UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

COLONISATION HISTORY OF A UBIQUITOUS INVASIVE, THE HOUSE MOUSE (*MUS MUSCULUS DOMESTICUS*): A GLOBAL, CONTINENTAL AND INSULAR PERSPECTIVE.

Sofia Isabel Vieira Gabriel

DOUTORAMENTO EM BIOLOGIA

(BIOLOGIA EVOLUTIVA)

2012

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

COLONISATION HISTORY OF A UBIQUITOUS INVASIVE, THE HOUSE MOUSE (*MUS MUSCULUS DOMESTICUS*): A GLOBAL, CONTINENTAL AND INSULAR PERSPECTIVE.

Sofia Isabel Vieira Gabriel

Tese orientada pela Profª. Doutora Maria da Luz Mathias e pelo Prof. Doutor Jeremy Byron Searle, especialmente elaborada para a obtenção do grau de doutor em Biologia (Biologia Evolutiva)

2012

This study was funded by Fundação para a Ciência e a Tecnologia (FCT) through a PhD grant – SFRH/BD/21437/2005 – attributed to Sofia Gabriel.

This Thesis should be cited as:

Gabriel SI (2012) *Colonisation history of a ubiquitous invasive, the house mouse (Mus musculus domesticus): a global, continental and insular perspective*. PhD Thesis, University of Lisbon, Lisbon, Portugal.

Nota Prévia

Nos termos do nº1 do Artigo 45, Capitulo V, do Regulamento de Estudos Pós-Graduados da Universidade de Lisboa, publicado no *Diário da República* – II Serie Nº 65, de 30 de Março de 2012, na elaboração desta dissertação foi efectuado o aproveitamento total de resultados de trabalhos já publicados, submetidos ou a submeter para publicação em revistas internacionais com factor de impacto e com arbitragem científica (indexadas no ISI Web of Knowledge – Journal Citation Report®), os quais integram alguns dos capítulos da presente tese. Tendo em conta que os referidos trabalhos foram realizados em colaboração com outros autores, a candidata esclarece que em todos eles liderou e participou activamente na concepção, recolha dos dados e execução do trabalho experimental, bem como na análise e discussão dos resultados obtidos, bem como na escrita dos respectivos artigos científicos.

Apesar desta tese integrar artigos científicos já publicados ou submetidos para publicação, o padrão de formatação apresentado foi uniformizado para todos os capítulos.

Lisboa, 11 de Junho de 2012

Sofia Isabel Vieira Gabriel

Para o meu avô Tó,

que teria gostado de ver este dia.

Table of contents

Acknowledgements *i*

Resumo *v*

Abstract *xiii*

Chapter 1 1

1. General Introduction 3

1.1. Phylogeography 3

1.2. Molecular markers for studies at different geographical scales 7

1.3. Molecular markers in the study of human mediated colonisations 9

1.4. The house mouse, *Mus musculus* Linnaeus 1758 11

1.4.1. Taxonomy and commensalism with humans 11

1.4.2. The western house mouse, *Mus musculus domesticus* Schwarz and Schwarz 1943 14

1.4.3. The house model as a model organism 16

1.4.4. Mice on Islands 17

1.5. Aims and thesis structure 21

1.5.1. Aims 21

1.5.2 Thesis structure 23

1.6. References 25

Chapter 2 43

Paper I Of mice and the 'Age of Discovery': the complex colonization history of the Azorean archipelago by the house mouse (*Mus musculus*) as revealed by mitochondrial DNA variation. 43

[Submitted to *'Journal of Evolutionary Biology'*]

Paper II Genetic structure of house mouse (*Mus musculus* Linnaeus 1758) populations in the Atlantic archipelago of the Azores: colonisation and dispersal. 97

[Submitted to '*Biological Journal of the Linnean Society*']

Chapter 3 127

Paper III Of mice and 'convicts': origin of the Australian house mouse, *Mus musculus*. 127

[Published in *PLoS ONE*, 6, e28622 (2011)]

Chapter 4 153

Paper IV History of a nearly global invasion: the worldwide colonisation of the western house mouse (*Mus musculus domesticus*). 153

[In preparation]

Chapter 5 183

Paper V Colonization, mouse-style. 183

[Published in *BMC Biology*, 8, 131. (2010)]

Chapter 6 193

6. General Discussion 195

Acknowledgements

Writing these words (in practice, the last ones of this thesis) seemed to be the easiest task. No need for references, statistical analysis, only a memory exercise to revisit this already long journey, which culminates here. This journey was obviously influenced by many people and many places, so I'll try to highlight the most important.

At the institutional level I would first like to thank 'Fundação para a Ciência e a Tecnologia' for funding me and this work through a PhD scholarship (SFRH/BD/21437/2005).

To my supervisors, thank you for accepting and accompanying me during this journey. To Prof. Maria da Luz Mathias I particularly thank the leap of faith, almost 10 years ago, when I first joined her group with zero hands-on experience. Thanks for the support, encouragement and confidence in me, which have happily made the 'small mammal group' my home ever since. To Prof. Jeremy Searle, I would like to thank for receiving me in his lab, introducing me to the all-new world of molecular genetics and phylogeography (well, all-new to me!). Your hospitability and friendship helped to make my periods in York much more enjoyable. Also, thanks for challenging me, for trusting me the job of "taking on the world" (the *domesticus* world, of course!) and for going the extra mile to help me have the best possible work presented here.

In York, I also have to mention all my lab mates, from both the 'Searle' and 'Douglas' labs, for their companionship, advice and pub trips: Dan Förster, Mabel Giménez, Friða Jóhannesdóttir, Rodrigo Vega, Ellie Jones, İslam Gündüz, Natália

Martínková, Jerry Herman, Tom White, Nancy Irwin and my fellow Portuguese, Joana Paupério. In particular, a special word to my dearest Mabel, Nancy and Natália, whom I particularly miss, and to Joana who became much more than a 'cultural link' with home.

Back in Lisbon, I would like to thank my colleagues and friends that have been sharing this journey with me. Either for picking me up or telling me off when necessary, or just for being there to help me keep my sanity at desirable levels, a big 'Thanks!' to all of you. Among them, a special acknowledgement to Rita Monarca, Joaquim Tapisso, Ana Cerveira, Cristiane Silveira and the Entomology gang. To my friends from the previous century that, despite the current worldwide dispersal, always make it so easy and familiar every time we're together: Guida, Andreia, Rita, Inês(es), Hugo, Paulo and Raul. To Catarina Castro and André Pestana, thanks for reminding me that "It's later than you think, so let's just go and enjoy!". Finally, to my 'big sister', my dear friend Paula Lopes, that always has the wisest words, and the loudest laugh!

When compared with other small mammals, working with a commensal species raises slightly different challenges when it comes to sampling. So, if my field trips resulted in success I also owe it to all the people in the Azores and southern Spain that allowed me to set the traps in their properties and to the 'Serviços de Desenvolvimento Agrário' from several Azorean islands. For the islands of São Miguel and Terceira, I'm grateful to Prof. Margarida Santos Reis and Ana Quaresma for providing me house mouse samples resulting from an ongoing project at the time. And of course, a huge 'Thank you!' to my friends that got a lot more work than they bargained for by joining me on the field trips to the Azores

(Mabel Giménez, Rita Monarca, Paula Lopes and Pedro Martins) and southern Spain (Joaquim Tapisso and Pedro Martins).

To my core family, which gives me the structure and the strength to carry on with this very unstable career choice: Carlos, Filipa, Santiago, Ana Maria, Céu, Avó Celeste, Avô João e Avó Maria. Specially, of course, thank you so, so much to my parents, that always supported (in more ways than they should…) my decision of becoming a Biologist. They always told me that work ethics and hard work would eventually pay off and I've tried to live up to that.

Finally, but certainly the most important, my partner in crime, Pedro. Building a life and a home while carrying the weight of our parallel PhDs was certainly a challenge, with all the travelling and the long periods of absence involved. A challenge we tried to take the most of, which I think we succeeded because some of our best memories will always be linked to this period of our lives. During these past months, in particular, thanks for your infinite patience during my moments of struggle and disbelieve. Thanks for being my right arm and my rock, for always making me laugh and for nurturing this side of us that is invisible to everybody else.

Resumo

Na Era do 'ADN', o uso de marcadores moleculares tem vindo a revolucionar a forma como olhamos para o mundo biológico que nos rodeia. Nos últimos 25 anos, a Filogeografia tem-se desenvolvido como um ramo da Biologia de eleição para explorar questões relacionadas com a história da expansão e colonização de inúmeras espécies de plantas e animais, incluindo o Homem. Uma das muitas aplicações práticas desta abordagem filogeográfica relaciona-se com o estudo de uma das mais devastadoras causas de perda de biodiversidade em todo o mundo: as espécies invasoras. Os seres humanos, seja involuntária ou propositadamente, têm vindo a transportar inúmeras espécies muito para além da sua área de distribuição e habitat natural, afectando a trajectória evolutiva quer das espécies introduzidas, quer das espécies nativas do ecossistema invadido. De entre as centenas de espécies invasoras mais bem-sucedidas, o ratinho-caseiro (*Mus musculus*) é certamente uma das que origina efeitos deletérios mais impactantes. Este pequeno roedor de hábitos comensais é um dos mamíferos mais ubíquos, apresentando uma distribuição practicamente global. Em estreita associação com o Homem, o ratinho-caseiro colonizou todos os continentes e uma grande proporção das ilhas de todo o mundo.

A sub-espécie de ratinho-caseiro mais comum, que ocorre na Europa Ocidental, *Mus musculus domesticus*, tornou-se comensal na região do Médio Oriente, há cerca de 10 000 – 12 000 anos, continuando o seu percurso evolutivo em paralelo com o Homem. Esta associação coincidiu com o início do processo de sedentarização das populações humanas na zona do antigo Crescente Fértil, onde a agricultura e a domesticação de animais davam os primeiros passos. O ratinho-

caseiro estava, então, no lugar e tempo certos para tirar partido das condições involuntariamente proporcionadas pelo Homem, nomeadamente, recursos alimentares ilimitados e abrigo contra predadores e condições climatéricas adversas. Mais recentemente, há cerca de 3000 – 4000 anos, na sequência da intensificação das trocas comerciais dos povos da Idade do Bronze, a sub-espécie *domesticus* iniciou então uma expansão da sua área de distribuição para Oeste, que foi reforçada com as movimentações marítimas dos povos da Idade do Ferro, estimuladas por intensas trocas comerciais ao longo de todo o Mediterrâneo. Sob estas condições, o ratinho-caseiro foi sendo transportado acidentalmente ao longo de ambas as costas do Mediterrâneo, progredindo para Oeste pelo Norte da África e Europa Ocidental. Contudo, é já no início do século XV, com o advento da Grande Expansão Marítima, inicialmente protagonizada pelos Portugueses e Espanhóis, seguidos dos Britânicos, Franceses e Holandeses, que a espécie testemunha a maior expansão da sua história evolutiva, sendo transportada 'clandestinamente' de forma global onde quer o Homem se deslocasse. Assim, para além do interesse intrínseco do estudo desta espécie, a história da colonização do ratinho-caseiro oferece uma oportunidade ímpar e empolgante de tirar ilacções acerca da história das colonizações e movimentações humanas.

Este estudo utilizou uma abordagem filogeográfica para examinar a história da colonização da sub-espécie Ocidental de ratinho-caseiro (*Mus musculus domesticus*) em três escalas geográficas distintas: regional (insular), continental e global.

Escala regional (insular). O estudo realizado à escala regional focou-se no arquipélago dos Açores, que foi objecto de uma análise detalhada, fazendo uso de diferentes marcadores moleculares, com o fim de obter resultados

complementares. Para tal, foram utilizados um marcador mitocondrial (D-loop) e dois tipos de marcadores nucleares (microssatélites e marcadores diagnosticantes do nível sub-específico de *M. musculus*). Assim, foram obtidas um total de 239 sequências de D-loop de animais capturados em todas as ilhas dos Açores (Santa Maria, São Miguel, Terceira, Graciosa, São Jorge, Pico, Faial, Flores e Corvo), confirmando o seu estatuto taxonómico de *M. m. domesticus*. Para fins comparativos, foi utilizado um conjunto de 76 amostras de ratinho-caseiro de Portugal continental, obtidas ao longo de todo o país na sequência de um trabalho anterior. Por outro lado, dada a total ausência de sequências publicadas de D-loop para esta espécie no sul de Espanha, e considerando a potencial importância histórica desta região como fonte colonizadora do arquipélago dos Açores, procedeu-se a uma amostragem ao longo da costa das províncias de Huelva e Cádiz, resultando em 66 novas sequências de D-loop de *M. musculus domesticus*. A totalidade dos haplótipos obtidos foi sujeita a análise filogenética juntamente com todos os haplótipos da sub-espécie *M. m. domesticus* disponíveis ou publicados à altura da análise, utilizando as sub-espécies *M. m. musculus* e *M. m. castaneus* como *outgroup*. Uma análise Bayesiana revelou a presença da maior parte das linhagens de D-loop previamente descritas para a sub-espécie, sugerindo uma complexa história da colonização dos Açores. Apesar de, como um todo, os resultados para o arquipélago revelarem múltiplas origens geográficas para as populações de ratinho-caseiro, ao nível de cada ilha, individualmente, foi observada uma menor heterogeneidade, sugerindo colonizações únicas em algumas delas. A esperada ligação histórica com Portugal continental reflectiu-se no padrão de variação de D-loop de algumas das ilhas, mas não todas. Uma associação mais inesperada com uma área distante do norte da Europa foi também detectada em três ilhas,

possivelmente reflectindo um contacto humano com os Açores anterior à sua descoberta oficial no século XV por exploradores Portugueses. Para além desta abordagem através da análise de DNA mitocondrial, o mesmo *dataset* foi analisado ao nível do genoma nuclear. Para tal, foram seleccionados dois genes localizados nos cromossomas sexuais, *Btk* (no cromossoma X) e *Zfy2* (no cromossoma Y), cujas sequências apresentam inserções/delecções características das sub-espécies *M. m. domesticus* e *M. m. musculus*. Desta forma, foi possível confirmar que a vasta maioria dos animais amostrados nos Açores possuem, de facto, um genoma característico da sub-espécie *M. m. domesticus*. A única excepção foi detectada em dois indivíduos (um oriundo da Terceira e outro de São Miguel) que, para estes marcadores, apresentaram uma componente do genoma nuclear característica da sub-espécie *M. m. musculus*, sugerindo a chegada ocasional de animais de regiões distantes, onde ocorre esta sub-espécie. Neste trabalho, os animais das ilhas dos Açores foram também genotipados para um conjunto de 19 microssatélites (um *locus* seleccionado por cada cromossoma autossómico). Para fins comparativos, foram também utilizados animais do arquipélago da Madeira e Portugal continental, perfazendo um total de 380 indivíduos genotipados. Desta forma, foi possível avaliar o nível de diferenciação genética entre as 12 populações estudadas, bem como o nível de fluxo genético entre ilhas do mesmo arquipélago, entre arquipélagos e entre estes e o continente. De destacar, os resultados revelaram elevados níveis de heterozigotia em todas as ilhas, sem evidências de reduções significativas no tamanho das populações de ratinho-caseiro. A análise de microssatélites revelou ainda que as populações açorianas constituem unidades genéticas distintas do continente e do arquipélago vizinho (Madeira e Porto Santo), não havendo evidências de migração recente entre estas três unidades geográficas.

Já dentro dos Açores, os grupos Ocidental, Central e Oriental tendem a revelar-se como unidades genéticas distintas, com algumas excepções. Em particular, foram detectados níveis consideráveis de migração entre algumas ilhas dos Açores, principalmente entre as do grupo Central, cujas populações eram relativamente indiferenciadas.

Escala continental. A história da colonização do ratinho-caseiro na Europa Ocidental tem vindo a ser estudada ao longo de cerca de duas décadas, tendo sido dada especial atenção à progressão da sub-espécie desde a sua origem, na região do Médio Oriente. Contudo, nos últimos seis séculos, o ratinho-caseiro expandiu significativamente a sua área de distribuição, tendo sido transportado acidentalmente pelo Homem, e como tal, colonizou continentes onde até então não ocorria naturalmente. A Austrália foi um desses continentes onde a espécie se estabeleceu, tornando-se uma praga de proporções inigualáveis e uma enorme ameaça tanto para a biodiversidade nativa como para a agricultura. Assim sendo, dado que a origem das populações de ratinho-caseiro da Austrália era até aqui desconhecida, foi assim avaliada a história da colonização deste país/continente, tão vasto como toda a Europa Ocidental. Através da análise filogenética das sequências de D-loop de um total de 77 amostras de ratinho-caseiro, foi possível estabelecer uma clara relação entre a história da colonização humana pós-Europeia e a origem das populações desta espécie invasora na remota Austrália. Contrariamente ao observado na Europa, a diversidade haplotípica foi mais reduzida, com um claro domínio de sequências representativas de apenas duas linhagens mitocondriais. Em suma, os resultados revelaram uma forte associação entre a Austrália e as Ilhas Britânicas, confirmando a longa ligação histórica e cultural entre ambas. Dada a proximidade da Austrália com inúmeras ilhas do

continente Asiático (onde ocorre outra sub-espécie – *M. m. castaneus*), é de notar a presença de haplótipos de *M. m. domesticus* em 100% das amostras analisadas. Este resultado reitera a forte influência das nações coloniais europeias na modelação das comunidades de animais e plantas em todo o mundo.

Escala global. Por último, procedeu-se a uma abordagem mais geral e extensiva dos padrões filogeográficos da sub-espécie *M. m. domesticus* a um nível global, no âmbito da sua vasta área de distribuição. Para tal, foram obtidas 760 novas sequências de D-loop, distribuídas por 5 continentes, de vários locais na Europa Ocidental, América do Norte, América Central, América do Sul, África, Australásia e várias ilhas em todo o mundo. Para esta componente do trabalho, a utilização de amostras de colecções museológicas revelou-se de extrema importância uma vez que permitiu uma cobertura geográfica ímpar, nunca antes obtida em estudos envolvendo esta espécie. Assim, foi possível reunir amostras de 47 países (ou suas dependências), 22 dos quais foram amostrados pela primeira vez neste estudo. A reconstrução das relações filogenéticas resultante da compilação dos haplótipos obtidos e previamente publicados revelou, de uma forma geral, uma complexa história da colonização global da sub-espécie *M. m. domesticus*. Este resultado não é de todo surpreendente dado o envolvimento obrigatório do Homem nos movimentos de dispersão de longa distância desta espécie comensal. Sendo a história da colonização do ratinho-caseiro de alguma forma um espelho da história da colonização humana, dada a globalização das movimentações humanas (quer histórica, quer actualmente) a complexidade observada seria expectável. Ainda assim, é notória a sobreposição entre a presença de determinadas linhagens mitocondriais de ratinho-caseiro e os padrões de colonização humana dos países e/ou regiões que inicialmente descobriram (ou se

estabeleceram) em certas regiões do mundo. Estes resultados, apoiados por estudos anteriores, reforçam a influência dos povos europeus de países como a Noruega e Dinamarca (durante a Era Viking) e de Portugal, Espanha, França e Reino Unido (na época dos Descobrimentos) na actual área de distribuição do ratinho-caseiro e, consequentemente, nos padrões de variação genética aí observados.

Dada a inédita amplitude geográfica coberta por este trabalho, foi ainda possível reiterar a adequabilidade de sequências mitocondriais de D-loop enquanto marcador molecular da colonização inicial de uma dada área geográfica pelo ratinho-caseiro, até aqui apenas avaliada em estudos à escala regional.

Em suma, o estudo da colonização desta sub-espécie permitiu inferências sobre as colonizações humanas, enfatizando o valor do ratinho-caseiro como um 'artefacto' vivo, passível de ser incorporado por historiadores no debate de cenários históricos, muitas vezes rodeados de controvérsia.

Palavras-chave: Ratinho-caseiro, *Mus musculus domesticus*, história da colonização, DNA mitocondrial, D-loop, microssatélites.

Abstract

For millennia, humans, either involuntarily or purposefully, have transported numerous species outside their natural range. Among these species, the house mouse (*Mus musculus*) has become one of the most successful and ubiquitous mammals worldwide, in particular its western subspecies *M. m. domesticus*. Due to its commensal relationship with humans, *M. m. domesticus* has colonised all 5 continents and a large proportion of islands, being transported as a stowaway. As well as being interesting in its own right, the colonisation history of house mice provides exciting and novel insights into movements and colonisations of humans. This study uses a 'phylogeographic' approach to examine the colonisation history of the western house mouse at three different geographical scales: regional (insular), continental and global.

Regional (insular) scale. Particular emphasis was given to the colonisation of the Azores archipelago, making use of mitochondrial DNA sequences (D-loop) and nuclear markers (microsatellites and subspecies-specific DNA sequences). A complex colonisation history was revealed, with the cultural and historical link with Portugal being reflected on some but not all islands, with a surprising northern European origin relating to Santa Maria, Terceira and Flores. On the other hand, microsatellites seem to have retained little phylogeographic signal, revealing instead generally high levels of genetic diversity within the Azorean house mouse populations. Evidence of recent migration was also detected mainly among islands of the Central group but not with mainland Portugal or the neighbouring Madeiran archipelago.

Continental scale. The colonisation of the Australian continent by the invasive house mouse was thought to have occurred soon after the establishment of the first British penal colony. Through the analysis of D-loop sequences from both Australia and potential source areas, the historical link with Britain was confirmed, helping to validate the use of phylogeographic patterns of house mice as a proxy of human settlement history.

Global scale. This approach was exploited further with D-loop sequences of the western house mouse being obtained from plentiful locations throughout Western Europe, North, Central and South America, Africa, Australasia and numerous islands worldwide, resulting in the widest sampling yet attained throughout the range of this subspecies. Despite the expected complexity in the patterns of D-loop variation, there is a striking overlap between the presence of certain D-loop lineages in particular regions of the world and the expectations from human history. Both Viking Age and Age of Discovery explorers and subsequent settlers appear to have had a particular influence in the current distribution and genetic makeup of house mouse populations worldwide.

Overall, the main achievements of this study comprise additional validation to the use of house mice as proxies of human movements and colonisations, based on both intense regional and broad global coverage. Accordingly, this study confirmed the suitability of the mitochondrial D-loop as a good molecular marker to assess initial colonisation of house mice in a given geographic area.

Keywords: Western house mouse, *Mus musculus domesticus*, colonisation history, mitochondrial DNA, D-loop, microsatellites.

Chapter 1

General Introduction

1. General Introduction

Trying to understand the biological world in all its diversity and complexity has long intrigued humans. Ever since Aristotle, more than 2000 years ago, many attempts have been made to organise life on Earth by implementing a classification system that is at the same time simple and universal. That task was accomplished by Karl von Linné in 1735 with his *Systema Naturæ*, a classification system that provided a basic background for taxonomy. While the core of the Linnaean system still survives, the rationale behind it has shown considerable transformations. What started as a system to rationally describe the unchanging order inherent to the Biblical creation has turned into a dynamic phylogenetic map of the tree of life, supported by the theory of evolution and the advent of DNA sequencing.

The DNA era is having more impact than any other time in the history of biological sciences in the way biologists address questions related with the diversity of living organisms and the processes that shape their geographic distribution. Over recent years it has become very clear that molecular markers can help to trace the history of expansion, spread and colonisation of plant and animal species, including humans. A field of biology has grown up to exploit these findings – 'Phylogeography'.

1.1. Phylogeography

The concept of 'phylogeography' was proposed 25 years ago (Avise *et al.* 1987) when the generation of genetic data from different species (mostly mitochondrial [mt] DNA) began to uncover striking patterns in the geographic

distribution of genetic lineages. Phylogeography was finally filling the gap between the fields of systematics and population genetics, two separate research areas until then. During the following years, the field quickly expanded, accompanied by the development of rigorous and powerful statistical tools, particularly derived from coalescent theory. These advances enabled a more accurate assessment of the historical forces and processes that shape the spatial distribution of genealogical lineages in a temporal context (Avise 2009, Hickerson *et al.* 2010). Phylogeography mainly focuses on how a certain species came to occupy its current range or part of its range through the analysis of intraspecific genetic variation. Occasionally, phylogeographic inference uncovers hidden biodiversity by revealing diverging genetic lineages 'within species', which turn out in fact to be different species (e.g. Gündüz *et al.* 2007, Charlton-Robb *et al.* 2011).

In the last decade, this young discipline has experienced an explosive growth and from the thousands of studies published under the umbrella of animal phylogeography, mtDNA studies have continued to dominate the field (see Beheregaray 2008 for a review).

Mitochondrial and nuclear DNA differ in many aspects, such as total length and sequence structure, ploidy, mutation rate, mode of inheritance, recombination, presence of introns, percentage of non-coding DNA and repair mechanisms (Fernández–Silva *et al.* 2003, Ballard and Whitlock 2004). The popularity of mtDNA as a molecular marker in phylogeographic studies is based on a combination of its biological and evolutionary properties as well as several technical ease-of-use advantages:

- The mitochondrial genome constitutes a relatively small circular DNA molecule (~ 16,000 – 17,000 base pairs on average) and each cell may have

thousands of copies, which substantially reduces the amount of sample required for analysis, facilitating its isolation even from degraded tissue samples (Hartl and Clark 2006).

- It is haploid and uniparentally (maternally) inherited, as virtually all mitochondria are transmitted from the mother's ovum (for exceptions see Kvist *et al.* 2003, Zhao *et al.* 2004 and references therein, and White *et al.* 2008 for a review). Therefore, as mtDNA is transmitted as a single locus (Avise 2000) it is particularly informative about ancestry, with female lineages being traceable far back in time.
- The high susceptibility to oxidative damage coupled with replication errors and a low fidelity of the mtDNA repair mechanisms leads to a higher mutation rate than nuclear DNA (Bogenhagen 1999, Lynch *et al.* 2006, Park and Larsson 2011), which is particularly valuable in studies addressing recently diverged lineages.
- In general, mtDNA does not undergo recombination (Avise 2000 but see also Ladoukakis and Zouros 2001, Innan and Nordborg 2002, Ballard and Whitlock 2004, White *et al.* 2008). This results in easily resolved differences between closely related individuals with a simpler branching structure of its gene trees when compared with those of nuclear genes.

The mammalian mitochondrial genome is very gene-dense, containing 22 transfer RNAs, 2 ribosomal RNAs and 13 mRNAs coding for essential peptides involved in subunits of the oxidative phosphorylation system (Avise 1986, Fernández–Silva *et al.* 2003). All the coding sequences of the mitochondrial

genome are contiguous, without introns, and the only major non-coding fraction, the control region, is functionally involved in the initiation and regulation of the molecule's replication and transcription process (Fernández–Silva *et al.* 2003). The control region (or D-loop) constitutes around 7% of the total length of the mitochondrial genome. This is also the region that shows the most variation among different species (both in sequence and size), although it contains some conserved elements with possible regulatory functions (Shadel and Clayton 1997). The D-loop is one of the most popular regions in phylogeographic studies because of its particularly high mutation rate (Hartl and Clark 2006).

Despite the numerous advantages of mtDNA, over recent years an increasing awareness of the limitations of the sole use of this marker have become apparent (e.g. Ballard and Whitlock 2004, Rubinoff and Holland 2005, Galtier *et al.* 2009, Balloux 2010), encouraging the inclusion of some form of analysis of the nuclear genome into the phylogeographic framework. Among the most commonly used nuclear markers, bi-parentally inherited microsatellites have proved effective in phylogeography studies, providing a relatively cheap and straightforward way of assessing genetic diversity and population structure that can be applied to questions of colonisation history.

Microsatellite loci, also known as short tandem repeats (STRs), are non-coding stretches of DNA widely distributed throughout the nuclear genome, in which a motif of 1 – 6 nucleotides is repeated in tandem (Schlötterer 2000). Microsatellites have emerged as versatile molecular markers in population studies because of their high mutation rate and ubiquitous occurrence (see Ellegren 2004 for a review). The mutation process does not occur through point mutations (individual base substitution) but instead through slippage of the polymerase

during DNA replication. The repeat motif is either inserted or deleted, resulting in a size polymorphism, which tends to be more frequent the greater the number of tandem repeats a microsatellite encloses. This results in high variability in sequence length and ultimately high levels of polymorphism (Weber and Wong 1993, Ellegren 2004, Selkoe and Toonen 2006). Some errors in slippage may be rectified by exonucleolytic proofreading mechanisms but some mutations escape repair and become established (Eisen 1999). The high mutation rate, which ranges from 10^{-6} to 10^{-2} per generation, is considerably higher than nucleotide substitution rates and results in a large number of alleles per microsatellite locus (Schlötterer 2000). Dinucleotide repeats are, in general, the most abundant in the genome and therefore the most common target for isolation for microsatellite analysis (Li *et al.* 2002).

The typical high level of variability in microsatellites makes them an asset of major importance when one is looking at recent changes in populations as a large number of mutations can occur in a relatively small number of generations.

1.2. Molecular markers for studies at different geographical scales

Phylogeographic studies have covered all geographical scales: local (e.g. White and Searle 2008), regional (e.g. Melo-Ferreira *et al.* 2005, Deffontaine *et al.* 2005), continental (e.g. Randi *et al.* 2004, Burbrink *et al.* 2008) and even global (e.g. Larson *et al.* 2005, Pineda *et al.* 2011, Ascunce *et al.* 2011, Daly-Engel *et al.* 2012). Phylogeographic studies have targeted all parts of the world (Hewitt 1999, Eggert *et al.* 2002, Luo *et al.* 2004, Swenson and Howard 2005, Ojeda 2010, Gongora *et al.* 2012), but there has been particular emphasis on postglacial

colonisation, exemplified by detailed studies on European biota, including the description of several glacial refugia (Taberlet *et al.* 1998, Hewitt 2004a,b, Sommer and Nadachowski 2006, Knopp and Merilä 2008). This includes many studies on European small mammals (e.g. Bilton *et al.* 1998, Jaarola and Searle 2002, Heckel *et al.* 2005, Vega *et al.* 2010, Wójcik *et al.* 2010, Herman and Searle 2011). Ancient DNA techniques have also allowed new insights into the paleo-phylogeographic patterns of some European mammals (Hofreiter *et al.* 2004). There has also been great interest in the colonisation of islands by a range of species. Islands offer unique features compared with continental systems, making them attractive environments for phylogeographic studies (Emerson and Hewitt 2005). The restricted scale and geographic isolation constitute important limitations to gene flow among islands and between islands and mainland, potentially leading to radiation phenomena (Emerson 2002). Over the years, numerous studies have addressed the colonisation history of island systems around the world, mostly by naturally dispersing species (e.g. insects from the Canaries). Molecular markers have also been used to trace the arrival of small mammals onto islands, including human mediated colonisations. The colonisation of the common vole, *Microtus arvalis*, onto Orkney (Haynes *et al.* 2003) or the pygmy shrew, *Sorex minutus*, onto Ireland (McDevitt *et al.* 2011) are just a couple of examples. A few studies on Mediterranean islands are also available, addressing the colonisation of the wood mouse, *Apodemus sylvaticus*, (Michaux *et al.* 2003) and North African shrews onto Sicily (Dubey *et al.* 2008).

1.3. Molecular markers in the study of human mediated colonisations

For most small mammals, large water bodies represent effective geographical barriers and overcoming them usually implicates human assistance. Therefore, understanding the colonisation over marine gaps to islands or distant landmasses is not only of interest for revealing the history of the introduced species, as it can also help to trace the movements of humans with whom the species has travelled as a stowaway (Searle 2008).

Numerous studies on human populations using molecular markers have improved our understanding of the fascinating history of human movements and colonisations (e.g. Cann *et al.* 1987, Templeton 2002, Crawford *et al.* 2010, Oppenheimer 2012). However, it is not always possible to accurately track these movements by making exclusive use of human genetics, or even combined with knowledge from human artefacts. Genetic studies of animals and plants that the humans have taken around with them are a valuable additional source of information. Most obviously, from the Neolithic period onwards, humans have taken domesticated animals and plants with them wherever they go. Several studies have now unravelled the domestication centres and colonisation pathways of the first farm animals using DNA analysis, which in turn imparts knowledge about the domesticators: the humans. Hence, molecular studies on sheep (Chessa *et al.* 2009), goats (Zeder and Hesse 2000), cattle (Ajmone-Marsan *et al.* 2010) and pigs (Larson *et al.* 2005) have provided a better understanding of our own history.

When people started to store food and livestock-rearing, that attracted pests and those pests, right from the early stages of agriculture and animal domestication, would also have been inadvertently transported as hitchhikers with

food stuffs and domesticates (Pascal *et al.* 2010). Among the most iconic examples are mice and rats that developed a close relationship with humans (living in the same place: commensalism), lasting until today. Because of their frequent transport by people, the colonisation history of these pest species (as revealed by phylogeographic studies with DNA) can provide valuable insights into historical movements of human, in the same way as domesticates have been used. For instance, commensal rodents have been used to examine the controversial origin of the first human inhabitants of the Pacific Islands. The arrival of rats and mice to remote islands is directly related with the arrival of humans that transport them in boats (Pocock *et al.* 2005, Cucchi *et al.* 2005). So, starting from this assumption, researchers used the Pacific rat (*Rattus exulans*) as a proxy for human mobility in the attempt to better understand the early colonisation of the islands of the Polynesian triangle (Matisoo-Smith *et al.* 1998, Matisoo-Smith 2002, Matisoo-Smith and Robins 2004, Matisoo-Smith 2009, Matisoo-Smith and Robins 2009). By analysing mitochondrial DNA sequences of both archaeological and modern samples of the Pacific rat, two often cited models of Polynesian history were clearly rejected (Matisoo-Smith and Robins 2004). This is a proof of principle that commensal mammals can provide an exciting and novel insight into human colonisations and maritime movements. The concept behind this approach is very simple: the commensal animals inhabiting a colonised area will be genetically related to those from their source. So, by revealing the colonisation history of the commensal species, it is also possible to trace the movements of the humans with whom they have travelled. Another striking example where introduced commensal rodents have provided a piece of human history, involves the western house mouse (*Mus musculus domesticus*). Two phylogeographic studies also involving

mitochondrial DNA sequences (Gündüz *et al.* 2001, Förster *et al.* 2009) suggest that house mice were probably introduced onto the Atlantic islands of Madeira from an unexpected source area in Northern Europe. The explanation for these surprising results seems to be that Viking seafarers, while sailing along the coast of Iberia over a thousand years ago, accidentally landed on Madeira island leaving mice originated from their homeland in northern Europe as evidence of the event, even though there is no other known historical record of their arrival on Madeira (Gündüz *et al.* 2001, Förster *et al.* 2009). This hypothesis is surprising given that the known main cultural link between Madeira and Europe involves mainland Portugal and the Portuguese discoveries in the 15th century. This exciting perspective created by DNA studies of a commensal species generates a whole new set of questions about human visitations of other Atlantic archipelagos and the house mouse appears to be an ideal model to investigate these possibilities further.

1.4. The house mouse, *Mus musculus* Linnaeus 1758

1.4.1. Taxonomy and commensalism with humans

Unravelling the taxonomy of the *Mus* complex has been quite a challenging task (Chevret *et al.* 2005). *Mus auctor*, believed to have been the common ancestor of all extant species of the genus *Mus*, lived in the Indian subcontinent about 5.5 million years ago (Auffray *et al.* 1990b, Boursot *et al.* 1993). Although many fossils have been identified as *Mus*, no clear paleontological evidence can be used to reconstruct the radiation of the genus. This is due to the rather conservative morphology of mice, and therefore, molecular dating has relied on fossil evidence of the split of *Mus* and *Rattus* estimated at ~10 Myr ago (Boursot *et al.* 1993,

Guénet and Bonhomme 2003). The emergence of the house mouse, *Mus musculus*, was generally accepted to have occurred in the northern Indian subcontinent (Boursot *et al.* 1993, Boursot *et al.* 1996, Din *et al.* 1996). Subsequent studies building on further genetic and morphological data have been suggesting slightly different origins and speciation scenarios, but all taking place in the vicinity of the Indo-Pak subcontinent, Afghanistan or the Iranian plateau (Prager *et al.* 1998, Duvaux *et al.* 2011, Siahsarvie *et al.* 2012). Despite the uncertainties surrounding the exact cradle region, the house mouse, *M. musculus*, seems to have originated about 0.5 – 0.9 million years ago, following an expansion characterized by the occupation of non-overlapping ranges to form three well defined subspecies – *Mus musculus domesticus, M. m. musculus* and *M. m. castaneus* (Boursot *et al.* 1993, 1996, Din *et al.* 1996, Geraldes *et al.* 2008) (Figure 1).

A more recent study, based on 27 autosomal loci provided a new estimation for the divergence time of all three subspecies, suggesting that it has occurred about 350 000 years ago, within a very short period of time (Geraldes *et al.* 2011).

Apart from the previous three well characterized subspecies (both on the mitochondrial and nuclear genome level), a fourth subspecies, *M. m. bactrianus*, was once described as the descendant form of the original ancestral house mice, remaining at the site of origin, in the Indian Peninsula (Silver 1995). However, it was later suggested that this taxon was synonymous with *M. m. castaneus*, occupying the whole of Southeast Asia (Gündüz *et al.* 2000, Baines and Harr 2007). A possible fifth taxa, *M. m. gentilulus* – so far based on mitochondrial DNA data alone – was initially detected in Yemen, in the Arabian Peninsula (Prager *et al.* 1998), and later on also found in Madagascar (Duplantier *et al.* 2002). A sixth subspecies, *M. m. molossinus*, has been described exclusively in Japan but it was

later concluded that it was the result of extensive hybridisation between the two Asian subspecies, *M. m. musculus* and *M. m. castaneus*, and not an independently arising subspecies (Yonekawa *et al.* 1988, Nunome *et al.* 2010).

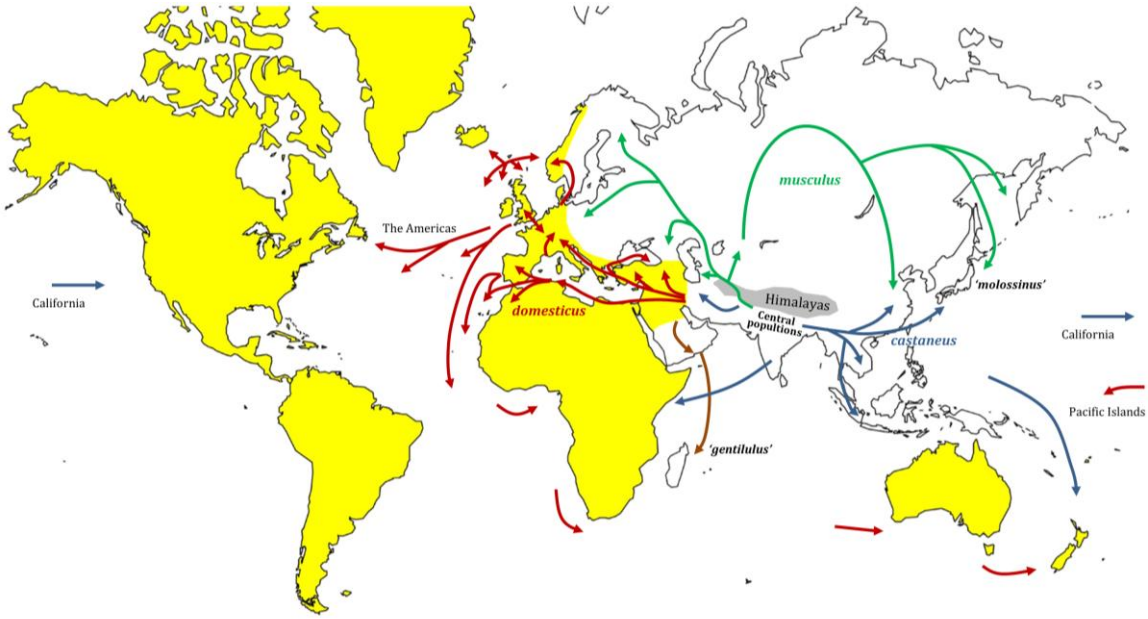


Figure 1. Geographic distribution of the different subspecies of house mouse *Mus musculus*, modified from Bonhomme and Searle (in press) and others. The range of the western European subspecies, *M. m. domesticus*, is highlighted in yellow.

There has been an ongoing debate about whether all these forms of house mouse should be classified taxonomically as separate species or considered as subspecies of *Mus musculus* (see Marshall and Sage 1981, Auffray *et al.* 1990a, Boursot *et al.* 1993, Sage *et al.* 1993, Searle 1998, Berry and Scriven 2005, Geraldes *et al.* 2008). Given that these taxa hybridise and exchange genes when they come into contact, they will here be referred to as subspecies. This contention is mostly supported by extensive work carried out in the secondary contact zone between *domesticus* and *musculus* in Europe (e.g. Payseur *et al.* 2004, Raufaste *et al.* 2005, Macholán *et al.* 2007, Teeter *et al.* 2008, 2010, Wang *et al.* 2011). In the region

where their ranges meet and partially overlap, a hybrid zone has been defined along approximately 2000 km from the Jutland Peninsula in Scandinavia to the Black Sea coast, through Central Europe and the Balkan Peninsula (Boursot *et al.* 1993).

1.4.2. The western house mouse, *Mus musculus domesticus* Schwarz and Schwarz 1943

The western house mouse *M. m. domesticus* (Figure 2, the focus of this dissertation) is now the most common and widespread subspecies, currently occupying Western Europe, Africa, Australasia and the Americas. This nearly global distribution was mostly attained through the commensal relationship developed with humans, which started about 10,000 – 12,000 years ago in the broad region of the ancient Fertile Crescent, in the Middle East when mice entered the domestic space of humans (Auffray *et al.* 1990b, Cucchi *et al.* 2005, Cucchi and Vigne 2006, Rajabi-Maham *et al.* 2008). This region corresponds to present-day Israel up through Lebanon and Syria curving back down through Iraq toward the Persian Gulf, and corresponds to the region where Neolithic people first started to build settlements and develop agriculture. Therefore, the western house mouse was present at the right place at the right time to be able to exploit the sedentary habits of humans. Mice that adapted to the new anthropic niche were then released from some ecological pressures of their natural habitat. By growing and storing grain, humans were involuntarily providing mice with unlimited food resources, shelter against predators and a buffered climate (Cucchi and Vigne 2006).

Figure 2. House mouse, *Mus musculus domesticus*, photographed and captured in a barn in Vila Franca de Xira, Portugal.

In fact, some of the earliest zooarchaeological records of house mice in a commensal setting were retrieved from Neolithic villages in the valley of the Euphrates river (Cucchi and Vigne 2006).

Much later, when humans migrated away from the steppes to colonise new areas, the house mouse went along as a stowaway. It was only about 3,000 – 4,000 years ago, that the *domesticus* subspecies made its way into Europe with Iron Age people (Cucchi *et al.* 2005). Archaeological deposits provides data on the progression of *domesticus* from the Middle East across south-western Europe, and the development of sea transport accelerated the sweep of both mice and people through the Mediterranean basin and North Africa (Cucchi *et al.* 2005, Cucchi 2008). Once in Western Europe, the house mouse just had to wait for the 'Age of Discovery', starting in the early 15th century, to use its commensal life style to go global.

1.4.3. The house model as a model organism

Long before Linnæus taxonomically classified the house mouse as *Mus musculus*, in *Systema Naturae* in 1758 (Linnaeus 1758), this small rodent was already influencing scientific research. In the late 16th century, the Flemish chemist Jan Baptist van Helmont performed an experiment to test a theory on the origin of life suggesting mice would appear by spontaneous generation through a simple recipe: "If you press a piece of underwear soiled with sweat together with some wheat in an open mouth jar, after about 21 days the odour changes and the ferment, coming out of the underwear and penetrating through the husks of the wheat, changes the wheat into mice" (Fox *et al.* 2007). Nowadays, this laughable misconception is perceived as a reflection of the species ubiquity in anthropogenic habitats, in this case opportunistically taking advantage of the wheat provided for the experiment. Indeed, the species common name 'house mouse' is quite appropriate and self-explanatory considering its scientific Latin name '*Mus musculus*', derived from two ancient Sanskrit words, 'musha' meaning 'thief' and 'musculus' meaning 'little mouse'.

Today, the house mouse simultaneously plays conflicting roles in people's lives. On one hand, it is a major economic pest, consuming and damaging crops and foodstuffs for human consumption, eating virtually anything available. Mice destroy much more food than they can eat as they nibble on a wide variety of items, leaving them partially eaten and contaminated with their urine and droppings (Silver 1995). Because of that, house mice are also vectors for a range of diseases, transmitting pathogens and parasites including leptospirosis, tularemia, bubonic plague, typhus and *Salmonella* (Silver 1995).

On the other hand, the house mouse is an icon of scientific research, and over the past century, despite its legacy as a pest, this small rodent has been a most valuable tool to help us make biological sense of ourselves. Used since early genetic studies by Mendel himself (Guénet 2005), the house mouse has become one of the most common laboratory model organisms, used in all sorts of biomedical studies, from cancer to stem cell research, from immunology to genetics (Guénet 2005, Rosenthal and Brown 2007). Since the beginning of the 20th century, breeding colonies have multiplied, starting from various founding stocks. Today, the laboratory mouse is in fact a mosaic derived from the relative proportion of the genome of the three parental house mouse subspecies, *Mus musculus domesticus*, *Mus musculus musculus* and *Mus musculus castaneus* (Guénet 2005, Yang *et al.* 2007, 2011). Not surprisingly, this was the first non-human species to have both the mitochondrial genome (Bibb *et al.* 1981) and the whole genome sequenced and published (Mouse Genome Sequencing Consortium 2002).

1.4.4. Mice on Islands

The 'Age of Discovery' which began in the early 15th century was the start of a long period of maritime expansion with major impacts on biodiversity in various regions of the world, through humans transporting species deliberately or passively around the globe. The deleterious effects of introduced plants and animals has been particularly notorious on native floras and faunas of remote islands, where in the past five centuries the number of biological invasions has increased dramatically, mostly human-mediated (Pascal *et al.* 2010). The introduction (accidental or intentional) of alien species onto islands has been

responsible for the reshaping of whole natural communities by disrupting ecosystem dynamics.

Certain mammal species thrive when introduced on islands, free of many ecological constraints present in their mainland native distribution. In extreme cases, the arrival of invasive species can cause the extinction of native forms, either by direct predation, herbivory, competition for the same resources or lack of defence mechanisms when facing the diseases and parasites carried by the invaders (Courchamp *et al.* 2003).

The western European house mouse (*Mus musculus domesticus*) is the perfect case-study to portray this as it constitutes one of the most successful and widespread mammals around the world and certainly one of the most successful island invaders (Angel *et al.* 2009). Darwin himself was aware of this reality, as highlighted in '*The Origin of Species*', where he acknowledged the high level of adaptability required for the survival of mice on islands with such distinct climates: " (…) The rat and mouse cannot be considered as domestic animals, but they have been transported by man to many parts of the world, and now have a far wider range than any other rodent, living free under the cold climate of Faroe in the north and of the Falklands in the south, and on many islands in the torrid zones (…)" (Darwin 1859). As Darwin indicates, the house mouse has reached and settled the most remote and challenging locations on the planet, from the Bering Sea region to sub-Antarctic islands (see Berry *et al.* 1978, Matthewson *et al.* 1994, Silver 1995, Cuthbert and Hilton 2004, Berry and Scriven 2005, van Vuuren and Chown 2007, Hardouin *et al.* 2010 and Jones *et al.* 2011b, 2012 for examples). The negative impacts of invasive house mice on island biotas (either direct or indirect) have been underestimated for a long time (Simberloff 2009). However, a growing

number of studies have now described a wide range of severe consequences associated with the presence of this mammal on remote islands (Le Roux *et al.* 2002, Smith *et al.* 2002, Wanless *et al.* 2007, Angel *et al.* 2009, Drake and Hunt 2009, Harris 2009, Hilton and Cuthbert 2010), particularly in the absence of its competitors and fellow invasives: the rats, *Rattus rattus* and *Rattus norvegicus* (Caut *et al.* 2007, Angel *et al.* 2009, Simberloff 2009). Mice have been associated with the extirpation and/or extinction of native species of plants, invertebrates and sea birds in ecosystems they have invaded and successfully occupied (Courchamp *et al.* 2003). One of the most iconic cases has been described in the remote South Atlantic Gough Island, where mice were shown to be implicated in the decline of Tristan albatrosses and Atlantic petrels through significant predation of their eggs and chicks (Wanless *et al.* 2007, Hilton and Cuthbert 2010). Similar evidence was also recently reported in the Sub-Antarctic Marion Island (Jones and Ryan 2010). On this same island, house mice have also been implicated in a mass mortality of fur seals presumably caused by the transmission of pathogens transported by the mice (de Bruyn *et al.* 2008).

In Australia, the introduced house mice have developed a unique trait that has been having catastrophic consequences for native biodiversity and agriculture for the past 100 years (Singleton *et al.* 2005). Despite the fact that mice only arrived about two centuries ago during European settlement, Australia quickly became the stage for the world's worst mouse plagues ever recorded in history. These plagues are characterised by widespread eruptions in grain growing areas, with densities reaching up to 3000 mice per hectare in grain stores (CSIRO 2003).

Despite their destructive features, island mice also provide opportunities for scientists to explore some of the driving mechanisms of evolutionary change on

islands. Basic questions in biology such as island gigantism and speciation have been addressed through the study of island mice. Both the St. Kilda and Faroese house mice have long been considered classic examples of extremely rapid morphological evolution. The first have even been granted species status, *Mus muralis* (Barrett-Hamilton 1899, Berry and Tricker 1969) on the basis of their large size and general robustness, when compared with continental mice.

Another extraordinary example of rampant evolution on island populations of house mice was described on the Atlantic island of Madeira. Here, six well differentiated chromosomal races thought to have arisen *in situ* were described, currently living on the island in parapatry (Britton-Davidian *et al.* 2000).

Despite the ubiquity and impact of these rodent invaders, until recent years the actual geographical origin of most oceanic island populations of the house mouse has remained unknown. However, important background information has become available. In the last couple of decades, studies have addressed the colonisation history of the western house mouse (*Mus musculus domesticus*) through a phylogeographical approach, with particular focus on the subspecies expansion into mainland western Europe and peripheral islands, and the Mediterranean basin area from the region of origin (Prager *et al.* 1993, 1996, 1998, Nachman *et al.* 1994, Gündüz *et al.* 2000, 2001, 2005, Ihle *et al.* 2006, Rajabi-Maham *et al.* 2008, Searle *et al.* 2009, Bonhomme *et al.* 2011, Jones *et al.* 2010, 2011a,b, 2012). As a consequence of all these studies, the number of available mtDNA sequences (mostly D-loop) has been increasing and the sampled geographical range has widened. This previous work is of vital importance considering that western Europe is the most likely source area of house mice for most oceanic islands worldwide. During the 15th – 17th centuries, countries like

the UK, France, the Netherlands, Portugal and Spain certainly played an important role in the spread of house mice all around the world, involuntarily transporting them as stowaways in their fleets.

The complex pattern of genetic variation observed in continental Europe and North Africa is not at all surprising considering that we are discussing a commensal species that mirrors human movements. On the mainland, the house mouse tends to be a lot more dependent of human-impacted habitats, i.e. a much stricter commensal, than house mice on islands. The house mouse is highly adaptable to new environments (e.g. has a wide diet, habitat breadth, short generation time, behavioural plasticity) (Ehrlich 1986) and very quickly is able to establish successful feral populations, even if only a small number of founder colonists successfully arrives at a mouse-free island. A pattern that has lately been observed in several oceanic islands is the resilience of mouse populations against re-invasion. After stable populations are established, secondary colonisation events seem to leave no (or limited) genetic signature on extant populations (see Förster *et al.* 2009, Hardouin *et al.* 2010, Jones *et al.* 2011b and Jones *et al.* 2012 for examples).

1.5 Aims and thesis structure

1.5.1 Aims

This thesis aimed to address the colonisation history of one of the most iconic invasive mammals on the planet, the house mouse, *Mus musculus*, based on molecular markers. The focus of this study was on the west European subspecies, *Mus musculus domesticus*, the most widespread taxon, at three different

geographical scales: regional (insular), continental and global. Accordingly, this thesis aimed to answer specific questions giving the scale under consideration:

i) To trace the origin of house mice on a remote Atlantic archipelago, the Azores, using both mitochondrial DNA and microsatellites data, comparing the results with previously studied Macaronesian archipelagos;

ii) To evaluate the level of post-colonisation gene flow among all Azorean islands;

iii) To assess the colonisation history of the Australian continent by the house mouse, comparing that colonisation pattern with the human historical details of the colonisation of Australia, taking advantage of earlier studies of a similar type relating to the colonisation of New Zealand;

iv) To draw a general picture on the global pattern of mitochondrial variation of the west European house mouse (*Mus musculus domesticus*) throughout its range. In particular we intended to assess, for the first time, the phylogeographic pattern of mtDNA D-loop outside the broadly screened Europe and Northern Africa, namely, in the Americas, sub-Saharan Africa and Australasia;

v) To assess the usefulness of the house mouse as a proxy for human movements and colonisations, relating the observed patterns of mitochondrial variation at different geographic scales (insular, continental and global) with aspects of documented human history.

1.5.2 Thesis structure

This thesis is composed of six chapters.

Chapter 1 corresponds to the 'General Introduction', where the research topic and key concepts are presented.

Chapter 2 consists of two papers (Papers I and II), both related to the colonisation history of the Azorean archipelago by the house mouse. The first (Paper I) involves an analysis of the colonisation of the nine Azorean islands based on mitochondrial D-loop data in the context of the vast previous knowledge using this marker [Gabriel SI, Mathias ML and Searle JB (submitted) Of mice and the 'Age of Discovery': the complex colonisation history of the Azorean archipelago by the house mouse (*Mus musculus*) as revealed by mitochondrial DNA variation]. The second paper (Paper II) covers an analysis of nuclear data, obtained from microsatellite markers, and is a significant progression from the previous paper in our understanding of the post-colonisation gene flow among the Azorean (and also Madeiran) islands [Gabriel SI, Mathias ML and Searle JB (submitted) Genetic structure of house mouse (*Mus musculus* Linnaeus 1758) populations in the Atlantic archipelago of the Azores: colonisation and dispersal].

Chapter 3 consists of a paper on the colonisation of Australia by the house mouse using mitochondrial D-loop data (Paper III). The results of this study have been published in the open access journal PLoS ONE [Gabriel SI, Stevens MI, Mathias ML and Searle JB (2011) Of mice and convicts: origin of the Australian house mouse, *Mus musculus*. *PLoS ONE*, 6, e28622]

Chapter 4 corresponds to the fourth paper (Paper IV), here presented as a manuscript under preparation with a general and global view of the pattern of mitochondrial D-loop variation of the western house mouse (*Mus musculus domesticus*), including a vast set of new samples from western Europe, Africa, the Americas, Australasia and numerous island worldwide.

Chapter 5 corresponds to an overview, in the form of a commentary (Paper V), on several studies where the colonisation history of the house mouse has been used as a proxy for human movements and colonisations. The particular emphasis of the commentary is on the colonisation of a remote archipelago in the south Indian Ocean but also provides a wider coverage of recent articles. It has been published in the open access journal *BMC Biology* and is classified as 'Highly Accessed' [Gabriel SI, Jóhannesdóttir F, Jones EP, Searle JB (2010) Colonization, mouse-style. *BMC Biology*, 8, 131].

Finally, **Chapter 6** corresponds to the 'General Discussion', where the main findings presented in the previous chapters are summarised and integrated. The main conclusions and implications of the overall results for the understanding of the colonisation process of the house mouse are also presented, as well as suggestions for future research on this topic.

1.6. References

Ajmone-Marsan P, Garcia JF, Lenstra JA (2010) On the origin of cattle: How aurochs became cattle and colonized the world. *Evolutionary Anthropology: Issues, News, and Reviews*, 19, 148–157.

Angel A, Wanless RM, Cooper J (2009) Review of impacts of the introduced house mouse on islands in the Southern Ocean: are mice equivalent to rats? *Biological Invasions*, 11, 1743–1754.

Ascunce MS, Yang C-C, Oakey J, Calcaterra L, Wu W-J, Shih C-J, Goudet J, Ross KG, Shoemaker DW (2011) Global invasion history of the fire ant *Solenopsis invicta*. *Science*, 331, 1066–1068.

Auffray J-C, Marshall JT, Thaler L, Bonhomme F (1990a) Focus on the nomenclature of European species of *Mus*. *Mouse Genome*, 88, 7–8.

Auffray J-C, Vanlerberghe F, Britton-Davidian J (1990b) The house mouse progression in Eurasia: A palaeontological and archaeozoological approach. *Biological Journal of the Linnean Society*, 41, 13–25.

Avise JC (1986) Mitochondrial DNA and the evolutionary genetics of higher animals. *Philosophical Transactions of the Royal Society, B* 312, 325–342.

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

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

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18, 489–522.

Baines JF, Harr B (2007) Reduced X–linked diversity in derived populations of house mice. *Genetics*, 175, 1911–1921.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, 13, 729–744.

Balloux F (2010) The worm in the fruit of the mitochondrial DNA tree. *Heredity*, 104, 419–420.

Barrett-Hamilton GEH (1899) On two recently described mice from St Kilda. *Annals of Scottish Natural History*, 31, 129–140.

Beheregaray LB (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, 17, 3754–3774.

Berry RJ, Jakobson ME, Peters J (1978) The house mice of the Faroe Islands: a study in microdifferentiation. *Journal of Zoology*, 185, 73–92.

Berry RJ, Scriven PN (2005) The house mouse: a model and motor for evolutionary understanding. *Biological Journal of the Linnean Society*, 84, 335–347.

Berry RJ, Tricker BJK (1969) Competition and extinction: the mice of Foula, with notes on those of Fair Isle and St Kilda. *Journal of Zoology*, 158, 247–265.

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell*, 26, 167–180.

Bilton DT, Mirol PM, Mascheretti S, Fredga K, Zima J, Searle JB (1998) Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonisation. *Proceedings of the Royal Society of London, B* 265, 1219–1226.

Bogenhagen DF (1999) Repair of mtDNA in vertebrates. *American Journal of Human Genetics*, 64, 1276–1281.

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffra J-C, 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.

Bonhomme F, Searle JB (in press) *House mouse phylogeography*. In *Evolution of the house mouse* (Cambridge Series in Morphology and Molecules). Eds Macholán M, Baird SJE, Munclinger P, Piálek J. Cambridge University Press, Cambridge.

Boursot P, Auffray J-C, Britton-Davidian J, Bonhomme F (1993) The evolution of house mice. *Annual Review of Ecology and Systematics*, 24, 119–152.

Boursot P, Din W, Anand R, Darviche D, Dod B, von Deimling F, Talwar GP, Bonhomme F (1996) Origin and radiation of the house mouse: mitochondrial DNA phylogeny. *Journal of Evolutionary Biology*, 9, 391–415.

Britton-Davidian, J, Catalan J, da Graça Ramalhinho M, Ganem G, Auffray JC, Capela R, Biscoito M, Searle JB, da Luz Mathias M (2000) Rapid chromosomal evolution in island mice. *Nature*, 403, 158.

Burbrink FT, Fontanella F, Pyron RA, Guiher TJ, Jimenez C (2008) Phylogeography across a continent: the evolutionary and demographic history of the North American Racer (Serpentes: Colubridae: *Coluber constrictor*). *Molecular Phylogenetics and Evolution*, 47, 274–288.

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature*, 325, 31–36.

Caut S, Casanovas JG, Virgos E, Lozano J, Witmer GW, Courchamp F (2007) Rats dying for mice: modelling the competitor release effect. *Austral Ecology*, 32, 858–868.

Charlton-Robb K, Gershwin L-A, Thompson R, Austin J, Owen K, McKechnie S (2011) A new dolphin species, the Burrunan Dolphin *Tursiops australis* sp. nov., endemic to Southern Australian coastal waters. *PLoS ONE*, 6, e24047.

Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, Kao RR, Pemberton JM, Beraldi D, Stear MJ, Alberti A, Pittau M, Iannuzzi L, Banabazi MH, Kazwala RR, Zhang YP, Arranz JJ, Ali BA, Wang Z, Uzun M, Dione MM,

Olsaker I, Holm LE, Saarma U, Ahmad S, Marzanov N, Eythorsdottir E, Holland MJ, Ajmone-Marsan P, Bruford MW, Kantanen J, Spencer TE, Palmarini M (2009) Revealing the history of sheep domestication using retrovirus integrations. *Science*, 324, 532–536.

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.

Courchamp F, Chapuis JL, Pascal M (2003) Mammal invaders on islands: impact, control and control impact. *Biological Reviews of the Cambridge Philosophical Society*, 78, 347–383.

Crawford MH, Rubicz RC, Zlojutro M (2010) Origins of Aleuts and the genetic structure of populations of the archipelago: Molecular and archaeological perspectives. *Human Biology*, 82, 695–717.

CSIRO (2003) *FAQs about mouse plagues*. CSIRO Sustainable Ecosystems, Canberra.

Cucchi T (2008) Uluburun shipwreck stowaway house mouse: molar shape analysis and indirect clues about the vessel's last journey. *Journal of Archaeological Science*, 35, 2953–2959.

Cucchi T, Vigne J-D (2006) Origin and diffusion of the house mouse in the Mediterranean. *Human Evolution*, 21, 95–106.

Cucchi T, Vigne J-D, Auffray J-C (2005) First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarcheaological revision of subfossil occurrences. *Biological Journal of the Linnean Society*, 84, 429–445.

Cuthbert R, Hilton G (2004) Introduced house mice *Mus musculus*: a significant predator of threatened and endemic birds on Gough Island, South Atlantic Ocean? *Biological Conservation*, 117, 483–489.

de Bruyn PJN, Bastos ADS, Eadie C, Tosh CA, Bester MN (2008) Mass mortality of adult male Subantarctic fur seals: are alien mice the culprits? *PLoS ONE*, 3, e3757.

Daly-Engel TS, Seraphin KD, Holland KN, Coffey JP, Nance HA, Toonen RJ, Bowen BW (2012) Global phylogeography with mixed–marker analysis reveals male–mediated dispersal in the endangered scalloped hammerhead shark (*Sphyrna lewini*). *PLoS ONE*, 7, e29986.

Darwin C (1859) *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray, London.

Deffontaine V, Libois R, Kotlík P, Sommer R, Nieberding C, Paradis E, Searle JB, Michaux JR (2005) Beyond the Mediterranean peninsulas: evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Molecular Ecology*, 14, 1727–1739.

Din W, Anand R, Boursot P, Darviche D, Dod B, Jouvin-Marche E, Orth A, Talwar GP, Cazanave P-A, Bonhomme F (1996) Origin and radiation of the house mouse: clues from nuclear genes. *Journal of Evolutionary Biology*, 9, 519–539.

Drake DR, Hunt TL (2009) Invasive rodents on islands: integrating historical and contemporary ecology. *Biological Invasions*, 11, 1483–1487.

Dubey S, Koyasu K, Parapanov R, Ribi M, Hutterer R, Vogel P (2008) Molecular phylogenetics reveals Messinian, Pliocene, and Pleistocene colonizations of islands by North African shrews. *Molecular Phylogenetics and Evolution*, 47, 877–882.

Duplantier J-M, Orth A, Catalan J, Bonhomme F (2002) Evidence for a mitochondrial lineage originating from the Arabian peninsula in the Madagascar house mouse (*Mus musculus*). *Heredity*, 89, 154–158.

Duvaux L, Belkhir K, Boulesteix M, Boursot P (2011) Isolation and gene flow: inferring the speciation history of European house mice. *Molecular Ecology*, 20, 5248–5264.

Eggert LS, Rasner CA, Woodruff DS (2002) The evolution and phylogeography of the African elephant inferred from mitochondrial DNA sequence and nuclear microsatellite markers. *Proceedings of the Royal Society*, *B* 269, 1993–2006.

Ehrlich PR (1986) Which animal will invade? In *Ecology of biological invasions of North America and Hawaii*. Edited by Mooney HA, Drake JA. Springer-Verlag, New York.

Eisen JA (1999) Mechanistic basis for microsatellite instability. In: *Microsatellites: evolution and applications*. Eds. Goldstein DB, Schlötterer C. Oxford University Press, Oxford.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5, 435–445.

Emerson BC (2002) Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology*, 11, 951–966.

Emerson BC, Hewitt GM (2005) Phylogeography. *Current Biology*, 15, R367–R371.

Fernández–Silva P, Enriquez JA, Montoya J (2003) Replication and transcription of mammalian mitochondrial DNA. *Experimental Physiology*, 88, 41–56.

Förster DW, Gündüz İ, Nunes AC, Gabriel S, Ramalhinho MG, Mathias ML, Britton-Davidian J, Searle JB (2009) Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology*, 18, 4477–4494.

Fox JG, Barthold SW, Davisson MT, Newcomer CE, Quimby FW, Smith AL (2007) *The mouse in biomedical research: normative biology, husbandry and models (Volume 3).* Academic Press, Burlington, Massachusetts.

Galtier N, Nabholz B, Glémin S and Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, 18, 4541–4550.

Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu H-T, Bulatova N, Ziv Y, Nachman MW (2008) Inferring the history of speciation in house mice from autosomal, X–linked, Y–linked and mitochondrial genes. *Molecular Ecology*, 17, 5349–5363.

Geraldes A, Basset P, Smith K, Nachman MW (2011) Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Molecular Ecology*, 20, 4722–4736.

Gongora J, Swan AB, Chong AY, Ho SYW, Damayanti CS, Kolomyjec S, Grant T, Miller E, Blair D, Furlan E, Gust N (2012) Genetic structure and phylogeography of platypuses revealed by mitochondrial DNA. *Journal of Zoology*, 286, 110–119.

Guénet JL (2005) The mouse genome. *Genome Research*, 15, 1729–1740.

Guénet JL, Bonhomme F (2003) Wild mice: an ever–increasing contribution to a popular mammalian model. *Trends in Genetics*, 19, 24–31.

Gündüz İ, Auffray J-C, Britton-Davidian J, Catalan J, Ganem G, Ramalhinho MG, Mathias ML, Searle JB (2001) Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology*, 10, 2023–2029.

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

Gündüz İ, Rambau RV, Tez C, Searle JB (2005) Mitochondrial DNA variation in the western house mouse (*Mus musculus domesticus*) close to its site of origin: studies in Turkey. *Biological Journal of the Linnean Society*, 84, 473–485.

Gündüz İ, Tez C, Malikov V, Vaziri A, Polyakov A, Searle JB (2000) Mitochondrial DNA and chromosomal studies of wild mice (*Mus*) from Turkey and Iran. *Heredity*, 84, 458–467.

Hardouin EA, Chapuis J-L, Stevens MI, van Vuuren BJ, Quillfeldt P, Scavetta RJ, Teschke M, Tautz D (2010) House mouse colonization patterns on the sub-

Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evolutionary Biology*, 10, 325.

Harris DB (2009) Review of negative effects of introduced rodents on small mammals on islands. *Biological Invasions*, 11, 1611–1630.

Hartl DL, Clark AG (2006) *Principles of population genetics (4th Edition)*. Sinauer Associates, Sunderland, Massachusetts.

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.

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

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

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

Hewitt GM (2004a) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society*, *B* 359, 183–195.

Hewitt GM (2004b) The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology*, 1, 4.

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010) Phylogeography's past, present, and future: 10 years after. *Molecular Phylogenetics and Evolution*, 54, 291–301.

Hilton GM, Cuthbert RJ (2010) The catastrophic impact of invasive mammalian predators on birds of the UK Overseas Territories: a review and synthesis. *Ibis*, 152, 443–458.

Hofreiter M, Serre D, Rohland N, Rabeder G, Nagel D, Conard N, Münzel S, Pääbo S (2004) Lack of phylogeography in European mammals before the last glaciation. *Proceedings of the National Academy of Science USA*, 101, 12963–12968.

Ihle S, Ravaoarimanana I, Thomas M, Tautz D (2006) An analysis of signatures of selective sweeps in natural populations of the house mouse. *Molecular Biology and Evolution*, 23, 790–797.

Innan H, Nordborg M (2002) Recombination or mutational hot spots in human mtDNA? *Molecular Biology and Evolution*, 19, 1122–1127.

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

Jones EP, Jóhannesdóttir F, Gündüz İ, Richards MB, Searle JB (2011a) The expansion of the house mouse into north–western Europe. *Journal of Zoology*, 283, 257–268.

Jones EP, Jensen J-K, Magnussen E, Gregersen N, Hansen HS, Searle JB (2011b) A molecular characterization of the charismatic Faroe house mouse. *Biological Journal of the Linnean Society*, 102, 471–482.

Jones EP, Skirnisson K, McGovern TH, Gilbert MTP, Willerslev E, Searle JB (2012) Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *BMC Evolutionary Biology*, 12, 35.

Jones EP, van der Kooij J, Solheim R, Searle JB (2010) Norwegian house mice (*Mus musculus musculus/domesticus*): Distributions, routes of colonization and patterns of hybridization. *Molecular Ecology*, 19, 5252–5264.

Jones MGW, Ryan PG (2010) Evidence of mouse attacks on albatross chicks on sub–Antarctic Marion Island. *Antarctic Science*, 22, 39–42.

Knopp T, Merilä J (2008). The postglacial recolonization of northern Europe by *Rana arvalis* as revealed by microsatellite and mitochondrial DNA analyses. *Heredity*, 102, 174–181.

Kvist L, Martens J, Nazarenko AA, Orell M (2003) Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Molecular Biology and Evolution*, 20, 243–247.

Ladoukakis ED, Zouros E (2001). Animal mitochondrial DNA recombination: evidence from published sequences. *Molecular Biology and Evolution*, 18, 2127–2131.

Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, 307, 1618–1621.

Le Roux V, Chapuis J-L, Frenot Y, Vernon P (2002) Diet of the house mouse (*Mus musculus*) on Guillou island, Kerguelen archipelago, Subantarctic. *Polar Biology*, 25, 49–57.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions, and mutational mechanisms: a review. *Molecular Ecology*, 11, 2453–2465.

Linnaeus C (1758) *Systema Naturae*. Staples Printing Group, Woking, Surrey (Photographic Facsimile).

Luo S-J, Kim J-H, Johnson WE, Walt Jvd, Martenson J, Yuhki N, Miquelle DG, Uphyrkina O, Goodrich JM, Quigley HB, Tilson R, Brady G, Martelli P, Subramaniam V, McDougal C, Hean S, Huang S-Q, Pan W, Karanth UK, Sunquist M, Smith JLD, O'Brien SJ (2004) Phylogeography and genetic ancestry of tigers (*Panthera tigris*). *PLoS Biology*, 2, e442.

Lynch M, Koskella B, Schaack S (2006) Mutation pressure and the evolution of organelle genomic architecture. *Science*, 311, 1727–1730.

Macholán M, Munclinger P, Sugerkova M, Dufková P, Bímová B, Božíková E, Zima J, Piálek J (2007) Genetic analysis of autosomal and X–linked markers across a mouse hybrid zone. *Evolution*, 61, 746–771.

Marshall JT, Sage RD (1981) Taxonomy of the house mouse. *Symposia of the Zoological Society of London*, 47, 15–25.

Matthewson DC, van Aarde RJ, Skinner JD (1994) Population biology of house mouse (*Mus musculus* L.) on sub-Antarctic Marion Island. *South African Journal of Zoology*, 29, 99–106.

Matisoo-Smith E (2002) Something old, something new: do genetic studies of contemporary populations reliably represent prehistoric populations of Pacific *Rattus exulans*? *Human Biology*, 74, 489–496.

Matisoo-Smith E (2009) The commensal model for human settlement of the Pacific 10 years on — what can we say and where to now? *The Journal of Island and Coastal Archaeology*, 4, 151–163.

Matisoo-Smith E, Roberts RM, Irwin GJ, Allen JS, Penny D, Lambert DM (1998) Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proceedings of the National Academy of Sciences, USA*, 95, 15145–15150.

Matisoo-Smith E, Robins JH (2004) Origins and dispersals of Pacific peoples: evidence from the mtDNA phylogenies of the Pacific rat. *Proceedings of the National Academy of Sciences, USA*, 101, 9167–9172.

Matisoo-Smith E, Robins JH (2009) Mitochondrial DNA evidence for the spread of Pacific rats through Oceania. *Biological Invasions*, 11, 1521–1527.

McDevitt AD, Vega R, Rambau RV, Yannic G, Herman JS, Hayden TJ, Searle JB (2011) Colonisation of Ireland – revisiting 'the pygmy shrew syndrome' using mitochondrial, Y chromosomal and microsatellite markers. *Heredity*, 107, 548–557.

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.

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.

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420, 520–562.

Nachman MW, Boyer SN, Searle JB, Aquadro CF (1994) Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. *Genetics*, 136, 1105–1120.

Nunome M, Ishimoh C, Aplin KP, Tsuchiya K, Yonekawa H, Mohwaki K, Susuki H (2010) Detection of recombinant haplotypes in wild mice (*Mus musculus*) provides new insights into the origin of Japanese mice. *Molecular Ecology*, 19, 2474–2489.

Ojeda AA (2010) Phylogeography and genetic variation in the South American rodent *Tympanoctomys barrerae* (Rodentia: Octodontidae). *Journal of Mammalogy*, 91, 302–313.

Oppenheimer S (2012) Out–of–Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. Philosophical Transactions of the Royal Society, B 367, 770–784.

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.

Park CB, Larsson N-G (2011) Mitochondrial DNA mutations in disease and aging. *The Journal of Cell Biology*, 193, 809–818.

Pascal M, Le Guyader H, Simberloff D (2010) Biological invasions and the conservation of biodiversity. *OIE Revue Scientifique et Technique*, 29, 367–403.

Pineda MC, López-Legentil S, Turon X (2011) The whereabouts of an ancient wanderer: global phylogeography of the solitary ascidian *Styela plicata*. *PLoS ONE*, 6, e25495.

Pocock MJO, Hauffe HC, Searle JB (2005) Dispersal in house mice. *Biological Journal of the Linnean Society*, 84, 565–583.

Prager EM, Orrego C, Sage RD (1998) Genetic variation and phylogeography of Central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics*, 150, 835–861.

Prager EM, Sage RD, Gyllensten U, Thomas WK, Hübner R, Jones CS, Noble L, Searle JB, Wilson AC (1993) Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biological Journal of the Linnean Society*, 50, 85–122.

Prager EM, Tichy H, Sage RD (1996) Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: comparison with other house mice and report of a 75-bp tandem repeat. *Genetics*, 143, 427–446.

Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe, *Molecular Ecology*, 17, 627–641.

Randi E, Alves PC, Carranza J, Milosevic-Zlatanovic S, Sfougaris A, Mucci N (2004) Phylogeography of roe deer (*Capreolus capreolus*) populations: the effects of historical genetic subdivisions and recent nonequilibrium dynamics. *Molecular Ecology*, 13, 3071–3083.

Raufaste N, Orth A, Belkhir K, Senet D, Smadja C, Baird SJE, Bonhomme F, Dod B, Boursot P (2005) Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society*, 84, 593–616.

Rosenthal N, Brown S (2007) The mouse ascending: perspectives for human–disease models. *Nature Cell Biology*, 9, 993–999.

Rubinoff D, Holland BS (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology*, 54, 952–961.

Sage RD, Atchley WR, Capanna E (1993) House mice as models in systematic biology. *Systematic Biology*, 42, 523–561.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma,* 109, 365–371.

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

Searle JB (2008) The genetics of mammalian invasions: a review. *Wildlife Research*, 35, 185–192.

Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry S, Giménez MD, Jóhannesdóttir F (2009) Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society, B* 276, 201–207.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, 9, 615–629.

Shadel GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. *Annual Reviews of Biochemistry*, 66, 409–435.

Siahsarvie R, Auffray J–C, Darvish J, Rajabi–Maham H, Vu H–T, Agret S, Bonhomme F, Claude J (2012) Patterns of morphological evolution in the mandible of the house mouse *Mus musculus* (Rodentia: Muridae). *Biological Journal of the Linnean Society*, 105, 635–647.

Singleton GR, Brown PR, Pech RP, Jacob J, Mutze GJ, Krebs CJ (2005) One hundred years of eruptions of house mice in Australia – a natural biological curio. *Biological Journal of the Linnean Society*, 84, 617–627.

Silver LM (1995) *Mouse genetics, concepts and applications*. Oxford University Press, Oxford.

Simberloff D (2009) Rats are not the only introduced rodents producing ecosystem impacts on islands. *Biological Invasions*, 11, 1735–1742.

Smith VR, Avenant NL, Chown SL (2002) The diet and impact of house mice on a Subantarctic island. *Polar Biology*, 25, 703–715.

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

Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist*, 166, 581–591.

Taberlet P, Fumagalli L, Wust-Saucy A, Cosson J (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 8, 1923–1934.

Templeton AR (2002) Out of Africa again and again. *Nature*, 416, 45–51.

Teeter KC, Payseur BA, Harris LW, Bakewell MA, Thibodeau LM, O'Brien JE, Krenz JG, Sans–Fuentes MA, Nachman MW, Tucker PK (2008) Genome–wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, 18, 67–76.

Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, Tucker PK (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, 64, 472–485.

van Vuuren B, Chown SL (2007) Genetic evidence confirms the origin of the house mouse on sub-Antarctic Marion Island. *Polar Biology*, 30, 327–332.

Vega R, Fløjgaard C, Lira-Noriega A, Nakazawa Y, Svenning J-C, Searle JB (2010) Northern glacial refugia for the pygmy shrew *Sorex minutus* in Europe revealed by phylogeographic analyses and species distribution modelling. *Ecography*, 33, 260–271.

Wang L, Luzynski K, Pool JE, Janoušek V, Dufková P, Vyskočilová MM, Teeter KC, Nachman MW, Munclinger P, Macholán M, Piálek J, Tucker PK (2011) Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. *Molecular Ecology*, 20, 2985–3000.

Wanless RM, Angel A, Cuthbert RJ, Hilton GM, Ryan PG (2007) Can predation by invasive mice drive seabird extinctions? *Biology Letters*, 3, 241–244.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics*, 2, 1123–1128.

White TA, Searle JB (2008) The colonization of Scottish islands by the common shrew, *Sorex araneus* (Eulipotyphla: Soricidae). *Biological Journal of the Linnean Society*, 94, 797–808.

White DJ, Wolff JN, Pierson M, Gemmell NJ (2008) Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, 17, 4925–4942.

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.

Yang H, Bell TA, Churchill GA, Pardo-Manuel de Villena F (2007) On the subspecific origin of the laboratory mouse. *Nature Genetics*, 39, 1100–1107.

Yang H, Wang JR, Didion JP, Buus RJ, Bell TA, Welsh CE, Bonhomme F, Yu AH, Nachman MW, Piálek J, Tucker P, Boursot P, McMillan L, Churchill GA, de Villena FP (2011) Subspecific origin and haplotype diversity in the laboratory mouse. *Nature Genetics*, 43, 648–655.

Yonekawa H, Moriwaki K, Gotoh O, Miyashita N, Matsushima Y, Shi LM, Cho WS, Zhen XL, Tagashira Y (1988) Hybrid origin of Japanese mice "*Mus musculus molossinus*": evidence from restriction analysis of mitochondrial DNA. *Molecular Biology and Evolution*, 5, 63–78.

Zhao X, Li N, Guo W, Hu X, Liu Z, Gong G, Wang A, Feng J, Wu C (2004) Further evidence for paternal inheritance of mitochondrial DNA in the sheep (*Ovis aries*). *Heredity*, 93, 399–403.

Zeder MA, Hesse B (2000) The initial domestication of goats (*Capra hircus*) in the Zagros Mountains 10,000 years ago. *Science*, 28, 2254–2257.

Chapter 2

Paper I

Of mice and the 'Age of Discovery': the complex colonization history of the Azorean archipelago by the house mouse (*Mus musculus*) as revealed by mitochondrial DNA variation

---

Sofia I Gabriel[1,2], Maria da Luz Mathias[1] and Jeremy B Searle[2,3]

[1] CESAM – Centre for Environmental and Marine Studies, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, 1749–016 Lisbon, Portugal

[2] Department of Biology, University of York, PO Box 373, York YO10 5YW, UK

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

Manuscript submitted to '*Journal of Evolutionary Biology*'

Abstract

Humans have introduced many species onto remote oceanic islands. The house mouse (*Mus musculus*) is a human commensal and has consequently been transported to oceanic islands around the globe as an accidental stowaway. The history of these introductions can tell us not only about the mice themselves but also about the people that transported them. Following a phylogeographic approach, we used mitochondrial D-loop sequence variation to study house mouse colonization of the Azores. A total of 239 sequences were obtained from all nine islands, and previous Iberian sequences and 70 new Spanish sequences helped interpretation. A Bayesian analysis revealed presence in the Azores of most of the D-loop clades previously described in the *domesticus* subspecies of the house mouse, suggesting a complex colonization history of the archipelago as a whole from multiple geographical origins, but much less heterogeneity (often single colonization?) within islands. The expected historical link with mainland Portugal was reflected in the pattern of D-loop variation of some of the islands but not all. A more unexpected association with a distant North European source area was also detected in three islands, possibly reflecting human contact with the Azores prior to the 15th century discovery by Portuguese mariners. Widening the scope to colonization of the Macaronesian islands, human linkages between the Azores, Madeira, the Canaries, Portugal and Spain were revealed through the sharing of mouse sequences between these areas. From these and other data we suggest mouse studies may help resolve historical uncertainties relating to the 'Age of Discovery'.

Introduction

Humans have been transporting numerous species outside their normal range for millennia (Hulme 2009), either accidentally or deliberately (Suarez and Tsutsui 2008). Such alien species can have particularly dramatic impacts as invasions on isolated ecosystems such as oceanic islands, being an important cause of extinction (Donlan and Wilcox 2008). Mice and rats are some of the most iconic invasive species worldwide and amongst the most common accidentally-introduced vertebrates on oceanic islands (Harris 2009), having accompanied humans since the early days of ocean-going navigation (Cucchi 2008). The unintentional introduction of commensal rodents has indisputably harmful consequences to insular ecosystems (Harris 2009, Drake and Hunt 2009, Tollenaere *et al.* 2010). But in terms of scientific enquiry, island invasions represent ecological and evolutionary experiments on the effects of the arrival and subsequent adaptation of invasives to their new environment (Michaux *et al.* 2007, Lawson-Handley *et al.* 2011). The history of colonization of introduced commensals also provides insights into the history of movements of the humans with whom they have travelled as stowaways (Searle 2008, Matisoo-Smith & Robin 2009, Tollenaere *et al.* 2010). Archipelagos provide a particularly interesting setting to study this human-mediated invasion process, affording a multiplicity of colonization scenarios and routes.

The house mouse (*Mus musculus*) is extraordinary not only for the numbers of oceanic islands on which it occurs (Angel *et al.* 2009), but also for the diversity of environment conditions it can tolerate (see Berry and Peters 1975, Berry *et al.* 1978, Matthewson *et al.* 1994, Cuthbert and Hilton 2004, van Vuuren and Chown

2007, Hardouin *et al.* 2010 and Jones *et al.* 2011b, 2012 for examples). Over the years, several studies have been able to establish links between the colonization of mice and humans by analysing mitochondrial D-loop sequences of insular populations of house mice (e.g. Gündüz *et al.* 2001, Duplantier *et al.* 2002, van Vuuren and Chown 2007, Förster *et al.* 2009, Searle *et al.* 2009a,b, Jones *et al.* 2010, 2011b, Hardouin *et al.* 2010, Gabriel *et al.* 2011). Among these, one of the most unexpected results arose from the work of Gündüz *et al.* (2001) and Förster *et al.* (2009) while analysing the house mouse populations of the Atlantic archipelago of Madeira. Their results revealed an extraordinary genetic similarity between the mice of Madeira and northern Europe despite, in terms of human history, by far the main cultural link with Madeira being mainland Portugal starting with the Portuguese explorations in the 15th century (Albuquerque and Vieira 1987). To explain the northern European connection, Gündüz *et al.* and Förster *et al.* suggested that Danish Vikings may have visited Madeira after having been blown off-course their route along the coast of Iberia, and inadvertently introduced a mouse population in advance of any accidental transport of mice from Portugal. As there are no historical records of a Viking presence on Madeira, the genetic data on house mice represent the only current evidence of the event (Gündüz *et al.* 2001, Förster *et al.* 2009).

Like the Madeira archipelago, the islands of the Azores were officially discovered by the Portuguese in the first half of the 15th century (Frutuoso 1590). However, in 1351, nearly a century earlier, a group of islands consistent with the location of the Azores were already placed on Genovese maps (Kelley 1979), giving rise to speculation on whether Phoenician, Scandinavian or Arabic navigators may have visited the Azores before the Portuguese arrival.

The official discovery of the Azores was part of an effort to find new trade routes and colonies overseas, which involved both Portugal and later Spain (Greene and Morgan 2008). However, during the early stages of this 'Age of Discovery', both Portugal and Spain were challenging each other's supremacy and it was strategic to keep new discoveries secret (Diffie and Winius 1977), contributing to the uncertainty of first arrivals at locations such as the Azores.

In terms of long-distance colonization of commensal species transported by humans, the Azores are one of the most intriguing archipelagos to study. With their strategic location in the middle of the Atlantic Ocean, the Azorean islands have been a popular stop-over that explorers, traders, settlers, whalers and pirates have used since colonial times. In this context, countless ships potentially infested with mice reached the islands originating from various parts of Europe. Today, the western subspecies of house mouse (*M. m. domesticus*) is one of the most abundant and widely distributed mammals in the archipelago. It is present in all nine islands (Mathias *et al.* 1998), having been recorded as common since the establishment of the first human settlements in the 15th century (Drouët 1861).

In this study, we investigate the colonization of the Azores using a phylogeographic approach (Avise 2000). By comparing D-loop sequences of mice from the Azores with mice from potential source areas, we consider whether the mice came from mainland Portugal or from some other origin. The extent to which the individual islands of the Azores were colonized independently or from other islands in the archipelago should also be reflected in the DNA variation of the extant island mouse populations.

The fast-evolving D-loop is by far the most extensively studied molecular marker in house mouse colonization studies (Prager *et al.* 1993, 1996, 1998;

Nachman *et al.* 1994; Gündüz *et al.* 2000, 2001, 2005; Ihle *et al.* 2006; Rajabi-Maham *et al.* 2008, Förster *et al.* 2009; Searle *et al.* 2009a, b; Jones *et al.* 2010, 2011a,b, 2012; Bonhomme *et al.* 2011; Hardouin *et al.* 2010; Gabriel *et al.* 2011). There is evidence that the D-loop is particularly good to study initial colonization of an area, with the marker (being maternally inherited) reflecting a lack of penetrance of later-arriving female mice into pre-existing populations (Förster *et al.* 2009, Gabriel *et al.* 2010, Bonhomme and Searle in press). Although the coverage of D-loop sequences from most of western Europe (including mainland Portugal) is excellent, there is a shortage of sequences from southern Spain where there were important ports for Spanish activities during the Age of Discovery (O'Flanagan 2008). Therefore, this paper presents new sequences from southern Spain as well as all nine Azorean islands. This work is put into the wider context of the role of northern Europe, Spain and Portugal in the colonization of Macaronesia (Madeira, the Azores and the Canaries) by house mice.

Materials and Methods

Sampling areas

Azores

The nine volcanic islands of the Azores are located approximately 1500 km from the European coast and 3900 km from the North American continent. The archipelago lies along a 600 km line running northwest-southeast between longitude -25.02º to 31.11º West and latitude 36.96º to 39.45º North. The islands are divided into three major geographic groups: 'Eastern' including Santa Maria and São Miguel; 'Central' including Terceira, Graciosa, São Jorge, Pico and Faial;

and 'Western' comprising Flores and Corvo (Figure 1). Altogether we analysed 239 mice from these nine islands collected during 2005 – 2007 (see Table S1, Supplementary Material for sample details).

Southern Spain

During the Age of Discovery, most of the Spanish documented expeditions had a departure point in the southern coast of Spain, usually from the historic ports of Palos de la Frontera or Cádiz, due to their strategic location on the Atlantic seaboard. Here we sampled this region for the first time, a distance of 200 km along the coast of Andalusia, from Aljaraque, Huelva (37.29º; -7.03º) to Algeciras, Cádiz (36.15º; -5.48º), collecting 70 mice from 15 localities during 2008 (Figure 1A; Table S1, Supplementary Material).

DNA extraction, amplification and sequencing

Genomic DNA was extracted from tail tips stored in ethanol at 4ºC using DNeasy™ Blood & Tissue Kits (Qiagen). The entire D-loop and flanking regions were amplified by Polymerase Chain Reaction (PCR) using the primer pair L15774 5' CGC CTG TTT ATC AAA AAC AT 3' and H2228 5' CTC CGG TTT GAA CTC AGA TC 3' as described in Searle *et al.* (2009a) and truncated between positions 15424 and 16276 of Bibb *et al.* (1981) to align with previously published sequences. The PCR protocol consisted of an initial 5 min denaturation step at 94ºC, followed by 35 cycles of denaturation at 94ºC for 1 min, annealing at 55ºC for 1 min, extension at 72ºC for 1 min, and a final 8 min extension step at 72ºC. Products resulting from successful amplifications were purified using the QIAquick PCR Purification Kit

(Qiagen) following the manufacturer's recommendations. Both strands were sequenced using the PCR primers by standard sequencing technology (Macrogen Inc., Korea).

Sequence data analyses

All sequence traces were checked for ambiguous bases in Sequencher v. 4.5 (Gene Codes Corp.) and aligned by eye using Bioedit v7.1.3.0. with 520 previously published *M. m. domesticus* haplotypes downloaded from GenBank (Prager *et al.* 1993, 1996, 1998; Nachman *et al.* 1994; Gündüz *et al.* 2000, 2001, 2005; Ihle *et al.* 2006; Rajabi-Maham *et al.* 2008; Förster *et al.* 2009; Searle *et al.* 2009a, b; Hardouin *et al.* 2010; Jones *et al.* 2010, 2011a,b, 2012; Gabriel *et al.* 2011) or directly from the publication (mouse sequences in Bonhomme *et al.* 2011, from countries neighbouring the Atlantic). All sequences resulting in new D-loop haplotypes were deposited in GenBank (accession numbers: xxxxx–xxxxx).

The combined haplotypes represent a substantial coverage of the geographical distribution of *M. m. domesticus* in western Europe and were used in the phylogenetic analysis along with two previously published *M. m. castaneus* (accession numbers AJ286322 and AF088879) and *M. m. musculus* (U47532 and U47504) sequences as outgroups (following Jones *et al.* 2011a).

The software jModelTest version 0.1.1 (Posada 2008) was used to determine the model of sequence evolution that best-fit the D-loop dataset (without outgroups). SYM+I+Γ was selected from 88 available models under the Akaike information criterion (AIC) method (Posada & Buckley 2004).

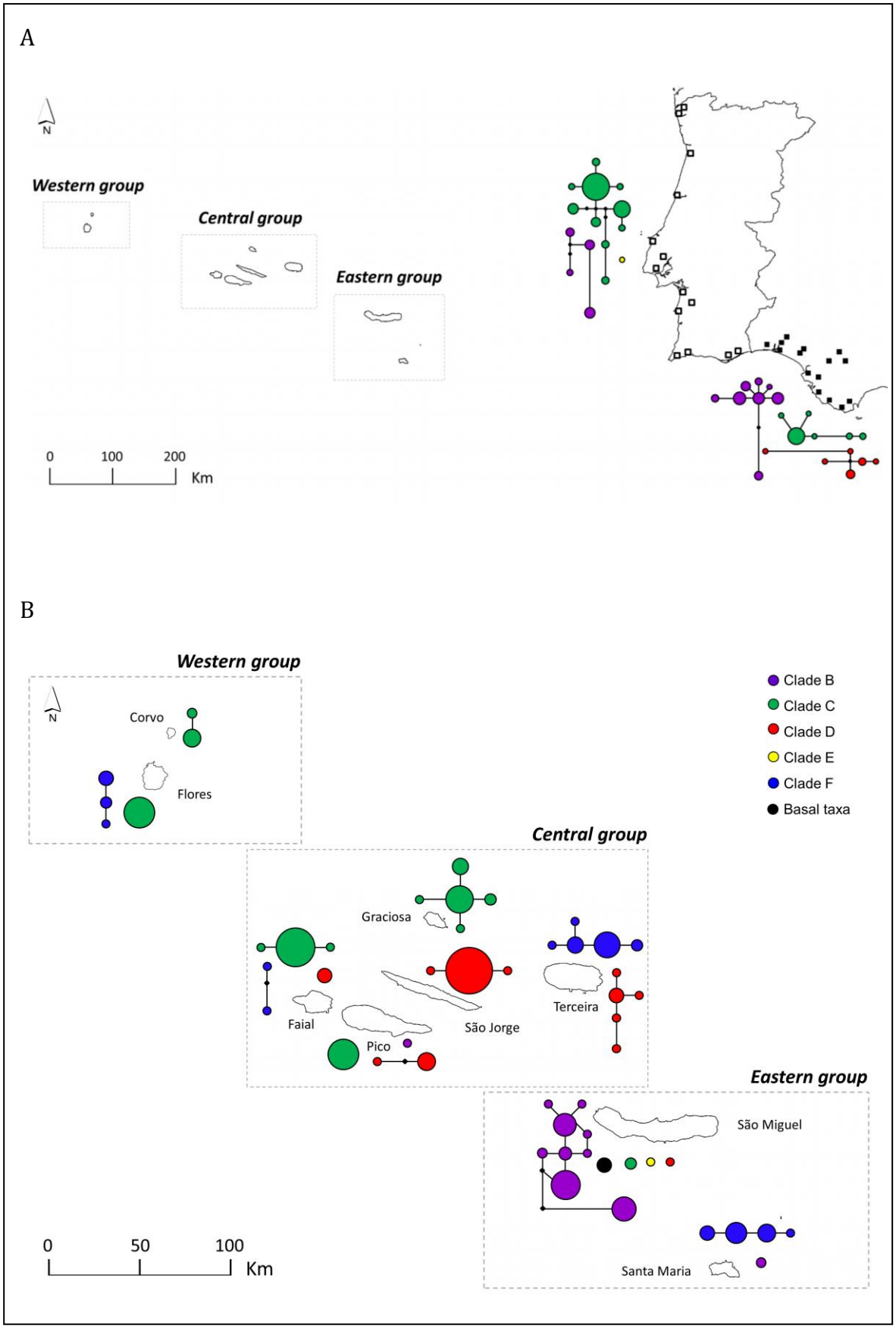

A
N
Western group
Central group
Eastern group
0
100
200
Km
B
Western group
N
Corvo
Flores
Clade B
Clade C
Clade D
Clade E
Clade F
Basal taxa
Central group
Graciosa
Faial
Pico
São Jorge
Terceira
Eastern group
São Miguel
Santa Maria
0
50
100
Km

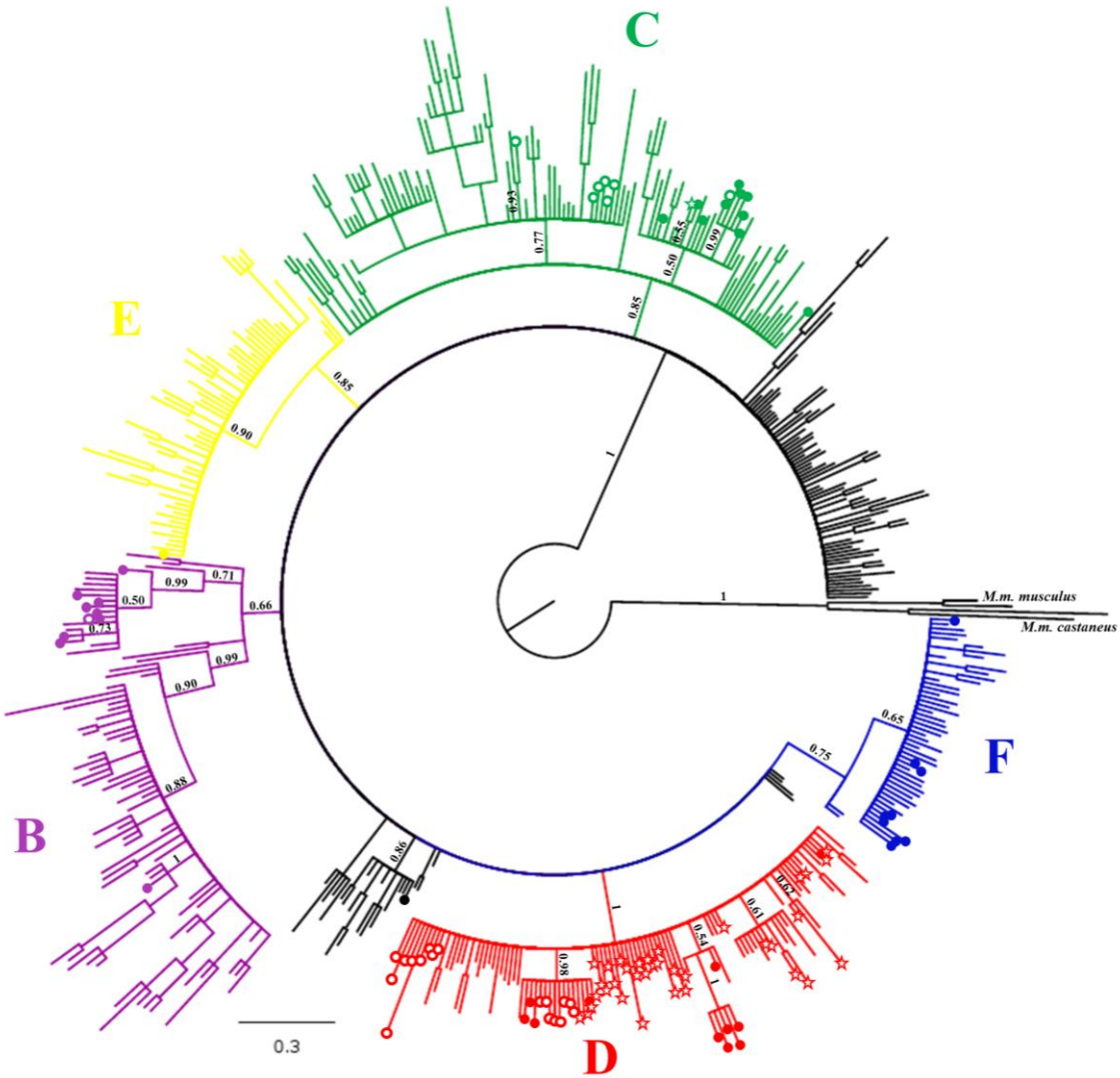
Figure 1. Median-joining haplotype networks obtained for house mouse D-loop sequences from Portugal, southern Spain and the Azores, displayed by clade (identified by colours relating to Figure 2). Each circle represents a unique haplotype and circle area is proportional to the frequency of occurrence. Solid lines connecting haplotypes represent mutational differences and solid black diamonds represent hypothetical historical haplotypes or current unsampled haplotypes. Details of new sampling locations in the Azores and southern Spain can be found in Table S1, Supplementary Material. **A.** Map of mainland Portugal and southern Spain (Huelva and Cadiz provinces) showing collection localities of all samples of house mouse, *Mus musculus domesticus*. Open squares: mainland Portugal (Förster *et al.* 2009); closed squares: southern Spain (this study). The distance between the Azores and mainland has been reduced for ease of visualisation. **B.** Map of the Azores showing the median-joining networks (per clade) of haplotypes found on each island. The distance between the Azores island groups has been reduced for ease of visualisation.

Phylogenetic reconstruction was carried out using Bayesian inference as implemented in MrBayes 3.2 (Ronquist and Huelsenbeck 2003). Two simultaneous MCMC (Markov Chain Monte Carlo) runs starting from random trees, each search consisting of 5 chains (1 cold and 4 heated, using an incremental heating parameter of 0.02), were run for 25 million generations, sampled every 1000 steps. After checking for convergence of both runs using the number of generations versus a log likelihood plot and PRSF statistics, the first 10,000 trees were discarded as burn-in (40%), to ensure tree sampling from a stationary posterior distribution. The remaining trees were summarised as a 50% majority rule consensus tree with nodal support expressed as posterior probabilities.

Over the years, the clades recovered in the D-loop phylogenies have been named in various ways. Jones *et al.* (2011a) suggested a unifying classification based on a letter-number labelling system long used in human nomenclature (Richards *et al.* 1998), which we will follow from here on.

Phylogenetic networks of the Azorean haplotypes were constructed under the Median Joining algorithm as implemented in Network 4.5.1.6 (Bandelt *et al.* 1999).

A)



B)

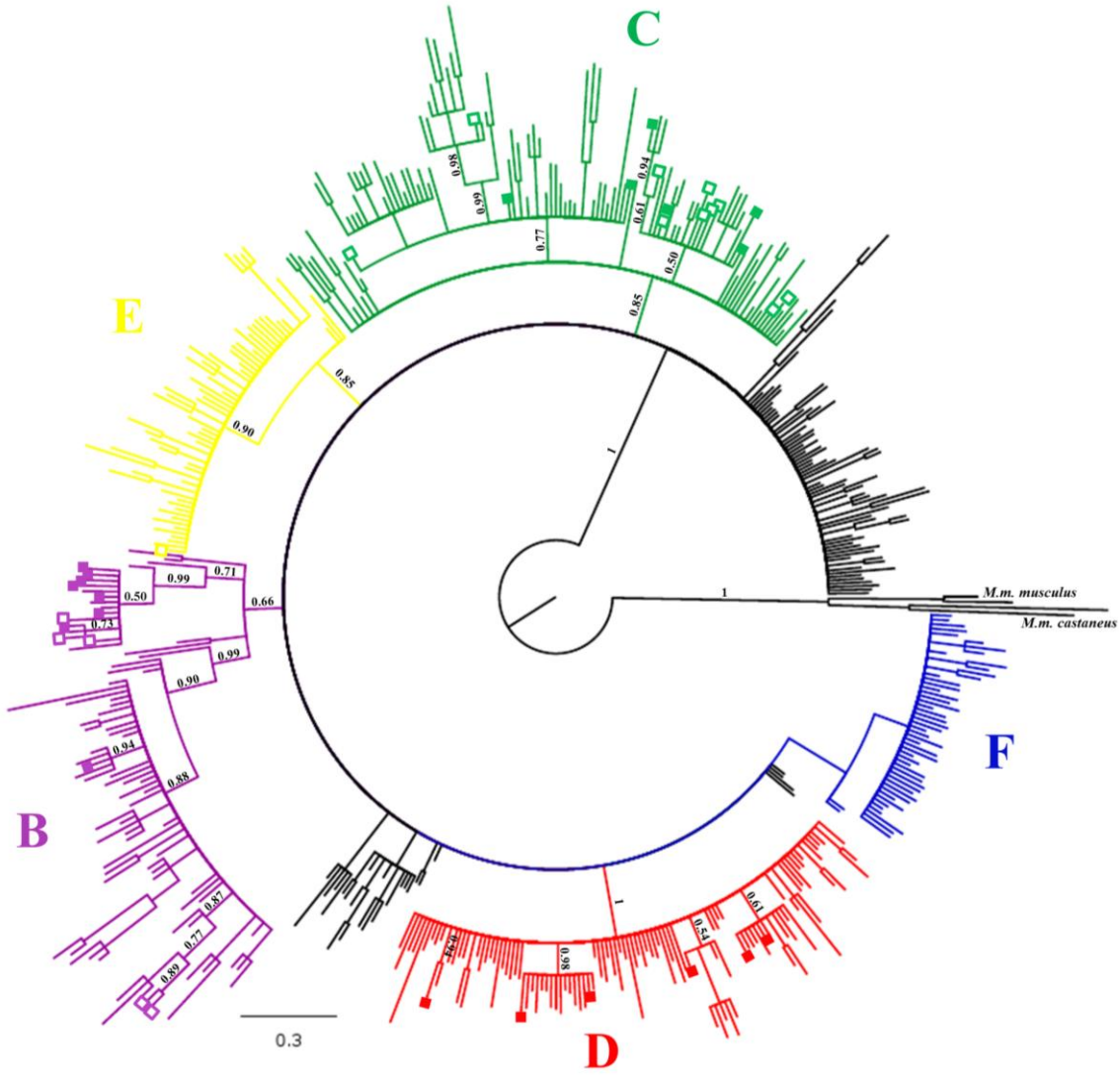


Figure 2. Bayesian phylogenetic tree of previously published house mouse D-loop haplotypes in addition to our new Azorean and Spanish sequences. **A.** Haplotypes detected on the three Macaronesian archipelagos are highlighted with the following symbols: closed circles (Azores, this study), open stars (Madeira, Gündüz *et al.* 2001, Förster *et al.* 2009) and open circles (Canaries, Bonhomme *et al.* 2011, Jones et al. 2011a). **B.** Haplotypes detected in southern Spain (this study) and mainland Portugal (Förster *et al.* 2009) are highlighted with closed and open squares, respectively. See Figure S1, Supplementary Material for the detailed tree.

Population genetic analysis

Genetic diversity parameters (nucleotide diversity, π and haplotype diversity, *h*) were estimated as described by Tamura and Nei (1993) in Arlequin 3.5.1.2 (Excoffier and Lischer 2010). Neutrality tests, D (Tajima 1989), F_S (Fu 1997) and R_2 (Ramos-Onsins and Rozas 2002), were estimated in DnaSP v.5.10.1 (Librado and Rozas 2009). Significantly negative values of D and F_S suggest the occurrence of population expansion, as well as significantly positive values of R_2. The levels of statistical significance of observed values of D, F_S and R_2 were obtained by comparisons to expected values obtained from 10,000 coalescent simulations (Librado and Rozas 2009). In order to assess the extent of differentiation between mtDNA sequences from insular and mainland populations, pairwise Φ_{ST} values were computed in Arlequin 3.5.1.2. Statistical significance was estimated using a simulation with 10,000 permutations (Excoffier *et al.* 1992).

Results

A total of 309 D-loop sequences were obtained from individuals that were morphologically assigned to the house mouse, *Mus musculus*, 305 of which could be classified as the western subspecies, *M. m. domesticus*, on the basis of their D-loop sequences. Four specimens collected in a single locality in southern Spain (Table S1, Supplementary Material) had D-loop sequences allocated to *M. spretus* and were removed for further analysis. These specimens had a nuclear gene sequence consistent with their *M. musculus* morphology (Barbosa *et al.* submitted), confirming their hybrid status.

Southern Spain

The 66 D-loop sequences grouped into 20 haplotypes within clades B, C and D (Figure 2B) and the haplotype diversity *h* ± SD was 0.927 ± 0.013 and nucleotide diversity π ± SD was 0.0118 ± 0.0060 (see Table 1). Among these haplotypes, there was a clear dominance of clade B, represented by a variety of haplotypes (SPAIN.04, SPAIN.06, SPAIN.08, SPAIN.10, SPAIN.12, Eu7, CanaryIsl.2, AZORES.09) comprising more than 50% of the sampled mice, found in 10 out of 14 locations from the southern coast of Spain (see Figure 1A and Table S1, Supplementary Material). Elsewhere, clade B is mostly found in the Near East, but it has also been detected throughout the northern Mediterranean and along the Portuguese coastline (Figure 1A; Jones *et al.* 2011a).

Azores

The 239 sequences from the nine islands collapsed into 41 haplotypes (haplotype diversity *h* ± SD was 0.910 ± 0.011 and nucleotide diversity π ± SD was 0.0144 ± 0.0072, see Table 1). From these, 24 haplotypes are so far unique to the Azorean islands, one is only shared between the Azores and southern Spain samples (AZORES.09), and the remaining 16 had been previously reported elsewhere, from the neighbouring archipelagos of Madeira (Gündüz *et al.* 2001, Förster *et al.* 2009), the Canaries (Jones *et al.* 2011a) and mainland Portugal (Förster *et al.* 2009), and as far away as Finland (Prager *et al.* 1993), Turkey (Gündüz *et al.* 2005) and the Falkland Islands (Hardouin *et al.* 2010) (Figure S1, Supplementary Material).

Table 1 – Genetic diversity indices (*h* and π) and neutrality tests (D, F_S and R_2) applied to the Azorean, Portuguese and Spanish populations of house mouse based on the D-loop dataset.

						Genetic diversity		Neutrality tests		
		N	N_H	Area (km^2)	Human Pop	*h*	*p*	D	F_S	R_2
Azorean Island										
	all islands	239	41	2 333	246 102	0.910 ± 0.012	0.0144 ± 0.0072			
E group	Santa Maria	24	5	97	5 547	0.754 ± 0.048	0.0030 ± 0.0019	-1.470	1.927	0.075
	São Miguel	52	13	747	137 699	0.865 ± 0.024	0.0079 ± 0.0042	-0.647	1.465	0.086
Central group	Terceira	31	10	402	56 062	0.822 ± 0.047	0.0129 ± 0.0067	1.474	4.510	0.175
	Graciosa	23	5	61	4 393	0.668 ± 0.079	0.0011 ± 0.0009	-0.268	-1.207	0.127
	São Jorge	26	3	246	8 998	0.151 ± 0.093	0.0002 ± 0.0003	-1.156	**-2.176**	0.192
	Pico	23	4	448	14 144	0.557 ± 0.083	0.0124 ± 0.0065	1.277	12.177	0.182
	Faial	28	6	173	15 038	0.521 ± 0.102	0.0090 ± 0.0048	-0.246	6.653	0.117
W group	Flores	23	4	142	3 791	0.589 ± 0.094	0.0064 ± 0.0035	2.649	6.678	0.242
	Corvo	9	2	17	430	0.389 ± 0.164	0.0005 ± 0.0005	0.156	0.477	0.194
	Portugal	76	15	92 090	10 561 614	0.883 ± 0.023	0.0081 ± 0.0043	-0.109	1.590	0.100
	Spain (Huelva + Cadiz)	66	20	17 584	1 765 487	0.927 ± 0.013	0.0118 ± 0.0060	-0.207	0.653	0.096

N – sample size, N_H – number of haplotypes, Human Pop – human population size, *h* – haplotype diversity ± SD, π – nucleotide diversity ± SD, D - after Tajima (1989), F_S - after Fu (1997), R_2 - after Ramos-Onsins and Rozas (2002). Statistically significant neutrality test result shown in bold ($p < 0.05$).

The Azorean haplotypes were evenly distributed among four of the six D-loop clades described by Jones *et al.* (2011a): clades B, C, D and F (see Figure 2A and Figure S1, Supplementary Material). In addition to these four clades, there was a single haplotype grouping within clade E (U47431), represented by a single individual collected in São Miguel. There was also a single haplotype basal to the clades B – F representing five individuals from São Miguel. The basal haplotypes in our tree included the basal sequences and the clade A sequences from Jones *et al.*

(2011a); clade A was a poorly supported lineage that was not recovered in our analysis. With each phylogenetic analysis of house mouse D-loop haplotypes, there are small adjustments in the final topology of the trees produced; our tree produces the same general consistency of clades as previous analyses (see Prager *et al.* 1993, 1996, 1998, Nachman *et al.* 1994, Gündüz *et al.* 2001, 2005, Rajabi-Maham *et al.* 2008, Searle *et al.* 2009a,b, Förster *et al.* 2009, Jones *et al.* 2010, 2011a,b, 2012, Gabriel *et al.* 2011).

Clade B included eight haplotypes (AZORES.05-09, AZORES.12, Turkey.34, Portugal.12 and CanaryIsl.2) representing 45 individuals, 42 of which were scattered around São Miguel (Figure 1B). The remaining individuals were found in the neighbouring island of Santa Maria (n=2) and in Pico (n=1). As described above, many haplotypes from southern Spain also grouped within this clade. In fact, there was a total of four haplotypes in common between the Azorean and Spanish datasets, two of which were in clade B (AZORES.9 and CanaryIsl.2), both mostly found in São Miguel. This preponderance of similar clade B haplotypes in Spain and São Miguel contributed to the small pairwise Φ_{ST} values between them (Table 2).

Clade C (AZORES.10, AZORES.19-22, AZORES.26-27, AM182702, Portugal.7, CanaryIsl.3) was the most widespread clade in the Azores, present on six of the nine islands - São Miguel, Graciosa, Pico, Faial, Flores and Corvo (Figure 1B). This was the only clade found in Graciosa (AZORES.19-22 and CanaryIsl.3) and Corvo (AM182702, Portugal.7). On these islands, as in the others with a high representation of clade C (Pico, Faial, Flores), there is only a single haplotype or a star-like phylogeny of a central common haplotype and other haplotypes separated by one nucleotide (Figure 1B).

Table 2 - Pairwise Φ_{ST} values based on D-loop sequences between the house mouse populations on the Azorean islands, mainland Portugal and southern Spain (Huelva + Cadiz provinces). Significance was tested by 10,000 permutations; all values are significant at the 0.05 level unless marked with an asterisk.

	Sta. Maria	S. Miguel	Terceira	Graciosa	S. Jorge	Pico	Faial	Flores	Corvo	Portugal
Sta. Maria	-									
S. Miguel	0.552	-								
Terceira	0.173	0.439	-							
Graciosa	0.859	0.579	0.606	-						
S. Jorge	0.940	0.796	0.660	0.977	-					
Pico	0.572	0.375	0.326	0.370	0.627	-				
Faial	0.579	0.375	0.368	0.270	0.771	0.036*	-			
Flores	0.529	0.378	0.353	0.445	0.881	0.245	0.118	-		
Corvo	0.824	0.488	0.500	0.572	0.990	0.196*	0.060*	0.244	-	
Portugal	0.552	0.314	0.466	0.253	0.783	0.219	0.101	0.131	0.084*	-
Spain	0.456	0.071	0.345	0.377	0.647	0.162	0.180	0.233	0.269	0.182

This is as expected from recent introduction and population expansion, but even Graciosa and Corvo (where clade C is the only lineage) failed to yield significant results with the neutrality tests (Table 1), perhaps reflecting small sample sizes. From the ten haplotypes grouping within clade C, AM182702 stood out as the most common in the whole Azorean dataset. This same haplotype had been previously reported as the most frequent and widespread in mainland Portugal (named Portugal.1 in Förster *et al.* 2009). In the Azores, it was present in four islands (Faial, Pico, Corvo and Flores) where it represented 61-78% of the samples. The high frequencies of AM182702 contributed to the non-significant pairwise Φ_{ST} values between various combinations of Corvo, Faial, Pico and Portugal (Table 2). As well as the Azores and mainland Portugal, clade C is the most widely distributed clade in central and southern Europe (Jones *et al.* 2011a).

Clade D, formerly named as the 'Madeira clade' (Gündüz *et al.* 2001, Förster *et al.* 2009) was detected in four Azorean islands: São Miguel, Terceira, São Jorge

and Pico (Figure 1B). From the eleven haplotypes identified (AZORES.11, AZORES.13-16, AZORES.18, AZORES.23-25, PSanto.2 and U47456), two had previously been reported in the Madeiran archipelago, one on Madeira island (U47456, Gündüz *et al.* 2001, Förster *et al.* 2009) and the other on Porto Santo (PSanto.2, Gündüz *et al.* 2001). The haplotype U47456 was detected in a single location in Faial and PSanto.2 was very widespread in São Jorge (24 of 26 individuals sequenced). Because of this high frequency of PSanto.2, São Jorge has by far the lowest nucleotide and haplotype diversity of all the Azorean islands (h=0.151±0.093 and π=0.0002±0.0003, see Table 1). It is also the only population which shows a significant signal of population expansion (Fu's F_S; Table 1). Despite occasional occurrences along the northern Mediterranean coast, clade D has mostly been found in continental northern Europe, being particularly common in the Faroe islands and through much of Scandinavia (Jones *et al.* 2010, 2011a,b).

The clade E individual found on São Miguel had haplotype U47431. This clade is found most commonly in the British Isles and neighbouring parts of continental Europe (Jones *et al.* 2010, 2011a).

Clade F (AZORES.01-04, AZORES.17, AZORES.28, 1334, ETL11, NTj1) was dominant and widespread in Santa Maria (22 of 24 specimens – AZORES.01-04) and Terceira (21 of 31 mice – AZORES.01-03, AZORES.17, NTji), and also represented by a small number of specimens on Faial and Flores (Figure 1B). Because haplotypes AZORES.01-03 were shared by Santa Maria and Terceira, they showed one of the lower Φ_{ST} values between island pairs (Table 2). Clade F has a clearly defined range in Europe bordering the north-east Atlantic: Ireland, northern Scotland, the coast of Norway and Iceland (Jones *et al.* 2010, 2011a, 2012).

Discussion

Genetic diversity

As commonly found for continental populations of house mice, the new data on D-loop sequences in Spanish mice showed high levels of diversity, reflecting the presence of several lineages and variation within those. This is consistent with the long lasting human presence in that region, rich in multicultural contacts and intense commercial trading (Valenzuela-Lamas *et al.* 2011).

The overall D-loop diversity observed in the Azorean archipelago was also high, again comparable to many continental populations, including Portugal and Spain (see Table 1). However, assessing each island separately, a different scenario emerges, with some very low values of haplotype and nucleotide diversity for certain islands (in particular São Jorge and Corvo: Table 1).

D-loop variation has been examined in house mice from other archipelagos. For example, seven small islands in the Faroes (Jones *et al.* 2011b), three small-to-medium islands in the Canaries (Bonhomme *et al.* 2011) and seven small islands in the Kerguelen archipelago have also been examined (Hardouin *et al.* 2010). Overall diversity in the Faroes was low (h=0.599, π= 0.00434) (Jones *et al.* 2011b), and only clades D and E were detected. Among the 143 house mice surveyed in the Canaries, clades B, C and D were detected (Bonhomme *et al.* 2011; Jones *et al.* 2011a) and only clades C and E were found among the 283 mice from small islands in the Kerguelen archipelago (Hardouin *et al.* 2010) (Figure S1, Supplementary Material). Our results for the nine small-to-medium Azorean islands revealed a much different scenario, with the presence of five clades throughout the archipelago, translating into a high haplotype and nucleotide diversity.

History and haplotype affiliations

Unlike the coast of Portugal, which has been well-sampled (Förster *et al.* 2009), the southern coast of Spain constituted a serious sampling gap that we needed to address in order to examine the colonization history of the Macaronesian archipelagos by the house mouse. Both Iberian countries played a decisive role in the human settlement of their insular dependencies (Portugal: Madeira, Azores; Spain: Canaries), so their respective coastlines constituted important potential source areas for mouse colonists. Due to their strategic location between the Atlantic and the Mediterranean, the historic harbours of Seville, Palos de la Frontera, Sanlúcar de Barrameda and Cádiz in southern Spain were major trading and setting off points for ships that would have included supplies intended for the Macaronesian islands and ships carrying settlers for the Canaries (O'Flanagan 2008).

There was a high genetic similarity of the house mice in southern Spain and Portugal, presumably reflecting a common history of colonization of house mice in western Iberia (Förster *et al.* 2009), as part of the expansion of *M. m. domesticus* around the Mediterranean during the Iron Age (Cucchi *et al.* 2005). The new Spanish samples revealed D-loop lineages previously recorded in Portugal (Förster *et al.* 2009), with the two most common clades in Portugal also being the most widespread in southern Spain (clades B and C). As in mainland Portugal, clade F was not detected, but interestingly clade D was found in 11 individuals in 5 localities in southern Spain, while this clade has never been recorded in Portugal.

The main focus of this paper is the Azores, and human history is critical in the colonization of these islands by house mice. Historical accounts suggest that

the archipelago lacked terrestrial mammals prior to the arrival of the first Portuguese sailors in the 15th century (Drouët 1861). If these records were accurate, mainland Portugal would be the expected source area for the first mouse colonists based on its cultural and historical links with the Azores. As a consequence of the remoteness of the islands, the peopling process of the Azores was slow, lasting over two centuries (Mendonça 1996). Settlers arrived mostly from mainland Portugal and the neighbouring Madeiran archipelago but there were also minor contributions from elsewhere in Europe (Flanders, Spain, Italy, France, England, Germany and Scotland) (Matos *et al.* 1989), which is corroborated by human genetic data (Santos *et al.* 2003, Montiel *et al.* 2005, Santos *et al.* 2006). Also, from the late 15th century onwards, due to its location in the middle of the Atlantic, the Azores started to play an essential role as a stop-over. The islands became a vital strategic base for both the Portuguese and Spanish maritime empires and their deeply intertwined trading networks (O'Flanagan 2008). From the 16th century onwards, the Azores became an important centre for trade and exchange for ships crossing between Europe and the Americas, or going around Africa towards India (Diffie and Winius 1977). During the 19th century, the whaling industry became substantially developed in the Azores along with fisheries and stockbreeding facilities (Greene and Morgan 2008).

Thus, it seems clear that throughout the centuries of human presence, there have been numerous opportunities for the introduction of house mice onto the islands of the Azores.

Indeed, the pattern of D-loop variation observed in Azorean mice suggests quite a complex colonization history of the archipelago, generally in agreement with this historically rich background. The house mice on different sets of islands

appear to have a different origin. Nevertheless, the link with mainland Portugal and/or southern Spain is clear for the majority of the islands (São Miguel, Graciosa, Pico, Faial, Flores and Corvo) with dominance of clades B and C in these islands as in the mainland areas. There are haplotypes in both clades that are found both in Portugal and the Azores (AM182702 [named Portugal.1 in Förster *et al.* 2009], Portugal.12 and Portugal.7), and in Spain and the Azores (AM182702, CanaryIsl.2 and AZORES.09). There are also many other Azorean, Portuguese and Spanish haplotypes within clade B and C which are not identical but are closely positioned in the phylogenetic tree. A closer inspection of clades B and C clearly shows sub-clustering within each of the clades, with haplotypes from Portugal, Spain and the Azores grouping together (Figure 2; Figure S1, Supplementary Material). This reinforces the affinity between these areas, suggesting that further sampling would possibly reveal an even higher degree of co-occurrence of some of these haplotypes.

The historical link with Madeira was also revealed particularly in the Central group but with a relatively limited impact on the genetic makeup of these populations except for one island. This was São Jorge, where 92% of the obtained sequences belonged to a single haplotype (PSanto.2), previously recorded in Porto Santo, in the Madeiran archipelago (Gündüz *et al.* 2001), and in two localities in southern Spain (this study).

Among the islands of the Azorean archipelago, Santa Maria produced the most unexpected result. As the closest island to the mainland (~1300 km), it is not surprising that it was the first to be officially colonized by the Portuguese (Diffie and Winius 1977). Previous studies based on zooarchaeological evidence (Cucchi *et al.* 2005, Valenzuela-Lamas *et al.* 2011) and molecular dating (Förster *et al.*

2009) showed that house mice were present in the Iberian Peninsula by the time of the documented discovery and settlement of the Azores in the 15th century. If Santa Maria had never been visited before, house mice arriving with Portuguese vessels would have had the opportunity to colonize the island at this stage. However, the mice on Santa Maria are predominantly clade F, which has not been found at all in Portugal or southern Spain. Clade F has a geographical range that matches well with the realm of influence of the Norwegian Vikings (coastal Norway, northern and western Scotland, Ireland, Iceland: Jones *et al.* 2012). Norwegian Vikings were certainly transporting mice of this clade (based on sequencing of Viking Age mouse specimens: Jones *et al.* 2012), and were excellent mariners, and crossed enormous water gaps, so it is not inconceivable that they visited the Azores, leaving mice as evidence of that event, as suggested for Danish Vikings and Madeira (Gündüz *et al.* 2001, Förster *et al.* 2009). The haplotypes found on Santa Maria have not so far been found in the area occupied by Norwegian Vikings. However, there are clear similarities between the representation of clade F on Terceira with that on Santa Maria and one of the clade F haplotypes found on Terceira is identical to one from Norway. A later introduction of clade F is also possible (e.g. in association with movements of ships involved in the whaling industry).

Long distance colonization or inter-island movements?

As the nine Azorean islands are geographically spread over 600 km of ocean, their discovery and peopling process was sequential, much like in other archipelagos with a similar geographic layout (e.g. the Aleutian Islands) (Crawford

et al. 2011, Balter 2012). In the case of the Azores, that process went from east to west, as expected from the proximity of Portugal, the primary source of human occupation. As with the human colonization, there is an expectation that house mice would have occupied the islands from east to west, with inter-island movements. A possible example of this is the comparison of the Eastern group island of Santa Maria and the Central group island of Terceira. The three most common D-loop haplotypes (AZORES.01–03) are shared by the islands, suggesting a stepwise colonization for the house mice as for the people.

The seven easternmost islands were colonised long before the Western group of islands, Flores and Corvo. The first written document with a reference to the peopling of the Azores dates back to 1439, according to which the Portuguese crown ordered "sheep to be released in the seven Azorean islands in preparation for human settlement" (Crosby 2004). During about two decades, Flores and Corvo remained unknown until their official discovery in 1452. The most frequent house mouse D-loop haplotype found both in Flores and Corvo (AM182702) is also the most frequent in Faial and Pico, the closest islands. This haplotype is the same as Portugal.1, the most common and widespread in mainland Portugal, as scored by Förster *et al.* (2009). Therefore, Flores and Corvo may have been colonized from more easterly islands, but direct colonization from Portugal is also possible.

Multiple colonization sources for the archipelago, often single colonization sources for individual islands

While a single D-loop clade is represented in the neighbouring Madeiran archipelago (clade D, Gündüz *et al.* 2001, Förster *et al.* 2009) and only three clades have been reported in the Canary Islands (clades B, C and D, Bonhomme *et al.* 2011

Jones *et al.* 2011a), in the Azores we obtained sequences grouping within 5 named clades of the *M. m. domesticus* D-loop phylogenetic tree (B, C, D, E and F: Jones *et al.* 2011a). This level of diversity is consistent with multiple successful colonizations from multiple origins. However, as pointed out earlier, much of this diversity is between islands rather than within them. São Miguel, the biggest and most populated island within the archipelago is the only island to have all five clades, but even there one clade is much commoner than all the others (Figure 1B). Thus, although there is a multitude of D-loop clades in the archipelago, and a greater tendency in the larger islands to have more of them, there is still a remarkable sorting of clades among the islands and groups of islands.

Indeed, the pattern of mitochondrial DNA variation observed in Santa Maria (Eastern group), São Jorge, Graciosa (Central group) and Corvo (Western group) is consistent with a scenario of a single colonization event per island. In each of these four islands we detected a very limited number of haplotypes, usually one or two mutation steps apart, either accumulated *in situ* from founding haplotypes due to the high D-loop mutation rate (e.g. Geraldes *et al.* 2008) or originally arriving from the same source population. The only exception occurs in Santa Maria, where only two specimens were sampled near the island's harbour carrying a haplotype belonging to a distinct clade, frequently found in the neighbouring island, São Miguel.

The history of the house mouse in the Macaronesian islands

The Macaronesian islands of the Azores, Madeira, and the Canaries are the Atlantic island groups off the coast of Iberia and North Africa, which, as a result of

this paper, have now all been studied with regards house mouse colonization (this study, Gündüz *et al.* 2001 and Förster *et al.* 2009, Bonhomme *et al.* 2011, respectively). These islands were particularly important during the period when mainly the Portuguese, Spanish, French, Dutch and British explored, exploited and settled around the world, from the 15th century onwards. At this stage, the Macaronesian islands started to be colonized by Europeans and were often used as stopping off points for ships, and played a pivotal role in commercial networks (Greene and Morgan 2008). Here we consider the extent to which the history of colonization of these islands by house mice matches this complex human history.

The house mouse subspecies that colonized all three archipelagos was *M. m. domesticus*, which is also present in Portugal, Spain, France, the Netherlands and Britain. This is the subspecies that is the predominant colonist of global territories, taken by those countries to places such as North and South America, Australia, New Zealand, French Antarctic islands, etc. (Boursot *et al.* 1993, Silver 1995, Searle *et al.* 2009b, Hardouin *et al.* 2010, Gabriel *et al.* 2011).

Here we have focused on the colonization of the Azores by house mice using D-loop sequences. Much of the genetic signal recovered is consistent with the first recorded colonization of the islands by Portuguese people. From the D-loop data, the mice on six of the nine islands could have colonized directly from Portugal based on the sharing of clades that are widespread both on these islands and mainland (Figure 1). As the discovery of the islands was sequential, from east to west, there is also the possibility of inter-island colonization based on haplotype sharing between mice on different islands/island groups.

The specific involvement of Spain is also possible in colonization of the Azores by house mice. Indeed, the close genetic similarity between the mice of

southern Spain and São Miguel suggests that Spain was more likely than Portugal to be the source area for this island (Table 2).

Madeira was actually discovered and settled by the Portuguese slightly earlier than the Azores (Diffie and Winius 1977), with the recorded discovery of Madeira at 1419 and the Azores at 1427. Despite this historical and cultural link with mainland Portugal, for the Madeira archipelago there is essentially no correspondence with the D-loop sequences found in Portugal (only one individual from Madeira had a sequence likely to derive from Portugal: see Figure 2; Förster *et al.* 2009). Instead there is an extraordinary and close correspondence between the sequences in Madeira and those in northern continental Europe (Germany and Scandinavia). Both share the same lineage (clade D), having five haplotypes in common, including the most frequent and widespread in both areas (Gündüz *et al.* 2001, Förster *et al.* 2009). Haplotypes found in the Madeiran archipelago (or haplotypes closely related to them) are also found in the Azores, the Canaries and southern Spain. However, among anywhere in southern Europe, Madeira is the only place that has such similarity in house mouse D-loop sequences (in terms of clade predominance and sharing of haplotypes) with northern continental Europe. This incredibly close correspondence was the reason why it was suggested that Madeira was colonized by house mice from northern Europe (Gündüz *et al.* 2001). In particular, it was proposed that the mice colonized before the Portuguese settlers had an influence on the island, and that despite much traffic of mice from Portugal, there is no D-loop signal of that, because of the difficulty of incoming female mice to displace female residents (Jones *et al.* 1995, Gabriel *et al.* 2010, Bonhomme and Searle in press).

We believe that an unrecorded visit of Danish Vikings to Madeira is the most likely explanation for the genetic make-up of the Madeira mice. Other scenarios are possible, but appear less plausible. For example, the similarity of Madeiran and north European mice could reflect mice going in the other direction, with mice from Madeira colonizing northern Europe, or alternatively, somewhere else in southern Europe colonizing both Madeira and northern Europe. However, that 'somewhere else' is entirely conjecture, because nowhere else in southern Europe have mice been shown to have the tight linkage with northern Europe as those on Madeira. We have argued elsewhere that clade D, which includes the haplotypes found in Madeira and northern Europe, colonized through central Europe and that the northern European representatives of clade D are derived from central European ancestral sequences (Jones *et al.* 2011a).

Another possibility is that the Danish Vikings transported mice to southern Spain, and from there mice got to Madeira. But, again, there is a much closer fit between the sequences found on Madeira and northern Europe, than those found in Spain. Southern Spain would have been colonized by house mice long before the Vikings made expeditions to the Mediterranean (Förster *et al.* 2009, Valenzuela-Lamas *et al.* 2011). It would appear much more likely that the small amount of clade D infiltration in southern Spain results from the long period of contact with Macaronesia rather than a transitory visitation by Vikings.

It is the mice on Madeira island itself that are closely similar to those in northern Europe. Those on the other major island in the archipelago, Porto Santo, are distinctive from the Madeira island mice, being characterised by a single nucleotide difference from the most common haplotype on Madeira (Förster *et al.*

2009). It can be suggested that the mice on Porto Santo derived from those on Madeira by a founder event, with the founding mice showing this variant.

Interestingly, it is mice of the Porto Santo lineage within the D clade that are particularly widespread; being found in the three Macaronesian archipelagos and also in southern Spain. A Porto Santo haplotype is the predominant D-loop sequence among mice on São Jorge in the Azores and the Porto Santo lineage is also well-represented on El Hierro and La Palma in the Canaries (Bonhomme *et al.* 2011).

Within the Madeiran archipelago, Porto Santo was settled earlier than Madeira island by the Portuguese, even though later on Madeira became more important than Porto Santo (Albuquerque and Vieira 1987). Therefore, the involvement of Porto Santo mice in the colonization of the Azores is not surprising.

The association between the Canaries and the Madeira archipelago was also strong in the 15th century (the Canaries were disputed territory between Spain and Portugal: Greene and Morgan 2008). So, again, representation of mice of the Porto Santo lineage in the Canaries is not surprising. Clade C is also present in the Canaries, being dominant on Tenerife island (Bonhomme *et al.* 2011), and probably represents a colonization from Spain, based on sequence similarity (Figure 2, Figure S1, Supplementary Material).

Finally, the occurrence of the Porto Santo lineage in southern Spain could represent migration of mice from Macaronesia.

It is interesting that the mouse D-loop haplotypes on the Canaries may relate to events in Macaronesia after the colonial expansion by the Portuguese and the Spanish. The level of divergence within clades suggests that mice in the Canaries have a more recent history than the mice on Madeira (Bonhomme *et al.*

2011). The fact that the mice in the Canaries may relate to a post-colonial colonization is surprising because there is ^{14}C dating archaeological evidence for house mice in the islands of La Palma, El Hierro, Fuerteventura and Lanzarote pointing out to an arrival time occurring between 756 cal BC and 313 cal AD (Alcover *et al.* 2009). It appears that the earlier population of house mice was in some fashion completely replaced by the post-colonial mice (Bonhomme *et al.* 2011); the most reasonable explanation is that the earlier mice went extinct for some reason.

It is possible that an extinction event also occurred on Santa Maria, leading to the loss of mice originally brought from Portugal or Porto Santo. Under this scenario, the presence of clade F would represent a 'turnover event', with mice carrying the new clade arriving with ships that happened to come from Norway or elsewhere in the distribution of clade F.

However, earlier colonization of Santa Maria by house mice brought from within the range of the Norwegian Viking kingdom is also an intriguing possibility. Santa Maria is a long way off the expected course of travel of Norwegian Vikings, but they were intrepid mariners, being the first European discoverers of Iceland, Greenland and North America (Forte *et al.* 2005). Much like the hypothesised scenario for the Madeiran archipelago (although involving a different D-loop clade, F instead of D), it is possible that Vikings would have landed or been shipwrecked on Santa Maria (leaving mice that evacuated the boat as evidence of that occurrence). This might seem an unlikely event but many documented island discoveries have been accidental, resulting from ships getting off course for any number of reasons. Based on evidence from medieval maps and Scandinavian

texts, Kelley (1979) has already speculated that Norwegian Vikings may have found their way to the Azores.

Interestingly, the three clade F haplotypes that are shared between Santa Maria and Terceira (AZORES.01-03) are also found on the Falkland Islands, a sub-Antarctic archipelago in the south Atlantic (Figure S1, Supplementary material; Hardouin *et al.* 2010). Additionally, there are three other clade F haplotypes closely similar to AZORES.01-03, and also three clade D haplotypes closely similar to those found on Madeira (Figure S1, Supplementary material; Hardouin *et al.* 2010). This strongly suggests that mice were taken from the Azores (and maybe Madeira) to the Falklands. This could have happened at various times in the colonial history of the Falklands. The Azores were, as we have said, important stopping off points for ships travelling from Europe via the Atlantic to places elsewhere in the world. The interesting thing about the Falklands is that the question of who discovered the islands is one of the issues relating to their disputed sovereignty, and one of the earliest potential discoverers was in the service of the Portuguese Crown (Amerigo Vespucci in 1501; Hope 1983). So, it is possible that the close linkage between Santa Maria, Terceira and the Falklands in terms of sharing of clade F haplotypes, could relate to a house mouse colonization of the Falklands in the 1500s. This would not only be interesting from the point of view of the human history of the Falklands, but also in relation to how long clade F haplotypes have occupied Santa Maria and Terceira, and therefore whether they arrived with the Norwegian Vikings. This discussion emphasises that, as well as the Azores, other island groups and continental areas elsewhere in the world were found by explorers during the Age of Discovery, and for which there was dispute among the competing European nations as who 'discovered' the place first. It

appears that house mice may have a role in helping to distinguish between competing claims.

In conclusion, this study has demonstrated once again that there is much to be learnt from D-loop sequences regarding the colonization history of house mouse populations, with the focus in this study being the Azores. We have been able to generate a regional colonization history, bringing together data on mice from the Macaronesia archipelagos of the Azores, Madeira and Canaries (whose human histories are much entangled) and the countries that govern them (Portugal and Spain). It has been possible to develop scenarios for the mouse colonization history in the light of what we know about human history. Ultimately, however, deeper genome analysis would be worthwhile to properly test those scenarios, providing opportunities to distinguish between alternative closely similar colonization sources (e.g. Spain versus Portugal), to apply accurate molecular dating methods and to infer demographic history of mouse populations.

Acknowledgments

We are grateful to Ana Quaresma and Margarida Santos Reis for samples from São Miguel and Terceira islands. We are also indebted to Pedro Martins da Silva, Mabel Giménez, Paula Lopes, Rita Monarca and Joaquim Tapisso for their help during fieldwork in the Azores and/or Southern Spain. This study was funded by the scholarship SFRH/BD/21437/2005 from Fundação para a Ciência e a Tecnologia (Portugal) to SIG.

References

Albuquerque L, Vieira A (1987) *O arquipélago da Madeira no século XV*. Colecção Atlântica. Secretaria Regional do Turismo e Cultura / Centro de Estudos de História do Atlântico, Região Autónoma da Madeira, Funchal, Portugal.

Alcover JA, Rando JC, García-Talavera F, Hutterer R, Michaux J, Trias M, Navarro JF (2009) A reappraisal of the stratigraphy of Cueva del Llano (Fuerteventura) and the chronology of the introduction of the house mouse (*Mus musculus*) into the Canary Islands. *Palaeogeography Palaeoclimatology Palaeoecology*, 277, 184–190.

Angel A, Wanless RM, Cooper J (2009) Review of impacts of the introduced house mouse on islands in the Southern Ocean: are mice equivalent to rats? *Biological Invasions*, 11, 1743–1754.

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

Balter M (2012) The peopling of the Aleutians. *Science*, 335, 158–161.

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

Berry RJ, Peters J (1975) Macquarie Island house mice: A genetical isolate on a Sub-Antarctic island. *Journal of Zoology*, 176, 375–389.

Berry RJ, Peters J, Van Aarde RJ (1978) Sub-Antarctic house mice: colonization, survival and selection. *Journal of Zoology*, 184, 127–141.

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell*, 26, 167–180.

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffra J-C, 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.

Bonhomme F, Searle JB (in press) *House mouse phylogeography*. In *Evolution of the House Mouse* (Cambridge Series in Morphology and Molecules). Eds Macholán M, Baird SJE, Munclinger P, Piálek J. Cambridge University Press, Cambridge, UK.

Boursot P, Auffray J-C, Britton-Davidian J, Bonhomme F (1993) The evolution of house mice. *Annual Review of Ecology and Systematics*, 24, 119–152.

Crawford MH, Rubicz RC, Zlojutro M (2010) Origins of Aleuts and the genetic structure of populations of the archipelago: Molecular and archaeological perspectives. *Human Biology*, 82, 695–717.

Crosby AW (2004) *Ecological Imperialism: The Biological Expansion of Europe, 900-1900*. Cambridge University Press, Cambridge, UK.

Cucchi T (2008) Uluburun shipwreck stowaway house mouse: molar shape analysis and indirect clues about the vessel's last journey. *Journal of Archaeological Science*, 35, 2953–2959.

Cucchi T, Vigne J-D, Auffray J-C (2005) First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarcheaological revision of subfossil occurrences. *Biological Journal of the Linnean Society*, 84, 429–445.

Cuthbert R, Hilton G (2004) Introduced house mice *Mus musculus*: a significant predator of threatened and endemic birds on Gough Island, South Atlantic Ocean? *Biological Conservation*, 117, 483–489.

Donlan CJ, Wilcox C (2008) Diversity, invasive species and extinctions in insular ecosystems. *Journal of Applied Ecology*, 45, 1114–1123.

Drake DR, Hunt TL (2009) Invasive rodents on islands: integrating historical and contemporary ecology. *Biological Invasions*, 11, 1483–1487.

Diffie BW, Winius GD (1977) *Foundations of the Portuguese Empire, 1415–1580*. University of Minnesota Press, Minneapolis, USA.

Drouët H (1861) *Éléments de la faune Açorienne*. Baillière et Fils, Paris, France.

Duplantier J-M, Orth A, Catalan J, Bonhomme F (2002) Evidence for a mitochondrial lineage originating from the Arabian peninsula in the Madagascar house mouse (*Mus musculus*). *Heredity*, 89, 154–158.

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

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

Förster DW, Gündüz İ, Nunes AC, Gabriel S, Ramalhinho MG, Mathias ML, Britton-Davidian J, Searle JB (2009) Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology*, 18, 4477–4494.

Forte A, Oram R, Pedersen F (2005) *Viking Empires*. Cambridge: Cambridge University Press, Cambridge, UK.

Frutuoso G (1590) *Saudades da Terra – Livro VI*. Instituto Cultural de Ponta Delgada Ed. 1998, Ponta Delgada, Portugal.

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

Gabriel SI, Jóhannesdóttir F, Jones EP, Searle JB (2010) Colonization, mouse-style. *BMC Biology*, 8, 131.

Gabriel SI, Stevens MI, Mathias ML, Searle JB (2011) Of mice and 'convicts': origin of the Australian house mouse, *Mus musculus*. *PLoS ONE*, 6, e28622.

Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu H-T, Bulatova N, Ziv Y, Nachman MW (2008) Inferring the history of speciation in house mice from autosomal, X–linked, Y–linked and mitochondrial genes. *Molecular Ecology*, 17, 5349–5363.

Greene JP, Morgan PD (2008) *Atlantic History: A Critical Appraisal*. Oxford University Press, Oxford,UK.

Gündüz İ, Auffray J-C, Britton-Davidian J, Catalan J, Ganem G, Ramalhinho MG, Mathias ML, Searle JB (2001) Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology*, 10, 2023–2029.

Gündüz İ, Rambau RV, Tez C, Searle JB (2005) Mitochondrial DNA variation in the western house mouse (*Mus musculus domesticus*) close to its site of origin: studies in Turkey. *Biological Journal of the Linnean Society*, 84, 473–485.

Gündüz İ, Tez C, Malikov V, Vaziri A, Polyakov A, Searle JB (2000) Mitochondrial DNA and chromosomal studies of wild mice (*Mus*) from Turkey and Iran. *Heredity*, 84, 458–467.

Hardouin EA, Chapuis J-L, Stevens MI, van Vuuren BJ, Quillfeldt P, Scavetta RJ, Teschke M, Tautz D (2010) House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evolutionary Biology*, 10, 325.

Harris DB (2009) Review of negative effects of introduced rodents on small mammals on islands. *Biological Invasions*, 11, 1611–1630.

Hope AFJ (1983) Sovereignty and decolonization of the Malvinas (Falkland) Islands. *Boston College and International and Comparative Law Review*, 6, 391–446.

Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46, 10–18.

Ihle S, Ravaoarimanana I, Thomas M, Tautz D (2006) An analysis of signatures of selective sweeps in natural populations of the house mouse. *Molecular Biology and Evolution*, 23, 790–797.

Jones CS, Noble LR, Jones JS, Tegelström H, Triggs GS, Berry RJ (1995) Differential male genetic success determines gene flow in an experimentally manipulated mouse population. *Proceedings of the Royal Society*, *B* 260, 251–256.

Jones EP, Jóhannesdóttir F, Gündüz İ, Richards MB, Searle JB (2011a) The expansion of the house mouse into north–western Europe. *Journal of Zoology*, 283, 257–268.

Jones EP, Jensen J-K, Magnussen E, Gregersen N, Hansen HS, Searle JB (2011b) A molecular characterization of the charismatic Faroe house mouse. *Biological Journal of the Linnean Society*, 102, 471–482.

Jones EP, Skirnisson K, McGovern TH, Gilbert MTP, Willerslev E, Searle JB (2012) Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *BMC Evolutionary Biology*, 12, 35.

Jones EP, van der Kooij J, Solheim R, Searle JB (2010) Norwegian house mice (*Mus musculus musculus/domesticus*): Distributions, routes of colonization and patterns of hybridization. *Molecular Ecology*, 19, 5252–5264.

Kelley Jr JE (1979) Non-Mediterranean influences that shaped the Atlantic in the early Portolan charts. *Imago Mundi*, 31, 18–35.

Lawson-Handley LJ, Estoup A, Thomas C, Lombaert E, Facon B, Aebi A, Evans D, Roy HE (2011) Ecological genetics of invasive alien species. *BioControl*, 56, 409–428.

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

Matisoo-Smith E, Robins JH (2009) Mitochondrial DNA evidence for the spread of Pacific rats through Oceania. *Biological Invasions*, 11, 1521–1527.

Mathias ML, Ramalhinho MG, Santos-Reis M, Petrucci-Fonseca F, Libois R, Fons R, de Carvalho GF, Oom MM, Collares-Pereira M (1998) Mammals from the Azores islands (Portugal): an updated overview. *Mammalia*, 62, 397–407.

Matos AT (1989) Settlement and colonization of the Azores. *Portugal no Mundo*, 12, 185–188 (in Portuguese).

Matthewson DC, Van Aarde RJ, Skinner JD (1994) Population biology of house mice (*Mus musculus* L.) on sub-Antarctic Marion Island. *South African Journal of Zoology,* 29, 99–106.

Michaux J, Cucchi T, Renaud S, Garcia-Talavera F, Hutterer R (2007) Evolution of an invasive rodent on an archipelago as revealed by molar shape analysis: the house mouse in the Canary Islands. *Journal of Biogeography*, 34, 1412–1425.

Mendonça L (1996) *História dos Açores*. Visão Geral (sec. XV-XIX). Centro de Apoio Tecnológico à Educação, Ponta Delgada, Portugal.

Montiel R, Bettencourt C, Silva C, Santos C, Prata MJ, Lima M (2005) Analysis of Y-chromosome variability and its comparison with mtDNA variability reveals different demographic histories between islands in the Azores archipelago (Portugal). *Annals of Human Genetics*, 69, 1–10.

Nachman MW, Boyer SN, Searle JB, Aquadro CF (1994) Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. *Genetics*, 136, 1105–1120.

O'Flanagan, P. 2008. *Port Cities of Atlantic Iberia, C. 1500 – 1900*. Ashgate Publishing, Aldershot, UK.

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

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53, 793–808.

Prager EM, Orrego C, Sage RD (1998) Genetic variation and phylogeography of Central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics*, 150, 835–861.

Prager EM, Sage RD, Gyllensten U, Thomas WK, Hübner R, Jones CS, Noble L, Searle JB, Wilson AC (1993) Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biological Journal of the Linnean Society*, 50, 85–122.

Prager EM, Tichy H, Sage RD (1996) Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: comparison with other house mice and report of a 75-bp tandem repeat. *Genetics*, 143, 427–446.

Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe, *Molecular Ecology*, 17, 627–641.

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

Richards M, Macaulay V, Bandelt H-J, Sykes B (1998) Phylogeography of mitochondrial DNA in western Europe. *Annals of Human Genetics*, 62, 241–260.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572–1574.

Santos C, Lima M, Montiel R, Angles N, Pires L, Abade A, Aluja MP (2003) Genetic structure and origins of peopling in the Azores Islands (Portugal): the view from mtDNA. *Annals of Human Genetics*, 67, 433–456.

Santos C, Montiel R, Bettencourt C, Prata MJ, Abade A, Aluja MP, Lima M (2006) Peopling, demographic history and genetic structure of the Azores Islands: Integrating data from mtDNA and Y-chromosome. *International Congress Series*, 1288, 85–87.

Searle JB (2008) The genetics of mammalian invasions: a review. *Wildlife Research*, 35, 185–192.

Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry S, Giménez MD, Jóhannesdóttir F (2009a) Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society, B* 276, 201–207.

Searle JB, Jamieson PM, Gündüz İ, Stevens MI, Jones EP, Gemmill CEC, King CM (2009b) The diverse origins of New Zealand house mice. *Proceedings of the Royal Society, B* 276, 209–217.

Silver LM (1995) *Mouse genetics, concepts and applications*. Oxford University Press, Oxford, UK.

Suarez AV, Tsutsui ND (2008) The evolutionary consequences of biological invasions. *Molecular Ecology*, 17, 351–360.

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

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, 10, 512–526.

Tollenaere C, Brouat C, Duplantier J-M, Rahalison L, Rahelinirina S, Pascal M, Moné H, Mouahid G, Leirs H, Cosson J-F (2010) Phylogeography of the introduced species *Rattus rattus* in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. *Journal of Biogeography*, 37, 398–410.

Valenzuela-Lamas S, Baylac M, Cucchi T, Vigne J-D (2011) House mouse dispersal in Iron Age Spain: a geometric morphometrics appraisal. *Biological Journal of the Linnean Society*, 102, 483–497.

van Vuuren BJ, Chown SL (2007) Genetic evidence confirms the origin of the house mouse on sub-Antarctic Marion Island. *Polar Biology*, 30, 327–332.

Supplementary Material – Table S1. Table of D–loop haplotypes found in *Mus musculus domesticus* from the Azores and southern Spain.

Location	Latitude	Longitude	N	Haplotype	Clade
AZORES ISLANDS					
Santa Maria	36.9752	-25.1373	1	AZORES.01	F
	36.9752	-25.1373	1	AZORES.02	F
	36.9706	-25.1158	2	AZORES.03	F
	36.9545	-25.1235	3	AZORES.02	F
	36.9462	-25.1525	1	AZORES.01	F
	36.9462	-25.1525	2	AZORES.02	F
	36.9553	-25.0397	1	AZORES.03	F
	36.9553	-25.0397	1	AZORES.01	F
	36.9602	-25.1360	2	CanaryIsl2	B
	36.9418	-25.0272	2	AZORES.01	F
	36.9753	-25.0417	1	AZORES.01	F
	36.9753	-25.0417	1	AZORES.04	F
	36.9861	-25.0595	1	AZORES.03	F
	36.9861	-25.0595	1	AZORES.02	F
	36.9879	-25.1543	1	AZORES.02	F
	36.9531	-25.0996	1	AZORES.03	F
	36.9531	-25.0996	1	AZORES.02	F
	36.9531	-25.0996	1	AZORES.01	F
São Miguel	37.7415	-25.6786	1	Turkey.34	B
	37.7646	-25.7043	1	CanaryIsl2	B
	37.8741	-25.8254	1	AZORES.05	B
	37.8928	-25.7882	1	Turkey.34	B
	37.7168	-25.4335	1	AZORES.06	B
	37.7221	-25.4864	1	CanaryIsl2	B
	37.8056	-25.6018	1	1083	BASAL
	37.7424	-25.5568	1	CanaryIsl2	B
	37.7622	-25.6027	1	U47431	E
	37.8071	-25.6481	1	1083	BASAL
	37.7951	-25.5086	1	AZORES.07	B
	37.8052	-25.4922	1	1083	BASAL
	37.8011	-25.7315	1	AZORES.08	B
	37.7489	-25.3992	1	AZORES.07	B
	37.8758	-25.7267	2	AZORES.09	B
	37.8833	-25.7244	1	AZORES.10	C
	37.7328	-25.3718	1	Turkey.34	B
	37.8528	-25.7708	1	AZORES.06	B
	37.7749	-25.4956	1	AZORES.07	B
	37.7552	-25.5176	1	CanaryIsl2	B

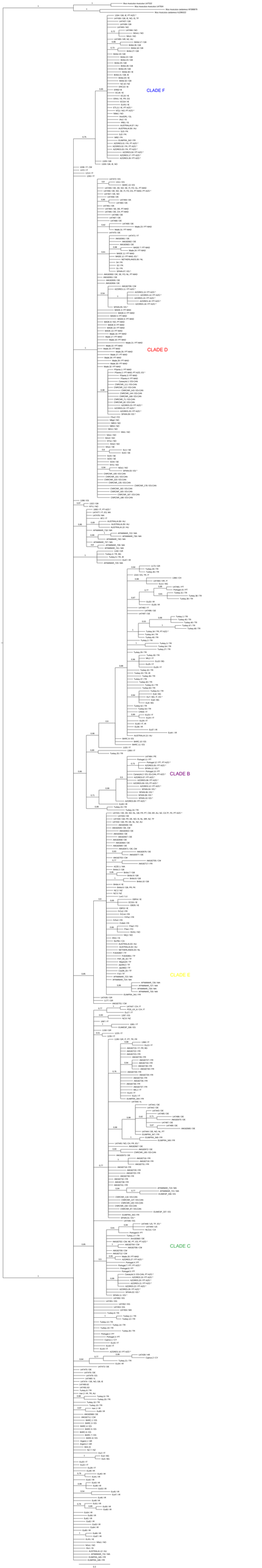
Location	Latitude	Longitude	N	Haplotype	Clade
AZORES ISLANDS					
São Miguel	37.8410	-25.7989	1	1083	BASAL
	37.8247	-25.8157	1	CanaryIsl2	B
	37.8345	-25.2178	1	Portugal.12	B
	37.7772	-25.2354	1	CanaryIsl2	B
	37.7749	-25.1862	1	AZORES.07	B
	37.8491	-25.2450	1	AZORES.09	B
	37.8319	-25.2826	1	AZORES.07	B
	37.8471	-25.2967	1	CanaryIsl2	B
	37.7924	-25.2963	1	CanaryIsl2	B
	37.8707	-25.7169	1	CanaryIsl2	B
	37.8291	-25.7458	2	AZORES.10	C
	37.7994	-25.3945	1	AZORES.07	B
	37.8388	-25.7433	1	Turkey.34	B
	37.7575	-25.4155	1	AZORES.07	B
	37.7867	-25.4646	1	Turkey.34	B
	37.7757	-25.3072	1	CanaryIsl2	B
	37.7773	-25.3586	1	Turkey.34	B
	37.7391	-25.3535	1	Turkey.34	B
	37.8370	-25.3464	1	AZORES.07	B
	37.7871	-25.5698	1	Turkey.34	B
	37.7998	-25.3387	1	CanaryIsl2	B
	37.8194	-25.3698	1	CanaryIsl2	B
	37.7600	-25.5401	1	AZORES.09	B
	37.7613	-25.6543	1	1083	BASAL
	37.7301	-25.5262	1	AZORES.07	B
	37.7925	-25.6364	1	CanaryIsl2	B
	37.8243	-25.6992	1	AZORES.11	D
	37.7913	-25.6928	1	Turkey.34	B
	37.7569	-25.6794	1	AZORES.12	B
	37.7541	-25.6634	1	Turkey.34	B
Terceira	38.7687	-27.1432	1	AZORES.01	F
	38.7425	-27.1070	1	AZORES.01	F
	38.7172	-27.3551	2	AZORES.02	F
	38.6985	-27.3436	1	AZORES.02	F
	38.6916	-27.3254	1	AZORES.02	F
	38.7919	-27.2247	1	AZORES.13	D
	38.7925	-27.2076	1	AZORES.01	F
	38.7799	-27.2505	1	AZORES.14	D
	38.7672	-27.2406	1	AZORES.02	F

Location	Latitude	Longitude	N	Haplotype	Clade
AZORES ISLANDS					
Terceira	38.7743	-27.2898	1	AZORES.03	F
	38.6871	-27.1191	1	AZORES.02	F
	38.6800	-27.1193	1	AZORES.15	D
	38.6674	-27.1110	1	AZORES.01	F
	38.6593	-27.0937	1	AZORES.16	D
	38.6993	-27.2651	1	AZORES.03	F
	38.7224	-27.3045	1	AZORES.03	F
	38.7206	-27.2898	1	AZORES.02	F
	38.7025	-27.2588	1	AZORES.17	F
	38.7412	-27.1605	1	AZORES.18	D
	38.7656	-27.1858	1	AZORES.01	F
	38.7412	-27.1605	1	AZORES.02	F
	38.7195	-27.2486	1	AZORES.02	F
	38.7408	-27.2620	1	AZORES.15	D
	38.7049	-27.0877	1	AZORES.15	D
	38.7320	-27.1338	1	AZORES.15	D
	38.7320	-27.1338	1	AZORES.01	F
	38.7146	-27.1995	1	NTj1	F
	38.7260	-27.1726	1	AZORES.02	F
	38.7024	-27.1515	1	AZORES.02	F
	38.6915	-27.1045	1	AZORES.15	D
Graciosa	39.0293	-27.9891	4	CanaryIsl3	C
	39.0293	-27.9891	1	AZORES.19	C
	39.0489	-28.0441	1	CanaryIsl3	C
	39.0489	-28.0441	3	AZORES.20	C
	39.0752	-28.0052	5	AZORES.21	C
	39.0507	-28.0100	3	CanaryIsl3	C
	39.0507	-28.0100	1	AZORES.21	C
	39.0579	-27.9780	3	CanaryIsl3	C
	39.0579	-27.9780	1	AZORES.22	C
	39.0463	-28.0289	1	CanaryIsl3	C
São Jorge	38.6453	-28.1263	2	PSanto.2	D
	38.6453	-28.1263	1	AZORES.23	D
	38.6865	-28.2036	2	PSanto.2	D
	38.6796	-28.1013	4	PSanto.2	D
	38.6785	-28.0939	5	PSanto.2	D
	38.6110	-27.9815	2	PSanto.2	D
	38.6847	-28.1079	2	PSanto.2	D
	38.6176	-28.0411	1	PSanto.2	D
	38.5394	-27.8147	5	PSanto.2	D
	38.6173	-28.0406	1	AZORES.24	D
	38.6173	-28.0406	1	PSanto.2	D

Location	Latitude	Longitude	N	Haplotype	Clade
AZORES ISLANDS					
Pico	38.5487	-28.4008	1	AM182702	C
	38.5487	-28.4008	1	PSanto.2	D
	38.4299	-28.4275	4	AM182702	C
	38.5079	-28.5289	5	AM182702	C
	38.4328	-28.4558	2	AM182702	C
	38.4147	-28.2727	1	AZORES.25	D
	38.4147	-28.2727	6	PSanto.2	D
	38.4147	-28.2727	2	AM182702	C
	38.5178	-28.3086	1	AZORES.07	B
Faial	38.5589	-28.6297	4	AM182702	C
	38.5751	-28.7816	1	AM182702	C
	38.5751	-28.7816	1	AZORES.02	F
	38.5890	-28.6105	1	AM182702	C
	38.5966	-28.6127	2	AM182702	C
	38.6349	-28.6933	2	AZORES.26	C
	38.5626	-28.6462	1	AM182702	C
	38.5606	-28.6247	1	AZORES.27	C
	38.5606	-28.6247	4	AM182702	C
	38.5606	-28.6247	1	AZORES.28	F
	38.5510	-28.6505	2	AM182702	C
	38.5751	-28.7816	2	AM182702	C
	38.5329	-28.7345	5	U47456	D
	38,634	-28,717	1	AM182702	C
Flores	39.4017	-31.1545	1	1334	F
	39.3890	-31.1662	3	AM182702	C
	39.4609	-31.1356	5	AM182702	C
	39.4670	-31.1482	1	ETL11	F
	39.5203	-31.2124	2	AM182702	C
	39.5216	-31.2109	5	AZORES.28	F
	39.3757	-31.1765	2	AM182702	C
	39.3886	-31.1668	2	1334	F
	39.3781	-31.1742	2	AM182702	C
Corvo	39.6799	-31.1022	4	AM182702	C
	39.6799	-31.1022	1	Portugal.7	C
	39.6738	-31.1127	2	AM182702	C
	39.6732	-31.1116	1	Portugal.7	C
	39.6732	-31.1116	1	AM182702	C

Location	Latitude	Longitude	N	Haplotype	Clade
SPAIN (Huelva and Cadiz)					
Villamartin	36.9553	-5.8199	5	AM182702	C
Jerez de la Frontera	36.7013	-6.0794	4	Eu7	B
	36.7013	-6.0794	1	AM182702	C
Prado del Rey	36.7966	-5.5481	2	PSanto.2	D
	36.7966	-5.5481	3	AM182702	C
Chiclana de la Frontera	36.3681	-6.1565	2	SPAIN.01	C
	36.3681	-6.1565	2	AM182702	C
	36.3681	-6.1565	2	CanaryIsl2	B
Sanlúcar de Barrameda	36.7627	-6.3637	4	*Mus spretus*	-
Arcos de la Frontera	36.7613	-5.8330	1	SPAIN.02	C
	36.7613	-5.8330	1	SPAIN.03	D
	36.7613	-5.8330	1	U47448	C
Vejer de la Frontera	36.2659	-5.9638	4	SPAIN.05	D
	36.2659	-5.9638	1	PSanto.2	D
	36.2659	-5.9638	1	MADE.12	D
Algeciras	36.1527	-5.4824	2	AZORES.09	B
	36.1527	-5.4824	3	SPAIN.06	B
Tarifa	36.1892	-5.7730	4	CanaryIsl2	B
	36.1892	-5.7730	1	SPAIN.04	B
	36.1892	-5.7730	1	SPAIN.07	D
Palos de la Frontera	37.2219	-6.8952	6	AZORES.09	B
La Rabida	37.2191	-6.9144	3	SPAIN.08	B
	37.2191	-6.9144	1	SPAIN.09	D
Aljaraque	37.2949	-7.0300	4	SPAIN.04	B
	37.2949	-7.0300	1	U47445	C
Moguer	37.2971	-6.8311	1	CanaryIsl2	B
	37.2971	-6.8311	1	SPAIN.10	B
El Rocio	37.1360	-6.4900	1	SPAIN.11	C
	37.1360	-6.4900	3	SPAIN.12	B
Doñana	37.1426	-6.4739	4	SPAIN.12	B

Supplementary Material – Figure S1. Phylogenetic tree of *Mus musculus domesticus*. Detailed 50% majority rule consensus tree after Bayesian inference, summarised in Figure 1 (based on a total of 520 *Mus musculus domesticus* D-loop haplotypes, derived from this study and previous data available at the time of analysis) (Prager *et al.* 1993, 1996, 1998; Nachman *et al.* 1994; Gündüz *et al.* 2000, 2001, 2005; Ihle *et al.* 2006; Rajabi-Maham *et al.* 2008; Förster *et al.* 2009; Searle *et al.* 2009a, b; Bonhomme *et al.* 2011; Jones *et al.* 2010, 2011a,b, 2012; Gabriel *et al.* 2011). Posterior probabilities are shown on branch nodes. All haplotypes detected in this study are highlighted with an asterisk in the tree (details of samples can be found in Table S1). Codes indicate GenBank number or code used in the publication where the sequence was first reported, followed by a list of all the countries (country codes) where the haplotype has been recorded: AR, Argentina; AU, Australia; BG, Bulgaria; CH, Switzerland; CM, Cameroon; CY, Cyprus; DE, Germany; DK, Denmark; EG, Egypt; ES, Spain; ES-CAN, Canary Islands (Spanish dependency); FI, Finland; FK, Falkland Islands; FO, Faroe Islands; FR, France; GB, Great Britain; GE, Georgia; GL, Greenland; GR, Greece; GS, South Georgia and South Sandwich Islands; HR, Croatia; IE, Ireland; IL, Israel; IR, Iran; IS, Iceland; IT, Italy; LU, Luxembourg; MA, Morocco; MR, Mauritania; NE, Niger; NO, Norway; NL, Netherlands; NZ, New Zealand; PE, Peru; PT, Portugal; PT-AZO, Azores Archipelago (Portuguese dependency); PT-MAD, Madeira Archipelago (Portuguese dependency); SE, Sweden; TF, French Southern Territories; TR, Turkey; US, United States.


Mus musculus musculus LAT532
Mus musculus musculus LAT534
Mus musculus castaneus WSB6875
Mus musculus castaneus AJ200022
CLADE F
CLADE D
CLADE B
CLADE E
CLADE C

Supplementary Material – References (Figure S1)

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffra J-C, 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.

Förster DW, Gündüz İ, Nunes AC, Gabriel S, Ramalhinho MG, Mathias ML, Britton-Davidian J, Searle JB (2009) Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology*, 18, 4477–4494.

Gabriel SI, Stevens MI, Mathias ML, Searle JB (2011) Of mice and 'convicts': origin of the Australian house mouse, *Mus musculus*. *PLoS ONE*, 6, e28622.

Gündüz İ, Auffray J-C, Britton-Davidian J, Catalan J, Ganem G, Ramalhinho MG, Mathias ML, Searle JB (2001) Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology*, 10, 2023–2029.

Gündüz İ, Rambau RV, Tez C, Searle JB (2005) Mitochondrial DNA variation in the western house mouse (*Mus musculus domesticus*) close to its site of origin: studies in Turkey. *Biological Journal of the Linnean Society*, 84, 473–485.

Gündüz İ, Tez C, Malikov V, Vaziri A, Polyakov A, Searle JB (2000) Mitochondrial DNA and chromosomal studies of wild mice (*Mus*) from Turkey and Iran. *Heredity*, 84, 458–467.

Hardouin EA, Chapuis J-L, Stevens MI, van Vuuren BJ, Quillfeldt P, Scavetta RJ, Teschke M, Tautz D (2010) House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evolutionary Biology*, 10, 325.

Ihle S, Ravaoarimanana I, Thomas M, Tautz D (2006) An analysis of signatures of selective sweeps in natural populations of the house mouse. *Molecular Biology and Evolution*, 23, 790–797.

Jones EP, Jóhannesdóttir F, Gündüz İ, Richards MB, Searle JB (2011a) The expansion of the house mouse into north–western Europe. *Journal of Zoology*, 283, 257–268.

Jones EP, Jensen J-K, Magnussen E, Gregersen N, Hansen HS, Searle JB (2011b) A molecular characterization of the charismatic Faroe house mouse. *Biological Journal of the Linnean Society*, 102, 471–482.

Jones EP, Skirnisson K, McGovern TH, Gilbert MTP, Willerslev E, Searle JB (2012) Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *BMC Evolutionary Biology*, 12, 35.

Jones EP, van der Kooij J, Solheim R, Searle JB (2010) Norwegian house mice (*Mus musculus musculus/domesticus*): Distributions, routes of colonization and patterns of hybridization. *Molecular Ecology*, 19, 5252–5264.

Nachman MW, Boyer SN, Searle JB, Aquadro CF (1994) Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. *Genetics*, 136, 1105–1120.

Prager EM, Orrego C, Sage RD (1998) Genetic variation and phylogeography of Central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics*, 150, 835–861.

Prager EM, Sage RD, Gyllensten U, Thomas WK, Hübner R, Jones CS, Noble L, Searle JB, Wilson AC (1993) Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biological Journal of the Linnean Society*, 50, 85–122.

Prager EM, Tichy H, Sage RD (1996) Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: comparison with other house mice and report of a 75-bp tandem repeat. *Genetics*, 143, 427–446.

Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe, *Molecular Ecology*, 17, 627–641.

Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry S, Giménez MD, Jóhannesdóttir F (2009) Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society, B* 276, 201–207.

Searle JB, Jamieson PM, Gündüz İ, Stevens MI, Jones EP, Gemmill CEC, King CM (2009b) The diverse origins of New Zealand house mice. *Proceedings of the Royal Society, B* 276, 209–217.

Chapter 2

Paper II

Genetic structure of house mouse (*Mus musculus* Linnaeus 1758) populations in the Atlantic archipelago of the Azores: colonisation and dispersal

---

Sofia I Gabriel[1,2], Maria da Luz Mathias[1] and Jeremy B Searle[2,3]

[1] CESAM – Centre for Environmental and Marine Studies, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, 1749–016 Lisbon, Portugal

[2] Department of Biology, University of York, PO Box 373, York YO10 5YW, UK

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

Manuscript submitted to '*Biological Journal of the Linnean Society*'

Abstract

We analysed the genetic structure and relationships of house mouse (*Mus musculus*) populations in the remote Atlantic archipelago of the Azores using nuclear sequences and microsatellites. We typed *Btk* and *Zfy2* to confirm that the subspecies *M. m. domesticus* was the predominant genome in the archipelago. Nineteen microsatellite loci (one per autosome) were typed in a total of 380 individuals from all nine Azorean islands, the neighbouring Madeiran archipelago (Madeira and Porto Santo islands) and mainland Portugal. Levels of heterozygosity were high on the islands, arguing against population bottlenecking. The Azorean house mouse populations were differentiated from the Portuguese and Madeiran populations and no evidence of recent migration between the three was obtained. Within the Azores, the Eastern, Western and Central island groups tended to act as separate genetic units for house mice, with some exceptions. In particular, there was evidence of recent migration events among islands of the Central island group, whose populations were relatively undifferentiated. Overall, the genetic structure as revealed with microsatellites appears to show recent movements of house mice around nearby islands rather than the original colonisation of the islands which is more readily interpreted from previously conducted mitochondrial DNA studies.

Introduction

The Azorean islands constitute the most remote archipelago in the North Atlantic, located ~1500 km from Europe and ~3900 km from the east coast of North America, comprising nine volcanic islands spread along 600 km (Figure 1). The islands were officially discovered and settled by the Portuguese in the 15th

century, although there may have been visitations by other mariners before that. Due to the strategic location of the Azores in the middle of the Atlantic, there has been much marine traffic through the islands since they have been settled.

As for other isolated oceanic islands, a large proportion of the current terrestrial fauna of the Azores arrived as a result of human-mediated introductions (Mathias *et al.* 1998). The western house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) is a human commensal that was introduced onto all nine islands, and is now one of the most common and widespread mammals in the archipelago (Mathias *et al.* 1998). Here we are interested in the history of this colonisation and any on-going interactions between the mouse populations within the archipelago. The colonisation history of *M. m. domesticus* throughout western Europe and the Mediterranean region has been studied in great detail using mitochondrial D-loop sequences (e.g. Gündüz *et al.* 2001, 2005, Rajabi-Maham *et al.* 2008, Searle *et al.* 2009a, Bonhomme *et al.* 2011, Jones et al. 2010, 2011a,b, 2012). This molecular marker appears to be particularly effective to address the first arrival of house mice, reflecting its maternal inheritance and the fact that resident female house mice may be difficult to displace by later immigrants (Gabriel *et al.* 2010, Bonhomme & Searle in press). D-loop studies of the house mouse in the Azores show a clear signal of colonisation from Portugal, Spain and the Madeiran archipelago, in line with recorded history (Gabriel *et al.* submitted) and geographical proximity (Figure 1). The Azores share most of their historical background with Madeira, as both are dependencies of Portugal that were officially discovered and settled by the Portuguese, Madeira slightly earlier than the Azores (Diffie and Winius 1977). The histories of all the Macaronesian island groups of the Azores, Madeira and the Canaries are interlinked with both Portugal and Spain.

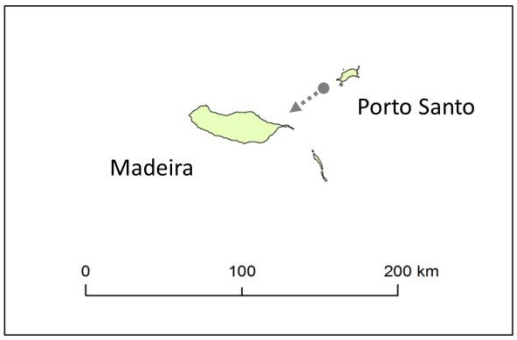
Although the house mice on most of the Azorean islands have D-loop sequences that reflect a direct or indirect history relating to Portugal, Santa Maria and Terceira show an unexpected association with a house mouse D-loop lineage otherwise found in Norway, Iceland and northern and western parts of the British Isles. Whether this represents a pre-Portuguese house mouse colonisation of the Azores, or a post-Portuguese replacement event, is unknown (Gabriel *et al.* submitted).

Nuclear markers have been used less frequently than the mitochondrial D-loop in studies of house mouse colonisation history, but can provide valuable additional information. A molecular study of the house mouse populations of the sub-Antarctic Kerguelen archipelago using both D-loop sequences and microsatellites confirmed the independence of colonisation and subsequent isolation of neighbouring islands (Hardouin *et al.* 2010). For Madeira, allozymes have generated a different outlook on house mouse colonisation from that obtained from D-loop sequences, with the mitochondrial marker probably reflecting first colonisation (persistence of female genetic lineage), and the nuclear markers primarily reflecting later immigrants (with genes transmitted on the male side) (Britton-Davidian *et al.* 2007, Förster *et al.* 2009). The differing perspectives created by nuclear markers for Kerguelen and Madeira presumably reflect aspects of house mouse history such as the intensity of immigration events and the duration over which they occur. Such features can also influence the D-loop genetic architecture (Bonhomme and Searle in press).

Here we use autosomal microsatellite markers to determine the genetic relations between house mouse populations among the islands of the Azores and to compare them with the populations on the Madeiran archipelago and mainland

Portugal. The 19 microsatellites are all centromeric and represent each of the autosomes. They therefore provide independent and comparable data for the analysis of population genetic structure.

The study will provide a further example, along with the study of Britton-Davidian *et al.* (2007) and Hardouin *et al.* (2010), comparing mitochondrial and nuclear data to examine relations of house mouse populations, and therefore interpret house mouse colonisation history. In particular, there is an interest to see if the nuclear data reflect in any way the differing colonisation histories of the Azorean islands as suggested by the mitochondrial data.



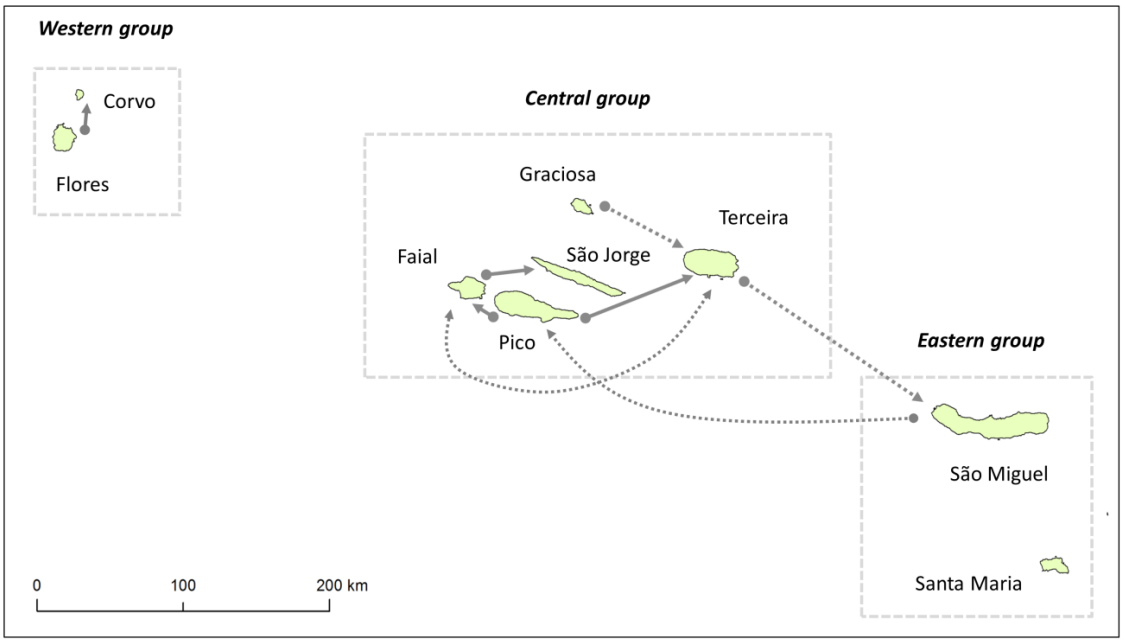


Figure 1. Map overview and location of the Azores archipelago relative to the Madeira archipelago and mainland Portugal (above), detail of the Madeira archipelago (middle) and detail of the Azores archipelago (below). Solid lines and arrows indicate significant gene flow between populations as determined by BAYESASS. Dotted lines and arrows indicate estimated individual mouse exchanges between island populations, as determined by GENECLASS (see text).

The impact that colonisation of small islands has on genetic diversity has also been examined previously in the house mouse using nuclear markers (Berry & Peters 1977, Navajas y Navarro *et al.* 1989, Jones *et al.* 2011b). Here we have the opportunity to consider the colonisation of nine islands within an archipelago, each of differing geographical area and human population size.

The house mouse is subdivided into a number of genetically distinctive subspecies (Boursot *et al.* 1993). Due to the commensal association with humans, another aspect of house mouse colonisation, therefore, is the opportunity for these subspecies to be transported to the same destination, potentially causing mixing of the subspecific genomes. On the basis of morphology (Mathias *et al.* 1998) and

mitochondrial DNA (Gabriel *et al.* submitted), the Azores islands are occupied by the western house mouse (*M. m. domesticus*). However, given the use of the Azores by ships from all parts of Europe, it is reasonable to expect a degree of representation of the eastern house mouse (*M. m. musculus*), which is found from eastern Germany and Sweden (including parts of Denmark and Norway) eastwards (Boursot *et al.* 1993, Jones *et al.* 2010). We test this using an X- and a Y-linked marker which distinguish *M. m. domesticus* and *M. m. musculus*. Thus, overall, we use markers from each one of the different chromosomes in the house mouse in order to examine the genetic structure and colonisation history of the populations in the Azores.

Materials and methods

Sample collection

DNA samples from a total of 380 house mice were typed (Table 1). Two hundred and fifty-two samples from all nine Azorean islands were collected in connection with a previous study (Gabriel *et al.* submitted). For purposes of comparison with the Azorean material, samples from mainland Portugal (n=64) and the neighbouring Madeiran archipelago were also included (Madeira, n=56; Porto Santo, n=8) (Gündüz *et al.* 2001, Förster *et al.* 2009).

Microsatellites amplification and typing

Nineteen proximal (centromeric) di-nucleotide microsatellite loci (one locus per autosomal chromosome) were selected from the Whitehead Institute database (no longer available) on the basis of their mapped position in terms of

centiMorgans from the centromere (in brackets), high specificity and polymorphism: D1Mit64 (3.67 cM), D2Mit1 (2.23 cM), D3Mit117 (1.96 cM), D4Mit103 (2.77 cM), D5Mit145 (3.37 cM), D6Mit86 (1.81 cM), D7Mit75 (3.81 cM), D8Mit58 (2.52 cM), D9Mit218 (2.46 cM), D10Mit188 (5.52 cM), D11Mit74 (3.52 cM), D12Mit145 (0.00 cM), D13Mit154 (13.12 cM), D14Mit48 (7.08 cM), D15Mit12 (1.8 cM), D16Mit32 (2.33 cM), D17Mit19 (2.68 cM), D18Mit219 (4.46 cM) and D19Mit32 (3.04 cM). All loci were amplified in multiplex reactions of at least four markers using FAM, PET, NED and VIC fluorescent-labelled reverse primers.

Template DNA was amplified via the polymerase chain reaction (PCR) in 5 μL reactions using the Qiagen Multiplex PCR Kit, following the manufacturer's guidelines. Cycling conditions were 15 min at 95°C, followed by 30 cycles of 30 s at 94°C, 90 s at 58°C and 1 min at 72°C, and a 10 min final elongation step at 72°C. The fluorescent PCR products were size-separated in an Applied Biosystems ABI 3130XL automated DNA sequencer using a Gene-Scan-600 LIZ size standard (ABI) and allele sizes were scored using GENEMAPPER v3.7 (Applied Biosystems). To control for scoring errors, a random selection of 5% of all individuals were rerun and rescored anonymously. For each locus, all peak profiles were checked by eye twice and histograms were built based on raw allele sizes. With the bins created for each of the nineteen loci, individual genotypes were assigned.

Btk and Zfy2 amplification and typing

House mice from all Azorean islands were scored at two loci with fixed nucleotide differences between the subspecies *M. m. musculus* and *M. m. domesticus*: *Btk* on the X chromosome and *Zfy2* on the Y chromosome.

Btk. This marker corresponds to an insertion in the Bruton agammaglobulinemia tyrosine kinase gene (*Btk*) on the X chromosome that has become fixed in *M. m. domesticus* but is absent from *M. m. musculus*. *Btk* was amplified using the primer pair 5'- AAT GGG CTA GCG TAG TGC AG -3' and 5'- AGG GGA CGT ACA CTC AGC TTT -3'. Cycling conditions included an initial step at 96ºC for 5 min, followed by 35 cycles of 95°C for 45 s, 62°C for 45 s and 72°C for 45 s, with a final extension step of 72°C for 10 min. PCR products were run on a 2% agarose gel and the allele sizes scored: a 342 bp fragment indicated the presence of the insertion and therefore, a *domesticus*-like specimen, while the screening of a 206 bp fragment indicated the absence of the insertion and the specimen scored as *musculus*-like (Munclinger *et al.* 2002).

Zfy2. This marker corresponds to an 18 bp deletion in *M. m. musculus* in the last exon of Zinc Finger protein 2 on the Y chromosome (*Zfy2*) which is absent in *M. m. domesticus*. The portion of the *Zfy2* gene containing the deletion was amplified using the primers 5'- CAT TAA GAC AGA AAA GAC -3' and 5'- GTG AGG AAA TTT CTT CCT -3'. Cycling conditions included an initial step of 96°C for 5 min, followed by 30 cycles of 95°C, 60°C, 72°C for 1 min and a final extension of 10 min. Amplification products were run on a high resolution 4% MetaPhor® gel and allele size differences scored: *domesticus*-like males should show a single 202 bp fragment (the product of both *Zfy1* and *Zfy2*) and *musculus*-like males should show the same 202 bp band in addition to a second fragment of 184 bp (corresponding to the 18 bp deletion on the *Zfy2* product) (Munclinger *et al.* 2002).

Data analysis

Within-population diversity was examined with the following indices: mean number of alleles per locus (A), mean number of private alleles per locus (A_P), and observed (H_O) and expected (H_E) heterozygosities, using the program FSTAT 2.9.3.2 (Goudet 2001).

The differentiation among the 12 populations on the Azorean and Madeiran islands and in mainland Portugal was assessed using pairwise F_{ST} (Weir and Cockerham 1984).

Genetic structure was also assessed through a Bayesian approach, as implemented in STRUCTURE version 2.3.3 (Pritchard *et al.* 2000, Falush *et al.* 2003), through the identification of the number of genetic clusters (K) within our dataset and by assigning individuals to these clusters on the basis of their multi-locus genotypes. All 12 populations were subject to a joint STRUCTURE analysis with multiple K values being tested (K=2 to K=20). Simulations were run under the 'admixture' ancestry model (each individual may have mixed ancestry, potentially carrying a fraction of its genome from each of the K populations) and the 'correlated allele frequencies' model (frequencies in different populations are likely to be similar due to common ancestry or migration).

As most populations were from discrete islands and all the sampling locations were known, the geographic origin of mice was incorporated as prior information in order to improve the accuracy of the inference. Each prior of K was run for 500 000 Monte Carlo Markov Chains (MCMC) replicates after a burn-in period of 100 000 iterations. Five independent simulations were run for each value of K to assess stability and the average posterior probability among runs and the

standard error were calculated for each value of K. ΔK statistics as developed by Evanno *et al.* (2005) and implemented in Dent & von Holdt (2011) were used to detect the highest hierarchical level of structure among all the tested values of K. A graphical visualisation of the results was prepared in DISTRUCT 1.1 (Rosenberg 2004).

Genetic structure was further assessed with GENETIX 4.05.2 (Belkhir *et al.* 1999) by performing a factorial correspondence analysis (FCA) on the allelic frequencies obtained for the nine Azorean island mouse populations.

In order to estimate the degree of on-going exchange of mice between all the screened populations, particularly within the Azorean archipelago, 'real-time' migration was assessed through a Bayesian analysis implemented in GENECLASS 2.0 (Piry *et al.* 2004). Based on individual multi-locus genotypes, the likelihood of each house mouse being a resident or a first generation migrant was estimated using the Bayesian method of Rannala and Mountain (1997) with a Monte Carlo resampling algorithm (10 000 simulated individuals, α set to 0.01) following Paetkau *et al.* (2004). The detection of migrants was based on the maximization of the ratio computed as $L=L_{home}/L_{max}$, where L_{home} describes the likelihood to sample an individual from its original population and L_{max} describes the corresponding highest likelihood if sampled from all populations including the one where the individual was sampled from (Paetkau *et al.* 2004).

The results obtained in the assignment tests in GENECLASS were compared with estimates of migration rates among populations calculated with the Bayesian method implemented in BAYESASS 1.3 (Wilson and Rannala 2003). The latter was used to obtain an estimate of the magnitude and direction of contemporary gene flow among all 12 populations. This method estimates mean immigration rates

among populations and their confidence intervals (CI). The MCMC algorithm was initially run for 1 000 000 iterations, discarded as burn-in, followed by 3 000 000 iterations to ensure stationarity, sampled every 2000 steps. These default settings proved sufficient to reach convergence based on visual inspection of likelihood scores.

Mantel tests were run with IBD 1.52 (Bohonak 2002) to investigate whether there was a correlation between the genetic differentiation (F_{ST}) and the geographical distance between pairs of populations (km), considering all populations and particularly the nine islands of the Azorean archipelago. The level of significance was estimated through 100 000 randomizations using the same software. Significantly positive correlations are indicative of isolation by distance (Bohonak 2002).

Results

Nuclear markers to identify subspecies

Based on the screening of *Btk* and *Zfy2*, the Azorean populations were entirely *M. m. domesticus*-like, with the exception of one mouse from Terceira which was *musculus*-like for the *Zfy2* marker on the Y-chromosome and another from São Miguel that was heterozygous for *Btk*, with one fragment representing each subspecies.

Microsatellite markers

Genetic diversity within populations

Overall, the 19 loci showed high levels of polymorphism with 8-25 alleles per locus, yielding a total of 295 alleles across all sampling locations. All populations showed moderate to high levels of both observed (H_O) and expected (H_E) heterozygosity with values of the former ranging from 0.412 – 0.667 and values of the latter from 0.483 – 0.796 (Table 1).

Table 1. Summary of genetic diversity measures based on the multilocus analysis of 19 microsatellite loci scored in house mouse populations in the Azores, Madeira, Porto Santo and mainland Portugal.

Population	N	A	A_P	H_O	H_E	Area (km^2)	Human Pop
Santa Maria	24	5.90	0.316	0.539	0.612	97	5 547
São Miguel	55	10.26	0.526	0.667	0.717	747	137 699
Terceira	37	8.74	0.105	0.562	0.670	402	56 062
Graciosa	23	5.95	0.105	0.533	0.623	61	4 393
São Jorge	26	4.53	0.053	0.412	0.502	246	8 998
Pico	26	4.79	0.105	0.428	0.531	448	14 144
Faial	28	7.16	0.368	0.537	0.690	173	15 038
Flores	24	5.42	0.053	0.556	0.666	142	3 791
Corvo	9	3.05	0.000	0.438	0.483	17	430
Madeira	56	9.53	0.316	0.569	0.729	741	262 302
Porto Santo	8	4.68	0.158	0.618	0.671	43	5 483
Portugal	64	11.00	1.211	0.555	0.796	92 090	10 561 614

N, number of individuals scored
A, mean number of alleles per locus
A_P, mean number of private alleles per locus
H_O, observed heterozygosity
H_E, unbiased expected heterozygosity
Human pop, human population size

The house mouse population on Corvo, the smallest island with the lowest human population size among those sampled, showed low diversity with all measures, including a complete absence of private alleles (Table 1). This compares with Porto Santo, which is the second smallest island sampled, and for which there

was a similar number of individuals scored as Corvo, but which shows mid-range numbers of private alleles and mid to high heterozygosities.

Also worthy of note are the high diversity values for São Miguel (Table 1). However, even though this population has the highest mean number of private alleles per locus (A_P) among the islands sampled, the value is still well below that recorded in Portugal with a similar number of individuals scored. Finally, São Jorge shows relatively low diversity values among the populations considered (Table 1).

There is little correspondence between H_O for each mouse population and the human population size for the same geographic unit (as revealed by a Spearman's rank correlation, $r_s = 0.371$, $p = 0.236$). Nor, when analysing the Azorean dataset alone, which had H_O values spanning the whole range (0.412 – 0.667), did this relationship hold ($r_s = 0.417$, $p = 0.265$). This is despite the lowest diversity values noted for the island with the smallest human population size (Corvo) and the highest values for the island with the largest human population size (São Miguel). A relationship between H_O and island area also failed to emerge ($r_s = 0.267$, $p = 0.488$); nor is there an east to west pattern of decreasing diversity as might be expected if there was stepwise colonisation of the Azores from Portugal.

Genetic structure

The smallest pairwise F_{ST} values were observed between certain islands in the Azores: São Miguel and Terceira, Graciosa, Pico and Faial; Terceira and Graciosa, Pico and Faial; Graciosa and Faial (Table 2). This same set of Azorean islands was also found to be closely similar in the STRUCTURE analysis (Figure 2). Using the test of Evanno *et al.* (2005) there were shown to be six genetic clusters

for this dataset. From this clustering, São Miguel, Terceira, Graciosa, Pico and Faial were found to show a largely common identity (with some admixture though), mirroring the F_{ST} results.

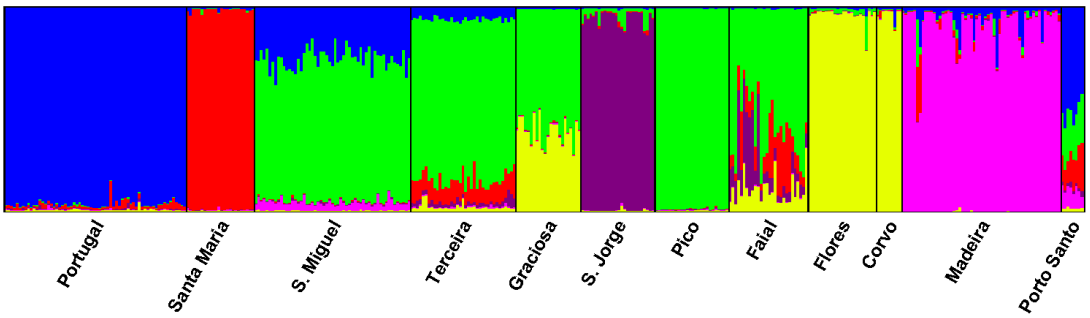


Figure 2. Population structure as inferred by Bayesian STRUCTURE clustering based on microsatellite genotypes of 12 house mouse populations (9 islands of the Azorean archipelago, 2 islands of Madeiran archipelago, and mainland Portugal). Each of the 380 individuals is represented by a thin vertical coloured line, representing the estimated proportion of membership of the genome originating from each of K inferred clusters. Black lines separate different pre-defined populations. This bar plot represents the most likely number of genetic clusters, K=6, as determined by the method of Evanno *et al*. (2005).

This represents the Central group of the Azorean islands to the exclusion of São Jorge and the inclusion of São Miguel from the Eastern group (see Figure 1). The mice on Flores and Corvo (the Western group of Azorean islands: Figure 1) were the main representatives of another genetic cluster. Santa Maria of the Eastern group, São Jorge of the Central group, and both Madeira and mainland Portugal each had their own genetic identity, while the small sample of Porto Santo mice showed substantial admixture.

Table 2. Pairwise F_{ST} values among house mouse populations from the Azores, Madeira, Porto Santo and mainland Portugal. Comparisons among the Azorean populations are highlighted in grey. POR – mainland Portugal, SMA – Santa Maria, SMI – São Miguel, TER – Terceira, GRA – Graciosa, SJO – São Jorge, PIC – Pico, FAI – Faial, FLO –Flores, COR – Corvo, MAD – Madeira, PSAN – Porto Santo.

	POR	SMA	SMI	TER	GRA	SJO	PIC	FAI	FLO	COR	MAD
SMA	0.127										
SMI	0.121	0.221									
TER	0.145	0.231	0.035								
GRA	0.148	0.276	0.069	0.100							
SJO	0.223	0.288	0.221	0.239	0.280						
PIC	0.222	0.336	0.088	0.057	0.141	0.334					
FAI	0.134	0.209	0.072	0.083	0.105	0.202	0.119				
FLO	0.120	0.262	0.143	0.173	0.141	0.252	0.237	0.157			
COR	0.188	0.376	0.243	0.295	0.223	0.410	0.368	0.231	0.186		
MAD	0.120	0.201	0.179	0.210	0.213	0.252	0.277	0.196	0.178	0.277	
PSAN	0.162	0.276	0.170	0.205	0.233	0.334	0.279	0.208	0.223	0.375	0.195

The same general results for the Azores were found with the FCA (Figure 3). This basically showed a separation of the Western, Central and Eastern island groups, except that São Miguel clustered with the Central group islands. Within that Central group cluster São Jorge was the most distinctive.

Current gene flow between populations

We applied two distinct tests to assess migration among the studied populations. BAYESASS and GENECLASS were used, respectively, to estimate migration rates and to test for statistically significant cases of assignment to a different population than the one specimen was sampled from.

BAYESASS revealed that current migration rates between all surveyed populations are generally very low. The only exceptions are among islands of the

Azorean Central group (Terceira, São Jorge, Pico and Faial) and among those of the Azorean Western group (Flores and Corvo) (Figure 1, Table 3). Nevertheless, migration events among these islands are mostly asymmetrical, occurring particularly from Pico to Terceira (mean migration rate = 0.292), from Pico to Faial (0.232), from Faial to São Jorge (0.283) and from Flores to Corvo (0.192) (Table 3).

Considering the whole dataset, GENECLASS detected ten first generation migrants ($p \leq 0.01$), eight of which were found to be exchanging between islands of the Azores, mostly from the Central group (Figure 1).

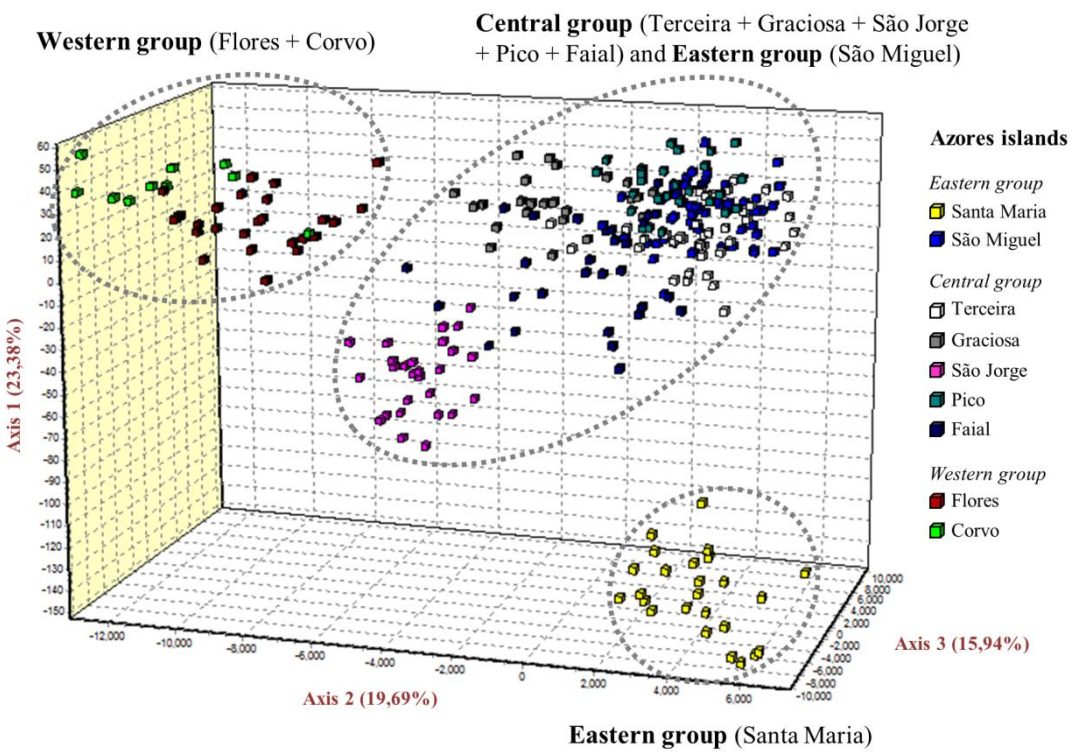


Figure 3. Population structure based on Factorial Correspondence Analysis (FCA) of all nine Azorean island populations of house mouse typed for 19 microsatellite loci. Each square represents an individual multilocus genotype coloured according to the island from where it was sampled.

Terceira was either the main source of migrating mice (two captured in São Miguel and two in Faial) but also the main recipient of incoming mice (two originally from Faial and one from Graciosa). Also, one mouse estimated to be originally from São Miguel was captured in Pico. In the neighbouring archipelago, two individuals collected on Madeira were assigned as incoming migrants from the nearby island of Porto Santo.

Both BAYESASS and GENECLASS detected no exchange of mice between the Azorean and Madeiran archipelagos or between mainland Portugal and any of the islands.

Isolation by distance

The joint analysis of all 12 populations (Azores, Madeira and mainland Portugal) showed no evidence of isolation by distance (Z=9777, r=0.052, p=0.409) while there was a nearly significant value when the analysis was limited to the Azorean islands (Z=2009, r =0.359, p=0.069).

Table 3. Means and 95% confidence intervals of the posterior distributions for migration rates among the Azorean islands, Madeira, Porto Santo and mainland Portugal house mouse populations as estimated by BAYESASS. Populations from which individuals were sampled are listed in rows, while the populations from which they migrated are listed in columns. Values in bold along the diagonal are self-recruitment rates, corresponding to the proportion of individuals derived from the source populations. Values in parentheses below migration rates correspond to the 95% CI. Migration rates between populations significantly different from zero are highlighted in grey.

Table 3. Means and 95% confidence intervals of the posterior distributions for migration rates among the Azorean islands, Madeira, Porto Santo and mainland Portugal house mouse populations as estimated by BAYESASS (see previous page).

Migration into \ Migration from	Portugal	Sta. Maria	S. Miguel	Terceira	Graciosa	S. Jorge	Pico	Faial	Flores	Corvo	Madeira	Porto Santo
Portugal	**0.995** (0.982 - 1.000)	0.000 (0 - 0.004)	0.000 (0 - 0.004)	0.000 (0 - 0.005)	0.000 (0 - 0.005)	0.000 (0 - 0.004)	0.000 (0 - 0.005)	0.000 (0 - 0.004)	0.000 (0 - 0.004)	0.000 (0 - 0.003)	0.000 (0 - 0.003)	0.000 (0 - 0.004)
Sta. Maria	0.001 (0 - 0.010)	**0.988** (0.954 - 1.000)	0.001 (0 - 0.017)	0.001 (0 - 0.010)	0.001 (0 - 0.011)	0.001 (0 - 0.010)	0.001 (0 - 0.010)	0.001 (0 - 0.010)	0.001 (0 - 0.012)	0.001 (0 - 0.010)	0.001 (0 - 0.011)	0.001 (0 - 0.010)
S. Miguel	0.001 (0 - 0.005)	0.001 (0 - 0.005)	**0.993** (0.973 - 1.000)	0.000 (0 - 0.005)	0.001 (0 - 0.008)	0.001 (0 - 0.006)	0.001 (0 - 0.007)	0.000 (0 - 0.005)	0.000 (0 - 0.005)	0.000 (0 - 0.004)	0.001 (0 - 0.007)	0.001 (0 - 0.007)
Terceira	0.003 (0 - 0.017)	0.003 (0 - 0.019)	0.004 (0 - 0.028)	**0.676** (0.667 - 0.698)	0.006 (0 - 0.033)	0.003 (0 - 0.017)	0.292 (0.245 - 0.327)	0.003 (0 - 0.017)	0.003 (0 - 0.017)	0.003 (0 - 0.018)	0.003 (0 - 0.018)	0.003 (0 - 0.020)
Graciosa	0.002 (0 - 0.020)	0.002 (0 - 0.018)	0.003 (0 - 0.026)	0.002 (0 - 0.021)	**0.950** (0.841 - 0.999)	0.002 (0 - 0.018)	0.026 (0 - 0.117)	0.002 (0 - 0.018)	0.002 (0 - 0.018)	0.002 (0 - 0.019)	0.002 (0 - 0.018)	0.002 (0 - 0.018)
S. Jorge	0.004 (0 - 0.023)	0.004 (0 - 0.023)	0.004 (0 - 0.027)	0.004 (0 - 0.027)	0.004 (0 - 0.024)	**0.678** (0.667 - 0.708)	0.004 (0 - 0.027)	0.283 (0.224 - 0.325)	0.003 (0 - 0.023)	0.004 (0 - 0.026)	0.004 (0 - 0.027)	0.004 (0 - 0.024)
Pico	0.001 (0 - 0.009)	0.001 (0 - 0.009)	0.001 (0 - 0.011)	0.001 (0 - 0.009)	0.001 (0 - 0.012)	0.001 (0 - 0.010)	**0.988** (0.960 - 1.000)	0.001 (0 - 0.011)	0.001 (0 - 0.012)	0.001 (0 - 0.010)	0.001 (0 - 0.009)	0.001 (0 - 0.008)
Faial	0.003 (0 - 0.022)	0.003 (0 - 0.023)	0.003 (0 - 0.021)	0.004 (0 - 0.021)	0.003 (0 - 0.022)	0.004 (0 - 0.022)	0.232 (0.167 - 0.286)	**0.733** (0.693 - 0.785)	0.003 (0 - 0.022)	0.004 (0 - 0.024)	0.003 (0 - 0.019)	0.003 (0 - 0.021)
Flores	0.001 (0 - 0.010)	0.001 (0 - 0.011)	0.001 (0 - 0.015)	0.001 (0 - 0.011)	0.001 (0 - 0.011)	0.001 (0 - 0.008)	0.001 (0 - 0.008)	0.001 (0 - 0.011)	**0.987** (0.954 - 1.000)	0.001 (0 - 0.014)	0.001 (0 - 0.012)	0.001 (0 - 0.010)
Corvo	0.011 (0 - 0.067)	0.011 (0 - 0.062)	0.011 (0 - 0.065)	0.011 (0 - 0.066)	0.010 (0 - 0.065)	0.011 (0 - 0.064)	0.010 (0 - 0.061)	0.012 (0 - 0.061)	0.192 (0.066 - 0.303)	**0.697** (0.668 - 0.768)	0.011 (0 - 0.067)	0.010 (0 - 0.057)
Madeira	0.001 (0 - 0.008)	0.001 (0 - 0.008)	0.005 (0 - 0.021)	0.001 (0 - 0.008)	0.001 (0 - 0.007)	0.001 (0 - 0.007)	0.001 (0 - 0.009)	0.001 (0 - 0.009)	0.001 (0 - 0.008)	0.001 (0 - 0.009)	**0.978** (0.952 - 0.994)	0.008 (0 - 0.028)
Porto Santo	0.003 (0 - 0.032)	0.003 (0 - 0.029)	0.003 (0 - 0.031)	0.003 (0 - 0.027)	0.003 (0 - 0.033)	0.003 (0 - 0.028)	0.003 (0 - 0.029)	0.003 (0 - 0.029)	0.004 (0 - 0.036)	0.003 (0 - 0.028)	0.003 (0 - 0.031)	**0.966** (0.876 - 1.000)

Discussion

Genetic diversity

The overall levels of heterozygosity throughout the Azorean archipelago were high (H_O=0.412–0.667) when compared with other insular populations of the western house mouse also studied with microsatellites (e.g. H_O=0.34–0.51 on islands of the sub-Antarctic Kerguelen archipelago, Hardouin *et al.* 2010; H_O=0.000–0.567 on the Faroe islands, Jones *et al.* 2011; H_O=0.28 on Iceland, Jones *et al.* 2012). The levels of heterozygosity on the neighbouring archipelago, Madeira (H_O=0.569) and Porto Santo (H_O=0.618) were also high. These latter values fit perfectly with the findings of Britton-Davidian *et al.* (2007) with allozymes. They noted that the genetic diversity of the Madeiran archipelago was considerably higher than typically expected for island populations, with values comparable to those found in continental Europe. Clearly, what applies to the Madeiran archipelago, also applies to the Azores.

Among the islands of the Azores, the lowest genetic diversity is found in Corvo (H_O=0.412). However, the drop in diversity was very minor compared with what was found by Jones *et al.* (2011b) in the Faroe Islands. Here, three of the islands had H_O values of less than 0.1, all very small islands with human population sizes of less than 50 inhabitants. Corvo had low values for all the measures of diversity relative to the other islands in the Azores. São Jorge also had relatively low values, mirroring low mitochondrial D-loop diversity (Gabriel *et al.* submitted).

Overall then, it appears that all the islands sampled in the Azorean and Madeiran archipelagos have managed to maintain long-lasting high population

sizes, so that there is no genetic signal of population bottlenecks. Mice have been recorded as common on the Azorean islands since the very early days of human settlement (Drouët 1861). In particular, on the westernmost Flores island (the last to be reached and populated), mice soon became a pest after human colonisation, as reported by a 16th century historian (Frutuoso 1590) and populations appear to have kept high numbers of individuals to present day. Clearly, it helps that none of the Azorean or Madeiran islands studied are exceptionally small in area. Also, the mild weather conditions year round and suitable habitat both in and around buildings (domestic, commercial and farm) and in the plentiful agricultural countryside (Nunes *et al.* 2005, Mathias *et al.* 1998) are conducive to constant large population sizes. In the Azores, these large populations are maintained despite the existence of a periodic pest management control policy (Mathias *et al.* 1998).

Genetic determination of colonisation history and dispersal

The subspecies of house mouse that clearly dominates the colonisation history of the Azores is *M. m. domesticus*. This was evident from morphology (Mathias *et al.* 1998) and mitochondrial DNA (Gabriel *et al.* submitted), and is now confirmed by two nuclear markers on the sex chromosomes. This is as expected given that Portugal dominated the settlement of the Azores by people (Frutuoso 1590) and that the subspecies of house mouse present in Portugal is *M. m. domesticus* (Mathias *et al.* 1999). However, the global nature of the colonisation process in the house mouse is illustrated by two out of the 252 mice studied being hybrids between *M. m. domesticus* and *M. m. musculus*, on the basis of genotype at

the X- and Y-linked sequences studied. These two individuals were from Terceira and São Miguel, which are two islands that have frequently been used as a stop-over on ship routes both during historical (Frutuoso 1590) and recent times. The *M. m. musculus* contribution to the Azores house mouse gene pool would ultimately have derived from eastern Europe or Asia, but it is unknown whether *M. m. musculus* arrived on the islands and hybridized, or whether the initial hybridization event took place elsewhere. Whatever the truth, there has been genetic mixing of mice that have come from geographically distant sources.

Previous mitochondrial DNA results have provided much detail on both relatively short- and long-distance colonisation of the islands (Gabriel *et al.* submitted). The relatively short-distance colonisation relates to the complex network of interactions involving the Azores, the Madeiran archipelago, Spain and Portugal, as revealed by the sharing of D-loop sequences among these entities. Thus, there is evidence that particular islands or groups of islands in the Azores were colonized from the Madeiran archipelago, Spain and Portugal. In terms of long-distance colonisation, there is a D-loop clade present in the Azores that is typically associated with Atlantic northern Europe, found on Santa Maria and Terceira (Gabriel *et al.* submitted).

For the microsatellites studied here, the only comparison is between Portugal, the Madeiran archipelago and the Azorean archipelago. Therefore, there was no chance to assess source areas of long-distance colonisation events. However, the data were amenable to reveal a situation where most Azorean islands had a history of house mouse colonisations linked to Madeira and Portugal, and other islands had a history completely separate.

This is not what was observed. Instead, the analyses of genetic structure showed Azorean island populations grouping together to the exclusion of other populations. On the basis of tests of migration, the clustering reflected current or recent gene flow, and is exemplified by the Central islands of the Azores. The two islands in the Western group (Flores and Corvo) also tended to cluster as a single entity in the analyses.

When numerous islands within an archipelago are occupied by a particular species, as for the house mouse in the Azores, the nature of any genetic subdivision among those island populations may be particularly illuminating. In the case of the Azores, the geographically separated groups of islands (Eastern, Central, Western) generally represent separate genetic entities of house mice. Given what is said above, this is not surprising, because higher levels of migration would be expected between closely located islands, and the large marine gaps would reduce the numbers of boat journeys between islands and the probability of mice to be transported.

The main exception is São Miguel, an Eastern group island that clustered together with the Central Group (Figure 3). This actually fits extremely well with what is known about human movements in the archipelago. São Miguel is the administrative centre of the Azores, is the largest island and has the largest human population (137,699 inhabitants). There is, and has long been, particularly intense human interactions (including ship traffic) between São Miguel and Terceira (da Costa 1978). Terceira has the second largest human population on the archipelago (56,062 inhabitants; all other islands have a human population size of 15,000 inhabitants or less). Two of the migrants detected by GENECLASS were between Terceira and São Miguel (Figure 1).

It is interesting to contrast the results we have obtained for the Azores with those obtained with mitochondrial DNA and allozymes relating to colonisation of Madeira (Gündüz *et al.* 2001, Britton-Davidian *et al.* 2007, Förster *et al.* 2009). There is great similarity between the mitochondrial DNA sequences found in Madeira and those found in northern continental Europe, and practically no overlap in sequences between Madeira and Portugal, despite 600 years of common history (the sequence of only one individual in Madeira suggested it had a derivation from Portugal: Förster *et al.* 2009). We take this to likely reflect original colonisation of Madeira from northern Europe, and difficulty for female mice coming from Portugal to displace resident females. In contrast there is greater similarity in allozymes between Madeira and Portugal than between Madeira and northern Europe, presumably resulting from secondary colonisations, with males originating from Portugal penetrating the pre-existing Madeiran population. For the Azores it appears that Portugal, Madeira, Spain and possibly northern Europe may have been involved in the initial colonisation as defined by mitochondrial DNA (Gabriel *et al.* submitted). In terms of secondary colonisation (penetration of males into pre-existing populations), our microsatellite analysis suggests that there have been recent inter-island movements. However, it is likely that such movements have been a frequent occurrence since the settlement of the Azores. With regards the current movements suggested by GENECLASS, it appears somewhat surprising to find so many recent migrants, and this may be an overestimate. Nevertheless, the case is made with both BAYESASS and GENECLASS for frequent migration within island groups in the Azores, and as for Madeira, the mitochondrial DNA pattern is apparently undisturbed.

With regards the particular role of males and females in gene flow in the house mouse, we are currently developing a phylogeography of the species involving Y-chromosome markers to allow us to examine the male component independently (Jones *et al.* in prep.).

In conclusion, the genetic structure of populations of house mice in the Azores islands as revealed by microsatellites most likely represents recent patterns of gene flow, with populations on islands within the same island group generally clustering together genetically. The permanent settlement of the Azorean islands by people and plentiful movement of livestock, agricultural materials and food supplies between them provides the opportunity for frequent migration of house mice (see Pocock *et al*. 2005), particularly given that the mice are living both outdoors and indoors and therefore are readily available to become part of whatever is being transported.

Acknowledgements

This study was funded by the scholarship SFRH/BD/21437/2005 from Fundação para a Ciência e a Tecnologia (Portugal) granted to SIG.

References

Belkhir K, Borsa P, Goudet J, Bonhomme F (1999) *GENETIX: logiciel sous Windows pour la génétique des populations.* Laboratoire Génome et Populations, CNRS-UPR, Université de Montpellier II, Montpellier.

Berry RJ, Peters J (1977) Heterogeneous heterozygosities in *Mus musculus* populations. *Proceedings of the Royal Society of London* B, 197, 485–503.

Bohonak AJ (2002) IBD (Isolation By Distance): a program for analyses of isolation by distance. *Journal of Heredity*, 93, 153–154.

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffra J-C, 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.

Bonhomme F, Searle JB (in press) *House mouse phylogeography*. In *Evolution of the house mouse* (Cambridge Series in Morphology and Molecules). Eds Macholán M, Baird SJE, Munclinger P, Piálek J. Cambridge University Press, Cambridge.

Boursot P, Auffray J-C, Britton-Davidian J, Bonhomme F (1993) The evolution of house mice. *Annual Review of Ecology and Systematics*, 24, 119–152.

Britton-Davidian J, Catalan J, Lopez J, Ganem G, Nunes AC, Ramalhinho MG, Auffray JC, Searle JB, Mathias ML (2007) Patterns of genic diversity and structure in a species undergoing rapid chromosomal radiation: an allozyme analysis of house mice from the Madeira archipelago. *Heredity*, 99, 432–443.

da Costa S (1978) *Revista de história económica e social*. Edições 1–4. Sá da Costa Editora, Lisbon.

Diffie BW, Winius GD (1977) *Foundations of the Portuguese Empire*, *1415-1580*. University of Minnesota Press, Minneapolis.

Drouët H (1861) *Eléments de la faune Açoréenne*. J.B. Baillière & Fils, Librairie de l'Académie de Médecine, Paris.

Earl DA, von Holdt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611–2620.

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

Förster DW, Gündüz İ, Nunes AC, Gabriel S, Ramalhinho MG, Mathias ML, Britton-Davidian J, Searle JB (2009) Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology*, 18, 4477–4494.

Frutuoso G (1590) *Saudades da Terra – Livro I - VI* (Ed. 1998). Instituto Cultural de Ponta Delgada, Ponta Delgada.

Gabriel SI, Jóhannesdóttir F, Jones EP, Searle JB (2010) Colonization, mouse-style. *BMC Biology*, 8, 131.

Gabriel SI, Mathias ML, Searle JB (submitted) Of mice and the 'Age of Discovery': the complex colonization history of the Azorean archipelago by the house mouse (*Mus musculus*) as revealed by mitochondrial DNA variation. *Journal of Evolutionary Biology.*

Goudet J (2001) FSTAT: a program to estimate and test gene diversities and fixation indices (version 2.9.3). Updated from Goudet J. 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, 86, 485–486.

Gündüz İ, Auffray J-C, Britton-Davidian J, Catalan J, Ganem G, Ramalhinho MG, Mathias ML, Searle JB (2001) Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology*, 10, 2023–2029.

Hardouin EA, Chapuis J-L, Stevens MI, van Vuuren BJ, Quillfeldt P, Scavetta RJ, Teschke M, Tautz D (2010) House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evolutionary Biology*, 10, 325.

Jones EP, van der Kooij J, Solheim R, Searle JB (2010) Norwegian house mice (*Mus musculus musculus/domesticus*): Distributions, routes of colonization and patterns of hybridization. *Molecular Ecology*, 19, 5252–5264.

Jones EP, Jóhannesdóttir F, Gündüz İ, Richards MB, Searle JB (2011a) The expansion of the house mouse into north–western Europe. *Journal of Zoology*, 283, 257–268.

Jones EP, Jensen J-K, Magnussen E, Gregersen N, Hansen HS, Searle JB (2011b) A molecular characterization of the charismatic Faroe house mouse. *Biological Journal of the Linnean Society*, 102, 471–482.

Jones EP, Skirnisson K, McGovern TH, Gilbert MTP, Willerslev E, Searle JB (2012) Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *BMC Evolutionary Biology*, 12, 35.

Mathias ML, Ramalhinho MG, Santos-Reis M, Petrucci-Fonseca F, Libois R, Fons R, de Carvalho GF, Oom MM, Collares-Pereira M (1998) Mammals from the Azores islands (Portugal): an updated overview. *Mammalia*, 62, 397–407.

Mathias ML (1999) *Mamíferos Terrestres de Portugal Continental, Açores e Madeira.* ICN, Lisbon.

Munclinger P, Božiková E, Šugerková M, Piálek J, Macholán M (2002) Genetic variation in house mice (*Mus*, Muridae, Rodentia) from the Czech and Slovak Republics. *Folia Zoologica*, 51, 81–92.

Navajas y Navarro M, Britton-Davidian J (1989) Genetic structure of insular Mediterranean populations of the house mouse. *Biological Journal of the Linnean Society*, 36, 377–390.

Nunes AC, Britton-Davidian J, Catalan J, Ramalhinho MG, Capela R, Mathias ML, Ganem G (2005) Influence of physical environmental characteristics and anthropogenic factors on the position and structure of a contact zone between two chromosomal races of the house mouse on the island of Madeira (North Atlantic, Portugal). *Journal of Biogeography*, 32, 2123–2134.

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, 13, 55–65.

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GeneClass2: A software for genetic assignment and first generation migrant detection. *Journal of Heredity*, 95, 536–539.

Pocock MJO, Hauffe HC, Searle JB (2005) Dispersal in house mice. *Biological Journal of the Linnean Society*, 84, 565–583.

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

Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe, *Molecular Ecology*, 17, 627–641.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Genetics*, 94, 9197–9201.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137–138.

Schwarz E, Schwarz HK (1943) The wild and commensal stocks of the house mouse, *Mus musculus*. *Journal of Mammalogy*, 24, 59–72.

Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry S, Giménez MD, Jóhannesdóttir F (2009) Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society, B* 276, 201–207.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163, 1177–1191.

Chapter 3

Paper III

Of mice and 'convicts': origin of the Australian house mouse, *Mus musculus*

Sofia I. Gabriel[1,2], Mark I. Stevens[3], Maria da Luz Mathias[2], Jeremy B. Searle[1,4]

[1] Department of Biology, University of York, Wentworth Way, York YO10 5DD, UK

[2] CESAM–Centre for Environmental and Marine Studies, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, 1749–016 Lisbon, Portugal

[3] South Australian Museum, and School of Earth and Environmental Sciences, University of Adelaide, SA 5000, Adelaide, Australia

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

Gabriel SI, Stevens MI, Mathias ML, Searle JB (2011) Of mice and 'convicts': origin of the Australian house mouse, *Mus musculus*. *PLoS ONE*, 6, e28622.

doi:10.1371/journal.pone.0028622

Abstract

The house mouse, *Mus musculus*, is one of the most ubiquitous invasive species worldwide and in Australia is particularly common and widespread, but where it originally came from is still unknown. Here we investigated this origin through a phylogeographic analysis of mitochondrial DNA sequences (D-loop) comparing mouse populations from Australia with those from the likely regional source area in Western Europe. Our results agree with human historical associations, showing a strong link between Australia and the British Isles. This outcome is of intrinsic and applied interest and helps to validate the colonization history of mice as a proxy for human settlement history.

Introduction

Invasive species have a major impact on Australia, threatening native biodiversity and causing massive costs to agriculture every year. The house mouse (*Mus musculus*) is a highly successful invader worldwide and particularly throughout mainland Australia and surrounding islands (Singleton 1995), but the source of this invasion is still unknown. Over the last century the house mouse has become a serious agricultural pest in Australia, particularly in grain growing regions where it shows aperiodic but increasingly frequent outbreaks (Singleton *et al.* 2005, 2007). This ability to respond rapidly to favorable environmental conditions, reaching plague proportions over large areas, is the most striking trait that distinguishes Australian house mice from their conspecifics around the world (Singleton *et al.* 2005). However, no attempt has yet been made to determine the geographical origin(s) of the first mouse colonists. Australia was probably the last continental landmass to be colonized by the house

mouse, presumably remaining mouse-free until the arrival and settlement of the first Europeans colonists, about two centuries ago (Singleton *et al.* 2005). The earliest Australian specimen registered at a museum was collected in Tasmania in 1884 (Cat. No. M106, Australian Museum). The arrival of the British First Fleet in 1788 is often cited as the most probable origin of house mice into Australia but earlier introductions were possible with Dutch ships charting the former "New Holland" coast since the early 1600s. To determine the source area of Australian mice we have followed a phylogeographic approach involving the analysis of mitochondrial D-loop sequences, as previously adopted elsewhere in Australasia when New Zealand and nearby islands were analyzed (Searle *et al.* 2009b). D-loop sequences are the only molecular marker for which there are substantial data throughout the distribution of the house mouse and mitochondrial DNA appears to be particularly useful as a phylogeographic tool to elucidate initial colonization events in this species (Gabriel *et al.* 2010). In our study, we compare the haplotypes of mice from Australia with available haplotypes of mice from likely geographical source areas in Western Europe, including published sequences from the British Isles (Searle *et al.* 2009a, Jones *et al.* 2011a) and new sequences from the Netherlands.

Materials and Methods

Samples

We obtained a total of 77 house mouse tissue and DNA samples representing 38 locations throughout mainland Australia, one from Kangaroo Island and six from Tasmania (Figure 2). The samples were provided by museums and private collectors (see Table S1 for details).

During the last two decades hundreds of D-loop sequences of *Mus musculus domesticus* have been published from locations all over Western Europe, including the British Isles, and our analysis includes all those available at the time (Figure S1). However, there were no sequences from the Netherlands and therefore, considering its potential as a source area for Australia, an additional set of 17 tissue samples from 10 Dutch localities were obtained from the Zoological Museum of Amsterdam and sequenced (Table S1).

Molecular methods

Genomic DNA was extracted from all samples using the Qiagen DNeasy Blood & Tissue extraction Kit according to the manufacturer's guidelines. For the Australian samples, complete mitochondrial D-loop and flanking regions were amplified by PCR with the primer pair L15774 - 5´-TGA ATT GGA GGA CAA CCA GT-3´ and H2228 - 5´-TTA TAA GGC CAG GAC CAA AC-3´, using standard concentrations of DNA and reagents as described in (Searle *et al.* 2009a). The DNA recovered from the Netherlands museum samples was degraded and was amplified in three overlapping fragments with the following primer pairs (designed here): dloopF1-5´-GCA CCC AAA GCT GGT ATT CT-3´ and dloopR1-5´-TTG TTG GTT TCA CGG AGG AT-3´, dloopF2 - 5´-ACT ATC CCC TTC CCC ATT TG - 3´ and dloopR2 - 5´-GAT TGG GTT TTG CGG ACT AA-3´, dloopF3-5´-ATA GCC GTC AAG GCA TGA AA-3´ and dloopR3-5´-GCT TTG CTT TGT TAT TAA GCT ACA-3´.

Data analysis

All haplotypes that were obtained are deposited in GenBank (JF277281-JF277300). These were combined with 367 previously published haplotypes of *Mus musculus domesticus* for phylogenetic analysis (Figure S1). For the broadest possible

comparison, the sequences were truncated to a standard length of 853 base pairs, including the whole D-loop. Based on the DNA substitution model retrieved from jModelTest0.1.1 (Posada 2008), the dataset was analyzed using the TIM2+I+Γ model for Bayesian MCMC inference (Ronquist and Huelsenbeck 2003) through two simultaneous runs of five chains each, during 10 million generations sampled every 1000 steps. After a 30% burn-in, both runs were used to generate a 50% majority rule consensus tree. *Mus musculus musculus* and *M. m. castaneus* sequences were used as outgroups. In comparisons involving Australian mice, haplotype (h) and nucleotide (π) diversities were determined using Arlequin 3.11 (Excoffier *et al*. 2005).

Results

All new sequences examined (Australian and Dutch) were typical of the western European house mouse *Mus musculus domesticus* rather than one of the other subspecies (*musculus*, *castaneus*, *gentilulus* (Prager *et al.* 1993, 1996, 1998) and therefore our phylogenetic tree was based purely on *domesticus* haplotypes and clades were labeled according to previous practice for this subspecies (Figures 1 and S1) (Jones *et al.* 2011a). The sequences from the potential source area of the Netherlands were largely clade E, which is also predominant in Britain and northern France (Figure S1). The Australian sequences harbored 13 distinct haplotypes belonging to four clades (Figure 1). The majority of sequences were evenly divided amongst two clades: E (N=32; haplotypes AUSTRALIA.01-04) and F (N=27; AUSTRALIA.05-07), each with a wide geographic distribution (Figure 2). Of the sequences assigned to clade E, 81% belonged to haplotype AUSTRALIA.01, previously described as U47431 found in the British Isles (Prager *et al.* 1993, Nachman *et al.* 1994, Searle *et al.* 2009a, Jones *et al.* 2011a), France (Jones *et al.* 2011a), Netherlands (this study, NETHERLANDS.02, N=1),

Germany (Prager *et al.* 1993), Denmark (Prager *et al.* 1993), Norway (Prager *et al.* 1996, Jones *et al.* 2010), Cameroon (Ihle *et al.* 2006), New Zealand (Searle *et al.* 2009b) and the sub-Antarctic Kerguelen Archipelago (Hardouin *et al.* 2010).

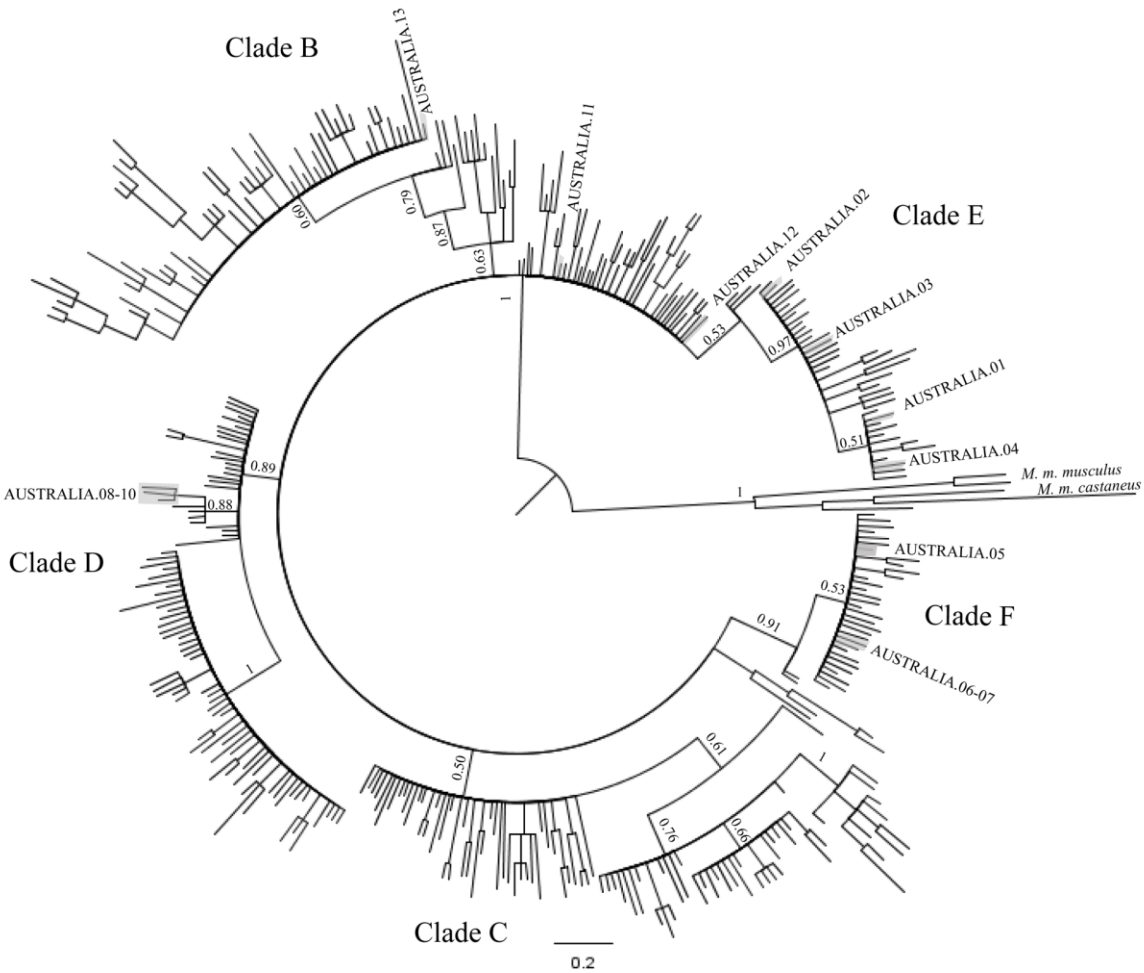


Figure 1. Phylogenetic tree for *Mus musculus domesticus* obtained after Bayesian analysis based on D-loop sequences (published and this study). Posterior probabilities ≥0.50 are shown at the nodes of each clade, labeled according to (Jones *et al.* 2011a). Australian haplotypes are highlighted in the tree (see Figure S1 for further details).

Likewise, 93% of the clade F sequences belonged to haplotype AUSTRALIA.05, previously described as U47495 and also found in Scotland (Jones *et al.* 2011a), Croatia (Prager *et al.* 1996) and New Zealand (Searle *et al.* 2009b). The remaining sequences corresponded to new haplotypes, distributed among clade D (*N*=9; AUSTRALIA.08-10),

restricted to a small area in Australia's east coast, clade B (*N*=1; AUSTRALIA.13), and eight sequences basal within the tree (AUSTRALIA.11-12), largely from Kangaroo Island and Tasmania (Figure 2). In addition to AUSTRALIA.01 and 05, only two other Australian haplotypes have previously been described elsewhere: AUSTRALIA.02 (*N*=4; mainland Australia) previously recorded as U47432 in Scotland (Prager *et al.* 1993), Germany (Prager *et al.* 1993), France (Hardouin *et al.* 2010, Jones *et al.* 2011a), the Netherlands (this study, NETHERLANDS.03, *N*=2) and New Zealand (Searle *et al.* 2009b), and AUSTRALIA.11 (*N*=6; Kangaroo Island and adjacent coast) previously recorded as Turkey.7 in Istanbul (Gündüz *et al.* 2005).

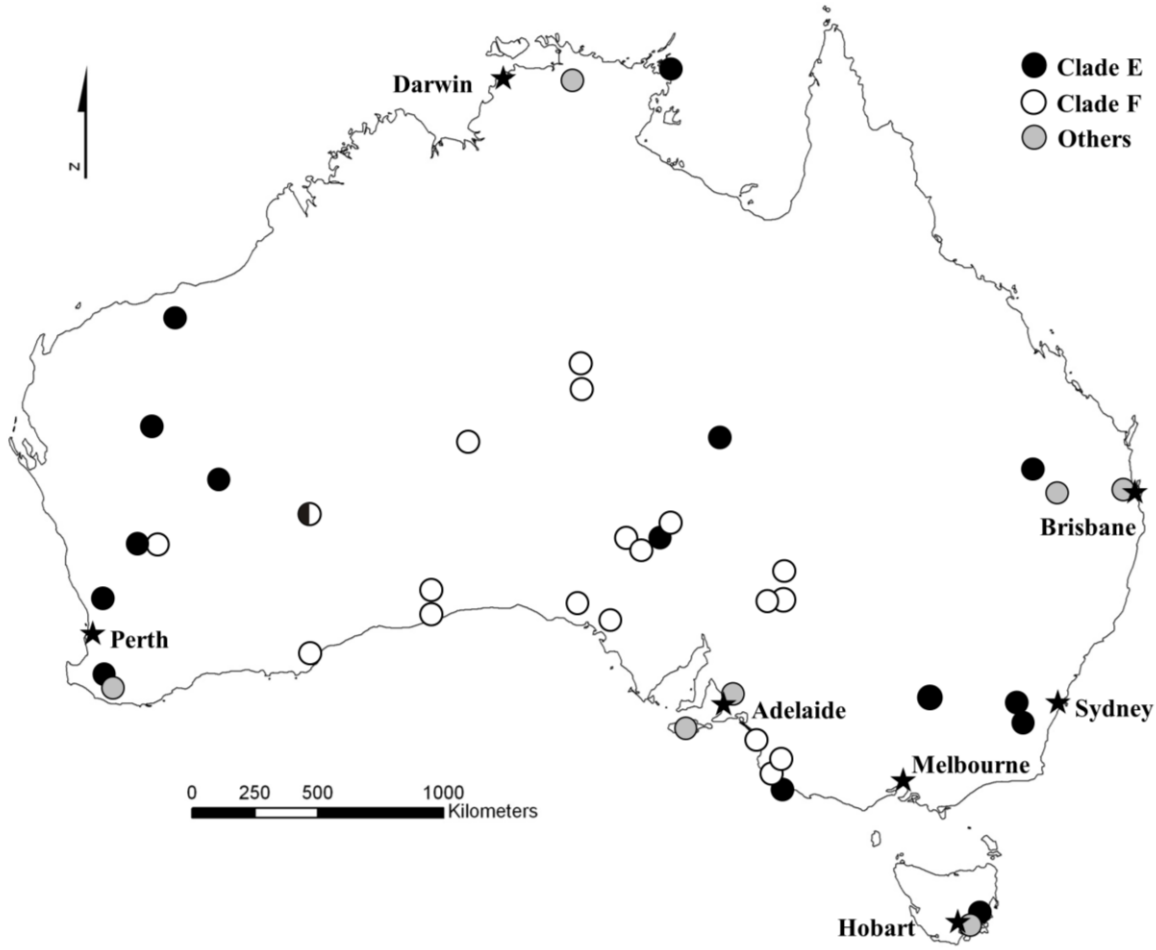


Figure 2. House mouse collection localities in mainland Australia, Kangaroo Island and Tasmania. Circles correspond to mtDNA haplotypes grouped by lineage. One locality was characterized by mice of both clade E and clade F.

Given that the predominant clades in Australia, E and F, are also found in the potential source areas of the Netherlands, Britain and Ireland, Table 1 compares nucleotide and haplotype diversity between these four regions plus New Zealand, as another colonized area in Australasia with a similar settlement history. The haplotype diversity is higher in the potential source areas, particularly on a clade-by-clade basis. The overall nucleotide diversity is not systematically higher in the source areas, but Australia in particular has much lower nucleotide diversities than the potential source areas within separate clades.

Discussion

Colonization of Australia by house mice

All Australian D-loop sequences examined belonged to the subspecies *Mus musculus domesticus*. This fits with the taxonomic status based on morphology (Schwarz and Schwarz 1943, Singleton and Redhead 1990) and lymphocyte antigens (Figueroa *et al.* 1986). This does not mean that other subspecies are not present. Some of our samples were deliberately chosen as *domesticus*, but in most cases the specimens we used were classified as '*Mus musculus*' with no attempt to define subspecies, yet all were characterized as *domesticus* on D-loop sequencing. Thus, from our wide-ranging studies, it appears that *M. m. domesticus* is the predominant subspecies of house mouse in Australia. This accords well with the hypothesis tested here that house mice arrived from within the range of the western European *domesticus* subspecies (Singleton and Redhead 1990), either on Dutch or British ships, and not from the closer potential source region (Southeast Asia) where a different subspecies is present, *Mus musculus castaneus* (Boursot *et al.* 1993).

Table 1. Genetic diversity indices applied to house mice with *Mus musculus domesticus* D-loop sequences by country and by prevalent mtDNA clade (E and F).

Country/ clade	*N*	Number of haplotypes	*h*	*π*	Reference
Britain					
Overall	89	31	0.932 ± 0.013	0.0087 ± 0.0045	
Clade E	37	9	0.757 ± 0.047	0.0033 ± 0.0020	[Jones *et al.* 2011a]
Clade F	45	18	0.900 ± 0.030	0.0033 ± 0.0020	
Ireland					
Overall	81	19	0.887 ± 0.017	0.0057 ± 0.0031	
Clade E	20	5	0.663 ± 0.069	0.0016 ± 0.0011	[Jones *et al.* 2011a]
Clade F	60	13	0.830 ± 0.027	0.0019 ± 0.0013	
Netherlands					
Overall	17	7	0.831 ± 0.065	0.0109 ± 0.0059	
Clade E	11	4	0.691 ± 0.128	0.0023 ± 0.0016	This study
Clade F	-	-	-	-	
Australia					
Overall	77	13	0.772 ± 0.031	0.0072 ± 0.0003	
Clade E	32	4	0.333 ± 0.100	0.0007 ± 0.0002	This study
Clade F	27	3	0.145 ± 0.090	0.0002 ± 0.0001	
New Zealand					
Overall	79	10	0.675 ± 0.049	0.0059 ± 0.0032	
Clade E	60	5	0.454 ± 0.060	0.0015 ± 0.0011	[Searle *et al.* 200b]
Clade F	6	2	0.333 ± 0.215	0.0004 ± 0.0005	

N – number of individuals analyzed
h – haplotype diversity
π – nucleotide diversity

Looking in fine detail at the D-loop clades and haplotypes present, the specific link with the British Isles becomes clearer. The two most widely distributed clades in Australia are also the two most widely distributed clades in the British Isles: clades E and F. The haplotype of clade E that has been found in most localities in Britain (nine

sites in southern Scotland, Wales and southern England (Searle *et al.* 2009a, Jones *et al.* 2011a) is also the most widespread and frequent haplotype of that clade in Australia (AUSTRALIA.01), but has only been detected in one out of 17 individuals from the Netherlands. Clade F is well documented around the northern and western periphery of the British Isles and the widespread clade F haplotype in Australia (AUSTRALIA.05) has been also found there (in a museum specimen dating from 1914 from the Isle of Lewis off the coast of Scotland (Jones *et al.* 2011a). On the other hand, clade F was not present among the Dutch mice we examined, nor has it been detected in the surrounding regions of continental Europe previously sampled in France, Germany and Denmark (Prager *et al*. 1993, 1996, Ihle *et al.* 2006, Jones *et al.* 2011a), except for three specimens out of 43 collected in northern France (taken from a single location in Abbeville (Jones *et al.* 2011a). The low frequency of AUSTRALIA.01 (= NETHERLANDS.02) and absence of clade F in our sample of contemporary Dutch mice does not fully exclude the Netherlands as a possible source area for the Australian house mouse. However, pending larger sample sizes and more detailed genetic data, a British Isles origin of Australian mice is the most reasonable interpretation of our results.

Therefore, our results agree with the historical routes of colonization, further suggesting that house mice were brought to Australia from at least two parts of the British Isles (northwestern Britain and/or Ireland [clade F] and somewhere in southern Britain [clade E]) probably early in the settlement of Australia. The early historical links between Britain and Australia date back to the claiming of Australia as part of the British Empire by James Cook in 1770. Although New Holland had already been discovered and named by the Dutch in 1606 and the west coast visited on a number of occasions, no trade or settlement was attempted at this time. The 11 ships

that constituted the British First Fleet arrived in January 1788 at Botany Bay, near the present site of Sydney, to establish the first European penal colony in Australia. During the following years two other convoys also comprising numerous ships carrying settlers and supplies arrived in the newly founded colony of New South Wales. Throughout these early stages of establishment, thousands of English, Scottish and Irish convicts were transported to Australia leaving mainly from southern England and Ireland, bringing provisions and livestock with them (Jupp 2002). Therefore, it seems reasonable to believe that the first mouse colonists, survivors of the months-long journey, arrived on this occasion, rapidly expanding their range throughout Australia, spreading the genetic signature of their geographic origin – clades E and F. Mice of other clades are restricted to coastal areas of Australia, near major harbors: Perth, Adelaide, Hobart, Brisbane and Darwin (Figure 2) and therefore most probably represent pockets of secondary colonization events. Thus, the genetic signature of the primary colonization appears to be retained in the extant mouse populations with a limited impact of subsequent introductions. This same phylogeographic pattern is found recurrently when house mouse populations from small isolated islands are surveyed (e.g. Förster *et al.* 2009, Searle *et al.* 2009b, Hardouin *et al.* 2010, Bonhomme *et al.* 2011, Jones *et al.* 2011b) but is particularly striking for a landmass as huge as Australia. Although Australia is nearly as large as Europe in area, the dynamics of arrival of house mice is still effectively like that of a small island, i.e. with a limited number of entry points and expansion from those.

The two most common haplotypes in Australia have also been found in New Zealand, with AUSTRALIA.01 being the most widespread and abundant haplotype there (scored as NZ.4) (Searle *et al.* 2009b). Given the shared history of activities involving British ships during settlement of the two countries this is not surprising (Ross 1987,

Searle *et al.* 2009b). It is interesting that genetic diversity within mitochondrial clades was higher for New Zealand than for Australia. That may be because New Zealand is a multitude of islands and the sampling reflected that (Searle *et al.* 2009b). In particular, among the small islands of the archipelago, the mice tended to differ in D-loop haplotype, increasing overall diversity, and with the inference that the different islands had different colonization histories (Searle *et al.* 2009b). The sampling of Australia was dominated by specimens from the mainland where haplotypes AUSTRALIA.01 and AUSTRALIA.05 managed to spread far and wide. The observed low haplotype diversity on the Australian landmass is either the reflection of low diversity among the founders or small numbers of them, or post-colonization population bottlenecking.

Overall then, the mitochondrial data presented here provide valuable information on the colonization history of Australia by the house mouse. There are strong indications that mitochondrial DNA is a particularly useful marker for first colonization (Searle *et al.* 2009a, Hardouin *et al.* 2010). However, the sequence that we used, the D-loop, is only a small part of the mitochondrial genome. More precision on the source and timing of initial colonization of Australia by house mice, associated population sizes, and the detail of secondary colonization events will be provided by studies involving larger numbers of individuals, complete mitochondrial genome sequencing, and substantive analysis of the nuclear genome (Begun *et al.* 2007, Gabriel *et al.* 2010, Jones *et al.* 2011a). Nuclear genome data has the potential to provide considerably more information than currently available, including the perspective of both male and female colonization (as the maternally inherited mitochondrial genome only reflects female colonization).

The house mouse as a model organism in invasion biology

Studies on species-invasion have provided invaluable insights into ecological and evolutionary processes, highlighting the utility of alien species as model organisms (Sax *et al.* 2007). In a mammalian context the house mouse is an iconic example of a successful invader (Searle 2008), capable of thriving in the most remote and inhospitable locations on the planet, particularly on islands. Given that house mice have colonized islands across all climatic and biogeographic regions, island invasions constitute valuable experiments to test for local adaptation to different biotic and abiotic factors. Australia, a continent with characteristics similar to an oceanic island in terms of its invasibility, constitutes one of the most dramatic house mouse invasion zones worldwide. Knowing that Australian house mice likely came from the British Isles allows comparison of populations from the source and colonized areas, with all the genomic tools available for this model species. This could result in a better understanding of characteristics of Australian house mice such as their distinctive predisposition to plagues (Singleton *et al.* 2005).

It has been claimed that through phylogeographic analysis house mice are good proxies for human movement patterns and settlement history over the last few thousand years because of their remarkable tendency to be transported wherever humans go (Searle 2008). The close match between post-European human settlement history and house mouse phylogeography in Australia supports this contention. It is noteworthy that house mice did not colonize from the closest geographical region, Southeast Asia, even though house mice (*castaneus* subspecies) have long been present there (Boursot *et al.* 1993). There is evidence of a long pre-European trade period between Australia, Indonesia and the Asian mainland, perhaps three to four centuries

prior to the arrival of the first Europeans (Nelson 1961). Apparently, despite this early nautical and commercial activity, propagule pressure of *castaneus* mice was not sufficient to allow them to invade and spread throughout Australia. It appears that only with the arrival and establishment of the British colony the invasion risk became significantly high, allowing the successful introduction of mice.

Acknowledgments

We are very grateful to Ken Aplin (Australia National Wildlife Collection), Michael Nachman, Kristin Ardlie, Michael Driessen, Belinda Bauer (Tasmanian Museum and Art Gallery), Claire Stevenson (Western Australian Museum), Ralph Foster (South Australian Museum) and the Australian Biological Tissue Collection (ABTC) at the South Australian Museum for samples of Australian mice. We also thank Adri Rol from the Zoological Museum of Amsterdam for access to that collection. We are grateful to the reviewers for their helpful comments on an earlier version of the manuscript.

References

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, Pachter L, Myers E, Langley CH (2007) Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology*, 5, e310.

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffra J-C, 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.

Boursot P, Auffray J-C, Britton-Davidian J, Bonhomme F (1993) The evolution of house mice. *Annual Review of Ecology and Systematics*, 24, 119–152.

Excoffier L, Laval G, Schneider S (2005) Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, 1, 47–50.

Figueroa F, Tichy H, McKenzie I, Hammerling U, Klein J (1986) Polymorphism of lymphocyte antigens-encoding loci in wild mice. *Current Topics in Microbiology and Immunology*, 127, 229–235.

Förster DW, Gündüz İ, Nunes AC, Gabriel S, Ramalhinho MG, Mathias ML, Britton-Davidian J, Searle JB (2009) Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology*, 18, 4477–4494.

Gabriel SI, Jóhannesdóttir F, Jones EP, Searle JB (2010) Colonization, mouse-style. *BMC Biology*, 8, 131.

Gündüz İ, Rambau RV, Tez C, Searle JB (2005) Mitochondrial DNA variation in the western house mouse (*Mus musculus domesticus*) close to its site of origin: studies in Turkey. *Biological Journal of the Linnean Society*, 84, 473–485.

Hardouin EA, Chapuis J-L, Stevens MI, van Vuuren BJ, Quillfeldt P, Scavetta RJ, Teschke M, Tautz D (2010) House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evolutionary Biology*, 10, 325.

Ihle S, Ravaoarimanana I, Thomas M, Tautz D (2006) An analysis of signatures of selective sweeps in natural populations of the house mouse. *Molecular Biology and Evolution*, 23, 790–797.

Jones EP, van der Kooij J, Solheim R, Searle JB (2010) Norwegian house mice (*Mus musculus musculus/domesticus*): Distributions, routes of colonization and patterns of hybridization. *Molecular Ecology*, 19, 5252–5264.

Jones EP, Jóhannesdóttir F, Gündüz İ, Richards MB, Searle JB (2011a) The expansion of the house mouse into north–western Europe. *Journal of Zoology*, 283, 257–268.

Jones EP, Jensen J-K, Magnussen E, Gregersen N, Hansen HS, Searle JB (2011b) A molecular characterization of the charismatic Faroe house mouse. *Biological Journal of the Linnean Society*, 102, 471–482.

Jupp J (2002) *From white Australia to Woomera: the story of Australian immigration*. Cambridge University Press, Cambridge.

Nachman MW, Boyer SN, Searle JB, Aquadro CF (1994) Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. *Genetics*, 136, 1105–1120.

Nelson JG (1961) Pre-European trade between Australia, Indonesia, and the Asiatic mainland. *The Canadian Geographer / Le Géographe canadien*, 5, 18–22.

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

Prager EM, Sage RD, Gyllensten U, Thomas WK, Hübner R, Jones CS, Noble L, Searle JB, Wilson AC (1993) Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biological Journal of the Linnean Society*, 50, 85–122.

Prager EM, Tichy H, Sage RD (1996) Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: comparison with other house mice and report of a 75-bp tandem repeat. *Genetics*, 143, 427–446.

Prager EM, Orrego C, Sage RD (1998) Genetic variation and phylogeography of Central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics*, 150, 835–861.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572–1574.

Ross JOC (1987) *William Stewart: sealing captain, trader and speculator*. Australia: Roebuck Society, Canberra.

Sax DF, Stachowicz JJ, Brown JH, Bruno JF, Dawson MN et al. (2007) Ecological and evolutionary insights from species invasions. *Trends in Ecology and Evolution*, 22, 465–471.

Schwarz E, Schwarz HK (1943) The wild and commensal stocks of the house mouse, Mus musculus Linnaeus. *Journal of Mammalogy*, 24, 59–72.

Searle JB (2008) The genetics of mammalian invasions: a review. *Wildlife Research*, 35, 185–192.

Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry S, Giménez MD, Jóhannesdóttir F (2009a) Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society, B* 276, 201–207.

Searle JB, Jamieson PM, Gündüz İ, Stevens MI, Jones EP, Gemmill CEC, King CM (2009b) The diverse origins of New Zealand house mice. *Proceedings of the Royal Society, B* 276, 209–217.

Singleton GR (1995) House mouse, *Mus musculus*. In: *Strahan R, ed. The mammals of Australia*. Australian Museum/Reed Books, Sydney, 646–647.

Singleton GR, Brown PR, Pech RP, Jacob J, Mutze GJ, Krebs CJ (2005) One hundred years of eruptions of house mice in Australia - A natural biological curio. *Biological Journal of the Linnean Society*, 84, 617–627.

Singleton GR, Redhead TD (1990) Structure and biology of house mouse populations that plague irregularly: an evolutionary perspective. *Biological Journal of the Linnean Society*, 41, 285–300.

Singleton GR, Tann CR, Krebs CJ (2007) Landscape ecology of house mouse outbreaks in Southeastern Australia. *Journal of Applied Ecology*, 44, 644–652.

Supplementary Material

Table S1. Details of all house mice obtained from Australia and the Netherlands subject to D-loop sequencing. Geographical coordinates are represented as decimal degrees. Most Australian samples were provided by Museum collections and the remaining samples belong to private collections of Michael Nachman, Kristin Ardlie [Ardlie KG, Silver LM (1998) Low frequency of t haplotypes in natural populations of house mice (*Mus musculus domesticus*). Evolution, *52*, 1185-1196] and Michael Driessen. Samples provided by Michael Driessen were collected during pest management work of the Resource Management and Conservation Division, Department of Primary Industries, and Water, Tasmania, following their standard ethical practice. All Dutch samples were provided by Adri Rol at the Zoological Museum of Amsterdam. 'Sample ID' corresponds to the original Museum Catalogue Number.

Sample ID	Sample provider	Latitude	Longitude	Haplotype	Clade
AUSTRALIA					
Western Australia					
36261	WAM	-31.2300	116.3017	AUSTRALIA.01	Clade E
49891	WAM	-31.2300	116.3017	AUSTRALIA.01	Clade E
44110	WAM	-28.2000	123.6000	AUSTRALIA.02	Clade E
44124	WAM	-28.2000	123.6000	AUSTRALIA.07	Clade F
56193	WAM	-26.9397	120.5581	AUSTRALIA.01	Clade E
48520	WAM	-25.0972	118.0083	AUSTRALIA.01	Clade E
49678	WAM	-29.2822	117.5742	AUSTRALIA.01	Clade E
49687	WAM	-29.2819	117.6428	AUSTRALIA.05	Clade F
56476	WAM	-34.2214	116.3741	AUSTRALIA.01	Clade E
56486	WAM	-34.0878	116.3194	AUSTRALIA.13	Clade B
ABTC07343	SAM	-30.8579	128.1011	AUSTRALIA.05	Clade F
ABTC07352	SAM	-30.8579	128.1011	AUSTRALIA.05	Clade F
ABTC07414	SAM	-33.4623	123.8421	AUSTRALIA.05	Clade F
ABTC07415	SAM	-33.4623	123.8421	AUSTRALIA.05	Clade F
ABTC07417	SAM	-33.4623	123.8421	AUSTRALIA.05	Clade F
ABTC07421	SAM	-31.8341	128.1776	AUSTRALIA.05	Clade F
ABTC63123	SAM	-21.3031	118.8611	AUSTRALIA.01	Clade E
Northern Territory					
ABTC41764	SAM	-25.5900	129.2858	AUSTRALIA.05	Clade F
ABTC24032	SAM	-23.7354	133.3307	AUSTRALIA.05	Clade F
ABTC24033	SAM	-22.7354	133.3307	AUSTRALIA.05	Clade F
ABTC28266	SAM	-12.1920	136.7742	AUSTRALIA.03	Clade E
M30774	ANWC	-12.7617	133.1033	AUSTRALIA.06	Clade F
South Australia					
ABTC37735	SAM	-37.0342	139.5117	AUSTRALIA.05	Clade F
ABTC37731	SAM	-37.1336	139.4804	AUSTRALIA.05	Clade F
ABTC37481	SAM	-37.0812	140.4142	AUSTRALIA.05	Clade F
ABTC37501	SAM	-37.5336	140.2228	AUSTRALIA.05	Clade F
ABTC37430	SAM	-38.0315	140.5635	AUSTRALIA.01	Clade E
ABTC26669	SAM	-32.0842	134.4352	AUSTRALIA.05	Clade F
ABTC26670	SAM	-32.0751	134.3943	AUSTRALIA.05	Clade F
ABTC26679	SAM	-31.4619	133.2820	AUSTRALIA.05	Clade F
ABTC26682	SAM	-31.4742	133.2545	AUSTRALIA.05	Clade F
ABTC35543	SAM	-29.3615	135.4335	AUSTRALIA.05	Clade F
ABTC35714	SAM	-29.1721	135.1100	AUSTRALIA.05	Clade F
ABTC35426	SAM	-29.0324	136.1556	AUSTRALIA.02	Clade E
ABTC36313	SAM	-28.5703	136.4655	AUSTRALIA.05	Clade F
ABTC36796	SAM	-31.2517	140.1515	AUSTRALIA.05	Clade F
ABTC36799	SAM	-31.2617	140.2807	AUSTRALIA.05	Clade F
ABTC36797	SAM	-31.2438	140.1035	AUSTRALIA.05	Clade F
ABTC36633	SAM	-30.2557	140.5826	AUSTRALIA.05	Clade F

Sample ID	Sample provider	Latitude	Longitude	Haplotype	Clade
South Australia (cont.)					
ABTC79657	SAM	-34.6004	138.7491	AUSTRALIA.11	Basal
ABTC33350	SAM	-35.8389	137.1679	AUSTRALIA.11	Basal
ABTC33353	SAM	-35.8389	137.1679	AUSTRALIA.11	Basal
ABTC33534	SAM	-35.8389	137.1679	AUSTRALIA.11	Basal
ABTC33610	SAM	-35.8389	137.1679	AUSTRALIA.11	Basal
ABTC33650	SAM	-35.8389	137.1679	AUSTRALIA.11	Basal
Queensland					
KA107	MN/KA	-27.5000	150.5000	AUSTRALIA.08	Clade D
KA112	MN/KA	-27.5000	150.5000	AUSTRALIA.08	Clade D
KA124aust	MN/KA	-27.5000	150.5000	AUSTRALIA.10	Clade D
KA1aust	MN/KA	-27.5000	150.5000	AUSTRALIA.08	Clade D
KA29aust	MN/KA	-27.5000	150.5000	AUSTRALIA.09	Clade D
KA35aust9205	MN/KA	-27.5000	150.5000	AUSTRALIA.08	Clade D
KA41aust	MN/KA	-27.5000	150.5000	AUSTRALIA.08	Clade D
KA9	MN/KA	-27.5000	150.5000	AUSTRALIA.08	Clade D
ABTC13890	SAM	-25.5023	138.4104	AUSTRALIA.02	Clade E
ABTC79716	SAM	-26.6422	149.6230	AUSTRALIA.02	Clade E
M30797	ANWC	-27.4900	152.8897	AUSTRALIA.08	Clade D
New South Wales					
KA35aust9204	MN/KA	-34.8058	145.8841	AUSTRALIA.04	Clade E
KA54	MN/KA	-34.8058	145.8841	AUSTRALIA.01	Clade E
M29998	ANWC	-35.5400	149.2100	AUSTRALIA.01	Clade E
Australia Capital Territory					
M30792	ANWC	-35.2192	149.1033	AUSTRALIA.01	Clade E
M30794	ANWC	-35.2192	149.1033	AUSTRALIA.01	Clade E
M30795	ANWC	-35.2192	149.1033	AUSTRALIA.01	Clade E
M30796	ANWC	-35.2192	149.1033	AUSTRALIA.01	Clade E
Tasmania					
A410	TMAG	-42.8817	147.3322	AUSTRALIA.01	Clade E
A504	TMAG	-42.9129	147.3545	AUSTRALIA.01	Clade E
A505	TMAG	-42.9129	147.3545	AUSTRALIA.01	Clade E
A1481a	TMAG	-42.8795	147.3275	AUSTRALIA.12	Basal
A1481b	TMAG	-42.8795	147.3275	AUSTRALIA.12	Basal
A1575	TMAG	-42.8731	147.3023	AUSTRALIA.01	Clade E
A3233	TMAG	-43.0624	147.8123	AUSTRALIA.01	Clade E
TAS1	MD	-42.9796	147.3262	AUSTRALIA.01	Clade E
TAS2	MD	-42.9796	147.3262	AUSTRALIA.01	Clade E
TAS3	MD	-42.9796	147.3262	AUSTRALIA.01	Clade E
TAS4	MD	-42.9796	147.3262	AUSTRALIA.01	Clade E
TAS5	MD	-42.9796	147.3262	AUSTRALIA.01	Clade E
TAS6	MD	-42.9796	147.3262	AUSTRALIA.01	Clade E
TAS7	MD	-42.9796	147.3262	AUSTRALIA.01	Clade E

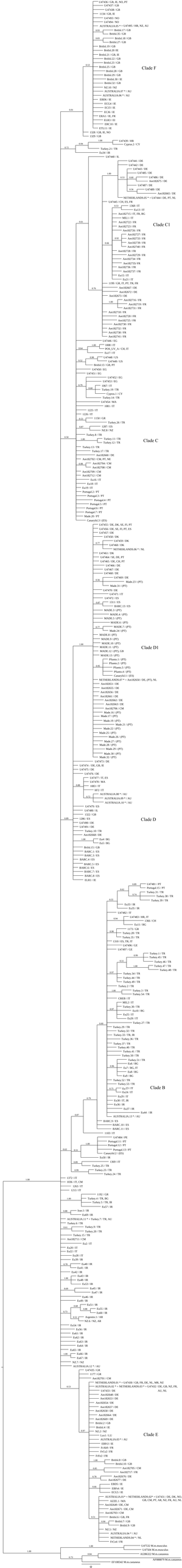
Sample ID	Sample provider	Latitude	Longitude	Haplotype	Clade
NETHERLANDS					
Rotterdam					
5902	ZMA	51.9226	4.4707	NETHERLANDS.04	Clade E
5897	ZMA	51.9226	4.4707	NETHERLANDS.05	Clade C
5898	ZMA	51.9226	4.4707	NETHERLANDS.04	Clade E
Oosthuizen					
8139	ZMA	52.5747	4.9985	NETHERLANDS.06	Clade D
Midsland noord					
27479	ZMA	53.3631	5.2110	NETHERLANDS.05	Clade C
Terschelling					
27476	ZMA	53.3631	5.2110	NETHERLANDS.01	Clade E
27477	ZMA	53.3631	5.2110	NETHERLANDS.01	Clade E
Montfort					
4674	ZMA	51.1173	5.9521	NETHERLANDS.01	Clade E
Maastricht					
4393	ZMA	50.8498	5.6873	NETHERLANDS.07	Clade D
Geulle					
4391	ZMA	50.9225	5.7484	NETHERLANDS.03	Clade E
Breda					
5369	ZMA	51.5893	4.7745	NETHERLANDS.01	Clade E
5368	ZMA	51.5893	4.7745	NETHERLANDS.01	Clade E
Groenekan					
5949	ZMA	52.1229	5.1519	NETHERLANDS.02	Clade E
Amsterdam					
21355	ZMA	52.3739	4.8941	NETHERLANDS.01	Clade E
4401	ZMA	52.3739	4.8941	NETHERLANDS.05	Clade C
4669	ZMA	52.3739	4.8941	NETHERLANDS.05	Clade C
11604	ZMA	52.3739	4.8941	NETHERLANDS.03	Clade E

Sample providers

WAM – Western Australia Museum
SAM – South Australian Museum
ANWC – Australia National Wildlife Collection
TMAG – Tasmanian Museum and Art Gallery
ZMA – Zoological Museum of Amsterdam
MN/KA – Michael Nachman / Kristin Ardlie
MD – Michael Driessen

Supplementary Material

Figure S1. Detailed phylogenetic tree of *Mus musculus domesticus*. This is the detailed 50% majority rule consensus tree after Bayesian inference that is summarized in Figure 1 in the paper (based on a total of 378 *Mus musculus domesticus* D-loop haplotypes, derived from this study and all previous data available at the time of analysis) (Prager *et al.* 1993, 1996, 1998; Nachman *et al.* 1994; Gündüz *et al.* 2000, 2001, 2005; Ihle *et al.* 2006; Rajabi-Maham *et al.* 2008; Förster *et al.* 2009; Searle *et al.* 2009a, b; Jones *et al.* 2011). Posterior probabilities of 0.50 and above are shown on branch nodes. All haplotypes detected in this study are highlighted with an asterisk in the tree (details of samples can be found in Table S1). Codes indicate GenBank number or code used in the publication where the sequence was first reported, followed by a list of all the countries (country codes) where the haplotype has been recorded: AR, Argentina; AU, Australia; BG, Bulgaria; CH, Switzerland; CM, Cameroon; CY, Cyprus; DE, Germany; DK, Denmark; EG, Egypt; ES, Spain; (ES), Canary Islands (Spanish dependency); FI, Finland; FR, France; GB, Great Britain; GE, Georgia; GR, Greece; HR, Croatia; IE, Ireland; IL, Israel; IR, Iran; IT, Italy; LU, Luxembourg; MA, Morocco; MR, Mauritania; NE, Niger; NO, Norway; NL, Netherlands; NZ, New Zealand; PE, Peru; PT, Portugal; (PT), Madeiran Archipelago (Portuguese dependency); SE, Sweden; TR, Turkey; US, United States.


Clade F
Clade C1
Clade C
Clade D1
Clade D
Clade B
Clade E

Chapter 4

Paper IV

History of a nearly global invasion: the worldwide colonisation of the western house mouse (*Mus musculus domesticus*)

Sofia I. Gabriel[1,2], Maria da Luz Mathias[2] and Jeremy B. Searle[1,3]

[1] Department of Biology, University of York, Wentworth Way, York YO10 5DD, UK

[2] CESAM–Centre for Environmental and Marine Studies, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, 1749–016 Lisbon, Portugal

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

Manuscript in preparation

Abstract

Global phylogeographic patterns of the western European house mouse, *Mus musculus domesticus*, were examined with mitochondrial DNA D-loop variation based on an unprecedented geographic coverage (760 new specimens). Samples from all over the known range of this subspecies – Western Europe, Africa, North, Central and South America, Australasia and numerous islands worldwide – were used to reconstruct the phylogenetic relationships of the haplotypes obtained. In all areas examined, the representation of house mouse clades matched well with the expectations from human history. The particular emphasis of this paper related to those parts of the globe that were 'discovered' and subsequently settled by western Europeans during the Age of Discovery onwards (from the beginning of the 1400s). The main mouse clades found in a particular geographic area almost always fitted well with either the country that discovered the area and/or the country that initially settled it. This analysis emphasises the value of the house mouse as a living artefact to aid historians.

Introduction

Along with humans, the house mouse is one of the most widespread and successful mammals living on earth today. It is generally accepted that the ancestors of what is collectively known as *Mus musculus* originated in the vicinity of the northern Indian subcontinent or the Iranian plateau about 0.5 million years ago (Boursot *et al.* 1993, Boursot *et al.* 1996, Din *et al.* 1996). From there, the species started to expand its range, undergoing divergence as it occupied non-overlapping areas, ultimately resulting in three well-defined subspecies, *M. m.*

musculus, *M. m. domesticus* and *M. m. castaneus*. Among these, the western European subspecies, *domesticus*, became the most widespread following a commensal association with humans that started about 10,000 years ago in the Fertile Crescent region, in the Middle East (Cucchi *et al.* 2005). This tight relationship with humans gave the *domesticus* subspecies a near global distribution, over all continents, with stable populations being established in Western Europe, North and South America, Africa, Australasia and many oceanic islands worldwide (Boursot *et al.* 1993). This extensive range was attained and facilitated by the involuntary transport of mice, mostly on ships (Pocock *et al.* 2005). The first direct evidence of house mice transportation as maritime stowaways was obtained on the shipwreck of a late Bronze Age vessel that was found just off the coast of Turkey; a mandible of a small mammal was recovered on the remains of the 'Uluburun' and identified as belonging to the house mouse (Cucchi 2008). This highlights the role of late Bronze Age/Iron Age shipping as a dispersal vector in the invasive process of house mouse expansion throughout the Mediterranean (Cucchi *et al.* 2005). Later on, the range expansion continued when humans started to master the challenges of ocean going navigation, with mice again able to occupy the ships - exploiting human and livestock food supplies. About a thousand years ago, the Vikings were the international traders of their time (Forte *et al.* 2005). Therefore, the Viking domain had the most profound impact in the establishment of new human settlements and trading routes in the northeast Atlantic, but also reached as far as Greenland and Newfoundland (westwards) and into the Mediterranean basin (eastwards) (Forte *et al.* 2005). Advances in the naval technology allowed the Vikings to build and use robust vessels that were able to endure long sea voyages as well as to carry livestock,

trade goods and other possessions - amenable conditions for the stowaway transport of house mice.

The early 1400s were the beginning of a new historical era known as the 'Age of Discovery'. The Western European colonial empires had their foundations with the development of Iberian technology in nautical sciences, cartography and ship development, coupled with the impulse of economically profitable trade (Greene and Morgan 2008). Portugal and Spain were the leading countries to establish the first global trade network, followed mainly by the French, the British and the Dutch. By extending the 'known' world as perceived by Europeans, these nations made not only Europe, but also Africa, Asia and the Americas, part of a global system of exchange (Russell-Wood 1992). This unprecedented worldwide expansion involved the use of big convoys of vessels with the ability to travel long distances in open sea, for long periods, carrying not only substantial numbers of people, but also livestock and other goods or merchandises. As a consequence, house mice were accidentally transported around the world and left wherever humans travelled to, leaving a living artefact relating to the place of origin of the journey (Gabriel *et al.* 2010).

The techniques of 'phylogeography' can be used to infer the colonisation source of a population that moves elsewhere (Avise 2000), whether those movements are natural or through human transport. Over the past two decades, plentiful studies have been assessing the phylogeographic patterns of the western house mouse, *M. m. domesticus*, which inform about those human-mediated movements to which the animals are subjected (Prager *et al.* 1993, 1996, 1998; Nachman *et al.* 1994; Gündüz *et al.* 2000, 2001, 2005; Ihle *et al.* 2006; Rajabi-Maham *et al.* 2008; Förster *et al.* 2009; Searle *et al.* 2009a, b; Jones *et al.* 2010,

2011a,b, 2012; Bonhomme *et al.* 2011; Hardouin *et al.* 2010; Gabriel *et al.* 2011, Gabriel *et al.* submitted). These studies have used mitochondrial D-loop sequences and there is growing evidence that this is a particularly good molecular marker to trace initial waves of colonisation, retaining the original signature of the ancestral house mouse founder colonists that successfully formed viable populations (Gabriel *et al.* 2010, Bonhomme and Searle in press). A recent study making use of both modern and 1000 years old house mouse samples from Iceland concluded that the exact same haplotypes survived on the island during the last millennium (Jones *et al.* 2012).

Despite the excellent coverage obtained so far in most of western Europe, there has been quite a significant underrepresentation of studies addressing the phylogeography of *M. m. domesticus* outside this range, with a few exceptions (e.g. Australasia: Searle *et al.* 2009b, Gabriel *et al.* 2011). In particular, little is known about North and South America, and most of sub-Saharan Africa. There has been some progress in our understanding of the colonisation of oceanic islands and archipelagos by house mice (Gündüz *et al.* 2001, Förster *et al.* 2009, Searle *et al.* 2009b, Bonhomme *et al.* 2011, Hardouin *et al.* 2010, Jones *et al.* 2011b, 2012, Gabriel *et al.* submitted), but again more work needs to be done.

The house mouse provides a unique opportunity not only to assess the colonisation history of a nearly global distributed mammal (and a major pest in some regions) but also to make inferences about the colonisation and trading history of the humans that transported them worldwide as stowaways. The contribution that house mice can make as a living artefact to support the work of historians is considerable. For example, during the Age of Discovery there was much secrecy surrounding the finding of new islands (Diffie and Winius 1977). This

creates considerable uncertainty as to who first discovered any particular island. The house mice on an island are likely to represent descendants of the mice taken by the first people to get to there, and therefore will likely provide a mitochondrial DNA signal that will relate to the source area of both the mice and the people.

In this study, hundreds of new tissue and DNA samples of *Mus musculus domesticus* were obtained and the resulting D-loop sequences added to previously published haplotypes. These samples came from throughout the range of the western house mouse, particularly from non-European regions that were under-represented or not characterised at all. With the widest sampling ever achieved for this subspecies, and building on what is already known from previous studies, this work will create a global picture of the colonisation of the western house mouse, through mitochondrial DNA variation. In this way we aim to construct a framework for the use of the western house mouse as a proxy for human history, elucidating discoveries, colonisations and settlements of islands and landmasses around the world.

Materials and Methods

Sampling

A total of 760 house mouse (*Mus musculus*) samples were obtained, with samples defined as tissue samples, DNA samples, or unpublished D-loop sequences. The vast majority of the samples were from within the range of the western house mouse (*Mus musculus domesticus*) and covered all 5 continents including 47 countries (or countries dependencies), 22 of which were sampled for the first time for this species (highlighted with an asterisk) – Argentina, Bolivia*,

Brazil*, Chile*, Cameroon, Cape Verde*, Cyprus, Germany, Ecuador*, Egypt, Spain, Canary Islands (Spanish dependency), France, the French dependencies French Guiana*, Guadalupe* and Martinique*, Great Britain, Greece, Honduras*, Croatia, India, Israel, Iran, Italy, Jordan*, Libya, Morocco, Madagascar, Malta*, Mauritius (Rodrigues Island)*, Mexico*, New Caledonia*, Peru, Portugal, Paraguay*, Senegal, Seychelles*, South Africa*, Syria, Tristan da Cunha*, Tunisia, Turkey, Tanzania*, United States of America, Vanuatu*, Yemen (Socotra Island)* and Zimbabwe*.

DNA extraction, amplification and sequencing

Genomic DNA was extracted from liver samples, tail tips or toe clippings, using Qiagen DNeasy Blood & Tissue Extraction Kit, following the manufacturer's guidelines. Dried skin samples retrieved from museum collections were soaked in water overnight at 37ºC and extracted with the same kit. Rigorous contamination controls were used for the museum-curated specimens, following Martínková and Searle (2006).

The mtDNA D-loop and flanking tRNAs were amplified by PCR using the primer pair L15774 (5´- TGA ATT GGA GGA CAA CCA GT -3') and H2228 (5'- TTA TAA GGC CAG GAC CAA AC -3') as described by Searle *et al.* (2009a). The resulting sequences were shortened to positions 15424–16276 of Bibb *et al.* (1981) to allow the alignment with the majority of previously published house mouse haplotypes.

The DNA recovered from most of the museum samples was degraded and failed to amplify with the primer pair L15774/ H2228. To accommodate this, three previously published primer pairs (Gabriel *et al.* 2011) were used to obtain three overlapping fragments: dloopF1 – 5´ GCA CCC AAA GCT GGT ATT CT 3´ / dloopR1 –

5´ TTG TTG GTT TCA CGG AGG AT 3´; dloopF2 – 5´ ACT ATC CCC TTC CCC ATT TG 3´ / dloopR2 – 5’ GAT TGG GTT TTG CGG ACT AA 3´; dloopF3 – 5´ ATA GCC GTC AAG GCA TGA AA 3´ / dloopR3 – 5´ GCT TTG CTT TGT TAT TAA GCT ACA 3´. Samples that failed to amplify at least one of the three fragments were excluded from the subsequent analysis. PCR products resulting from successful amplifications were purified with the Qiagen QIAquick PCR Purification Kit and sequenced commercially in both directions using the PCR primers (Macrogen Inc., Korea).

Sequence analysis

All DNA sequence traces were checked in Sequencher v. 4.5 (Gene Codes Corp.) for ambiguities and aligned by eye on Bioedit v7.1.3.0 along with 625 previously published *M. m. domesticus* haplotypes retrieved from GenBank or other publications (Bonhomme *et al.* 2011). Two *M. m. castaneus* (AJ286322 and AF088879) and two *M. m. musculus* sequences (U47532 and U47504) were used as outgroups (following Jones *et al.* 2011a).

Unique D-loop haplotypes were obtained on DnaSP ver. 5.10.01 and deposited in GenBank under the accession numbers xxxxx to xxxxx.

In total, 797 unique haplotypes were used to conduct a phylogenetic analysis using Bayesian inference, as implemented in MrBayes 3.2 (Ronquist and Huelsenbeck 2003). The selection of the nucleotide substitution model SYM+I+Γ, with gamma distribution and proportion of invariable sites, was based on the results obtained from jModelTest 0.1.1 (Posada 2008) using the Akaike information criterion, AIC (Posada & Buckley 2004). MrBayes was run for 30

million generations with two parallel sets of five chains with an incremental heating parameter of 0.02. The convergence diagnosis and assessment of burn-in length was estimated by visual inspection of the time-series plot and through the use of PSRF statistics (Gelman and Rubin 1992), making sure that all parameters approached 1.000. The burn-in was set to 40% and the resulting tree 50% majority rule consensus tree was visualised and edited in FigTree 1.3.1.

Results and Discussion

A total of 760 individual D-loop complete sequences were obtained, 741 of which were assigned to the western European house mouse, *Mus musculus domesticus*. The remaining 19 sequences were either allocated to *M. m. musculus* (found in Germany and South Africa), *M. m. castaneus* (found in the east coast of USA, Peru, India, Zimbabwe and the Seychelles) and *M. m. gentilulus* (detected in Madagascar). The 741 *domesticus* sequences generated 211 haplotypes, 161 of which had never been recorded before.

The full phylogenetic tree that we obtained is given in Figure S1, Supporting Material. Despite the repeatedly reported poor lineage support within the *domesticus* branch of the *M. musculus* phylogenetic tree (Jones *et al.* 2010, 2011a), the Bayesian analysis returned the same structure with six major clades (A – F) as previously obtained by Jones *et al.* (2011a). In the present analysis, support for the clades ranged from 0.53 – 1.00. A smaller additional clade arose from the analysis (posterior probability = 0.82), mostly with haplotypes occurring in the Near East, from Syria, Lebanon, Tunisia and Cyprus. This has not been labelled, but the clade has close affinity to the lineage Hg 6 proposed by Bonhomme *et al.* (2011). With

further sequences and analyses, it will be interesting to see if this is a consistent clade that should be given its own designation as clade 'G'. Clearly, like clade A, this lineage has affinity with the populations that represent the earliest expansion of *M. m. domesticus* in connection with early human settlements and agriculture (Jones *et al.* 2011a).

As well as haplotypes grouped into the above-mentioned clades, a considerable number of ungrouped haplotypes were located in a basal position on the tree. These include haplotypes that in previous analyses had been grouped into one of the clades. We have found previously that adding new sequences to the phylogeny may change affinities and, for instance, clade A has been lost in earlier analyses (Jones *et al.* 2011b, Gabriel *et al.* 2011, Gabriel *et al.* submitted). As an example of the sequences involved this time, some of those from Portugal and Spain in clade B in our earlier analysis (Gabriel *et al.* submitted) have now become basal (e.g. Portugal. 11-13, SPAIN.04, SPAIN.06, SPAIN.08, SPAIN.10,).

Figure 1 affiliates geographical areas to the phylogenetic tree. The name of a nation or island group is placed against a lineage if that clade represents more than 20% of the sequences obtained for that geographical region. Figures 1a and 1b show the haplotype affinities of western European countries that have been colonisation sources for other parts of the world. Figures 1c – j represent the rest of the distribution of *M. m. domesticus*, including continental areas (Figure 1e – i) and well-sampled island systems (Figure 1c, d, j). The geographic areas highlighted in Figures 1c – j would largely have been colonised from western Europe. The only exception is North Africa and the Arabian peninsula (Figure 1e), which is shown here as a contrast with the rest of the African continent (Figure 1f). This figure also provides an interesting comparison with all other geographical areas, because it

represents not only the Near Eastern source area of the *M. m. domesticus* expansion but also the Iron Age Mediterranean colonisation (about 3000 years ago), which was much earlier than the other colonisations examined (Cucchi *et al.* 2005, Bonhomme *et al.* 2011).

Figure 1a shows that only 3 lineages are well-represented in the continental Nordic area. The distribution of clades in this area has already been discussed in earlier papers (Jones *et al.* 2010, 2011a). Clade E has been associated with Iron Age (2000 – 3000 years ago) colonisation of southern Britain and nearby areas of continental Europe (Jones *et al.* 2011a), and the presence of this clade in Denmark may be viewed as a geographic continuation of that colonisation, while its presence in Norway may represent activity of Norwegian Vikings (ca. 1000 years ago), inadvertently bringing mice back from the British Isles (Jones *et al.* 2010). Clade D is one that appears to have colonised through central Europe in the Iron Age, and clade F appears to have arrived via a maritime route (Jones *et al.* 2011a).

Figure 1b shows the clades that are associated with countries that were important in colonial expansion around the world from the Age of Discovery (starting 600 years ago) onwards. The clades in the British Isles, E and F, have already been mentioned as associated with Iron Age and Norwegian Viking activity, respectively (Searle *et al.* 2009a, Jones *et al.* 2011a). The occurrence of clade E in northern France and the Netherlands apparently represents part of those same Iron Age movements (Jones *et al.* 2011a, Gabriel *et al.* 2011). Other sequence types (basal, clade C) more associated with Mediterranean Europe are found further south in France. As described in Gabriel *et al.* (submitted), Portugal and Spain have good representation of Mediterranean clades B and C (sometimes lost from those clades as basal sequences: see above), and there is a minor

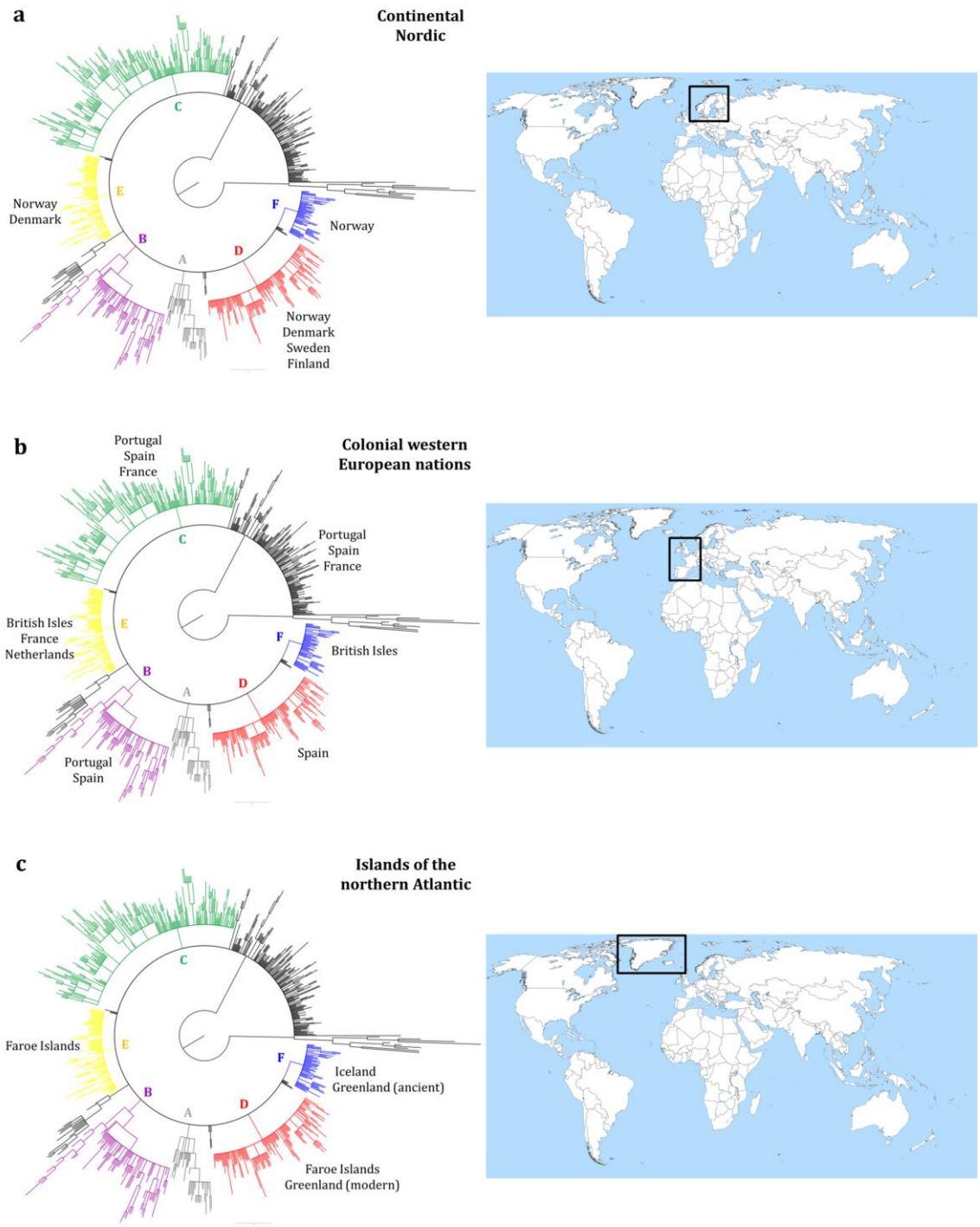
representation of clade D in Spain. Both clades B and C are likely to have arrived in Iberia during the Iron Age (Förster *et al.* 2009, Jones *et al.* 2011a).

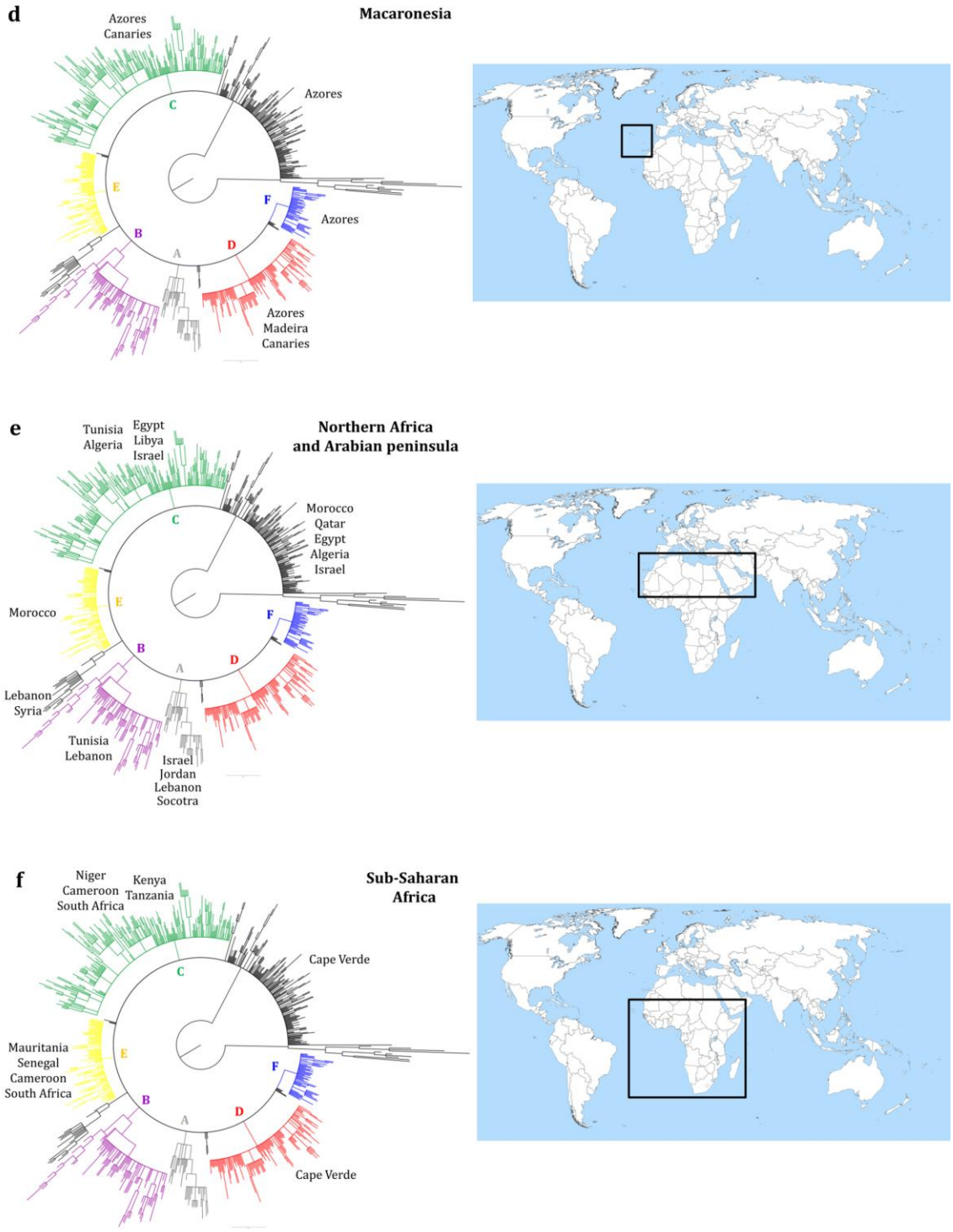
Having established the clade associations of the main expected source countries for *M. m. domesticus* colonisations around the globe, we can see if the expected clade associations are seen for the geographical areas that were discovered, subdued and settled by people from western Europe. The structuring of D-loop sequences in western Europe fits astonishingly well with historical events over the period 3000 – 1000 years ago; the distribution of clades around the rest of the world are largely expected to relate to events that are more recent.

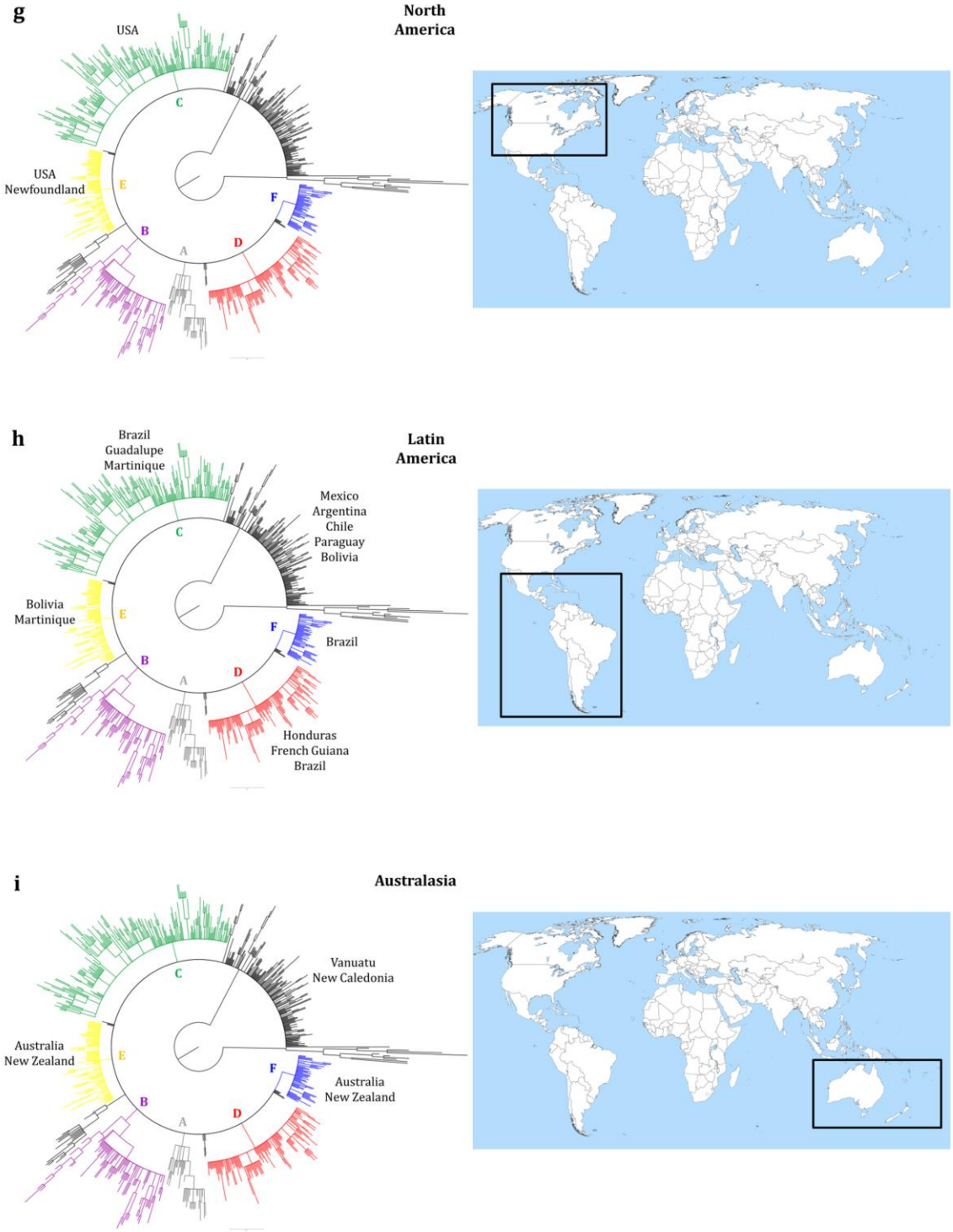
The interpretation of the data on the islands of the North Atlantic (Figure 1c) has already been made by Jones *et al.* (2011b, 2012) relating primarily to the continental Nordic area (Figure 1a) and also the British Isles (Figure 1b) during the Viking Age. Iceland and Greenland were colonised by Norwegian Vikings, and the house mice reflect that by presence of the clade primarily associated with Norwegian Viking activity – clade F. The Faroe Islands were also colonised by Norwegian Vikings, but from an area of southwestern Norway where clade D is prevalent, and that is the clade represented on the islands (Jones *et al.* 2011b), together with clade E, which as mentioned earlier for parts of Norway, may also have been transported from the British Isles by Norwegian Vikings. Modern Greenland was largely settled in the 1700s onwards from Denmark - where clade D is prevalent - explaining that association as well (Jones *et al.* 2012).

For the Macaronesian archipelagos (Figure 1d), we here follow the account of Gabriel *et al.* (submitted). The Azores and Canaries show the expected presence of the Iberian clade C and basal sequences (or clade B as obtained by Gabriel *et al.* submitted) (see Figure 1b). It has been well-argued in various papers that the

occurrence of clade D in Madeira relates to a visitation by Danish Vikings (Gündüz *et al.* 2001, Förster *et al.* 2009), and Gabriel *et al.* (submitted) further suggest that the distribution of this lineage elsewhere in Macaronesia and also Spain (Figure 1b) may relate to spread of house mice from the Madeiran archipelago.




d
Macaronesia
Azores
Canaries
Azores
Azores
Azores
Madeira
Canaries
C
E
B
A
D
F
e
Northern Africa
and Arabian peninsula
Tunisia
Algeria
Egypt
Libya
Israel
Morocco
Qatar
Egypt
Algeria
Israel
Morocco
Lebanon
Syria
Tunisia
Lebanon
Israel
Jordan
Lebanon
Socotra
f
Sub-Saharan
Africa
Niger
Cameroon
South Africa
Kenya
Tanzania
Cape Verde
Mauritania
Senegal
Cameroon
South Africa
Cape Verde


g
North America
USA
USA
Newfoundland
C
E
B
A
D
F
h
Latin America
Brazil
Guadalupe
Martinique
Mexico
Argentina
Chile
Paraguay
Bolivia
Bolivia
Martinique
Brazil
Honduras
French Guiana
Brazil
i
Australasia
Vanuatu
New Caledonia
Australia
New Zealand
Australia
New Zealand

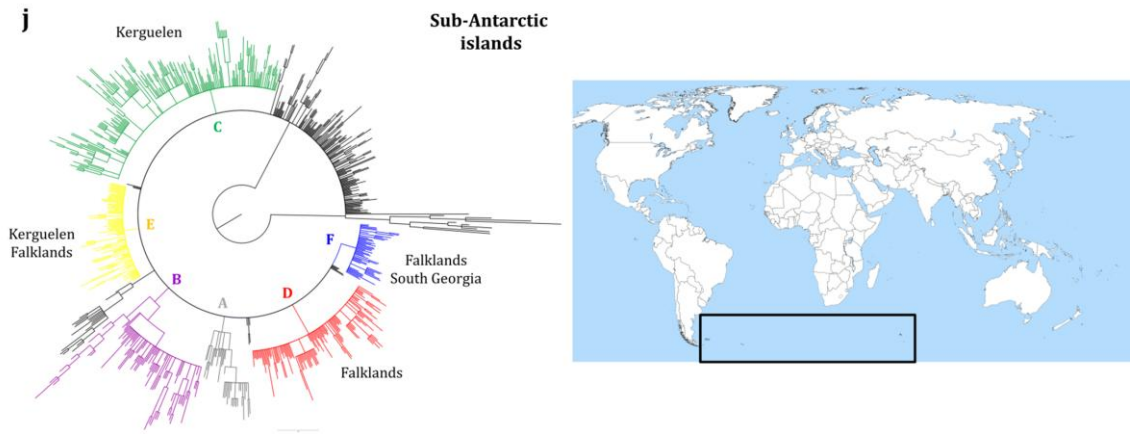


Figure 1. Phylogenetic tree of D-loop sequences of *Mus musculus domesticus* after Bayesian inference. Clades as defined by Jones *et al.* (2011a) are coloured as follows: outgroup and basal sequences: black, clade A: grey, clade B: purple, clade C: green, clade D: red, clade E: yellow, clade F: blue. The major lineages present in each potential colonising source (**a**, **b**) and each colonised area (**c** - **j**) are highlighted, where a major lineage is defined as having a frequency greater than 20%. **a**: continental Nordic countries, **b**: colonial western European nations, **c**: islands of the northern Atlantic, **d**: Macaronesia, **e**: North Africa and the Arabian peninsula, **f**: sub-Saharan Africa, **g**: North America (north of Mexico), **h**: Latin America, **i**: Australasia, **j**: sub-Antarctic islands.

The distribution of house mouse clades in North Africa and the Arabian Peninsula (as far north as Syria; Figure 1e) has been documented in particular detail by Bonhomme *et al.* (2011), but not using the same nomenclature as here. Clade A and the new minor clade are well-documented in the Arabian peninsula, suggesting that these (together with basal sequences) represent the first expansion of the western house mouse associated with initial commensalism, about 12,000 years ago (Cucchi *et al.* 2005, Jones *et al.* 2011a). Clades B, C and basal sequences are found in North Africa, just as they are in the northern Mediterranean (Bonhomme *et al.* 2011, Jones *et al.* 2011a). The presence of clade E

in Morocco may relate to its presence at the Atlantic margin, and therefore a stopping off point for boats from France (Figure 1a), as the country used to be a French protectorate.

Britain and France have been enormously influential in the settling of sub-Sarahan Africa (Figure 1f) and interestingly the sequences represented there (clades C and E) are those associated with southern Britain and France (Figure 1a; Jones *et al.* 2011a). Also the Netherlands and nearby parts of Germany are mostly clade E (Figure 1a; Gabriel *et al.* 2011, Jones *et al.* 2011a), and Portugal is predominantly clade C (Figure 1a). All these countries were involved in the exploration, exploitation and settling of Africa by Europeans. More detailed genomic studies are needed to establish the precise source of house mice to the different countries. Nevertheless, some associations can still be made, e.g. the presence of basal and clade D sequences in the Cape Verde archipelago, which has a long historical background with Portugal. Clade D is very common on the remaining Macaronesian archipelagos, particularly in Madeira and the Canaries (Figure 1d; Förster *et al.* 2009, Bonhomme *et al.* 2011), reinforcing the influence of these islands as potential source areas for subsequent colonisations of other geographic regions, during the Age of Discovery (Gabriel *et al.* submitted). The occurrence of clade C in Kenya and Tanzania may first seem surprising but early in the European history of these countries there was a major Portuguese influence along the coast of East Africa, following the discovery of the maritime route to India by Vasco da Gama in 1498 (Napoli 2010). It is therefore very likely that the presence of clade C relates to that. The later history of Kenya and Tanzania has been associated particularly with Germany and Britain (countries which do not have clade C). Further up along the coast of East Africa, mice on the island of

Socotra were *M. m. domesticus*, grouping within clade A (Figure 1e). Apparently, mice on this island did not originate on its mainland dependency (Yemen) as a different subspecies is found there, *M. m. gentilulus* (Prager *et al.* 1998). In South Africa it appears that there may be both an early Portuguese signal (clade C) and a later British/Dutch/German signal (clade E and *M. m. musculus*).

For North America, there are mouse D-loop data relating to Newfoundland and the USA (Figure 1g). Newfoundland was sampled because of the settlement by Norwegian Vikings (Jones *et al.* 2012). However, in Newfoundland and all 10 states of the USA sampled (California, Arizona, Iowa, Michigan, Indiana, Tennessee, New Jersey, Delaware, Maryland and New Hampshire), clade E was dominant. This fits well with a history that involved Britain, France and the Netherlands, all countries where clade E is common or predominant (Figure 1b). Clade C was also found at lower frequency in the USA, likely coming from either France or Spain.

A different picture is found in Latin America (Figure 1h), whose history was mostly tied to Spain and Portugal. The majority of sequences found in Iberia are found in clades B, C, D and the basal sequences (Figure 1b). In Latin America, the basal sequences are particularly well-represented. Brazil, which has had strongest historical links with Portugal, is characterised by clades C, D and F and neighbouring French Guiana had haplotypes grouping within a small sub-clade within clade D. Interestingly, once again, these are the clades present in Macaronesia (Figure 1d), and so the mouse genetic signal in Brazil and French Guiana may more closely represent the Atlantic stop-off points rather than the original source of the ships going there. The association of clades C and E with Guadalupe and Martinique is not surprising, given the close linkages between those islands and France (in fact, they are still French dependencies). The

occurrence of clade E in Bolivia is interesting, because this clade is associated with northern Europe and not Iberia (Figure 1b). The sample size is small (n=5) and we may have sampled a small set of mice deriving from the USA, with whom Bolivia has long had considerable trade (as with everywhere else in Latin America).

Previous studies (Searle *et al.* 2009b, Gabriel *et al.* 2011) have documented the occurrence of clades E and F in Australasia (Figure 1i). Both Australia and New Zealand have a settlement history closely associated with the British Isles where these two clades are also predominant (Figure 1b). Vanuatu and New Caledonia are located on the Pacific Ocean side of Australia and from the D-loop sequences it is not evident that the *M. m. domesticus* colonised via Australia. The mice of Vanuatu and New Caledonia had basal haplotypes. Gabriel *et al.* (2011) have also previously found two basal haplotypes in other islands closer to Australia: Kangaroo Island and Tasmania, which may or may not be relevant. The early European history of Vanuatu relates to Portuguese explorers in command of Spanish vessels (Richardson 2006), which may explain the occurrence of basal haplotypes (which are also present in Portugal and Spain: Figure 1b), although they could also have arrived with the later French association. The British Captain James Cook was the first to name New Caledonia in 1774, but the official history of these Pacific islands is more closely associated with France among the western European nations (Angleviel and Levine 2009) and again the basal haplotypes may have come from France (Figure 1b).

There is a reasonably close match between the clades found on the sub-Antarctic islands (Figure 1j) and those in the western European colonising countries (Figure 1b). Kerguelen is a French colony and displays clades C and E, while the Falklands and South Georgia are British and display clades E and F.

However, clade D is also present in the Falklands and the precise clade F haplotypes found in the Falklands actually match those in Santa Maria and Terceira in the Azores. The clade D and F haplotypes may, therefore, have got to the Falklands with the Portuguese during their voyages to South America rather than the British and the discovery of the islands may have been directly by the Portuguese (Gabriel *et al.* submitted). Alternatively, the colonisation of the Falklands may have been in fact by the British but the mice that first colonised the islands could have possibly been picked up from the Azores (from Santa Maria or Terceira, clade F) and Madeira or the Canaries (clade D) during a stop-over to replenish fresh food, water or even livestock for the remaining journey. Thus, the future analysis of distributions of individual haplotypes around the world may provide a much more nuanced picture of the colonisation history of mice linked to human history.

In conclusion, the colonisation history of the western house mouse *M. m. domesticus* matches extraordinarily well with human history throughout its range. This has already been demonstrated within the Mediterranean area and mainland western Europe (Bonhomme *et al.* 2011, Jones *et al.* 2011a). Here we have looked at regions of the world that western Europe mariners 'discovered', leading to subsequent European settlement. So, we have documented the mouse colonisation in these areas and have revealed, with reasonable confidence, that Norway, Denmark, Britain, France, Germany, Portugal and Spain have been sources of colonisation of those mice. The genetic signal of the mice has matched well either what is expected on the basis of initial discovery or later settlement. The contention that D-loop sequences represent well first colonisation by house mice in a particular area (Gabriel *et al.* 2010, Bonhomme and Searle in press) has been

upheld. House mice will provide historians with valuable data relating to Viking settlement and the Age of Discovery, possibly helping to elucidate some unconfirmed and speculative historical scenarios.

References

Angleviel F, Levine S (2009) *New Zealand – New Caledonia: Neighbours, Friends, Partners*. Victoria University Press, Wellington, New Zealand.

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

Boursot P, Auffray J-C, Britton-Davidian J, Bonhomme F (1993) The evolution of house mice. *Annual Review of Ecology and Systematics*, 24, 119–152.

Boursot P, Din W, Anand R, Darviche D, Dod B, von Deimling F, Talwar GP, Bonhomme F (1996) Origin and radiation of the house mouse: mitochondrial DNA phylogeny. *Journal of Evolutionary Biology*, 9, 391–415.

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell*, 26, 167–180.

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffra J-C, 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.

Cucchi T (2008) Uluburun shipwreck stowaway house mouse: molar shape analysis and indirect clues about the vessel's last journey. *Journal of Archaeological Science*, 35, 2953–2959.

Cucchi T, Vigne J-D, Auffray J-C (2005) First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarcheaological revision of subfossil occurrences. *Biological Journal of the Linnean Society*, 84, 429–445.

Diffie BW, Winius GD (1977) *Foundations of the Portuguese Empire, 1415–1580*. University of Minnesota Press, Minneapolis, USA.

Din W, Anand R, Boursot P, Darviche D, Dod B, Jouvin-Marche E, Orth A, Talwar GP, Cazanave P-A, Bonhomme F (1996) Origin and radiation of the house mouse: clues from nuclear genes. *Journal of Evolutionary Biology*, 9, 519–539.

Förster DW, Gündüz İ, Nunes AC, Gabriel S, Ramalhinho MG, Mathias ML, Britton-Davidian J, Searle JB (2009) Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology*, 18, 4477–4494.

Forte A, Oram R, Pedersen F (2005) *Viking Empires*. Cambridge University Press, Cambridge, UK.

Gabriel SI, Jóhannesdóttir F, Jones EP, Searle JB (2010) Colonization, mouse-style. *BMC Biology*, 8,131.

Gabriel SI, Mathias ML, Searle JB. submitted. Of mice and the 'Age of Discovery': the complex colonization history of the Azorean archipelago by the house mouse (*Mus musculus*) as revealed by mitochondrial DNA variation. *Journal of Evolutionary Biology*.

Gabriel SI, Stevens MI, Mathias ML, Searle JB (2011) Of mice and 'convicts': origin of the Australian house mouse, *Mus musculus*. *PLoS ONE*, 6, e28622.

Gelman A, Rubin D (1992) Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-511.

Greene JP, Morgan PD (2008) *Atlantic History: A Critical Appraisal*. Oxford University Press, Oxford, UK.

Gündüz İ, Auffray J-C, Britton-Davidian J, Catalan J, Ganem G, Ramalhinho MG, Mathias ML, Searle JB (2001) Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology*, 10, 2023–2029.

Gündüz İ, Rambau RV, Tez C, Searle JB (2005) Mitochondrial DNA variation in the western house mouse (*Mus musculus domesticus*) close to its site of origin: studies in Turkey. *Biological Journal of the Linnean Society*, 84, 473–485.

Gündüz İ, Tez C, Malikov V, Vaziri A, Polyakov A, Searle JB (2000) Mitochondrial DNA and chromosomal studies of wild mice (*Mus*) from Turkey and Iran. *Heredity*, 84, 458–467.

Hardouin EA, Chapuis J-L, Stevens MI, van Vuuren BJ, Quillfeldt P, Scavetta RJ, Teschke M, Tautz D (2010) House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evolutionary Biology*, 10, 325.

Ihle S, Ravaoarimanana I, Thomas M, Tautz D (2006) An analysis of signatures of selective sweeps in natural populations of the house mouse. *Molecular Biology and Evolution*, 23, 790–797.

Jones EP, Jóhannesdóttir F, Gündüz İ, Richards MB, Searle JB (2011a) The expansion of the house mouse into north–western Europe. *Journal of Zoology*, 283, 257–268.

Jones EP, Jensen J-K, Magnussen E, Gregersen N, Hansen HS, Searle JB (2011b) A molecular characterization of the charismatic Faroe house mouse. *Biological Journal of the Linnean Society*, 102, 471–482.

Jones EP, Skirnisson K, McGovern TH, Gilbert MTP, Willerslev E, Searle JB (2012) Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *BMC Evolutionary Biology*, 12, 35.

Jones EP, van der Kooij J, Solheim R, Searle JB (2010) Norwegian house mice (*Mus musculus musculus/domesticus*): Distributions, routes of colonization and patterns of hybridization. *Molecular Ecology*, 19, 5252–5264.

Martínková N, Searle JB (2006) Amplification success rate of DNA from museum skin collections: a case study of stoats from 18 museums. *Molecular Ecology Notes*, 6, 1014–1017.

Napoli T (2010) *Vasco Da Gama: Discovering the Sea Route to India*. Great explorers of the World Series. Enslow Publishers, Berkeley Heights, New Jersey, USA.

Nachman MW, Boyer SN, Searle JB, Aquadro CF (1994) Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. *Genetics*, 136, 1105–1120.

Pocock MJO, Hauffe HC, Searle JB (2005) Dispersal in house mice. *Biological Journal of the Linnean Society*, 84, 565–583.

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

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53, 793–808.

Prager EM, Orrego C, Sage RD (1998) Genetic variation and phylogeography of Central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics*, 150, 835–861.

Prager EM, Sage RD, Gyllensten U, Thomas WK, Hübner R, Jones CS, Noble L, Searle JB, Wilson AC (1993) Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biological Journal of the Linnean Society*, 50, 85–122.

Prager EM, Tichy H, Sage RD (1996) Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: comparison with other house mice and report of a 75-bp tandem repeat. *Genetics*, 143, 427–446.

Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe, *Molecular Ecology*, 17, 627–641.

Richardson WAR (2006) *Was Australia Charted Before 1606?: The Java La Grande Inscriptions*. National Library Australia, Canberra, Australia.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572–1574.

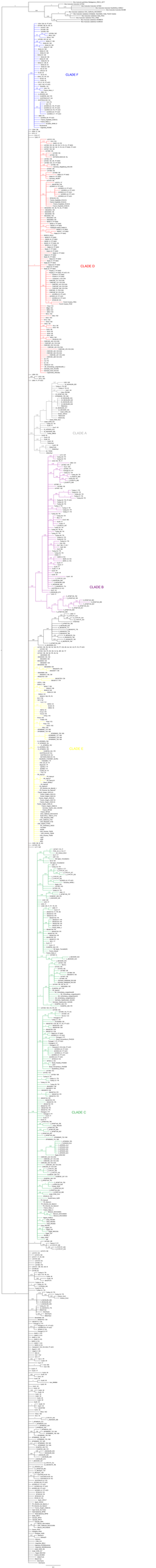
Russell-Wood AJR (1992) *The Portuguese Empire, 1415-1808: a World on the Move*. Manchester, England, UK, in association with The Calouste Gulbenkian Foundation.

Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry S, Giménez MD, Jóhannesdóttir F (2009) Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society, B* 276, 201–207.

Searle JB, Jamieson PM, Gündüz İ, Stevens MI, Jones EP, Gemmill CEC, King CM (2009b) The diverse origins of New Zealand house mice. *Proceedings of the Royal Society, B* 276, 209–217.

Supplementary material

Figure S1. Phylogenetic tree of *Mus musculus domesticus*. Detailed 50% majority rule consensus tree after Bayesian inference, summarised in Figure 1 (based on a total of 786 *Mus musculus domesticus* D-loop haplotypes, derived from this study and previous data available at the time of analysis) (Prager *et al.* 1993, 1996, 1998; Nachman *et al.* 1994; Gündüz *et al.* 2000, 2001, 2005; Ihle *et al.* 2006; Rajabi-Maham *et al.* 2008; Förster *et al.* 2009; Searle *et al.* 2009a, b; Bonhomme *et al.* 2011; Jones *et al.* 2010, 2011a,b, 2012, Gabriel *et al.* 2011, Gabriel *et al.* submitted). Posterior probabilities are shown on branch nodes. Clades are coloured as in Figure 1: outgroup and basal sequences: black, clade A: grey, clade B: purple, clade C: green, clade D: red, clade E: yellow, clade F: blue. Codes indicate GenBank number or code used in the publication where the sequence was first reported.


Rice accession gambica Madagascar_ERR11_2277
Rice accession 100_Rice
Rice accession Kaolack 147B2B
Rice accession Kaolack 147B2B
Rice accession Kaolack 147B2B

Chapter 5

Paper V

Colonization, mouse-style

Sofia I Gabriel[1], Fríða Jóhannesdóttir[2], Eleanor P Jones[3] and Jeremy B Searle[2]

[1] CESAM - Centre for Environmental and Marine Studies, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, 1749-016 Lisbon, Portugal.

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

[3] Population Biology and Conservation Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18 D, SE-752 36 Uppsala, Sweden.

Gabriel SI, Jóhannesdóttir F, Jones EP, Searle JB (2010) Colonization, mouse-style. *BMC Biology*, 8, 131.

doi:10.1186/1741-7007-8-131

Abstract

Several recent papers, including one in *BMC Evolutionary Biology*, examine the colonization history of house mice. As well as background for the analysis of mouse adaptation, such studies offer a perspective on the history of movements of the humans that accidentally transported the mice.

Commensals of humans are likely to share the human global distribution. Being easily noticed, and surviving and reproducing well in environments that humans create, they also include some of the most favored model organisms. A prime example of this is the house mouse, *Mus musculus*, which is both the 'classic' mammalian model organism and a globally present commensal. Through its association with humans, the house mouse is even found in the remotest archipelagos, such as Kerguelen, a group of sub-Antarctic islands with a mean summer temperature as low as 8°C. It is the mouse populations inhabiting this inhospitable place that are the focus of a study by Hardouin *et al.* (2010) in *BMC Evolutionary Biology*.

With all the genomic tools available, there is currently a scramble to study the genetics of adaptation in house mice. What better place to study that than Kerguelen? Here, human occupancy is restricted to the few inhabitants of a research station on the main island (Grande Terre). The mice on these islands live outdoors in extreme conditions and, in contrast to the typical seed-eating of house mice elsewhere, they feed primarily on invertebrates. To understand the adaptations for this exceptional lifestyle, it is important to know something about

the history of the mice. In particular: where did the mice come from? Is it a genetically mixed population? Is the population young or old? Hardouin *et al.* (2010) investigated all these questions for Kerguelen house mice which belong to the western subspecies, *Mus musculus domesticus*.

The Kerguelen study

Hardouin *et al.*'s (2010) study involved 437 mice from Kerguelen, an unprecedented coverage for the analysis of colonization history of such a small area. They found remarkable consistency in the mitochondrial DNA (mtDNA) sequences on Grande Terre and most of the surrounding small islands, suggesting that these populations are the product of a single relatively recent colonization (ultimately deriving from Europe). This fits with the recorded discovery of the archipelago in 1772 (by a Frenchman called Kerguelen-Trémarec) and settlement by mice either at the time or with subsequent human arrivals. Two of the other small islands in the archipelago (Cochons and Cimetière) may have been colonized in a second, separate introduction, as their mice belong to a different mtDNA lineage (also ultimately European). Over the archipelago as a whole there was no evidence of within-island heterogeneity in terms of mtDNA lineage. This is surprising given the large number of ships carrying mice that would have visited the islands (coming from many different places and therefore carrying mice of many different mtDNA lineages). These results are consistent with other data (Searle *et al.* 2009) suggesting that mouse populations are resistant to secondary invasion by females (mtDNA is a maternally inherited marker). Presumably, newly arriving females coming into an established population are generally unable to

survive or gain mates, and in consequence do not contribute to the population's gene pool. All of this means that mtDNA may be a very good marker for initial colonization by house mice within a given area.

Studies on European mice

The association of *Mus musculus domesticus* with humans in a European context has long been studied by zooarcheologists, and genomic tools are now being deployed to study adaptation in some of these European populations. With regard to colonization history, there is much zooarcheological evidence on the progression of mice from their site of first commensalism with humans in the Near East through the Mediterranean region, providing a good test on the match between mtDNA sequences and the historical record. Gratifyingly, Bonhomme *et al.* (2011) have identified a discontinuity in mtDNA lineages that fits very well with the two phases of mouse colonization of the Mediterranean revealed by zooarcheologists (Figure 1). The eastern Mediterranean was colonized by mice during the Neolithic when they were first able to exploit stored grain. However, the western Mediterranean could not be colonized by house mice until the Iron Age (Figure 1), when settlements reached a sufficient size for the house mice not to be outcompeted by local mice living outdoors, and when seafarers such as the Phoenicians carried cargoes of sufficiently large size to inadvertently transport house mice (Cucchi *et al.* 2005, Bonhomme *et al.* 2011).

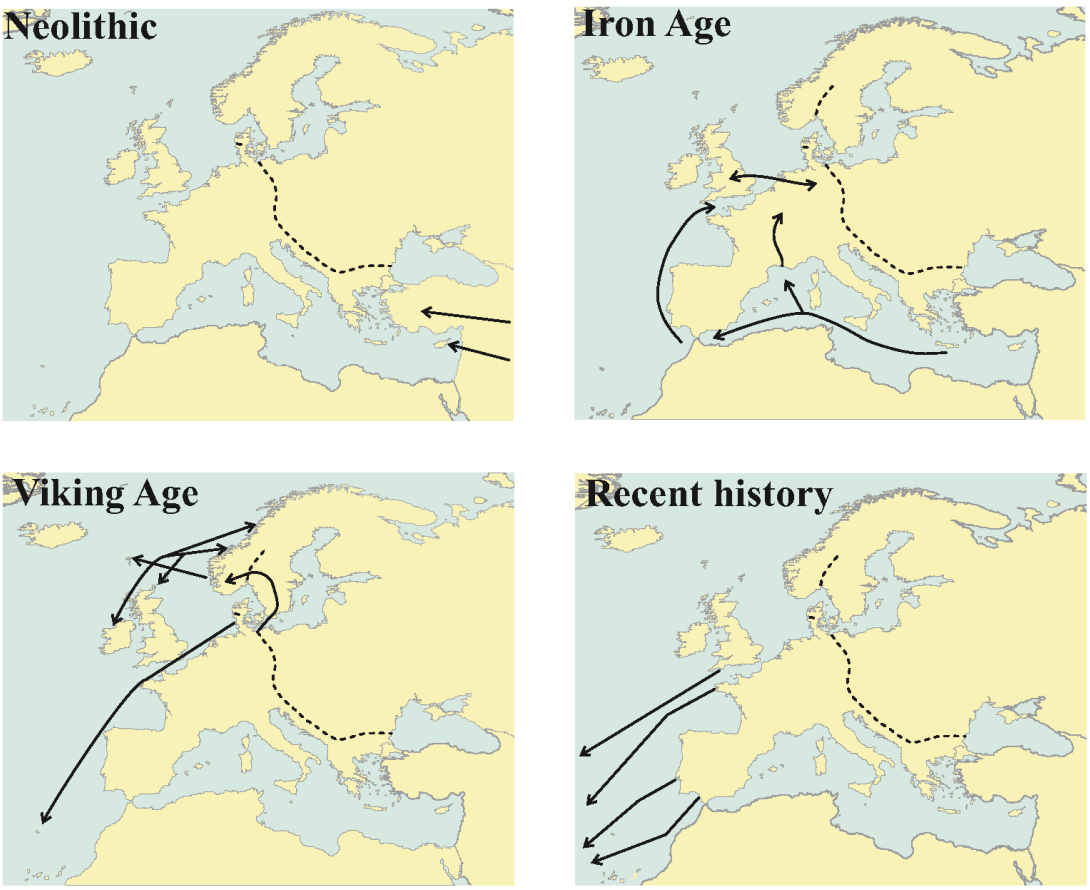


Figure 1. Maps showing possible colonization routes taken by the western house mouse *Mus musculus domesticus*, based on mtDNA evidence. Neolithic (starting 12,000 years ago): colonization restricted to the eastern Mediterranean close to where this subspecies first became commensal (Bonhomme *et al.* 2011). Iron Age (starting 3,000 years ago): colonization westwards along the Mediterranean and then into north-west Europe by overland and coastal routes (Bonhomme *et al.* 2011, Jones *et al.* 2011a). Viking Age (around 1,000 years ago): movements around the periphery of north-west Europe and colonization of Scandinavia and Madeira (Searle *et al.* 2009, Jones *et al.* 2010, 2011a,b). [The colonization of Scandinavia may have been earlier (Jones *et al.* 2010)]. Recent history (a few hundred years ago): mice were taken substantial distances from western Europe, including to Kerguelen (Hardouin *et al.* 2010). The dashed line shows the location of the hybrid zone between the subspecies *M. m. domesticus* and *M. m. musculus*.

Studies by ourselves and others have looked at the mtDNA lineages of house mice in northern Europe. A different lineage from those typically seen in Mediterranean Europe has been found further north in the area between Britain and Germany (Searle *et al.* 2009, Jones *et al.* 2011a). The mouse mtDNA again matches a regional sphere of influence of Iron Age people (Cunliffe 2004), and, unlike other mouse mtDNA lineages, it appears that this Anglo-German lineage did not arrive in northern Europe by an overland route; instead it probably came along the Atlantic coast (Figure 1).

MtDNA studies suggest another pulse of detectable mouse colonizations during Viking times (Figure 1). Like the Phoenicians, the Vikings were impressive seafarers, carrying substantial cargoes ideal for stowaway mice, and there are mtDNA signals of maritime colonization events (Searle *et al.* 2009, Jones *et al.* 2010, 2011a,b).

How did *Mus musculus domesticus* get from Europe to Kerguelen? This subspecies had been in the right place at the right time to make use of the first storage of grain by Neolithic humans in the Fertile Crescent in the Middle East, and to adapt to changing human cultural practices. Good fortune struck again when the subspecies found itself in western Europe at the time that British, Dutch, French and Iberian seafarers were 'discovering', exploiting and taking settlers to the rest of the world. Kerguelen-Trémarec and his crew may have been the first humans to see the archipelago that now bears his name, but the colonization route of the first mice to arrive there is still uncertain, although their starting point was certainly western Europe (Figure 1).

Mice as a proxy for human history

It is intriguing how far the linkage between human history and mouse history may go. Jones *et al.* (2011b) found a correlation between mouse genetic diversity and human population size (proportional to amount of mouse habitat) in discrete areas of the Faroe Islands in the northeastern Atlantic Ocean, another archipelago where house mice have been studied. This supports the expectation that the population genetics (in terms of genetic response to population expansions and contractions) of house mice is likely to reflect rather closely the population genetics of humans.

We have been considering how the history of humans impacts on the genetics of the house mouse, but that can be turned around. If the history of house mice is so intimately determined by humans, then the genetics of house mice may be useful to answer human historical questions; for example, the details of human affiliations in the Iron Age are sometimes imprecise - might house mice be able to indicate associations between Iron Age people from different geographical areas?

House mice are equivalent to an artifact that an archeologist discovers and uses to determine human colonization or trading routes. The provenance of the mice is established from their DNA sequence and that is a very powerful tool, given its extraordinary information content. Not only can the DNA sequence help to establish the source of the mice found in a particular place but it can be used to date the original colonization and subsequent population history (including secondary colonizations), following approaches used for human DNA (see, for example (Fagundes *et al.* 2007)). However, it is clear from all the recent papers considered here (Hardouin *et al.* 2010, Bonhomme *et al.* 2011, Jones *et al.* 2010,

2011a,b) that archeogenetics using house mice is at an early stage, and that, in particular, calibrations to generate an accurate mouse mtDNA molecular clock are urgently needed.

Hardouin *et al.* (2010) comment that, for the mtDNA region analyzed, they found a much higher mutation rate than suggested by previous studies. Further work should follow up this finding and also use other subspecies to globalize the opportunities for applying mice as a proxy to study humans, following the lead of another recent paper (Nunome *et al.* 2010).

Acknowledgements

SIG was funded by the scholarship SFRH/BD/21437/2005 from Fundação para a Ciência e a Tecnologia (Portugal), FJ and JBS by Cornell University (USA) and EPJ by the Carl Trygger Foundation (Sweden). We are grateful to Angela Douglas for her comments on the manuscript.

References

Hardouin EA, Chapuis J-L, Stevens MI, van Vuuren BJ, Quillfeldt P, Scavetta RJ, Teschke M, Tautz D (2010) House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evolutionary Biology*, 10, 325.

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffra J-C, 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.

Cucchi T, Vigne J-D, Auffray J-C (2005) First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarcheaological revision of subfossil occurrences. *Biological Journal of the Linnean Society*, 84, 429–445.

Cunliffe BW (2004) *Iron Age communities in Britain: an account of England, Scotland and Wales from the seventh century BC until the Roman conquest*. 4th edition. Routledge; London, UK.

Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Science USA*, 104, 17614–17619.

Jones EP, Jóhannesdóttir F, Gündüz İ, Richards MB, Searle JB (2011a) The expansion of the house mouse into north–western Europe. *Journal of Zoology*, 283, 257–268.

Jones EP, Jensen J-K, Magnussen E, Gregersen N, Hansen HS, Searle JB (2011b) A molecular characterization of the charismatic Faroe house mouse. *Biological Journal of the Linnean Society*, 102, 471–482.

Jones EP, van der Kooij J, Solheim R, Searle JB (2010) Norwegian house mice (*Mus musculus musculus/domesticus*): Distributions, routes of colonization and patterns of hybridization. *Molecular Ecology*, 19, 5252–5264.

Nunome M, Ishimoh C, Aplin KP, Tsuchiya K, Yonekawa H, Mohwaki K, Susuki H (2010) Detection of recombinant haplotypes in wild mice (*Mus musculus*) provides new insights into the origin of Japanese mice. *Molecular Ecology*, 19, 2474–2489.

Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry S, Giménez MD, Jóhannesdóttir F (2009) Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society, B* 276, 201–207.

Chapter 6

General Discussion

6. General Discussion

Over the last couple of decades it has become very clear that molecular markers can help to trace the history of expansion, spread and colonisation of plant and animal species, including humans. Thousands of studies have been carried out under a phylogeographic framework (Avise 2009), also helping to address questions related with species conservation, hybridisation and biological invasions. However, few species with a nearly global distribution have been under phylogeographic scrutiny over their complete range, or most of it.

The main focus of this thesis was to investigate the colonisation history of one of the most successful invasive mammals in the world, the western house mouse (*Mus musculus domesticus*), at three different geographic scales: insular (Azores archipelago, Chapters 2 – Papers I and II), continental (Australia, Chapter 3 – Paper III) and global (Chapter 4 – Paper IV). In order to accomplish these goals, different molecular markers were used: mitochondrial DNA (mtDNA) sequences (Chapters 2, 3 and 4), nuclear microsatellite loci (Chapter 2) and other subspecies-specific nuclear markers (Chapter 2).

Contrary to the assessment of the phylogeographic patterns of many other species, the study of the house mouse allows inferences beyond those relevant to the species itself. Due to the commensal nature of this rodent, the colonisation history of the house mouse is intimately linked to the history of long-distance movements of humans, creating a unique opportunity to indirectly address aspects of human colonisation and trade. Finding congruence between the current distribution of mitochondrial DNA variation of mice and the known history of humans as defined by historical documentation and archaeological evidence is of

extreme importance for the validation of this phylogeographic approach. Only under this premise, can later unexpected phylogeographic patterns in house mice become decisive evidence of early historical or cultural linkage between geographic locations, in the absence of any human artefacts.

House mice as proxies for human movements and colonisations

As humans we have long been fascinated by the history of our own species, how we evolved, and ultimately how we came to occupy our current global distribution. The rise of the DNA era allowed a whole new perspective on the historical peopling of continents and islands by modern humans, most widely analysed using mitochondrial and Y-chromosome data (see Oppenheimer 2012 for a review). However, high-precision inference of human movements within the last few thousand years has not always been possible to achieve with these markers (Soares *et al.* 2010).

Molecular studies on livestock (e.g. cattle, Ajmone-Marsan *et al.* 2010) have provided important insights into the process, timing and centres of domestication of these species but have also enabled crucial insights into recent human history. In the ancient Fertile Crescent region, about 10,000 years ago, along with the development of agriculture, domestication had a huge impact in human demography and cultural development. Later on, by the Iron Age, as a consequence of large settlements, substantial food storage capability and extensive maritime movement, people were incidentally setting up a parallel evolutionary experiment: the massive expansion of commensalism of house mice (Cucchi *et al.* 2005).

During the Iron Age, the Viking Age and lastly the Age of Discovery people appear to have left an indirect biological signature on the mitochondrial variation of extant house mice descending from those they transported throughout their areas of influence (Chapter 2 – 5). Regardless of the geographical scale at which one is analysing the colonisation history of house mouse populations (either a whole continent or a small island), it seems clear that there is always some underlying pattern that can be related to human expansions – colonisations or enhanced trading (Chapters 4 and 5). The distribution of the subspecies illustrates the remarkable discreteness and clarity of the phylogeographic signal in the house mouse. Despite the occasional detection of a non-*domesticus* genomic signature outside the established range of the subspecies (see Chapters 2 and 4, Bonhomme and Searle in press), an overall analysis throughout the western house mouse range (Chapter 4) has confirmed, at the mitochondrial DNA level, remarkably little introgression.

Also at smaller geographic scales there tends to be a very clear-cut pattern, with exclusively or predominantly single mitochondrial DNA clades occupying separate islands or discrete parts of islands or continents. Several recent island surveys of house mouse populations have consistently obtained this same phylogeographic pattern when using mitochondrial DNA sequences (e.g. New Zealand, Searle *et al.* 2009b; Madeira, Förster *et al.* 2009; Canary Islands, Bonhomme *et al.* 2011; Kerguelen, Hardouin *et al.* 2010; Faroe Islands, Jones *et al.* 2011; Iceland, Jones *et al.* 2012; Azores, Chapter 2). As with other biological systems, there are also exceptions to this rule of mitochondrial DNA exclusivity but, as with the other cases, the reasons for those exceptions are informative. Thus, there is an exceptional mitochondrial DNA mixture on the Mediterranean island of

Cyprus involving 6 lineages (Bonhomme *et al.* 2011). This unusually heterogeneous mitochondrial structure probably reflects an exceptionally long house mouse presence on Cyprus, with repeated introductions dating back to Neolithic times (Cucchi *et al.* 2002). Here, the normal resistance of female populations to future immigrants has apparently been overcome, leading to the presence of multiple lineages. Even on some islands of the Azores there was a degree of mixing (e.g. Terceira, Chapter 2), may be due to a local extinction event allowing penetration. Having emphasised these exceptions it should be said that in a globalised world, with mice being transported all over, it is quite astonishing that higher levels of admixture are not reported. That refers to island populations, but also continental populations (e.g. Australia: Chapter 3)

The special significance of island populations

Different islands around the world offer very distinct biotic and abiotic conditions to invasive house mice. In particular, islands where the human presence is very restricted or even nonexistent (and therefore, the commensal niche is highly confined or totally absent) may be thought to represent greater challenges for mice. Their arrival onto islands or remote landmasses is inevitably tied to a mandatory human transportation, suggesting that the first mouse colonists were commensals to start with (and therefore presumably adapted to that lifestyle), and that they may have difficulties in surviving in an outdoor habitat. However, in actual fact, one of the most characteristic traits of house mice is their resilience and ability to use new environments (Triggs 1991); they have managed to survive and thrive in some of the most inhospitable locations on the planet. Therefore, even a

single propagule that survives the selective pressures associated with the transport and arrival to the new range may form a viable population in the absence of humans and their dwellings (Berry and Scriven 2005). Under these circumstances, mice adopt a lifestyle that involves a new type of habitat to nest and move around in and a new diet. They are aided in doing that by high breeding rates and lack of predators and/or competitors. All these factors combined culminate in such a successful establishment of these island invaders that attempts to eradicate them become a very challenging task. Over the past two decades, numerous house mouse eradication attempts on small islands have been tried as a conservation tool to preserve native flora and fauna, many of which have proved unsuccessful (Howald *et al.* 2007).

So, with regards the response of humans to island house mice, in a conservation sense, the task is to ensure that house mice do not cause a threat to native species. But the human interest in island house mice goes far beyond that. They provide numerous opportunities to address interesting scientific problems relating to rapid adaptation and evolution. A few studies have used insular populations of house mice to elucidate fundamental aspects of evolutionary genetics. One of the most iconic studies was carried out on the Atlantic island of Madeira, where an extraordinary chromosomal radiation was observed (Britton-Davidian *et al.* 2000). Among the Madeiran house mouse populations, six highly divergent chromosomal races were described, departing from the standard all-acrocentric karyotype. All these races seem to have arisen *in situ* only within the last 500 – 1000 years, neighbouring each other but geographically isolated by mountain barriers (Britton-Davidian *et al.* 2005). Another example of rapid divergence among island populations of house mice was recently described in the

Sub-Antarctic archipelago of Kerguelen (Boell and Tautz 2011). Here, despite the very recent invasion, mice inhabiting two neighbouring islands already revealed a degree of morphological divergence at the mandible shape level.

This thesis has clearly emphasised the extreme interest in island systems of house mouse from the viewpoint of colonisation history. An understanding of colonisation history of house mice goes hand-in-hand with evolutionary studies. In trying to understand the evolution of house mice on a particular island, it is critical to know from where it derives and what sort of adaptive environment it may have been exposed to. It is hugely significant that mitochondrial DNA appears to provide a signature of initial colonisation, allowing colonisation history to be inferred.

Challenges for the study of house mouse colonisations

Molecular dating is a powerful statistical tool that has been widely used to determine the rate of diversification of species or test between alternative hypotheses relating to evolutionary history and specific climatic or geological events. It would be very interesting to track the recent colonisations of house mice using molecular dating but this is quite a challenging task because the timeframe of the investigated events falls somewhere between a few thousand years, at most, and the present time. As seen from the results obtained on Chapters 2, 3 and 4, the level of divergence observed between mitochondrial D-loop sequences from the source and sink populations is very low. In fact, in all sampled continents outside Europe many of the detected sequences were 100% identical to haplotypes found in the likely source areas. This lack of resolution could potentially be overcome by sequencing the whole mitochondrial genome, uncovering more mutational

differences. Whole genome studies would take interpretation to an even higher level, as illustrated by recent analyses with humans (e.g. Gronau *et al.* 2011). Numerous genomic tools are now available in a more affordable, cost-effective and high-throughput fashion. Current genomic approaches in the house mouse include two genome SNP arrays commercially available (the Mouse Universal Genome Array with ~8000 SNPs, from Geneseek; and the Mouse Diversity Array with ~620,000 SNPs, from Jax).

Another big limitation of any phylogeographic study is related to adequate sample size and geographic coverage of the surveyed species, two challenging but critical aspects for the most accurate inferences to be drawn. Therefore it is clear that adequate sampling is crucial for a valid use of house mice as proxies of humans, both to minimize the likelihood of the source populations remaining unsampled and also to provide a better measure of the real extension of the range of each clade. Obviously, this is one of the biggest challenges in the assessment of the house mouse colonisation history, considering the ubiquitous distribution of the species. It is notable that the increasing number of published studies using the mitochondrial D-loop as a molecular marker has greatly improved our understanding of the colonisation routes that have shaped the patterns of genetic variation of modern house mouse populations. This dissertation has substantially contributed to this effort, covering, for the first time, extensive parts of the *domesticus* subspecies range (e.g. Australia, Chapter 3), filling important gaps in many poorly sampled regions (e.g. North and South America, sub-Saharan Africa, Chapter 4), and surveying previously unsampled remote islands (e.g. New Caledonia and Vanuatu, Chapter 4; the Azores archipelago, Chapter 2). The phylogeographic patterns of the western European subspecies, *M. m. domesticus*, as

addressed here at different geographical scales (Chapters 2 – 5), have emphasised and reinforced the use of the house mouse as a reliable proxy of human movements and colonisations.

Future research

The achievements of this thesis can be extended in several directions. Building on the results obtained, new questions can be suggested, warranting further investigation. The house mouse is a privileged species to work with as it lacks many limitations of non-model species. As the genomic era flourishes, with techniques and analytical tools becoming more common and accessible, the knowledge of the whole mouse genome sequence (Mouse Genome Sequencing Consortium, 2002) is still a major asset. As such, future research projects using house mouse populations (either insular or continental) could address research questions as follows:

- Given the nearly global distribution of the species, phylogeography of the house mouse can always be seen as work in progress. Despite the effort to significantly enlarge the geographical coverage of the western house mouse, a lot more could be done to continue to unveil the colonisation routes that help to explain the current distribution of this subspecies. In particular, better sampling of sub-Saharan Africa would provide new insights into the undoubtedly complex colonisation history of this continent.
- This thesis has focused exclusively on the most common and widespread *domesticus* subspecies, the western house mouse. Over the past couple of decades, a clear bias against the other most common subspecies, *M. m. musculus*

and *M. m. castaneus*, has been translated in the lack of literature addressing their phylogeography. The *musculus* subspecies has been thoroughly studied, but mostly in a very narrow geographic range in Eastern Europe and mostly focusing on the hybridization aspects resulting of the secondary contact with *domesticus*. A phylogeography of the whole species would be fascinating and valuable. There are many issues of human history relating to the areas occupied by *M. m. musculus* and *M. m. castaneus*, that could reasonably be addressed using phylogeography. For example, studies of *M. m. castaneus* could help document the extent of human movements across the Indian Ocean between South Asia and East Africa, going back in time from modern historical periods (Fuller and Boivin in press).

- Both the exclusively maternally inherited mitochondrial DNA and bi-parentally inherited microsatellites have been applied to the study of house mice in the Azores (Chapter 2) and islands elsewhere (Faroe, Jones *et al.* 2011; Iceland, Jones *et al.* 2012). They have provided perspectives both on colonisation of islands and later movements between islands in archipelagos. Despite the known low diversity of the mammalian Y-chromosome (Hellborg and Ellegren 2004), it would be interesting to use this marker on island house mice. The addition of these data would provide a good measure of both male and female origins and the better assessment of migration patterns (and whether there is evidence of sex-biased dispersal movements).

- One of the challenges of the phylogeographic approach employed in this thesis relates to the impossibility to confirm whether the modern distribution of mitochondrial haplotypes represents the original one, resulting from the first

successful colonisations. To some extent, this is a limitation for all phylogeography studies as, for whatever reason, complete replacements of populations or lineages of a certain species can always occur. A way to appraise this is to use the latest ancient DNA techniques on samples retrieved from archaeological excavations to confirm if the current genetic makeup is reflecting the original one. Resembling the approach followed by Jones *et al.* (2012), it would be interesting to assess through carbon dating and genetic analysis if the mitochondrial signature of the primary Azorean colonists persisted to the modern populations inhabiting the islands today. However, ancient DNA analysis has notorious limitations that increase with the age of the archaeological sample and the warmer the climate of the region from where the samples were retrieved. Also, it not always easy to obtain archaeological bone samples that are available for destructive sampling.

- Using the mitochondrial and microsatellite markers for the Azorean house mouse populations, particular groups of islands were particularly divergent from each other (Chapter 2). It would be interesting to examine whether the genetic differentiation between these island populations is reflected on the morphological level. As a replacement of classic morphometric analysis, a new set of techniques, so-called 'Geometric Morphometrics' (Zelditch *et al.* 2004), is now available to capture more objective shape information, decoupling size from shape, which allows subtle but important differences to be found. Using a landmark-based approach on the 2D or 3D digital surface of the morphological structure (skull, mandible or teeth), a multivariate statistical analysis could be applied on the totality of the landmark configuration in order to extract as much of the shape variation as possible.

Final remarks and future prospects for the field

We have reached an era in which available data sets of large numbers of sequences are more and more common (e.g. the 1000 Human Genomes Project, sequencing across 27 human populations for a full picture of genomic diversity – 1000 Genomes Project Consortium, 2010). The technology for the generation of complete genome data of model and non-model species has seen a massive development in the past few years, rapidly replacing the traditional individual genes approach. As a consequence, some of the major challenges for the immediate future lie in the development of more efficient next generation tools to analyse, display and integrate large-scale genomic data (e.g. SNP-chips, next generation sequencing) (Hawkins 2010). Of course, large datasets and powerful statistic tools require computational resources and storage space that are nowadays a bigger limitation than the data production itself. Therefore, it is becoming a tangible prospect that a shift will soon occur in the way phylogeography studies are conducted worldwide and not just in a limited number of top laboratories. As soon as the exploitation of these new resources of genomic variation become affordable and consequently the rule and not the exception (not just within reach of large consortia), the somewhat excessive reliance on small fractions of the genome, such as the popular uniparental markers, will come to an end. Therefore, not too far in the future, one can expect that phylogeographic studies using the house mouse as a model species will be taken to a much deeper level. Questions related to how all the different subspecies originated and diverged, coming to occupy their current range (if not a reassessment of their taxonomic status) will most likely obtain robust answers. Tracking the colonisation of the house mouse onto islands and

archipelagos worldwide, determining the origin of those populations or understanding the different local adaptions are among the numerous questions that would greatly benefit from a genomic approach. Islands around the world are located throughout all latitudes and longitudes, all biogeographic regions, are under the influence of very distinct climatic conditions, and have all sorts of resident communities of plants and animals. This diversity of biotic and abiotic conditions translates into distinct challenges that mice have to face and certainly no other mammal is in a similar position of ubiquity to allow such a vast coverage of variables to study. Over recent years we have witnessed a cascade of new methodologies that allow an integrative analysis of genome-wide scale data. This would certainly reinforce the use of the house mouse as a research model in numerous areas of field biology, namely, adaptation, invasion or evolution, and not just the biomedical analyses of physiology, immunology or genetics.

References

1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature*, 467, 1061–1073.

Ajmone-Marsan P, Garcia JF, Lenstra JA (2010) On the origin of cattle: How aurochs became cattle and colonized the world. *Evolutionary Anthropology: Issues, News, and Reviews*, 19, 148–157.

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

Berry RJ, Scriven PN (2005) The house mouse: a model and motor for evolutionary understanding. *Biological Journal of the Linnean Society*, 84, 335–347.

Boell L, Tautz D (2011) Micro-evolutionary divergence patterns of mandible shapes in wild house mouse (*Mus musculus*) populations. *BMC Evolutionary Biology*, 11, 306.

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffray J-C, 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.

Bonhomme F, Searle JB (in press) *House mouse phylogeography*. In *Evolution of the house mouse* (Cambridge Series in Morphology and Molecules). eds Macholán M, Baird SJE, Munclinger P, Piálek J (Cambridge University Press, Cambridge).

Britton-Davidian J, Catalan J, Ramalhinho M, Ganem G, Auffray J-C, Capela R, Biscoito M, Searle JB, Mathias ML (2000) Rapid chromosomal evolution in island mice. *Nature*, 403, 158.

Britton-Davidian J, Catalan J, Ramalhinho G, Auffray J-C, Nunes AC, Gazave E, Searle JB, Mathias ML (2005) Chromosomal phylogeny of Robertsonian races of the house mouse on the island of Madeira: testing between alternative mutational processes. *Genetical Research*, 86, 171–183.

Cucchi T, Vigne JD, Auffray J-C, Croft P, Peltenburg E (2002) Passive transport of the house mouse (*Mus musculus domesticus*) to Cyprus at the Early Preceramic Neolithic (late 9th and 8th millennia cal. BC). *Comptes Rendus Palevol*, 1, 235–241.

Cucchi T, Vigne J-D, Auffray J-C (2005) First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the western Mediterranean: a zooarchaeological revision of subfossil occurrences. *Biological Journal of the Linnean Society*, 84, 429–445.

Förster DW, Gündüz İ, Nunes AC, Gabriel S, Ramalhinho MG, Mathias ML, Britton-Davidian J, Searle JB (2009) Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology*, 18, 4477–4494.

Fuller DQ, Boivin N in press. Crops, cattle and commensals across the Indian Ocean: Current and potential archaeobiological evidence. In *Plants and people in the western Indian Ocean*. Etudes Océan Indien, Centre d'Études et de Recherche sur L' Océan Indien Occidental (CEROI).

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, 43, 1031–1034.

Hardouin EA, Chapuis J-L, Stevens MI, van Vuuren BJ, Quillfeldt P, Scavetta RJ, Teschke M, Tautz D (2010) House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evolutionary Biology*, 10, 325.

Hawkins RD, Hon GC, Ren B (2010) Next-generation genomics: an integrative approach. *Nature Reviews Genetics*, 11, 476–486.

Hellborg L, Ellegren H (2004) Low levels of nucleotide diversity in mammalian Y chromosomes. *Molecular Biology and Evolution*, 21, 158–163.

Howald G, Donlan CJ, Galván JP, Russell J, Parkes J, Samaniego A, Wang Y, Veitch D, Genovesi P, Pascal M, Saunders A, Tershy B (2007) Invasive rodent eradication on islands. *Conservation Biology*, 21, 1258–1268.

Jones EP, Jensen J-K, Magnussen E, Gregersen N, Hansen HS, Searle JB (2011) A molecular characterization of the charismatic Faroe house mouse. *Biological Journal of the Linnean Society,* 102, 471–482.

Jones EP, Skirnisson K, McGovern TH, Gilbert MTP, Willerslev E, Searle JB (2012) Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *BMC Evolutionary Biology*, 12, 35.

Oppenheimer S (2012) Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map*. Philosophical Transactions of the Royal Society B,* 367*,* 770–784*.*

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420, 520–562.

Searle JB, Jamieson PM, Gündüz İ, Stevens MI, Jones EP, Gemmill CE, King CM (2009) The diverse origins of New Zealand house mice. *Proceedings of the Royal Society B,* 276, 209–217.

Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt H-J, Torroni A, Richards M (2010). The archaeogenetics of Europe. *Current Biology*, 20, R174–R183.

Triggs GS (1991) The population ecology of house mice (*Mus domesticus*) on the Isle of May, Scotland. *Journal of Zoology*, 225, 449–468.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) *Geometric morphometrics for biologists: a primer* (Elsevier, San Diego).