

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL



**Ciências
ULisboa**

História da colonização do ratinho-caseiro, *Mus musculus domesticus*, em ilhas Atlânticas (Madeira, Açores e Cabo Verde): uma abordagem multilocus.

Ana Sofia Carromeu dos Santos

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada por:
Doutora Sofia Gabriel
Professor Doutora Maria da Luz Mathias

2017

Agradecimentos

Parece que chegou a altura de agradecer a todos os que contribuíram para esta etapa da minha vida, não só aos que para ela contribuíram ativamente como aqueles que a viveram comigo.

Queria começar por agradecer às minhas orientadoras Sofia Gabriel e Maria da Luz Mathias por me terem orientado durante este projeto. Queria agradecer em especial à Sofia por todos os ensinamentos científicos e pessoais e pelo apoio dado durante todo o mestrado. Queria ainda deixar um agradecimento ao grupo pelas questões levantadas e sugestões dadas durante o desenvolvimento da tese.

Quero agradecer à minha família, pois sem ela não estaria aqui hoje.

Aos meus pais que fizeram todos os esforços para me dar tudo, o que podiam e até o que não podiam. Nunca haverá palavras suficiente para agradecer a educação e todo o apoio que me deram durante toda a minha vida.

À minha irmã preferida (não tenho outra, mas fica sempre bem :P) por toda a educação que me deu, todas as questões que levantou durante as minhas escolhas, por todas as idas à praia, caipirinhas, compras e férias, mas principalmente por estar sempre onde eu precisei de ti. Obrigada por seres a melhor mana do mundo!

Aos meus avós por todas as gracinhas que tiveram de aturar, todos os “jogos da bola” e todas as flores (partidas) perdoadas, todos os “almocinhos para a menina”, e todos os sorrisos com que me recebiam cada vez que chegava a casa ao fim-de-semana.

Agora quero passar a agradecer à minha outra família. Não são irmãos e irmãs de sangue (felizmente, porque não imagino o que seria partilhar genes com esta gente) mas também não serão só amigos, são família. Queria agradecer a todos os passaram por esta fase comigo. Agradecer todas as viagens, passeios, jantares e arruaças, todo o meu mau humor e energia a mais que aturaram, todas as piadas sem graça e todos os meus estados de graça. Queria agradecer também todas as vezes que me desencaminharam, pois sem dúvida que esta aventura não teria tido metade da graça se não fossem vocês. Por isso obrigada:

- Às Bacanas do TUPPERWARE por todo o apoio dado mesmo quando fico semanas sem dar notícias da minha existência – Mariya Kosac, Carolina Condenço, Telma Silva, Carolina Nunes, Carolina Alves e Rita Rosas, principalmente às últimas 3 estalarolas por me aturarem desde 2011 (a serio, eu sinto a vossa dor :P).

- Ao meu afilhado Leo, por ser um chato do pior e por tentar manter a madrinha na linha. E à Martinha por ser sempre preocupada e uma querida.

- A todos os Turbinados e mestres da Arruaça por alegrarem a minha vida, por serem iguais a mim e piores (bem piores) e por todas as memórias construídas que jamais esquecerei. Ao Team MAC/Kel Team/Modern Family, um agradecimento especial por aguentarem quase viver comigo continuamente durante este tempo, sabe Deus que não foi fácil. Não foi fácil, para mim né? Porque para vocês foi um privilégio certamente. Espero que venham mais viagens das nossas com tudo a que temos direito: “roubos” (de molas), carros por cima de pedras, carros em que não cabem nem mais uma agulha, aldeias interessantes, faroladas e observação de passarinhos/aves, trutas, subidas a palco etc etc. OBRIGADO POR ME ATURAREM - Mariana/30, Vans/Vanilha, Bá/Babaloo, Ritinha/Denise, Maria/Inês/Clara/Clarice, Maria/Maraiazz/e Gamado/Gamadix.

- À Tânia Lampreia por todo o incentivo e confiança cega nas minhas capacidades e por todas as viagens a ouvir as minhas lamurias.

- À Inês José por me encorajar a fazer aquilo que gosto, mesmo quando isso parece uma loucura. E naturalmente por todo o apoio desde a escolha do curso a toda a minha vida.

- À Beatriz Painho pelo apoio incondicional, pela força e motivação e principalmente por acreditar em mim mesmo quando eu própria não acreditava. Certamente, eu também não estaria aqui hoje se não fosse pelo teu apoio.

Por último, quero fazer um agradecimento especial à Sissoca, Marianinha, Lulinha e Rita pela ajuda com a tese, foram um excelente controlo de qualidade.

Obrigado mais uma vez a todos, eu pago quando tiver trocos ;)

Resumo

As ilhas oceânicas são consideradas laboratórios naturais para o estudo da evolução. Os humanos são também agentes de colonização de várias espécies, nomeadamente, mamíferos terrestres, ao transportarem, ativa ou passivamente, estes animais para as ilhas. Uma das introduções mediadas pelo Homem mais icónicas envolve a disseminação mundial do ratinho-caseiro, principalmente da sub-espécie *Mus musculus domesticus*, com a qual desenvolveu uma associação comensal, iniciada com o a sedentarização humana há 15000 anos. Esta associação permitiu que o ratinho-caseiro quebrasse todas as barreiras geográficas e atingisse as ilhas oceânicas, aquando das primeiras expansões marítimas, colonizando-as.

Neste estudo, foi abordada a colonização desta sub-espécie nos arquipélagos da Macaronésia (Madeira, Açores e Cabo Verde) de colonização histórica portuguesa, através de marcadores moleculares de transmissão exclusivamente materna (D-loop, mtDNA), exclusivamente paterna (microsatélites – cromossoma Y) e bi-parental (intrão *Ocrl* – cromossoma X), numa abordagem filogeográfica. A análise dos vários marcadores permitiu avaliar a origem da colonização inicial através do mtDNA e seguir possíveis vagas secundárias de colonização através da dispersão diferencial de machos e fêmeas.

Os resultados obtidos para o marcador D-loop permitiram reforçar a hipótese de que os países do Norte da Europa constituem a fonte provável de colonização do arquipélago da Madeira. Na ilha da Madeira, para os microsatélites do cromossoma Y, foi detetada uma associação a Portugal continental, que se considera posterior à primeira colonização da ilha. Este sinal de machos com uma origem possivelmente diferente da primeira colonização levou à divergência das ilhas da Madeira e Porto Santo para este marcador. O arquipélago dos Açores reflete na sua maioria a ligação histórica, a Portugal continental, numa colonização inicial. Porém, as ilhas de São Jorge e Santa Maria representam uma exceção ao nível do marcador mitocondrial (já descrita) que é agora refletido também para o cromossoma Y, neste estudo. Este padrão aponta para uma colonização diferente destas ilhas em relação às restantes ilhas do arquipélago dos Açores e entre si.. No arquipélago de Cabo Verde, a análise da população da Ilha de Santa Luzia originou padrões diferentes em cada marcador, tendo sido sugerida uma primeira colonização oriunda das populações residentes da ilha da Madeira, à qual se seguiram várias ondas de colonização.

A utilização conjunta dos diferentes marcadores moleculares permitiu avançar no conhecimento da história da colonização do ratinho-caseiro nas ilhas estudadas, mas revelou um cenário bastante complexo e de difícil interpretação em algumas delas, principalmente pela falta de dados para algumas áreas geográficas que poderão ter funcionado como fonte colonizadora. A ligação a Portugal continental, país de onde partiram os descobridores e colonizadores destas ilhas, surge na colonização inicial da maioria das ilhas dos Açores, mas também numa colonização secundária no arquipélago da Madeira, reforçando a ocorrência de uma passagem anterior à portuguesa por esta ilha. Por último, a colonização de Santa Luzia sugere uma colonização inicial através de indivíduos com origem no arquipélago da Madeira, refletindo no entanto, um segundo sinal de colonização de machos com associação à ilha de Santa Maria. Este cenário prevê um padrão de colonização complexo para o arquipélago de Cabo Verde, sendo que esta ilha deve refletir, pelo menos em parte, o padrão das ilhas adjacentes devido à sua utilização como porto piscatório.

Palavras chave

Mus musculus domesticus, D-loop, *Ocrl*, microsatélites do cromossoma Y, colonização de ilhas.

Abstract

Oceanic islands are considered natural laboratories for the study of evolution. Humans are also colonizing agents for several species, mainly terrestrial mammals, by actively or passively transporting the animals to the islands. One example of species introduction mediated by humans is the worldwide dissemination of the house mouse, most notably of the subspecies *Mus musculus domesticus*, with which it developed a commensal relationship that began 15.000 years ago. This relationship allowed the house mouse to overcome all geographic barriers and colonize oceanic islands, ever since the beginning of the European maritime expansion.

The present study focuses on the colonization history of the house mouse in the Macaronesian archipelagos historically colonized by the Portuguese (Madeira, Azores and Cape Verde) under a phylogeographic approach. To accomplish this, different molecular markers: exclusively maternal inheritance (D-loop, mtDNA), exclusively paternal inheritance (microsatellites – Y chromosome) and biparental inheritance (*Ocrl* intron – X chromosome). The overall analysis of the different markers allowed the inference of the origin of the initial colonization through the mtDNA analysis and the identification of possible secondary routes of colonization through the results of the X and Y chromosome markers, allowing the assessment of the differential dispersion of males and females.

The results obtained for the D-loop reinforced that Northern European countries are the most likely source of initial colonization of the Madeira archipelago. From the microsatellites of the Y chromosome it was detected an association between Madeira and mainland Portugal, which was interpreted as the result of secondary colonization events. These male-inherited markers also revealed a separation between the Madeira and Porto Santo islands. The Azores archipelago displays, in its majority, the historical connection to mainland Portugal through the initial wave of colonization. However, São Jorge and Santa Maria islands represent an exception at the level of the mitochondrial marker (previously described), which was also observed in this study for the Y chromosome markers. This pattern points towards the possibility of a different wave of colonization of these island, separate from the colonization of the rest of the archipelago. In the Cape Verde archipelago, the analysis of the population from the Santa Luzia island originated different patterns for each marker, which led to the suggestion of a first wave of colonization from the populations occupying Madeira, followed by several secondary waves of colonization.

The joint analysis of the different molecular markers contributed to advances in the knowledge of the dissemination, throughout history, of the house mouse in the studied archipelagos. However, in some islands it revealed a very complex scenario of difficult interpretation, especially due to the scarcity of data for some of the geographical areas that may have contributed as colonization sources. The connection to mainland Portugal, the historical discoverer and colonizer, arises in the initial colonization of most Azorean islands but also in a secondary colonization of the Madeira archipelago, reflecting the posterior and continuous presence of the Portuguese over the centuries. Finally, the colonization of Santa Luzia suggests an initial colonization through individuals originated from the Madeira archipelago, showing, nonetheless, a signal of male colonization associated with Santa Maria island. This scenario predicts a complex pattern of colonization of the Cape Verde archipelago, as the results for Santa Luzia seem to be somewhat reflecting the diversity patterns of the adjacent islands.

Key words

Mus musculus domesticus, D-loop, *Ocrl*, Y chromosome microsatellites, island colonization.

Índice

Introdução.....	1
Ratinho-caseiro, <i>Mus musculus</i>	1
Marcadores moleculares.....	2
Colonização em ilhas.....	4
Objetivos	6
Materiais e métodos.....	7
Amostragem	7
Extração de DNA	8
Sexagem	8
<i>mtDNA – D-loop</i>	10
Amplificação e Sequenciação	10
Análise filogenética.....	10
Análise da Estrutura Populacional e Demográfica	11
Microssatélites (cromossoma Y).....	11
Amplificação e Genotipagem	11
Análise de diversidade e estrutura populacional	12
<i>Ocr1</i> (cromossoma X).....	13
Amplificação e Sequenciação	13
Análise filogenética.....	14
Análise da estrutura populacional e demográfica.....	14
Resultados	15
<i>mtDNA</i>	15
Análise filogenética.....	15
Análise da estrutura populacional e demográfica.....	17
Microssatélites.....	23
Análise de diversidade.....	23
Análise da Estrutura Populacional.....	24
<i>Ocr1</i>	29
Análise filogenética.....	29
Análise da estrutura populacional e demográfica.....	30
Discussão.....	35
Bibliografia.....	43
Anexos.....	47

Anexo I – Quadro resumo de todas as amostras de <i>Mus musculus domesticus</i> utilizadas e análises laboratoriais realizadas por população/ilha. As amostras cujo sexo foi determinado através do protocolo de sexagem estão assinaladas com um asterisco.....	48
Anexo II - <i>Primers</i> utilizados neste estudo para cada marcador (d-loop, cromossoma Y, <i>Ocr1</i>). Descrição da sequência e referência correspondente. <i>Primers</i> utilizados para cada marcador.....	56
Anexo III– Dados de microssatélites (cromossoma Y) corrigidos, no programa TADEM, para a sub-espécie <i>Mus musculus domesticus</i> nas ilhas atlânticas de colonização portuguesa.	57
Anexo IV - Quadro de haplótipos para o marcador d-loop, amostrados em todas as ilhas do Arquipélago dos Açores e da Madeira, ilha de Santa Luzia (Cabo Verde), Portugal e Espanha, para a sub-espécie de <i>Mus musculus domesticus</i> . As amostras utilizadas neste estudo estão assinaladas com um asterisco.	63
Anexo V - Quadro de haplótipos obtidos para os microssatélites do cromossoma Y para os indivíduos da sub-espécie <i>Mus musculus domesticus</i> amostrados em todas as ilhas do Arquipélago dos Açores (exceto a ilha do Corvo), Madeira, Porto Santo, Santa Luzia (Cabo Verde) e Portugal continental.	67
Anexo VI - Quadro de haplótipos do gene <i>Ocr1</i> obtidos para os indivíduos da sub-espécie <i>Mus musculus domesticus</i> amostrados em todas as ilhas do Arquipélago dos Açores, Madeira, Porto Santo, Santa Luzia (Cabo Verde) e Portugal continental (assinalados com um asterisco). Foram também incluídos os haplótipos já publicados num estudo anterior, originários de Espanha, Israel, Itália, Grécia, Reino Unido (Geraldés <i>et al.</i> 2008).	70

Índice de Figuras

Figura 1 - Distribuição das sub-espécies de <i>Mus musculus</i> no Mundo: <i>Mus musculus domesticus</i> – azul; <i>Mus musculus castaneus</i> – verde; <i>Mus musculus musculus</i> – vermelho. As zonas axadrezadas representam zonas de hibridação. Setas a vermelho – representam as rotas históricas de migrações e recentes movimentações em associação com o Homem. Retirado de Phifer-Rixey & Nachman 2015...2	
Figura 2 – Eficácia relativa dos marcadores moleculares: taxa mutacional vs. tempos de coalescência (adaptado de Zink & Barrowclough, 2008).....3	
Figura 3 – A - Mapa da área geográfica onde foi realizada a amostragem de ratinho-caseiro para este estudo: Portugal continental, arquipélado da Madeira, arquipélago dos Açores e arquipélago de Cabo Verde. B – Mapa do arquipélago dos Açores; C – Mapa do arquipélago da Madeira; D – Mapa do arquipélago de Cabo verde, com indicação da ilha de Santa Luzia (única ilha amostrada).....7	
Figura 4 – Análise de picos de microssatélites. A – Verificação da qualidade e altura dos picos de microssatélites no Programa PEAK SCANNER. A amostra representada ilustra o resultado da amplificação de dois microssatélites do cromossoma Y (Y21, a verde – fluorocormo HEX; e Y23, a azul – fluorocormo FAM). A vermelho encontra-se a escada interna GS500(-250), com marcação ROX. B – Correção dos alelos dos microssatélites obtidos através do programa TANDEM. As barras pretas representam a frequências do tamanhos de alelos obtidos e as barras cinzentas indicam os tamanhos dos alelos corrigidos.....12	
Figura 5 – A figura representa (a) a distribuição geográfica das sub-espécies de ratinho-caseiro: <i>Mus musculus domesticus</i> (azul), <i>Mus musculus musculus</i> (vermelho) and <i>Mus musculus castaneus</i> (amarelo) e (b) a genealogia do gene <i>Ocr1</i> para as mesmas sub-espécies (adaptado de Nachman & Payseur 2012).13	
Figura 6 – Árvore Filogenética Bayesiana para o marcador D-loop para todas as sequências publicadas e não publicadas até à data da sub-espécie <i>Mus musculus domesticus</i> . Os haplótipos novos amostrados neste estudo encontram-se marcados com um símbolo abertos e os haplótipos já amostrados com símbolo fechados. A forma do símbolo representa onde foi amostrado: estrelas – Cabo Verde; círculos – Madeira; quadrados - Porto Santos.16	
Figura 7 – Mapa das clades de D-loop de <i>M. m. domesticus</i> amostradas na Europa Ocidental e ilhas atlânticas. As clades estão representadas em círculos coloridos. clade B – roxo; clade C – verde; clade D – vermelho; clade E – amarelo e clade F - azul. Adaptado de Jones & Searle (2015).17	
Figura 8 - Rede de haplótipos <i>Neighbor-Joining</i> para sequências D-loop de <i>Mus musculus domesticus</i> de Portugal, Espanha, Açores, Madeira e Cabo Verde. Cada círculo representa um haplótipo único e o seu tamanho representa a frequência de ocorrência. O comprimento das linhas que ligam os haplótipos é proporcional ao número de mutações entre os mesmos e os pontos vermelhos são haplótipos intermédios hipotéticos ou não amostrados.....19	
Figura 9 - Estrutura populacional de <i>Mus musculus domesticus</i> inferida pelo STRUCTURE baseados no genótipo dos microssatélites Y6, Y12, Y24, Y21 e Y23 para o arquipélago dos Açores (exceto o Corvo), e da Madeira, ilha de Santa Luzia (Cabo Verde) e Portugal continental. Cada indivíduos está representado por uma linha colorida correspondente a proporção de semelhança de genótipos. As linhas pretas marcam a delimitação das ilhas. Este gráfico de barras representa o número mais provável de populações genéticas, K=6.25	
Figura 10 – Análise de Coordenadas Principais (PCoA) para as populações de ratinho-caseiro estudadas para os microssatélites Y6, Y12, Y24, Y21 e Y23 nas ilhas do Arquipélago dos Açores (exceto o Corvo), Madeira, Porto Santo, Santa Luzia e Portugal continental. Os três primeiros eixos explicam 72.12%, 12.04% e 8.52% da variação, respetivamente. Cada símbolo representa um genótipo, colorido de acordo com a ilha em que foi amostrado.....26	

Figura 11 – Rede de haplótipos baseada nas distâncias de alelos partilhados entre a todas as ilhas do arquipélago dos Açores (exceto a ilha do Corvo), as ilhas da Madeira, Porto Santo, Santa Luzia e Portugal Continental. Apenas ligações com distâncias iguais ou inferiores ao <i>percolation threshold</i> (0.4) estão representadas. Os genótipos estão identificados consoante a ilha em que foram amostrados.....	27
Figura 12 - Árvore filogenética Bayesiana para o intrão <i>Ocrl</i> no cromossoma X para todas as sequências obtidas neste estudo e publicadas da sub-espécie <i>Mus musculus domesticus</i>	29
Figura 13 - Rede de haplótipos <i>Neighbour-Joining</i> para o marcador do gene <i>Ocrl</i> de <i>Mus musculus domesticus</i> . Cada círculo representa um haplótipo único sendo o seu tamanho proporcional à frequência de ocorrência. As cores são referentes às ilhas/arquipélagos e países onde foram detetados. O tamanho das linhas pretas que ligam os haplótipos é proporcional à distância mutacional. A – Representação de todos os haplótipos amostrados no arquipélago dos Açores e da Madeira, ilha de Santa Luzia (Cabo Verde), Portugal, Espanha, Grécia, Itália, Israel e Reino Unido. A linha a preto realça parte da rede de haplótipos referente aos haplótipos detetados nos Açores. B – Representação dos haplótipos encontrados no arquipélago dos Açores, por ilha, realçada em A.....	31

Índice de Quadros

Quadro 1- Número de indivíduos amostrados em cada população (ilha/arquipélago/continente) para os vários marcadores analisados: D-loop (mtDNA), microssatélites (cromossoma Y) e gene <i>Ocr1</i> (cromossoma X). As sequências publicadas utilizadas estão assinaladas com asterisco (*).....	9
Quadro 2 - Variabilidade genética do marcador D-loop (mtDNA) para cada população de <i>Mus musculus domesticus</i> em análise. O número de haplótipos, número de sítios variáveis, diversidade haplotípica (Hd) e diversidade nucleotídica (π), bem como os testes de neutralidade D de Tajima e FS de Fu foram determinados no programa Arlequin v 3.5.2.2. Os valores de P, quando significativos, dos testes D de Tajima e FS de Fu não estão assinalados com um asterisco (*) ($P < 0.05$).....	20
Quadro 3 - Matriz de diferenças <i>pairwise</i> para o marcador D-loop. Todos os valores da Quadro são significativos com exceção dos representados com um asterisco.	21
Quadro 4 - Análise Molecular da Variância entre arquipélagos e ilhas.	22
Quadro 5 – Quadro das medidas de diversidade genética dos microssatélites para o cromossoma Y.	23
Quadro 6 – Diversidade haplotípica dos microssatélites.....	24
Quadro 7 - Matriz de diferenças <i>pairwise</i> para os microssatélites Y6, Y12, Y24, Y21, Y23 do cromossoma Y. Todos os valores da Quadro são significativos com exceção dos representados com um asterisco.....	25
Quadro 8 - Variabilidade genética para o intrão (960pb) do gene <i>Ocr1</i> na sub-espécies <i>Mus musculus domesticus</i> para todas as ilhas do Arquipélago dos Açores, ilha da Madeira e Porto Santo do Arquipélago da Madeira, ilha de Santa Luzia do Arquipélago de Cabo Verde, Portugal, Espanha, Grécia, Israel, Itália e Reino Unido. O número de haplótipos (Hap), número de sítios variáveis (S), diversidade haplotípica (Hd) e diversidade nucleotídicas (π), bem como os testes de Neutralidade D de Tajima e FS de Fu foram determinados no programa Arlequin v 3.5.2.2. Os valores de P, quando significativos, dos testes D de Tajima e FS de Fu, estão assinalados com um asterisco (*) ($P < 0.05$).	32
Quadro 9 - Matriz de diferenças <i>pairwise</i> para o marcador <i>Ocr1</i> . Os valores significativos ($p < 0.05$) estão representados com um asterisco.....	33
Quadro 10 – Análise molecular da variância (AMOVA) para o gene <i>Ocr1</i> . Os valores de p estão definidos para um nível de significância de 0.05.	34

Lista de Abreviaturas

DNA – ácido desoxiribonucleico

dNTP – Desoxirribonucleotídeos Fosfatados

mtDNA – DNA mitocondrial

Ocrl - inositol polyphosphate 5-phosphatase OCRL-1

PCR – Polymerase Chain Reaction)Polymerase Chain Reaction

pb – Par de bases

TBE – tris-borato-EDTA

SMA – Santa Maria

SM – São Miguel

T - Terceira

GR - Graciosa

SJ – São Jorge

P – Pico

F - Faial

FL - Flores

CR – Corvo

M – Madeira

PS – Porto Santo

SL – Santa Luzia

PT - Portugal

NOTA: o separador decimal utilizado nesta tese segue a representação com ponto (e não vírgula).

Introdução

As ilhas têm sido consideradas laboratórios naturais de investigação de diversos fenómenos biológicos devido a diversos fatores como a sua escala reduzida, isolamento geográfico, diferentes topologias e idades, entre outros (Babiker & Tautz 2015). O isolamento imposto por grandes massas de água constitui uma barreira natural à dispersão de várias espécies, sendo que a sua presença em ilhas permite o estudo de padrões evolutivos de uma forma menos complexa e com menos variáveis (nomeadamente bióticas) quando comparada com as que se encontram nos continentes. Assim, a dinâmica de colonização das ilhas é de grande interesse em biologia evolutiva, não só pela oportunidade do estudo das radiações adaptativas mas também pela possibilidade de seguir os padrões de colonização de espécies invasoras (Hardouin *et al.* 2010). Às espécies invasoras é reconhecido o profundo impacto que podem ter na fauna e flora nativas das ilhas colonizadas por interferirem com a sua dinâmica natural, podendo mesmo levar à extinção de algumas espécies (Suarez & Tsutsui 2008).

Ratinho-caseiro, *Mus musculus*

De entre as espécies de mamíferos invasores mais comuns destaca-se o ratinho-caseiro, *Mus musculus*, em grande parte devido à sua associação comensal com os humanos. Esta espécie subdivide-se em 3 sub-espécies: *Mus musculus musculus*, *M. m. castaneus* e *M. m. domesticus*, sendo esta última a mais bem-sucedida e de distribuição mais alargada (Figura 1) (Gündüz *et al.* 2001; Searle *et al.* 2009a; Bonhomme *et al.* 2010; Gabriel *et al.* 2010; Jones *et al.* 2011b; Gabriel *et al.* 2013; Jones & Searle 2015). A relação comensal estabelecida entre *M. m. domesticus* e o Homem remonta ao período Neolítico (há aproximadamente 15 000 anos, Weissbrod *et al.* 2017), com o início da sedentarização das populações humanas e o desenvolvimento da agricultura na região do antigo Crescente Fértil no Médio Oriente (Cucchi *et al.* 2005). A sub-espécie *M. m. domesticus* iniciou a invasão da Europa Ocidental há cerca de três mil anos, transportada passivamente pelo Homem (Gabriel *et al.* 2010). Esta estreita associação com o Homem proporcionou-lhe várias vantagens como uma disponibilidade contínua de alimento e abrigo (Searle *et al.* 2009a), destacando-se ainda uma capacidade de dispersão, que apesar de passiva, permitiu ultrapassar largamente a capacidade intrínseca de dispersão da espécie. Sendo um mamífero terrestre com reduzida capacidade natatória, os mares e oceanos constituem uma barreira geográfica intransponível, tornando a maioria das ilhas inacessíveis (Searle *et al.* 2009a). Porém, com o início das expansões marítimas dos povos da Europa Ocidental no século XV, o transporte involuntário deste pequeno roedor pelo Homem resultou numa expansão da espécie, sem precedentes à escala mundial (Hardouin *et al.* 2010) (Figura 1), tendo-se tornado uma das espécies de mamíferos mais ubíquas (Gündüz *et al.* 2001; Searle *et al.* 2009a; Gabriel *et al.* 2010; Jones *et al.* 2011a). O sucesso da espécie *M. musculus* havia já sido referenciado por Darwin (1869) como um exemplo de adaptabilidade a ambiente extremos. Esta espécie não só é extraordinária pela quantidade de ilhas que alcançou e colonizou com sucesso mas também pela grande diversidade de condições ambientais que pode tolerar (Gabriel *et al.* 2013), sendo que alguns exemplos estão descritos em Hardouin *et al.* 2010; Jones *et al.* 2011a.

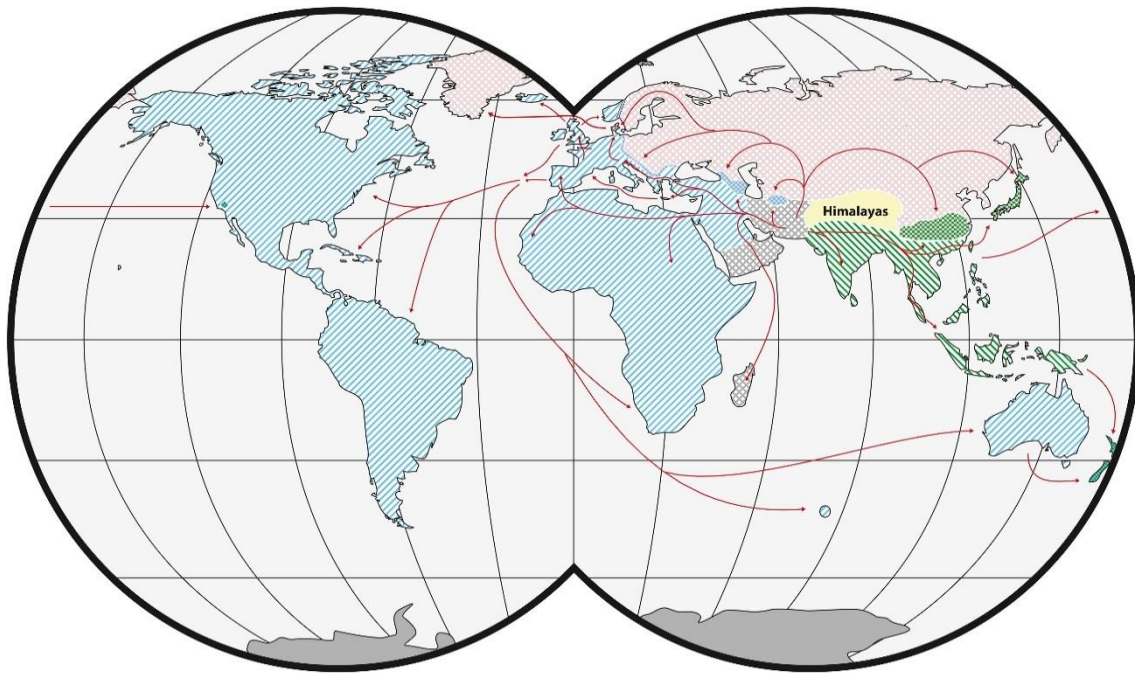


Figura 1 - Distribuição das sub-espécies de *Mus musculus* no Mundo: *Mus musculus domesticus* – azul; *Mus musculus castaneus* – verde; *Mus musculus musculus* – vermelho. As zonas axadrezadas representam zonas de hibridação. Setas a vermelho – representam as rotas históricas de migrações e recentes movimentações em associação com o Homem. Retirado de Phifer-Rixey & Nachman 2015.

A história da dispersão e colonização das espécies tem sido largamente estudada através da filogenética com recurso a marcadores moleculares num contexto geográfico – o que levou à criação da disciplina ‘Filogeografia’ em 1987 por Avise. Esta disciplina estuda os princípios e processos da distribuição geográfica das linhagens moleculares, principalmente num contexto específico. A análise e interpretação da distribuição das linhagens requer uma grande contribuição de dados moleculares, populacionais, demográficos, entre outros (Avise 2000). A abordagem filogeográfica implica a comparação de sequências de determinado(s) marcador(es) moleculares de indivíduos de populações com origens geográficas distintas e as suas potenciais fontes colonizadoras (Searle 2008). Ao longo das últimas duas décadas, esta abordagem filogeográfica tem sido utilizada com bastante sucesso para traçar a história da colonização da sub-espécie de rato-doméstico na Europa Ocidental, Austrália, Nova Zelândia e em várias ilhas oceânicas mundiais. A maioria dos estudos nesta área tem privilegiado como marcador molecular o D-loop do DNA mitocondrial (mtDNA) (Gündüz *et al.* 2001; Förster *et al.* 2009a; Searle *et al.* 2009a; Bonhomme *et al.* 2010; Hardouin *et al.* 2010; Jones *et al.* 2010, 2011a, b, 2012; Gabriel *et al.* 2011, 2015; Jones & Searle 2015; Lippens *et al.* 2017). Para além destes, os estudos recorrem também a microsatélites autossómicos (Hardouin *et al.* 2010; Jones *et al.* 2011b; Gabriel *et al.* 2013) e menos frequentemente ao cromossoma Y (Hardouin *et al.* 2010; Jones & Searle 2015).

Marcadores moleculares

O mtDNA (DNA mitocondrial), mais propriamente a região controlo D-loop, tem sido apontado como um marcador bastante eficiente no registo do primeiro evento de colonização bem-sucedido (Gabriel *et al.* 2010). As suas características não-recombinantes e um baixo efetivo populacional aliado a uma elevada taxa mutacional potenciam a sua utilização em estudos de filogeografia, principalmente quando se pretende distinguir espécies próximas ou populações da

mesma espécie (Awise *et al.* 1987) num universo temporal recente (Figura 2). Este é também um marcador fácil de obter em laboratório, uma vez que é haplóide e existe em elevado número de cópias por célula, o que permite uma fácil amplificação e análise quando comparado com outros marcadores (Zink & Barrowclough 2008). Porém, a sua transmissão exclusivamente materna (Awise *et al.* 1987) que facilita a inferência do primeiro evento de colonização bem-sucedido impossibilita a deteção da contribuição paterna no *pool* genético, não permitindo a avaliação de situações de dispersão diferencial dos sexos.

Por outro lado, marcadores localizados em zonas não recombinantes do cromossoma Y permitem aceder à genealogia paterna (Jones & Searle 2015). Deste modo o número de marcadores polimórficos para o cromossoma Y tem vindo a aumentar, permitindo uma caracterização da influência de ambos os sexos numa colonização. Para além disso, o reduzido efetivo populacional do cromossoma Y, em relação aos cromossomas autossómicos, torna-o mais sensível à deriva genética permitindo identificar diferenças entre populações próximas com tempos de divergência recentes (Perez-Lezaun *et al.* 1999; Mesa *et al.* 2000; Tosi *et al.* 2003). Os microssatélites são uma escolha recorrente para análise de zonas não codificantes, principalmente devido aos elevados níveis de variabilidade. Estes marcadores consistem em unidades repetitivas de sequência de nucleótidos (1-6 nucleótidos) que evoluem através do ganho ou perda da unidade de repetição, podendo ser altamente polimórficos (Shamjana *et al.* 2015). Assim, as mutações não ocorrem por substituição de uma base individual, mas sim por falha da polimerase durante a replicação, adicionando ou removendo motivos de repetição, resultando em polimorfismos de tamanho (Ellegren 2004). Os microssatélites dinucleotídicos são geralmente os mais abundantes no genoma, e conseqüentemente, o principal alvo para análises de microssatélites (Li *et al.* 2002). Devido à sua elevada variabilidade e rápida evolução (Figura 2), os microssatélites revelam-se eficazes na identificação de mudanças recentes na população. Numa outra abordagem, os marcadores moleculares localizados no cromossoma X permitem uma análise de hereditariedade bi-parental. Por ser um marcador nuclear, o tamanho populacional efetivo é quatro vezes superior ao do mtDNA, a taxa mutacional é mais lenta e, conseqüentemente, a variabilidade genética é mais baixa. Assim, este marcador consegue identificar diferenças que ocorreram num passado menos recente que o identificado pelo mtDNA (Zink & Barrowclough 2008). O gene *Ocrl* (*Inositol Polyphosphate-5-Phosphatase*) codifica uma enzima que modifica os fosfolípidos da membrana celular. Este gene possui 4 intrões e está identificado como sendo um gene monofilético para as diferentes sub-espécies de *Mus musculus* (Gerald *et al.* 2008). Tal facto, aliado às suas características de marcador nuclear, com uma maior taxa de acumulação de mutações nos intrões quando comparado com cromossomas autossómicos (Gerald *et al.* 2008), sugerem-no como um bom marcador no estudo da filogeografia desta espécie.

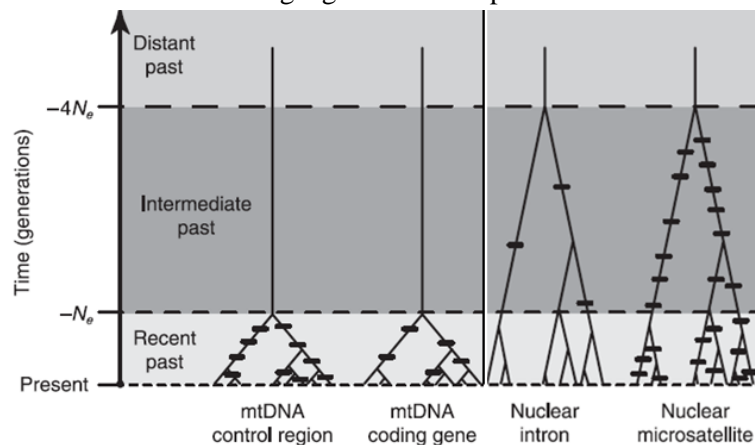


Figura 2 – Eficácia relativa dos marcadores moleculares: taxa mutacional vs. tempos de coalescência (adaptado de Zink & Barrowclough, 2008).

Colonização em ilhas

Durante as últimas décadas, têm sido realizados vários estudos ao longo da área de distribuição de *M. m. domesticus*, particularmente na parte continental da Europa ocidental. Estes estudos ganham ênfase graças à associação desta espécie com a história das colonizações e movimentações humanas, que tiveram como consequência indireta e involuntária a dispersão deste pequeno invasor para as ilhas atlânticas. O estudo da colonização da espécie introduzida (o ratinho-caseiro, neste caso) pode corroborar a hipótese da origem mais provável para a colonização, prevista a partir de dados históricos ou, contrariamente, levantar novas questões sobre a história das movimentações humanas, assentando assim na premissa de que a história da colonização desta espécie reflete as migrações/movimentações humanas para as ilhas (Searle *et al.* 2009a). Porém, apesar de muitas vezes os resultados das análises filogeográficas do ratinho-caseiro serem congruentes com o esperado historicamente, como é o caso da colonização do continente Australiano pelos Britânicos (Gabriel *et al.* 2011) ou a colonização das ilhas do Atlântico Norte (Jones *et al.* 2012), nem sempre essa concordância se verifica (ver abaixo).

Algumas das ilhas da região da Macaronésia (Figura 3A) foram já alvo de estudos filogeográficos com ratinho-caseiro, particularmente os arquipélagos dos Açores (Gabriel *et al.* 2013, 2015) e da Madeira (Gündüz *et al.* 2001; Förster *et al.* 2009), com dados obtidos para todas as ilhas. O arquipélago das Canárias foi alvo de apenas um estudo focado, contudo, num pequeno número de ilhas (Bonhomme *et al.* 2010).

O arquipélago dos Açores (Figura 3B) foi alvo de dois estudos dos quais se concluiu que este é um arquipélago com elevada diversidade genética quando considerado como um todo (Gabriel *et al.* 2015). Na maioria das ilhas foi identificada a esperada ligação histórica com Portugal continental, existindo, no entanto, ilhas que apresentaram uma associação ao arquipélago da Madeira e outras uma ligação mais inesperada ao Norte da Europa (Gabriel *et al.* 2015). Apesar da elevada diversidade no arquipélago, a diversidade em cada ilha é baixa. Este cenário é consistente com a hipótese de múltiplas origens de colonização em múltiplas ocasiões (Gabriel *et al.* 2015). No que diz respeito ao arquipélago da Madeira (Figura 3C), o cenário é bastante diferente do observado no arquipélago dos Açores. A associação com Portugal continental ao nível do mtDNA é praticamente inexistente, com apenas um indivíduo amostrado (num total de 112) que representa esta ligação. As sequências de D-loop detetadas na ilha da Madeira mostram uma forte identidade com as sequências encontradas em países do Norte da Europa (Noruega, Suécia, Filândia, Dinamarca e Alemanha), contrariamente ao que se verifica em Portugal continental (Gündüz *et al.* 2001; Förster *et al.* 2009). Esta associação foi altamente inesperada uma vez que, segundo todos os registos históricos disponíveis, o povo português foi o oficial descobridor e primeiro colonizador da ilha da Madeira em 1419 (Betthencourt & Chaudhuri 1998). O único dado histórico que apontava para um potencial conhecimento do arquipélago, anterior à descoberta portuguesa, prende-se com a existência de um mapa datado de 1351 onde surgiam ilhas com uma localização e nomenclatura coerentes com as ilhas do arquipélago da Madeira (Kelley 1979.). Deste modo, dada a identidade genética ao nível do mtDNA com as populações de ratinhos do Norte da Europa, foi levantada a hipótese de que poderão ter sido os Vikings os responsáveis pela introdução do ratinho-caseiro na Madeira (Gündüz 2001; Förster *et al.* 2009). Esta hipótese revelou-se controversa, porém, a adição de dados paleontológicos reforçaram as mesmas conclusões. Em 2011, foi descoberto um esqueleto parcial de *Mus musculus domesticus* na Ponta de São Lourenço, na ilha da Madeira, estimado ser de idade anterior à chegada dos portugueses à ilha em quatrocentos anos (Rando *et al.* 2014). Este dado paleontológico veio reforçar a possibilidade de que um outro povo anterior aos portugueses tivesse navegado/naufragado até ao arquipélago da Madeira, proporcionando a colonização do mesmo pelo ratinho-caseiro.

Dos três arquipélagos da Macaronésia de colonização portuguesa, o arquipélago de Cabo Verde (Figura 3D) é o único que ainda não foi alvo de qualquer estudo sobre a história da colonização

do ratinho-caseiro, sendo uma das novidades deste projeto. As ilhas do arquipélago de Cabo Verde foram descobertas pelos portugueses, vários anos após a descoberta e povoamento das ilhas da Madeira e dos Açores, entre 1460 e 1462, em duas fases. Inicialmente foram descobertas as ilhas de sotavento (Maio, Santiago e Fogo) e as ilhas de barlavento (Sal e Boa Vista), mais tarde as restantes ilhas de barlavento (Brava, Santo Antão, São Vicente, Santa Luzia e São Nicolau) (Oliveira Marques 1973). As ilhas foram povoadas maioritariamente por colonos portugueses (vindos de Portugal continental e das ilhas Açores e Madeira) e escravos africanos, mas recebendo também povos nórdicos, espanhóis, franceses e holandeses (Oliveira Marques 1973). Pela primeira vez, serão analisadas amostras de uma das ilhas do arquipélago de Cabo Verde, da parte ocidental, a ilha de Santa Luzia. Esta é uma ilha desabitada de pequenas dimensões (35 km²), situada entre a ilha de São Vicente e São Nicolau (Figura 3D). Pertencente ao município de São Vicente e já albergou alguns agricultores mas neste momento funciona apenas como porto piscatório.

Para analisar com maior detalhe a história da colonização do ratinho-caseiro (*Mus musculus domesticus*) nas ilhas Atlânticas historicamente colonizadas pelos Portugueses (Madeira, Açores e Cabo Verde), utilizaram-se marcadores moleculares de herança uniparental (mtDNA de herança exclusivamente materna e microssatélites do cromossoma Y de herança exclusivamente paterna), assim como um marcador nuclear de herança bi-parental (gene *Ocr1* do cromossoma X). Estes três marcadores foram escolhidos por apresentarem diferentes modos de transmissão e diferentes tamanhos efetivos populacionais, o que lhes confere diferentes padrões evolutivos. Estes padrões evolutivos, em conjunto com a ecologia do roedor em estudo, permitem, não só inferir o modo como as populações se estabelecem após a colonização de uma nova ilha, mas também determinar a contribuição de machos e fêmeas nessa mesma colonização.

Objetivos

Este projeto tem como principal objetivo contribuir com novas camadas de informação no estudo da colonização do ratinho-caseiro nos arquipélagos da Macaronésia de colonização histórica portuguesa, através de dois pontos principais:

Análise de marcadores moleculares de herança não exclusivamente materna localizados nos cromossomas X (intrão gene *Ocr1*) e Y (microssatélites) em populações de ratinho-caseiro dos arquipélagos da Madeira e dos Açores.

Iniciar o estudo de uma população de ratinho-caseiro do arquipélago de Cabo Verde, na Ilha de Santa Luzia, através da análise de marcadores do mtDNA (D-loop) e dos cromossomas X (gene *Ocr1*) e Y (microssatélites).

A análise conjunta destes três tipos de marcadores permitirá, pela primeira vez, uma avaliação mais abrangente e integrada da história da colonização das ilhas em estudo, através da avaliação da estrutura genética presente nas mesmas. Os dados obtidos poderão levar a uma melhor compreensão dos padrões de colonização do ratinho-caseiro nas ilhas destes 3 arquipélagos, nomeadamente, na contribuição relativa de machos e fêmeas para a constituição de populações insulares. Este conhecimento permitirá uma melhor perceção do fenómeno de colonização insular desta espécie invasora, com implicações no conhecimento dos processos ecológicos evolutivos associados.

Materiais e métodos

Amostragem

Um total de 259 amostras de tecido biológico foram utilizadas neste estudo. Essas amostras, originárias dos arquipélagos da Madeira (Gündüz *et al.* 2001; Förster *et al.* 2009), Açores (Gabriel *et al.* 2013, 2015) e Portugal continental (Förster *et al.* 2009), foram obtidas por investigadores em trabalhos anteriores e as amostras da ilha de Santa Luzia, Cabo Verde, foram enviadas por colaboradores. A Figura 3 A-D mostra o mapa dos arquipélagos que foram alvo de estudo, sendo que todas as ilhas dos arquipélagos da Madeira e dos Açores foram amostradas enquanto que no arquipélago de Cabo Verde apenas foi amostrada a ilha de Santa Luzia.

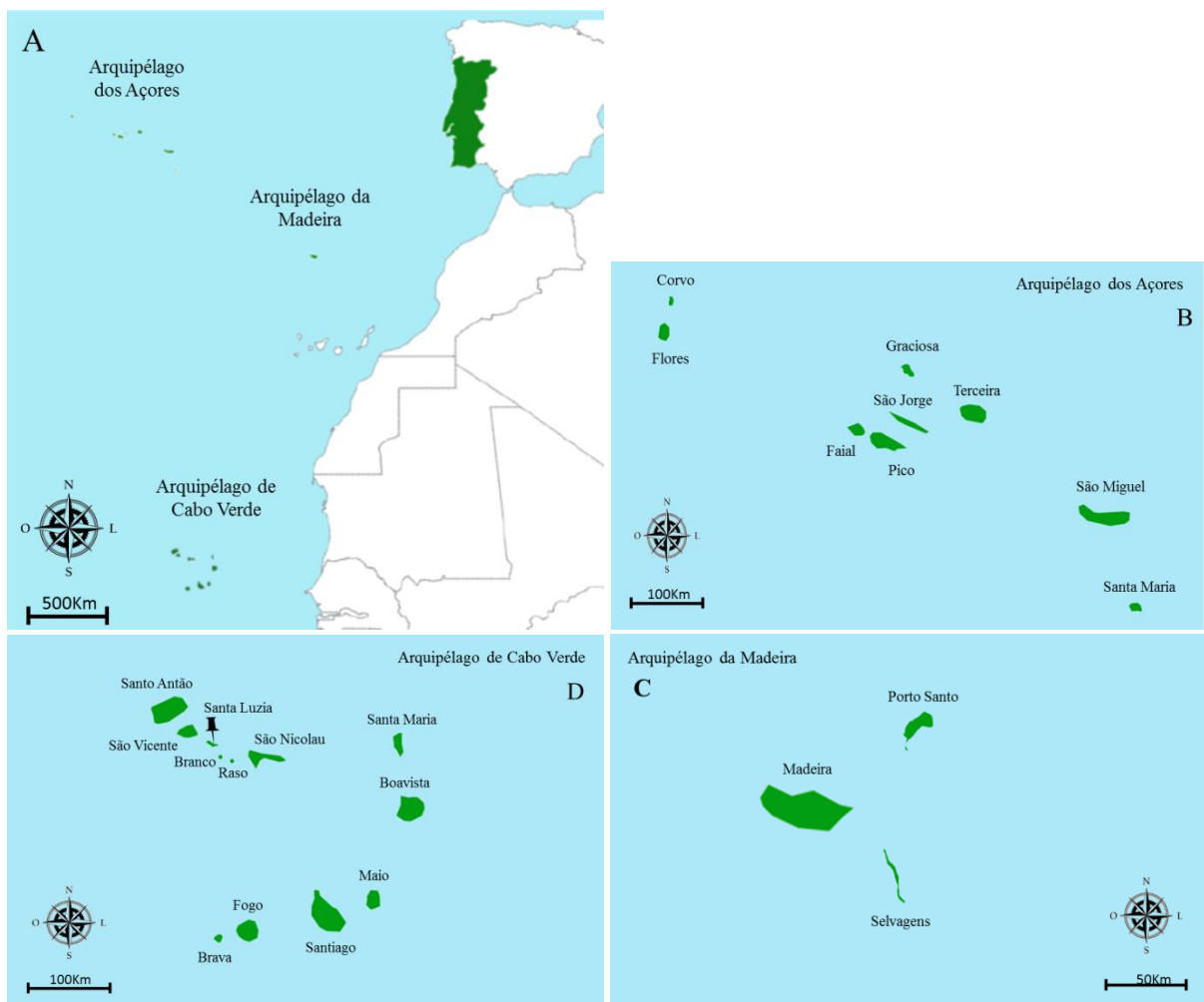


Figura 3 – A - Mapa da área geográfica onde foi realizada a amostragem de ratinho-caseiro para este estudo: Portugal continental, arquipélago da Madeira, arquipélago dos Açores e arquipélago de Cabo Verde. B – Mapa do arquipélago dos Açores; C – Mapa do arquipélago da Madeira; D – Mapa do arquipélago de Cabo verde, com indicação da ilha de Santa Luzia (única ilha amostrada).

Relativamente à análise de D-loop (ver abaixo), foram incorporados neste trabalho todos os dados já publicados para algumas dessas amostras. O número de amostras utilizadas para cada ilha e para cada marcador estão definidos na Quadro 1 (ver também Anexo I para mais detalhes).

Extração de DNA

Um total de 259 amostras biológicas (pontas de cauda, rim ou fígado) de *Mus musculus domesticus* originários de todas as ilhas dos arquipélagos dos Açores, Madeira, ilha de Santa Luzia (Cabo Verde), Portugal continental foram alvo de extração de DNA genómico (ver Anexo I para detalhes sobre a origem geográfica das amostras).

Estas amostras estavam conservadas em etanol absoluto a -20°C. A extração foi realizada usando o kit de extração ‘EZNA tissue DNA purification’ (Omega), seguindo as instruções do fabricante para o protocolo “Extração de tecidos”. De modo a verificar a qualidade do DNA, todas as amostras foram corridas em gel de agarose a 1% em TBE (tris-borato-EDTA), durante trinta minutos a 100 voltes.

Sexagem

Antes da análise dos marcadores em estudo foi necessário determinar o sexo dos indivíduos de Portugal Continental e das ilhas do Corvo e Santa Luzia, para os quais esta informação não estava disponível. Para tal, foi amplificado um fragmento no ‘*Dead box gene*’ que apresenta um fragmento de 180 pares de base (pb) no cromossoma X e um fragmento de 150 pb no cromossoma Y – *PCR-based gonosomal sexing method* (Christian Roos, DPZ Goettingen, dados não publicados). Assim, os machos apresentam duas bandas referentes a dois fragmentos, um de 150pb e outro de 180pb e as fêmeas apenas uma banda, referente aos dois fragmentos de 180pb.

Para a amplificação foram utilizados dois primers: *forward* e *reverse*, seguindo o protocolo de PCR que consiste em 15 minutos iniciais de desnaturação a 95°C, seguido de 35 ciclos de 60 segundos de desnaturação a 95°C, 60 segundos de *annealing* a 57°C e 60 segundos de extensão a 72°C, com um extensão final de 8 minutos a 72°C. Após o PCR, os produtos obtidos foram sujeitos a uma electroforese em gel de agarose 1%, durante 30 minutos a 100 voltes.

Quadro 1- Número de indivíduos amostrados em cada população (ilha/arquipélago/continente) para os vários marcadores analisados: D-loop (mtDNA), microssatélites (cromossoma Y) e gene *Ocrl* (cromossoma X). As sequências publicadas utilizadas estão assinaladas com asterisco (*).

	Arquipélago dos Açores									Arquipélago da Madeira		Arquipélago Cabo Verde	Portugal Continental
	Santa Maria	São Miguel	Terceira	Graciosa	São Jorge	Pico	Faial	Flores	Corvo	Madeira	Porto Santo	Santa Luzia	-
área (km ²)	97	747	402	61	246	448	173	142	17	801	43	35	92090
mtDNA D-loop	24*	52*	31*	27*	26*	23*	28*	23*	9*	142 (30+112*)	24 (14+10*)	29	76*
Cromossoma Y microssatélites	16	34	22	6	6	15	14	13	3	30	12	16	8
Cromossoma X gene <i>Ocrl</i> (intrão)	8	11	11	5	7	8	8	6	2	11	6	5	14

mtDNA – D-loop

Amplificação e Sequenciação

Um total de 72 amostras de indivíduos da sub-espécie *Mus musculus domesticus*, foram utilizadas para a análise deste marcador, sendo que 28 são provenientes da ilha de Santa Luzia (Cabo Verde), 30 da ilha da Madeira e 14 da ilha de Porto Santo.

Toda a sequência D-loop (entre as posições 15 424 e 16 276) foi amplificada através da técnica de PCR (*Polymerase Chain Reaction*) usando o par de *primers* L15774 - 5'TGA ATT GGA GGA CAA CCA GT3' e H2228 - 5'TTA TAA GGC CAG GAC CAA AC3', conforme o disposto em Searle *et al.* (2009a). A reação foi realizada em 25µl, contendo 12.5 µl de Master Mix (DNA polimerase (0.05U/µl), tampão de reação, 4mM MgCl₂ e 0.4mM de cada dNTP), 11 µl de H₂O, 0.25 µl de cada primer (*forward* e *reverse*) e 1 µl de DNA. O programa de PCR foi realizado num termociclador ARKTIK Thermal cycler da Thermo Scientific e consiste em 5 minutos iniciais de desnaturação a 95°C, seguido de 35 ciclos de desnaturação de 1 minuto, *annealing* a 57°C durante 1 minuto, extensão a 72°C durante 90 segundos e, por último, extensão a 72°C durante 8 minutos. Para confirmar a amplificação do fragmento pretendido, os produtos de PCR foram sujeitos a uma electroforese em gel de agarose a 1% durante trinta minutos a 100 volts. As amostras cuja amplificação foi bem sucedida foram posteriormente purificadas de modo a eliminar todos os metabolitos que não tenham o tamanho da sequência de interesse e possam interferir com a reação de sequenciação. A purificação foi levada a cabo com o sistema enzimático ExoProStar™ 1-Step, Enzymatic PCR and Sequence Reaction Clean-up, ILLUSTRATE™ (3 µl de mistura enzimática a cada 5 µl de produto de PCR) ou ExoSAP-IT (Affimatrix) (1 µl de exosap e 2 µl de fast AP por cada 5 µl de produto de PCR) conforme a disponibilidade em stock, para ambos foram seguidas as instruções do fabricante. O programa correu no mesmo termociclador da amplificação com 15 min a 37°C e 15min a 85°C para inativação das enzimas.

As amostras, depois de purificadas, foram enviadas para sequenciação (em ambos os sentidos) com os mesmos *primers* da amplificação, na empresa StabVida.

Análise filogenética

As sequências de cada indivíduo (*forward* e *reverse*) foram alinhadas com a sequência de referência do genoma mitocondrial de *Mus musculus domesticus* (Bibb *et al.* 1981). Seguidamente procedeu-se à análise dos cromatogramas das sequências de modo a identificar potenciais incongruências, sendo esta análise realizada no programa SEQUENCHER 4.1.4. (Sequencer, GeneCodes Corporation, EUA - <http://www.genecodes.com>). De seguida, todas as sequências *consensus* foram alinhadas manualmente, através do programa BIOEDIT v7.1.3.0 (Hall 1999), com todas as sequências disponíveis à data da análise, e disponíveis na base de dados pública, GenBank. Duas sequências pertencentes a indivíduos da ilha de Santo Antão e São Nicolau (dados não publicados) foram também inseridas na análise. Posteriormente, todas as sequências individuais foram colapsadas em haplótipos únicos através do *software* DNAsp (Librado & Rozas 2009), de modo a verificar o surgimento de novos haplótipos.

Para a construção da árvore filogenética foi realizada uma análise Bayesiana através do programa MR.BAYES v3.2.6 (Ronquist & Huelsenbeck 2003) com recurso ao método de *Monte Carlo Markov Chain* (MCMC) para 100 milhões de gerações com amostragem a cada 10 000 iterações. Como modelo evolutivo foi seleccionado o modelo Tamura e Nei (Tamura & Nei 1993) - TN93 + I + G conforme o disposto pelo *software* JMODELTEST1.0 (Posada 2008). Nesta abordagem foram utilizadas como *outgroup* duas sequências das sub-espécies *Mus musculus musculus* e *Mus musculus castaneus* (Gündüz *et al.* 2001).

Análise da Estrutura Populacional e Demográfica

A informação haplotípica das ilhas e arquipélagos em estudo, juntamente com as potenciais fontes colonizadoras, Portugal continental e Espanha, permitiram a construção de uma *median-joining network* (Bandelt *et al.* 1999). Esta permite avaliar as relações evolutivas entre os indivíduos, ou seja, o modo como a variabilidade genética se encontra distribuída, e seguir a sequência de eventos que está na origem do padrão genético observado (Avise 2000). Para além disso, permite ainda saber a associação entre as ilhas e as fontes colonizadoras. Esta análise foi feita com recurso ao programa NETWORK 5.0.0 (Fluxus-engineering, www.fluxus-engineering.com).

A diversidade genética haplotípica (H_d) e nucleotídica (π) e o número de sítios variáveis (S) foram, também, estimados. Para verificar se a maior parte da variação da subespécie em estudo se concentrava entre os arquipélagos ou entre as ilhas ou dentro destas foi realizada uma análise molecular da variância (AMOVA). Por outro lado, para avaliar se existe estrutura genética para este marcador, foram implementadas estatísticas F (F_{ST}). A AMOVA e as estatísticas F baseiam-se na frequência genética e na distância molecular entre haplótipos. Estes testes foram realizados no programa ARLEQUIN v3.5.2.2 (Excoffier & Lischer 2010) agrupando-se as ilhas por arquipélagos. Para verificar a existência de possíveis eventos de expansão ou seleção foram realizados testes de desvio à neutralidade, como D de Tajima (Tajima 1989) e F_s de Fu (Fu 1997), com um nível de significância de 0.05, no programa ARLEQUIN v3.5.2.2.

Microssatélites (cromossoma Y)

Para o cromossoma Y foram utilizadas amostras de 195 machos provenientes de todas as ilhas do arquipélago dos Açores e Madeira, ilha de Santa Luzia (Cabo Verde) e Portugal continental, segundo o Quadro 1 (ver Anexo I para mais detalhes).

Baseado nos trabalhos de Hardouin *et al.* (2010) e Jones & Searle (2015), foram selecionados 5 microssatélites di-nucleotídicos (de 6 microssatélites disponíveis), localizados no cromossoma Y e designados por Y6, Y12, Y21, Y23 e Y24. O locus Y22 foi excluído *a priori* por apresentar problemas de amplificação e de *scoring* segundo a autora da publicação Hardouin *et al.* 2010 (E. Hardouin, comunicação pessoal).

Amplificação e Genotipagem

Os 5 microssatélites selecionados foram amplificados por PCR com os *primers* descritos no Anexo II. Os *primers forward* dos microssatélites Y6, Y12 e Y23 apresentam um fluorocromo FAM associado enquanto os *primers forward* dos loci Y21 e Y24 apresentam um fluorocromo HEX. Os *primers reverse* não apresentam qualquer marcação.

De modo a otimizar a sua amplificação, tendo em atenção o fluorocromo FAM ou HEX e o tamanho médio conhecido dos alelos de cada microssatélite obtidos em trabalhos anteriores (Hardouin *et al.* 2010; Jones & Searle 2015) foram testadas várias conjugações de microssatélites em *multiplex*, tendo por fim obtido os *multiplex* - A (Y12 e Y24) e B (Y21 e Y23). O microssatélite Y6 acabou por ser amplificado isoladamente por não produzir resultados satisfatórios quando conjugado com qualquer um dos outros *multiplex*.

A reação de PCR foi realizada para um volume total de 15 μ l contendo 7.5 μ l de Master Mix (Taq DNA *polymerase* (0.05U/ μ l), tampão de reação, 4mM $MgCl_2$ e 0.4mM de cada dNTP, nuclease), 1.38 μ l de H_2O , 0.03 μ l de cada *primer* e 6 μ l de DNA. O programa de PCR consistiu em 15 minutos iniciais de desnaturação a 95°C, seguido de 28 ciclos de 30 segundos de desnaturação a 95°C, 60 segundos de *annealing* a 60°C e 90 segundos de extensão a 72°C, com uma extensão final de 10 minutos a 72°C. Para o microssatélite Y6, a reação de PCR foi realizada para o mesmo volume de 15

μl contendo 7.5 μl de Master Mix, 1.44 μl de H₂O, 0.03 μl de cada *primer* e 6μl de DNA. O programa de PCR utilizado foi o mesmo utilizado nos dois *multiplex*.

Os perfis alélicos dos microssatélites foram verificados através do *software* PEAK SCANNER v2.0 (Applied Biosystems, EUA), para verificação do tamanho e forma dos picos (Figura 4A). Este *software* calcula o tamanho dos alelos por comparação com uma escada interna, GS500(-250), de fragmentos de tamanhos conhecidos. De modo a corrigir o tamanho dos alelos foi utilizado o programa TANDEM (Matschiner *et al.* 2009) que aplica um modelo de minimização de erros de arredondamento (Figura 4B). Para cada microssatélite foram repetidas 5% das amostras como controlo de qualidade. Após a correção do tamanho dos alelos no programa Tandem, os valores dos alelos corrigidos foram utilizados para todos os cálculos subsequentes (Anexo III).

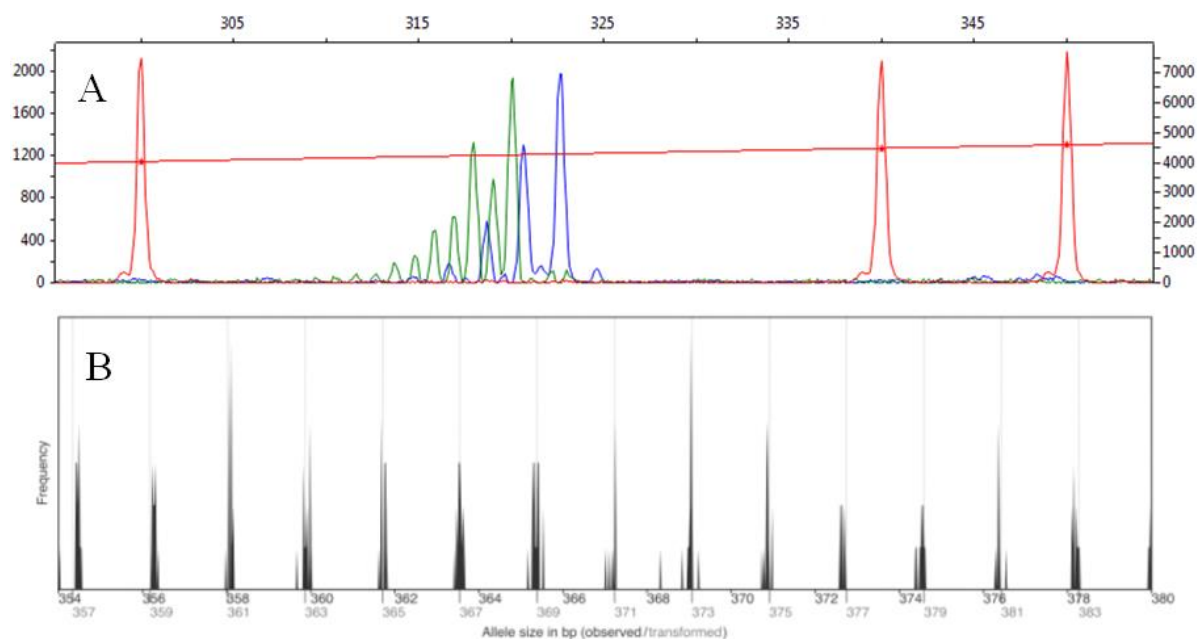


Figura 4 – Análise de picos de microssatélites. A – Verificação da qualidade e altura dos picos de microssatélites no Programa PEAK SCANNER. A amostra representada ilustra o resultado da amplificação de dois microssatélites do cromossoma Y (Y21, a verde – fluorocromo HEX; e Y23, a azul – fluorocromo FAM). A vermelho encontra-se a escada interna GS500(-250), com marcação ROX. B – Correção dos alelos dos microssatélites obtidos através do programa TANDEM. As barras pretas representam a frequências dos tamanhos de alelos obtidos e as barras cinzentas indicam os tamanhos dos alelos corrigidos.

Análise de diversidade e estrutura populacional

As estatísticas de diversidade genética *standard* (diversidade haplotípica e alélica, número de alelos privados) bem como a média de alelos por locus e alelos raros foram calculados no programa FSTAT (Goudet 2001). A determinação do conjunto de haplótipos foram avaliados no programa ARLEQUIN e depois realizada uma rede de haplótipos com os mesmos, através do programa EDENETWORK v 2.18 (Kivelä *et al.* 2014).

De modo a inferir o grau de diferenciação entre as populações foram realizadas comparações par a par, por estatísticas F (F_{ST}). Também foi realizada uma AMOVA para verificar como se encontrava distribuída a variação nas diferentes ilhas. Para estes testes foram eliminadas as amostras para as quais não havia sido possível obter dados para os 5 *loci* de modo a evitar que a distância entre as populações fosse subestimada. Ambos os testes foram desenvolvidos através do programa ARLEQUIN.

A estrutura genética das populações amostradas nas ilhas do arquipélago da Madeira e dos Açores (com a exceção da ilha do Corvo) e na ilha de Santa Luzia foi inferida através do programa

STRUCTURE (Pritchard *et al.* 2000; Falush *et al.* 2003, 2007; Hubisz *et al.* 2009) para K de 3 a 12, numa corrida com 500 000 cadeias de Markov (MCMC) e um *burn in* de 10% programado com “*no admixture model*” e “*sampling location as prior*”. Foi ainda realizada uma Análise de Coordenadas Principais (ACoP) englobando os dados de todos os microssatélites, recorrendo à macro GENALEX no programa Microsoft Excel 2010.

Ocrl (cromossoma X)

Para o estudo do cromossoma X foi selecionado o gene *Ocrl* (região intrónica), por ser um gene monofilético para as subespécies de *Mus musculus* e por existirem alguns dados publicados para comparação, provenientes da área de distribuição da sub-espécie *Mus musculus domesticus* (Figura 5). O gene OCRL está localizado no cromossoma X e possui 4 intrões com um total de 2480 pb (Gerald *et al.* 2008).

Para este marcador foram analisadas 121 amostras (ver Quadro 1 e Anexo I para mais detalhes), incluindo amostras de todas as ilhas dos Açores, Madeira, Porto Santo e Santa Luzia (Cabo Verde).

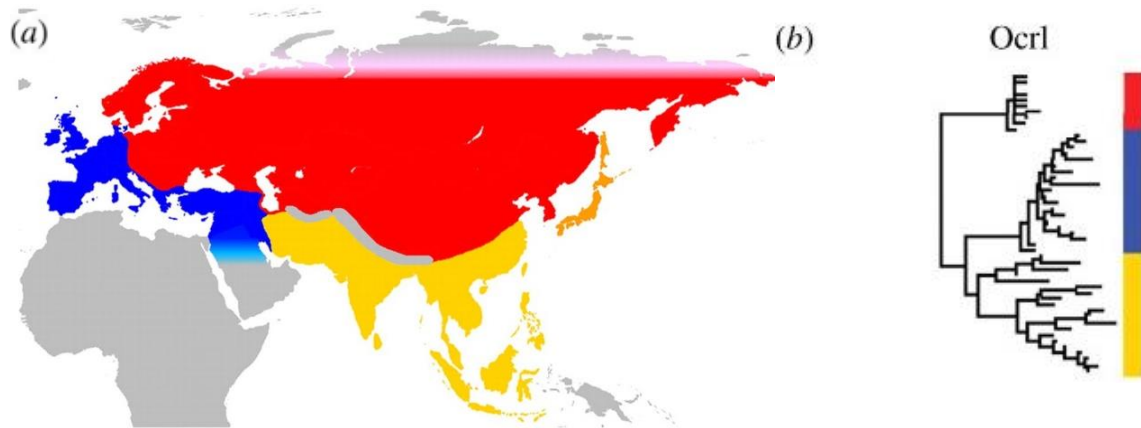


Figura 5 – A figura representa (a) a distribuição geográfica das sub-espécies de rato-caseiro: *Mus musculus domesticus* (azul), *Mus musculus musculus* (vermelho) and *Mus musculus castaneus* (amarelo) e (b) a genealogia do gene *Ocrl* para as mesmas sub-espécies (adaptado de Gerald *et al.* 2008).

Amplificação e Sequenciação

Para a amplificação deste marcador foi testado o protocolo de amplificação descrito no estudo de Gerald *et al.* (2008), mas sem sucesso. Ao analisar-se as sequências publicadas nesse estudo verificou-se que os 4 intrões do gene *Ocrl* apresentavam reduzida variabilidade pelo que foi selecionado um fragmento de 1083 pb compreendido entre o segundo e o terceiro intrão, incorporando alguns sítios polimórficos. Definida a região de interesse foram desenhados os seguintes *primers*: *forward* 5' TTG ACC AAA TTG AAG GGT GA 3' e *reverse* 5' CTC TTC TCT TTC GGC ACA CA 3', com recurso ao programa Primer3 (Untergasser *et al.* 2007). Posteriormente, procedeu-se à amplificação deste fragmento através da técnica de PCR, numa reação de 25µl com 12.5µl de Master Mix (Taq DNA *polymerase* (0.05U/µl), tampão de reacção, 4mM MgCl₂ e 0.4mM de cada Dntp, nuclease), 10µl de H₂O, 0.25µl de cada primer (*forward* e *reverse*) e 2µl de DNA utilizando o seguinte programa no termociclador: 5 minutos iniciais de desnaturação a 95°C, seguido de 35 ciclos com 60 segundos de desnaturação a 95°C, 60 segundos de *annealing* a 56°C, e 90 segundos de extensão a 72°C, com uma extensão final de 8 minutos a 72°C. Após a amplificação, as amostras foram sujeitas a uma electroforese em gel de agarose a 1% durante 30 minutos a 100 voltes, para confirmação do

sucesso da amplificação e nesse caso purificadas com recurso ao kit ExoProStar™ 1-Step, Enzymatic PCR and Sequence Reaction Clean-up, ILLUSTRATE™ (3 µl de mistura enzimática a cada 5 µl de produto de PCR) ou ExoSAP-IT (Affimetrix) (1 µl de exosap e 2 µl de fast AP por cada 5 µl de produto de PCR) conforme a disponibilidade em stock. Para ambos foram seguidas as instruções do fabricante. O programa no termociclador envolveu 15 minutos a 37°C e 15 minutos a 80°C para inativação das enzimas.

Os produtos de PCR purificados foram enviados para sequenciação (*forward* e *reverse*) na empresa STABVIDA com os mesmos *primers* da amplificação.

Análise filogenética

As sequências obtidas (*forward* e *reverse*) foram alinhadas com uma sequência já publicada, para verificação da ocorrência de ambiguidades nucleotídicas, tal como realizado para o marcador D-loop do DNA mitocondrial. Para isso procedeu-se à análise dos cromatogramas das sequências com recurso ao programa SEQUENCHER 4.1.4. Por se tratar de um marcador diplóide, as sequências foram inspecionadas visualmente base a base de forma a identificar os indivíduos heterozigóticos. Às sequências dos indivíduos heterozigóticos foi realizado um *phase* manual por envolver apenas uma única posição heterozigótica. Todas as sequências *consensus* obtidas foram alinhadas manualmente, através do programa BIOEDIT v7.1.3.0, com todas as sequências publicadas da sub-espécie *Mus musculus domesticus*, do estudo de Geraldine *et al.* (2008), provenientes da Grécia, Espanha, Israel, Itália e Reino Unido. Duas sequências de *M. m. musculus* e *M. m. castaneus* publicadas no mesmo estudo foram utilizadas como *outgroup*. Todas as sequências obtidas depois de alinhadas foram colapsadas em haplótipos através do *software* DNASP.

A árvore filogenética foi construída através de uma abordagem Bayesiana recorrendo ao programa Mr. Bayes através do método de *Monte Carlo Markov Chain* (MCMC) para 10 milhões de gerações com amostragem a cada 1000 iterações. O modelo evolutivo que melhor explica a variação nucleotídica das sequências obtidas foi determinado no programa JMODELTEST 1.0 (Posada 2003).

Análise da estrutura populacional e demográfica

Através da informação haplotípica de todas as ilhas em estudo, juntamente com as amostras portuguesas e as amostras já publicadas provenientes da Espanha, Grécia, Itália, Israel e Reino Unido foi construída uma rede de haplótipos com recurso ao programa NETWORK 5.0.0. Esta permite avaliar se existe estrutura populacional na Macaronésia e saber sua associação com as amostras do continente Europeu.

De modo a caracterizar os locais em estudo, foi estimada a diversidade haplotípica (H_d) e nucleotídica (π), assim como o número de sítios variáveis (S) para todas as amostras tratadas neste estudo e para as amostras já publicadas. De modo a verificar como se encontra distribuída a diversidade para este marcador nas ilhas foi realizada uma análise molecular da variância (AMOVA), bem como estatísticas F (F_{ST}), avaliando o nível de diferenciação genética entre as populações amostradas. Ambos os testes foram realizados no programa ARLEQUIN v3.5.2.2, agrupando-se as ilhas por arquipélagos. De modo a tentar detetar eventuais eventos de expansão ou seleção foram realizados os testes de desvio à neutralidade D de Tajima e F_s de Fu, para um nível de significância de 0.05. Estes testes foram realizados no programa ARLEQUIN v3.5.2.2.

Resultados

mtDNA

Análise filogenética

A análise haplotípica dos setenta e dois novos indivíduos amostrados nas ilhas da Madeira e Porto Santo para o marcador molecular D-loop resultou em 26 haplótipos, dos quais 14 nunca tinham sido amostrados. Destes, onze foram detetados pela primeira vez na ilha da Madeira (Made.33-Made.42) e três da ilha de Porto Santo (PSanto5-PSanto7) (ver Anexo IV). As sequências redundantes pertencem na maioria a haplótipos já identificados na Madeira (Made.1/2/5/8/12/18/22/25) e Porto Santo (Psanto.2/4) (Förster *et al.* 2009, Gündüz *et al.* 2001), sendo amostrado um indivíduo pertencente ao haplótipo Portugal.1 e um indivíduo com o haplótipo U47461, anteriormente apenas detetado na Dinamarca. Na ilha de Santa Luzia, que foi pela primeira vez amostrada neste estudo, foram detetados três haplótipos, todos eles coincidentes com haplótipos presentes no arquipélago da Madeira (Made.1, PSanto2 e MADE.18). Os indivíduos da ilha de Santo Antão e São Nicolau originaram, cada um, um novo haplótipo. Todos os haplótipos podem ser consultados no anexo IV.

Para a análise filogenética o modelo evolutivo que mais se adequa aos dados é o de Tamura & Nei 93 + I + G. Segundo este modelo, foi realizada uma análise Bayesiana, da qual resultou a árvore filogenética representada na Figura 6. Nesta, formaram-se 6 clades monofiléticas e uma zona basal onde existem haplótipos que não apresentam uma estruturação em clade.

Os resultados filogenéticos obtidos são consistentes com estudos anteriores (Förster *et al.* 2009; Gabriel *et al.* 2015; Jones *et al.* 2011b; Jones & Searle 2015) onde foram descritas as clades B a F (ver Figura 6). Neste estudo existe uma exceção na Clade B (representada a roxo) quando comparada com a análise mais recente em Gabriel *et al.* 2015. Neste caso, a clade designada por B aparece dividida em duas clades, sendo que a maior parte dos haplótipos amostrados em Portugal Continental, nos Açores e o haplótipo único da ilha de Santo Antão se isolam num grupo e os haplótipos predominantes da Turquia, Itália, Espanha noutro grupo. O valor de suporte desta divisão é apenas 0.52, pelo que não se considera como uma nova clade. Os haplótipos integrados na clade C apresentam uma forte representatividade em Portugal continental e em algumas ilhas dos Açores. A clade D é composta por haplótipos que ocorrem mais a norte da Europa e nas ilhas da Madeira e dos Açores. Por último, a clade F é composta por haplótipos amostrados em larga escala no Norte da Europa mas com uma dispersão diferente da clade D. Na figura.6 pode ser consultada a amostragem de cada clade por país/ilhas, sendo que esta informação para a Europa Ocidental e ilhas atlânticas está também representada geograficamente na Figura.7.

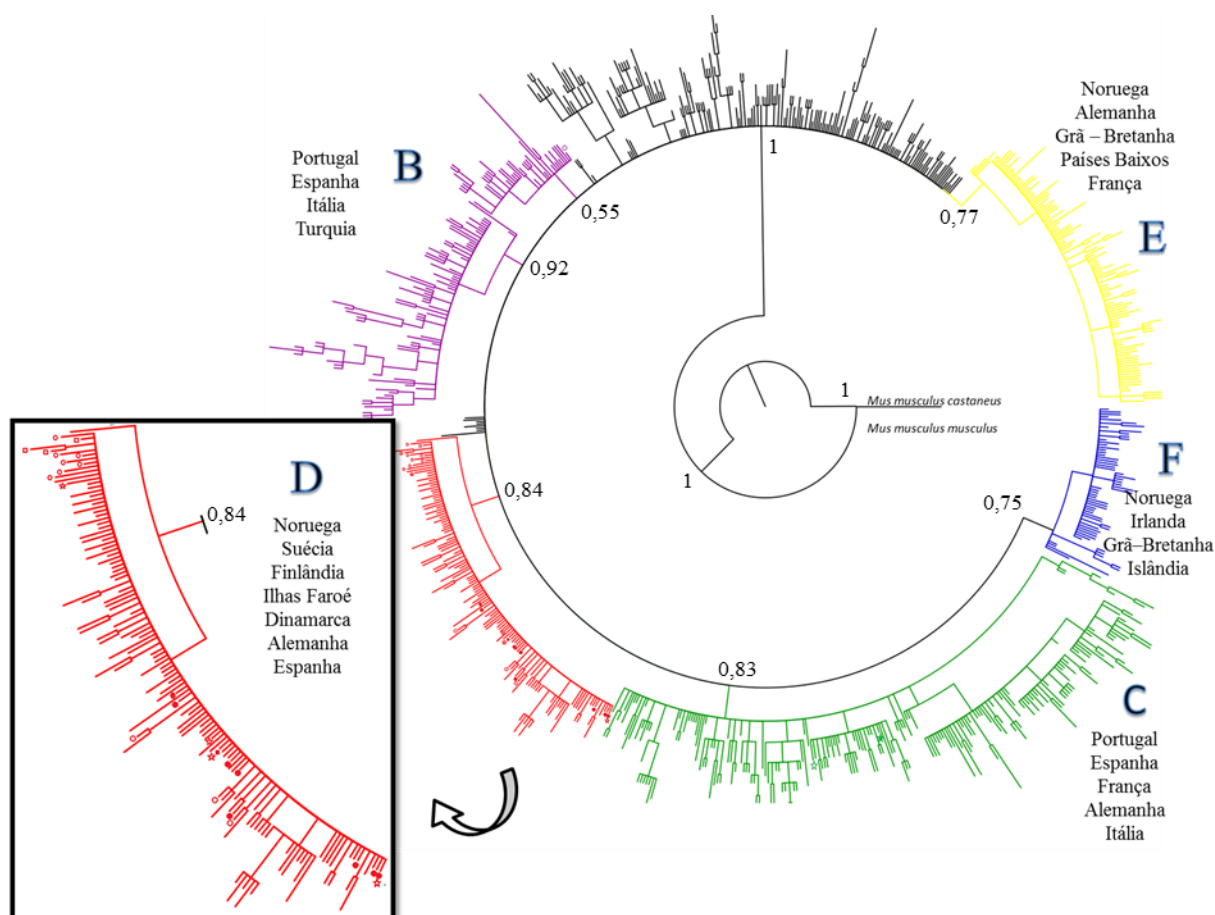


Figura 6 – Árvore Filogenética Bayesiana para o marcador D-loop para todas as sequências publicadas e não publicadas até à data da sub-espécie *Mus musculus domesticus*. Os haplótipos novos amostrados neste estudo encontram-se marcados com um símbolo aberto e os haplótipos já amostrados com símbolo fechados. A forma do símbolo representa onde foi amostrado: estrelas – Cabo Verde; círculos – Madeira; quadrados - Porto Santos.

O arquipélago dos Açores apresenta cinco das seis clades descritas em Jones *et al.* 2011a: clade B, C, D, E e F. Neste estudo a clade A não foi encontrada e a clade E resume-se a um único haplótipo, representado por um único indivíduo amostrado em São Miguel, ilha com maior diversidade contendo haplótipos de todas as clades. A clade B apresenta oito haplótipos estando presente nas ilhas de São Miguel, Santa Maria e Pico. A clade mais disseminada no arquipélago dos Açores é a clade C presente em seis das nove ilhas São Miguel, Graciosa, Faial, Flores, Pico e Corvo com dez haplótipos. A clade D está representada no arquipélago dos Açores com dez haplótipos e está presente nas ilhas de São Miguel, Terceira, Pico e São Jorge. Por último, a clade F está presente em quatro ilhas dos Açores (Flores, Faial, Santa Maria e Terceira) com nove haplótipos. Esta não está presente em Portugal Continental mas é predominante na ilha de Santa Maria, estando também bem representada na ilha da Terceira.

O arquipélago da Madeira está representado pelas ilhas da Madeira e de Porto Santo. Ambas as ilhas apresentam quase exclusivamente haplótipos da clade D, sendo que a ilha da Madeira é composta por 45 haplótipos (U47455, U47456, MADE.03-13, MADE.16-17, MADE.19, MADE.21-41, MADE.43, U47461, U47464, U47465, MADE.18, U47461) e a ilha de Porto Santo apresenta 7 haplótipos (PSanto.1-7). Na ilha da Madeira existem ainda 3 haplótipos pertencentes à clade C (MADE.20, MADE.42 e Portugal.1) representados por um indivíduo cada. Estes são os únicos que mostram associação direta com Portugal continental para o marcador D-loop. Os haplótipos MADE.33-MADE.42 e PSanto5-PSanto7 são haplótipos novos, pela primeira vez amostrados neste

estudo e com um N=1 com a exceção de MADE.34 que conta com dois indivíduos. O haplótipo U47461 até agora exclusivamente detetado na Dinamarca foi, também, pela primeira vez amostrado na Madeira.

No arquipélago de Cabo Verde apenas foi amostrada a ilha de Santa Luzia, porém existe uma amostra da ilha de São Nicolau e outra amostra de Santo Antão (não publicadas) que também foram analisadas. Assim, na ilha de Santa Luzia existem 3 haplótipos - MADE.1, PSanto2 e MADE.18, todos pertencentes à clade D. Na ilha de São Nicolau obtivemos o haplótipo SNicolau.1 pertencente a clade D e na ilha de Santo Antão o haplótipo StoAntao.1 pertencente ao Clade B. Assim, os haplótipos pertencente à clade D aproximam-se filogeneticamente das populações da ilha Madeira enquanto o haplótipo da clade C aproxima-se das populações de Portugal continental.



Figura 7 – Mapa das clades de D-loop de *M. m. domesticus* amostradas na Europa Ocidental e ilhas atlânticas. As clades estão representadas em círculos coloridos. clade B – roxo; clade C – verde; clade D – vermelho; clade E – amarelo e clade F – azul. Adaptado de Jones & Searle (2015).

Análise da estrutura populacional e demográfica

A relação entre haplótipos presentes em cada ilha e a sua frequência foi definida através de uma rede haplotípica, representada na Figura 8, com todos os indivíduos presentes nos três arquipélagos em estudo Açores, Madeira e Cabo Verde. Para além destes, foram ainda adicionados todos os indivíduos publicados de Portugal e Espanha como potenciais fontes colonizadora.

Podemos verificar a formação de quatro agrupamentos de haplótipos, que são representativos das clades referidas anteriormente. A clade D incorpora principalmente a Madeira e Cabo Verde, mas conta ainda com a presença de algumas ilhas dos Açores. A clade F incorpora apenas ilhas dos Açores enquanto a clade B incorpora indivíduos de Espanha, Portugal, Açores e um indivíduo de Cabo Verde

(ilha de Santo Antão). Por último a Clade C, informalmente designada de Clade Portuguesa, incorpora Portugal, Espanha, Açores e 3 indivíduos da ilha da Madeira. A clade E é representada por um único indivíduo referente a São Miguel e que se encontra isolado.

Os novos haplótipos que surgiram no arquipélago da Madeira associam-se maioritariamente com a clade D, existindo apenas a exceção de um haplótipo que surge na Clade C, o haplótipo MADE.42, que aparece no grupo com haplótipos dos Açores, Portugal e Espanha, próximo do haplótipo MADE.20, até então o único haplótipo encontrado na Madeira com associação a Portugal continental. Um outro indivíduo amostrado na ilha da Madeira neste estudo apresenta o haplótipo Portugal.1 que é dos mais representativos dos Açores, aparecendo também no Pico, Flores, Faial e Corvo juntamente com Portugal, sendo assim o terceiro indivíduo da ilha da Madeira entre 142 amostrados a ter associação com Portugal continental para o marcador de mtDNA – D-loop. Importa salientar que o haplótipo MADE.42 dista em apenas uma mutação do haplótipo Portugal.1.

A Macaronésia apresenta-se assim com 92 haplótipos distribuídos pelos três arquipélagos em estudo. No arquipélago dos Açores contabilizam-se 41 haplótipos pertencentes 5 clades B,C,D,E e F com amostragem nas 9 ilhas, existindo haplótipos que ocorrem em mais que uma ilha. A estruturação genética é evidente pela distribuição das clades pelas ilhas. Três ilhas apresentam apenas uma clade, como é o caso de São Jorge que apresenta 3 haplótipos, o haplótipo PSanto.2, dominante, e dois haplótipos que divergem deste numa mutação, pertencendo todos ao clade D. As ilhas de Graciosa e Corvo também apresentam apenas uma clade, a C, contando cada uma com 5 e 2 haplótipos respectivamente. Com duas clades temos a ilha das Flores, que conta com três haplótipos da clade F e um haplótipo Portugal.1 (dominante na ilha) pertencente à clade C. Para além dessa ilha, a ilha de Santa Maria também apresenta duas clades com um haplótipo da clade B e 5 haplótipos da clade F e por último a Terceira que apresenta 5 haplótipos da clade F e 5 haplótipos da clade D.

No arquipélago de Cabo Verde temos cinco haplótipos, três deles referentes à ilha de Santa Luzia, um referente à ilha de São Nicolau e o último da ilha de Santo Antão. Todos os haplótipos apresentam associação com o arquipélago da Madeira, clade D, com a exceção do haplótipo de Santo Antão que aparece na classe B. Os haplótipos que surgiram em Santa Luzia são exactamente iguais a haplótipos encontrados na Madeira, sendo que a maioria dos indivíduos (N=17) caem no haplótipo dominante deste arquipélago MADE.1. Para além disso, verifica-se ainda que a segunda maior percentagem de indivíduos (N=7) pertence ao haplótipo PSanto.2, o haplótipo dominante da ilha de Porto Santo. Os restantes indivíduos amostrados (N=4) apresentam o haplótipo MADE.18. Estes três haplótipos são bastante próximos entre si, sendo que o mais comum em Santa Luzia, MADE.1 difere em apenas uma mutação de Psanto.2 e de duas mutações de MADE.18.

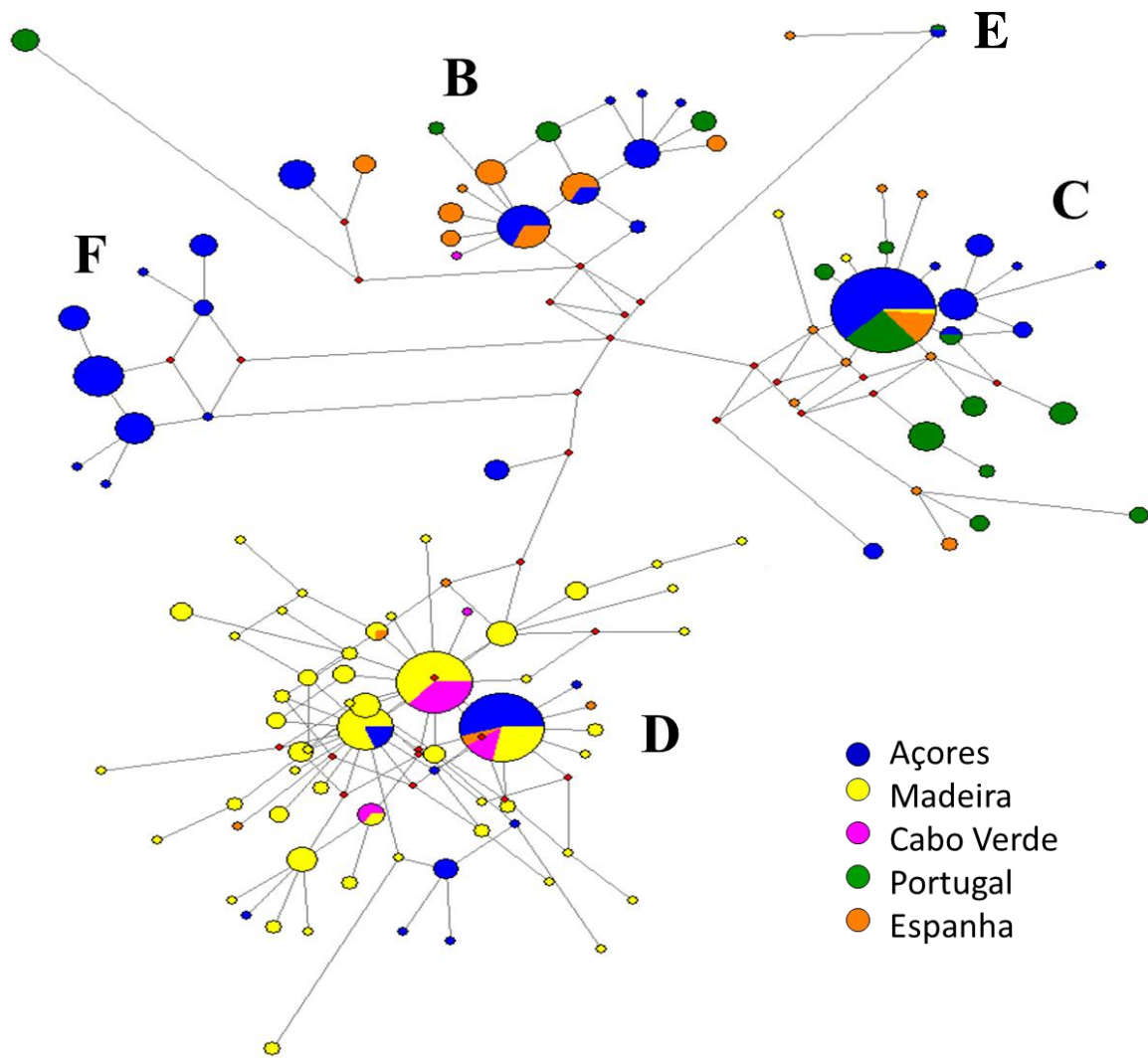


Figura 8 - Rede de haplótipos *Neighbor-Joining* para sequências D-loop de *Mus musculus domesticus* de Portugal, Espanha, Açores, Madeira e Cabo Verde. Cada círculo representa um haplótipo único e o seu tamanho representa a frequência de ocorrência. O comprimento das linhas que ligam os haplótipos é proporcional ao número de mutações entre os mesmos e os pontos vermelhos são haplótipos intermédios hipotéticos ou não amostrados.

A diversidade haplotípica e nucleotídica, de cada ilha, para o marcador D-loop encontra-se sumarizada na Quadro 2.

Os valores da diversidade haplotípica são bastante contrastantes entre as ilhas em estudo. A ilha que apresenta maior diversidade é a ilha da Madeira ($H_d=0.928$) seguindo-se as ilhas de São Miguel e Terceira ($H_d=0.865$ e $H_d=0.822$, respetivamente). Estes valores são elevados em relação ao esperado em ilhas, já que se aproximam dos valores encontrados no continente, como é o caso de Portugal, com uma diversidade haplotípica de 0.8828, e de Espanha, com uma diversidade haplotípica de 0.933. Por outro lado, existem ilhas com uma diversidade haplotípica muito reduzida como é o caso de São Jorge ($H_d=0.151$) e Corvo ($H_d=0.389$). A ilha de Santa Luzia apresenta, apesar das suas dimensões reduzidas, uma diversidade haplotípica intermédia ($H_d=0.569$). Quanto à diversidade nucleotídica, as ilhas de São Jorge e Corvo apresentam os valores mais baixos ($\pi=0.0002$ e $\pi=0.0005$, respetivamente), sendo a ilha do Pico e Terceira que apresentam os valores mais elevados ($\pi=0.0124$ e $\pi=0.0129$, respetivamente). A ilha de São Miguel apresenta uma diversidade intermédia em relação às populações em estudo ($\pi=0.0075$) enquanto a ilha de Santa Luzia apresenta a terceira diversidade nucleotídica mais baixa ($\pi=0.0010$).

Quadro 2 - Variabilidade genética do marcador D-loop (mtDNA) para cada população de *Mus musculus domesticus* em análise. O número de haplótipos, número de sítios variáveis, diversidade haplotípica (Hd) e diversidade nucleotídica (π), bem como os testes de neutralidade D de Tajima e FS de Fu foram determinados no programa Arlequin v 3.5.2.2. Os valores de P, quando significativos, dos testes D de Tajima e FS de Fu não estão assinalados com um asterisco (*) ($P < 0.05$).

População	N	N Haplótipos	Número de Sítios Variáveis	Hd	π	D de Tajima	FS de Fu
Santa Maria	24	3	12	0.754±0.048	0.0032±0.0019	-1.47	1.927
São Miguel	52	10	26	0.865±0.024	0.0075±0.004	-0.708	1.247
Terceira	31	6	16	0.822±0.047	0.0129±0.0067	1.474	4.51
Graciosa	23	4	3	0.668±0.079	0.0011±0.0009	-0.268	-1.207
São Jorge	26	3	1	0.151±0.093	0.0002±0.0003	-1.156	-2.176*
Pico	23	4	14	0.557±0.083	0.0124±0.0065	1.277	12.177
Faial	28	6	17	0.521±0.102	0.0090±0.0048	-0.246	6.653
Flores	23	3	10	0.589±0.094	0.0064±0.0035	2.649	6.678
Corvo	9	2	1	0.389±0.164	0.0005±0.0005	0.156	0.477
Madeira	142	45	34	0.928±0.013	0.0062±0.0034	-2.124*	-23.123*
Porto Santo	24	7	5	0.5368±0.123	0.0012±0.0009	-1.457	-1.937
Santa Luzia	28	3	1	0.569±0.077	0.0010±0.0008	0.819	1.28
Portugal	76	15	28	0.883±0.023	0.0081±0.0043	-0.002	1.588
Espanha	69	39	21	0.933±0.013	0.0119±0.0061	-0.594	-0.528

Por meio das estatísticas F (F_{ST}), realizaram-se comparações par a par para averiguar o grau de diferenciação entre as diferentes populações, estas estão representadas na Quadro 3.

O valor de F_{ST} que indica significativamente maior semelhança entre ilhas refere-se à ilha de Santa Luzia com a ilha da Madeira ($F_{ST}=0.076$). Este resultado era esperado dado que os três haplótipos que surgem na ilha de Santa Luzia aparecem também na ilha da Madeira. É também possível verificar que as ilhas que apresentam maior diferenciação das potenciais fontes colonizadoras são a Madeira, Porto Santo, Santa Luzia e também São Jorge ($F_{ST}=0.713/0.767/0.771$, respetivamente). Esta última apresenta grande diferenciação em relação às restantes ilhas do arquipélago dos Açores (valores F_{ST} entre 0.627 e 0.990), porém verifica-se uma maior semelhança com Porto Santo, ao nível dos haplótipos como verificado anteriormente na rede haplotípica, ainda assim, esta não se mostra significativa na estatística F ($F_{ST}=0.023$).

Quadro 3 - Matriz de diferenças *pairwise* para o marcador D-loop. Todos os valores da Quadro são significativos com exceção dos representados com um asterisco.

	São Miguel	Terceira	São Jorge	Santa Maria	Pico	Faial	Flores	Corvo	Graciosa	Madeira	Porto Santo	Santa Luzia	Portugal	Espanha
São Miguel	0.000													
Terceira	0.449	0.000												
São Jorge	0.804	0.660	0.000											
Santa Maria	0.565	0.173	0.940	0.000										
Pico	0.387	0.326	0.627	0.572	0.000									
Faial	0.389	0.368	0.771	0.579	0.036*	0.000								
Flores	0.394	0.353	0.881	0.529	0.245	0.118	0.000							
Corvo	0.506	0.500	0.990	0.824	0.196	0.060*	0.244	0.000						
Graciosa	0.593	0.606	0.977	0.859	0.370	0.270	0.445	0.572	0.000					
Madeira	0.707	0.569	0.224	0.769	0.514	0.641	0.735	0.761	0.787	0.000				
Porto Santo	0.785	0.627	0.023*	0.918	0.584	0.739	0.853	0.959	0.957	0.207	0.000			
Santa Luzia	0.789	0.638	0.521	0.923	0.607	0.749	0.863	0.962	0.959	0.076	0.383	0.000		
Portugal	0.325	0.466	0.783	0.552	0.219	0.101	0.132	0.084	0.253	0.713	0.767	0.771	0.000	
Espanha	0.073	0.339	0.644	0.448	0.156	0.164	0.216	0.255	0.362	0.592	0.625	0.627	0.169	0.000

Com o objetivo de perceber como estava distribuída a variação na sub-espécie nos arquipélagos de colonização portuguesa foi realizada uma análise molecular da variância (AMOVA) que está representada na Quadro 4. O resultado desta análise não foi significativo para explicar a variação entre arquipélagos. No entanto, obteve-se que 43% da variação é explicada significativamente pela variação encontrada entre ilhas do mesmo arquipélago e 39% da variação é explicada pela variância entre ilhas. Sugerindo que a maior parte da diversidade encontra-se entre e dentro das ilhas do que entre arquipélagos, sendo consequência da partilha de haplótipos entre arquipélagos.

Quadro 4 - Análise Molecular da Variância entre arquipélagos e ilhas.

Fonte de Variação	% Variação	Estatísticas ϕ	Valor de p
Entre arquipélagos	17.88	$\phi_{CT}=0.1788$	0.0763
Entre ilhas (por arquipélago)	43.41	$\phi_{SC}=0.5286$	0.0000
Entre ilhas	38.71	$\phi_{ST}=0.6129$	0.0000

Os testes de neutralidade estão representados na Quadro 2 e foram testados para um valor de significância de 0.05, sendo que os valores que cumprem este requisito estão assinalados com asterisco. A ilha da Madeira é a única que apresenta valores significativos para ambos os testes enquanto a ilha de São Jorge apresenta valor significativo apenas para o teste de FS de Fu. Para as restantes ilhas não obtivemos valores significativos. Os valores dos dois testes obtidos para a ilha da Madeira (D de Tajima= -2.124 e Fs de Fu= -23.12) e o valor do teste FS para a ilha de São Jorge (Fs de Fu = -2.176) são negativos, o que indica uma possível expansão demográfica ou sinal de seleção, dado existirem mais haplótipos do que o esperado para o tamanho de cada ilha.

Microsatélites

Para o cromossoma Y foram amostrados 195 indivíduos, destes apenas foram tratados 180, por exclusão dos indivíduos que continham informação para apenas 2 microsatélites. A amplificação dos microsatélites não teve igual taxa de sucesso para todos os loci, uma elevada percentagem de *missing data* (Quadro 5) para os microsatélites Y6 e Y12 levaram a que mais alguns indivíduos fossem eliminados das análises que não têm capacidade para tratar a ausência de dados. Assim, foram utilizados todos os dados (N=180) para as análises de diversidade e para o STRUCTURE e foram utilizados os dados sem *missing data* (N=135) para as estatísticas F, AMOVA, rede de haplótipos e PCoA.

O *score* dos microsatélites foi realizado de modo a minimizar ao máximo o erro associado a categorização dos alelos. No Anexo III, encontra-se a Quadro com os dados de microsatélites corrigidos.

Análise de diversidade

A diversidade de cada microsatélite encontram-se sumarizada na Quadro 5, sendo de um modo geral elevada, variando entre 0.626 para o Y21 e 0.876 para o Y23. A percentagem de *missing data* é elevada para os microsatélites Y6 e Y12 (%MD=14.4% e %MD=11.7%, repetivamente) pelo que a sua diversidade pode estar subestimada. A média de alelos por locus por população é bastante inferior ao número de alelos amostrados para cada locus, o que indica que cada ilha possui um conjunto de alelos específicos.

Quadro 5 – Quadro das medidas de diversidade genética dos microsatélites para o cromossoma Y.

Locus	N	Na	A	Ar	Diversidade	%MD	Tamanho (pb)
Y6	154	8	2.33	2	0.688	14.4	116-130
Y12	159	13	3.17	2	0.784	11.7	120-146
Y24	177	10	3.17	2	0.631	1.7	382-408
Y21	179	9	3.25	3	0.626	0.6	311-327
Y23	175	16	3.83	3	0.876	2.8	298-330

N – número de indivíduos; Na – número de alelos por locus; A – média de alelos por locus; Ar – alelos raros; MD – *missing data*.

Ao estimar o número de haplótipos foi necessário retirar os indivíduos com *missing data* de modo a não enviesar os dados obtidos. Todas as restantes análises de diversidade utilizaram todos os indivíduos e os resultados encontram-se representados na Quadro.6. A Quadro com os haplótipos de cada indivíduo pode ser consultada no Anexo V.

No total, obteve-se 77 haplótipos, em que apenas cinco foram amostrados em mais do que uma ilha. Os cinco haplótipos partilhados ocorrem entre: São Miguel, Terceira, Pico e Faial; São Miguel, Terceira e Flores; São Miguel e Terceira; Pico e Faial; Pico e Graciosa, ou seja, existe alguma partilha de haplótipos entres as ilhas do grupo central e São Miguel. Somente 21 haplótipos têm um N>1 o que nos deixa um total de 56 haplótipos únicos. As ilhas que apresentam maior número de haplótipos e maior diversidade haplotípica são as ilhas de São Miguel (N=19 e H=0.657) e da Madeira (N=16 e H=0.692). Por outro lado as ilhas que apresentam menor número de haplótipos e menor diversidade haplotípica são as ilhas de Santa Luzia (N=2 e H=0.258) e São Jorge (N=2 e H=0.200) com apenas dois haplótipos. Os dois haplótipos de Cabo Verde apresentam apenas dois locus com alelos em comum enquanto os haplótipos de São Jorge apresentam três locus com alelos comuns. As

ilhas com maior diversidade são também aquelas que apresentam mais alelos privados, com exceção da ilha da Terceira que apesar da baixa diversidade apresenta um alelo privado, sendo a ilha da Madeira a que possui maior número de alelos privados ($A_p=4$).

Quadro 6 – Diversidade haplotípica dos microssatélites

População	N	Haplótipos	Lp	H	Ap
Santa Maria	16	5	5	0.242	0
São Miguel	29	19	3	0.657	2
Terceira	18	9	3	0.331	1
Graciosa	4	4	3	0.325	0
São Jorge	6	2	5	0.2	0
Pico	15	6	3	0.362	0
Faial	12	6	2	0.61	0
Flores	6	5	3	0.416	0
Madeira	19	16	5	0.692	4
Porto Santo	9	7	5	0.494	2
Santa Luzia	15	2	3	0.258	0
Portugal	5	5	3	0.588	3

N – número de indivíduos; Lp – loci polimorficos; H – diversidade; Ap – alelos privados.

Análise da Estrutura Populacional

Para a análise F_{ST} não foram utilizados indivíduos com *missing data*, o que fez com que os indivíduos de Portugal fossem removidos desta análise.

Os valores de F_{ST} são relativamente baixo entre as populações das ilhas de São Miguel, Terceira, Pico, Faial, Flores, Graciosa e Madeira (Quadro 7). As ilhas de São Jorge e Santa Maria são as que apresentam maiores valores de maior diferenciação com todas as outras ilhas, sendo que a ilha mais próxima de São Jorge é o Faial ($F_{ST} = 0.332$), e a mais próxima de Santa Maria é a ilha da Madeira ($F_{ST} = 0.345$). Porto Santo também se mostra relativamente afastado de todas as ilhas sendo a ilha que se encontra mais próxima a Madeira ($F_{ST} = 0.299$). A ilha de Santa Luzia apresenta maior relação com a Madeira ($F_{ST} = 0.333$) e depois Santa Maria ($F_{ST} = 0.399$). A análise do programa STRUCTURE (ver Figura 9) corrobora a análise F_{ST} para um $K=6$. As ilhas das Flores, Graciosa, Pico e Terceira apresentam a mesma identidade genética, sendo que a mesma aparece ainda no Faial e São Miguel. O Faial mostra uma segunda identidade genética, a mesma que ocorre em São Jorge, tal como assinalado pelo valor de F_{ST} . São Miguel também apresenta uma segunda população genética, porém esta não ocorre em mais nenhum local em estudo. A ilha de Santa Luzia apresenta dois sinais, um igual a Santa Maria e outro igual à Madeira e São Jorge. A ilha de Porto Santo destaca-se por evidenciar uma identidade genética diferente de todas as outras amostradas. Na ilha da Madeira foi detetado o mesmo sinal genético que em Portugal Continental (ver Figura 9).

Quadro 7 - Matriz de diferenças *pairwise* para os microssatélites Y6, Y12, Y24, Y21, Y23 do cromossoma Y. Todos os valores da Quadro são significativos com exceção dos representados com um asterisco.

	São Miguel	Terceira	São Jorge	Santa Maria	Pico	Faial	Flores	Graciosa	Madeira	Porto Santo	Santa Luzia
São Miguel	0.000										
Terceira	0.122	0.000									
São Jorge	0.458	0.684	0.000								
Santa Maria	0.361	0.549	0.742	0.000							
Pico	0.106	0.015*	0.646	0.571	0.000						
Faial	0.127	0.152	0.332	0.438	0.179	0.000					
Flores	0.074*	0.045*	0.607	0.547	0.050*	0.074*	0.000				
Graciosa	0.152	0.242	0.642	0.581	0.200	0.072*	0.152*	0.000			
Madeira	0.126	0.238	0.400	0.345	0.250	0.125	0.136	0.178*	0.000		
Porto Santo	0.345	0.526	0.644	0.433	0.526	0.381	0.438	0.512	0.299	0.000	
Santa Luzia	0.412	0.570	0.720	0.399	0.585	0.429	0.562	0.586	0.333	0.417	0.000

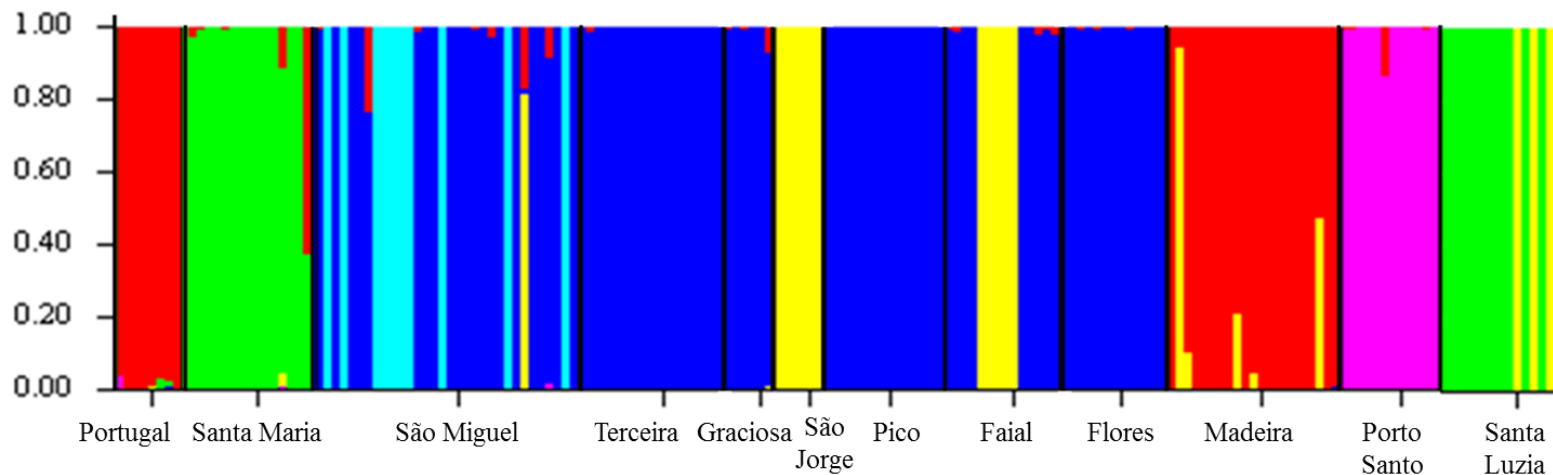


Figura 9 - Estrutura populacional de *Mus musculus domesticus* inferida pelo STRUCTURE baseados no genótipo dos microssatélites Y6, Y12, Y24, Y21 e Y23 para o arquipélago dos Açores (exceto o Corvo), e da Madeira, ilha de Santa Luzia (Cabo Verde) e Portugal continental. Cada indivíduos está representado por uma linha colorida correspondente a proporção de semelhança de genótipos. As linhas pretas marcam a delimitação das ilhas. Este gráfico de barras representa o número mais provável de populações genéticas, K=6.

A análise de coordenadas principais (PCoA), representada na Figura.10, foi realizada apenas com os indivíduos sem *missing data* (N=135) e mais uma vez corrobora os resultados da F_{ST} e do STRUCTURE. Os três eixos desta análise explicam, respetivamente, 72%, 12% e 9% da variação observada nos dados, o que se considera bastante explicativo. O primeiro eixo separa a maioria das ilhas dos Açores e Porto Santo da ilha da Madeira, Santa Maria, São Jorge e Portugal continental (Portugal continental tem N=1, indivíduo de Lisboa que é o único a ter informação para todos os microssatélites). Pode verificar-se que com a exceção dos genótipos de Santa Maria e São Jorge todos os outros genótipos amostrados nas ilhas do Açores se encontram no lado esquerdo do primeiro eixo do gráfico. Um genótipo amostrado na ilha do Faial ocorre juntamente com as exceções no lado direito do eixo principal. A ilha de São Miguel aparece firmemente dividida em duas populações pelos dois eixos com maior explicação representando as duas populações identificadas pelo STRUCTURE em que uma representa a população dominante no arquipélago dos Açores e outra uma população que não ocorre em nenhum outro local amostrado neste estudo. Os dois genótipos da ilha de Santa Luzia aparecem também eles separados, um mais próximo da população de Santa Maria e outro mais próximo da população de São Jorge e Madeira. A ilha de Porto Santo apresenta-se separada do *cluster* principal dos Açores e da Madeira, Santa Maria e São Jorge, apoiando a hipótese do STRUCTURE de ser uma população independente. A existência de um migrante na população de Porto Santo é visível na Figura 10, em que um indivíduo amostrado na Madeira apresenta um genótipo de Porto Santo.

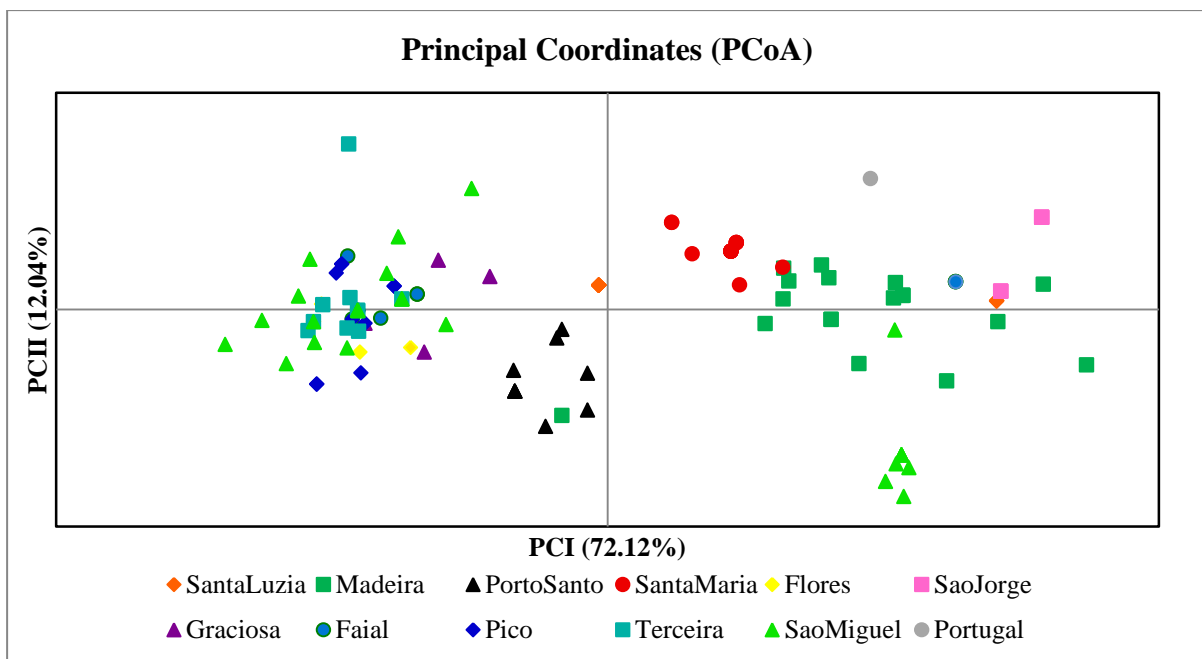
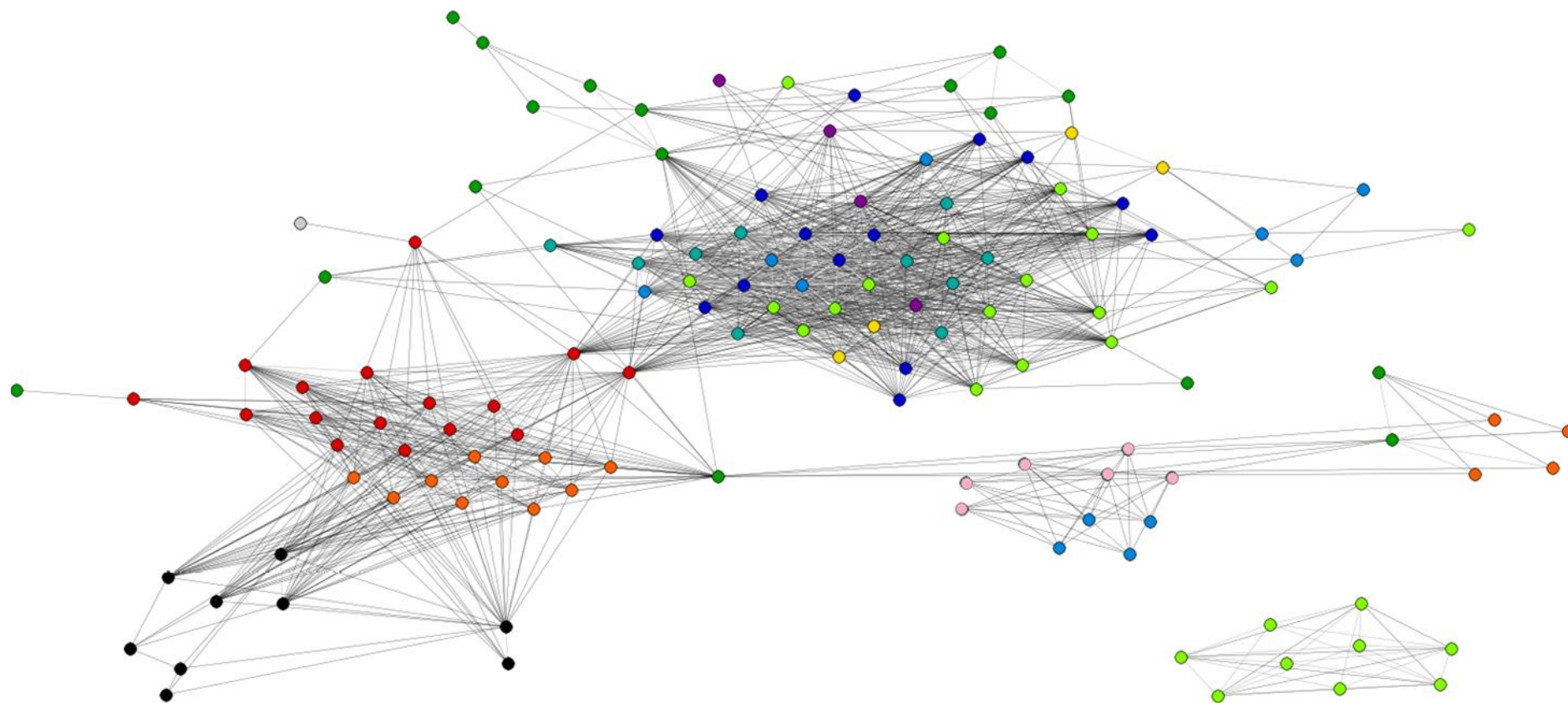


Figura 10 – Análise de Coordenadas Principais (PCoA) para as populações de ratinho-caseiro estudadas para os microssatélites Y6, Y12, Y24, Y21 e Y23 nas ilhas do Arquipélago dos Açores (exceto o Corvo), Madeira, Porto Santo, Santa Luzia e Portugal continental. Os três primeiros eixos explicam 72.12%, 12.04% e 8.52% da variação, respetivamente. Cada símbolo representa um genótipo, colorido de acordo com a ilha em que foi amostrado.

A relação entre os indivíduos de cada ilha está representada na rede de haplótipos da Figura 11. Esta análise forma 5 grupos, o superior é constituído pelas ilhas dos Açores que formam uma população genética: Flores, Graciosa, Pico, Terceira, parte do Faial e parte de São Miguel. Alguns indivíduos de São Miguel que já tinham revelado ser de uma população genética diferente, formam um agrupamento isolado (em baixo à direita). À esquerda temos o agrupamento dos genótipos de parte dos indivíduos da ilha de Santa Luzia e dos indivíduos da ilha de



Legenda:



Figura 11 – Rede de haplótipos baseada nas distâncias de alelos partilhados entre a todas as ilhas do arquipélago dos Açores (exceto a ilha do Corvo), as ilhas da Madeira, Porto Santo, Santa Luzia e Portugal Continental. Apenas ligações com distâncias iguais ou inferiores ao *percolation threshold* (0.4) estão representadas. Os genótipos estão identificados consoante a ilha em que foram amostrados.

Santa Maria, e mais a baixo surge o agrupamento dos indivíduos da ilha de Porto Santo sem ligação clara com outras ilhas, formando também este agrupamento uma população. Na parte inferior central da rede de haplótipos forma-se o agrupamento com os restantes indivíduos dos Açores representantes das ilhas de São Jorge e Faial, e os restantes indivíduos da ilha de Santa Luzia, os quais formam outra população genética. Todos estes dados suportam as populações genéticas inferidas pelo STRUCTURE. A única população que não é recuperada na rede de haplótipos é a associação dos indivíduos da Madeira e de Portugal continental que nesta análise está representado com um único indivíduo. Os indivíduos da Madeira também não formam um agrupamento entre si, dado serem bastante diferenciados uns dos outros, no entanto verifica-se pelo menos três pequenos agrupamentos espalhados pela rede haplótipos.

Ocr1

Análise filogenética

A análise da sequência de interesse das 121 amostras para o gene *Ocr1* determinou a formação de vinte e dois haplótipos (Anexo VI), sendo dois deles claramente dominantes (*Ocr1_1* e *Ocr1_2*) e onze haplótipos únicos. O haplótipo mais disseminado, *Ocr1_1*, contém 66 indivíduos provenientes de todas as ilhas dos Açores (N= 37, ainda que exista apenas um indivíduo da ilha de Santa Maria), da ilha de Porto Santo (N=5), da ilha da Madeira (N=2), da ilha de Santa Luzia (N=1) e ainda de Portugal continental (N=9), Espanha (N=8), Israel (N=1) e Itália (N=4). O haplótipo *Ocr1_2*, também bastante disseminado, inclui 50 haplótipos, não aparecendo nas ilhas de Porto Santo, São Jorge e Corvo. Este está, assim, presente nas restantes ilhas dos Açores (N=24), na ilha da Madeira (N=8), em Portugal continental (N=1), em Espanha (N=1), Reino Unido (N=2) e em Itália (N=12). Os haplótipos únicos aparecem na Grécia, Israel, Itália e depois nas ilhas da Madeira, Santa Luzia, Pico e Santa Maria (um em cada).

A árvore filogenética obtida para este marcador encontra-se na Figura 12, e seguiu o modelo evolutivo TIM2+I selecionado pelo programa JMODELTEST. Esta apresenta uma baixa resolução no estabelecimento das relações evolutivas entre os haplótipos amostrados, conseguindo apenas separar, para além dos *outgroups*, dois haplótipos únicos provenientes de Israel (*Ocr1_11*) e Santa Maria (*Ocr1_19*) dos restantes haplótipos amostrados.

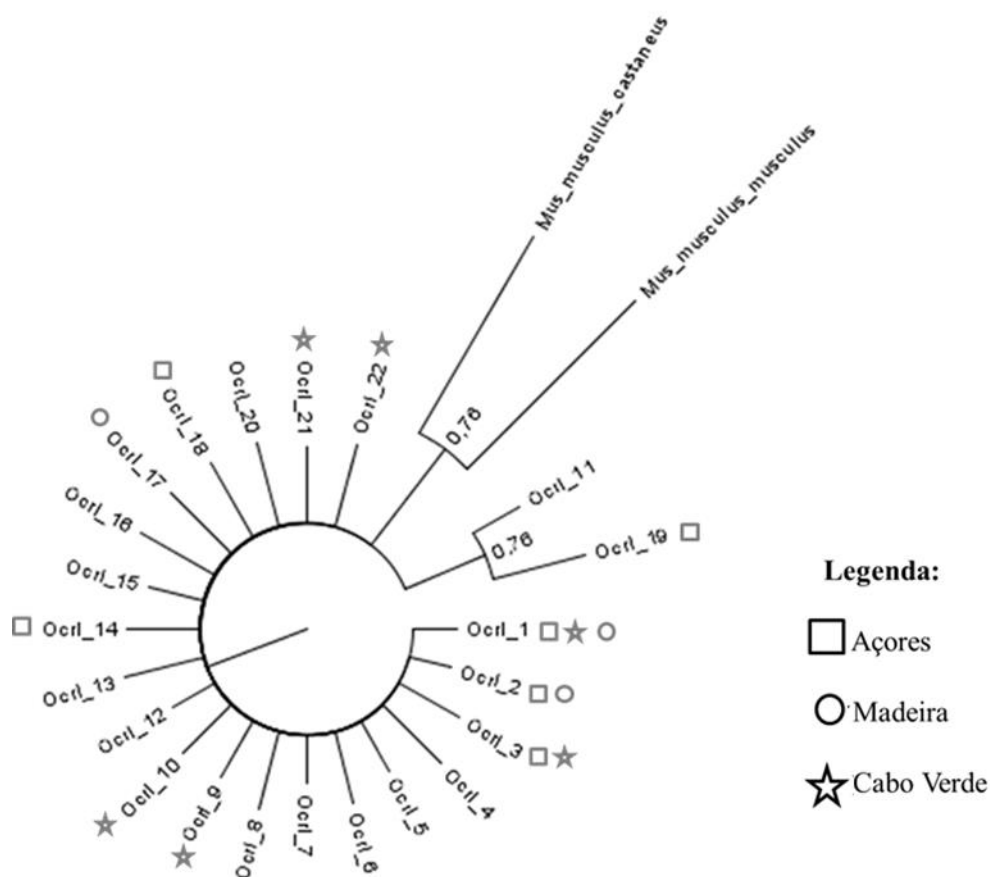
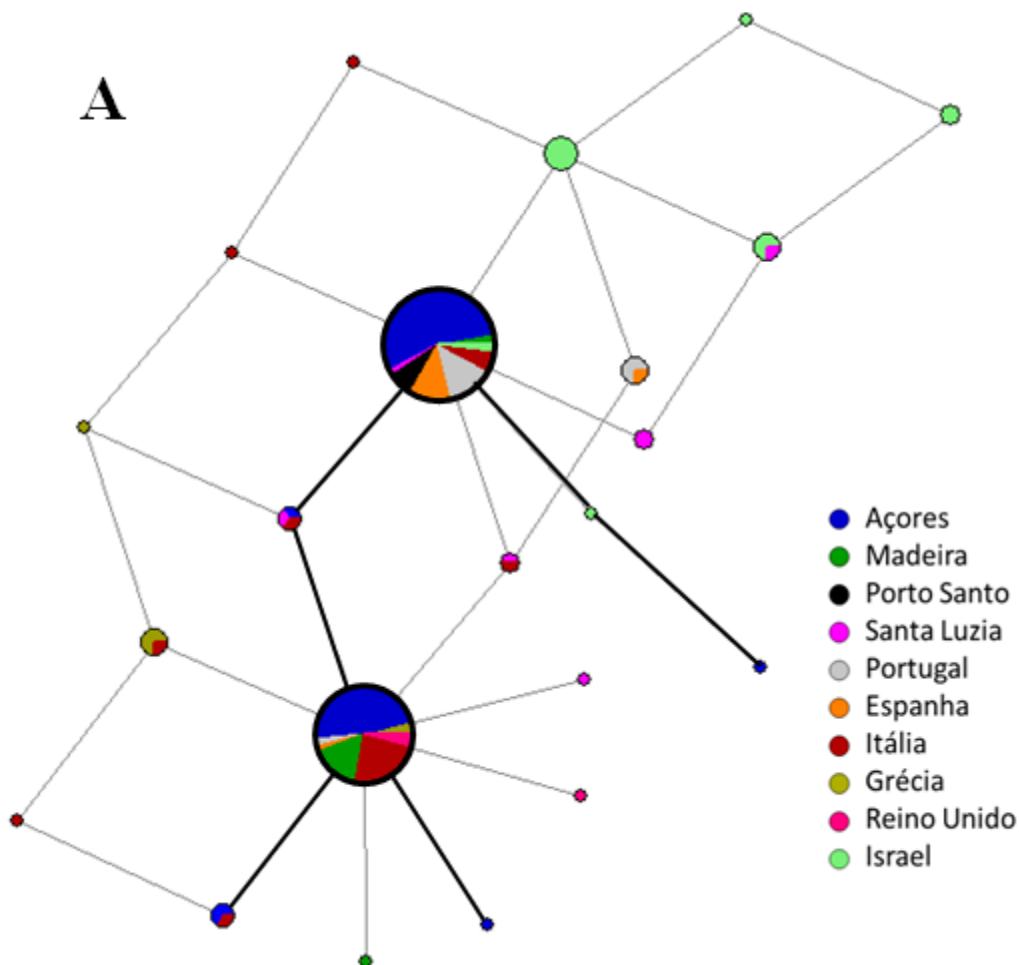


Figura 12 - Árvore filogenética Bayesiana para o intrão *Ocr1* no cromossoma X para todas as sequências obtidas neste estudo e publicadas da sub-espécie *Mus musculus domesticus*.

Análise da estrutura populacional e demográfica

A rede haplotípica para o marcador do cromossoma X está representada na figura 13 A – para todos os locais de estudo e B – para o arquipélago dos Açores. Apresenta dois haplótipos dominantes, Ocr1_1 (N=66) e Ocr1_2 (N=50) sem ligação direta, e com vários haplótipos descendentes apresentando, ambos, uma topologia em forma de estrela. A ligação entre os dois haplótipos dominantes é através dois haplótipos, Ocr1_3 e Ocr1_12 (Figura 13A), compostos por indivíduos dos Açores, Santa Luzia e Itália. Verifica-se, assim, dois agrupamentos de haplótipos em torno dos haplótipos dominantes. Em que os indivíduos dos Açores aparecem igualmente distribuídos por ambos os agrupamentos (Figura 13 A-B), com a exceção das ilhas de São Jorge e Corvo que aparecem apenas no Ocr1_1 (Figura 13B). As ilhas do arquipélago da Madeira aparecem separadas, com a ilha da Madeira a aparecer maioritariamente no Ocr1_2 enquanto a ilha de Porto Santo se encontra exclusivamente no Ocr1_1 juntamente com a generalidade dos indivíduos portugueses (Figura 13A) . Assim sendo, os haplótipos da Madeira aparecem na sua maioria no grupo do Ocr1_2 e os de Portugal continental no grupo do Ocr1_1. Os haplótipos da ilha de Santa Luzia encontram-se distribuídos por ambos os agrupamentos, bem como os haplótipos italianos, aparecendo nos dois haplótipos dominantes e em haplótipos descendentes de ambos (Figura 13A). Relativamente aos haplótipos que foram detetados nas ilhas dos Açores, apesar de estarem bem distribuídos por ambos os agrupamentos, algumas ilhas apresentam uma forte dominância de um dos haplótipos (Figura 13B). É o caso da ilha das Flores que apresenta 5 indivíduos no Ocr1_1 e apenas 1 indivíduo no Ocr1_2, e a ilha de Santa Maria que apresentam o padrão inverso 1 indivíduo no Ocr1_1 e 4 indivíduos no Ocr1_2.



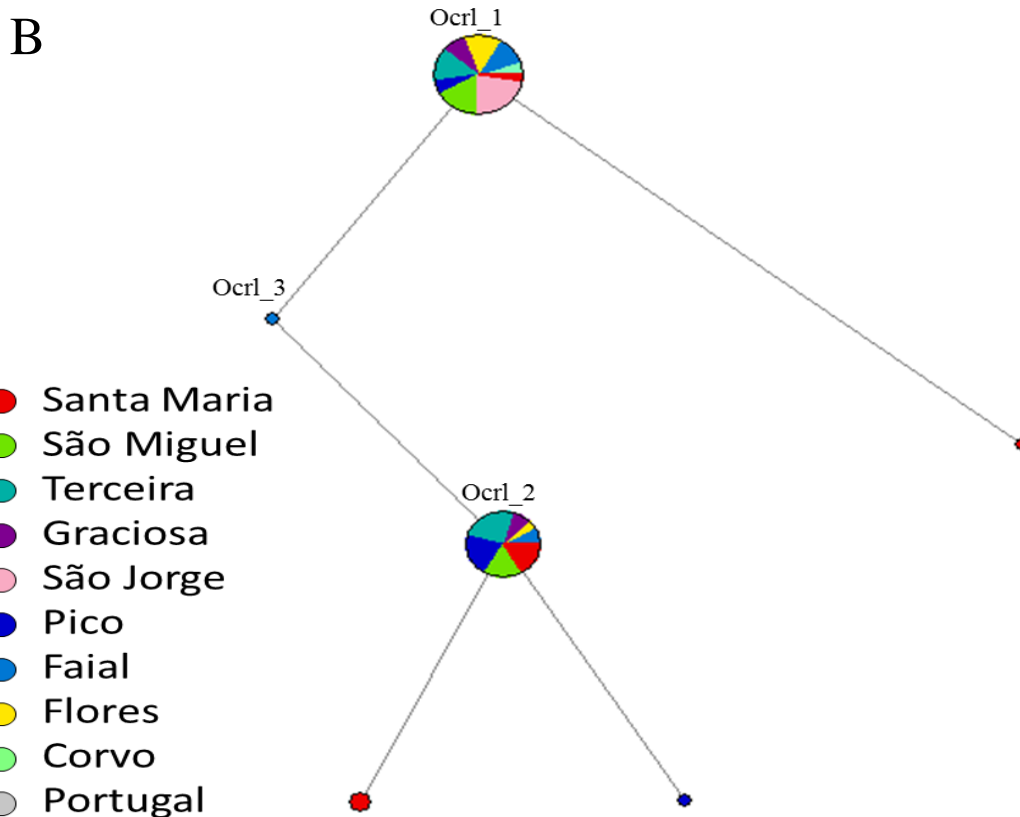


Figura 13 - Rede de haplótipos *Neighbour-Joining* para o marcador do gene *Ocr1* de *Mus musculus domesticus*. Cada círculo representa um haplótipo único sendo o seu tamanho proporcional à frequência de ocorrência. As cores são referentes às ilhas/arquipélagos e países onde foram detetados. O tamanho das linhas pretas que ligam os haplótipos é proporcional à distância mutacional. **A** – Representação de todos os haplótipos amostrados no arquipélago dos Açores e da Madeira, ilha de Santa Luzia (Cabo Verde), Portugal, Espanha, Grécia, Itália, Israel e Reino Unido. A linha a preto realça parte da rede de haplótipos referente aos haplótipos detetados nos Açores. **B** – Representação dos haplótipos encontrados no arquipélago dos Açores, por ilha, realçada em A.

A variabilidade haplotípica e nucleotídica do gene *Ocr1*, representada na Quadro 8, não apresenta grandes diferenças entre a maioria das ilhas e o continente. O local de maior diversidade haplotípica é Israel (0.7912 ± 0.0894), sendo inferior à ilha que apresenta maior diversidade haplotípica, a ilha de Santa Luzia ($Hd=0.9524 \pm 0.0955$). Os restantes países amostrados apresentam uma diversidade intermédia, sendo que Portugal é o que apresenta menor diversidade ($Hd=0.4725 \pm 0.1358$). Nas ilhas verifica-se uma grande discrepância de diversidades, com as ilhas de São Jorge, Corvo e Porto Santo com diversidade 0 ao apresentarem apenas um haplótipo, seguindo as ilhas das Flores e da Madeira ($Hd=0.3333 \pm 0.2152$ e 0.4727 ± 0.1617 , respetivamente). As restantes ilhas apresentam uma diversidade ligeiramente superior a Portugal continental, como se pode verificar pela Quadro 8. Quanto à variabilidade nucleotídica esta é relativamente baixa, apresentando apenas dois indivíduos heterozigóticos na ilha de Santa Luzia, sendo esta a ilha que possui maior diversidade ($\pi=0.002090 \pm 0.001561$).

Quadro 8 - Variabilidade genética para o intrão (960pb) do gene *Ocr1* na sub-espécies *Mus musculus domesticus* para todas as ilhas do Arquipélago dos Açores, ilha da Madeira e Porto Santo do Arquipélago da Madeira, ilha de Santa Luzia do Arquipélago de Cabo Verde, Portugal, Espanha, Grécia, Israel, Itália e Reino Unido. O número de haplótipos (Hap), número de sítios variáveis (S), diversidade haplotípica (Hd) e diversidade nucleotídicas (π), bem como os testes de Neutralidade D de Tajima e FS de Fu foram determinados no programa Arlequin v 3.5.2.2. Os valores de P, quando significativos, dos testes D de Tajima e FS de Fu, estão assinalados com um asterisco (*) ($P < 0.05$).

Zona geográfica	População	N	Hap	S	Hd	Π	D de Tajima	FS de Fu
Arquipélago dos Açores	Santa Maria	8	4	4	0.7500±0.1391	0.001897±0.001418	0.283	-0.240
	São Miguel	11	2	2	0.5091±0.1008	0.001061±0.000865	1.502	2.343
	Terceira	11	2	2	0.5455±0.0722	0.001136±0.000909	1.828	2.500
	Graciosa	5	2	2	0.6000±0.1753	0.001250±0.001106	1.459	1.688
	São Jorge	5	1	0	0.0000±0.0000	0.0000±0.0000	0.000	NA
	Pico	8	3	3	0.6071±0.1640	0.001153±0.000954	-0.177	0.390
	Faial	7	3	2	0.6667±0.1598	0.001091±0.000936	1.168	0.110
	Flores	6	2	2	0.3333±0.2152	0.000694±0.000706	-1.132	0.952
	Corvo	2	1	0	0.0000±0.0000	0.000000±0.000000	0.000	NA
Arquipélago da Madeira	Madeira	11	3	3	0.4727±0.1617	0.000871±0.000755	-0.628	0.242
	Porto Santo	5	1	0	0.0000±0.0000	0.0000±0.0000	0.000	NA
Arquipélago de Cabo Verde	Santa Luzia	7	6	4	0.9524±0.0955	0.002092±0.001562	0.518	-3.554*
Continente	Portugal	14	3	3	0.4725±0.1358	0.000984±0.000804	0.006	0.735
Continente	Espanha	9	3	3	0.4167±0.1907	0.000868±0.000770	-0.936	0.016
Continente	Grécia	6	3	2	0.7333±0.1552	0.001001±0.000934	-0.050	-0.427
Continente	Italia	23	9	5	0.7154±0.0941	0.001661±0.001171	0.182	-4.331*
Continente	Reino Unido	3	2	1	0.6667±0.3143	0.000771±0.000961	0.000	0.201
Continente	Israel	14	6	4	0.7912±0.0894	0.001461±0.001098	0.016	-2.249*

O cálculo da matriz de diferenças entre ilhas/países amostrados foi realizado por meio de estatísticas F. Este apresentou vários valores não significativos entre as ilhas em estudo e os potenciais colonizadores, Portugal e Espanha, que não serão analisados. De um modo geral, os valores de F_{ST} são relativamente elevados, verificando-se uma grande divergência entre as populações dos locais amostrados.

Os valores mais baixos de F_{ST} representam a ligação entre Portugal continental e a ilha de Santa Luzia ($F_{ST}=0.169$) e Portugal continental e a ilha da Terceira ($F_{ST}=0.203$), sendo os que refletem a maior semelhança. A ilha da Madeira apresenta uma divergência acentuada em relação a todos os outros locais de amostragem. No entanto, o valor de F_{ST} que indica maior homologia é com Grécia e a ilha de Santa Luzia ($F_{ST}=0.308$ e $F_{ST}=0.394$, respetivamente). Pelo contrário, os locais com maior divergência são as ilhas de São Jorge e Porto Santo ($F_{ST}=0.711$ e $F_{ST}=0.618$, respetivamente). Quanto às ilhas do arquipélago dos Açores apenas existem três valores significativos, dois deles apresentando grande divergência entre São Jorge e Santa Maria e São Jorge e Pico ($F_{ST}=0.568$ e $F_{ST}=0.640$, respetivamente). A maior semelhança no arquipélago dos Açores ocorre entre a ilha de Santa Maria e a ilha das Flores ($F_{ST}=0.347$). A ilha de Santa Luzia para além da elevada semelhança a Portugal continental apresenta também semelhança com Itália.

Quadro 9 - Matriz de diferenças *pairwise* para o marcador *Ocrl*. Os valores significativos ($p < 0.05$) estão representados com um asterisco.

	São Miguel	Terceira	São Jorge	Santa Maria	Pico	Faial	Flores	Corvo	Graciosa	Madeira	Porto Santo	Santa Luzia	Portugal	Espanha	Grécia	Itália	Reino Unido	Israel
São Miguel	0.000																	
Terceira	-0.029	0.000																
São Jorge	0.230	0.430	0.000															
Santa Maria	0.199	0.069	0.568*	0.000														
Pico	0.160	-0.018	0.640*	0.000	0.000													
Faial	-0.120	-0.043	0.267	0.176	0.150	0.000												
Flores	-0.045	0.148	0.028	0.347*	0.378	-0.064	0.000											
Corvo	-0.010	0.233	0.000	0.367	0.464	-0.027	-0.304	0.000										
Graciosa	-0.167	-0.121	0.331	0.098	0.062	-0.189	-0.067	-0.017	0.000									
Madeira	0.267	0.067	0.711*	0.039	-0.079	0.269	0.493*	0.596*	0.190	0.000								
Porto Santo	0.179	0.382	0.000	0.515*	0.593	0.200	-0.034	0.000	0.250	0.678*	0.000							
Santa Luzia	0.135	0.192	0.296*	0.290*	0.296*	0.103	0.120	0.007	0.063	0.394*	0.230*	0.000						
Portugal	0.056	0.203*	0.089	0.390*	0.390*	0.066	-0.041	-0.155	0.029	0.478*	0.046	0.169*	0.000					
Espanha	0.002	0.175	0.028	0.370*	0.385*	0.000	-0.130	-0.259	-0.023	0.487*	-0.023	0.140	-0.079	0.000				
Grécia	0.494*	0.372*	0.841*	0.266*	0.275*	0.488*	0.655*	0.724*	0.443*	0.308*	0.810*	0.501*	0.626*	0.640*	0.000			
Itália	0.116	-0.003	0.444*	0.008	-0.041	0.101	0.263*	0.315	0.030	-0.005	0.414*	0.273*	0.306*	0.285*	0.144*	0.000		
Reino Unido	0.428*	0.232	0.926*	0.044	0.020	0.431*	0.667	0.813	0.353	0.017	0.902*	0.423*	0.597*	0.626*	0.320	0.053	0.000	
Israel	0.483*	0.543*	0.543*	0.594*	0.615*	0.468*	0.433*	0.333*	0.460*	0.688*	0.447*	0.300*	0.340*	0.402*	0.716*	0.563*	0.702*	0.000

A análise molecular da variância determinou significativamente que a maior parte da variação, 68%, do gene *Ocr1* se encontra entre todas as ilhas, sem estruturação por arquipélago, como se pode verificar na Quadro 10.

Quadro 10 – Análise molecular da variância (AMOVA) para o gene *Ocr1*. Os valores de p estão definidos para um nível de significância de 0.05.

Fonte de Variação	% Variação	Estatísticas ϕ	p
Entre arquipélagos	-5.02	$\phi_{CT}=-0.05022$	0.68524
Entre ilhas (por arquipélago)	37.05	$\phi_{SC}=0.35984$	0.00000
Entre ilhas	68.24	$\phi_{ST}=0.32769$	0.00000

Os testes de desvio à neutralidade apenas encontraram valores significativos para Itália, Israel e para a ilha de Santa Luzia, sendo que todos são valores negativos para o teste FS de Fu (FS=-4.331, FS=-2.249 e FS=-3.554, respetivamente). Tal indica uma possível expansão populacional ou ação da seleção natural, o que seria suportado pelos níveis de diversidade encontrados para ambas as localizações. Os valores dos testes de neutralidade encontram-se representados na Quadro 8.

Discussão

A colonização das ilhas oceânicas foi desde sempre um dos focos de interesse da Biologia Evolutiva. Ao permitir seguir a evolução de uma espécie, em meio natural, num contexto de menor complexidade ecológica comparativamente ao encontrado nos continentes, proporciona excelentes oportunidades para estudar diferentes processos evolutivos. No entanto, se por um lado a dispersão natural de alguns organismos, como plantas ou aves, para ilhas oceânicas localizadas a vários quilómetros dos continentes é possível, a colonização de mamíferos terrestres implica normalmente o transporte (ativo ou passivo) por um vetor, os humanos. O ratinho-caseiro é um dos exemplos em que a colonização de ilhas oceânicas implica necessariamente o transporte passivo pelo homem, permitindo assim, inferências não só sobre a origem do ratinho mas também dos humanos com quem foi transportado.

Este trabalho visou adicionar informação ao estudo da colonização do ratinho-caseiro nas ilhas de colonização histórica portuguesa, através da análise de marcadores moleculares como o D-loop (mtDNA) de transmissão exclusivamente materna, microssatélites do cromossoma Y, de transmissão exclusivamente paterna, bem como da zona não codificante do gene *Ocrl* (cromossoma X), de transmissão bi-parental. Os arquipélagos da Madeira e Açores já tinham sido alvo de estudos anteriores focados no D-loop (Förster *et al.* 2009; Gündüz *et al.* 2001; Gabriel *et al.* 2015), tendo sido aqui realizada uma análise maioritariamente focada nos restantes marcadores. No caso do arquipélago de Cabo Verde, foi iniciado o estudo da sua colonização pelo ratinho-caseiro na ilha de Santa Luzia, com a aplicação de todos os marcadores referidos.

Arquipélago dos Açores

Dos dados já publicados relativamente às populações de ratinho-caseiro no arquipélago dos Açores, verificou-se que este apresenta uma elevada diversidade para o marcador D-loop, tendo sido detetados cinco das seis clades descritas em Jones *et al.* 2011a - B, C, D, E e F (Gabriel *et al.* 2015). Estas clades (Figura 8) apresentam, contudo, expressões diferentes no arquipélago. A clade C, muito comum em Portugal continental, é a que se encontra mais disseminada no arquipélago, tendo sido detetada em seis das nove ilhas (São Miguel, Graciosa, Pico, Faial, Flores e Corvo), com maior representação nas ilhas dos grupos Central e Ocidental. A clade D foi encontrada em cinco ilhas (São Miguel, Terceira, São Jorge, Pico, Faial), com a quase totalidade dos indivíduos originários do grupo Central, enquanto a clade F foi amostrada em 4 ilhas (Santa Maria, Terceira, Faial, Flores) e a clade B em três (Santa Maria, São Miguel e Pico), sendo absolutamente dominante na ilha de São Miguel (ver Gabriel *et al.* 2015 para mais detalhes). Os resultados obtidos, principalmente com as clades B e C, representam a ligação histórica esperada das ilhas dos Açores com Portugal continental, dado serem as clades aí dominantes. No entanto, existem exceções quando se observa a presença das outras clades no arquipélago. A clade D, única clade amostrada em São Jorge, nunca foi detetada em Portugal continental, porém é altamente representada nos países do Norte da Europa (Jones *et al.* 2011b). Outra exceção surge na ilha de Santa Maria que apresenta maioritariamente sequências da clade F e a ilha da Terceira que é, também ela, bem representada pela mesma clade. A clade F, para além de não ter sido detetada em Portugal continental, apresenta uma distribuição geográfica bastante circunscrita, ocorrendo na Irlanda, Islândia, Noruega, Escócia e ilhas Faroé (Jones *et al.* 2011b, 2012), indicando assim um possível elo destas ilhas dos Açores com o Norte da Europa (Gabriel *et al.* 2015).

Na análise do cromossoma Y para o arquipélago dos Açores, encontrou-se um padrão de alguma forma semelhante ao registado no mtDNA, na medida em que continua a existir um sinal filogeográfico predominante na maioria das ilhas dos Açores, como identificado pelas diferentes análises de microssatélites (ver Figuras 9 – 11). No entanto, são de referir como grandes exceções as ilhas de São Jorge e Santa Maria, que apresentam, cada uma, a sua própria identidade genética

(Figuras 9 – 11). A maioria dos indivíduos amostrados nas ilhas de São Miguel, Terceira, Graciosa, Faial, Pico e Flores, exibem a mesma identidade genética. A ilha de São Miguel para além de apresentar uma identidade partilhada com a maioria das ilhas dos Açores, apresenta ainda um segundo sinal que não apresenta qualquer relação com nenhuma das populações amostradas neste estudo, sugerindo uma outra vaga de colonização de ratinhos (machos) de origem externa às áreas de estudo avaliadas (Figuras 9 – 11). A ilha de São Jorge, à semelhança do que acontece com o mtDNA, apresenta-se como uma população geneticamente diferenciada das restantes ilhas, mostrando apenas alguma associação com a vizinha ilha do Faial ($F_{ST}=0.332$), na qual foram amostrados alguns indivíduos com a mesma identidade genética que os indivíduos de São Jorge (Figura 9 – 10). A ilha de Santa Maria apresenta, também, uma população geneticamente distinta de qualquer outra ilha do arquipélago dos Açores (os valores de F_{ST} variam entre 0.361 e 0.742), indicando assim uma possível colonização independente. Tanto os dados de D-loop como os do cromossoma Y apontam para uma colonização única e de origem diferente para as ilhas de São Jorge e Santa Maria, o que é ainda corroborado pelo estudo de Gabriel *et al.* (2013), onde foram analisados 19 microssatélites autossómicos. Estes dados são compatíveis com uma origem única e comum de machos e fêmeas, que terão fundado a população insular e mantido o seu *pool* genético ao longo das gerações, sem contribuições significativas de colonizações secundárias. Este cenário, de múltipla colonização ao nível das diferentes ilhas do mesmo arquipélago, já tinha sido detetado noutras ilhas oceânicas. No arquipélago de Kerguelen todas as ilhas amostradas apresentam o mesmo haplótipo de D-loop com a exceção de duas ilhas em que ocorre um haplótipo completamente diferente, indicando assim uma colonização com origem diferente em relação às restantes ilhas do arquipélago (Hardouin *et al.* 2010). Um panorama semelhante ocorre nas ilhas Faroé em que duas ilhas apresentam colonização pelo Norte da Europa (Finlândia, Suécia e Dinamarca) e a terceira ilha apresenta um sinal de colonização das ilhas Britânicas (Jones *et al.* 2011b). No que toca às restantes ilhas dos Açores, os dados destes marcadores não apontam no mesmo sentido uma vez que, apesar do marcador D-loop indiciar uma primeira colonização portuguesa, os dados do cromossoma Y não apresentam uma ligação direta com Portugal continental. Esta ausência de ligação pode decorrer da falta de amostragem para o continente ou de uma contribuição diferencial de machos e fêmeas na colonização das ilhas, em que o sinal do cromossoma Y aparece de colonizações posteriores à colonização pelos portugueses. No entanto, a ausência de amostras de outras possíveis fontes colonizadoras não permitiu identificar a origem desses machos. Também é possível que a colonização há 600 anos, juntamente com o isolamento das populações de ratinho das diferentes ilhas e a rapidez da evolução dos microssatélites seja suficiente para que o sinal filogeográfico de ligação a Portugal continental não seja detetável.

Os resultados da análise do gene *Ocr1* não apresentam estruturação geográfica dado que as sequências encontradas nas ilhas dos Açores distribuem-se maioritariamente pelos dois haplótipos dominantes (*Ocr1_1* e *Ocr1_2*). Mais uma vez, a exceção verificou-se ao nível da ilha de São Jorge em que ocorre apenas um haplótipo, *Ocr1_1* (Figura 13-B), e em que foram detetados os valores significativos mais elevados de F_{ST} face às restantes ilhas dos Açores ($F_{ST}=0.568$ com Santa Maria e $F_{ST}=0.640$ com o Pico), contribuindo mais uma vez para a hipótese de uma colonização única e diferente das restantes ilhas dos Açores. A ilha de Santa Maria também apresenta um cenário um pouco diferente do arquipélago dos Açores, não só por ocorrer quase exclusivamente num dos haplótipos dominantes (*Ocr1_2*) mas também por apresentar dois haplótipos únicos à parte dos haplótipos dominantes. A ilha de Santa Maria apresenta-se bastante diferenciada das restantes ilhas dos Açores, mostrando maior proximidade com a ilha de Santa Luzia ($F_{ST}=0.290$) do que com as ilhas do mesmo arquipélago. Este resultado, tal como o encontrado para o D-loop e microssatélites do cromossoma Y, reforça a hipótese de uma colonização de origem diferente para esta ilha.

Arquipélago da Madeira

A colonização inicial do ratinho-caseiro no arquipélago da Madeira parece ter a sua assinatura genética na presença quase exclusiva da clade D, para o marcador D-loop, detetada em larga escala no norte da Europa (Förster *et al.* 2009). Esta evidência, juntamente com a ausência desta mesma clade em Portugal continental sugere uma colonização de ratinho-caseiro anterior a descoberta histórica do arquipélago pelos portugueses. A análise de 30 novas amostras da Madeira e 14 de Porto Santo obtidas entre 2013 e 2015 serviram, por um lado, para verificar se o padrão de associação genética com o Norte da Europa, para o marcador D-loop, se mantinha, devido à amostragem anterior ter ocorrido há 15 anos e por não cobrir todas as áreas da ilha. Por outro para ser possível fazer uma comparação atual com a nova amostragem para ilha de Santa Luzia. Para além disso, a ilha da Madeira apresenta uma diversidade genética bastante elevada para uma ilha oceânica colonizada há cerca de 600 anos, assim este maior esforço de amostragem serviria também para verificar se já tínhamos atingido a saturação da diversidade genética para este marcador. Este estudo veio então comprovar que ainda não foi amostrada toda a diversidade existente na ilha da Madeira, pois das novas 44 amostras surgiram 14 novos haplótipo, nunca antes amostrados. O sinal da clade D também continua a ser quase exclusivo na ilha da Madeira, ainda assim, a ligação a Portugal continental, anteriormente resumida à presença de um único indivíduo portador de um haplótipo pertencente à clade C (MADE.20), surge agora associada à presença de mais 2 haplótipos (MADE.42 e Portugal.1), amostrados em 2 indivíduos entre os 30 novos indivíduos amostrados. Assim, em todas as amostragens levadas a cabo até hoje, dos 142 indivíduos analisados apenas foram detetados 3 indivíduos portadores de haplótipos associados a Portugal continental. Neste estudo foi também amostrado pela primeira vez um haplótipo, até então, apenas detetado na Dinamarca (U47461). Esta falta de evidências de associação entre o arquipélago da Madeira e Portugal continental, aliada às várias evidências sobre colonização Viking afasta a hipótese histórica de uma colonização inicial portuguesa para sub-espécie *Mus musculus domesticus*, suportando a hipótese de uma passagem humana no arquipélago da Madeira (voluntária ou acidental) anterior à portuguesa, durante a qual pode ter ocorrido a colonização do ratinho-caseiro. Esta hipótese surgiu inicialmente em Gündüz *et al.* 2001, tendo sido reforçada com maior esforço de amostragem em Förster *et al.* 2009. Mais recentemente, dados paleontológicos vieram apoiar esta hipótese com a datação de estruturas ósseas de um esqueleto parcial de ratinho-caseiro, datado com ^{14}C (Carbono-14) cuja idade situou a colonização do ratinho em pelo menos quatrocentos anos antes da colonização portuguesa e coincidente com o auge da expansão Viking (Rando *et al.* 2014). Apesar destes indícios, dados de aloenzimas (Britton-Davidian *et al.* 2007) e de microssatélites autossómicos (Förster *et al.* 2013; Gabriel *et al.* 2013) indiciam uma ligação entre a Madeira e Portugal continental. De fato, os dados obtidos neste estudo para o cromossoma Y indicam que os ratinhos da ilha da Madeira representam na sua maioria a mesma população genética encontrada nos indivíduos amostrados em Portugal continental (Figura 9). A divergência de resultados entre os marcadores mtDNA e cromossoma Y poderá, então, ser reflexo das diferenças comportamentais dos sexos ao nível da capacidade de dispersão e introgressão numa população estável. Tal poderá dever-se à existência de uma grande resiliência à entrada de fêmeas migrantes a uma população pré-estabelecida (Searle *et al.* 2009b; Jones *et al.* 2011b;). Os machos têm, por sua vez, relativa facilidade em introgridir em populações estabelecidas (Jones & Searle 2015), o que leva a que o mtDNA seja um melhor marcador do evento de primeira colonização e o cromossoma Y permita seguir potenciais colonizações subsequentes. Esta hipótese também explica a diminuta presença de outras clades no arquipélago da Madeira, apesar do fluxo das rotas comerciais constantes durante longos períodos de tempo. Assim, poderá explicar-se que potenciais colonizações subsequentes não tiveram impacto na assinatura genética para o marcador mitocondrial. Outro dado que suporta esta hipótese é a existência de um maior número de linhagens de cromossoma Y do que de mtDNA. Deste modo, pressupõe-se uma

colonização inicial da ilha da Madeira por indivíduos do Norte da Europa e subsequente invasão de machos oriundos de Portugal continental aquando a colonização portuguesa. Os dados do cromossoma X voltam a parecer indicar uma colonização da Madeira independente de Portugal continental (Figura 13A). Este resultado poderia ser considerado esperado dado a primeira colonização bem-sucedida ser atribuída aos Vikings, e as fêmeas que chegaram carregariam dois cromossomas X com a mesma origem. No entanto, o sinal de uma segunda colonização de machos oriundos de Portugal é corroborada pelos dados do cromossoma Y (Figura 9), e verifica-se também uma introgressão de genes autossómicos (Gabriel *et al.* 2013), o que já poderia ter levado à alteração do sinal presente no cromossoma X.

Os dados do D-loop para as ilhas da Madeira e Porto Santo apresentam uma colonização com a mesma origem geográfica e identificada pela clade D. Porém, estas ilhas não apresentam haplótipos partilhados e a rede de haplótipos apresenta uma distribuição em estrela para os haplótipos dominantes de cada ilha. O haplótipo mais comum em Porto Santo difere em apenas uma mutação do haplótipo mais comum da ilha da Madeira, sugerindo que a colonização de Porto Santo poderá ter tido origem nesta (Förster *et al.* 2009) com pelo menos uma fêmea grávida apresentando um haplótipo divergente. No entanto, o padrão apresentado pelos resultados do cromossoma Y é muito diferente dos resultados do mtDNA. Estes indicam uma acentuada divergência entre Porto Santo e Madeira, com a formação de populações completamente independentes (Figuras 9 e 10). Este fenómeno pode ser explicado por origens diferentes de ratinhos machos de cada ilha em colonizações posteriores ao estabelecimento da primeira população de ratinho nesta ilha, sendo que a proximidade entre os haplótipos do cromossoma Y da ilha de Porto Santo e o seu isolamento de outros haplótipos sugere uma origem única e diferente dos machos desta ilha (Figura 9-11). Num estudo anterior com microssatélites autossómicos (Gabriel *et al.* 2013) verificou-se a existência de diversas influências na ilha de Porto Santo nomeadamente da ilha da Madeira, Portugal continental, Santa Maria e o grupo central dos Açores. Contudo, a amostragem aí levada a cabo era muito reduzida, não permitindo inferências robustas. Quanto aos resultados do gene *Ocr1* apontam novamente para uma maior diferenciação para estas ilhas do que seria de esperar à partida, através da não partilha de haplótipos ($F_{ST}=0.678$; $p<0.05$ – Quadro 9). Este resultado suporta a hipótese da contribuição diferencial dos sexos na colonização desta ilha, pois machos com origens diferentes ao introgridirem o seu cromossoma X na população poderão ter levado ao aumento da diferenciação entre estas duas ilhas. A deriva genética inerente ao isolamento de ambas as populações insulares pode também ter contribuído para esta diferenciação. Assim, todos estes dados apontam para que a história da colonização desta ilha seja mais complexa do que anteriormente se considerava, sendo que uma hipótese possível terá envolvido duas vagas de colonização na ilha de Porto Santo, uma inicial através de ratinhos trazidos da Madeira e uma segunda vaga, de origem desconhecida, na qual só contribuíram os machos.

A diversidade haplotípica do D-loop registada na população de ratinho-caseiro da ilha da Madeira é mais elevada do que seria expectável para uma ilha oceânica colonizada há cerca de 600–1000 anos. Esta diversidade é superior à encontrada em Portugal continental e equiparada à diversidade encontrada em Espanha e outras áreas continentais. Esta situação pode ser explicada por uma significativa expansão populacional, sem ocorrência de *bottlenecks* severos, o que é corroborado pelo valor e significância dos testes D de Tajima=-2.214 e FS de Fu = -23.123. A elevada diversidade encontrada na ilha da Madeira repete-se para os microssatélites do cromossoma Y, apresentando a maior diversidade observada ($Hd=0.692$) entre todos os locais de estudo. No que toca à estruturação populacional, os microssatélites apresentam uma fraca estruturação na ilha da Madeira, com haplótipos muitos dispersos. Contrariamente ao que se verifica com o mtDNA, em que todos os haplótipos amostrados são muito semelhantes, com apenas algumas mutações de diferença (Figura 8 e 11, respetivamente).

Ilha de Santa Luzia

A colonização do arquipélago de Cabo Verde tem uma associação histórica com Portugal, tendo este sido o oficial descobridor, em 1460 (Oliveira Marques 1973). No entanto, o povoamento deste arquipélago ocorreu lentamente e por fases. Foi recebendo colonos de várias regiões portuguesas, incluindo das ilhas previamente colonizadas e povoadas dos Arquipélagos dos Açores e da Madeira, tendo recebido também colonos europeus: nórdicos, franceses, ingleses, holandeses, mas também africanos (devido à rota dos escravos). Desta forma, espera-se que possam existir várias influências nas diversas ilhas deste arquipélago. A ilha de Santa Luzia pertence ao grupo barlavento juntamente com as ilhas de Santo Antão, São Nicolau, São Vicente, Sal e Boavista sendo a única não povoada (ver Figura 4 D).

O resultado do marcador D-loop para a ilha de Santa Luzia estabelece um elo com o arquipélago da Madeira, com a deteção de dois dos haplótipos mais frequentes da ilha da Madeira (Made.1 e Made.18) e o mais frequente da ilha de Porto Santo (PSanto2). Nenhuma das sequências encontradas revelou qualquer associação com Portugal continental. Esta associação entre o arquipélago de Cabo Verde e o arquipélago da Madeira não era, de todo inesperada, pela origem de alguns colonos durante o povoamento de Cabo Verde ou pelo papel da Madeira como ponto de reabastecimento nas viagens rumo ao Atlântico Sul. O facto da ilha de Santa Luzia não ser uma ilha povoada, não permite identificar um evento de colonização único, mas a sua utilização como porto piscatório e ocupação temporária por parte de habitantes das ilhas vizinhas, permite a inferência de que parte da diversidade existente nesta ilha poderá refletir a diversidade das ilhas circundantes. Os microssatélites do cromossoma Y apresentam duas populações distintas, como evidenciado pela análise realizada no software STRUCTURE representada na Figura 9, uma delas associada com a ilha de Santa Maria e outra associada a São Jorge e a Madeira. Se pelas rotas de navegação a associação com a Madeira seria algo esperado, a ligação com Santa Maria e São Jorge seria menos expectável devido à distância geográfica. A análise PCoA apresenta o mesmo sinal (Figura 10). Apenas a análise F_{ST} entre São Jorge e Santa Luzia ($F_{ST}=0.720$) não recuperam este sinal de proximidade, no entanto, este resultado pode ser reflexo de uma grande diferenciação entre os dois haplótipos encontrados na ilha de Santa Luzia. Por outro lado, a mesma análise suporta mais uma vez alguma associação de Santa Luzia com Santa Maria ($F_{ST}=0.399$) e com a Madeira ($F_{ST}=0.333$), apesar dos valores de F_{ST} não serem baixos. Estas duas populações, com histórias aparentemente diferentes, sugerem a influência de machos de duas origens geográficas distintas. Deste modo, é possível que a população oriunda da Madeira tenha chegado primeiro, estabelecendo a primeira colonização bem-sucedida, revelando o sinal da clade D para o marcador mitocondrial, e uma segunda vaga de colonização tenha alterado amplamente o sinal filogeográfico do cromossoma Y pela introgressão de machos de uma população migrante.

Ao nível do marcador do cromossoma X os dados foram surpreendentes tendo em atenção os resultados obtidos para as restantes populações analisadas. Por um lado, foi encontrada uma elevada diversidade haplotípica para uma ilha de tão reduzidas dimensões, como Santa Luzia ($H_d=0.952$), e que não foi encontrada nos marcadores haplóides e mais variáveis (mtDNA - $H_d=0.569$ e cromossoma Y - $H_d=0.258$). Apresenta uma grande similaridade com Portugal continental ($F_{ST}=0.169$) ao contrário do que foi registado para o marcador D-loop ($F_{ST}=0.771$). A análise F_{ST} para os microssatélites do cromossoma Y não foi possível de realizar devido a falta de dados de Portugal continental, porém a análise de estrutura populacional realizada no software STRUCTURE não revela a associação desta ilha com o seu potencial colonizador histórico. Este resultado sugere a hipótese de uma contribuição diferencial de machos e fêmeas nesta ilha, juntamente com a ocorrência de um processo de expansão populacional ($O_{crl} - F_s$ de $F_u = -3.554$). No entanto, se a contribuição diferencial dos sexos é algo

expetável, o evento de uma expansão populacional é mais improvável dada a sazonalidade do clima, que ao passar períodos de extrema seca, levaria a *bottlenecks* severos. A particularidade do clima leva à assunção de que a população de ratinho-caseiro na ilha de Santa Luzia não terá um efetivo populacional estável, sendo periodicamente alvo da chegada de migrantes das ilhas adjacentes. Estas continuarão provavelmente a manter o sinal de colonização pela clade D, caso as ilhas circundantes também tivessem sido, elas próprias, colonizadas por animais portadores desta clade. Esta inferência é, contudo, meramente especulativa devido ao desconhecimento acerca das linhagens mitocondriais presentes nessas ilhas. Regista-se a possível introgressão de machos em colonizações subsequentes pelos diferentes haplótipos do cromossoma Y registados. Por sua vez, a entrada de machos na população, levaria também à alteração do sinal obtido no gene *Ocr1* no cromossoma X, por mistura da influência materna e paterna. A elevada diversidade encontrada para um marcador tipicamente com menor taxa de evolução, como o cromossoma X, em relação aos marcadores haplóides levanta várias hipóteses. Por um lado, pode ser fruto de uma maior resiliência do marcador diplóide em perder diversidade em *bottlenecks*, por outro pode ser sinal de uma grande estruturação na ilha de Santa Luzia. O clima sazonal, com vários períodos de seca, podem isolar pequenas populações levando a uma evolução diferencial das mesmas por ação da deriva ocorrendo uma maior diversificação do que seria de esperar. No entanto, qualquer uma destas hipóteses deve ser avaliada cuidadosamente e carece naturalmente de dados mais robustos para ser avaliada com rigor. Para tal, é imprescindível a análise filogeográfica das restantes ilhas do arquipélago de Cabo Verde, principalmente as ilhas adjacentes de Santa Luzia (Figura 4D), para se definir a origem de colonização desta ilha, para a sub-espécie *Mus musculus domesticus*, e determinar a influência subsequentes passagens humanas.

Esta ponderação na análise dos dados é válida para todos dados filogeográficos, obtidos para os vários marcadores, neste estudo. Por um lado, porque nem sempre a amostragem é proporcional ao tamanho da ilha ou suficientemente alargada para aceder a todas as regiões da ilha, levando assim à deteção de um padrão que pode ser diferente do existente na natureza. Por outro lado, por não termos dados comparáveis para todos os possíveis locais de fonte de colonização das ilhas. Este ponto é particularmente importante na análise dos dados dos microssatélites do cromossoma Y, em que os dados obtidos para Portugal continental são insuficientes para estabelecer qualquer relação, e para os quais não existem dados de outras possíveis fontes colonizadoras, como o Norte da Europa. Apesar de existir um estudo prévio, com estes microssatélites (Jones & Searle 2015), estas amostras não puderam ser comparadas com as amostras utilizadas neste estudo, pois após a repetição da análise de alguns indivíduos genotipados em Jones & Searle (2015) obtiveram-se tamanhos de alelos diferentes para os mesmos microssatélites.

A análise conjunta de vários marcadores é bastante complexa. Os resultados obtidos, em vários estudos, com os diferentes marcadores utilizados em filogeografia, têm revelado incongruências entre si, nomeadamente entre marcadores nucleares autossómicos e o mtDNA (Förster *et al.* 2009; Jones *et al.* 2010; Gabriel *et al.* 2013) e entre o mtDNA e marcadores do cromossoma Y (Boursot *et al.* 1993; Jones & Searle 2015). As primeiras podem ser explicadas por vários factores, tal como o diferente tamanho efetivo populacional do mtDNA quando comparado com o genoma nuclear, o efeito relativo da deriva genética nos dois e, também, a taxa mutacional, muito mais baixa em marcadores diploides (Tosi *et al.* 2003; Zink & Barrowclough 2008). As incongruências provenientes da análise dos marcadores de herança uni-parental, cromossoma Y e mtDNA, podem ser o resultado de diferentes capacidades de dispersão dos dois sexos e diferentes capacidades de introgressão em populações estabelecidas (Jones & Searle 2015). Assim, estas diferenças de resultados entre marcadores não devem ser abordadas directamente como incongruências mas sim como diferentes perspectivas de evolução. Consoante o marcador analisado esta “incongruência” é consequência de uma diferente taxa evolutiva e/ou tipo de herança dos marcadores, sendo que a sua análise permite aceder a uma diferente parte da história evolutiva. Desta forma, reforça-se a necessidade de analisar diferentes

marcadores moleculares de modo a permitir uma avaliação mais rigorosa e menos enviesada da história da colonização das espécies em estudo (Rokas *et al.* 2003; Brito *et al.* 2009;). O estudo de Tosi *et al.* (2003) apresenta um exemplo de uma abordagem multilocus em que foram utilizados marcadores do mtDNA, cromossoma Y e genes autossómicos para aceder aos padrões evolutivos do género *Macaca*, enfatizando a importância relativa de cada marcador nas descobertas e realçando que marcadores com diferentes taxas de evolução apresentam em alguns pontos incongruências, mas que apenas analisados como um todo podem aproximar-se da história evolutiva da espécie. Este ponto é especialmente importante para espécies com diferenças comportamentais e de dispersão entre sexos e/ou que tenham sofrido processos adaptativos. Outro estudo de Mcdevitt *et al.* 2009, mtDNA e microssatélites para inferir a origem de colonização da espécie *Sorex minutus* na Irlanda, com dados concordantes para os dois marcadores. O caso da sub-espécie *Mus musculus domesticus*, neste trabalho, não é exceção. Muitos dos estudos até agora publicados basearam-se maioritariamente em dados de mtDNA, que como já referido, apenas permitem avaliar a colonização das fêmeas, sendo portanto, uma avaliação incompleta da colonização da espécie.

Perspetivas futuras

Para aceder mais detalhadamente aos aspectos da colonização deste roedor, nas ilhas atlânticas de colonização portuguesa, é necessário continuar a investigação no sentido de aumentar a amostragem para as ilhas mais fracamente amostradas e amostrar igualmente as diversas potenciais fontes colonizadoras. Isto é válido principalmente para os microssatélites do cromossoma Y, onde é prioritário uma análise mais robusta em Portugal continental, que neste estudo contou com uma elevada falta de informação para alguns dos microssatélites. Estes dados são de elevada importância de modo a melhor compreender o papel de Portugal continental na colonização das ilhas descobertas e povoadas pelos portugueses, nomeadamente em relação à dispersão dos machos. Para realizar o estudo filogeográfico do arquipélago de Cabo Verde é essencial amostrar as restantes ilhas, com maior foco nas ilhas de maiores dimensão e com ocupação humana, inicialmente para o marcador mitocondrial, D-loop, para verificar se o sinal obtido na ilha de Santa Luzia referente à ilha da Madeira se mantém nas restantes ilhas do arquipélago. Da mesma forma, o estudo dos microssatélites para o cromossoma Y é igualmente importante para inferir a sua origem e aceder à informação de potenciais introgressões por parte de machos nas populações.

Quanto ao estudo da zona intrónica do gene *Ocr1* é necessária uma revisão ao seu estudo visto ter-se verificado ser pouco informativo devido à falta de variabilidade encontrada. No entanto, considera-se que o esforço empregado neste marcador justifica estender a sua análise pelo menos aos países do Norte da Europa que se considera poderem ter agido como colonizadores de algumas das ilhas atlânticas. É também razoável tentar aceder a uma maior parte da região intrónica deste marcador de forma a aceder a mais pontos de variação e verificar se esses identificam alguma estruturação não identificada neste estudo.

Conclusões finais

A sub-espécie *Mus musculus domesticus* vive em associação com o homem desde a sedentarização e acompanhou-o desde então, eliminando assim barreiras geográficas à sua dispersão. Seguir o seu processo de colonização em ilhas oceânicas não é fácil, dado o complexo padrão das movimentações humanas e à informação passível de ser extraída de cada marcador molecular. Ainda assim, as análises filogeográficas, conduzidas ao longo deste estudo, permitiram clarificar o processo de colonização no que respeita à contribuição diferencial de machos e fêmeas nos arquipélagos dos Açores e da Madeira e iniciar o estudo do vasto arquipélago de Cabo Verde.

O sinal do primeiro evento de colonização bem-sucedido parece ficar registado no D-loop sem sofrer grandes distúrbios com colonizações subsequentes, permitindo comparar os padrões obtidos com potenciais fontes colonizadoras. No entanto, os marcadores de transmissão não exclusivamente materna permitem ainda determinar possíveis colonizações posteriores através da detecção da introgressão de material genético, dando ênfase ao papel da contribuição diferencial dos sexos no padrão de colonização.

A análise destes marcadores tornou possível inferir um ou vários eventos de colonização para cada ilha. Com os dados não mitocondriais analisados, as ilhas de São Jorge e Santa Maria, no arquipélago dos Açores, apresentaram, consistentemente, diferenças em relação às potenciais fontes colonizadoras analisadas bem como em relação às ilhas adjacentes, indicando um evento de colonização único e diferente das restantes ilhas e entre si, que não sofreu alterações mesmo com o constante fluxo de rotas comerciais e de passageiros. Discutiram-se, ainda, as possíveis origens de colonização para as ilhas da Madeira, Porto Santo e Santa Luzia consoante os marcadores analisados, inferindo-se uma origem de colonização inicial com o mesmo sinal, a qual sofreu diferentes vagas de colonizações secundárias, em cada ilha, com uma origem diferente da primeira. Pelos dados obtidos, a ilha de Santa Luzia foi colonizada originalmente a partir do arquipélago da Madeira sofrendo posteriormente introgressão de machos com duas origens geográficas distintas.

Bibliografia

- Arnaud E, Philippe J, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591-1604.
- 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.
- Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA.
- Babiker H, Tautz D (2015) Molecular and phenotypic distinction of the very recently evolved insular subspecies *Mus musculus helgolandicus* ZIMMERMANN, 1953. *BMC Evolutionary Biology*, **15**, 160–174.
- Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.
- Bentthencourt F, Chaudhuri K (1998) *História da Expansão Portuguesa*, vol I, p. 145–153. Círculo de Leitores, Lisboa.
- 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, Catalan H, Boursot J, Auffray P, Britton-Davidian J (2010) Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization. *Proceedings of Royal Society B*, **278**, 1034–1043.
- Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F (1993) The evolution of house mice. *Annual Review of Ecology and Systematics*, **24**, 119–52.
- Brito PH, Edwards VS (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, **135**, 439-455.
- 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**, 432–442.
- Cucchi T, Vigne JD, Auffray JC (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.
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**, 435–445.
- 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.
- 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.

- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.
- Förster DW, Gündüz İ, Nunes A, Gabriel SI, Ramalinho M, 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.
- Förster DW, Mathias ML, Britton-Davidian J, Searle JB (2013) Origin of the chromosomal radiation of Madeiran house mice: a microsatellite analysis of metacentric chromosomes. *Heredity*, **110**, 380–388.
- Fu YX (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.
- Gabriel SI, Mathias ML, Searle JB (2013) Genetic structure of house mouse (*Mus musculus* Linnaeus, 1758) populations in the Atlantic archipelago of the Azores: Colonization and dispersal. *Biological Journal of the Linnean Society*, **108**, 929–940.
- Gabriel SI, Mathias ML, Searle JB (2015) Of mice and the “Age of Discovery”: The complex history of colonization of the Azorean archipelago by the house mouse (*Mus musculus*) as revealed by mitochondrial DNA variation. *Journal of Evolutionary Biology*, **28**, 130–145.
- Geraldes A, Basset P, Gibson B, Kimberly LS, Bettina H, Hon Tsen Y, Nina B, Yaron Z, Michael WN (2008) Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology*, **17**, 5349–5363.
- 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 JC, Britton-Davidian J, Catalan J, Ganem G, Ramalinho MG, Mathias ML, Searle JB (2001) Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology*, **10**, 2023–2029.
- Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.
- Hardouin EA, Chapuis J-L, Stevens MI, Van V, Bettine J, 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.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

- 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, Jensen JK, Magnussen E, Gregersen N, Hansen HS, Searle JB (2011a) A molecular characterization of the charismatic Faroe house mouse. *Biological Journal of the Linnean Society*, **102**, 471–482.
- Jones EP, Jóhannesdóttir F, Gündüz İ, Richards MB, Searle JB (2011b) The expansion of the house mouse into north-western Europe. *Journal of Zoology*, **283**, 257–268.
- Jones E, Skirnisson K, MCGovern T, Gilbert M, 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, Searle JB (2015) Differing Y chromosome versus mitochondrial DNA ancestry, phylogeography, and introgression in the house mouse. *Biological Journal of the Linnean Society*, **115**, 348–361.
- Kelley Jr JE (1979) Non-Mediterranean influences that shaped the Atlantic in the early Portolan charts. *Imago Mundi*, **31**, 18–35.
- Kivelä M, Arnaud-Haond S, Saramäki J (2015) EDENetworks: A user-friendly software to build and analyse networks in biogeography, ecology and population genetics. *Molecular Ecology Resources*, **15**, 117–122.
- Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Lippens C, Estoup A, Hima MK, Loiseau A, Tatard C, Dalecky A, Bâ K, Kane M, Diallo M, Sow A, Niang Y, Piry S, Berthier K, Leblois R, Duplantier J-M, Brout C (2017) Genetic structure and invasion history of the house mouse (*Mus musculus domesticus*) in Senegal, West Africa: a legacy of colonial and contemporary times. *Heredity*, **18**, 1–12.
- Matschiner M, Salzburger W (2009) TANDEM: Integrating automated allele binning into genetics and genomics workflows. *Bioinformatics*, **25**, 1982–1983.
- Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortíz-Barrientos D, García LF, Velez ID, Bravo ML, Múnera JG, Bedoya G, Bortolini MC, Ruiz-Linares A (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *American Journal of Human Genetics*, **67**, 1277–1286.
- Mcdevitt AD, Rambau RV, O'Brien J, Mcdevitt CD, Hayden TJ, Searle JB (2009) Genetic variation in Irish pygmy shrews *Sorex minutus* (Soricomorpha: Soricidae): implications for colonization history. *Biological Journal of the Linnean Society*, **97**, 918–927.
- Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 409–421.
- Oliveira Marques AH (1973) *História de Portugal*, vol I, p. 263–264. Editora Àgora, Lisboa.

- Perez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martinez-Arias R, Clarimon J, Fiori G, Luiselli D, Facchini F, Pettener D, Bertranpetit J (1999) Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *American Journal of Human Genetics*, **65**, 208–219.
- Phifer-Rixey M, Nachman MW (2015) Insights into mammalian biology from the wild house mouse *Mus musculus*. *eLife*, **4**, 1–13.
- Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Rando JC, Pieper H, Alcover JA (2014) Radiocarbon evidence for the presence of mice on Madeira Island (North Atlantic) one millennium ago. *Proceedings of the Royal Society B*, **281**, 20133126.
- Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, **425**, 798–804.
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Searle JB (2008) The genetics of mammalian invasions: A review. *Wildlife Research*, **35**, 185–192.
- Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones E, Herman J, Rambau R, Noble L, Berry R, Gimenez M, Johannesdottir F (2009) Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society*, **276**, 201–207.
- Shamjana U, Bharadwaj T, Grace T (2015) Microsatellites: a Versatile Marker for Genetic / Evolutionary / Ecological Studies. *International Journal of Advanced Biological Research*, **5**, 86–95.
- 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**, 512–26.
- Tosi AJ, Morales JC, Melnick DJ (2003) Paternal, Maternal, and biparental molecular markers provide unique windows onto evolutionary history of Macaque monkeys. *Evolution*, **57**, 1419–1435.
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen J (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, **35**, 71–74.
- Weissbrod L, Marshall FB, Valla FR, Khalaily H, Bar-Oz G, Auffray JC, Vigne JD, Cucchi T (2017) Origins of house mice in ecological niches created by settled hunter-gatherers in the Levant 15,000 y ago. *Proceedings of the National Academy of Sciences*, **114**, 4099–4104.
- Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107–2121.

Anexos

Anexo I – Quadro resumo de todas as amostras de *Mus musculus domesticus* utilizadas e análises laboratoriais realizadas por população/ilha. As amostras cujo sexo foi determinado através do protocolo de sexagem estão assinaladas com um asterisco.

	ID População	ID Amostra	Sexo	Extração DNA	mtDNA	Microssatélites – Crom. Y					Crom. X
					d-loop	Y6	Y12	Y21	Y23	Y24	Ocr1
Arquipélago dos Açores	Santa Maria (n=20)	SMA 2	M	X		X	X	X	X	X	X
		SMA 3	M	X		X	X	X	X	X	
		SMA 7	M	X		X	X	X	X	X	
		SMA 13	M	X		X	X	X	X	X	
		SMA 14	M	X		X	X	X	X	X	
		SMA 16	F	X							X
		SMA 21	M	X		X	X	X	X	X	
		SMA 22	M	X		X	X	X	X	X	
		SMA 24	F	X							X
		SMA 25	M	X		X	X	X	X	X	
		SMA 26	M	X		X	X	X	X	X	
		SMA 30	M	X		X	X	X	X	X	
		SMA 32	M	X		X	X	X	X	X	X
		SMA 33	M	X		X	X	X	X	X	
		SMA 34	M	X		X	X	X	X	X	X
		SMA 35	F	X							X
		SMA 36	F	X							X
		SMA 37	M	X		X	X	X	X	X	
		SMA 38	M	X		X	X	X	X	X	
		SMA 40	M	X		X	X	X	X	X	X
São Miguel (n=43)	SM3	M	X		X	X	X	X	X		
	SM9	M	X		X	X	X	X	X	X	
	SM33	M	X		X	X	X	X	X	X	
	SM57	M	X		X	X	X	X	X		
	SM67	?	X							X	
	SM70	M	X		X	X	X	X	X		
	SM83	M	X		X	X	X	X	X		
	SM95	M	X		X	X	X	X	X	X	
	SM109	F	X							X	
	SM110	M	X		X	X	X	X	X		
	SM113	M	X		X	X	X	X	X		

	ID População	ID Amostra	Sexo	Extração DNA	mtDNA	Microsatélites – Crom. Y					Crom. X	
					d-loop	Y6	Y12	Y21	Y23	Y24	Ocrl	
Arquipélago dos Açores (continuação)	São Miguel (n=43) continuação	SM143	M	X		X	X	X	X	X		
		SM165	M	X		X	X	X	X	X		
		SM170	M	X		X	X	X	X	X		
		SM181	F	X								X
		SM182	M	X		X						
		SM188	F	X								X
		SM192	M	X		X	X	X	X	X		
		SM201	M	X		X	X	X	X	X		X
		SM213	M	X		X	X	X	X	X		
		SM230	M	X		X	X	X	X	X		
		SM247	F	X								X
		SM275	M	X		X	X	X	X	X		
		SM287	M	X		X	X	X	X	X		X
		SM306	M	X		X	X	X	X	X		
		SM329	M	X		X	X	X	X	X		
		SM348	M	X		X	X	X	X	X		
		SM350	M	X		X		X	X	X		
		SM362	M	X		X		X	X	X		
		SM376	M	X		X		X	X	X		
		SM383	M	X		X	X	X	X	X		
		SM405	M	X		X		X	X	X		
		SM406	M	X		X	X	X	X	X		
		SM415	M	X		X		X	X	X		X
		SM442	M	X		X	X	X	X	X		
		SM459	M	X		X	X	X	X	X		
		SM472	M	X		X	X	X	X	X		
		SM494	M	X		X	X	X	X	X		
	SM501	F									X	
	SM514	F	X									
	SM519	F	X									
	SM533	F	X									
	SM545	F	X									
	Terceira (n=27)	T145	M	X				X	X	X		
T156		M	X				X	X	X			
T170		M	X			X	X	X	X			
T172		F	X							X		

	ID População	ID Amostra	Sexo	Extração DNA	mtDNA d-loop	Microssatélites – Crom. Y					Crom. X		
						Y6	Y12	Y21	Y23	Y24	Ocr1		
Arquipélago dos Açores (continuação)	Terceira (n=27) continuação	T174	M	X			X	X	X	X			
		T 175	F	X							X		
		T194	M	X					X	X	X		
		T195	M	X				X	X	X	X		
		T197	M	X				X	X	X	X		
		T198	M	X			X	X	X	X	X	X	
		T241	M	X			X	X	X	X	X		
		T258	M	X				X	X	X	X	X	
		T259	M	X				X	X	X	X		
		T 273	F	X								X	
		T294	M	X									
		T304	M	X			X	X	X	X	X	X	
		T307	M	X			X	X	X	X	X		
		T330	M	X			X	X	X	X	X	X	
		T332	M	X			X	X	X	X	X		
		T355	M	X			X	X	X	X	X		
		T362	M	X			X	X	X	X	X	X	
		T391	M	X				X	X	X	X		
		T397	M	X				X	X	X	X		
		T402	M	X			X	X	X	X	X		
		T 403	F	X								X	
		T413	M	X				X	X	X	X		
		T 426	F	X								X	
		Graciosa (n=10)	GR 3	M	X			X	X	X	X	X	X
			GR 6	M	X			X	X	X	X	X	X
			GR 7	M	X			X		X	X	X	X
			GR 11	F	X								
GR 12	F		X								X		
GR 18	M		X			X	X	X	X	X	X		
GR 21	M		X			X	X	X	X	X			
GR 24	F		X										
GR 25	F		X								X		
GR 22	M		X			X	X	X	X	X	X		

	ID População	ID Amostra	Sexo	Extração DNA	mtDNA d-loop	Microsatélites – Crom. Y					Crom. X
						Y6	Y12	Y21	Y23	Y24	Ocr1
Arquipélago dos Açores (continuação)	São Jorge (n=9)	SJ 1	M	X		X	X	X	X	X	X
		SJ 8	F	X							X
		SJ 12	M	X		X	X	X	X	X	X
		SJ 13	M	X		X	X	X	X	X	
		SJ 20	M	X		X	X	X	X	X	
		SJ 29	F	X							X
		SJ 31	M	X		X	X	X	X	X	X
		SJ 33	M	X		X	X	X	X	X	X
		SJ 39	F	X							X
	Pico (n=19)	P 1	F	X							X
		P 2	F	X							X
		P 3	M	X		X	X	X	X	X	X
		P 5	M	X		X	X	X	X	X	
		P 6	M	X		X	X	X	X	X	
		P 8	M	X		X	X	X	X	X	
		P 9	M	X		X	X	X	X	X	
		P10	M	X		X	X	X	X	X	X
		P 14	M	X		X	X	X	X	X	
		P 16	M	X		X	X	X	X	X	
		P 17	M	X			X	X	X	X	
		P 18	M	X		X	X	X	X	X	X
		P 19	F	X							X
		P 23	M	X		X	X	X	X	X	
		P 25	M	X		X	X	X	X	X	X
		P 29	M	X		X	X	X	X	X	
		P 30	M	X		X	X	X	X	X	
	P 33	M	X		X	X	X	X	X		
	P 35	F	X							X	
	Faial (n=13)	F 1	M	X		X	X	X	X	X	X
		F 4	M	X		X	X	X		X	
		F 6	M	X			X	X	X	X	X
		F 7	F	X							X
		F 8	M	X		X	X	X	X	X	
		F 12	?	X							X
		F 13	M	X		X	X	X	X	X	

	ID População	ID Amostra	Sexo	Extração DNA	mtDNA	Microssatélites – Crom. Y					Crom. X
					D-loop	Y6	Y12	Y21	Y23	Y24	Ocr1
Arquipélago dos Açores (continuação)	Faial (n=13) continuação	F 14	M	X		X	X	X	X	X	
		F 18	M	X		X	X	X	X	X	
		F 21	M	X		X		X	X	X	
		F 22	M	X		X	X	X	X	X	
		F 23	M	X		X	X	X	X	X	X
		F 26	M	X		X	X	X	X	X	X
		F 28	F	X							X
		F 30	M	X		X	X	X	X	X	X
		F 32	M	X		X	X	X	X	X	
		F 35	M	X		X	X	X	X	X	X
	F 36	F	X							X	
	Flores (n=16)	FL 1	F	X							X
		FL 2	F	X							X
		FL 3	M	X		X		X	X	X	
		FL 4	M	X		X		X	X	X	
		FL 6	M	X		X				X	X
		FL 10	M	X			X	X	X	X	X
		FL 11	M	X		X		X	X	X	
		FL 12	M	X		X	X	X	X	X	
		FL 13	M	X		X		X	X	X	
		FL 14	M	X		X		X	X	X	
		FL 16	M	X		X		X	X	X	
		FL 17	M	X		X		X	X	X	
		FL 18	M	X		X	X	X	X	X	X
		FL 23	M	X		X	X	X	X	X	X
	FL 26	M	X		X	X	X	X	X		
	FL 27	F	X							X	
	Corvo (n=5)	CR1	M*	X							
		CR 3	F*	X							X
		CR4	M*	X							
		CR 5	F*	X							X
		CR7	M*	X							
	Arquipélago da Madeira	Madeira (n=33)	P5.2	M	X	X					
P9.2			M	X	X						

	ID População	ID Amostra	Sexo	Extração DNA	mtDNA	Microssatélites – Crom. Y					Crom. X
					D-loop	Y6	Y12	Y21	Y23	Y24	Ocr1
		P10.2	M	X	X	X	X	X	X	X	X
		P13.2	M	X	X						
		P17.1	M	X	X	X	X	X	X	X	
		P19.6	M	X	X	X	X	X	X	X	
		P20.2	M	X	X	X	X	X	X	X	X
		P22.4	M	X	X	X	X	X	X	X	X
		P24.2	M	X	X						
		P25.3	M	X	X						
		P30.10	F	X							X
		P31.4	M	X	X	X	X	X	X	X	X
		P33B.5	M	X	X	X	X	X	X	X	
		P34.3	F	X							X
		P34.7	M	X	X	X	X	X	X	X	
		P36.9	M	X	X						
		P38.1	M	X	X			X	X	X	
		P39.2	M	X	X	X	X	X	X	X	
		P43.2	M	X	X	X	X	X	X	X	X
		P47.1	M	X	X	X	X	X	X	X	
		P50.3	M	X	X	X	X	X	X	X	
		P51.5	M	X	X	X	X	X	X	X	X
		P52.5	M	X	X	X	X	X	X	X	
		P54.2	M	X	X	X	X	X	X	X	X
		P55.3	M	X	X	X	X			X	
		P64.1	M	X	X	X	X	X	X	X	
		P65.1	M	X	X						
		P66.6	F	X							X
		P67.3	M	X	X			X			
		P70.6	M	X	X	X	X	X	X	X	
		P72.4	M	X	X	X	X	X	X	X	
		P75.1	M	X	X						
		FuncTal1	M	X	X	X	X	X	X	X	

	ID População	ID Amostra	Sexo	Extração DNA	mtDNA	Microsatélites – Crom. Y					Crom. X	
					D-loop	Y6	Y12	Y21	Y23	Y24	Ocr1	
Arquipélago da Madeira (continuação)	Porto Santo (n=15)	PS1.7	M	X	X	X	X	X	X	X	X	X
		PS1.8	M	X	X	X	X	X		X		
		PS1.9	M	X	X	X	X	X	X	X		
		PS2.5	F	X	X							X
		PS2.7	M	X	X	X	X	X	X	X	X	X
		PS2.9	M	X	X	X	X	X	X	X		
		PS2.10	?	X	X							
		PS2.14	M	X	X	X	X	X		X		
		PS3.2	M	X	X	X	X	X	X	X	X	
		PS3.3	M	X	X	X	X	X	X	X	X	X
		PS3.4	M	X	X	X	X	X	X	X	X	
		PS4.5	M	X	X	X	X	X	X	X	X	
		PS4.7	F	X								X
		PS4.9	M	X	X	X	X		X	X		
		PS4.12	M	X	X	X	X	X	X	X	X	X
Arquipélago de Cabo Verde	Santa Luzia (n=29)	SL1.9	F*	X								
		SL1.10	M*	X	X	X	X	X	X	X		
		SL1.14	F*	X	X							
		SL1.15	F*	X	X							
		SL1.16	F*	X	X							
		SL1.17	F*	X	X							X
		SL1.18	M*	X	X	X	X	X	X	X		
		SL1.19	F*	X	X							
		SL1.20a	M*	X	X	X	X	X	X	X		
		SL1.20b	M*	X	X	X	X	X	X	X		
		SL1.21	M*	X	X			X	X	X		
		SL1.22	M*	X	X	X	X	X	X	X	X	
		SL1.23	F*	X	X							
		SL1.24	F*	X	X							
		SL1.25	M*	X	X	X	X	X	X	X	X	X
		SL1.30	M*	X	X	X	X	X	X	X	X	
SL2.11	M*	X	X	X	X	X	X	X	X			
SL2.14	F*	X	X									

	ID População	ID Amostra	Sexo	Extração DNA	mtDNA	Microssatélites – Crom. Y					Crom. X
					D-loop	Y6	Y12	Y21	Y23	Y24	Ocr1
Arquipélago de Cabo Verde (continuação)	Santa Luzia (n=29) (continuação)	SL3.5	M*	X	X		X	X	X	X	
		SL3.12	M*	X	X	X	X	X	X	X	X
		SL3.13	M*	X	X	X	X	X	X	X	
		SL3.14	F*	X	X						X
		SL3.18	F*	X	X						
		SL4.1	F*	X	X						
		SL4.2	M*	X	X	X	X	X	X	X	
		SL4.3	F*	X	X						
		SL58.17	M*	X	X	X	X	X	X	X	X
		SL61.15	M*	X	X	X	X	X	X	X	
		SL75.16	M*	X	X	X	X	X	X	X	
Portugal continental	Portugal (n=14)	Sines49	M	X			X	X	X	X	X
		VFXira60 C	M	X			X	X	X	X	X
		Porto68	M	X			X	X	X	X	X
		VCastelo6 4	M	X			X	X	X	X	X
		Lisboa26 C	M	X		X	X	X	X	X	X
		PenicTe69	M	X		X		X	X	X	X
		PenicTe70	M	X				X	X	X	
		Tavira81	M	X				X	X	X	X
		Portimão8 7	F	X							X
		VNMFontes51	F	X							X
		Lagoa22	F	X							X
		FigFoz57	F	X							X
		Setúbal35	F	X							X
		Elvas1.1	F	X							X
Espanha	Espanha (n=1)	Cáceres1	F	X						X	

Anexo II - *Primers* utilizados neste estudo para cada marcador (d-loop, cromossoma Y, *Ocr1*).
 Descrição da sequência e referência correspondente. *Primers* utilizados para cada marcador.

Marcadores	Primers	Label	Sequência	Referência
D-loop	<i>Forward</i> L15774	-	5' TGA ATT GGA GGA CAA CCA GT3'	Searle <i>et al.</i> 2009
	<i>Reverse</i> H2228	-	5'TTA TAA GGC CAG GAC CAA AC3'	
Y6	<i>Forward</i>	FAM	5' – AACCACCACTATCTTCATTC '3	Hardouin <i>et al.</i> 2010)
	<i>Reverse</i>	-	5' – ACAGAGTATACGTACGTGTG '3	
Y12	<i>Forward</i>	FAM	5' – CCCAATCTAGGCATTTAATT '3	
	<i>Reverse</i>	-	5' – ATTCACCATTCTCCAGTGTG '3	
Y21	<i>Forward</i>	HEX	5' – ACCATCAGATGATCACCAAGTGC '3	
	<i>Reverse</i>	-	5' – TCCAGCATTCAATGGTACAGGCT '3	
Y23	<i>Forward</i>	FAM	5' – ACCTCACTCAGGATGATGCCCTC '3	
	<i>Reverse</i>	-	5' – AGCCTGTGCGCACGTGTGTG '3	
Y24	<i>Forward</i>	HEX	5' – TCTGGGGGTTTCGGGTGGAGCCT '3	
	<i>Reverse</i>	-	5' – GCATCACAGCTGAGGCTCTGTGG '3	
<i>Ocr1</i>	<i>Forward</i>	-	5' TTG ACC AAA TTG AAG GGT GA 3'	Desenhados para este estudo
	<i>Reverse</i>	-	5' CTC TTC TCT TTC GGC ACA CA '3	

Anexo III– Dados de microssatélites (cromossoma Y) corrigidos, no programa TANDEM, para a sub-espécie *Mus musculus domesticus* nas ilhas atlânticas de colonização portuguesa.

	ID População	ID amostra	Y6	Y12	Y24	Y21	Y23
Arquipélago dos Açores	Santa Maria (N=16)	SMA 2	118	128	398	321	308
		SMA 3	118	128	398	321	310
		SMA 7	120	124	398	321	308
		SMA 13	120	128	398	321	308
		SMA 14	120	128	398	321	308
		SMA 21	120	124	398	321	308
		SMA 22	118	128	398	321	308
		SMA 25	120	128	398	321	308
		SMA 26	120	128	398	321	308
		SMA 30	120	128	398	321	308
		SMA 32	118	128	398	321	308
		SMA 33	118	128	398	321	308
		SMA 34	120	128	400	319	308
		SMA 37	120	128	398	321	308
		SMA 38	120	128	398	321	308
		SMA 40	120	132	398	323	308
	São Miguel (N=32)	SM3	122	124	392	325	316
		SM9	124	138	404	313	308
		SM33	122	124	392	321	320
		SM57	126	138	404	313	308
		SM70	122	124	398	321	322
		SM83	122	124	398	321	324
		SM95	122	126	402	325	326
		SM110	124	138	404	313	308
		SM113	124	138	404	313	308
		SM143	124	138	404	313	308
		SM165	130	138	404	313	308
		SM170	124	138	408	313	308
		SM192	122	124	398	321	328
		SM201	124	124	402	323	322
		SM213	122	124	392	321	324
		SM230	124	138	392	313	308
SM275	122	124	398	325	324		
SM287	122	124	400	321	324		
SM306	122	124	398	321	322		
SM329	122	124	398	323	326		
SM348	122	124	398	321	322		
SM350	122	-	398	321	326		
SM362	-	-	398	321	322		
SM376	126	-	404	313	308		
SM383	122	124	398	321	320		

	ID População	ID amostra	Y6	Y12	Y24	Y21	Y23
Arquipélago dos Açores (continuação)	São Miguel (N=32) continuação	SM405	-	120	400	317	300
		SM406	122	124	398	321	322
		SM415	-	124	398	321	322
		SM442	122	122	398	313	318
		SM459	122	124	398	325	320
		SM472	124	138	404	311	308
		SM494	122	124	398	321	320
	Terceira (N=21)	T145	-	-	398	321	322
		T156	-	-	400	321	322
		T170	-	126	398	321	320
		T174	-	124	398	321	322
		T194	-	-	400	321	322
		T195	-	124	400	321	322
		T197	-	124	398	321	322
		T198	124	122	398	321	322
		T241	126	124	398	321	322
		T258	122	124	398	321	322
		T259	-	124	398	321	324
		T304	124	124	398	321	324
		T307	122	124	398	321	322
		T330	122	124	398	323	322
		T332	122	124	398	321	320
		T355	122	124	398	321	324
		T362	122	124	400	321	322
		T391	-	124	382	321	322
		T397	-	124	398	321	322
		T402	122	124	382	321	322
	T413	-	124	398	321	316	
	Graciosa (N=6)	GR 3	122	128	398	319	322
		GR 6	122	124	398	319	322
		GR 7	124	-	398	319	322
		GR 18	122	124	398	321	316
		GR 21	122	-	398	319	322
GR 22		122	120	398	319	316	

	ID População	ID amostra	Y6	Y12	Y24	Y21	Y23
Arquipélago dos Açores (continuação)	São Jorge	SJ 1	116	136	400	319	302
		SJ 12	116	136	394	319	300
		SJ 13	116	136	394	319	300
		SJ 20	116	136	394	319	300
		SJ 31	116	136	400	319	302
		SJ 33	116	136	400	319	302
	Pico	P 3	122	124	398	321	322
		P5	122	124	398	321	322
		P6	122	124	398	321	322
		P8	122	124	396	325	322
		P9	122	124	398	323	320
		P10	124	124	396	325	322
		P 14	122	124	404	321	324
		P 16	122	124	398	321	322
		P 17	-	124	398	321	322
		P 18	122	124	404	321	322
		P 23	124	124	398	321	322
		P 25	122	124	398	319	322
		P 29	122	124	404	321	324
		P 30	122	124	398	323	320
		P 33	122	124	396	325	322
	Faial	F 1	122	124	394	323	322
		F 4	122	124	394	323	0
		F 6	-	124	398	321	324
		F 8	122	124	398	321	322
		F 13	116	136	398	319	304
		F 14	116	136	398	319	304
		F 18	116	136	398	319	304
		F 21	116	-	398	319	304
		F 22	116	136	398	319	304
		F 23	124	124	398	321	322
		F 26	122	124	398	321	322
		F 30	122	122	400	321	318
		F 32	122	122	400	319	320
		F 35	122	122	400	321	318

	ID População	ID amostra	Y6	Y12	Y24	Y21	Y23
Arquipélago dos Açores (continuação)	Flores	FL 3	122	-	400	323	322
		FL 4	122	122	400	323	322
		FL 6	122	-	400	-	-
		FL 10	-	124	398	319	322
		FL 11	122	-	402	321	322
		FL 12	122	124	402	321	322
		FL 13	122	-	398	321	324
		FL 14	122	-	398	321	324
		FL 16	122	-	398	321	324
		FL 17	122	-	398	321	324
		FL 18	122	124	398	321	324
		FL 23	122	124	398	321	324
		FL 26	122	128	400	323	322
Arquipélago da Madeira	Madeira	P10.2	122	134	394	317	310
		P17.1	120	136	400	317	302
		P19.6	120	136	402	319	308
		P20.2	122	132	402	321	306
		P22.4	122	132	396	321	308
		P31.4	126	130	398	321	306
		P33B.5	122	132	398	323	306
		P34.7	122	138	398	325	306
		P38.1	-	-	396	317	302
		P39.2	122	132	402	321	306
		P43.2	118	142	392	317	304
		P47.1	122	138	398	327	306
		P50.3	122	138	398	325	306
		P51.5	122	138	398	325	306
		P52.5	124	134	398	321	310
		P54.2	120	136	398	321	306
		P55.3	122	132	398	317	-
		P64.1	122	146	396	321	310
		P67.3	120	142	400	317	302
		P72.4	122	132	398	321	306
Funchal	122	144	392	321	326		

	ID População	ID amostra	Y6	Y12	Y24	Y21	Y23
Arquipélago da Madeira (continuação)	Porto Santo	PS1.7	120	130	408	321	318
		PS1.8	120	130	402	313	-
		PS1.9	120	128	404	313	316
		PS2.7	118	128	402	321	316
		PS2.9	120	128	402	321	316
		PS2.14	128	130	-	323	-
		PS3.2	120	130	404	321	316
		PS3.3	120	128	406	321	318
		PS3.4	120	128	406	321	318
		PS4.5	120	128	404	321	318
		PS4.9	118	-	-	321	318
		PS4.12	120	128	406	321	318
		ZR1 10	120	128	398	321	314
Arquipélago de Cabo Verde	Santa Luzia	SL1 18	120	128	398	321	314
		SL1 20a	120	128	398	321	314
		SL1 20b	120	128	398	321	314
		SL1 21	-	128	-	321	314
		SL1 22	120	128	398	321	314
		SL1 25	120	128	398	321	314
		SL1 30	120	128	398	321	314
		SL2 11	120	128	398	321	314
		SL3 5	-	136	398	317	302
		SL3 12	120	128	398	321	314
		SL3 13	120	136	398	317	302
		SL4 2	120	128	398	321	314
		SL58 17	120	136	398	317	302
		SL61 15	120	136	398	317	302
		SL75 16	120	136	398	317	302

	ID População	ID amostra	Y6	Y12	Y24	Y21	Y23
Portugal	Portugal	Sines 49	-	134	402	315	318
		VilaFrancaXira60 C	-	142	394	321	328
		Porto68	-	134	402	321	308
		VianaCastelo64	-	132	394	323	314
		Lisboa26C	120	126	398	323	300
		Peniche 69	120	-	398	321	328
		Peniche 70	-	-	398	321	330
		Tavira81	-	-	396	317	298

Anexo IV - Quadro de haplótipos para o marcador d-loop, amostrados em todas as ilhas do Arquipélago dos Açores e da Madeira, ilha de Santa Luzia (Cabo Verde), Portugal e Espanha, para a sub-espécie de *Mus musculus domesticus*. As amostras utilizadas neste estudo estão assinaladas com um asterisco.

	Clade	Haplótipo	N	Amostras (Ind1; Ind2;...)	Referência
Hap_1:	B	Turkey.34	10	São Miguel (N=10)	Gabriel <i>et al.</i> 2015
Hap_2:	B	Canary112	22	São Miguel (N=13), Santa Maria (N=2), Espanha (N=7)	
Hap_3:	B	AZORES.0 5	1	São Miguel (N=1)	
Hap_4:	B	AZORES.0 6	2	São Miguel (N=5)	
Hap_5:	-	1083.C.It	5	São Miguel (N=2)	
Hap_6:	E	1316.Engl	2	São Miguel (N=1), Portugal (N=1)	Gabriel <i>et al.</i> 2015; Förster <i>et al.</i> 2009
Hap_7:	B	AZORES.0 7	10	São Miguel (N=9), Pico (N=1)	Gabriel <i>et al.</i> 2015
Hap_8:	B	AZORES.0 8	1	São Miguel (N=1)	
Hap_9:	B	AZORES.0 9	12	São Miguel (N=4), Espanha (N=8)	
Hap_10:	C	AZORES.1 0	3	São Miguel (N=3)	
Hap_11:	B	Portugal.12	5	São Miguel (N=1), Portugal (N=4)	Gabriel <i>et al.</i> 2015; Förster <i>et al.</i> 2009
Hap_12:	D	AZORES.1 1	1	São Miguel (N=1)	Gabriel <i>et al.</i> 2015
Hap_13:	B	AZORES.1 2	1	São Miguel (N=1)	
Hap_14:	F	AZORES.0 1	13	Terceira (N=6), Santa Maria (N=7)	
Hap_15:	F	AZORES.0 2	21	Terceira (N=11), Santa Maria (N=9), Pico (N=1)	
Hap_16:	D	AZORES.1 3	1	Terceira (N=1)	
Hap_17:	D	AZORES.1 4	1	Terceira (N=1)	
Hap_18:	F	AZORES.0 3	8	Terceira (N=3), Santa Maria (N=5)	
Hap_19:	D	AZORES.1 5	5	Terceira (N=5)	
Hap_20:	D	AZORES.1 6	1	Terceira (N=1)	
Hap_21:	F	AZORES.1 7	1	Terceira (N=1)	
Hap_22:	D	AZORES.1 8	1	Terceira (N=1)	

	Clade	Haplótipo	N	Amostras (Ind1; Ind2;...)	Referência
Hap_23:	F	NTj1	1	Terceira (N=1)	Gabriel <i>et al.</i> 2015
Hap_24:	D	Psanto.2	57	São Jorge (N=24), Pico (N=7), Espanhas (N=3), Porto Santo (N=6), PS1_8*; PS2_5*; PS2_7*; PS2_9*; PS2_10*; PS2_14*; PS3_2*; PS3_3*; PS3_4*; SL1_19*; SL1_25*; SL2_11*; SL3_5*; SL3_12*; SL3_14*; SL3_18*; PS1_9*]	Gabriel <i>et al.</i> 2015; Presente estudo*
Hap_25:	D	AZORES.2 3	1	São Jorge (N=1)	Gabriel <i>et al.</i> 2015
Hap_26:	D	AZORES.2 4	1	São Jorge (N=1)	
Hap_27:	F	AZORES.0 4	1	Santa Maria (N=1)	
Hap_28:	C	Portugal.1	87	Pico (N=14), Flores (N=14), Faial (N=19), Corvo (N=7), Portugal (N=21), Espanha (N=11), P50_3*]	Gabriel <i>et al.</i> 2015; Förster <i>et al.</i> 2009; Presente estudo*
Hap_29:	D	AZORES.2 5	1	Pico (N=1)	Gabriel <i>et al.</i> 2015
Hap_30:	C	Canary113	12	Graciosa (N=12)	
Hap_31:	C	AZORES.1 9	1	Graciosa (N=1)	
Hap_32:	C	AZORES.2 0	3	Graciosa (N=3)	
Hap_33:	C	AZORES.2 1	6	Graciosa (N=6)	
Hap_34:	C	AZORES.2 2	1	Graciosa (N=1)	
Hap_35:	F	1334.Scot	3	Flores (N=3)	
Hap_36:	F	ETL11	1	Flores (N=1)	
Hap_37:	F	AZORES.2 8	6	Flores (N=5), Faial (N=1)	
Hap_38:	C	AZORES.2 6	1	Faial (N=1)	
Hap_39:	C	AZORES.2 7	1	Faial (N=1)	
Hap_40:	D	Made.2	25	Faial (N=5), Madeira (N=17), P24_2*; P52_5*; P75_1*]	
Hap_41:	C	Portugal.7	4	Corvo (N=2), Portugal (N=2)	Gabriel <i>et al.</i> 2015
Hap_42:	C	Portugal.2	11	Portugal (N=11)	Förster <i>et al.</i> 2009
Hap_43:	C	Portugal.3	2	Portugal (N=2)	
Hap_44:	C	Portugal.4	6	Portugal (N=6)	
Hap_45:	C	Portugal.5	5	Portugal (N=5)	
Hap_46:	C	Portugal.6	3	Portugal (N=3)	
Hap_47:	C	Portugal.8	2	Portugal (N=2)	
Hap_48:	C	1199.Gree	3	Portugal (N=3)	
Hap_49:	C	U47444.16	3	Portugal (N=3)	

	Clade	Haplótipo	N	Amostras (Ind1; Ind2;...)	Referência
Hap_82:	D	MADE.16	2	Madeira (N=2)	Förster <i>et al.</i> 2009
Hap_83:	D	MADE.17	2	Madeira (N=2)	
Hap_84:	D	MADE.18	6	Madeira (N=2), SL1_14*; SL1_21*; SL1_30*; SL58_17*]	Förster <i>et al.</i> 2009; Presente estudo*
Hap_85:	D	MADE.19	2	Madeira (N=2)	Förster <i>et al.</i> 2009
Hap_86:	C	MADE.20	1	Madeira (N=1)	
Hap_87:	D	MADE.21	1	Madeira (N=1)	
Hap_88:	D	MADE.22	4	Madeira (N=2), P5_2; P38_1*]	Förster <i>et al.</i> 2009; Presente estudo*
Hap_89:	D	MADE.23	2	Madeira (N=2)	Förster <i>et al.</i> 2009
Hap_90:	D	MADE.24	1	Madeira (N=1)	
Hap_91:	D	MADE.25	7	Madeira (N=6), P43_2*]	Förster <i>et al.</i> 2009; Presente estudo*
Hap_92:	D	MADE.26	1	Madeira (N=1)	Förster <i>et al.</i> 2009
Hap_93:	D	MADE.27	1	Madeira (N=1)	
Hap_94:	D	MADE.28	2	Madeira (N=2)	
Hap_95:	D	MADE.29	3	Madeira (N=3)	
Hap_96:	D	MADE.30	1	Madeira (N=1)	
Hap_97:	D	MADE.31	1	Madeira (N=1)	
Hap_98:	D	MADE.32	1	Madeira (N=1)	
Hap_99:	D	PSanto.1	2	Porto Santo (N=2)	
Hap_100:	D	PSanto.3	1	Porto Santo (N=1)	
Hap_101:	D	PSanto.4	2	Porto Santo (N=1); PS4_5*]	
Hap_102:	B	StoAntão.1	1	[CapeVerde.1_StoAntao]	Dados não publicados
Hap_103:	D	SNicolau.1	1	[CapeVerde.2_SNicolau]	
Hap_104:	C	SPAIN13	1	[BDZ1_2*]	Presente estudo*
Hap_105:	E	SPAIN14	1	[BDZ1_8*]	
Hap_106:	C	SPAIN15	1	[BDZ2_2*]	

	Clade	Haplótipo	N	Amostras (Ind1; Ind2;...)	Referência
Hap_107:	D	MADE.33	1	[Funchal1*]	Presente estudo*
Hap_108:	D	MADE.34	2	[P9_2*; P31_4*]	
Hap_109:	D	MADE.35	1	[P13_2*]	
Hap_110:	D	MADE.36	1	[P17_1*]	
Hap_111:	D	MADE.37	1	[P25_3*]	
Hap_112:	D	MADE.38	1	[P47_1*]	
Hap_113:	D	MADE.39	1	[P51_5*]	
Hap_114:	D	U47461.33	1	[P64_1*]	
Hap_115:	D	MADE.40	1	[P65_1*]	
Hap_116:	D	MADE.41	1	[P70_6*]	
Hap_117:	D	PSanto.5	1	[PS1_7*]	
Hap_118:	D	PSanto.6	1	[PS4_9*]	
Hap_119:	C	MADE.42	1	[P22_4*]	
Hap_120:	D	MADE.43	1	[P72_4*]	
Hap_121:	D	PSanto.7	1	[PS4_12*]	

Anexo V - Quadro de haplótipos obtidos para os microssatélites do cromossoma Y para os indivíduos da sub-espécie *Mus musculus domesticus* amostrados em todas as ilhas do Arquipélago dos Açores (exceto a ilha do Corvo), Madeira, Porto Santo, Santa Luzia (Cabo Verde) e Portugal continental.

Haplótipo	Ocorrência	N	ID Amostra
122124392325316	São Miguel	1	SM3
124138404313308	São Miguel	4	SM9; SM110; SM113; SM143
122124392321320	São Miguel	1	SM33
126138404313308	São Miguel	1	SM57
122124398321322	São Miguel Terceira Pico Faial	12	SM70; SM306; SM348; SM406; T258; T307; P3; P5; P6; P16; F8; F26
122124398321324	São Miguel Terceira Flores	4	SM83; T355; FL18; FL23
122126402325326	São Miguel	1	SM95
130138404313308	São Miguel	1	SM165
124138408313308	São Miguel	1	SM170
122124398321328	São Miguel	1	SM192
124124402323322	São Miguel	1	SM201
122124392321324	São Miguel	1	SM213
124138392313308	São Miguel	1	SM230
122124398325324	São Miguel	1	SM275
122124400321324	São Miguel	1	SM287
122124398323326	São Miguel	1	SM329
122124398321320	São Miguel Terceira	3	SM383; SM494; T332
122122398313318	São Miguel	1	SM442
122124398325320	São Miguel	1	SM459
124138404311308	São Miguel	1	SM472
124122398321322	Terceira	1	T198
126124398321322	Terceira	1	T241
124124398321324	Terceira	1	T304
122124398323322	Terceira	1	T330
122124400321322	Terceira	1	T362
122124382321322	Terceira	1	T402
116136400319302	São Jorge	3	SJ1; SJ31; SJ33
116136394319300	São Jorge	3	SJ12; SJ13; SJ20
118128398321308	Santa Maria	4	SMA2; SMA22; SMA32; SMA33

Haplótipo	Ocorrência	N	ID Amostra
118128398321310	Santa Maria	1	SMA3
120124398321308	Santa Maria	2	SMA7; SMA21
120128398321308	Santa Maria	7	SMA13; SMA14; SMA25; SMA26; SMA30; SMA37; SMA38
120128400319308	Santa Maria	1	SMA34
120132398323308	Santa Maria	1	SMA40
122124396325322	Pico	2	P8; P33
122124398323320	Pico	2	P9; P30
124124396325322	Pico	1	P10
122124404321324	Pico	2	P14 P29
122124404321322	Pico	1	P18
124124398321322	Pico/Faial	2	P23; F23
122124398319322	Pico Graciosa	2	P25; GR6
122124394323322	Faial	1	F1
116136398319304	Faial	4	F13; F14; F18; F22
122122400321318	Faial	2	F30; F35
122122400319320	Faial	1	F32
122122400323322	Flores	1	FL4
122124402321322	Flores	1	FL12
122128400323322	Flores	1	FL26
122128398319322	Graciosa	1	GR3
122124398321316	Graciosa	1	GR18
122120398319316	Graciosa	1	GR22
122134394317310	Madeira	1	P10.2
120136400317302	Madeira	1	P17.1
120136402319308	Madeira	1	P19.6
122132402321306	Madeira	2	P20.2; P39.2
122132396321308	Madeira	1	P22.4
126130398321306	Madeira	1	P31.4
122132398323306	Madeira	1	P33B.5
122138398325306	Madeira	3	P34.7; P50.3; P51.5
118142392317304	Madeira	1	P43.2
122138398327306	Madeira	1	P47.1
124134398321310	Madeira	1	P52.5
120136398321306	Madeira	1	P54.2
122146396321310	Madeira	1	P64.1
120142400317302	Madeira	1	P67.3

Haplótipo	Ocorrência	N	ID Amostra
122132398321306	Madeira	1	P72.4
122144392321326	Madeira	1	FuncT1
120130408321318	Porto Santo	1	PS1.7
120128404313316	Porto Santo	1	PS1.9
118128402321316	Porto Santo	1	PS2.7
120128402321316	Porto Santo	1	PS2.9
120130404321316	Porto Santo	1	PS3.2
120128406321318	Porto Santo	3	PS3.3; PS3.4; PS4.12
120128404321318	Porto Santo	1	PS4.5
120128398321314	Santa Luzia	10	SL1_10; SL1_18; SL120a; SL120b; SL1_22; SL1_25; SL1_30; SL2_11; SL3_12; SL4_2;
120136398317302	Santa Luzia	4	SL3_13; SL58_17; SL61_15; SL75_16
120126398323300	Portugal	1	Lisboa26C

Anexo VI - Quadro de haplótipos do gene *Ocr1* obtidos para os indivíduos da sub-espécie *Mus musculus domesticus* amostrados em todas as ilhas do Arquipélago dos Açores, Madeira, Porto Santo, Santa Luzia (Cabo Verde) e Portugal continental (assinalados com um asterisco). Foram também incluídos os haplótipos já publicados num estudo anterior, originários de Espanha, Israel, Itália, Grécia, Reino Unido (Gerald *et al.* 2008).

Haplótipo	Ocorrência	N	ID Amostra (Ind1;Ind2;...)	Referência
Ocr1_1:	Santa Maria/São Miguel/Terceira/Graciosa/São Jorge/Pico/Faial/Flores/Corvo/Madeira/Porto Santo/Santa Luzia/Portugal/Espanha/Itália/Israel	66	[CR3*; CR5*; F1*; F12*; F26*; F30*; FL2*; FL6*; FL10*; FL18*; FL23*; Elvas1*; Cáceres1*; GR3*; GR6*; GR7*; T304*; T330*; T332*; T362*; T426*; IS13; IT4; IT11; IT21; IT22; Lagoa22*; Lisboa26*; M54.2*; M74.1*; P2*; P3*; PenicTe69*; Portimão87*; Porto68*; PS1_7*; PS2_5*; PS2_7*; PS3_3*; PS4_12*; SM9*; SM33*; SM95*; SM201*; SM247*; SM287*; SM415*; Sines49*; SJ1*; SJ8*; SJ12*; SJ29*; SJ31*; SJ33*; SJ39*; SMA40*; SP2; SP3; SP4 ; SP5; SP7; SP8; SP9; Tavira81*; VilaFrancaXira60*; SL3_12*]	Gerald <i>et al.</i> 2008; Presente estudo*
Ocr1_2:	Santa Maria/São Miguel/Terceira/Graciosa/Pico/Faial/Flores/Madeira/Porto Santo/Portugal/Espanha/Itália/Grécia/Reino Unido	50	[F6*; F23*; FL1*; GR18*; GR22*; GRE5; GRE6; T172*; T175*; T198*; T258*; T273*; T403*; IT1; IT7; IT10; IT12; IT13; IT14; IT15; IT16; IT17; IT18; IT19; IT23; M10.2*; M20.2*; M22.4*; M30_10*; M34.3*; M43.2*; M51.5*; M66.6*; P10*; P18*; P19*; P25*; P35*; SM109*; SM181*; SM188*; SM501*; Setúbal35*; SMA2*; SMA16*; SMA32*; SMA35*; SP1; UK1; UK2]	Gerald <i>et al.</i> 2008; Presente estudo*
Ocr1_3:	Faial/Santa Luzia/ Itália	3	[F7*; IT2; SL3_14_ii*]	Gerald <i>et al.</i> 2008; Presente estudo*
Ocr1_4:	Portugal/Espanha	4	[FigueiraFoz57*; SP6; VianaCastelo64*; VilaNovaMilFontes51*]	Gerald <i>et al.</i> 2008; Presente estudo*

Haplótipo	Ocorrência	N	ID Amostra (Ind1;Ind2;...)	Referência
Ocrl_5:	Grécia/Itália	4	[GRE1; GRE2; GRE4; IT6]	Geraldes <i>et al.</i> 2008
Ocrl_6:	Grécia	1	[GRE3]	
Ocrl_7:	Israel	6	[IS1; IS3; IS4 IS5; IS9b; IS16a]	
Ocrl_8:	Israel	2	[IS7; IS12]	
Ocrl_9:	Israel/Santa Luzia	4	[IS9a; IS14; IS18; SL1_17_i*]	Geraldes <i>et al.</i> 2008; Presente estudo*
Ocrl_10:	Israel	1	[IS16b]	Geraldes <i>et al.</i> 2008
Ocrl_11:	Israel	1	[IS17]	
Ocrl_12:	Israel/Santa Luzia	2	[IT3; SL1_17_ii*]	Geraldes <i>et al.</i> 2008; Presente estudo*
Ocrl_13:	Itália	1	[IT5]	Geraldes <i>et al.</i> 2008
Ocrl_14:	Santa Maria/Itália	3	[IT8; SMA24*;SMA36*]	Geraldes <i>et al.</i> 2008; Presente estudo*
Ocrl_15:	Itália	1	[IT9]	Geraldes <i>et al.</i> 2008
Ocrl_16:	Itália	1	[IT20]	
Ocrl_17:	Madeira	1	[M31.4*]	Presente estudo*
Ocrl_18:	Pico	1	[P1]	
Ocrl_19:	Santa Maria	1	[SMA34*]	
Ocrl_20:	Reino Unido	1	[UK3]	Geraldes <i>et al.</i> 2008
Ocrl_21:	Santa Luzia	2	[SL1_25*; SL58_17*]	Presente estudo*
Ocrl_22:	Santa Luzia	1	[SL3_14_i*]	