OF THE FIELD VOLE IN EURASIA

The detection and delimitation of cryptic or nascent divergent lineages provide a major contribution to the understanding of speciation and its underlying processes (Sites and Marshall 2003). In addition, such detection has important consequences in biodiversity assessment and conservation planning (Bickford *et al.* 2007, Pfenninger and Schwenk 2007). In fact, many of these cryptic or nascent evolutionary lineages are designated as Evolutionarily Significant Units (ESU, Moritz 1994), and may later be even considered as different species.

The molecular study presented on chapter 2 (paper I), which relies on a multilocus dataset obtained from 160 field voles distributed throughout the species range, confirmed the existence of three highly divergent evolutionary units (Northern, Southern and Portuguese) with neighbouring geographic distributions. The Northern lineage has a very wide range, occurring through northern and central Europe and into Eurasia, going from the Alps to Fennoscandia and from Britain to eastern Siberia. The Southern lineage occupies a narrow area in Southern Europe, which includes northeastern Iberia, much of France and extends as far as Hungary through the north of Italy. The Portuguese lineage has a restricted distribution limited to western Iberia (ranging from central Portugal to northwestern Spain).

The results obtained in paper I (chapter 2) show that the seven loci analysed, which have different modes of inheritance, are generally consistent in the support of the three lineages. The three lineages show in general high levels of differentiation, though they diverge more substantially for mtDNA than for the nuclear loci. On the mitochondrial gene, cyt*b*, a divergence from 3 to 6% was estimated, while on the nuclear loci, the divergence between the lineages ranged from 0.3% (for RAG1 between the Southern and Northern lineage) to 1.9% (for DBX5 between the Portuguese and the Southern lineage). Although there is congruence in the distinction of the lineages, there are differences in the topology of the individual gene trees. However, the application of recently developed coalescent methods that model the membership of the individuals to the evolutionary lineages detected (Ence and Carstens 2011) showed that the fully resolved tree had an estimated probability of one. These results provide overwhelming support for considering the Northern, Southern and Portuguese lineages as different evolutionary units.

Despite of these high genetic differentiation observed, these three lineages were not previously known to differ in morphology, being in that sense 'cryptic'. Moreover the previous studies on chromosome variation in field voles had detected little variation in the structure of their giant sex chromosomes and no variation in the autosomes. In fact, Fredga and Jaarola (1997) that performed an extensive review of sex chromosome variation in field voles over central and northern Europe only detected variability in the structure of the Y chromosome, which has a substantial short arm ('the Lund Y') in southern Sweden. However, all studies that have focused on the variability of the field vole' karyotype have largely been restricted to the range of the Northern lineage (Chapter 2, paper II, Fig. 1). From the data presented by Fredga and Jaarola (1997) it seems that only one of the localities sampled in the past was most likely from within the range of the Southern lineage. Yet, no published information was available from northwestern Iberia (Portuguese lineage).

Our analyses of the Portuguese field vole karyotype (Chapter 2, paper II) showed that they possess giant sex chromosomes, but revealed the existence of subtle variation in sex chromosome morphology. The male specimen karyotyped had a Y chromosome with a distinct short arm, but with half the length of that on the Lund Y. The short arm of the Lund Y represents about 17.8% of the total length of the Y chromosome (Fredga and Jaarola 1997), while the Y chromosome from the Portuguese male had a short arm that comprised about 8.3% of the total length (Chapter 2, paper II, Fig. 3). In addition, the total lengths of the X and Y chromosomes were somewhat larger than the average size reported in the previous studies of field voles (Chapter 2, paper II, Fig. 2). Even though these results are based in only two specimens (one male and one female) and more individuals need to be examined to determine the 'typical' structure of the giant sex chromosomes of the Portuguese lineage, they suggest that there are some chromosome differences between, at least, the Portuguese and the Northern lineage. The comparison with the karyotype of the Southern lineage is not possible, since the data available for this lineage is very limited. Considering that the locality that was examined in Fredga and Jaarola (1997) is fairly close to what is thought to be distributional limits of the Northern and Southern lineage, that some introgression was detected in that area (Beysard *et al.* 2011) and that there is no genetic information on the karyotyped individuals, some uncertainty about the lack of variability between the Southern and Northern lineage arises. Therefore, a more extensive analysis of the karyotypic variability in the Southern lineage, coupled with genetic analysis for lineage assignment would be very valuable.

In a more detailed analysis of the lineages from the Iberian Peninsula (Portuguese and Southern, Chapter 3, paper IV), some morphometric differentiation

was detected. Geometric morphometrics analysis of the mandible, considered as a powerful tool for the identification of patterns of morphological variation below the species level (Loy 1996; Zelditch *et al.* 2004), revealed subtle differences in mandible shape between the lineages, while maintaining an identical pattern of size (no variability in size was observed). These results indicate that there is already some slight morphometric differentiation congruent with the molecular data, at least between the two Iberian lineages. Nevertheless, the detected differences are not sufficient for the designation of diagnostic morphological features for these two groups. Situations like this, where fine-scale morphometric analysis revealed subtle differences congruent with proposed taxonomic assignments in 'cryptic species' have already been reported, for example, in studies of 'cryptic bat species' (*Myotis*: Evin *et al.* 2008; *Pipistrellus*: Sztencel-Jabłonka *et al.* 2009) or rodents (Gündüz *et al.* 2007). Although our study concerns only two lineages (Portuguese and Southern) occurring in the Iberian Peninsula, the comparison of the mandible shape of the three lineages of this species complex would be desirable since we believe that it may also reveal morphometric differentiation consistent with the molecular data.

5.2 THE EVOLUTIONARY HISTORY OF *MICROTUS AGRESTIS SENSU LATO*

The species tree inference analyses (Chapter 2, paper I) performed with the multilocus dataset provided remarkable insights into the evolutionary history of the Eurasian field vole. The results obtained were concordant across methods (maximum likelihood and Bayesian inference analysis) providing congruent phylogenies of the species. The first proposed event is the separation of the Portuguese lineage from a combined Northern/Southern lineage, which represents a different branching order than the previously estimated with mitochondrial DNA analysis (Jaarola and Searle 2002, 2004). In this earlier study, based on cyt*b* analysis, the Portuguese lineage appeared to be closer to the Southern lineage, with the separation of the Northern and Southern/Portuguese cluster being the first estimated split. In fact, looking into the cyt*b* gene tree that we generate (Chapter 2, paper I, Fig. 2), we can observe the same phylogeny. However, when taking into account all genealogies (Chapter 2, paper I, Fig. 4) some differences in the topology of the individual gene trees are identified as well as in the support of the lineages within them. As it is well recognised, retention and incomplete lineage sorting of ancestral polymorphism can cause differences between

gene trees and the estimated species trees, particularly in the case of recently diverged species or populations (Avise *et al.* 1983; Tajima 1983; Pamilo & Nei 1988; Maddison 1997).

The divergence events that led to the three recognised lineages were estimated to have occurred quite recently, during the last glacial period and within a short time of each other. The first estimated split, between the Portuguese and Northern/Southern cluster, was inferred to have occurred at ca. 70 000 years BP (± 30 000 years BP), having most reasonably taken place in the vicinity of the Iberian Peninsula. The Northern and Southern lineage separated somewhat latter, at around the Last Glacial Maximum (LGM; estimated at ca. 18 500 years BP ± 6 000 years BP).

The divergence times of the Portuguese lineage interestingly seem to fit the maximum glacial extent in northern Iberia during the Weichselian, estimated from glacial records and lacustrine sediments at around 50 000 to 80 000 years BP (Lewis *et al.* 2009; García-Ruíz *et al.* 2010; Moreno *et al.* 2012). Moreover, both fossil records (Póvoas *et al.*, 1992; Moreno-García and Pimenta, 2002) and ice age species distribution modelling (Fløjgaard *et al.*, 2009) demonstrated that field voles could have persisted in the northwest of the Iberian Peninsula during the LGM. This evidence suggests that the Portuguese lineage probably remained endemic to western Iberia.

Periglacial conditions, in which temperate species like the field vole would not survive, prevailed in the ice-free parts of northwestern Europe during the LGM. Continuous or discontinuous permafrost is also thought to have prevailed in the high altitude areas of northern, central and eastern areas of France (Rensen & Vandenberghe, 2003). These very harsh conditions in France together with the glaciers in the Alps, could have create a barrier splitting the Northern and Southern field vole lineages. Nonetheless, fossils dated to the LGM indicate that the Southern lineage refugium might be located in southern France, northern Italy and in northeastern Spain (The Stage Three Project, http://www.esc.cam.ac.uk/research/research-groups/oistage3). The Northern lineage refugium during the LGM was probably located further to the east (Jaarola and Searle 2002 suggest the Carpathians).

After the split, the historical demography of the three lineages is quite disparate. A recent study by Herman and Searle (2011) described in detail the post glacial history of the Northern lineage, based on mtDNA analysis. The results shown in paper I (Chapter 2) based on a multilocus dataset are similar to those obtained with mtDNA alone, showing a substantial expansion for the Northern lineage that apparently followed the Younger Dryas cold period 12 000 years ago and carried on through the temperate conditions that continued after that (Chapter 2, paper I, Fig. 6a). In contrast, the demographic history of the Southern and Portuguese lineages seems to reflect

constraints on their distributional ranges. The results for the Southern lineage suggest a somewhat limited expansion, probably reflecting the narrow range occupied in southern Europe. The Portuguese lineage is limited to a restricted area in northwestern Iberia, with the demographic analysis showing very limited signs of expansion (Chapter 2, paper I, Table 4, Fig. 6c; Chapter 3, paper III).

5.3 MULTIPLE EVIDENCE SUGGEST THE REVISION OF SPECIES STATUS

Overall, multiple evidence gathered during the present study suggest that the field vole, *Microtus agrestis*, is indeed a complex of three distinct evolutionary units, genetically distinct (in all genomic compartments) and that also seem to show some karyotypic variability and some subtle morphological differentiation. However their distributions do not closely match those of any currently recognised subspecies.

Several taxa have been described that were latter synonymised with *Microtus agrestis* (38 according to Wilson and Reeder, 2005), but currently only a few of these are recognised as valid subspecies (Krapp and Niethammer 1982). In 1865, Bocage described *Arvicola rozianus*, based on a specimen from a locality in central Portugal. This species was later synonymised with *M. agrestis* by Lataste (1983) and afterwards considered as a subspecies, *M. agrestis rozianus*, with a distribution that resembles that of the Portuguese lineage (Trouessart 1910; Miller 1912; Cabrera 1914). However, 50 years later, Niethammer (1964, 1970) disregarded it as a valid subspecies and considered it to be a part of the more widespread *M. agrestis bailoni* (de Sélys-Longchamps, 1841). In fact, a study by Almaça (1993) that included the analysis of variation in pelage colour, morphometric traits of the skull and mandible and dental traits of three of the described taxa from southern Europe (*M. a. rozianus*, *M. a. bailoni* and *M. a. levernedii* Crespon, 1844) found little support for their designation as subspecies, since the detected variability was attribute to clinal variation between highland and lowland populations. However, as it is shown in Chapter 3 (paper IV) the application of highly powerful techniques of analysis of morphometric variation revealed that some morphometric variability may be attributed to the differentiation of the three lineages.

Genetically these three lineages are at the borderline between being separate populations or even species. For cyt*b*, the molecular marker for which there is the most comparative data available, the observed divergences vary from 3 to 6% which are

within the range of both intra- and interspecific divergences within the genus *Microtus* (Conroy and Cook 2000; Jaarola *et al.* 2004). The sibling species pairs *M. duodecimcostatus / M. lusitanicus*, *M. arvalis / M. rossiaemeridionalis* [*epiroticus*] and *M. bavaricus* / *M. liechtensteini* have their species status supported by independent data (*e.g.* fertility, karyotypes, morphology) and exhibit cyt*b* divergences of 2 - 8% (Jaarola *et al.* 2004; Martínková *et al.* 2007). However, there are also species such as *M. arvalis* and *M. oeconomus* that show high cyt*b* variability between clades (1.0 - 4.3%; Brunhoff *et al.* 2003; Fink *et al.* 2004; Heckel *et al.* 2005). There are also instances of other small mammal species that exhibit high degrees of mitochondrial and nuclear differentiation similar to those detected within the field vole (Serizawa *et al.*, 2000; Brändli *et al.*, 2005; Yannic *et al.*, 2008; Geraldes *et al*, 2008). Brändli *et al.* (2005) recently supported the taxonomic separation of two lineages of *Crocidura russula*, which show similar values of divergences in mtDNA and the sex chromosomes. Moreover, the *Sorex araneus* species complex, which includes four recognised species (*S. araneus s.s.*, *S. coronatus*, *S. granarius*, *S. antinorri*), presents a striking similarity with *M. agrestis*, in both the degree of genetic divergence and its European distribution (Yannic *et al.*, 2008).

This high level of genetic differentiation, together with the karyotypic and morphological evidence found in our study, would be sufficient for some mammalogists to consider that these three lineages should be elevated to species status (following for example the 'Genetic Species Concept'; Baker and Bradley 2006). However, others rely on a species definition that adheres more closely to Mayr's (1963) Biological Species Concept (BSC), considering forms to be separate species if there is absence of gene flow between them. Recent studies on the areas where the Northern and Southern lineages come into contact, revealed the existence of interbreeding in a narrow contact zone, but with little genetic interchange between them (Hellborg *et al.* 2005; Beysard *et al.* 2012). Actually, Beysard *et al.* (2012) detected a small number of hybrids, with higher introgression at mitochondrial than nuclear loci and without any sign of introgression of the Y chromosome, which suggested that at least some post-mating isolation mechanisms was already in place. Our detailed study of the Iberian populations of field vole through noninvasive genetic sampling (Chapter3, paper III) also revealed the existence of barriers to gene flow between the Southern and Portuguese lineages. Although a wide area with hybrids was detected, a small number of hybrids as well as a strong selection against hybrid males were observed.

Even though reproductive isolation is clearly important in speciation, there are several examples of taxa considered as valid species that do hybridise (Mallet 2008; Melo-Ferreira *et al.* 2012; Bastos-Silveira *et al.* 2012). Therefore, if one would follow a

more relaxed BSC, as do, at present, most evolutionary biologists (Mallet 2008; Hendry 2009), and consider that different groups should be different species if they could maintain their genetic integrity in nature (if they could remain distinct), one would be inclined, in face of the results here obtained, to consider these three field voles lineages as different species. If this would be the case, care should be given to the taxonomic designations of the three lineages. There are clear priorities in these designations. The Northern lineage clearly represents the field vole *M. agrestis* first described by Linnaeus in Uppsala in 1971. The Portuguese lineage, relates, as it was mentioned above, to the species described by Bocage, *Arvicola rozianus*, in 1865 based on a specimen from central Portugal. Finally, the Southern lineage should be designated as *M. levernedii*, which is the oldest available. *M. levernedii* was first described by Crespon (1844) from a locality in Southern France.

Nevertheless, even if the BSC would be followed in a more strict way, rejecting the species status, the Northern, Southern and Portuguese lineages should undoubtedly be designated as ESUs in the sense of Moritz (1994). Consequently they should be recognised as such for conservation purposes. The Northern lineage has a very wide range and very large population sizes are currently observed in northern Europe, especially in Britain and Scandinavia (our data on chapter 2, paper I; Harris *et al.* 1995; Myllymäki *et al.* 1977). However, the Southern and Portuguese lineage have much smaller geographical ranges and correspondently lower effective population sizes as revealed by the demographic analysis presented in Chapter 2 (paper I). In particular, the Portuguese lineage has a very restricted range (northwestern Iberia) and seems to have low population sizes, as revealed by the difficulties in collecting samples within its range, specifically in Galicia, where we were unable to collect samples by live trapping, despite the substantial sampling effort conducted. Considering that, like other Iberian endemic species, this lineage is liable to extinction or severe reduction in future distribution under climate change scenarios (*e.g.* Carvalho *et al.* 2010), particular efforts should be made to recognise it within the context of Iberian rodent conservation planning.

5.4 THE VALUE OF NONINVASIVE SAMPLING

Noninvasive genetic sampling (NIGS) has been increasingly used in animal conservation genetics for species identification (*e.g.* Oliveira *et al.* 2010; Alasaad *et al.* 2011) and also for the estimation of population parameters (*e.g.* population size,

Solberg *et al.* 2009; paternity, Morin *et al.* 1993). It is a particularly valuable technique in studies of illusive, rare or small species, such as rodents, since it obviates the need to capture and handle the individuals.

During the present study, some difficulties arose during sample collection of *M. agrestis* due to low live trapping successful rates. In particular, in northwestern Iberia it was very difficult to collect specimens, which explains the low quantity of samples analysed in paper I (Chapter 2) from that area and may even reflect the small population size. Fortunately, nowadays there are several techniques available for obtaining DNA from diverse biological sources, including from faeces or bone material in owl pellets (Taberlet & Fumagalli 1996; Beja-Pereira et al, 2009). Owl pellets can be an important noninvasive source of rodent species, since usually they constitute a major part of barn owls' diet. Since owl pellets are easy to collect, they can constitute an abundant and inexpensive source of samples for genetic analysis (Poulakakis *et al.* 2005). In our study a large quantity of bone samples was collected throughout the Iberian Peninsula, revealing with more precision the distributional range of the Portuguese and Southern lineages in that area. Moreover, it allowed the analyses of the secondary area of contact revealing a wide area of introgression together with evidence of the existence of barriers to gene flow between these two groups.

As detailed in Chapter 1, the need to use NIGS techniques lead to the definition of an additional goal: the development of a method for Iberian rodent genetic identification using both mitochondrial and nuclear loci that could be applied to noninvasive samples like bones recovered from owl pellets and faeces. This method (Chapter 4, paper V), allowed the unambiguous identification of all Iberian species, except for the sibling species *Microtus lusitanicus* and *M. duodecimcostatus*. It was also successfully applicable to the noninvasive samples tested, namely faecal pellets and bones from owl pellets. By allowing the identification of the great majority of Iberian rodents from noninvasive samples, this method can be considered as a valuable approach in studies of distribution, spatial ecology, population dynamics and for conservation.

5.5 THE FIELD VOLE AS A MODEL SYSTEM

The main results of this study represent an important contribution to the understanding of evolutionary processes that lead to the diversity of small mammals in Eurasia, stressing the field vole as an informative model system.

Within the time span of the last glacial period (up to ca. 100 000 years BP) the single species of field vole (*Microtus agrestis*) has differentiated into three distinct evolutionary units that might even be considered species. This implies a more rapid pace of speciation than is generally suggested. In fact, speciation has been usually considered as a very slow process, perhaps requiring millions of years (Hendry 2009). However, it is generally thought that cryptic species are of more recent origin such that morphological and other traits have not yet evolved (Bickford *et al.* 2007). Even so, considering typical estimated speciation times for small mammals, speciation in *Microtus agrestis* appears to represent an unusually rapid process. As detailed in Chapter 2 (paper I) the four closely related species of the *Sorex araneus* species complex show similar levels of divergence as found in the field vole but the estimated separation times are older (54 000 to 580 000 years BP; Yannic *et al.* 2008). Moreover, other small mammals that are also considered as close to being species, show earlier estimates of divergence times, such as *Mus musculus musculus* and *M. m. domesticus*, at ca. 250 000 years BP (Bonhomme and Searle 2012), and even one example that belongs to the *Microtus* genus, *Microtus arvalis arvalis* and *M. a. obscures*, with divergence times around 237 000 to 587 000 years BP (Fink *et al.* 2004).

The demographic analysis, which reveals relatively small population sizes for all lineages before the Holocene (Chapter 2, paper I, Fig. 6), in the context of ecological and biological characteristics of the species, may help understand this rapid pace of speciation. The low tolerance of *Microtus agrestis* to dry conditions, when compared to the dominant sympatric vole species, *Microtus arvalis* (Dienske 1979), as well as the presence of other competitors within the genus *Microtus* may have only permitted the survival of the field vole in small populations during the cold periods at the end of the last glaciation. Thus, population bottlenecking and strong genetic drift in small populations may have been important processes in this rapid cryptic speciation.

Furthermore, the fact that the nucleotide substitution rate used was a faster rate than that generally used in other phylogeography studies (Herman and Searle, 2011), also contributed to the recentness of the divergence times estimates. Estimating the nucleotide substitution rate is a major challenge in evolutionary studies, and according to Herman and Searle (2011) the substitution rates used in previous studies on *Microtus* may have been too low. The clarification of this aspect as well as the increased amount of genetic information available will most certainly allow more robust divergence times to be inferred (Ho *et al.* 2005). Then, it will be possible to assess if the field vole's rapid divergence is representative of other Eurasian small mammals.

The study of the contact zone between the Portuguese and Southern lineage also revealed some remarkable characteristics. It appears that these recently separated lineages show hybrid male unfitness. Though the lack of male mediated gene flow fits Haldane's rule of sterility or inviability of the heterogametic sex (Haldane 1922), it seems that there may be other issues related to male hybrids, since there is a significantly lower proportion of males within the area of occurrence of hybrids (Chapter 3, paper III). However, the currently available data are not sufficient to understand the basis of this phenomenon. Moreover, the width of the area where hybrids were detected together with the asymmetric introgression observed and the evidence of strong barriers to gene flow between the two lineages suggest that the contact zone has moved, probably due to climatic shifts in the end of the last glaciations. These mechanisms have already been proposed as responsible for the characteristics of a hybrid zone between subspecies of *Salamandra salamandra,* located in the same area (northwestern Iberia, García-Paris *et al.* 2003).

All the evidence collected throughout this study, molecular, karyotypic and morphometric, highlight the field vole species complex as a remarkable system for understanding speciation. The detected rapid pace of speciation (which translates into a relatively sorted phylogeny), high reproductive isolation, small morphological and karyotypic variability, together with the presence of astonishingly large sex chromosomes, makes the field vole an obvious candidate for more comprehensive studies aimed at investigating both the cryptic speciation process and the nature of species boundaries. Furthermore, *M. agrestis* is also interesting as a representative of a widespread and speciose genus, where it seems to hold a relatively basal position, thus providing the prospects to extend the findings on a particular system to continent-wide species radiations.

5.6 FUTURE RESEARCH PROSPECTS

Although this PhD thesis has made an important contribution to the evolutionary history of the field vole and to the knowledge of cryptic speciation processes in small mammals in a Eurasian context, more work would be desirable in several aspects.

The multilocus molecular dataset collected has allowed the inference of both the times of divergence and the evolutionary history of *M. agrestis* in Eurasia, providing insights into the speciation process in small mammals. Thus, *M. agrestis* is highlighted as an obvious candidate system for the further investigation of cryptic speciation

processes and species boundaries through genomic approaches. Furthermore, deepening the genomic analysis (using next generation sequencing techniques) and expanding to other *Microtus* species would allow for a more complete understanding of the origin of the diversity of this genus over its Eurasian and North American distribution.

The karyotypic analysis of the Portuguese lineage that revealed some differences in the size and structure of the sex chromosomes was based on only a few specimens. Moreover, data on the Southern lineage are practically non-existent. Thus, an extensive analysis of karyotypic variability within both the Portuguese and Southern lineages is urgently needed and may have important relevance to understand male hybrid unfitness.

Considering this last point, in this work some evidence suggest the existence of reproductive isolation between the Southern and Portuguese lineages. However, a more detailed analysis of the secondary contact between these lineages using highly polymorphic markers, like microsatellites, would provide a better characterisation. It would be also desirable to perform a comprehensive analysis of the possible area of contact between the Northern and Southern lineages, most likely located in central and southern France. The implementation of ecological studies in these areas would also greatly contribute to understand the reproductive isolation processes in place. Moreover, to clearly determine the extent of reproductive isolation between these three lineages, captive breeding studies involving specimens from all lineages would probably be the best approach.

Although subtle morphometric differences were detected in mandible shape, the morphometric analysis of other structures, like the skull and the molar teeth, would provide more information related with the morphological divergence between the two lineages. Moreover, it would be important to examine also the morphometric variability of the Northern lineage in combination with the data already collected for the Portuguese and Southern lineage. This would allow a detailed analysis of the correspondence between morphometric and molecular differentiation between the three lineages.

5.7 REFERENCES

Alasaad S, Soriguer RC, Jowers MJ, Marchal JA, Romero I, Sánchez A (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.

Almaça C (1993) On the Portuguese population of *Microtus agrestis* (Linnaeus, 1761). *Arquivos do Museu Bocage Nova Série*, **2**, 319-335.

Avise JC, Shapira JF, Daniel SW, Aquadro CF, Lansman RA (1983) Mitochondrial DNA differentiation during speciation process in *Peromyscus*. *Molecular Biology and Evolution*, **1**, 38-56.

Baker RJ, Bradley RD (2006) Speciation in mammals and the genetic species concept. *Journal of Mammalogy*, **87**, 643-662.

Bastos-Silveira C, Santos SM, Monarca R, Mathias ML, Heckel G (2012) Deep mitochondrial introgression and hybridization among ecologically divergent vole species. *Molecular Ecology*. DOI: 10.1111/mec.12018.

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.

Beysard M, Perrin N, Jaarola M, Heckel G, Vogel P (2012) Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (*Microtus agrestis*). *Journal of Evolutionary Biology*, **25**, 400-408.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**, 148-155.

Bonhomme F, Searle JB (2012) House mouse phylogeography. In *Evolution of the House Mouse* (eds Macholán M, Baird SJE, Munclinger P, Piálek J). Cambridge University Press, Cambridge, 278-296.

Brändli L, Handley LJL, Vogel P, Perrin N (2005) Evolutionary history of the greater white-toothed shrew (*Crocidura russula*) inferred from analysis of mtDNA, Y and X chromosome markers. *Molecular Phylogenetics and Evolution*, **37**, 832-844.

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

Cabrera A (1914) *Fauna Iberica*. Mamíferos. Madrid.

Carvalho SB, Brito JC, Crespo EJ, Possingham HP (2010) From climate change predictions to actions – conserving vulnerable animal groups in hotspots at a regional scale. *Global Change Biology*, **16**, 3257-3270.

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

Dienske H (1979) The importance of social interactions and habitat in competition between *Microtus agrestis* and *Microtus arvalis*. *Behaviour*, **71**, 1-126.

Ence DD, Carstens BC (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**, 473-480.

Evin A, Baylac M, Ruedi M, Mucedda M, Pons JM (2008) Taxonomy, skull diversity and evolution in a species complex of *Myotis* (Chiroptera: Vespertilionidae): a geometric morphometric appraisal. *Biological Journal of the Linnean Society*, **95**, 529-538.

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

Fløjgaard C, Normand S, Scov F, Svenning JC (2009) Ice age distributions of European small mammals: insights from species distribution modelling. *Journal of Biogeography*, **36**, 1152-1163.

Fredga K, Jaarola M (1997) The origin and distribution of the Lund Y chromosome in *Microtus agrestis* (Rodentia, Mammalia). *Hereditas*, **126**, 25-34.

García-París M, Alcobendas M, Buckley D, Wake DB (2003) Dispersal of viviparity across contact zones in Iberian populations of fire salamanders (*Salamandra salamandra*) inferred from discordance of genetic and morphological traits. *Evolution*, **57**, 129-143.

García-Ruiz JM, Moreno A, González-Sampériz P, Valero Garcés B, Martí Bono C (2010) La cronologia del último ciclo glaciar en las montañas del sur de Europa. Una revisión. *Cuaternario y Geomorfología*, **24**, 35-46.

Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu HT, Bulatova N, Ziv Y, Nachman M (2008) Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology*, **17**, 5349-5363.

Gündüz I, Jaarola M, Tex C, Yeniyurt C, Polly PD, Searle JB (2007) Multigenic and morphometric differentiation of ground squirrels (*Spermophilus*, Sciuridae, Rodentia) in Turkey with the description of a new species. *Molecular Phylogenetics and Evolution*, **43**, 916-935.

Harris S, Morris P, Wray S, Yalden D (1995) A review of British mammals: population estimates and conservation status of British mammals other than cetaceans. Joint Nature Conservation Committee, Peterborough, UK.

Heckel G, Burri R, Fink S, Desmet JF, Excoffier L (2005) Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. *Evolution*, **59**, 2231-2242.

Hellborg L, Gündüz I, Jaarola M (2005) Analysis of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology*, **14**, 2025–2031.

Hendry AP (2009) Speciation. *Nature*, **458**, 162-164.

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.

Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, **12**, 101-109.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561-1568.

Jaarola M, Martínková N, Gündüz I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedaski-Tsolis S, González-Esteban J, López-Fuster MJ, Kandaurov AS, Kefelioğlu H, Mathias ML, Villate I, Searle JB (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.

Krapp F, Niethammer J (1982) *Microtus agrestis* (Linnaeus, 1761) - Erdmaus. In *Handbuch der Säugetiere Europas Band 2/I*, Nagetiere II (eds Niethammer J, Krapp F) Akademische Verlagsanstalt, Wiesbaden, 349-373.

Lataste F (1883). *Arvicola rozianus* Barboza du Bocage est synonyme de *Mus agrestis* Linné. *Le Naturaliste*, **47**, 373-374.

Lewis CJ, McDonald EV, Sancho C, Peña JL, Rhodes EJ (2009) Climatic implications of correlated Upper Pleistocene glacial and fluvial deposits on the Cinca and Gállego Rivers (NE Spain) based on OSL dating and soil stratigraphy. *Global and Planetary Change*, **67**, 141-152.

Loy A (1996) An introduction to geometric morphometrics and intraspecific variation: a fascinating adventure. In *Advances in morphometrics* (eds Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE). Plenum, New York, 271-273.

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

Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B*, **363**, 2971-2986.

Martínková N, Zima J, Jaarola M, Macholán M, Spitzenberger F (2007) The origin and phylogenetic relationships of *Microtus bavaricus* based on karyotype and mitochondrial DNA sequences. *Folia Zoologica*, **56**, 39-49.

Mayr E (1963) *Animal Species and Evolution.* Harvard University Press, Cambridge, Massachusetts.

Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farelo L, Alves PC (2012) Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulation. *Systematic Biology*, **61**, 367–381.

Miller GS (1912) *Catalogue of Mammals of Western Europe.* British Museum, London.

Moreno A, González-Sampériz P, Morellón M, Valero-Garcés BL, Fletcher, WJ (2012) Northern Iberian abrupt climate change dynamics during the last glacial cycle: a view from lacustrine sediments. *Quaternary Science Reviews*, **36**, 139-153.

Moreno-García M, Pimenta C (2002) The paleofaunal context. In *Portrait of the Artist as a Child: The Gravettian Human Skeleton from the Abrigo do Lagar Velho and its Archeological Context* (eds Zilhão J, Trinkaus E). Instituto Português de Arqueologia, Lisboa, 112-131.

Morin PA, Wallis J, Moore JJ, Chakraborty R, Woodruff DS (1993) Non-invasive sampling and DNA amplification for paternity exclusion, community structure and phylogeography in wild chimpanzees. *Primates*, **34**, 347-356.

Moritz C (1994) Defining “Evolutionarily Significant Units” for conservation. *Trends in Ecology and Evolution*, **9**, 373-375.

Myllymäki A, Christiansen E, Hansson L (1977) Five-year surveillance of small mammal abundance in Scandinavia. *EPPO Bulletin*, **7**, 385-396.

Niethammer J (1964) Ein Beitrag zur Kenntnis der Kleinsauger Nordspaniens. *Zeitschrift für Säugetierkunde*, **29**, 193-220.

Niethammer J (1970) Uber Kleinsauger aus Portugal. *Bonner Zoologische Beiträge*, **21**, 89-118.

Oliveira R, Castro D, Godinho R, Luikart G, Alves PC (2010) Species identification using a small nuclear gene fragment: application to sympatric wild carnivores from southwestern Europe. *Conservation Genetics*, **11**, 1023-1032.

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

Pfenninger M, Schwenk K (2007) Cryptic animal species are homogeneously distributed among taxa and geographical regions. *BMC Evolutionary Biology*, **7**, 121-127.

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.

Póvoas L, Zilhão J, Chaline J, Brunet-Lecomte P (1992) La faune de rongeurs du Pléistocene supérieur de la grotte de Caldeirão (Tomar, Portugal). *Quaternaire*, **3**, 40-47.

Rensen H, Vandenberghe J (2003) Investigation of the relationships between permafrost distribution in NW Europe and extensive winter sea-ice cover in the North Atlantic Ocean during the cold phases of the Last Glaciation. *Quaternary Science Reviews*, **22**, 209-223.

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.

Sites Jnr JW, Marshall JC (2003) Delimiting species: a renaissance issue in systematic biology. *Trends in Ecology and Evolution*, **18**, 462-470.

Solberg KH, Bellemain E, Drageset OM, Taberlet P, Sweson JE (2006) An evaluation of field and noninvasive methods to estimate brown bear (*Ursus arctos*) populations size. *Biological Conservation*, **128**, 158-168.

Sztencel-Jabłonka A, Jones G, Bogdanowic ZW (2009) Skull morphology of two cryptic bat species: *Pipistrellus pipistrellus* and *P. pygmaeus* - a 3D geometric morphometrics approach with landmark reconstruction. *Acta Chiropterologica*, **11**, 113-126.

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

Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics*, **105**, 437-460.

Trouessart EL (1910) *Faune des Mammiféres d'Europe*. R. Friedlander und Sohn. Berlin.

Wilson DE, Reeder DM (2005). *Mammal Species of the World. A Taxonomic and Geographic Reference* (3rd ed), Johns Hopkins University Press, Maryland.

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.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) *Geometric Morphometrics for Biologists: a Primer*. Elsevier Academic Press, London.

CHAPTER 6

CONCLUSIONS

The specific goals set for this work were fully accomplished and correspond to the five scientific papers presented in Chapters 2 to 4. In the following paragraphs the main conclusions obtained in the different chapters are summarised.

1. The analysis of a multilocus dataset, consisting of loci with different modes of inheritance (mitochondrial, sex-linked and autosomal loci) corroborated the two lineages of *Microtus agrestis* previously described in Eurasia. In addition, it allowed the confirmation of the existence of a third distinct lineage in northwestern Iberia. These lineages show high divergence at every loci analysed, although the level of differentiation was found to be more substantial for mtDNA than for the nuclear loci. Although differences in the topologies of the individual gene trees were found, the application of recently developed coalescent methods that model the membership of the individuals to the lineages gave overwhelming support for considering the three lineages as different evolutionary units.

2. Therefore, three cryptic, parapatric lineages of *Microtus agrestis* were confirmed in Eurasia by a multilocus molecular approach: Northern, Southern and Portuguese. The Northern lineage has a wide range throughout Eurasia, occurring through northern and central Europe and into Asia. The Southern lineage occupies a narrow belt in Southern Europe, including northeastern Iberia, southern and central France and through the south of the Alps into Hungary. The Portuguese lineage has a limited range restricted to western Iberia.

3. Results obtained using species tree inference methods suggest that the divergence events occurred during the last glacial cycle (up to ca. 100 000 years BP), highlighting a rapid pace of speciation in this species complex. The Portuguese lineage is identified as the first to diverge (at ca. 70 000 years BP), and has apparently remained endemic to Iberia ever since. The separation of the Southern and Northern lineages is dated to the Last Glacial Maximum (LGM; at ca. 18 500 years BP). Some possible refugial areas are identified for the Southern lineage, namely southern France, northern Italy and northeastern Spain, while the Northern lineage refugium was probably located further to the east in Europe.

4. The analysis of karyotypes of the Portuguese field voles revealed some minor variation in sex chromosome morphology compared to existing data on individuals of the Northern lineage. It appears that the giant sex chromosomes, already well-known in the field vole, may be particularly large in animals of the Portuguese

lineage. Moreover, the male specimen examined had a Y chromosome with a distinct short arm, but with half the length of the Lund Y variant in the Northern lineage. These results, although based on only a few specimens, point to a possible karyotypic differentiation between the lineages of the field vole and stress the interest in performing an extensive analysis of karyotypic variability within the Portuguese and Southern lineages, to match that already carried out on the Northern lineage.

5. The noninvasive genetic analysis of bones from owl pellets from Iberia allowed the distribution ranges of the Portuguese and Southern lineage within the Iberian Peninsula to be assessed. The Portuguese lineage occupies a wider area than indicated by earlier studies, ranging from central Portugal, through Galicia and encompassing a considerable area of the provinces of Asturias and Léon in northern Spain.

6. This noninvasive genetic study revealed the existence of a wide area of introgression between the Portuguese and Southern lineages in northwestern Spain, going from Galicia into the Cantabria and Burgos regions. However, some evidence of barriers to gene flow between the Portuguese and Southern lineages was also observed. Only a small number of hybrids was detected and introgression was found to be asymmetric, with a higher introgression of Southern lineage alleles into the Portuguese lineage in northwestern Spain. No hybrid males were detected, suggesting strong selection against them.

7. Geometric morphometrics revealed significant difference in mandible shape between the Portuguese and Southern field vole lineages. However, no differences in the size of the mandible were detected. These results suggest that, although these forms are considered ‘cryptic’ they already show some subtle morphometric differentiation in informative structures like the mandible.

8. A method for rodent species identification that can be applied to noninvasive genetic sampling, namely faeces and bones from owl pellets, was established. This method allowed for the unambiguous identification of all Iberian rodents with the exception of the sibling species *Microtus lusitanicus* and *M. duodecimcostatus*. Its application to noninvasive samples provided reasonable amplification success. This method can be viewed as having great value in studies of distribution, abundance and population dynamics as well as for conservation purposes.

Overall, multiple evidence suggests that the field vole, *M. agrestis*, is indeed a complex of three distinct evolutionary units, genetically distinct (in all genomic compartments) and that also show some karyotypic variability and some subtle morphological differentiation. Collecting additional information both on karyotypic and morphological variability as well as genomic data would nonetheless be desirable as it could provide further support for the species status of these lineages. Regardless of that, the Northern, Southern and Portuguese lineages should be recognised undoubtedly as Evolutionarily Significant Units for conservation purposes.

U.PORTO

FC FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO